2-1-1

2-1-1:　微弱周期信号と動的同期加算を用いたサンプリング周波数変動下でのインパルス応答推定

Estimation of impulse response using dynamic synchronous averaging and weak periodic signal under sampling frequency variation

☆住吉 亨介, 山岡 洸瑛, 若林 佑幸, 小野 順貴 (都立大)

- この論文では，既知の周期信号を長時間録音することで雑音環境下のインパルス応答を推定する新しい枠組みを提案する．
- 同期加算を適用することで雑音下でも周期信号を強調することができる．しかし長時間録音においては録音デバイスのサンプリング周波数が時間的に変動し，同期加算の同期性を劣化させ，強調性能も低下させる．
- 本研究では，周期ごとに相互相関を用いて観測信号と再生した周期信号との時間差を推定し，時間差を動的に補償する同期加算を提案する．
- また，時間差推定の精度を向上させるための反復的な参照信号の更新を提案する．
- シミュレーション実験によって提案法が周期信号を強調し (Fig. 1参照)，高い精度でインパルス応答を推定できること (Fig. 2参照) を確認した．

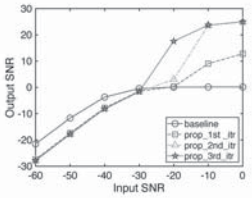


Fig. 1: Relationship between input and output SNRs with proposed method and baseline.

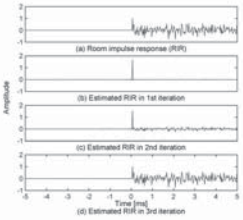


Fig. 2: Change in estimated impulse response at 1st, 2nd, and 3rd iterations when the input SNR is −20 dB.

2-1-2

2-1-2　アトミックノルム最小化によるスパース時間周波数表現を利用したノイズ除去

Denoising by sparse time-frequency representation via atomic norm minimization

◎草野翼, 矢田部浩平, 及川靖広(早大理工)

◆多くの音響信号は時間周波数領域でスパースな構造を持つので，時間周波数領域での処理が広く行われる．

◆スパース性を利用したノイズ除去では ℓ_1 ノルムを用いたスパース最適化が用いられているが，連続パラメータをグリッド状に離散化する必要があり，このグリッドの不一致の影響でスパースな解が得られない恐れがある．

◆このグリッドの不一致の影響を回避するために，我々はアトミックノルムを用いた周波数方向にグリッドレスなスパース時間周波数表現の推定手法を提案してきた．

◆本研究では提案手法を音声信号のノイズ除去に適用し，従来の ℓ_1 ノルムを用いた手法よりも局在化した時間周波数表現が得られることを確認した（Fig. 1）．

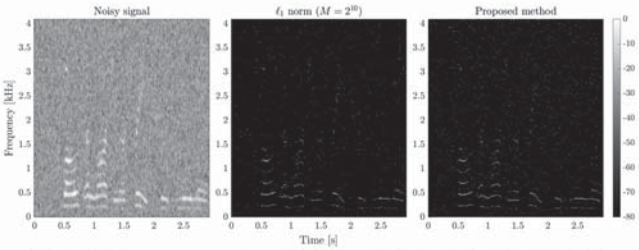


Fig. 1　Time-frequency representations of a noisy speech signal, a denoised signal using ℓ_1-norm, and a denoised signal using the proposed method.

2-1-3

2-1-3　時間周波数線形結合ビームフォーマの空間フィルタ数に対する音源強調性能の評価

Performance evaluation of time-frequency-bin-wise linear combination beamformer with respect to the number of spatial filters

◎山岡洸瑛, 小野順貴（都立大）

◆これまでに，劣決定条件下における音源強調手法として，時間周波数線形結合ビームフォーマ（TFLC）を提案してきた．

➢ TFLC は，それぞれ異なる雑音の組を抑圧する複数の空間フィルタの線形結合により音源強調を行うことで，劣決定条件下でも，目的音源を歪ませることなく，高い強調性能を達成する．

◆これまで，TFLC で使用する空間フィルタの数 K は，音源数 N とマイク数 M に対して，$K = {}_{N-1}C_{M-1}$ に固定されていた．

➢ K はユーザが指定可能なパラメータであり，提案法の性能は K の設定に依存すると考えられる．

◆本発表では，フィルタ数に対する提案法の性能を評価し，報告する．

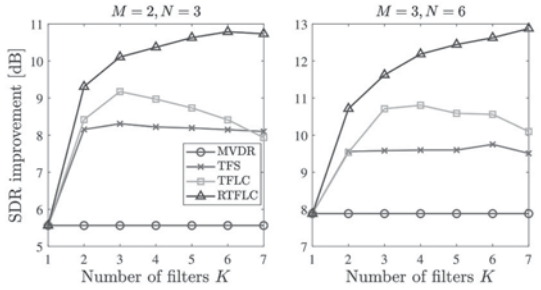


Fig. 1: Results of speech enhancement. TFS: our previous method, RTFLC: algorithmic variant of TFLC

2-1-4

2-1-4　要素選択による低演算量化を用いたDNN マスク推定に基づく音声強調処理

Time-frequency mask-based speech enhancement by low computational DNN with element selection

◎春田智穂, 小野順貴（都立大）

◆難聴者の雑音下での言葉の聞き取りに対して、**DNN** を用いた**時間周波数マスク**による音声強調処理が高い効果をあげており**補聴器**への応用が期待されている。

◆一方で補聴器には演算量の制約があり、乗算の回数が多いと大きな計算負荷となる。よって大規模な DNN は搭載することができない。

◆本研究では、同処理で発生する**乗算の回数を低減**するため、**要素選択による次元削減**を行ってから DNN で時間周波数マスクの推定を行い音声強調を行う方法を検討する。

◆次元削減は元のデータからの要素の選択のみで行われるため、新たな乗算は発生しない。

◆提案法により次元削減を行った場合、次元削減を行わなかった場合と比較し、乗算回数を削減しつつ高い音声強調性能が得られた。

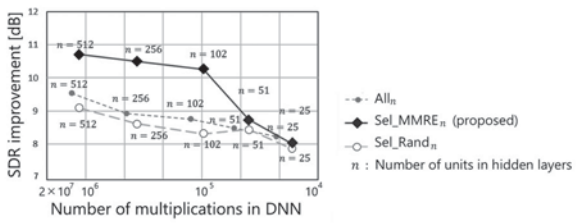


Fig.1 : The relationship between the number of multiplications and speech enhancement performance.

2-1-5

2-1-5 スイッチング畳み込みビームフォーマ

Improved Switching Convolutional Beamformer

○中谷智広，池下林太郎，△加茂直之，木下慶介，荒木章子，澤田宏（NTT）

◆本稿では，少ない数 (2～3 個) のマイクで効果的に雑音抑圧・残響抑圧・音源分離を同時実現する方法について議論する．

◆従来，十分な数のマイクがある場合，畳み込みビームフォーマ (CBF) を重み付き最小パワー無歪み応答 (wMPDR) 基準で最適化した wMPDR CBF が，効果的であると報告されている．しかし，マイクが少なくなると十分な性能が得られないという問題があった．

◆一方，山岡らは，劣決定条件（音源数＞マイク数）下で目的音を効果的に推定する最小分散無歪み応答スイッチング BF (MVDR swBF) を提案している．また，池下らは，WPE に時変スイッチを導入した swWPE を提案し，大幅な性能改善が得られることを示している．

◆そこで，本稿では，wMPDR CBF に時変スイッチを導入した wMPDR スイッチング CBF (swCBF) を構築し，少数のマイクを用いて効果的に目的音を推定する方法を提案する．

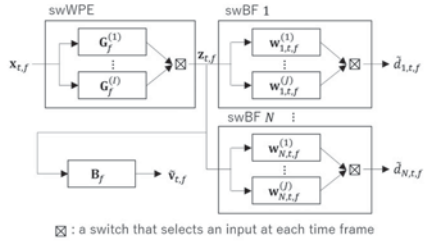


Fig. 1: Processing flow of wMPDR swCBF

2-2-1

2-2-1 回転楕円体アンビソニックス

Spheroidal ambisonics

☆金子昌賢，△Ramani Duraiswami，(University of Maryland, College Park)

◆立体音場の収音・再現において広く用いられている立体音場の球面調和領域表現に立脚したアンビソニックスを，球座標から回転楕円体座標系に一般化した回転楕円体アンビソニックスを提案する．

◆また，すでに普及しつつある従来の球基底方式との互換性を確保するために，回転楕円体アンビソニックスから球基底アンビソニックスへの解析的変換式を導出した．

◆提案法により回転楕円体型マイクロホンアレイを用いてアンビソニックス形式の音場収音が可能になる (Fig. 1).

◆数値実験により扁長回転楕円体アレイを用いた場合の再現領域が回転楕円体の長軸方向に伸長した形を有することが確認された．

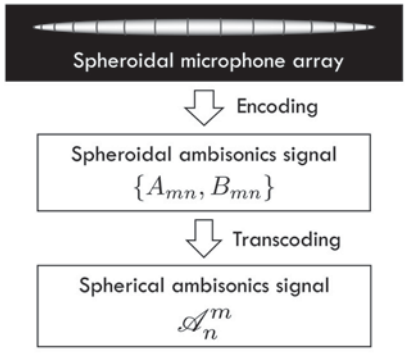


Fig. 1: The overview of spheroidal ambisonics.

2-2-2

2-2-2 多重散乱アンビソニックス: 剛球マイクロホンアレイ群による多重散乱を考慮した多自由度立体音場収音

Three-dimensional sound field estimation using a grid of interacting rigid spherical microphone arrays for higher degrees-of-freedom spatial audio

☆金子昌賢，△Ramani Duraiswami，(University of Maryland, College Park)

◆剛球マイクロホンアレイによるアンビソニックス方式に基づく立体音場収音は広く用いられているが，音場再現時のスウィートスポットの狭さが課題である．本研究では、多重散乱によるアレイ間相互作用を考慮した剛球アレイ群を用いた三次元音場のアンビソニックス収音のための方法である多重散乱アンビソニックスを提案する．

◆数値実験により，提案法により密配置の剛球アレイ群においても再現時のスウィートスポットの広い音場表現を取得できることが確認された．提案法は多自由度立体音響への応用が期待される．

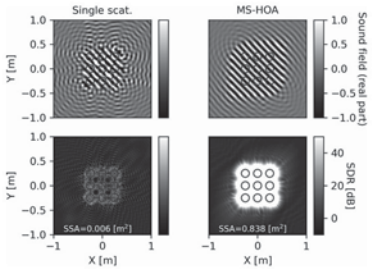


Fig. 1: The estimated incident fields and SDR maps using a grid of nine interacting rigid spherical microphone arrays, for an incident field of 4kHz. Results for the single-scattering case (left) and MS-HOA (right), respectively.

2-2-3

2-2-3 マルチカーネル学習に基づく音場推定の検討

Sound field estimation based on multiple kernel learning

☆堀内亮佑，小山翔一，Juliano G. C. Ribeiro（東大院），
植野夏樹（都立大），猿渡洋（東大院）

◆空間的に分散配置されたマイクロフォンで得られる観測信号を元に，音源が存在しない領域内の連続的な音場を補間する音場推定の手法に関して検討を行う．

◆本研究では，音源方向の事前情報を組み入れることが可能な方向重み付けカーネルを用いる手法において，事前情報がない場合でもそのパラメタを最適化する手法を提案する．

◆提案法において，カーネルを複数のサブカーネルの重み付き和でモデル化した．その際，重みに制約を課す必要があり，L1 制約と，それを緩和した L2 制約に関して検討を行った．

◆本実験において，以下の図に示すように，提案手法，特に，L1 制約を用いた場合に推定精度が大きく改善した．

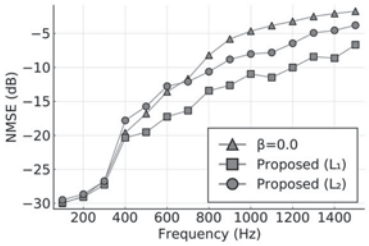


Fig. 1: Experimental results of the baseline and the proposed methods. NMSEs are plotted against frequency.

2-2-4

2-2-4 Region-to-region interpolation of the acoustic transfer function: Directional weighting and its parameter optimization

Juliano G. C. Ribeiro, Ryousuke Horiuchi, Shoichi Koyama, and Hiroshi Saruwatari (The University of Tokyo)

◆We propose a method of region-to-region interpolation for the acoustic transfer function (ATF) using kernel ridge regression (KRR) with a kernel fitted for the data.

◆In a previous work, we have shown that a region-to-region KRR approach using the acoustic properties of the ATF was possible.

◆We now expand the previous definition of the kernel function to include the directionality of the problem.

◆The model is tuned to the data by optimizing a leave-one-out loss criterion using a gradient descent method.

◆This method was compared to the previously established model through numerical simulations.

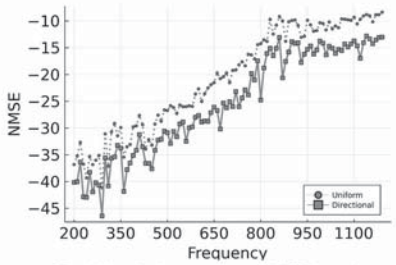


Fig.1: Normalized mean square error (NMSE) comparison.

2-2-5

2-2-5 仮想的焦点形成と瞬時スペクトル推定法による同一到来方向の音源分離に関する考察

Consideration on separation of multiple sound sources in the same direction by instantaneous spectral estimation with virtual focusing

○小澤賢司（山梨大・工），渡邉貫治（秋田県立大・システム科技），坂本修一（東北大・通研）

◆マイクロホンアレイの出力信号に関する 2 次元スペクトルについて連立方程式を解くことで，複数音源のスペクトルを推定する。これを，分析の時間フレームごと独立に行うため，瞬時スペクトル推定法と称している。

◆同一方向にあり距離が異なる複数音源の信号分離を，前報では円弧状アレイを用いることで焦点位置を固定して実施したが，本稿では直線状アレイで収録した音に位相回転を施すことで仮想的に焦点を設けて実施した。4 マイクロホンを用いて 4 音源の分離を試みた。

◆4 音源をアレイから 1 m 間隔で配置し，仮想焦点位置を推移させて音源分離性能を調べる計算機実験を行った。Fig. 1 に信号対歪比 (SDR) を示すとおり，焦点の位置に依らず良好な分離を達成した（誤差棒は 4 種の音源信号を交換して配置したことによる標準偏差）。音源位置による SDR の違いは，振幅スペクトルの大きさに対応している。

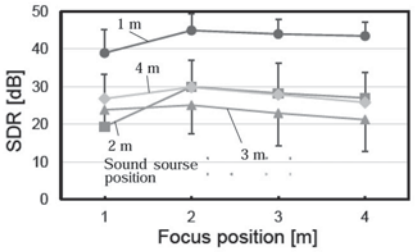


Fig.1 : Effects of the focus position on the performance of sound source separation.

2-2-6

2-2-6 IEC のオーディオ標準化動向

The standardization trends of audio in IEC

○由雄淳一（音響芸術専門学校）

◆IEC (International Electrotechnical Commission) は電気電子分野の国際標準を策定しており、その規格策定委員会 (Technical Committee) は分野ごとに 110 の TC がありその中の TC 100 はタイトルを Audio, video and multimedia systems and equipment としオーディオ分野の標準化を行っている。

◆その歴史は古く、初期の TC 12: Radiocommunications、TC 60: Recording からオーディオ分野の標準化を行っており、1995 年にそれら TC などが統合し TC 100 が発足してオーディオ分野の標準化を行い今に至っている。

◆このように初期よりオーディオの基本的な標準を策定しつつ、現在では技術の進歩によりオーディオの対応分野は、3D オーディオ、ハイレゾ、Internet、IP、IT、Cyber Physical System、Cloud また AAL (Active Assisted Living) に及ぶ。

◆本講演では TC 100 でのオーディオ分野の標準化について 1980 年代より IEC で規格策定を行い TC 100 下で Project Leader、議長、幹事を務めてきた筆者がその標準化と今後の動向について解説する。

2-2-7

2-2-7 ITU-R SG6 における高臨場感オーディオに関する標準化の動向

Trend of international standardization on high-reality audio in ITU-R SG6

○大出訓史（NHK）

◆国際電気通信連合 無線通信部門（ITU-R）では，無線通信技術に関する国際規則や勧告を策定しており，放送サービスに関する研究委員会（SG6）では，次世代音声サービスに対応する先進的音響システムの勧告を策定した。

◆先進的音響システムは，5.1 サラウンドを上回るチャンネルベース音響，もしくは音響メタデータを用いたチャネルベース音響，オブジェクトベース音響，シーンベース音響とこれらの組み合わせによって構成される。

◆オブジェクトベース音響は音声信号と音響メタデータで構成され，レンダラーによって再生するスピーカ配置に合わせた信号に変換されるため，スピーカ配置を自由に選択でき，音声オブジェクトの組合せや音量調整など，ユーザが番組音声をカスタマイズできる。

• ITU-R SG6 では，これまでにスピーカ配置，音響メタデータ，音声ファイル形式やレンダラーなどの勧告を策定しており，欧米諸国では実用化が進んでいる。

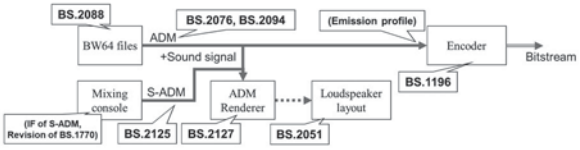


Fig.1 : ITU-R Recommendations related to Advanced Sound System.

2-2-8

2-2-8　ITU-WHO の Safe Listening に関する新国際標準について

On the new international standard on safe-listening

○川森雅仁（慶大）、 Shelly Chadha (WHO)

◆WHO の発表によれば 12 歳～35 歳の全世界人口のうち 4300 万人が難聴で、さらに約１１億人が難聴になる危険性があるとされている。特に世界的な若年難聴者の増加が問題視されている。

◆この状況を受けて WHO は、2015 年に Make Listening Safe キャンペーンを開始し、安全な音楽視聴方法を啓発すると共に、2015 年 10 月から、同じ国連組織である ITU (世界電気通信連合)と共同でイヤホンなどで視聴する、スマートホンや MP3 プレーヤなどの個人用音楽視聴用携帯端末(Personal Audio System ; PAS)を対象にした標準化活動を開始し、2018 年に ITU-T Rec. H.870: ”Guidelines for safe listening devices/system” として勧告化した。WHO は同勧告を、WHO の新標準として、2019 年 2 月に採用し発表した。

◆この標準は PAS 音量の安全基準値を、エネルギー等価原理（Equal Energy Principle）とそれに基づく音曝露量(Dose)という概念によって規定している。また音曝露量の単位として **Pa^2h** (平方パスカル時)を新たに導入している。

◆この標準は一週間単位での音曝露量の安全敷居値を成人ための標準レベル）と子供などのための安全標準の２基準を推奨している。

◆この安全基準に基づいて、PAS メーカは、一週間の音曝露量をモニターし、適切な形でユーザに警告を与えたり、”スマート”な形で自動的に音量を安全なレベルまで下げるなどの実装することが期待されており、すでに複数のメーカ実装している。

2-2-9

2-2-9　MPEG Audio Coding における標準化動向

Standardization activities of MPEG Audio Coding

○杉本岳大（NHK）

◆高臨場感オーディオをスコープとした最新の MPEG 規格である MPEG-H 3D Audio (3DA)と MPEG-I Immersive Audio (IA)を中心に紹介する。

◆MPEG-H 3DA は，ITU-R で標準化された 3 次元に拡張されたスピーカ配置およびチャネルベース，オブジェクトベース，シーンベースの各音響システムに対応した符号化方式である。図 1 にデコーダの構成を示す。

◆MPEG-I IA は，VR や AR などのイマーシブメディア用のオーディオ技術である。MPEG-H 3DA が 3DoF のオーディオ技術であるのに対し，MPEG-I IA はリスナーの移動にも追随可能な 6DoF のオーディオ技術である。6DoF コンテンツではユーザごとに異なる映像・音響を体験できる点が，従来の 3DoF コンテンツとの違いである。図 2 に基本構成を示す。

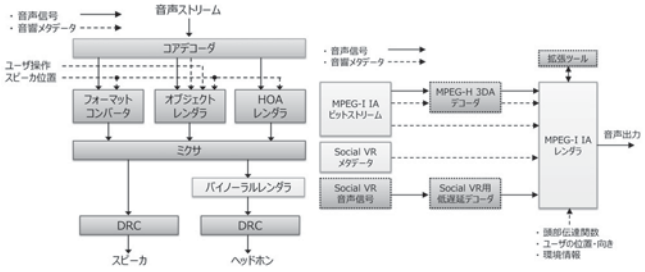


図 1　MPEG-H 3DA デコーダの構成

図 2　MPEG-I IA の基本構成

2-2-10

2-2-10　測定時間が異なる複数のラウドネス値を基準とした副音声制作におけるダイアログレベルの自動調整法

Dialogue level auto adjustment based on multiple loudness with individual measurement periods on the production of second audio program

◎久保弘樹，西口敏行，大出訓史（NHK）

◆効率的にオブジェクトベース音響による副音声制作を行うため、著者らはこれまでに、ITU-R 勧告であるラウドネス値に基づく客観指標を用いて主音声のダイアログレベル(MDL)に合わせることで自動的に副音声のダイアログレベル(SDL)を調整する手法を提案してきたが、従来法では主音声と副音声の発話内容が大きく異なるときに過剰な調整となる可能性があった。

◆本稿では、測定時間が異なる複数の客観指標を組み合わせて用いるダイアログレベルの自動調整法を提案する（図１）。提案法では、測定時間をそれぞれ 60 秒、20 秒、2 秒とした 3 種類の客観指標を測定し、MDL と SDL の客観指標の差分に主音声と副音声の発話タイミングの類似度によって求めた重みをかけて加算して調整値を求める。

◆DMOS 法による主観評価実験により提案法の効果を確認した。従来法を含め、客観指標の組み合わせ方を変えて比較した結果、提案法が比較した組み合わせの中では最も有効であった。

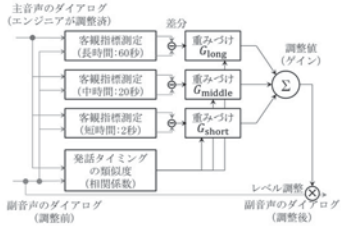


図1:ダイアログレベルの自動調整法

2-2-11

2-2-11　MPEG-H 3D Audio によるモノ信号の圧縮符号化

Coding of mono signal using MPEG-H 3D Audio

○杉本岳大（NHK）

◆オブジェクトベース音響システムの仕組みを活かした音声サービスであるダイアログエンハンスメントやセカンドデバイス再生では，オブジェクト（多くの場合モノ）が単体で再生される場合が想定される。

◆伝送路や放送装置の設計において，オブジェクトを単体で再生した場合でも符号化劣化が顕在化しないモノ信号のビットレートの値が必要になるが，これまでに 3DA に関する調査はなされていない。

◆12 種類のモノの音素材（人声 4 種類，楽器 4 種類，アンサンブル 4 種類）と，5 種のビットレート（24，32，40，48，64 kbit/s）を組み合わせて刺激を作成し，隠れ基準付き三刺激二重盲検法で主観評価した。

◆48 kbit/s では 12 種類の音素材中，11 種類が放送品質を満たし，64 kbit/s では 12 種類の音素材全てが放送品質を満たすことが明らかになった（表 1）。

音素材	ビットレート [kbit/s]				
	24	32	40	48	64
女声／英語				●	●
女声／日本語			●	●	●
男声／英語				●	●
男声／日本語			●	●	●
グロッケン				●	●
タンバリン		●	●	●	●
ウインドチャイム					●
三味線		●	●	●	●
バンド				●	●
長唄		●	●	●	●
ピアノトリオ				●	●
弦楽四重奏	●	●	●	●	●

表 1　放送品質を満たすモノ信号のビットレート

2-2-12

2-2-12　人声の 3 次元放射特性の推定法における検討

Estimation of three-dimensional radiation characteristics of Japanese speech

◎木下光太郎，杉本岳大，中山靖茂（NHK）

◆AR / VR などの視聴時に**音源の放射特性**を再現することで，音源の向きが自由に変化した場合に自然な可聴化を行う**ステアリング**を検討している。

◆放射特性データベースから任意の聴取位置における周波数特性を再現するためには，データベースの有限個の測定点から球面上の任意の収音・聴取位置での周波数特性を推定する必要がある。

◆本稿では Fig.1 に示す 4 種類の推定手法について推定誤差を算出し，**3 次元放射特性の推定法**を比較した。

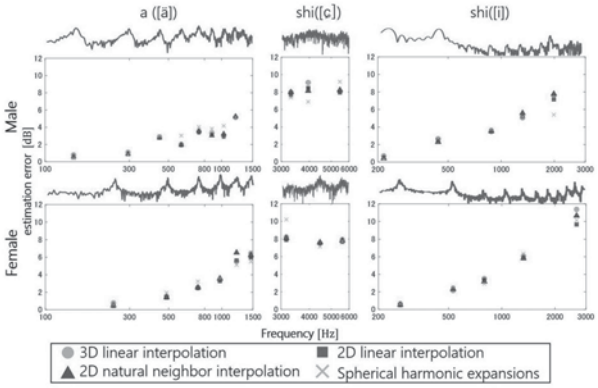


Fig.1 Comparison of estimation error

2-3-1

2-3-1　深層エネルギーベースモデルによる音声の音響特徴量の生成

Generation of acoustic feature of speech based on deep energy-based model

◎岸田拓也，中鹿亘（電通大）

- エネルギーベースモデルのエネルギー関数部分を深層ニューラルネットワーク化したモデルは、他の確率的生成モデルにはない特性を持つことから注目を集めている。
- 音声信号処理分野では、拡散確率モデルや雑音除去スコアマッチングを用いた関連手法がいくつか提案されているが、深層エネルギーベースモデルを直接用いた手法は提案されていない。
- 音声のメルケプストラムの生成確率を推定する深層エネルギーベースモデルを３種類の学習法で学習し、ノイズ除去課題でそれぞれのモデルの性能を評価した。

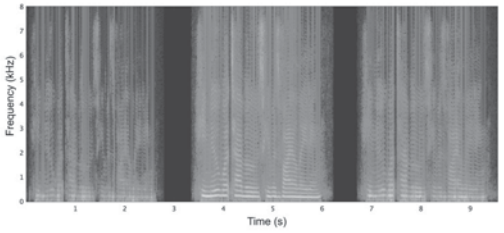


Figure: Spectrograms of perturbed (left), original (center), and denoised (right) speech by Simple EBM.

2-3-2

2-3-2　敵対的学習による統合型ソースフィルタネットワーク

Unified Source-Filter Network with Adversarial Learning

☆米山怜於，呉宜樵，戸田智基（名大）

本稿ではソースフィルタモデルを単一のニューラルネットワークで再現する Unified Source-Filter GAN（uSFGAN）を提案する．Quasi-Periodic Parallel WaveGAN の生成器を音源生成と声道フィルタを担う２つの連結したネットワークに分解するため，音源信号から抽出したスペクトル包絡に対する正則化項を補助損失に追加する．また，ネットワークが周期構造を推定し易くするための最小限の手掛かりとして，正弦波基底信号を補助入力に用いる．実験的評価から，提案手法は音質及びF_0制御性能において QPPWG 及び Neural Source-Filter を優位に上回ることを示す．また，図１より，音源のスペクトル包絡に関する正則化が妥当な音源出力に有効であることが分かる．

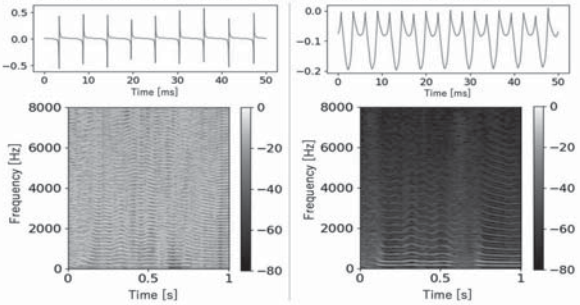


Fig.1 : Comparison of waveforms and spectrograms of output source signals from the source-networks of uSFGAN (left) and uSFGAN w/o spectral envelope regularization loss (right).

2-3-3

2-3-3　DNN歌声合成のための調子はずれ補正

Automatic pitch correction of tuneless singing for DNN-based singing voice synthesis

◎法野行哉，橋本佳，大浦圭一郎，南角吉彦，徳田恵一（名工大）

- **深層ニューラルネットワーク** (deep neural network; DNN) **に基づく歌声合成のための調子外れ補正を提案する．**
- **提案手法では，音高に関する事前分布と擬似音符音高を，** DNN **音響モデルの学習部に導入する．**
- DNN **歌声合成の学習データに調子外れな歌声波形が含まれている場合においても，調子外れが補正された歌声の合成が可能となり，提案手法の有効性を示した．**

Table 1: Systems for evaluation of pitch correction techniques.

Note pitch at the training	w/o prior	w/ prior
Original note pitch	Org	Org+Prior
Pseudo note pitch		
Heuristic	Heur	Heur+Prior
Pitch bias-based	Bias	Bias+Prior

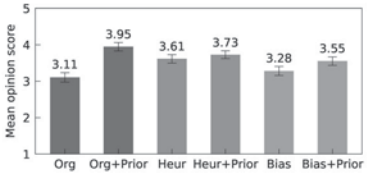


Fig. 1: Subjective evaluation results.

2-3-4

2-3-4　発声タイミングのずれを考慮した sequence-to-sequence 歌声合成

Sequence-to-sequence singing voice synthesis considering vocal timing fluctuation

◎法野行哉，橋本佳，大浦圭一郎，南角吉彦，徳田恵一
（名工大）

- 擬似音素境界を基準に生成した楽譜特徴量を入力系列として用い，アテンション機構により発声タイミングのずれを考慮するsequence-to-sequence (seq2seq) 歌声合成を提案する．
- 実験により，外部の発声タイミングモデル及び継続長モデルを用いることなく，自然な歌声の生成を実現し，提案法の有効性が示された．

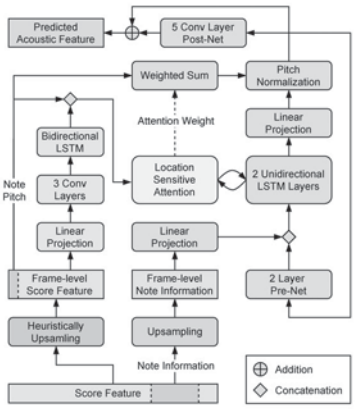


Fig. 1: The model architecture of proposed seq2seq singing voice synthesis system.

2-3-5

2-3-5　隠れセミマルコフモデルによる構造化アテンションを用いた自己回帰型 VAE に基づく sequence-to-sequence 音声合成

Autoregressive variational autoencoder-based sequence-to-sequence speech synthesis using a hidden semi-Markov model based structured attention mechanism

◎藤本崇人，橋本佳，南角吉彦，徳田恵一（名工大）

本稿では，HSMM による構造化アテンションを自己回帰型 VAE に取り入れた sequence-to-sequence 音声合成を提案する．言語特徴量と音響特徴量に対応する潜在変数間の対応関係を HSMM 構造化アテンションでモデル化する．また，自己回帰型 VAE の枠組みにより自然性を向上させる．自然性に関する評価実験と継続長を変更した音声合成実験を行い，提案法の有効性を示した．

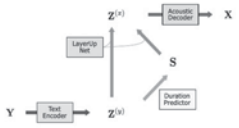


Fig. 1: 生成モデルの概要

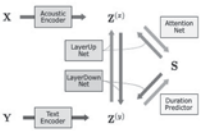


Fig. 2: 推論モデルの概要

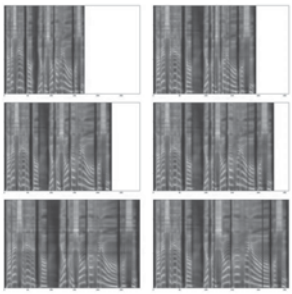


Fig. 3: 継続長の制御

2-3-6

2-3-6　音声区間検出不要の実用的なストリーミング End-to-End 音声認識

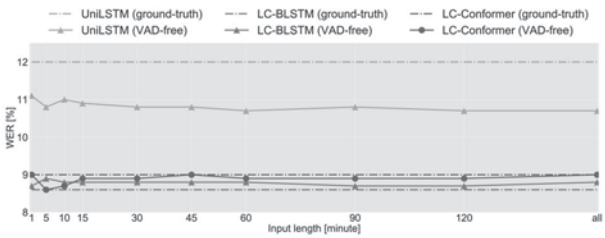
VAD-free Practical Streaming End-to-End ASR

◎稲熊寛文，河原達也（京都大）

◆注意機構型音声認識モデルは CTC や RNN トランスデューサなどに比べて，長い音声系列に対して頑健ではない

◆本稿では，monotonic chunkwise attention (MoChA) をベースとし，CTC 確率を用いた音声区間検出不要のストリーミング推論アルゴリズムを提案

◆提案手法により，精度を維持したまま数時間の長さの音声をストリーミング認識可能



2-3-7

2-3-7　How language similarity affects the Mongolian ASR using cross-lingual transfer learning

Narangerel Purevdorj, Ryota Nishimura (Tokushima University)
Altangerel Ayush (Mongolian University of Science and Technology)
Kengo Ohta (National Institute of Technology, Anan College)
Norihide Kitaoka (Toyohashi University of Technology)

In this paper, we investigate how the phoneme similarity of languages affects the low-resource ASR.
We reveal the similarity of the two different languages (target and training set). It is essential to measure similarities in many ways. Therefore, the following three characteristics were investigated.

1. Data Ratio (DR) - ratio of the amount of data between target and training language.
2. Shared Phoneme Ratio (SUR) - ratios of shared and unique phonemes among each of target and training set.
3. Phoneme Similarity (PS) - similarities of most frequent shared diphonemes in the target and training set.

Table 1. Experiment results for mono-lingual and cross-lingual ASR

Dataset	MN	MN + EN	MN + JP	MN + TR
CER	19.3	11.0	10.7	10.0

Table 2. Correlations of CER prediction and experiment (CER-cross) results

	DR	SUR	PS	CER
DR	1			
SUR	0.4597	1		
PS	0.5278	0.9969	1	
CER	-0.9345	-0.7457	-0.7955	1

Table 1 shows the monolingual and cross-lingual ASR results. Table 2 shows the correlations among our proposed similarity measures and CERs. Between each proposed measure and the CER, the values are inversely well correlated.
This paper presented to determine whether language similarity affects the low-resource ASR using cross-lingual transfer learning.

2-3-8

2-3-8 半教師あり学習に基づくリッチトランスクリプション型 End-to-End 音声認識の検討

Rich Transcription-style End-to-End Automatic Speech Recognition Based on Semi-supervised Learning

◎田中 智大, 増村 亮, 庵 愛, 高島 瑛彦, 折橋 翔太, 牧島 直輝
（日本電信電話株式会社）

◆人間の発話には，笑いや言い淀み，咳など様々な現象が含まれ，音声認識精度を低下させることが想定される

◆End-to-End 音声認識の枠組みを利用して，文字や単語とそのような現象を同時推定するリッチトランスクリプション音声認識（RT 音声認識）が実現可能

◆しかし，現象をラベリングすることは非常にコストが大きいため，高精度に RT 音声認識を実現するのは困難

◆本稿では，擬似ラベリングを利用した半教師あり学習に基づき，擬似ラベル付きデータを作成して利用

◆現象ラベル付きデータ、ラベルなしデータの両者を利用するために，ラベル付きか否か（スタイルトークン）としてモデルに与えて学習

Figure: Character error rate (%) on evaluation set. C: unlabeled data, R: labeled data and PR: pseudo labeled data.

システム	半教師あり学習	学習データ	スタイルトークン	%CER w/o 現象	%CER w/ 現象
音声認識	-	C	-	24.7	31.2
	-	C+R	-	23.5	34.1
RT 音声認識	no	R	no	46.7	48.2
	no	C+R	yes	23.2	26.8
	yes	C+R+PR	yes	**22.7**	**26.2**

2-3-9

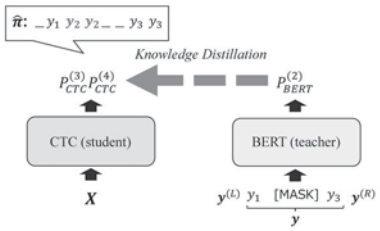
2-3-9 BERT による CTC 音声認識への知識蒸留

Knowledge Distillation from BERT to CTC-based ASR

☆二見颯, 稲熊寛文, 三村正人, 河原達也(京大)

◆BERT を teacher モデル, CTC 音声認識モデルを student モデルとした知識蒸留法 (Knowledge Distillation) を提案.

- BERT に学習されたテキストデータの知識の利用.
- CTC によるフレーム単位の予測と BERT による単語単位の予測の対応をとるため, 最尤の CTC パス $\hat{\pi}$ を利用.



◆推論速度を維持したまま, WER を改善できた.

- さらに, Shallow Fusion, Rescoring を適用する効果も得られた.

Table 1. ASR Results on CSJ

	WER(%)	Runtime
CTC baseline	9.29	×1
+Shallow Fusion (TLM)	8.57	×18.4
+20-best Rescoring (TLM)	8.82	×20.2
+20-best Rescoring (BERT)	8.57	×20.5
CTC w/Knowledge Distillation (BERT)	9.01	×1
+Shallow Fusion (TLM)	8.44	×18.4
+20-best Rescoring (TLM)	8.36	×20.2
+20-best Rescoring (BERT)	**8.15**	×20.5

2-3-10

2-3-10 Hybrid RNN-T/Attention 構造を用いたストリーミング型 End-to-End 音声認識の検討

A Study on Hybrid RNN-T/Attention Architecture for Streaming End-to-End Speech Recognition Systems

◎森谷崇史, 田中智大, 芦原孝典, 落合翼, 佐藤宏, 安藤厚志
増村亮, デルクロアマーク, 浅見太一（日本電信電話株式会社）

◆先行研究では Hybrid CTC/Attention 構造の CTC を RNN-T に置き換えたモデル構造を提案

◆Attention モデル（S2S）はオフラインを前提としているため，RNN-T と同時にストリーミング動作をさせることが困難

◆本研究では，Triggered Attention（TA）を応用することでストリーミング動作可能な Hybrid RNN-T/Attention 構造を提案

◆ストリーミング動作の制約を維持ながら RNN-T の認識仮説を，TA を応用した S2S（S2S+TA）でスコアリングすることで RNN-T 単体の認識性能よりも改善することを確認

Table 1 : Results of each RNN-T-based System.

Encoder	System	Switchboard		TED-LIUM2	NVS
		SWBD	CH		
ULSTM	RNN-T	20.8	35.8	18.5	12.2
	+S2S+TA	**19.6**	**35.0**	**17.5**	**11.2**
LCBLSTM	RNN-T	16.7	29.4	14.3	7.1
	+S2S+TA	**15.9**	**28.5**	**13.6**	**6.8**

2-3-11

2-3-11 ゲーム場面の笑い声・叫び声に対する音響的識別可能性の検証

Acoustic discriminability of laughter and scream unconsciously yielded during game-play

☆松田匠翔, 有本泰子(千葉工大)

◆社会的シグナルである笑い声と叫び声は同様の音響的傾向を持っており，その識別可能性の検証が依然として行われていない。

◆笑い声，叫び声が音響的に識別できるか識別実験を行い検証した。

◆音響的特徴量を入力，笑い声，叫び声，それ以外の発話の３種類を含む発話タイプを出力とした DNN，SVM の識別モデルを構築した。

◆Accuracy が DNN：93.5%，SVM：95.5%と高い値を示した。これは笑い声，叫び声，発話の間で傾向が異なる音響的特徴があると言える。

◆笑い声，叫び声，発話の音響的特徴量のヒストグラムを視察した結果，HNR が笑い声と叫び声を識別できる音響的特徴量の１つと言える。

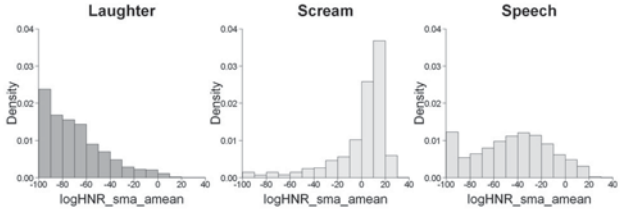


Fig.1 : Distribution of Harmonics-to-Noise Ratio in laughter, scream, and speech

2-3-12

2-3-12 話者混在音声からの fo 軌跡の分離

fo contour separation from mixed speech

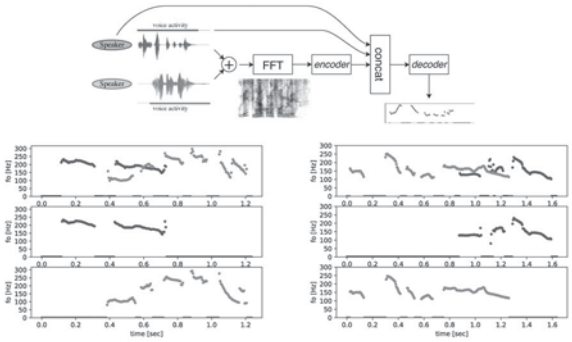
○森 大毅 (宇都宮大)

◆日常的な生活環境で収録した話者混在音声から fo 情報を話者ごとに分離する技術の確立を目指している

◆提案するニューラルネットは、話者混在音声のスペクトログラム、対象話者の埋め込みベクトル、対象話者の発話区間を入力とし、対象話者の fo 軌跡の推定結果を出力する

◆訓練時は、２話者のクリーン音声を人工的に重畳させることで話者混在音声を模擬し、元音声から推定した fo 軌跡を教師信号として損失を評価しネットワークを学習する

◆実験の結果、提案するモデルは話者混在音声から各話者の fo を高精度に推定できることがわかった



2-3-13

2-3-13 プリエンファシスの適用回数が IAIF に基づく声門流推定の精度に与える影響

Effect of the number of times pre-emphasis is applied on the accuracy of glottal flow estimation based on IAIF

☆平野雄大，モクタリ・パーハム，森川大輔（富山県立大）

◆Iterative Adaptive Inverse Filtering (IAIF) での正確な声門流の推定にはプリエンファシス適用回数によって制御される音声のスペクトル傾斜が重要であることが示唆されている.

◆プリエンファシスの適用回数を変化させて声門流を推定し，誤差が最小となる際 (Best) の総スペクトル傾斜 ST_{total} を算出することで声門流の推定に最適なプリエンファシスの適用回数を調べた.

◆標準の IAIF (プリエンファシス１回) に比べて複数回のプリエンファシスは声門流推定精度を向上させ，IAIF の改善手法である IOP によって最適な ST_{total} の上限を決定できることがわかった.

◆最適な ST_{total} は，声質が弱く気息性から強く圧迫した発声になるにつれて小さくなり，弱く気息性な発声ほど最適な ST_{total} は大きくなることがわかった.

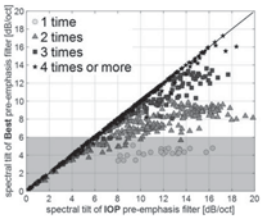


Fig. 1 Relation between ST_{total} yielded by IOP vs Best

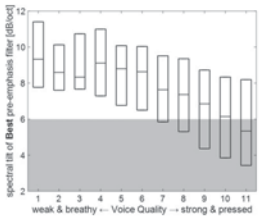


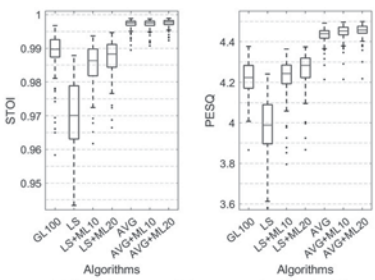
Fig. 2 Distribution of Best ST_{total} depending on Voice Quality

2-3-14

2-3-14 Maximum likelihood estimation for phase reconstruction from its derivatives

☆ Nguyen Binh Thien (Ritsumeikan Univ.), Yukoh Wakabayashi (Tokyo Metropolitan Univ), Kenta Iwai and Takanobu Nishiura (Ritsumeikan Univ.)

This paper presents an improvement to a two-stage algorithm estimating the phase from only the amplitude. Instead of directly estimating the phase, the two-stage algorithm estimates phase derivatives, i.e., instantaneous frequency (IF) and group delay (GD), by using deep neural networks, and it then reconstructs the phase from those derivatives using a least-squares (LS) method. A problem is that the periodicity of the IF and GD significantly affects the results of LS estimation. Tackling this problem, in this paper, we propose a maximum likelihood estimation using *von Mises* distribution for reconstructing the phase from the IF and GD. The error function is minimized using the Adagrad algorithm, in which results of the conventional methods can be used as initializations. Experimental results confirm the effectiveness of the proposed method in improving results of conventional methods.



Algorithms:
- GL100: [conv.] Griffin-Lim method with 100 loops
- LS: [conv.] Least-squares method
- AVG: [conv.] Average method
- XX+MLyy: proposed method where XX and yy are replaced by the conv. method used for initialization and number of optimization loops

Fig.1: Performances of phase reconstruction algorithms

2-4-1

2-4-1 情報理論に基づく蝸牛モデルの評価法の提案

Information theory-based approach for evaluation of cochlear models

○村上泰樹（九州大学芸術工学研究院）

◆背景と目的
- 蝸牛モデルの評価は、線形システムを前提として行われてきた。
- 蝸牛モデルは非線形であるため、評価で必要とされる前提を満たしていない。
- 本研究では、非線形システムへも対応可能な情報理論を基盤とした新たな蝸牛モデルの評価法の確立を目指す。

◆方法と結果
- 蝸牛モデル（生理モデルと機能モデル）から情報量を計算（図１）
- 圧縮特性が得られる条件で情報量が入力よりも大きくなる。

◆考察
- 圧縮特性は、蝸牛の応答特性を複雑にすることで情報量を増加させる働きが示唆される。

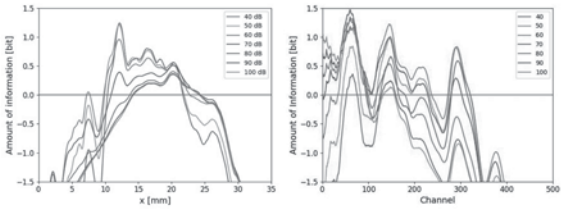


Fig.1 : Amount of Information obtained from models
(Left) Physiological-motivated model. (Right) Functional model

2-4-2

2-4-2　基底膜を含む蝸牛構造への蝸牛定在波モデルの適用に関する検討

Discussion on relationship between traveling-wave model and cochlea standing wave model

☆大野　夏凜，堀井　康史（関西大）

◆Békésy は基底膜の分散性により、高周波音は蝸牛の基部近くに、低周波音は頂部近くに進行波が発生し、これを外有毛細胞が知覚すると考えて「進行波説」を唱えた。著者らは蝸牛のマクロ／ミクロな構造から類推して「蝸牛定在波理論」と「外有毛細胞共鳴説」を提唱した。本報告では全く異なる２つのアプローチの関連性について検討する。

◆図１は 1950 年代に提案された蝸牛の等価回路モデルであり、基底膜の物理定数が $Z_{CF}(x)$に集約されている。図２は著者らの理論に基づいて構成した回路で、外有毛細胞共鳴が LC 直列共鳴回路で表現されている。これらの等価回路は本質的に同じものであり、共通項が多い。

◆同様な回路構成で音波（速波）から進行波（遅波）の励振を説明しようとしていた 1950 年代の研究者に共感を覚える。生物の細胞は常温で 3nm 程度の熱ゆらぎがあるといわれている一方で、近年のシミュレーションでは、基底膜の初期変位は熱ゆらぎと同程度か、それよりも小さく見積もられることが多い。聴覚メカニズムに関しては、今後も様々な角度から検討を進める必要があると著者らは考えている。

Fig.1：Traveling-wave Model

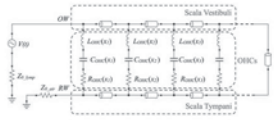


Fig. 2:Cochlea standing wave model

2-4-3

基底膜を含む蝸牛モデルを用いた内耳気腫による難聴の発生メカニズムに関する検討

Simulation study on mechanism of hearing loss due to presence of air bubbles in cochlea

☆佐藤　圭祐，堀井　康史（関西大）

◆蝸牛の前庭階に気泡が生じると聴力が著しく落ちるが、鼓室階に生じてもその影響は比較的小さいという。本稿では、外リンパ液を圧縮性媒質に設定した直線状蝸牛モデルの前庭階、あるいは鼓室階の、蝸牛基部から 10 mm、20 mm、30 mm の位置に 1 mm^3 の気泡を配置した６種類のモデルと、気泡を含まない正常モデルを Multiphysics を用いて作成し、基底膜の振動状態を評価することで聴力特性を検討した。

◆図２は 2000 Hz の純音を卵円窓より入力したときの、蝸牛位置に対する基底膜の最大変位量を表したもので、進行波のエンベロープに相当する。図２左は気泡が前庭階に、右は鼓室階にあるときの特性で、破線は気泡を含まないときの特性である。グラフより、蝸牛のトノトピーにおいて、2000 Hz に対応する点よりも基部側に気泡があるとき（図２左、La=10mm 時）に聴力特性が著しく低下することがわかる。

◆この結果は、前庭階に気泡が入ると気泡位置に応じて低周波域で難聴をきたす可能性があること（図２左）、また鼓室階に気泡が入っても聴力に大きな影響を及ぼさないこと（図２右）を示唆している。

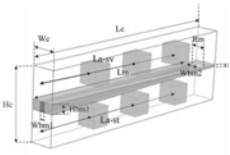


Fig.1：Cochlea model

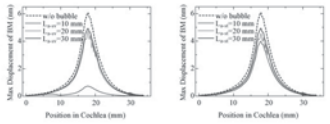


Fig.2：Maximum displacement of basilar membrane

2-4-4

2-4-4　圧縮性リンパ液を想定した蝸牛モデルにおける進行波の励振メカニズムに関する検討

Examination of traveling wave excitation mechanism in cochlea model based on compressible lymph

☆洪　文甲，堀井　康史（関西大）

◆1950 年代，蝸牛等価回路モデルに関する論文では，前庭階，鼓室階の外リンパ液は圧縮性媒質として扱われ，特に 7000 Hz 以上の周波数ではこの取り扱いが必要であることも指摘された。しかし，近年の研究では，速波が進行波を生み出すメカニズムについて十分な検討がなされないまま、外リンパ液を非圧縮媒質として近似的に扱うことが習慣となった。

◆本稿では、外リンパ液を圧縮性媒質に設定し，蝸牛構造によって生じる聴覚生理を取り扱える蝸牛モデルを新たに構築し，偶奇モード理論の観点から進行波の生成メカニズムを明らかにする。Fig.1 に外リンパ液を圧縮性媒質とする直線形状の蝸牛モデルを示す。Fig.2 に示すように，基底膜上を伝搬する進行波は，蝸牛の前庭階と鼓室階に生じる速波の音圧差によって励振され（奇モード励振），音圧差が生じないとき，進行波は励振されない（偶モード励振）ことを確認した。

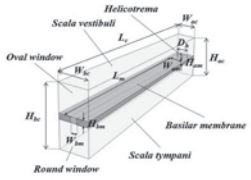


Fig.1：Cochlea model

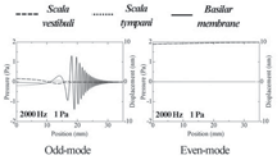


Fig.2：Excitation of traveling waves by odd / even model excitation

2-4-5

ZEST による振幅変調検出のコンピューターシミュレーション

Computer simulations of ZEST procedure of amplitude modulation detection

○森周司，村田悠登（九州大），森本隆司（リオン）

◆時間分解能の検査開発のために、検査で用いる無音検出と振幅変調（AM）検出の閾値測定の高速化を目指している。

◆高速化の手段として、適応法の ZEST を採用した。ZEST はベイズの定理を応用した適応法の一種であり、使用するパラメータが測定対象者の特性を適切に反映していれば、15 試行程度で十分な精度で閾値が測定される。

◆本研究では、AM 検出の閾値測定で用いるパラメータ値を検討するために、ZEST のコンピューターシミュレーションを行った。特に、既存のデータの健聴者群（NH）、難聴者群（HI）、その二つを合わせた集合（PL）のそれぞれから得られた平均と標準偏差をパラメータとして用いた。

◆その結果、PL の平均と標準偏差を用いた場合に、15 試行までで最も誤差が少なく閾値が推定されることが分かった（Figure 1）。

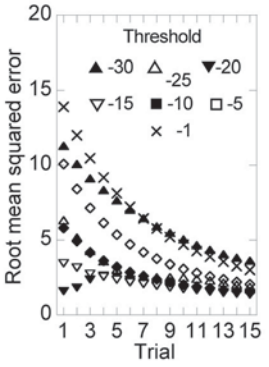


Fig.1：Root mean squared errors of estimated threshold values with mean and standard deviation computed from PL

2-5-1

講演取消

2-5-2

2-5-2　音響インテンシティと FDTD 法を用いた音響インピーダンスの推定

Estimation of acoustic impedances using sound intensities and FDTD method

☆大川祐貴子，渡辺靖明，池田雄介(東京電機大)，及川靖広(早大理工)

◆目的：

少数の測定データと音場シミュレーションによる音響インピーダンス推定手法における測定位置誤差に対する推定精度の改善

◆提案手法：

音響インテンシティと FDTD 法を用いた音響インピーダンスの推定

◆２次元シミュレーション実験の結果から，従来手法と比べ，測定位置誤差を含む場合において最大約 1500 程度の改善が見られた。

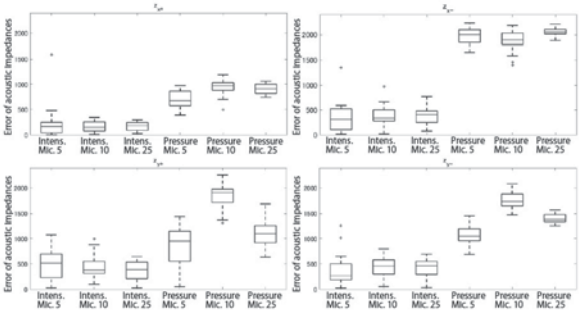


Fig. 1 Boxplot of Estimation Errors of acoustic impedances (Intens. : Proposed method, Pressure : Conventional Method, Mic. : The number of Mic.)

2-5-3

2-5-3　一般的なスピーカを用いた楽器の指向性模擬の試み ―可聴化における聴感印象の変化―

An attempt of simulating musical instrument directivities with a common loudspeaker: Auditory impression variations in auralization

◎原量子，岩見貴弘，尾本章(九大・芸工)

◆楽器の指向性を考慮した室内インパルス応答を得ることを目的とし，一般的なスピーカを様々な方向に回転させ，球面調和領域で指向性を模擬する試みを行っている。

◆音場再生システムに実装し，ホールでのトランペットの演奏を対象として，広がり感・包まれ感・奥行き感・定位感・音色の５項目について，再現した音源と可聴化した音源の聴感印象の違いを調査した。

◆再現した音源と可聴化した音源には知覚できる程度の印象変化が生じており，特に音色については，可聴化の手法間にも有意差が生じていることが明らかとなった。

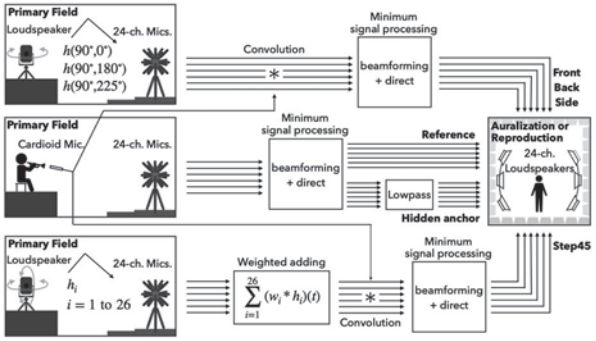


Fig. 1 :　A flowchart for preparing stimuli

2-5-4

2-5-4　虚像法とスパース等価音源法を用いた局所領域の室内インパルス応答推定

Estimation of Room Impulse Response in Local Region using Image Source Method and Sparse Equivalent Source Method.

☆松橋遼，津國和泉，池田雄介(東京電機大)

◆背景

- 効率的に室内インパルス応答 (RIR) を取得するための RIR 推定手法が提案されている。
- これまで，虚像法とスパース等価音源を用いて，直接音と初期反射音を対象に少数固定マイクロホンの周辺の RIR を推定してきた。

◆提案手法

- 初期反射音を等価音源法，後期反射音を平面波分解でモデル化し局所多点 RIR を推定する。

◆実験

- 無響室における実測実験によって推定精度を評価した。
- 直接音・初期反射音成分の推定は 10dB 以上で等価音源法が有効であること，また平面波分解では後期反射音成分のモデル化の精度が不十分であることがわかった。

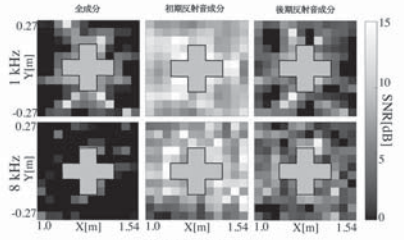


Fig.1 : SNR map

2-5-5

2-5-5　音源の指向性を考慮した等価音源を用いた局所領域室内インパルス応答のモデル化

Modeling of room impulse response with source directivity in local area using equivalent sources

◎津國和泉，松橋遼，池田雄介（東京電機大）

◆問題点
- 室内インパルス応答（RIR）の多点測定は困難
- 等価音源法を用いた多点局所 RIR のモデル化において，指向性をもつ音源のモデル化が不十分

◆目的
- 指向性をもつ音源の多点局所 RIR のモデル化精度の向上

◆提案手法
- 事前に指向性を持つ音源自体のモデル化を行い，その後音場のモデル化を行う手法を提案
- RIR はスパースな等価音源の線形和によって表現

◆シミュレーション結果
- 指向性を持つ音源から放射される一次反射音を含む音場を推定
- 提案手法は従来手法よりも，8 kHz で約 14 dB の誤差の改善

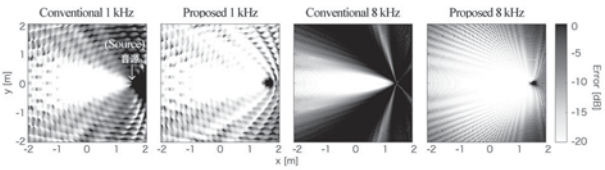


Fig.1: Error distribution of estimated room impulse responses including primary reflections

2-5-6

2-5-6　Partition of Unity FEM による室内音場解析 ー頑健な平面波数の設定に関する検討ー

Partition of Unity FEM based room acoustic simulations -Study on robust plane-wave number setup-

☆向江俊一，奥園健，阪上公博（神戸大・院）

◆本研究では、エンリッチメントの関数として平面波を用いる Partition of Unity FEM（PW-FEM）による室内音場解析手法の開発を行っている。

◆本稿では、PW-FE 解析において、重畳する平面波数を各節点で可変とする Variable q について検討した。

◆まず、Fig.1 に示す様々なサイズの要素を含む非一様メッシュを用いた室内音場解析を対象に、重畳する平面波数を全節点で同一とする Constant q との性能比較を通して、Variable q の有効性を明らかにした（Fig. 2）。

◆次に、モデル形状の複雑性とメッシュの空間分解能が解析の精度やロバスト性に与える影響を検討し、形状に関しては形状が単純な方が、メッシュの空間分解能に関しては分解能が高い方が、よりロバストな解析が可能であることを明らかにした。

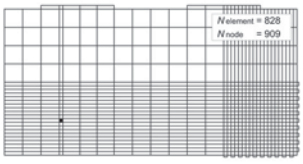


Fig.1 A used mesh for PW-FE analysis.

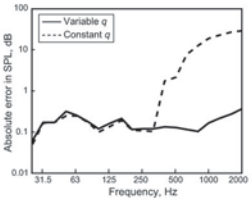


Fig.2 Comparison of absolute error between variable q and constant q.

2-5-7

2-5-7　修正アダムス法を用いた陽的時間領域 FEM による室内音場解析 ー周波数依存の吸音境界の実装ー

An explicit time-domain FEM for room acoustic simulations using modified Adams method -Implementation of frequency-dependent absorption boundaries-

☆杉本優衣(神戸大院・工)，吉田卓彌(安藤ハザマ技研/神戸大院・工)，奥園健，阪上公博（神戸大院・工）

◆陽的な時間領域 FEM による室内音場解析手法の有用性の向上のため、周波数依存の吸音境界を ADE 法により考慮した、修正アダムス法を用いたスキームを提案する。

◆提案法の妥当性を音響管モデルを対象とした数値実験により検証し、次に、実大の 2 次元オフィスモデルを対象に空間４時精度の周波数領域 FEM(FD-FEM)との比較から性能を検証した。

◆音響管において、提案法と理論値は良好に一致した。

◆オフィスモデルにおいて FD-FEM との周波数応答の比較では、1/3 oct. band 音圧レベルの絶対誤差は最大でも 0.81dB で、実用上十分な精度を持つといえる。

◆提案法は FD-FEM と比べ、約 20.4 倍高速、かつ、1/6 倍のメモリ容量で解析が可能である。

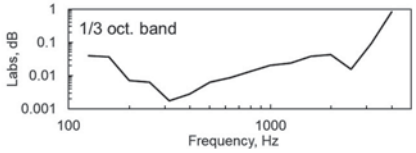


Fig.1 : Absolute error in 1/3 oct. band SPL between the proposed explicit TD-FEM and FD-FEM.

2-5-8

2-5-8　伝達マトリクス法による建築躯体の固体伝搬音解析 ―その１ 解析理論の定式化―

Structure-borne sound analysis of a building frame by the transfer matrix method - A formulation of an analysis theory

○井上尚久(前橋工科大学)

◆床・壁の構造振動を適切に模擬するためには，境界条件を適切に設定することが非常に重要である。

◆一方で，端部におけるインピーダンスの具体的な与え方が不明確であり，方法論も未整備である。

◆L. Cremer らによって示された構造体を伝わる曲げ振動・縦振動の解析方法を取り上げ，以下の２つを提案する。
- 一連の代数操作を伝達マトリクス法により置き換える方法論
- インピーダンスマトリクスによる端部境界条件の表し方

◆インピーダンスマトリクスは理論的に妥当と考えられる値とほぼ完全に一致する。

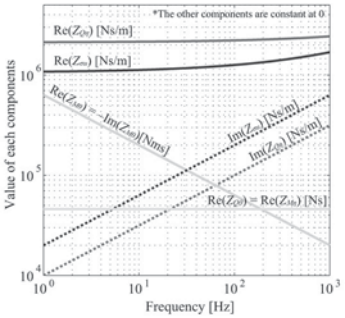


Fig. 1 Each components of impedance matrix calculated at orthogonal cross-shape junction.

2-5-9

2-5-9　CLT の数値解析モデル構築に向けた振動特性および物性値に関する基礎的検討

Fundamental study on vibration characteristics and physical properties for numerical analysis model development of CLT.

○笠井祐輔（日総試），豊田政弘（関西大学），田中学（日総試）

◆建築物の遮音特性の予測に関してはインピーダンス法や FDTD 法などをベースとした数値解析モデルが実用化されているが、木造建築物では実務的にはあまり活用されていない。筆者らは CLT の数値解析モデルの作成を目指し、本稿では、振動特性の把握および各種物性値の設定方法に関する実験と解析の比較を行った。

◆CLT の直交異方性の程度を把握することを目的とし、立方体形状の角の CLT を用いて振動特性の測定を行い、面外方向（ラミナの積層方向）と面内方向という２軸の異方性を持つと考えた。

◆等方性媒質においては、振動伝搬速度と各種物性値との関係式がある。Fig. 1 に示すような 90 mm 角の長さ 2070 mm の棒状の CLT で振動伝搬速度を実測し、その結果から等方性媒質における関係式を適用して、物性値を算出した。

◆算出した物性値を用いて、Fig. 1 と同じ棒状のモデルを作成し、FDTD 法による数値解析を行った。数値解析における振動伝搬速度を算出し、実測との比較を行い、Table 1 に示す結果が得られた。一部で実測値と解析値に違いが見られるが、精度よく再現することができた。

(a) Specimen A　(b) Specimen B

Fig. 1 Diagram of a bar-shaped test specimen. (unit: mm)

Table 1 Vibration velocity

		Longitudinal wave C_P[m/s]	Transverse wave C_S[m/s]
Specimen A	Measurement	3683.6	1170.4
	Analysis	3661	1167
Specimen B	Measurement	1334.4	638.9
	Analysis	1343	1070

2-5-10

2-5-10 共同住宅における給排気口を用いた生活音のマスキングに関する実験的検討

An experimental study on the sound masking of living noise using air supply and exhaust vents in multi-family housings

○黒木　拓，大脇雅直（熊谷組）

近年は自住戸内で発生する音が他住戸や外部でどのように聞こえているか気になる居住者が増えてきている。特に共同住宅の内部共用廊下が非常に静かな環境となる場合が多いため，住戸内の会話や生活音が聞こえていないか気になるという相談が増えてきている。

本稿では，共同住宅の内部共用廊下に設置されている給排気口を用いた生活音のマスキングに関する検討を行った。

◆無響室において検討を行い，給気量や対象音に合わせて背景音付加装置を用いたマスキング音を調整できることを確認した。

◆実建物において，背景音付加装置を設置することで，共用廊下で居室内の会話の内容を聴感上確認することはできなかった。

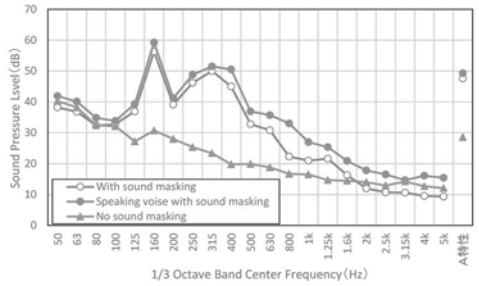


Fig.1　Measurement results of sound masking in the common corridor of an actual building

2-5-11

2-5-11　自宅でのオンライン会議における会話しやすさの印象とプライバシー感の関係

The relationship between ease of conversation and privacy on online meeting at home

☆仁平真由，辻村壮平（茨城大院）

◆自宅においてオンラインで知識創造活動を行う状況における音環境に着目し、自宅を想定した音源を用いて音源の種類、提示レベルを変化させた空間で主観評価実験を実施し、これらの音環境の要因がオンライン会議での会話しやすさに与える影響を検討した。

◆主観評価実験の結果から、オンライン会議を行う自宅の音環境として、「会話しやすさ」には静かな音環境が求められることが示唆された。また、家族の話し声やテレビの音声という自身の家族の存在を表す音や提示レベルが大きくなるに従って「プライバシーが守られている」印象が低下しやすく、「会話しやすさ」も低下していることから、「プライバシー感」が「会話しやすさ」に影響を及ぼす可能性が示された。

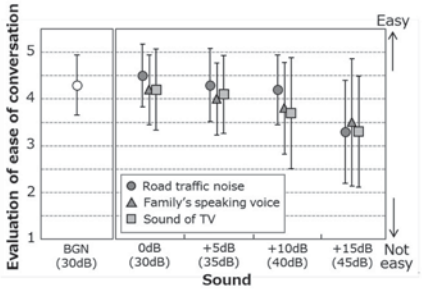


Fig.1 :"Evaluation of ease of conversation" for each sound environmental condition

2-5-12

2-5-12　音線法を用いた都市街区における屋外拡声システムの明瞭性の面的評価

Surface assessment based on sound ray-tracing method for clarity of outdoor public address system in urban area

○佐藤逸人（神戸大院・工学研）

◆都市街区における屋外拡声システムについて，実際の 3D 都市モデルを取り込んだ音線法によるシミュレーションにより，実効的な拡声音の明瞭性分布のマップ化を試みた。

◆明瞭性は暗騒音の影響も考慮した U_{50} で評価するものとし，子局スピーカが設置されている公園周辺の 450m×450m のエリアについて，スピーカからの拡声音に対する U_{50} を算出し，カラーマップを作成した（Fig. 1）。

◆Fig. 1 より，特に建物の遮蔽の影響を受けない見通しの良い道路では，遠方まで明瞭性が確保されていることが確認できる。

◆したがって，都市街区では道路を音の伝搬経路として利用したスピーカ配置計画が有効であると考えられる。

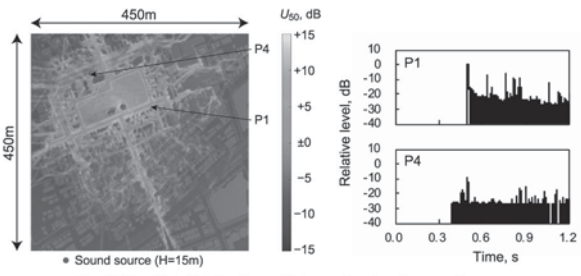


Fig.1 :Spatial distribution of U_{50} and echo-time patterns obtained from the sound ray-tracing method.

2-6-1

2-6-1　打音法を用いた長崎県端島の鉄筋コンクリート柱部材の劣化調査

Deterioration survey of reinforced concrete columns on Hashima Island in Nagasaki using hammering method

◎菅原彬子，△岸本一蔵，平栗靖浩（近畿大学）

◆2015 年に世界遺産の構成遺産の一部として登録された長崎県の端島(軍艦島)には，築数十年から百年の鉄筋コンクリート(RC)造建築が多く存在する．これらの劣化性状を継続的に調査することは，RC 部材の劣化メカニズムの解明に有用である．

◆本研究では，内部劣化性状の影響を明らかにすることを目的として，端島の RC 柱を対象に打音法による調査を行った．

◆結果より，部材表面に現れない内部劣化の影響を検出できることを確認した．また，目視評価結果によらず，内部劣化の進行により低周波数帯域の音圧レベルが上がるという傾向を確認した．

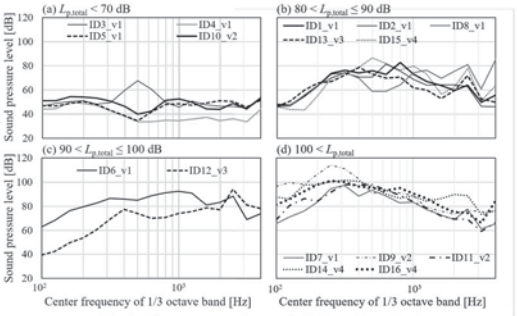


Fig.1：Frequency characteristics classified by the total levels of 100 Hz- 4 kHz bands.

2-6-2

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

○中川裕，杉本恒美，杉本和子，上地樹（桐蔭横浜大　大学院工学研究科）

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

◆探査対象の表面状態に起因するレーザの受光量の低下よるノイズの影響により、欠陥位置の推定精度が悪化する事例が確認されている。

◆測定された信号に欠陥部の共振による信号とそれ以外のノイズを判別する事で、欠陥位置の推定精度を向上する事について検討を行った。

◆測定対象は下面にスチレンシートもしくは発泡シートを埋設した外壁タイル供試体である。

◆Fig.1 は従来手法と提案手法による振動エネルギー分布である。白い破線は欠陥位置を示している。算出結果から従来法では欠陥位置以外でも振動エネルギーが高い場所が散見しているが、提案手法ではそれらが改善され欠陥位置が識別しやすくなった。

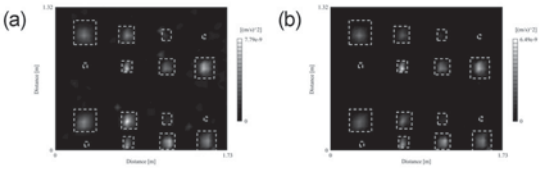


Fig.1：Vibration energy distribution.
(a) Conventional method, (b) Proposed method

2-6-3

2-6-3　弾性表面波伝搬特性による コンクリート内部の火害深度推定

Estimated of depth of fire damage of concrete by surface acoustic wave propagation characteristics

○池谷友秀，大隅歩，伊藤洋一，（日大・理工）

◆本研究は，空中超音波照射によって励起させた弾性表面波を利用してコンクリートの火害診断を行う方法について検討を行っている。火災で残存したコンクリート構造物に対する火害深度診断は，火元の特定や再利用のための補修規模を決める上で極めて重要である。

◆本報告では，空中超音波を利用した新たな火害深度診断手法として，弾性表面波伝搬特性を用いたコンクリート内部の火害の深さ分布(深度)を非接触で推定する方法について実験的に検討を行った。

◆その結果，本手法によりモルタル内部の火害領域の深度推定を非接触で行える可能性を明らかにした。

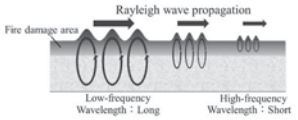


Fig.1：Surface acoustic wave propagation in concrete at each harmonic.

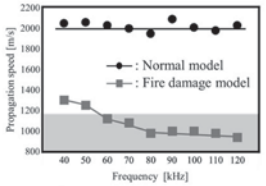


Fig.2：Experimental result.

2-6-4

2-6-4　FDTD による層構造を有する媒質における横波伝搬の基礎検討

Basic examination of shear wave propagation on layered structure by Finite-difference time-domain method

☆大里晃大[1]，西東若菜[2]，大栗拓真[2,3]，神山直久[3]，
平田慎之介[4]，吉田憲司[4]，山口匡[4]
（[1]千葉大・工，[2]千葉大院・融合理工，[3]GE ヘルスケア・ジャパン，
[4]千葉大・CFME）

◆筋組織などの層構造がせん断波の伝搬と shear wave velocity（SWV）の評価に与える影響について，FDTD 法を用いた計算機シミュレーションで検証した．

◆各層の固有せん断波音速の差および境界面でのアーチファクトにより，時間経過に伴いせん断波の伝搬方向が変化し，SWV にばらつきが生じ，評価精度が低下した（Fig. 1）．

◆今後，より生体の線維構造に近づけた空間モデルでのせん断波シミュレーションを実施し，SWV 算出プロトコルの違いによる評価精度差についても検証する．

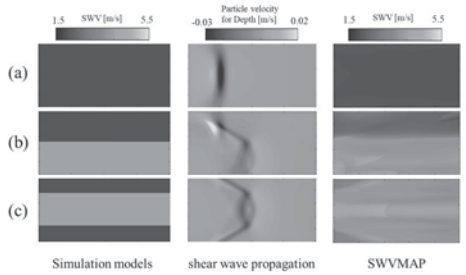


Fig. 1 Simulation models, shear wave propagation and SWVMAP of homogeneous (a), two-layer (b), three-layer models (c).

2-6-5

2-6-5　層構造を持つ媒質における アニュラアレイでの後方散乱特性評価

Evaluation of backscattering properties in tissue with layered structure by annular array

☆西東若菜（千葉大・院融合），大村眞朗（富山大・工/千葉大・CFME），
大澤侑暉（千葉大・工），△Jeffrey A. Ketterling（Riverside Research），
平田慎之介，吉田憲司，山口匡（千葉大・CFME）

◆中心周波数 20 MHz のアニュラアレイで，実際の真皮および皮下組織における音響インピーダンスの違いを模擬した層構造ファントムを対象とし，後方散乱係数（BSC）を評価することで，上層の厚さの違いが後方散乱係数の評価精度に与える影響について検討した．

◆上層では，厚くなるほど BSC が大きくなることが確認された．また，下層では BSC が理論値に比して 2～3 dB 程度大きく評価される結果となった（Fig. 1）．

◆簡便なシステムであるアニュラアレイによる表在組織の組織性状評価の可能性が示唆された．今後は，臨床応用に向けたリアルタイム性の向上を踏まえたシステム構築を行う．

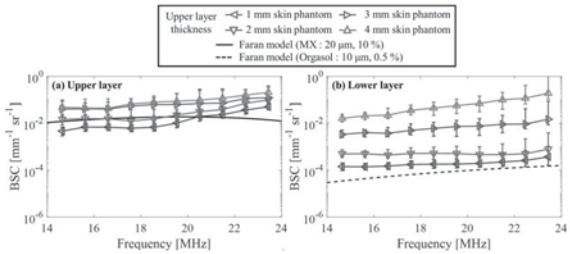


Fig.1 : Frequency dependency of BSC in layer phantoms

2-6-6

2-6-6　後方散乱特性解析における 送受信音場と組織構造の関係性の検証

Varification of relationship between tissue structure and transmitting/receiving sound field on backscatter coefficient analysis

☆八木風音，西東若菜（千葉大・院融合），
大村眞明（富山大・工，千葉大・CFME），
平田慎之介，吉田憲司，山口匡（千葉大・CFME）

◆4 種の超音波センサを用いて取得した 6 種の散乱条件の異なる均質ファントムのエコー信号を対象に後方散乱係数（BSC）解析を行い，送受信音場と組織構造の関係性について考察した．

◆散乱体径 10 μm，体積分率 3 %のファントムを参照媒質とした各ファントムの後方散乱係数 BSC（Fig.1）推定において，10 MHz の振動子においては，BSC は理論とほぼ一致し，参照媒質と評価媒質の散乱強度差がそのまま BSC に表れた．

◆25 MHz でも基本的に同様の傾向が確認されたが，20 μm のファントムでは参照媒質と評価媒質の条件の乖離が大きいため，BSC の評価精度が低く，音場特性に適した参照媒質の設定の重要性が確認された．

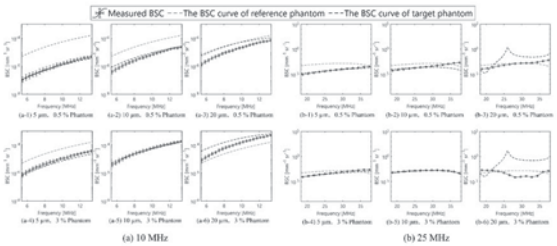


Fig.1 : Evaluated BSCs referred with 10 μm, 3 % phantom.

2-6-7

2-6-7　聴覚保護具の遮音性能評価における 規格化された単一評価指標の比較

Comparison between standardized single number ratings for evaluating the sound attenuation of hearing protectors

○横山 栄（小林理研）

甚大な騒音環境における作業者の聴覚を保護するためにはイヤーマフや耳栓等の適切な保護具を選定することが重要であり、そのためには比較可能な評価指標が示されることが不可欠である。わが国では 2020 年に ISO 4869 シリーズを対応国際規格として、聴覚保護具の測定法および評価法を規定した JIS 規格「聴覚保護具（防音保護具）」が制定され、国際規格との整合化が図られたが、欧米における法規制の違いにより現在も統一には至っていない。本報では、各規格に準拠した REAT（real-ear attenuation at threshold）測定によって得られるオクターブバンド毎の遮音値および標準偏差の結果から求められる単一評価指標について、国内外で規格化されている指標を比較した結果を報告する。

Table 1 : Comparison of standardized ratings.

Rating	Standard	Frequency weighting	3M™ PELTOR™, H9A	Country
SNRx single number rating	ISO 4869-2: 2018 (JIS T 8161-2: 2020)	C-A' in pink noise	31 dB	Europe, Japan, China, Korea
H_x, M_x, L_x **value** high/medium/ low-freq. attenuation value		C-A'	H=36 dB M=29 dB L=20 dB	
NRR noise reduction rating	(by U.S. EPA, 1979)	C-A'-3 in pink noise	(25 dB) *16 subjects	U.S.A., Canada
NRS_A noise reduction statistic for A-weighted noise	ANSI/ASA S12.68-2017	A-A' *NIOSH 100 noise spectra	—	U.S.A., Canada
***CSA* class** *A/AL,B/BL,C	CSA Z94.2-14	A-A'	(Class A)	Canada
SLC_{80} sound level conversion	AS/NZS 1269-3:2005	C-A'	(31 dB) *experienced-subject	Australia, new Zealand

2-6-8

2-6-8　ダミーヘッド（HATS)を用いた VHF 領域の耳介周り音響計測（その 2）

Acoustic measurement around the pinna in the VHF region using a dummy head -Part2

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

◆ 先行研究では、HATS を用いて頭部周辺の影響で VHF 領域の音がどのように変化するかを耳介近傍で計測し、HRTF を求めた。

◆ 今回の研究では、別の HATS を加えて 2 体の結果を比較し、頭部周辺の VHF 音への影響を確認した。

◆ 結果、前回の 1/4 inch.マイクと比較した場合、周波数間のレベル差のばらつきが大きく、背面側のバラつきは似ているものの細かなピークディップは異なっていた。

◆ MEMS マイクと比較した場合、バラつき自体は似通っているが同一周波数間で比較すると HRTF のピークが異なり、高周波になるほどその傾向が見られた。

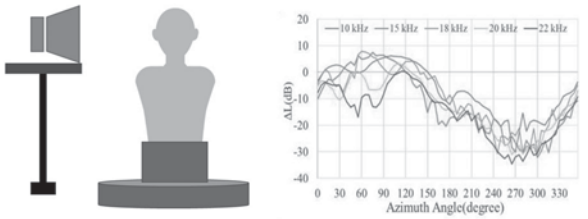


Fig. 1 A measuring system in this study and HRTF of 'SAMURAI' HATS.

2-6-9

2-6-9　大学生を対象とした 20 kHz 近傍の聴覚閾値計測

Auditory threshold measurement near 20 kHz for university students

☆佐藤龍之介，春澤恒輝，五十嵐康起，藤田惣（神奈川工科大），
廣江正明（小林理研），中村健太郎（東工大），長谷川英之（富山大），
神崎晶（慶応大），桐生昌吾（東京都市大），上田麻理（神奈川工科大）

◆ ヒトの可聴域は 20 Hz～20 kHz と言われている。しかし、近年の研究で、若年層の場合 100 dB 超の音圧であれば 28 kHz 前後まで聴取できることが報告されている。

◆ こうした閾値に関する報告事例はまだまだ少ない。本研究では、大学生 20 名を対象に、16 kHz～22 kHz の聴覚閾値を計測した。

◆ 計測の結果、16 kHz は 18 名、18 kHz は 13 名、20 kHz は 1 名の閾値が得られた。今回確認した 16 kHz、18 kHz の最小閾値は 19 dB、49 dB であった。

◆ また、先行研究との比較（Fig. 1）から、今回の閾値は妥当な結果であると判断された。

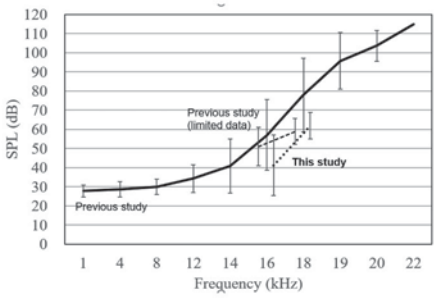


Fig.1 Comparison of auditory thresholds between previous studiy and this study.

2-6-10

2-6-10　音響用語としての「サウンドスケープ」をめぐって

On soundscape as an acoustic term

〇永幡幸司（福島大），森長誠（神奈川大）

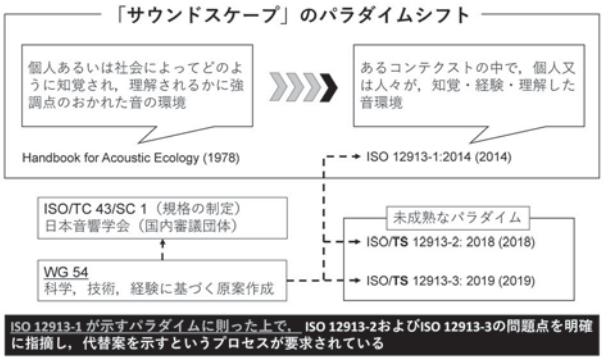


Fig.1 : Outline of the paper: This paper shows the paradigm shift of “sound scape” as an acoustic term caused by ISO 12913-1: 2014 and shows the relationships with ASJ.

2-6-11

2-6-11　身の周りの VHF 音その 3 －LED 照明の発生源探査－

VHF sounds around us part-3
- Source detection of VHF sound from desktop LED light stands -

☆五十嵐康起，春澤恒輝，佐藤龍之介，藤田惣（神奈川工科大），
廣江正明（小林理研），中村健太郎（東工大），長谷川英之（富山大），
上田麻理（神奈川工科大）

◆先行研究で家電製品を対象とした VHF 音の発生調査を行い，LED 卓上照明から調波成分をもつ VHF 音の発生を確認した．本研究では，調光段階や電源供給タイプの異なる 10 種類の LED 卓上照明を対象に，VHF 音の発生の有無とその発生箇所を探査した．

◆VHF 音の発生を確認した 8 種類のうち，先行研究の LED 照明を除く 7 種類中の 5 種類は，VHF 音が最低の調光段階で発生し，最高の調光段階では発生しなかった．また，各 LED 照明の VHF 音の周波数特性は異なっていた．

◆音源探査の結果，先行研究の LED 照明以外に，USB 電源タイプ 3 機種，および AC 電源タイプ 2 機種で，最低調光段階の時に電源部付近から VHF 音が発生していることを確認した．

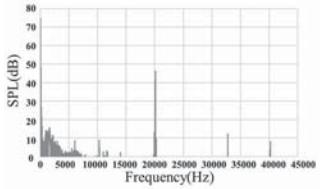


Fig. 1 Frequency characteristics of VHF sound from Semlos LED for low level of brightness with. USB power supply (θ =0 deg.).

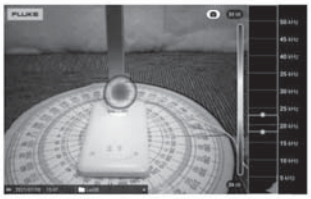


Fig. 2 A result of source detection of Semlos LED with USB power supply (θ =0 deg.)

2-6-12

2-6-12　走行中の列車及び列車内における高周波音の計測

Measurement of high-frequency sound on running trains and curves inside trains

☆藤田惣，春澤恒輝，五十嵐康起，佐藤龍之介（神奈川工科大），
廣江正明（小林理研），中村健太郎（東工大），長谷川英之（富山大），
上田麻理（神奈川工科大）

◆先行研究では曲線区間の走行列車から VHF 音が発生していることが明らかとなっている。今回の研究では、曲線走行する列車内外で同時に発生音の収録を行い、VHF 音の周波数特性を分析した。また、同時に音カメラを用いて VHF 音の音源探査も行った。

◆測定の結果、車内外ともに数多くの列車で 15 kHz～35 kHz の周波数帯域に VHF 音が発生していることが判明した。

◆音カメラによる音源探査の結果、車輪とレールの接触部あるいは車両搭載機器が VHF 音の発生箇所であることが示唆された。

◆車内で観測された VHF 音は車両下部から伝搬したものと推察される。

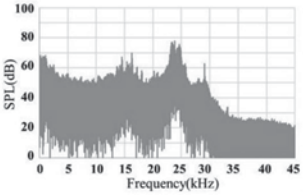


Fig.1 Frequency characteristics of VHF sound from running train (4/21, 12:34, 3000 Series).

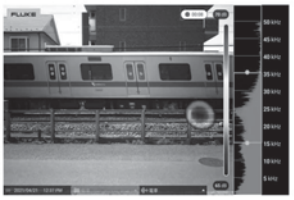


Fig.2 Results of source detection on VHF sounds — 4/21, 12:34, 3000 Series —

2-6-13

2-6-13　聴覚特性を考慮した紙幣処理機の騒音評価とその改善

Noise evaluation and countermeasure of currency processing machine considering auditory characteristics.

○寺田幹，吉田準史（阪工大・エ），△松川岳嗣

◆　本研究では紙幣処理機の稼働時騒音を紙幣処理機の基本機能を損なわない対策により改善することを目的としている．

◆　稼働時騒音の放射音に対してラウドネスや音圧レベルを用いた分析の結果 730 Hz のギヤ次数成分と 3.5 kHz 以上の高周波成分が騒音増大の要因であることが明らかとなった．このことを考慮した上で紙幣処理機稼働時の動作に着目したところ，730 Hz の次数成分の発生原因は搬送ベルトと当該ベルトを駆動させるギヤによるものであり，また高周波成分は搬送ベルトと搬送路側ローラによる摩擦音であることが判明した．そこで騒音低減策として搬送路側ローラを歯付きプーリに変え，ベアリングを搭載する対策を施した．その結果，対策前後における放射音のラウドネス分析 (Fig. 1) より，30 %程度の音の大きさの低減を行なうことができた．

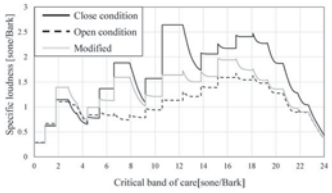


Fig.1 : Comparison of specific loudness between original and modified conditions.

2-6-14

2-6-14　見回り移動型ロボットの音響パワーレベル計測その 1

Measuring the sound power level of a mobile robot part1

☆森川結斗，高橋達樹（神奈川工科大），廣江正明（小林理研），上田麻理（神奈川工科大）

<概要>

■ 近年医療施設や高齢者施設などで見回りを目的としたロボットが普及している。

■ ロボットの駆動音に関する知見は多くなく、その音響的特徴などもあまり知られていない。

■ そこで本研究では、ロボット（今回は移動型の見回り介護用ロボット）の音響特徴を明らかにする。

■ 音響インテンシティ法(Fig.)により音響パワーレベルを計測した。

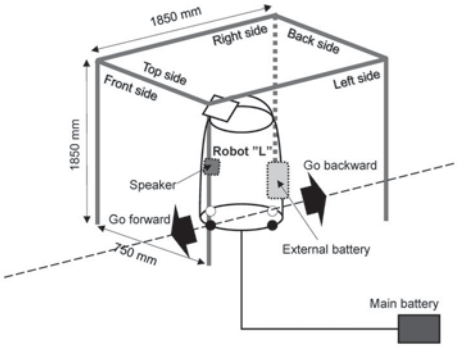


Fig. Rectangular frame and robot movement direction

2-7-1

2-7-1　縦振動子を用いた自走式スイマー

Self-propelled swimmer using longitudinal vibrator.

☆谷村　瞭, 孔　徳卿, 青柳　学（室蘭工大・院）

◆本研究では縦振動をする振動子を用いて水中を自走するスイマーの製作を行った。

◆Fig.1 に本研究で使用した振動子とそのアドミタンス特性の測定結果を示す。使用した振動子は，(株)富士セラミックス社の圧電セラミックエレメント(型式：Z10T5D-LYX(C-213))である。また，アドミタンス特性の測定より，150 kHz 付近に共振があることが確認できた。本研究では 149.1 kHz を駆動周波数として使用する。

◆Fig.2 にスイマーを前進・後退させた時の速度を示す。最も速度が出たのは電圧を 292 V_{pp} 印加したときで，23 mm/s が得られた。また，スイマー後退時の速度は 2.2 mm/s であった。

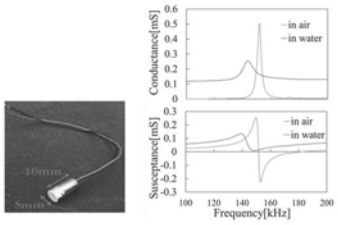


Fig.1 Longitudinal transducer and Admittance characteristics of transducer in air and water.

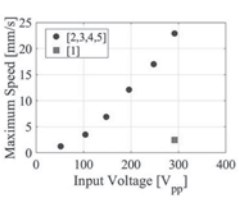


Fig.2 Speed of swimmer by diverse voltage in water.

2-7-2

2-7-2　弾性表面波を用いた SAW スイマーの VHF 帯における推進力の検討

Examination of propulsive force of surface acoustic wave (SAW) swimmers in VHF band.

○相原一鷹（東工大），孔徳卿（室蘭工大），梶井将，黒澤実（東工大）

◆弾性表面波(Surface Acoustic Wave, SAW)デバイスは超音波モータや、スマートフォンの内部にある SAW フィルタなどに利用されている。この SAW デバイスを用いて液中を駆動するアクチュエータを SAW スイマーと呼び現在研究を行なっている。SAW デバイスの特徴の一つに駆動周波数を高くすることによって小型化が可能であることが挙げられ、血管内を動くマイクロロボットへの応用が期待される。

◆先行研究では駆動周波数 9.64MHz のスイマーの駆動に成功しているが、本研究では駆動周波数を 38.45MHz 及び 96.13MHz に上げて小型化したスイマーを実際に作成し、その推進力の検討を行った。

◆その結果、周波数上昇に伴って単位面積当たりの推進力が向上していることが確認できた。

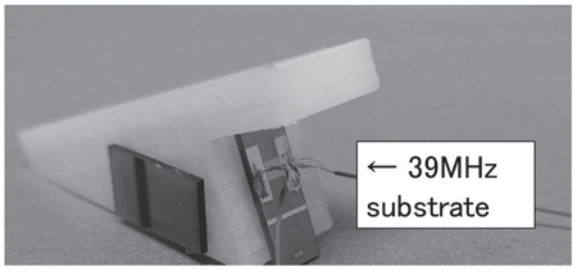


Fig.1 : Prototype SAW swimmer at 39MHz

2-7-3

2-7-3　厚み振動モードの非鉛性振動子(BNBTM, SCNN)の液中推進システムの検討

Thickness-Vibration-Mode lead-free ultrasonic transducers (BNBTM, SCNN) for underwater propulsion system.

☆平田　拓己, 孔　徳卿(室蘭工大･院), 土信田　豊(足利大),
黒澤　実(東京工大), 青柳　学(室蘭工大･院)

◆血管内を自走で推進できるスイマーアクチュエータの作製のため，厚み振動モードで，人体に無害である非鉛性の振動子(BNBTM，SCNN)の推力を測定し，比較・検討を行った結果について報告する。

◆それぞれの振動子のアドミタンス特性の測定結果をFig.1とFig.2に示す。また，推力測定結果をFig.3に示す。

◆BNBTM振動子は，優れた推進性能を持ち，血管内治療などに適した振動子である。また，SCNN振動子は，現段階では推進性能が低いが，材料の細かな調整により，更なる性能の向上が期待できる。

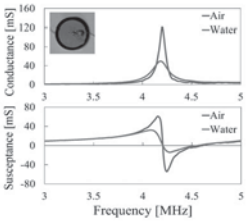


Fig.1 Admittance characteristics of BNBTM.

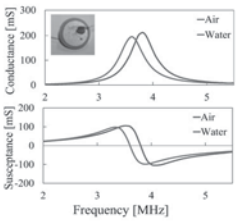


Fig.2 Admittance characteristics of SCNN.

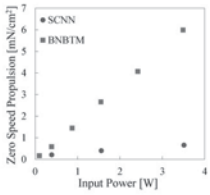


Fig.3 The result of propulsion measurements.

2-7-4

2-7-4　振動振幅及び駆動周波数による圧電素子を用いた液中推進システムの推進力の変化

Variation of Propulsive Force of underwater Propulsion System Using Piezoelectric Elements with Vibration Amplitude and Driving Frequency

☆銭源(東京工業大学), 孔徳卿（室蘭工業大学), 黒澤実(東京工業大学)

◆圧電素子を用いた新たな液中推進システムは、構造がシンプルで小型化が容易なため、小型液中ロボットへの応用が期待されている。現在、液中推進システムによく使われている圧電素子は、円板状厚み振動モードPZT　($Pb(Zr,Ti)O_3$)トランスデューサである。

◆新たな液中推進システムの基礎的な実験データとして、３つの円板状厚み振動モード PZT トランスデューサ（富士セラミックス、PZT1:c-204, Φ9 × 1mm,2.07 MHz、PZT2:c-204, Φ9 × 0.7mm,3.03 MHz、PZT3:c-213, Φ9 × 1mm,2.02 MHz）の水中での振動振幅と推進力を測定した。

◆推進力の測定結果をみると、PZT(c-204, 3.03 MHz)は駆動電圧の増加に対して，推進力の増加量がもっとも大きいことがわかる。

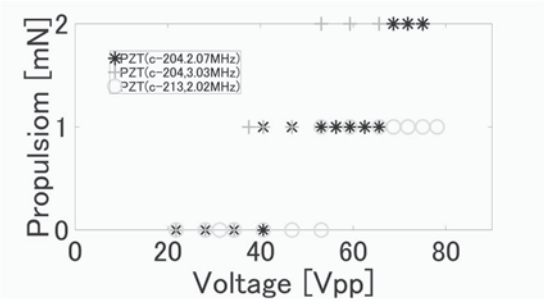


Fig.1　Measurement of propulsion

2-7-5

2-7-5　３つの空中超音波の干渉による二次元超音波マニピュレーション

Two-dimensional ultrasonic manipulation in air by crossing three sound beams

○小塚晃透, 芳本拓也(愛工大), △豊田昌弘(本多電子), 畑中信一(電通大)

◆空中超音波を３方向から照射して、２次元状に広がる定在波音場を形成し、音圧の節に微小物体を捕捉することを試みた。

◆３方向の音波の位相を制御することで、各音軸の方向に音場を移動させて、捕捉した物体を移動させることが可能であった。

◆２方向の音波に FM 変調を掛けて周期的に位相を振動させることで、捕捉した物体を円を描く様に移動させることが可能であった。

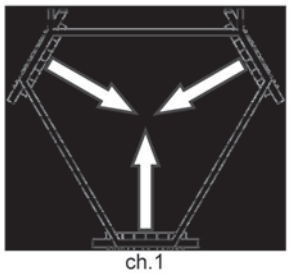


Fig. 1. Superimposing device for sound beams crossing from three directions.

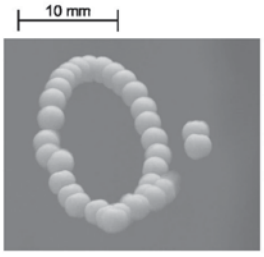


Fig. 2. Manipulation of particle (superimposed photograph)

2-7-6

2-7-6　小径穴付き円筒から噴出する流体による浮揚時に円筒に作用する保持力特性

Characteristics of the holding force acting on a levitated cylinder with a small hole by an acoustic streaming ejected downward from its own hole

◎青野浩平，孔徳卿，青柳学 (室蘭工大・院)

◆平面上に置かれた小径穴付き円筒に対して上方から振動面を近づけることにより，穴から平面に対して空気が噴出し円筒が浮揚する。このとき浮揚円筒には保持力が作用し振動面下に保持される。保持力特性について実験的に調査した。

◆Fig.1 に小径穴付き円筒に作用する保持力の測定方法を示す。平面，浮揚した円筒および振動面を含む装置全体を傾け，浮揚円筒が振動面外へずれ落ちるときの角度 φ の測定結果から保持力 F_h を算出した。

◆厚み T が 3，5 および 8mm のときの保持力測定結果を Fig.2 に示す。空気ギャップ H が増加すると保持力が低下した。また，厚み T が 5mm のときが他の厚みに対して保持力が最も大きいことから最適な厚みがあると考えられる。

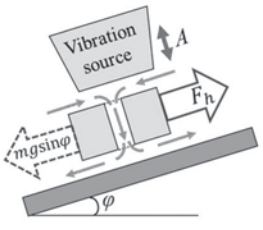


Fig.1 Measurement method of holding force.

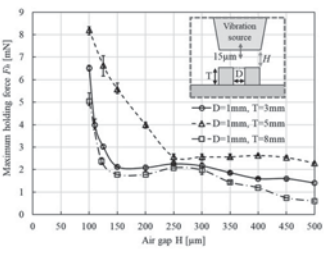


Fig.2 Maximum holding force F_h vs. air gap H.

2-7-7

2-7-7　超音波圧入加工における工具ホーンの駆動力変成振動系の検討

—電動アクチュエータを用いた超音波圧入加工(10)—

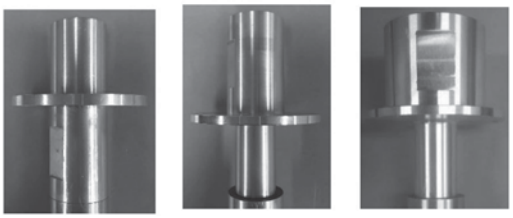
Examination of vibration system using horn with driving force transformation,
-Ultrasonically forced insertion device using an electric actuator (10)-

☆高橋功喜，青柳学（室蘭工大院）

◆本報告では，超音波圧入加工における工具ホーンの振動変成比を１よりも下げ発生力を大きくした際の圧入効果の違いを調査した。

◆駆動力無変成型ホーン（Straight-horn，速度変成比 1，共振周波数 27,983 Hz），駆動力２倍変成型ホーン（2F-horn，速度変成比 0.46，共振周波数 27,068 Hz），駆動力４倍変成型ホーン（4F-horn，速度変成比 0.23，共振周波数 27,071 Hz）を作製し，28 kHz のボルト締めランジュバン振動子と締結して圧入実験を行った。(Fig.1)

◆圧入量は 2F-horn，Straight-horn，4F-horn の順で大きく，圧入棒の締結側端部の振動速度は，2F-horn と Straight-horn で共に近い値になった。



(a) Straight-horn　(b) 2F-horn　(c) 4F-horn

Fig.1 Trial horns.

2-7-8

2-7-8　電極の厚みを考慮した円柱型多自由度超音波モータの設計精度の向上

Improvement of design accuracy of multi degrees of freedom in the consideration of thickness electrode thickness

☆山梨一成、小池義和(芝浦工大)

◆球形のロータを 2 つのステータで両側から挟み込む構成の多自由度超音波モータについて検討を行った。

◆電極の厚みを検討するために電極の厚さを考慮した有限要素法(FEM)により設計を行い、実現する多自由度超音波モータ（MDOF-USM）の性能向上を目指している。

◆電極の厚みを考慮して解析を行った場合、要素数が 60 万要素であるのに対し、電極を PZT の一部として解析を行った場合は要素数が 5 万要素であった。

◆縦振動が主体の振動モードでは電極の有無の影響が少ないことが確認できたが、他のモードでは無視できない影響が現れている。

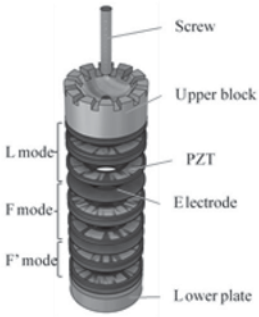


Fig.1 Composition of MDOF-USM

(a) with electrode　(b) without electrode

Fig.2 Comparison of meshed model with electrode to model without electrode.

Table 1 Comparison ofresonance frequencies in L, F and F' mode with electrode to frequencies without electrodes.

Mode	Resonance frequency with electrode [kHz]	Resonance frequency without electrode [kHz]
L mode	199.3	200.9
F mode	160.1	168.2
F' mode	160.1	168.2

T

2-7-9

2-7-9　NASH 肝エコーの振幅包絡特性解析における組織分布と評価パラメータの関係性に関する検証

Verification of relationship between tissue distribution and evaluation parameters in amplitude analysis of NASH liver echoes

☆氏原裕貴(千葉大・院融合)，田村和輝(浜松医科大)，
森翔平(東北大院・工学研)，平田慎之介，吉田憲司，山口匡(千葉大・CFME)

◆NASH 肝の特性評価の基礎段階として，計算機シミュレーションにより，正常肝の管腔構造を模擬した弱散乱体と，脂肪滴を模擬した強散乱体をランダムに配置した 10 種類の脂肪肝模擬ファントムを作成し，各ファントムの三次元 RF エコー信号について multi-Rayleigh (MRA) モデルによる振幅包絡特性解析を行った.

◆MRA モデルで推定される，エコー信号内の成分数に特に着目し，散乱体構造との対応付けを行った.

◆脂肪滴の数密度と管腔構造の散乱体数密度の比率に応じて，それぞれが独立した信号であると評価されるケースといずれかが優位な信号と評価されるケースに分けられることを確認した（Fig. 1）.

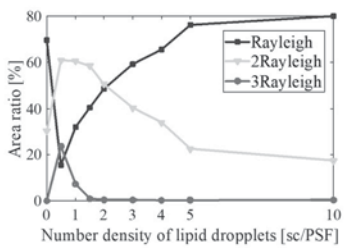


Fig.1 Relationship between number density of lipid droplets and component number of MRA model.

2-7-10

2-7-10　臓器不全マウス・ラットにおける多種組織の音響インピーダンス評価

Evaluation of acoustic impedance of various tissues in organ-deficient mice and rats

☆橋本明楽（千葉大・工)，大澤侑暉（千葉大・工)，平田慎之介，
吉田憲司（千葉大・CFME)，丸山紀史（順天堂)，山口匡（千葉大・CFME)

◆ラットおよびマウスの肝臓および筋組織を対象として，中心周波数 250 MHz の ZnO 振動子を組み込んだ超音波顕微鏡を用いて，音響インピーダンスを評価し，組織構造との比較を行った.

◆肝臓の音響インピーダンスは，正常肝に比して軽度脂肪肝では低値を示す領域が多く，全体的に不均質になっていることが確認された．一方で，筋肉の音響インピーダンスでは，肝疾患に伴う筋肉の組織性状変化が小さかったため，差異は確認されなかった（Fig.1）.

◆現在，筋疾患モデルマウスの摘出組織を用いての追検討を実施している.

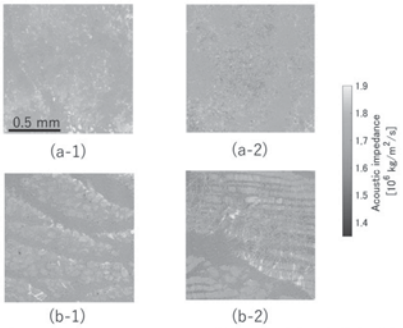


Fig.1 : Acoustic Impedance maps of normal liver (a-1), mild fatty liver (a-2), muscles in normal (b-1) and mild fatty liver rats (b-2).

2-7-11

2-7-11　臨床から顕微レベルの周波数帯におけるラット臓器の音速・減衰特性の評価

Evaluation of speed of sound and attenuation characteristics of rat organs in clinical to microscopic frequency band

☆伊能舞，野口和馬，瀬戸駿（千葉大・院融合），
平田慎之介，吉田憲司，山口匡（千葉大・CFME）

◆3.5～300 MHz の超音波で脂肪肝モデルラットの肝臓を計測し，音速と減衰を評価した．

◆透過法（3.5 MHz）と反射法（3.5～15 MHz）で評価した音速に大きな差はないが，超高周波（300 MHz）の反射法では音速が高値を示した．正常肝での検討では超高周波でも差が認められていない．

◆減衰係数は，反射法の結果で周波数依存性が示されているが，透過法との差が確認された．

◆脂肪の有無により音速・減衰の周波数依存性が正常肝とは異なることが確認されたが，評価試料作成プロトコルや信号解析法の安定性についての再評価が必要．

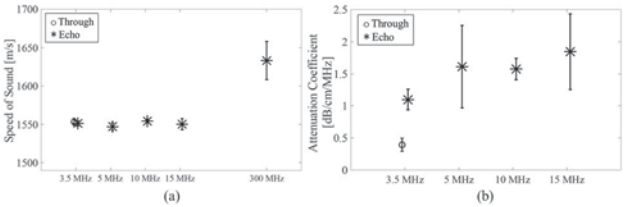


Fig. 1：Speed of sound (a) and attenuation coefficient (b) of fatty rat livers.

2-7-12

2-7-12　試料作製法の違いが生体組織の音速評価に与える影響の検討

Examination of the effect of different sample preparation methods for speed of sound evaluation in biological tissues

☆瀬戸駿，野口和馬（千葉大・院融合工），田村和輝（浜松医科大），
平田慎之介，吉田憲司，山口匡（千葉大・CFME）

◆性状の異なる複数の組織において，中心周波数 300 MHz の振動子を用いて音速を評価した．ホルマリンまたはハーフカルノフスキ液に浸漬する工程，濃度 30 %のスクロース液に 24 時間浸漬する工程の有無の違いが，音速評価に与える影響について検討した．

◆5 種の試料作製法と音速評価結果の関係性について，肝臓と腎臓で同様の結果となった．パラフィン包埋試料では一般に知られる両臓器の音速に比して速いが，熱処理を加えない他の 4 種では音速が遅くなることが確認された（Fig. 1）．

◆ハーフカルノフスキ固定後にスクロース置換を施すことで，ホルマリン固定の試料と同程度に組織損傷を小さくできることと，熱処理を除することによる音速評価の妥当性の向上を確認した．

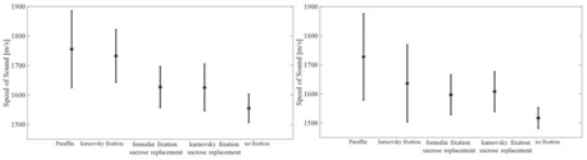


Fig.1 Speed of sound with different sample fixation methods (a:liver, b:kidney).

2-7-13

2-7-13　多周波数における超音波位相差を用いた速度推定法の検出最大速度と局所性の関係

Relationship between detectable maximum velocity and locality in multifrequency phased-tracking method

☆小原優，森翔平，荒川元孝，金井浩（東北大）

◆【目的】心筋層の局所的な厚み変化計測を目指し，多周波位相差トラッキング法（提案法）の検出最大速度と窓幅の関係を理論的に検討した上で，水槽実験系を用いて実験的に検証した。

◆【原理】提案法では，多周波数における位相差の傾きを推定することで，速い変位を推定することができる。しかし，窓長が短く，周波数分解能が低いとき，受信信号の中心周波数以外では位相差に偏り誤差が生じ，位相差の周波数特性の傾きが真値より小さく誤推定される。

◆【結果・考察】水槽実験における提案法による変位推定の誤差を Fig. 1 に示す。提案法の検出最大速度と局所性の関係は，理論的な検討結果（黒線）と整合することが確かめられた。この関係は，トランスデューサのインパルス応答から事前に推定できるため，本手法により心筋層の局所的な厚み変化計測に最適な窓長を決定できる。

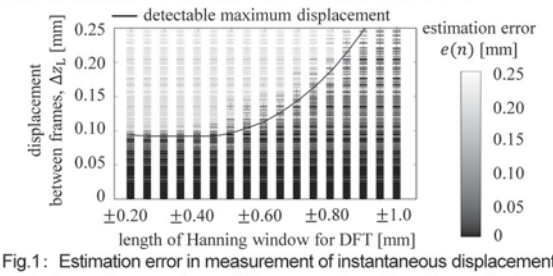


Fig.1：Estimation error in measurement of instantaneous displacement between frames using multifrequency phased-tracking method.

2-7-14

2-7-14　超音波の反射・散乱特性の差異を用いた胸椎描出における集束波送信好条件の検討

A study on transmission conditions of focused waves for thoracic spine imaging using differences in ultrasonic reflection and scattering characteristics

☆阪東泰河，森 翔平，金井 浩，△大西詠子，△山内正憲，荒川元孝（東北大学）

◆硬膜外麻酔施行の際，麻酔針の穿刺位置の特定の補助として医用超音波を用いることがある．我々は骨の反射特性と筋組織の散乱特性の差異を利用した胸椎描出法を提案している．本報では，反射・散乱の振幅角度特性の差異をより強調できるリニアプローブの送信素子数について検討した．

◆送信素子数を変化させたときの，反射体からの振幅角度特性を Fig. 1 に，焦点付近のラテラル方向の音場分布を Fig. 2 に示す．送信素子数が減少すると反射体からの振幅角度特性の形状は変化するが，散乱体からの振幅角度特性の形状はほとんど変化しないため，振幅角度特性の差異が大きくなる．一方，送信素子数の減少に伴い方位分解能は低下することから，トレードオフの関係が成り立つ．実際の計測においては胸椎の間隙が確認できるよう，適切な送信素子数を決定する必要がある．

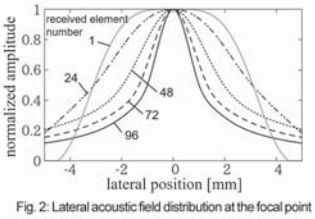


Fig. 2: Lateral acoustic field distribution at the focal point

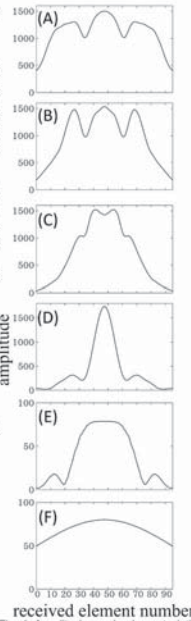


Fig. 1: Amplitude-angle characteristics.
(A), (B), (C), (D), (E): reflector, the numbers of transmitted elements are 96, 72, 48, 24, 1, respectively.
(F): scatterer with the transmitted element number of 96.

2-7-15

2-7-15　超音波後方散乱特性計測による赤血球集合度推定手法の検討 —点散乱体から計測したパワースペクトルによる散乱体サイズ推定—*

A method for estimating degree of red blood cell aggregation by measuring ultrasonic backscattering characteristics
-Estimation of scatterer size using power spectrum measured from a point scatterer-

☆東山京平，森 翔平，荒川元孝（東北大），
△八代 諭，△石垣 泰（岩手医大），金井 浩（東北大）

◆本研究グループでは，非侵襲的かつ定量的な血液性状評価のため，超音波計測により血管内腔（深さd）の赤血球に対して計測したパワースペクトル $P_s(f,d)$ を，反射体とみなした血管後壁のパワースペクトル $P_r(f,d)$で正規化することでサイズ固有の周波数特性 $S(f)$を抽出し，赤血球集合体サイズを推定してきた．しかし血管後壁の反射特性が周波数依存性を持ち，サイズ推定に影響すると考えられる．本報告では，$P_s(f,d)$ と直接比較してサイズを推定するため，予め用意するモデル散乱パワー $\{P_{MD}(f,d)\}$ の作成の検討を行った．

◆$P_{MD}(f,d)$は，直径Dの理論散乱特性 $S_D(f)$の他に，印加信号の周波数特性 $X(f)$，超音波プローブの送受信特性 $G(f)$，超音波プローブの深さdでの音圧特性 $H(f,d)$，伝播減衰特性 $A(f,d)$ を含む．本手法では直径ϕの単一球散乱体から取得したパワースペクトル $P_\phi(f,d)$ から式(1)のように，任意の直径Dの $P_{MD}(f,d)$ を作成できる．

$$P_{MD}(f,d) = P_\phi(f,d) \times \left|\frac{S_D(f)}{S_\phi(f)}\right|^2 \quad (1)$$

◆先端が球状のワイヤを単一球に見立てて超音波計測を行うことで $P_\phi(f,d)$ を取得した．また，異なる直径のワイヤから取得したパワースペクトルの互いの比 $P_{\phi_1}(f,d)/P_{\phi_2}(f,d)$ と理論散乱パワーの互いの比 $|S_{\phi_1}(f)|^2/|S_{\phi_2}(f)|^2$ の比較により式(1)による散乱パワースペクトルの精度を評価した結果，両者はほぼ同様の周波数特性を示し，$P_\phi(f,d)$ がモデル散乱パワー $\{P_{MD}(f,d)\}$ の高精度な作成に適している可能性を示した．

2-7-16

2-7-16　超音波伝達系の深さ依存性を考慮した特異値分解に基づく広帯域化フィルタの設計手法の検討

Design of broadband filter based on singular value decomposition considering depth dependence of transfer function in ultrasound measurement system

☆川又健太，森 翔平，荒川元孝，金井 浩（東北大）

◆本報では、ロバストなフィルタ設計のために、特異値分解を用いた広帯域化フィルタ設計における特異値の打ち切り次数の合理的決定手法について検討した。

◆細径ワイヤからの受信 RF 信号行列$\mathbf{Y}_0$に、ガウス性白色雑音$\mathbf{N}_k$を重畳させた。特異値の打ち切り次数Rの評価指標 $\beta(R) = E_k[|(\mathbf{Y}_0 + \mathbf{N}_k) - \mathbf{m}|^2$ （$\mathbf{m}$：インパルスのモデルベクトル、$E_k[\cdot]$：kに関する平均操作）より、$\beta(R)$が最小となる最適な特異値の打ち切り次数$\hat{R}$をSN 比ごとに推定し、$(\mathbf{Y}_0 + \mathbf{N}_k)$行列にフィルタを適用した。

◆Fig. 1 に SN 比ごとに算出した$\beta(R)$を示す。SN 比の低下に伴い、最適な打ち切り次数$\hat{R}$が低次にシフトし、Fig. 2 に示すように特異値を$\hat{R}$で打ち切ることにより(Fig. 2(c))、特異値をフルランク用いた際(Fig. 2(b))は観察できなかったインパルスを観測できた。

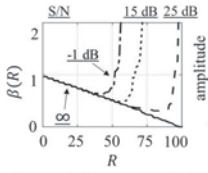


Fig.1：$\beta(R)$ calculated for each SN ratio.

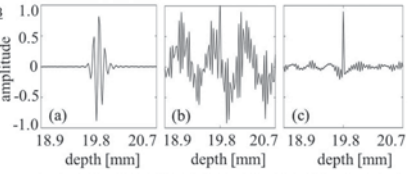


Fig.2：(a) Before applying filter, (b) After applying filter with full singular values, (c) After applying filter with truncated singular values at $\hat{R}$. (S/N: 25 dB).

2-7-17

2-7-17　超音波プローブの押圧に起因する血管断面形状変化を考慮した橈骨動脈壁弾性率推定の理論的検討

Theoretical study on elasticity measurement of radial arterial wall by considering change in cross-sectional shape of vessel from circular shape caused by pushing pressure of ultrasound probe

◎庄司悠人，森 翔平，荒川元孝，
△大庭茂男（東北大），小林和人（本多電子），金井 浩（東北大）

◆我々は血管内皮機能の直接的な評価を目指し，単一超音波プローブを用いて橈骨動脈壁の粘弾性率を計測してきた．しかし，プローブ押圧による血管の変形の影響を考慮できていないという課題があった．本報告では，押圧による血管断面形状の円形から楕円形への変化を仮定し，弾性の推定式を導出し，数値計算によってその妥当性を検討した．

◆図 1(a), (b), (c)はそれぞれ周方向応力の楕円上の位置(楕円の中心からの角度θ)に対する依存性，楕円の断面形状依存性，押圧依存性の計算結果を示す．図 1 より，(心臓の)拡張期から収縮期への増分応力は楕円短軸上で最大であり，血管が潰れるほど大きくなる．さらに，押圧に対しては一定値をとる．以上の結果は過去の in vivo 計測結果と傾向が一致し，導出した式の有効性を示している．

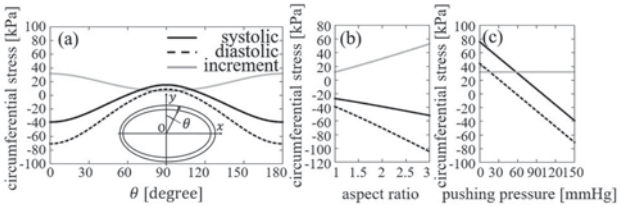


Fig.1：Calculation results of circumferential stress.
(a) positional dependence, (b) shape dependence, (c) pushing pressure dependence.

2-8-1

2-8-1　声帯膜物理モデルの発声実験 —ニホンザル摘出喉頭との比較—

Experimental study on physical model of vocal membranes
-Comparison with excised larynx of macaques-

☆金谷麻由佳，宮﨑琳太郎，植村泰佑，△川端怜，△松本拓磨（立命館大），
△西村剛（京都大），徳田功（立命館大）

◆動物声帯に見られる声帯膜を模擬する物理モデルを構築した。

◆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、声帯膜のある物理モデルと声帯膜のない物理モデルを比較した。

◆マカクザル(ニホンザル *Macaca fuscata*; アカゲザル *Macaca mulatta*)の摘出喉頭でも同様の実験を行い、振動特性を比較した。

◆物理モデルとニホンザル摘出喉頭の両方で声帯膜があることでオンセット圧が減少することが分かった。

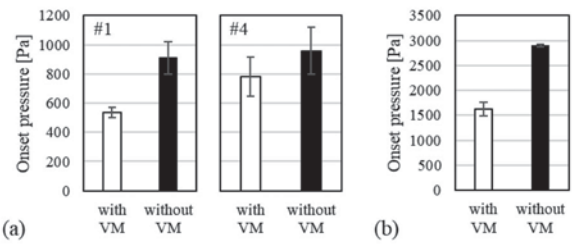


Fig.1：Comparison of vocal fold oscillations with and without vocal membranes (VM). (a) Onset pressures of macaque#1, and #4, and (b) Onset pressures of physical model.

2-8-2

2-8-2　鼻周期や首の角度が鼻音の音響特性に与える影響

Effects of nasal and neck angle on acoustic characteristics of nasals

☆伯田亜海，△加地優太，竹本浩典

◆鼻の粘膜の状態は経時変化し，鼻周期により左右の鼻の通気も変化する。また，首の角度の変化は咽頭の形状や長さを変化させる。これらは，鼻音のスペクトルに影響を与えると考えられる。

◆本研究では，鼻周期による鼻の通気の変化と首の角度の変化を考慮して鼻音を録音し，その音響特性の変化を検討した。

◆その結果，通気の変化によるスペクトル変動には，一定の傾向が見られなかった。一方，首の角度の変化については，上を向くとスペクトルのピークが高周波数領域に，下を向くと低周波数領域にシフトする傾向が見られた。

◆また，咽頭長は上を向くと短縮し，下を向くと伸長する。咽頭長を短縮・伸長した声道形状を用いてシミュレーションを行った結果，伝達関数でもスペクトルと同様の傾向が見られた。

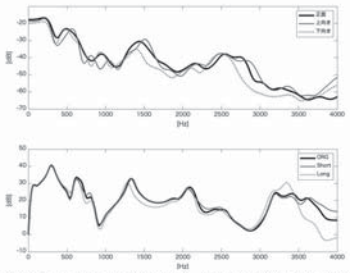


Fig.1：首の角度の変化とスペクトル変動（上），咽頭長の変化と伝達関数の変動（下）

2-8-3

2-8-3　声道内の狭窄と気流がリード式人工声帯の母音生成に及ぼす影響

Effects of airflow in constricted vocal tracts on vowel productions of the reed-type artificial vocal fold

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

◆これまで声帯の音発生を模擬するリード式人工声帯を用いて，声道の形状が母音の音響特性に与える影響について調べられてきた．

◆一方，声道の調音点の狭めにおける気流が声帯振動の特性に影響を与えるという非線形の音響作用が知られている．

◆本研究ではリード式人工声帯に対して，リードの振動，気流と音発生の連成シミュレーションにより自励振動による音の発生を表現し，声帯の後流に狭窄流路を付した円筒型声道を設置することにより，狭窄流路内の気流が発生する音と声帯の振動に与える影響を調べる．

◆シミュレーションにより，声道に狭窄を付した場合において，リードの振動特性への影響が確認された．声道内の流れ場から，どのように影響が与えられたのか解析する．

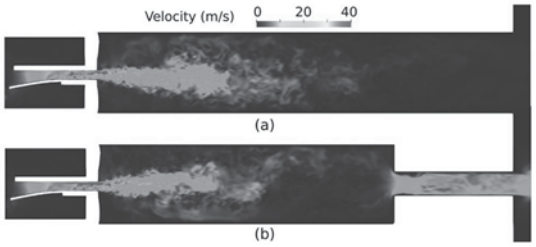


Fig.1：Velocity magnitudes in the artificial vocal fold and cylindrical vocal tract: without constriction (a), with constriction of diameter 10 mm (b).

2-8-4

2-8-4　鼻副鼻腔の模擬手術が伝達関数に及ぼす影響

Effects of simulated nasal sinus surgery on transfer function

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

◆鼻副鼻腔の手術により形状が変化し，術後の音声の個人性に変化が生じる。

◆患者の術前・術後の鼻腔・副鼻腔の形状をCTで計測し，その形状と音響特性の変化をシミュレーションで検討することで，手術による音声の変化の予測に役立てようとしている。

◆本研究では，内視鏡下鼻副鼻腔手術を受けた被験者２名の術前・術後の形状データを用いて，鼻副鼻腔の模擬手術と音響シミュレーションを行い，形状変化が伝達関数に及ぼす影響を検討した。

◆その結果，副鼻腔と自然孔の開放によって，術前には確認されなかった新たなディップが生成され，音圧分布においても術後と近似した連成振動パタンが確認された。

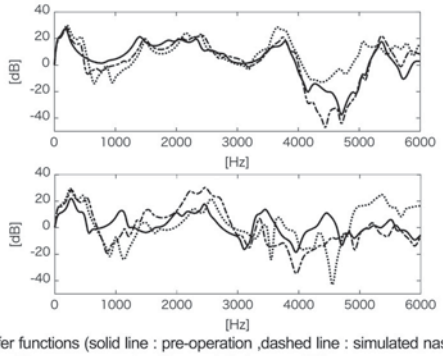


Fig.1：Transfer functions (solid line : pre-operation ,dashed line : simulated nasal sinus surgery, dotted line : post-operation, Upper : M1, Lower : M2)

2-8-5

2-8-5　長母音の発音における時間的な伸長と構音運動の相関性

Correlation between vowel lengthening and articulatory movements in production of long vowels of Japanese

○白勢 彩子（東京学芸大），北村 達也（甲南大），能田由紀子（ATR-P）

目的・概要　日本語の長・短母音のモーラリズムの生成機構を明らかにすることを目的に，磁気式計測システム（EMA）を用いて日本語共通語話者の母音発声時のデータを収集し，実証的に研究を進めてきた。本稿では，解析結果に基づき，生成のモデルを検討したい。

手続　発声リストは，/kabu/（下部）–/kaRbu/（カーブ）（/R/は長音拍）のような母音の短・長の対である。共通語話者を対象に，EMA システムの一つである，Northern Digital 社の Wave Speech Research System (以降，単に「Wave」) を用いて，舌上など５点の構音運動を計測した。

結果　短母音に比して，長母音で動作が大きくなる現象が認められた。図１に，母音アの短母音（「aS」），長母音（「aL」）の結果を示す。図２に，得られた結果を解釈するための生成モデルを提案する。横軸は時間，縦軸は運動の大きさを模式的に示す。A. は，短母音の生成の概念図で，青の図形は短母音の構音動作に対応する。B. はいずれも長母音の生成を示すモデルの候補で，赤の図形が，長音の構音動作に対応する。

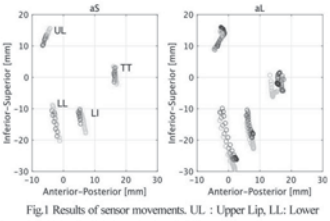


Fig.1 Results of sensor movements. UL : Upper Lip, LL: Lower Lip, LI: Lower Incisor, TT: Tongue Tip.

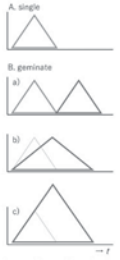


Fig.2 Scheme of vowel lengthening.

2-8-6

2-8-6　逆推定に基づく調音運動の考察 −可視化と分析−

Study of articulatory behavior based on inverse estimation: Visualization and analysis

☆伊東麻沙美，緒方公一（熊本大）

◆声道音響管マッピングインタフェースに基づく声道形状逆推定の研究を進めている。子音を含む音形において，母音から子音部，子音部から母音部への過渡状態の声道形状の振る舞いを把握するため，効果的な可視化などを検討している。

◆本稿では，逆推定結果から得られる声道音響管の擬似速度変化の可視化を取り扱い，そこから見出される運動特徴について考察する。

◆Fig. 1 に，発話 /abi/についての逆推定結果の例を示す。図の下段は，ホルマント周波数から逆推定された結果について，子音/b/直前までのマップ上の推定点を示している。またその右側は，子音/b/直前の逆推定結果を 20 区間（左：正門，右：口唇部）の声道音響管として示している。上段には，各音響管について，直径の変化から算出した疑似速度を，時間の経過に沿って重ねて表示したものである。この発話例では，口唇閉鎖の先行（赤楕円）と，破裂後の口唇部のみの動き（青楕円）が認められる。

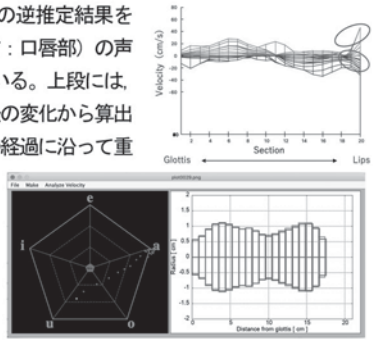


Fig. 1 Pseudo velocity, map, and vocal tract shape for /abi/.

2-8-7

2-8-7　教育用ロボットアームと拡張された時間伸長パルスを用いた音場測定と音声生成研究への応用について

Application of an educational robot arm and a measurement procedure based on an extended time-stretched pulse for acoustic field measurement and speech production research

○河原 英紀（和歌山大）

◆教育用に安価なロボットアームが利用できるようになった。著者らが提案する拡張された時間伸長パルスに基づく同時計測法を利用することで、無響室などの設備を用いずに様々な音響計測を行い、可視化することができるようになった。ここでは、音声の収録に用いられるポップガードの影響を、伝播する音波をムービーとして可視化する事例と、音声生成研究への応用について紹介する。

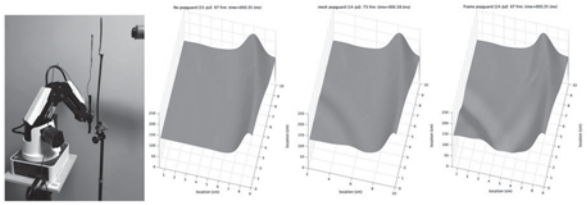


Fig.1 :(Left image) Setting of measuring equipment. From left to right; Robot arm (DOBOT Magician) and mounted omnidirectional miniature microphone (DPA-4066), a mesh-type popguard, and a loudspeaker (AURA SOUND NSW1- 205-8A). (Right three images) Snap shots from movie visualization of sound wave propagation. (Left) without popguard, (Middle) mesh-type popguard, and (Right) screen-type popguard. Middle and right images shows reflected wave fronts.

2-8-8

2-8-8　任意の 3 次元放射方向の人声の再現に必要な角度の分解能 —水平面および正中面における検討—

Angular resolution required for reproduction of speech on the arbitrary radiation direction: Examination in the horizontal and median planes

○杉本岳大，木下光太郎（NHK）

◆AR や VR ではユーザごとにコンテンツ空間内の任意の視聴位置からコンテンツを体験することが前提になる。そこで，発音源とユーザの位置・方向の関係に応じて変化する放射音の特性を再現することで，コンテンツのリアリティや没入感を向上させる手法を開発している。

◆任意の 3 次元放射方向の人声の再現に必要となる角度の分解能については明らかになっていないため，放射方向が異なる人声に関して、違いを知覚できる最小の角度間隔を主観評価実験で導出した。図1に水平面，図2に正中面における角度間隔 180°，90°，45°，30° の評価結果をボックスチャートで示す。

◆実験の結果から，任意の 3 次元放射方向の人声の再現に必要な角度の分解能は，水平面・正中面ともにおよそ 30° と結論付けた。

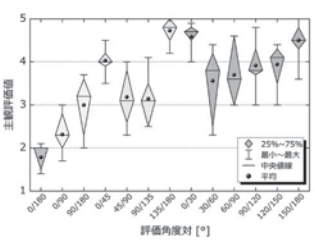


図 1 水平面における角度間隔 180°，90°，45°，30° の評価結果

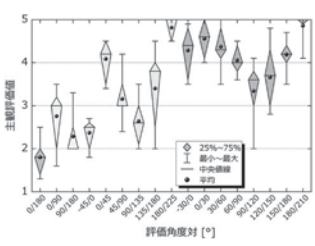


図 2 正中面における角度間隔 180°，90°，45°，30° の評価結果

2-8-9

2-8-9　用例ベース対話システムのための文ベクトルを用いた柔軟な応答生成手法の検討

Sentence representations for example-based dialogue systems to respond naturally

☆矢作凌大，能勢隆，伊藤彰則（東北大学工学研究科）

◆用例ベースの音声対話システムでは，応答を開発者が設計できる一方，入力に対応する応答が固定であるため応答に柔軟性がない。

◆文をベクトルで表現し，加法構成性を持たせることで，用例外のユーザ入力に柔軟に対応する応答生成手法を提案。

◆Transformer オートエンコーダをベースにした高次元の文ベクトル生成モデル “m-query” を提案。

◆提案モデルは従来のオートエンコーダより文の復元精度が非常に高く，生成される文の分散表現には加法構成性がみられたため，提案する応答生成手法の実現可能性が示された。

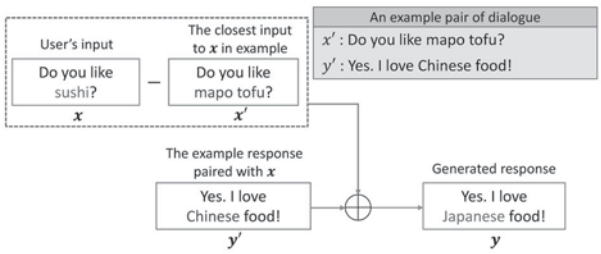


Fig.1 : Abstract of our response generation method

2-8-10

2-8-10　会話支援システムのための簡易な目標発声特定方法

A target utterance identification method
for a practical free conversation support system

☆小坂直人，脇田由実（大阪工大）

◆人同士の会話を支援するために、会話の雰囲気を自動推定し、雰囲気の悪い時に支援を行う会話支援システムの開発を行っている。
◆周辺の会話音声レベルが大きい使用環境で、目標会話音声のみを抽出するために、マイクロホン近くで発声された音声の音響特徴を利用して、目標音声と周辺音声を簡易に識別する方法を提案する。
◆基本周波数と音圧レベルの発声ごとの標準偏差値をパラメータとして、識別器にニューラルネットワークを用いて識別性能を評価した結果、識別率が67.60 %、適合率が79.44 %という結果となった。
◆今後は、学習データの増量と、実際の使用を想定した自動発声切り出し部分の開発を行っていく。

Table 1 Target utterance Identification rates for six speakers

	A	B	C	D	E	F	average
recall[%]	100.00	80.00	70.00	80.00	80.00	10.00	70.00
precision[%]	62.50	80.00	87.50	80.00	66.67	100.00	79.44
f-measure[%]	76.92	80.00	77.78	80.00	72.73	18.18	67.60
accuracy[%]	70.00	80.00	80.00	80.00	70.00	55.00	72.50

2-8-11

2-8-11　Speech laugh が発生しやすい調音に関する音声学的分析

Phonetic characteristics of speech laugh occurrence

○有本泰子，△阿部将士，△津田逸成（千葉工大）

◆背景：
- Speech-laugh（sl）とは笑いながらの発話
- 発話中にどのような声道形状になったら笑いが生じ始めるのか？

◆目的
- sl が発生しやすい声道形状があるかを明らかにする

◆分析内容
- 自発対話音声コーパスに含まれる sl 冒頭の分節音と発話冒頭の分節音の分布に差があるかカイ二乗検定と残差分析により検証

◆結果：　sl 冒頭は…
- 口腔内が狭めである i と e が多い
- 歯茎破裂音 d と t，軟口蓋破裂音 g が多い

◆結論
- 口腔内や喉で声道が狭められたときに sl は発生しやすい

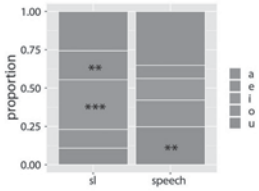


Fig.1: The proportion of vowels at initial speech-laugh and speech

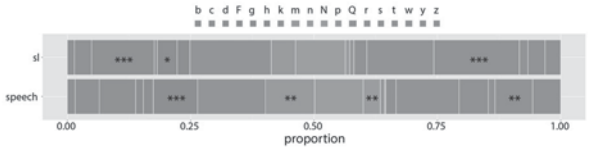


Fig.2: The proportion of consonants at initial speech-laugh and speech

2-8-12

2-8-12　声の高さとイントネーションの大きさの分離と聴覚印象に留意した操作

Separation of voice pitch and intonation:
Voice conversion that pays attention to the auditory impression.

○内田照久（大学入試センター・研究開発部）

問題　音声の基本周波数(f_0)の昇降操作は定数倍による変換が多い
イントネーションの記述に 平均律に基づく cent も用いられる
これらはいずれも 対数軸上での操作と表現

提案　聴覚的知覚に留意するなら，人が同じと感じる音高の差が，どの音域でも同じ値になるように作られた mel 尺度上で f_0 を表現して，それを操作する方が整合的

考察　従来の log 軸上の操作では，聴感上，抑揚の変動幅が拡大
mel 軸上で操作した聴覚実験や作品から 適用の有用性を指摘

課題　平均率に基づくピッチとの関係や，聴覚末梢系での周波数軸表現との整合性などの整理が課題

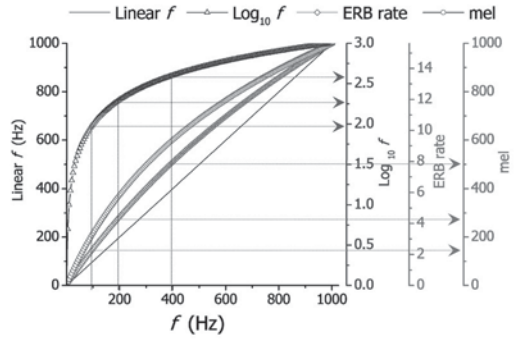


Fig. 1. Relationship between log scale, mel scale, and ERB $_{rate}$ scale.

2-8-13

2-8-13　オペラ歌唱における音高と声道形状に関する検討

Examination of pitch and vocal tract shape in opera singing

☆戸田菜月，竹本浩典（千葉工大），高橋純（大阪芸大）

◆オペラ歌唱のよく響く歌声は，声道形状を制御する特殊な技術を習得することによって実現されている。前報では，歌唱法や音高による形状の相違を検討した。しかし，どのような運動でこれを実現しているかを明らかにすることはできなかった。
◆本研究では，歌唱中の頭頸部を撮像したリアルタイム MRI 動画の全 500 フレームから，機械学習を用いて声道形状の輪郭を点群として抽出して歌唱に伴う声道形状の変動の主成分分析を行った。
◆抽出した第１～３主成分（PC1～PC3）が示す声道形状の変動と主成分スコアの時間変化から，各成分はそれぞれ，/a/の口の構え，音高の制御，吸気に伴う運動に対応づけられることが明らかになった。
◆喉頭の位置など歌唱法の違いを特徴づける運動は，複数の主成分にまたがって表れていたため，各主成分にそのスコアを乗じて加算することによって，声楽的な発声と非声楽的な発声による声道形状の違いを再構成することができた。

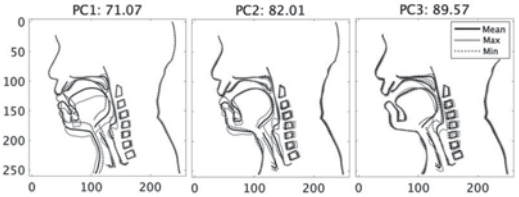


Fig.1：Results of principal component analysis of vocal tract shape during singing
(black: mean, gray: maximum score, dotted: minimum score)

2-8-14

2-8-14　人はなぜハミングを行うのか
—日本語話者における使用状況のアンケート調査—

Why do people hum?
-Questionnaire analysis of the usage status among Japanese speakers-

○浅野恵子（順天大・医）

◆ハミング音は個別他言語の域を超えて理解できなくても普遍的に発せられる特殊発声音である。日本語では「鼻歌」と言われる。

◆同じ音声的及び音響的特徴をもちながら、文化や気候風土によって変化するものがあり、音声行動としては無意識の場合が少なくない。

◆ハミング実施について日本人の20代から60代以上の年代ごとにアンケート調査を男女合計770名に実施した。

◆どの年代も女性の方が行うと回答した数が多かった。

◆6 つの質問：①ハミング行動の有無、②実施頻度、③実施時の感情、④実施時・場所、⑤実施理由、⑥利点と欠点を質・量的にアンケート

◆⑤の実施理由の結果は自由記述してもらった回答からキーワードを抽出し頻度順に図に示している。

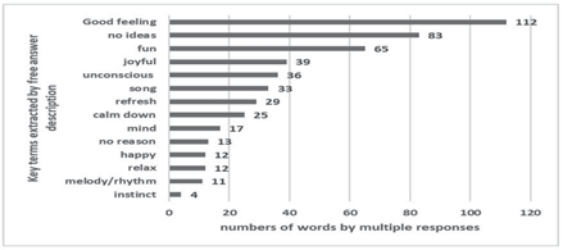


Fig.：Key terms extracted by free answer response in the question 5; “Why do you hum?”

2-8-15

2-8-15　東京方言においてアクセント型が後続助詞の持続時間に与える影響

The effect of accent on the duration of the following particle in Tokyo Japanese

○杉山由希子（慶應大），荒井隆行（上智大）

東京方言話者に、音素配列が同じでアクセント型のみが異なる、2 モーラの尾高型(final-accented)と平板型(unaccented)の最小対を「この単語は＿だ」と「この単語は＿と言う」という２種類のキャリア文に入れて発話してもらった。日本語ピッチアクセントの特徴は、アクセントのあるモーラから後続するモーラに現れることより、テスト語に続く助詞(/da/と/to/）の持続時間を計測したところ、発話末の助詞においてのみ、平板型に続く方が、尾高型に続くよりも持続時間が著しく長かった(Fig. 1)。この結果は、先行する単語のアクセント情報を後続助詞の持続時間で実現した現象と解釈することができ、モーラの等時性よりもアクセント情報の伝達を優先した accentual boost の一種と捉えられる。

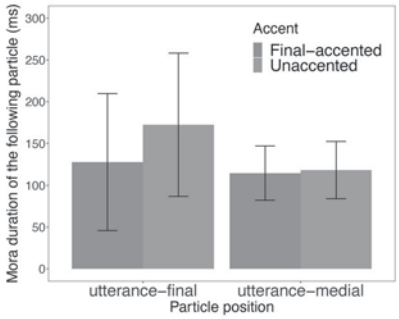


Fig.1　Particle duration in utterance final and word-medial positions following test words

2-9-1

2-9-1　「懐かしさ」を喚起する楽曲の音楽的要因の分析

Investigation of musical factors for music with a nostalgia.

○根木 心大，河合 熊輔，石川 大，中島 弘史（工学院大・情報）

◆本研究では，楽曲またコード進行印象調査を行い，懐かしさ感情を喚起させる音楽的要因を検討した。楽曲印象調査より，“夏”に関連する自伝的記憶が想起されやすいこと、また楽曲聴取時に「懐かしい体験」を想起した場合において、「おかしさ」の評価値が低くなることが示された(Fig.1)。懐かしさ感情を与える音楽的要因は，カントリー調のアコースティックギター，遅いテンポ（80～85 BPM）でのオルガン，シンセサイザーの演奏，レコード音源に含まれるノイズであることが明らかになった。コード進行印象調査より，下降系のセブンスコード進行が懐かしさを与える要因として挙げられた。

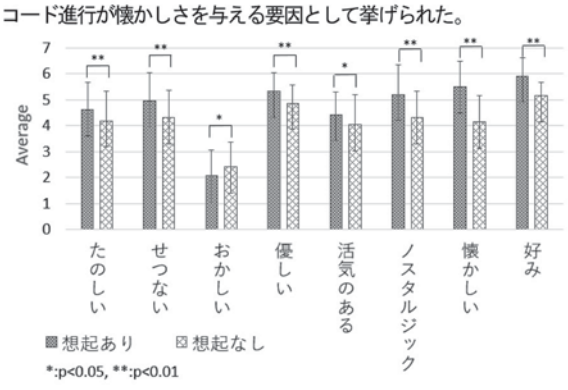


Fig.1：Evaluation values of music in nostalgic recollections.

2-9-2

2-9-2　メロディー輪郭の変化に対するミスマッチ・ネガティビティ：輪郭形状及び被験者の音楽経験の影響

Measurement of Mismatch Negativity for Changes of Melody Contour: Effects of Contour Shape and Music Training

☆内藤柚菜（千葉大），柴玲子（慶應大），大塚翔（千葉大），湯本真人（東京大），中川誠司（千葉大）

◆音楽経験者と非音楽経験者を対象に，メロディーの輪郭（連続する音列の音高の変化パターン）の変化に対するミスマッチ陰性電位（MMN）を計測した．

◆5 つの連続するトーンバーストから構成される音列を刺激音とし，各構成音の音高を変化させて山形/谷形/上昇形/下降形の輪郭を持つ４つの刺激音を作成した．これらの刺激音を標準刺激とし，各刺激音列の第 3 音を変化させることでそれぞれに 7 種類の逸脱刺激を作成した．逸脱音列に対して誘発される MMN を観察した．

◆その結果，音楽経験者ではメロディーの輪郭形状の複雑さに応じて MMN 振幅の出現パターンが変化する傾向が確認された．MMN 潜時は輪郭形状に依存した変化は認められなかった．

◆非音楽経験者においては，山形に対する MMN が最も大きかったが，輪郭に依存したMMNの出現パターンを見いだすことは出来なかった．MMN 潜時は輪郭形状の複雑さに応じて変化した．

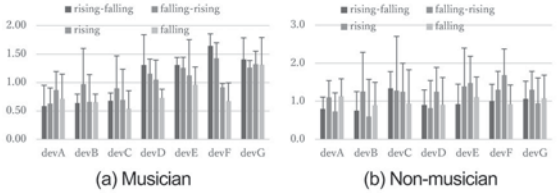


Fig.1：Peak amplitudes of MMNs

2-9-3

拍時刻推定のための オンセット時刻の知覚に対する検討 ~演奏音を構成する楽器数に関する調査~

Perception of onset time for beat time estimation ~Effect on the number of played instruments~

◎桶本まどか（国音大・院），三浦雅展（国音大）

◆本報告では，オンセット知覚において演奏音に含まれる楽器の数が変化した場合，その知覚がどのように変化するのかについて検討した．

◆ラベル作曲のボレロの演奏を用いて４名の被験者を対象として，クリック音を付与した演奏音について，そのクリック音が演奏音に対しずれているか，そろっているかを回答させることで，オンセット知覚に対する調査を行なった．図１に用いた３つの演奏音のエンヴェロープを示す．図１は，上から演奏音に含まれる楽器の数が少ない，普通，多い演奏音のエンヴェロープである．

◆その結果，４人の被験者のオンセット知覚の時間的許容幅については統一的な結果を確認することはできなかったものの，被験者の演奏経験がオンセット知覚に影響を及ぼす可能性が示唆された．

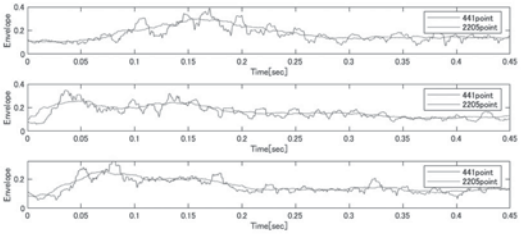


図１　用いた演奏音のエンヴェロープ

2-9-4

2-9-4　モーションキャプチャを用いた感情表現による複数のピアニストの演奏動作解析*

* Performance motion analysis of pianists by emotional expression using motion capture

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

我々は，モーションキャプチャシステムを用いた鍵盤楽器の演奏動作の計測ならびに分析を行ってきた．複数の演奏者を比較することで演奏動作の傾向が把握できるのではないかと考え，ピアニスト３名の感情表現に対する演奏動作を計測し，その演奏動作の分析を行った．今回，ピアニストの左右の手首，肘，肩の関節部位の動きに着目し，鉛直方向の速度の二乗平均平方根(root mean square, RMS) を用いて分析を行い，感情表現に対する演奏動作について考察を行った．

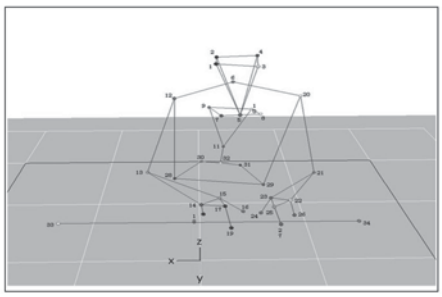


Fig. 1 Marker adhesion position

2-9-5

2-9-5 位相の混合微分に基づく調波打撃音分離

Harmonic and percussive sound separation based on mixed partial derivative of phase spectrogram

☆赤石夏輝，矢田部浩平，及川靖広　（早大理工）

◆従来法では，振幅スペクトログラムの特徴に注目

→　位相の情報を積極的に利用するものはなかった

◆提案法：**位相の混合微分**に基づく時間周波数マスキング

・位相の時間偏微分（**瞬時周波数**）の周波数偏微分

・位相の周波数偏微分（**群遅延**）の時間偏微分

→ それぞれによって調波音と打撃音の特徴付けが可能

◆提案法は，振幅のみの従来法に比べて打撃音の分離が得意

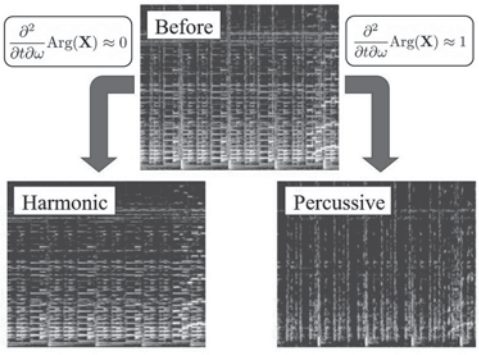


Fig 1. An example of separated power spectrogram

2-9-6

2-9-6　深層学習と非負値行列因子分解を利用した楽曲からの高精度楽器音分離

Accurate instrumental sound separation from music by using deep learning and non-negative matrix factorization

☆前田梨江，星野博之（愛知工業大学）

◆ 目的　複数の楽器音で構成された楽曲から個々の楽器音を抽出することを目的とした。非負値行列因子分解(NMF)と深層学習(DNN)を使用した高精度な音源分離における自動化と分解性能向上を行う。

◆ 方法　ある楽曲の約 11 秒間を切り取り、NMF による分割音がどの楽器音かを判別し各楽器音に合成する。NMF の基底数は 10, 20, 30 の 3 通りである。今回、聴感により行っていた楽器音判別を DNN で行うようにした。DNN のテストデータは聴感で楽器音ごとに振り分けた対象楽曲の分割音で、学習データは各楽器（ピアノ、コントラバス、ドラム）の奏法・音高・強弱の違う楽器単独音である。NMF を用いて楽器単独音を分解する際の基底数を 12 通り試して学習した結果、最適な基底数は 5 であった。

◆ 結果　Fig. 1 に楽器単独音を基底数 5 で学習した DNN を用いたときの対象楽曲の各楽器音の分離再合成波形を示す。例えばドラムでは不要な音が減っており、聴感では正しく判断が出来なかった音が DNN を用いると判別出来たことが分かる。

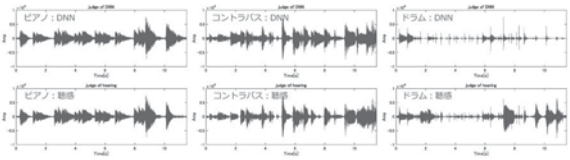


Fig. 1 Results of evaluation of sound separation by using DNN and hearing

2-9-7

音響的特徴に着目した採譜精度改善手法

Improvement of Music Transcription Accuracy Based on Acoustic Features

○黒川達也，上岡英史，金丸真奈美（芝浦工大）

◆ 自動採譜とは音声デジタルデータから楽譜を自動で生成するものである。既存研究では、データ駆動型の機械学習を用いることが多い。しかしながら、データ駆動型のアルゴリズムで求められる入力データの多様性は楽曲の著作権保護と相反する。また、既存研究ではピアノロールという楽譜形式を推定することが多い。既存研究のピアノロールでは、一般的な楽譜が有する拍子、テンポ、小節線などの音楽的要素が含まれていない。これらの情報がない場合、演奏者は時刻、および速度管理をすることが困難になる。

◆ 本研究ではデータ駆動型機械学習を使用せず、ピアノロールに不足している音楽的要素を過不足なく含む自動採譜の実現を目的する。本稿では BPM を導出するまでを扱う。

◆ 本研究において、ドラムセットの発音間隔を k-means++でクラスタリングを行い、クラスタの重心値となる発音間隔から BPM を推定する手法を提案する。k-means++は非階層型クラスタリングであり、教師なし機械学習に分類されるため、データ駆動を伴わない。よって、本研究では k-means++を使うことで楽曲の多様性がなくともクラスタリングによって BPM の推定を可能としている。

◆ 検証実験としてドラム音源を含んだ wav ファイルを 4 つ使用し BPM を導出した。その結果、全ての楽曲において確定値と等しい BPM の値を導出することに成功している。

2-9-8

2-9-8　微分可能な信号処理モジュールを用いたシンセサイザーによる楽音マッチング

Synthesizer Sound Matching with Differentiable DSP

☆増田尚建，齋藤大輔（東大）

問題

シンセサイザーの操作は複雑で難しい。

パラメータ（ツマミ）をどう設定すれば出したい音が得られるか？

楽音マッチング

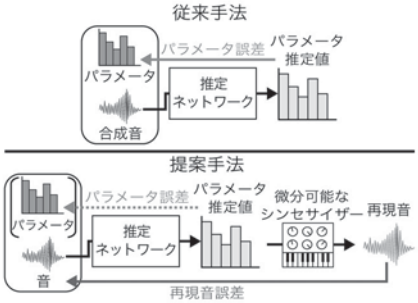
深層学習などによって、ある音を再現する合成パラメータを推定。

提案手法

微分可能な信号処理モジュールによってシンセサイザーを実装。

深層学習フレームワークに直接シンセサイザーを組み込み、出力音の実際の再現精度に関する誤差を最適化。

→主観評価によってマッチングの品質向上を確認。



2-9-9

2-9-9　楽器音の倍音列における位相差の理論的説明

Theoretical explanation of phase differences of harmonics of a sound of instruments

○三谷尚，△宮本百音、△高尾侑未、△新津充、△中畠瑞希（福岡教育大学）

■木管を主とする種々の音高の音圧関数において、各高調波（倍音列の）フーリエ成分における振幅比と位相差を測定し，さらに位相差の理論的説明を行った。測定によれば、クラリネットでは奇数倍音に加え、偶数倍音が微弱ながら同時存在、位相差は倍音列と共に振動的、オーボエ、フルート等の位相差の傾向も求められた。（測定結果は過去の研究会等でほぼ発表済み。）

■クラリネットにおける奇数偶数倍音の存在から、奇数偶数倍音を有するモデル楽器（範囲 $x = 0 \sim x = L$）を想定し、倍音列ごとの定在波を考えた。ここで、奏鳴開始時刻 $t = 0$以降、閉管条件（$\tilde{n} = 1, 3, \cdots$）では $x > 0$ での変位が増大し、開管条件（$\tilde{n} = 2, 4, \cdots$）では $x = 0$ の変位が増大すると考えた（Fig. 1）。これに連動して、定在波の挙動により、x=L での変位の振動の時間依存性を観察。偶数倍音のみの位相差をピックアップ（Fig. 2）すると、フルートの位相差の実測値と良い一致を見た。

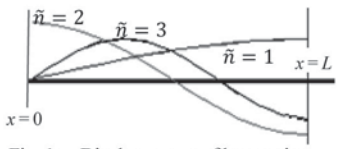


Fig. 1:　Displacements of harmonics

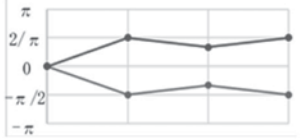


Fig. 2 :　Phase differences of even harmonics

2-10-1

2-10-1　スポーツにおける音楽・リズムの活用と今後の可能性

Utilization of music and rhythm in sports and future possibilities

○伊坂忠夫　△正田 悠　△菅 唯志（立命館大学）

スポーツならびに日常の運動やエクササイズにおいても、「音楽」は伴奏（伴走）する友であり、極めて身近な存在である。本報告では、身体運動に関わり、音楽の活用が認知科学や知覚心理学的にどのような効果をもたらすのか、さらに、音楽におけるシンクロニーについての現状の研究成果を述べた上で、スポーツ分野における音楽・リズムの活用の可能性と展望について著者らのアイデアを示す。

2-10-2

2-10-2 モータースポーツと音

Motorsports and Sound

△地頭所光（東京大学大学院），上田麻里（神奈川工科大）

現役レーシングドライバーの視点から、自動車レースにおいてドライバーがどのような聴覚情報を頼りに走行しているかを定性的にまとめた。さらに今後の次世代自動車によるモータースポーツの課題を述べた。

2-10-3

2-10-3 野球中継動画における打球音の音響的特徴と印象評価

Acoustic Characteristics of Hit Ball Sound in Baseball Live Video

☆磯川 昇汰，二村 亮平，新田 晃司（神奈川工大），三浦 貴大（産総研），春日 秀雄，上田 麻理（神奈川工大）

◆野球の試合中には，選手の掛け声や観客の歓声以外に，打球音や捕球音が発生する．そのような音は選手にとって大事な判断材料とされる．

◆これまでに我々は，野球選手においてアンビエントな音や競技中の音における利用や注意意識に関するアンケート調査を行うなどして，野球における競技力向上のための聴覚利用の可能性について解明を進めてきた．しかし，主な対象者はアマチュア選手であった．

◆本研究では，プロ野球の試合における打球音の音響解析を行い，強打における音の違いを明らかにすることを目指す．

◆このため，ウェブ上に公開されているダイジェスト版の動画から打球音を抽出の上で，周波数成分に関して分析を行った．

Fig.1：Schematic image of a baseball player just before swinging

2-10-4

2-10-4 野球競技における打球音の音響学的特徴 －守備位置ごとの違い－

Acoustic characteristics of ball hit sounds in a baseball game: Differences in fielding position

☆二村亮平，△磯川昇汰，△新田晃司（神奈川工科大）
三浦貴大（産総研），上田麻理（神奈川工科大）

野球において競技者は，ボールの音や人の声など様々な音を基に次の身体動作決定の他，ボールの挙動の予測を行う．先行研究より，聴覚情報は競技力向上に欠かせない重要な情報源であると考えられる．競技中は視覚情報のみならず周囲の声や音にも無意識に注意を向けているようだ．

本研究では引き続き，野球競技において聴覚情報を最大限に活用し打球音・捕球音などから実用的に競技力の向上を図るための科学的知見を得ることを目的とする．本稿では，野球競技における守備位置の違いに着目し，打球音の音響的特徴を明らかにするための音響計測を行った．

2-10-5

2-10-5 バレーボール競技における聴覚情報の利用に関する実験的検討 ～聴覚遮断によるプレーへの影響～

An Experimental Study on the Use of Auditory Information in Volleyball -Effect of hearing deprivation on play-

☆髙橋隼人（神奈川工科大），三浦貴大（産総研），上田麻理（神奈川工科大）

＜概要＞
バレーボール競技において人間は、絶えず連携を取りながら一瞬の間に様々な選択を行う。その選択において重要となる感覚情報の一つとして、チームプレイとしての声での連携、打球音の強さの把握や、足音、声などによる位置情報の確認を行う為の聴覚情報がある。本研究では、バレーボール競技における聴覚情報の有効性を明らかにした上で、バレーボール技術向上に音響学的側面からの貢献可能性を探る。

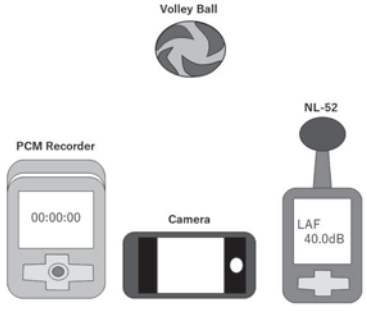


Fig 1. Overview of the volleyball experiment

2-10-6

2-10-6　スポーツ音響への応用に適したアンサンブル平均を用いた吸音特性の現場測定法 ―体育館における現場測定（その 2）

An in-situ measurement method using ensemble averaging technique for sound absorption of materials installed in athletic fields
- Applications of the EA method to materials in realistic gymnasiums Part-2

○大鶴 徹，富来礼次，岡本則子，△ラワワンデークル シワット（大分大・理工）

◆EA 法を用い、大分市内２箇所の体育施設メインアリーナで、鉛直壁面の吸音特性を試行的に測定した。

◆コロナ禍のため、測定時間等に大きな制約があった。。

◆Gym-X（公共体育館）の測定では、一般客がバドミントンをプレー中、かつ、空調機が通常稼働中であった。

◆4 種の壁面：コンクリート壁（反射性）、木板（板振動型）、吸音性壁（多孔質型）、穿孔板開口部（共鳴器型）を測定。

◆概ね文献値に近似の吸音率、並びに、4 種それぞれの吸音機構の特徴を反映した周波数特性が得られた（Fig.1）。

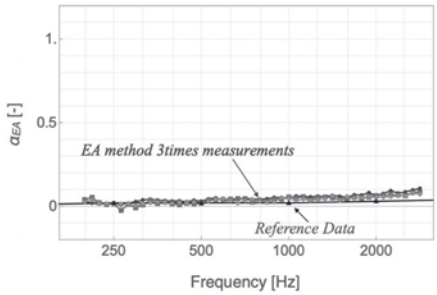


Fig.1：One-twelfth octave band average values of α_{EA} of reflective concrete wall measured three times by EA method at Gym-OU.

2-10-7

2-10-7　スピーカの性能の違いが運動表現に与える影響その 1

Effects of differences in speaker performance on sports performance
-Behavior analysis using OpenPose-

◎三春知史（神奈川工科大），松石遼太（TOA），春日秀雄，上田麻理（神奈川工科大）

スピーカの音質評価の多くは，物理評価や主観評価による評価方法をとっている。物理評価では音圧のピークや立ち上がり速度などに着目した評価方法をとっている。また，心理評価では音質評価語を用いた主観評価実験にて音質評価を行っている。しかしながら，音響分野の専門の知識がない人々にとって，こうした評価方法からスピーカの性能の良し悪しを判断することは難しい。そこで，本研究ではスピーカの音質の違いが運動動作に影響を与えるかどうかを，OpenPose を使用した動作解析によって明らかにすることで，客観的にスピーカの音質評価を可能にすることを目的とする。

Fig.1 Captured image at OpenPose runtime.

2-10-8

e スポーツと地域連携

e-sports　Regional　cooperation

○志和 康弘　小野 康佑（NTT 東日本）

◆近年，e スポーツが世界的に普及し始めている。 国内外の市場も急速に伸びており、2019 年には国体において史上初となる e スポーツ大会も開催された。COVID-19 の影響でフィジカルスポーツの活動が制限される中、オンラインで実施可能な e スポーツはますます注目を浴びており、オリンピックの事前プログラムにも e スポーツが採用された。地域からの問い合わせも多く、このような社会的ニーズに応えるため，NTT 東日本グループでは「株式会社 NTTe-Sports」を立ち上げる運びとなった。そこで本稿では，「e スポーツ」の概要とともに「e スポーツ」が地域連携でどのように広がりを見せているのかを紹介するとともに，神奈川工科大学をはじめとする各学術機関との e スポーツのかかわり方についても述べる。

◆「NTTe-Sports」は施設事業や教育、イベントソリューション等を幅広く展開している。その中でも、特に自治体も含めた地域連携に力を入れており、神奈川県横須賀市とは連携協定も締結し観光周遊や新ジャンルスポーツへのチャレンジ等様々な分野での取り組みを中長期に行っている（内閣府の地方創生推進交付金も採択）。さらに、大学（神奈川工科大学、筑波大学）やプロスポーツチーム（横浜Ｆ・マリノス）の連携など様々な分野の団体とのアライアンスを推進している。今後の e スポーツの展望に関しても紹介していく。

2-10-9

2-10-9　e スポーツにおけるレーサー目線での聴覚情報の有効性に関する検討その 1

Basic study on the effectiveness of auditory information in e-sports from the perspective of the racer part 1.

☆中田翔斗、△眞嶋紘太郎、日向昭吾（神奈川工科大）、地頭所光（東京大学大院）、岩田一、瀬林克啓、 丸山充、 塩川茂樹、 上田麻理（神奈川工科大）

◆本研究では e スポーツおける聴覚情報を遮断する事でどの様に行動・心理面で影響が出るのかをレーシングゲームを使い聴感評価実験を行った。

◆その結果、経験者は聴覚情報あり条件でラップタイムが最も速く、未経験者は聴覚情報あり条件で最も遅い結果となった。

Figure: ドライビングシミュレータによる実験の様子

2-10-10

2-10-10 リズムゲームにおける基礎的考察:その１ e-sports のネットワーク遅延と音の影響

Basic consideration in the rhythm game : Part 1
- A relevance of the network delay and the sounds in e-sports -

〇岩田 一，△青柳 和希，中田 翔斗，△瀬林 克啓，△丸山 充，
△塩川 茂樹，上田 麻理（神奈川工科大）

コンピューターゲームを利用したスポーツ競技である e スポーツが，COVID-19 の影響によりフィジカルスポーツの活動が制限されることに伴い，注目が高まっている．e スポーツの競技対象として様々な種類のコンピューターゲームが取り扱われているが，本研究では聴覚および音響が直接ゲーム性に関連するリズムゲームに着目する．リズムゲームの e スポーツにおける展開としては，リアルタイムによるネットワークを利用した大会の実施が期待されている．

本稿では，リズムゲームに影響があると考えられる，ネットワークの遅延と音の影響について基礎的な実験の結果を考察し，解決すべき課題を明らかにすることを目的とする．リモートアクセスによる画面表示遅延のリズムゲームへの影響を検証するため，音楽あり，消音でプレイした結果と，リモートデスクトップを経由したプレイ結果および被験者へのアンケートから考察した．

今回の実験結果により，音声の有無よりもネットワーク遅延の違和感が大きいことがうかがえた．今後の課題として，ネットワーク遅延の影響を許容できる範囲の確認，ゲームの習熟度，難易度の高低，入力デバイスの種類による影響など，実際の e スポーツ大会の環境に近い状況についての検証などが挙げられる．

2-10-11

2-10-11 野球における歓声等の環境音が選手に与える影響 その 2 ー無観客試合での音響 AR とタイミングー

Effects of environmental sounds such as cheers in baseball on players
-Acoustic AR and timing in unattended games-

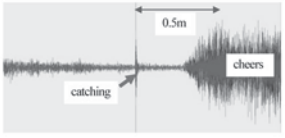
〇春日秀雄，磯川昇汰，二村亮平，△新田晃司，△塩川茂樹，
田中哲雄，上田麻理（神奈川工科大）

◆新型コロナウイルス流行の影響で、様々なスポーツで無観客試合が実施された。先行研究では、観客の有無は選手のプレーに直接的な影響を与えるものではないが、**観客の声援が選手のモチベーションに影響を与えることを明らかにした**。

◆神奈川工科大学の KAIT スタジアムに歓声等の環境音を発生させる音響 AR 環境を構築したが、「タイミングに違和感がある」という評価が得られた。そこで、本研究では**音響 AR の臨場感の向上**を目的に、**環境音の発生タイミングについて調査**を行った。

◆応援や歓声といった声、拍手などの環境音が、どのような状況で発生するかを 2019 年以前の試合映像から調べ、そのタイミングを音声データから分析した。

◆今回の研究で、歓声が上がるタイミングというものが、状況確定から 0.5 秒程度経過した後という事が分かった。このタイミングの**遅延を抑えることが音響 AR の臨場感の向上につながる**ことが判明した。



2-11-1

2-11-1 サウンドスケープからデザインを考える

What is Soundscape Design?

鳥越けい子（青山学院大学 総合文化政策学部）

◆豊かな聴取文化実現のために生み出された（デザインされた）のが「サウンドスケープ」という言葉である。

◆サウンドスケープは、いくつかのデザイン概念を内包している。

◆サウンドスケープ・デザインとは単なる「音のデザイン」ではない。音の世界を通じて環境との豊かな関係を取り結ぶことである。

◆『世界の調律』(1977)には、最近知られるようになった「人新世」的視座がある。その論点に、今ようやく時代が追いついてきた！

◆環境の内実は実に複雑で奥深い。「音の世界」もまた同様である。サウンドスケープ概念の特色とその存在意義のひとつは、異なる学問分野の知見や技術を統合するところにある。

◆「サウンドスケープ研究の本領はさしずめ、科学、社会、芸術の中間領域ということになろう。」by R.M.Schafer

◆『音響生態学ハンドブック』(1978)におけるサウンドスケープの定義に多様な要素が含まれているのはそのためである。

◆ISO によるサウンドスケープの定義からは物理的な音が排除されている。それは何故か？

◆サウンドスケープからデザインを考え、デザインからサウンドスケープを考える。

2-11-2

サウンドスケープデザインと都市

Case study of “soundscape design” in urban environment

〇鷲野宏（都市楽師プロジェクト）

◆都市におけるサウンドスケープデザインの実践には、都市環境を聴覚文化的に捉える研究的・教育的フェーズ、芸術的・祝祭的表現を行うフェーズとともに、日常的・社会的側面から都市環境デザインを行うフェーズが相互に関係しながらある。

◆1960 年代末のサウンドスケープ概念の提唱以前にも、サウンドスケープデザインの考え方と通底する実践は歴史の中に多く見受けられるので、時代横断的にこれ等の事象を観察することは意識的な実践に役立つものと思う。取り上げる事例は以下の通り。

- 上野寛永寺の時の鐘の可聴範囲
- 柴又帝釈天参道の賑わい
- 渋谷スクランブル交差点の喧騒
- 神田川の聖橋エリアの鉄道走行音
- おゆみ野ニュータウンでの歩車分離
- 代官山ヒルサイドテラスのパブリックスペース

◆本項では、「界隈」や「結界」などの感覚に結びつくものとして、聴覚的要素が、場の「らしさ」の醸成に効果をあげていると認められる幾つかの都市環境について、特に、現代都市の全体を均質的に覆っている車の走行音が作りだす雑踏との聴覚的な距離の取り方に焦点を当てながら、それぞれの要素の配置関係を図化・分析し、都市部におけるサウンドスケープデザインの意義や可能性について考える。

2-11-3

2-11-3　書式と時間を超えた新たなオンガクの創生 R.マリー・シェーファーのサウンド・エデュケーション

Difference and Deferral in Designing Music:
Of R. Murray Schafer's Sound Education

○今田匡彦（弘前大学）

西洋近代は，和声，対位法などのエクリチュール（書式）を専門的に学んだ作曲家の手になる作品を，音楽と考えた。本発表では，従来の西洋音楽が自明とした書式，時間の概念を，音環境を基盤とした子どもたちによる新たなオンガクの創生により脱構築しようとしたカナダの作曲家R.マリー・シェーファーのサウンドスケープ思想，及びサウンド・エデュケーションについて，以下を踏まえて検証する。西洋の調性音楽は，プラトンやアリストテレス以来，音そのものの形式としての音楽と修辞的内容として言語により説明される音楽の 2 つに分離されてきた。西洋哲学，美学が依拠してきた修辞学的内容の習慣は，20 世紀に入ると構造主義，ポスト構造主義により脱構築される。言語はシニフィアンとシニフィエの恣意的な関係による記号であり，差異のシステムであるとしたソシュール，ソシュール言語学を更に修正することで，言語の余剰性，曖昧性を指摘したデリダは，言語と音楽との関係を検証する上で重要な役割を果たす。このような，構造主義者，ポスト構造主義者たちが脱構築した言語の網目「書式」による小世界，直観，意味，現前と続く量的時間の流れを，シェーファーはサウンド・エデュケーションによって軽々と乗り越え，戯れた，という痕跡である。このオルタナティヴな実践を，これまで聞かれることのなかった，奏でられることのなかった新たなオンガクの創生と考える。本発表は以上のディスコースに貢献するものである。

2-11-4

2-11-4　サウンドスケープの記述に関する方法論の展開

Development of Methodologies for Notation of Soundscapes

○堤遼（日建総研），佐久間哲哉（東大・工）

◆これまでサウンドスケープ記述は、定性的な記述に始まり、次第に定量化され、記述対象範囲が広域化する方向に展開されてきた。筆者らはICT 技術・携帯端末を援用した記述法を提案し、**サウンドスケープの重層的な記述法**へと新たに展開した。

◆第一に、携帯端末を援用し位置情報と紐づいた音環境の意味的側面を多面的に記録する**自由行動にも対応した記録方法**を開発した。今後、街歩きやサウンドウオークへの活用などが期待される。

◆第二に、個人の音環境体験の視覚化、**データ蓄積による重層的な面的記述**を試みた。地域分析・評価への応用が期待される。Fig.1 に面的記述の一例を示す。

◆提案手法の拡張性として、携帯端末の機能を使用し、音圧レベルなどの**物理計測を行い意味的データと組合せることも可能**である。

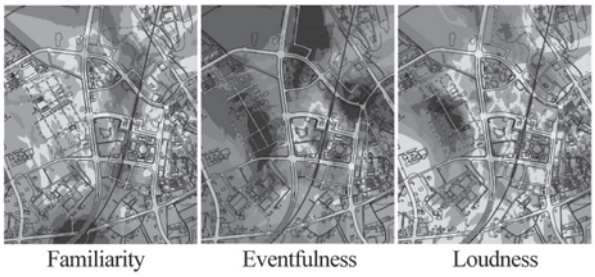


Fig.1 : Maps of Superimposed Sound Environment Experiences

2-11-5

2-11-5　幼稚園の音環境デザイン・プロジェクト

-環境音楽を用いたアクションリサーチを通して-

Sound Environment Design Project for Kindergarten

-Through action research using ambient music-

○小松正史（京都精華大学メディア表現学部）

◆本稿では、学校法人菊の花学園 菊の花幼稚園（京都市左京区上高野）を対象にした学校空間の音環境デザインの成果を報告する。

◆園全体に落ち着きや安心感を提供する楽曲づくりを試みた。

◆環境音楽の導入によって、静的な心理状態を促進させ、園内に穏やかさと落ち着きをもたらせる可能性が確かめられた。

◆制作した環境音楽が、園の環境にマッチしていることや、現場で好意的に活用されている状況を確認することができた。

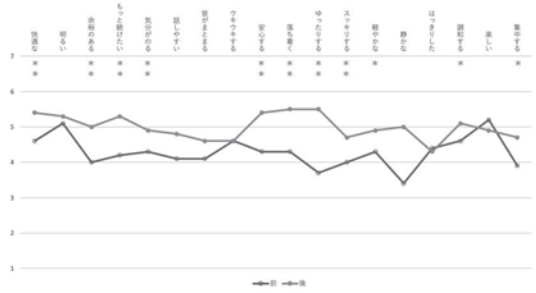


Fig. Profile chart before and after the introduction of ambient music

2-1P-1

2-1P-1　独立深層学習行列分析を用いたランク制約付き空間共分散行列推定による音声強調

Speech enhancement using rank-constrained spatial covariance matrix estimation with independent deeply learned matrix analysis

☆三澤颯大，中村友彦，高宗典玄（東京大学），
北村大地（香川高専），猿渡洋（東京大学）

◆ランク制約付き空間共分散推定（RCSCME）[Kubo+, 2019]は，拡散性雑音中の指向性目的音声に対して，ブラインド音源分離手法である独立低ランク行列分析（ILRMA）で推定するランクが欠落した空間共分散行列の欠落ランク補完に基づく手法である。

◆本稿では，RCSCMEの前段処理のILRMAを教師あり音源分離手法である独立深層学習行列分析（IDLMA）に置き換える手法を提案する。音声強調のためのIDLMAでは各音源を推定する学習済み深層ニューラルネットワーク(DNN)を必要とするため，代表的な音声強調DNNであるDenoiserを用いて音声と雑音を推定する手法を提案する。

◆シミュレーション実験によって，提案手法が従来手法より高い分離性能を持つことを確認する。

Table 1 The largest SDR improvements [dB] among the iterations

	Noise			
	Babble	Cafe	Station	Traffic
ILRMA	6.97	6.20	5.58	6.04
Denoiser	6.69	8.86	9.62	10.14
Conventional RCSCME	7.94	8.85	8.82	8.49
IDLMA	6.53	7.68	7.81	7.77
Proposed RCSCME	**8.39**	**10.10**	**10.19**	**10.21**

2-1P-2

2-1P-2　定誤警報確率により作成したマスクによる距離スペクトルの偽ピーク除去

False Peak Removal in Range Spectrum by Masking Using Constant False Alarm Rate

☆山門　樹，篠原　寿広，上保　徹志，中迫　昇（近畿大）

◆対象物までの距離情報は様々な応用が期待されているため，多くの距離センシング技術が提案されている．

◆至近距離から遠距離にある対象物の距離推定を正確に行う手法として，送信波と反射波の位相干渉に基づいた音響測距法(ADM)が提案されており，一般的に 1ch あるいは 2ch マイクロホンを用いる．

◆1chADM では測定系影響によって 0m 付近に偽ピークが生じ，2chADM ではクロススペクトル法による歪の影響で真のピークより遠距離側に偽ピークが生じることがある．

◆本研究では，偽ピークの除去によって近接 2ch マイクロホンにおける ADM の測定性能を高めることを目的とする．

◆本稿では，試みとして 定誤警報確率(CFAR) に基づき作成したマスクパターンを距離スペクトルに適用することによって偽ピークの除去を行った．

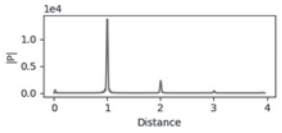


Fig.1: Range spectrum in 2chADM.

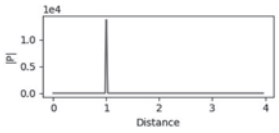


Fig.2: Masked ADM.

2-1P-3

2-1P-3　血管狭窄検出における複数位置での特徴量の評価

Evaluation of acoustic features at multiple positions in angiostenosis detection

○西島恵介，古家賢一（大分大）

◆血液透析患者はバスキュラーアクセス（患者と透析装置との間で血液循環するための仕組み）を管理し，狭窄の早期発見が必要である。

◆MFCC を音響特徴量とした分散分析により，正常音と狭窄症例中の狭窄特徴が明記されていない音（狭窄以外の音）の母平均に差があるとは言えない，狭窄音と狭窄以外の音の母平均に差があるとは言えないことがわかった。

◆狭窄以外の音が正答率を低下されている可能性を考慮し，識別のための学習モデルとして，全てを使用したモデル，狭窄以外の音を使用しないモデル，位置推定の結果のみを使用したモデルの３つで比較した。

◆複数位置間の差分平均特徴量により，症例中の最も特徴が現れている位置を推定し，その位置の音響特徴量により識別する手法を提案した。

◆識別には２次カーネルを用いたサポートベクターマシンを使用し，２分割交差検証を行った。

◆提案法の正答率は 65%（狭窄症例の正答率 77%）となった。

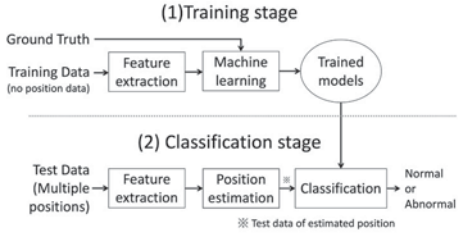


Fig. 1 Classification method

2-1P-4

2-1P-4　コサイン類似度罰則条件付き非負値行列因子分解に基づく音源分離の実験的評価

Experimental evaluation of source separation based on nonnegative matrix factorization with cosine similarity penalty

☆岩瀬佑太，北村大地（香川高専）

音源分離の分離精度を向上させるために罰則項を半教師あり非負値行列因子分解（SNMF）に付与する手法がいくつか提案されている．本稿では新しくコサイン類似度罰則項を SNMF に付与する手法（cos 型 PSNMF）を提案する．提案手法のコスト関数は下記である．

$$\min_{\boldsymbol{G},\boldsymbol{H},\boldsymbol{U}} \mathcal{D}(\boldsymbol{Y}\|\boldsymbol{F}\boldsymbol{G}+\boldsymbol{H}\boldsymbol{U}) + \mu \sum_{k,l} \frac{\sum_i f_{ik} h_{il}}{\left(\sum_i f_{ik}^2\right)^{\frac{1}{2}} \left(\sum_i h_{il}^2\right)^{\frac{1}{2}}}$$

$$\text{s.t. } g_{kj}, h_{il}, u_{lj} \geq 0 \; \forall i,\, j,\, k,\, l$$

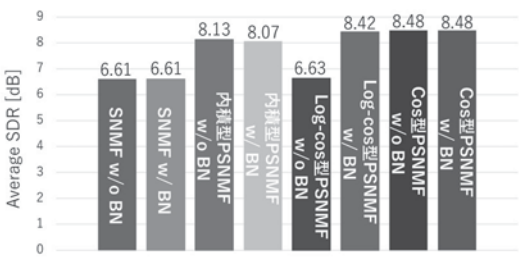


Fig.1 : Average SDR of test data for each method.

2-1P-5

2-1P-5　雑音スペクトルの振幅・位相同時推定に基づくスペクトル減算法の複数雑音源への拡張

Extension of spectral subtraction based on simultaneous estimation of amplitude and phase of noise spectrum of multiple noise sources

☆竹中幸輝，小澤賢司（山梨大・院・医工農総教）

◆マイクロホンアレイの観測信号を画像とみなして2D FFTすることにより２次元スペクトルを得る．その空間方向の直流成分には，(1) アレイに直交する方向 (0º) から到来した目的音の全スペクトルと，(2) その他の方向から到来した雑音の遅延和アレイ出力に相当する成分が混在している．ニューラルネットワーク（以下 NN と略記）により，(2) の成分を推定し減算することで雑音抑圧を行う．

◆前報（2021 春 音講論）では，位相特性における振幅特性への依存性を反映させるため，振幅と位相を同時に推定可能な NN を使用した．

◆学習データの正規化手法を検討することにより，雑音成分の推定精度の向上を図った．結果として，２つの白色雑音に対して前報を上回る抑圧量を得ることができた．

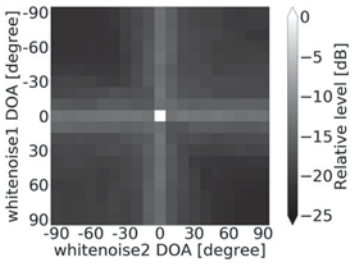


Fig.1: Noise suppression for two noise sources.

2-1P-6

2-1P-6　円周上の音場補間を利用した移動音源のステアリングベクトル推定

Steering vector estimation of moving source using sound field interpolation in the circumference

◎若林佑幸，山岡洸瑛，小野順貴（都立大）

◆多くのアレイ信号処理で仮定されている音源とマイクの位置（系）の時不変性は実応用では常に満たされない．

◆我々はこの問題に対し，これまでに円状マイクロホンアレイ（CMA）の回転移動に頑健な音場補間とビームフォーマへの応用を提案した．

◆本研究では，これまでに想定したCMAの移動でなく音源の移動に注目し，移動音源のステアリングベクトルを移動前のステアリングベクトルから音場補間によって推定する．上記想定の差に起因する誤差を許容し，どれだけステアリングベクトル推定に応用可能の確認を行う．

◆実験結果（Fig.１参照）より，提案法（Int）による推定性能は，補間を行わないとき（NoInt）や平面波仮定による推定（Plane）よりも高いものの，周波数平均で見ると，大きな改善は見られなかった．

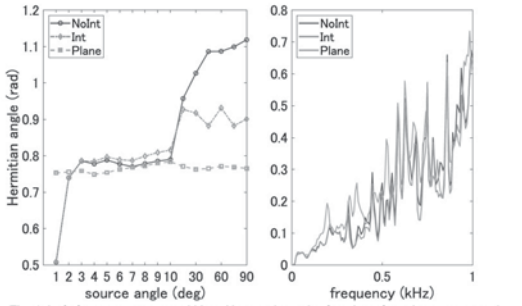


Fig. 1: Left: frequency-averaged Hermitian angle as the function of moving source angles, Right: angle-averaged Hermitian angle as the function of the frequency.

2-1P-7

2-1P-7 音響用電子透かしを用いた音声改ざん検出法の検討

A Speech Manipulation Detection Method using Audio Watermark

☆室井孝大，近藤和弘（山形大院・理工学研究科）

◆取り調べの録音音声を改ざん検出するシステムの開発

◆フィンガープリントとして**線スペクトル対**と時間情報として**フレーム番号**を音声の各フレームに電子透かしを用いて埋め込む

◆音声に置換改ざん、削除の改ざんを行い、それに対する改ざん検出の感度、特異度を評価した

◆結果

（１）　感度，特異度は98%以上

（２）　１音声あたりの誤検出フレームは平均２フレーム以下

◆正確な改ざん位置を特定できる検出精度が得られた

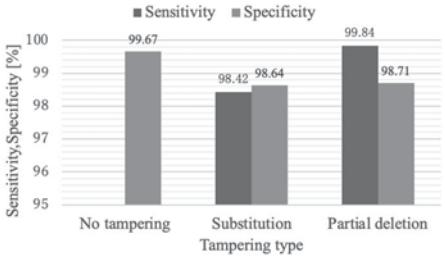


Fig.1: Tamper detection accuracy

2-1P-8

2-1P-8　再構成誤差に基づく教師無し機械異音検出のマルチステージ化検討

A multi-stage model for unsupervised anomalous sound detection based on reconstruction error

○鄭澤，△高椋佐和，△髙橋友紀子，△名取直毅（アイシン）

Auto-Encoder に基づく教師無し機械異音検出手法では，変動の大きい背景音が性能低下の一要因であった．本研究では,マルチステージモデルを導入し従来課題であった低SN比環境での異音検出性能改善への影響について考察する．

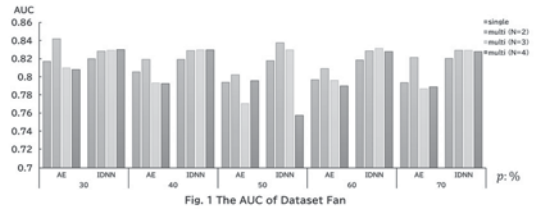


Fig. 1 The AUC of Dataset Fan

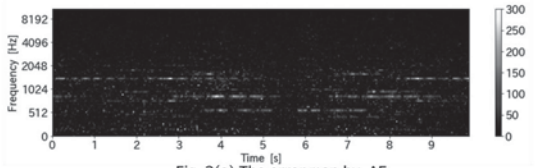


Fig. 2(a) The error map by AE

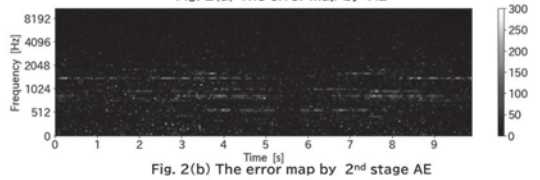


Fig. 2(b) The error map by 2nd stage AE

2-1P-9

2-1P-9　ヘビーテイル生成モデルに基づく独立深層学習テンソル分析

Independent deeply learned tensor analysis using heavy-tailed generative model

☆成澤直輝（東大），池下林太郎（NTT），高宗典玄（東大），
北村大地（香川高専），中村友彦，猿渡洋（東大），中谷智広（NTT）

◆独立深層学習テンソル分析 (IDLTA) [Narisawa+, 2021] は優決定条件における教師あり多チャネル音源分離手法であり，音源の生成モデルとして時変複素 Gauss 分布を仮定している．

◆独立深層学習行列分析 [Makishima+, 2019] を始めとする従来の多チャネル音源分離手法の多くは，生成モデルをヘビーテイルな分布へと拡張することによる性能向上が確認されている．

◆本研究では，IDLTA における生成モデルを，時変複素 Gauss 分布からヘビーテイルな分布である多変量複素 Student's t 分布及び多変量複素一般化 Gauss 分布 (GGD) に拡張した手法を提案し，その更新アルゴリズムを補助関数法に基づいて導出する．

◆音源分離実験によって提案手法の性能を評価する．

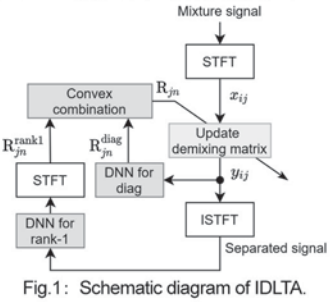


Fig.1: Schematic diagram of IDLTA.

2-1P-10

2-1P-10: ハンガリー法を用いたパーミュテーション解法に基づくブラインド音源分離

Blind source separation using Hungarian algorithm based permutation solver

☆山地修平，中嶋大志，若林佑幸，小野順貴（都立大）

- 本研究では，音源数 N ＝マイク数 M である優決定条件下のブラインド音源分離における，パーミュテーション問題を扱う．
- 既存の解法では，パーミュテーション探索にかかる計算量が $O(N!)$ である．そのため，音源数の増加に伴い計算量が増大する．
- 本研究ではハンガリー法を用いた解法を提案する．これにより，計算量を $O(N^3)$ に抑えることで，高速なパーミュテーション解決を実現する．
- 実験の結果，音源数が増えた場合に，提案手法が高速であることを示した．

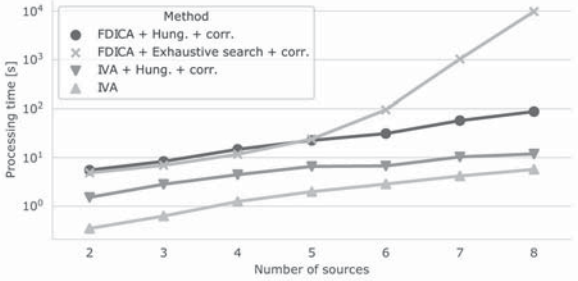


Fig. 1: Processing time against the number of sources using the Hungarian method or the exhaustive search.

2-1P-11

2-1P-11　日本語版 DRT 用音声を用いた劣化音声からの原音声推定

Estimating the Original Speech From Degraded Speech Using Japanese DRT　Sentence Utterance.

☆中澤和司，近藤和弘、上嶋雄大（山形大）

◆加算雑音、残響により劣化した音声に対して、原音声推定を用いた了解度推定を行うための原音推定部分に用いる手法の検討

◆Wiener フィルタリング、MMSE-STSA、SPENDRED、BiLSTM スペクトル強調を比較

◆残響時間 0.3、0.7、1.2 秒でさらにノイズを SNR0、-3、-6 で重畳した劣化音声に対する強調結果を評価

◆ノイズ、残響除去の面では BiLSTM が有効であるが、音声の品質が低下

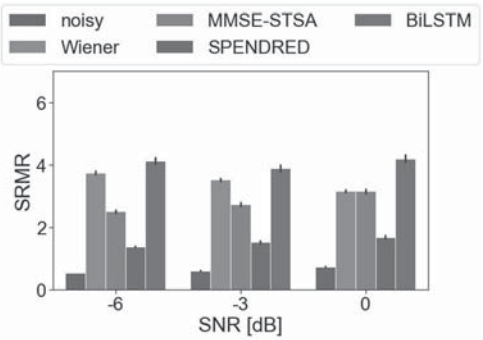


Fig. 1　Average SRMRvs. SNR when the reverberation time is 1.2 seconds

2-1P-12

2-1P-12　スペクトル重心の移動平均を用いた配管伝達音の異音検知

Anomaly sound detection for delivery pipe using moving average of spectral centroid.

富樫 敏侑，小林洋介（室蘭工大），榊原健太郎，矢澤伸弘（月島機械）

◆産業機器の自動運転において，異常検知は重要な課題である。本稿で対象とした配管伝搬音は振幅特性の時間変化が少なく，一定間隔で録音した音源信号のスペクトル重心の時系列解析を試みた。

◆録音信号のスペクトル重心を 0.5 s ごとに計算し，時間長の異なる二つの移動平均値の差から異音検知を行った。Fig.1 に異音発生時のスペクトル重心の変化と短時間と長時間の二つの移動平均を示す。また，提案法による異音検出時に星印（⋆）を付した。図では 4 区間の異音を検出しており，このうち，最初の 12000 s 付近の異音は機器オペレータが，この後の異常対策として装置停止操作を始めるタイミングと一致した。また，異音発生時以外を過剰に検知していない。

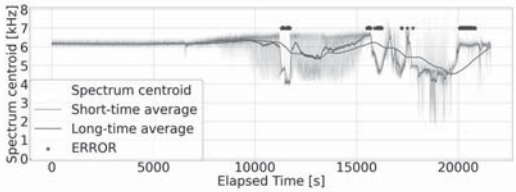


Fig.1: Anomaly detection results by spectral centroid and moving average

2-1P-13

2-1P-13　MIMII DUE: Sound Dataset for Malfunctioning Industrial Machine Investigation and Inspection with Domain Shifts due to Changes in Operational and Environmental Conditions

Ryo Tanabe, Harsh Purohit, Kota Dohi, Takashi Endo,
△Yuki Nikaido, △Toshiki Nakamura, ◎Yohei Kawaguchi (Hitachi, Ltd.)

◆稼動音に基づいて機械の異常を検知する「異音検知」技術は、機械の稼動品質の向上に寄与すると期待され、研究が盛んに行われている。

◆しかし、実用においては、正常/異常間の変化と独立に訓練時/運用時の間で音響特徴量の分布が変化する「ドメインシフト」が起こり、異音検知性能が低下する問題が存在する。ドメインシフトに頑健な異音検知方式はまだ明らかになっておらず、その開発に適するデータセットも公開されていない。

◆そこで、様々なドメインシフトを含んだ 5 種類の実際の機械の稼動音からなる異音検知評価用データセット MIMII DUE を構築し、公開した。

◆本データセットは、CC BY-NC-SA 4.0 の下で、無償利用可能である。
https://zenodo.org/record/4740355

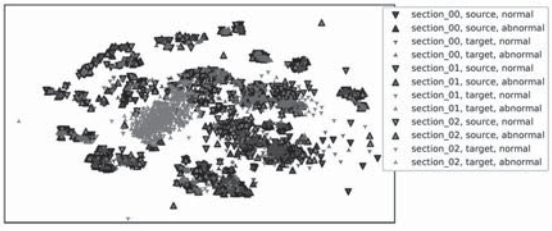


Figure: t-SNE visualization for acoustic features (machine type: fan)

2-1Q-1

2-1Q-1　耳道近傍の血管の拍動による音波に着目した心拍測定法

Heart rate measurement method focusing on sound waves caused by the pulsation of blood vessels near the ear canal

☆佐藤颯空，工藤彰洋（苫高専）

◆心拍測定法には、医療現場で用いられる ECG と呼ばれる方法がある。本稿では、ECG より簡便で衛生的な心拍測定法として、耳道近傍の血管の拍動による音波に着目し、作成したマイクセンサによって心拍を測定する方法を提案する。

◆Fig.1 のような心拍測定システムを構成し、ECG 信号を基準信号・マイクセンサ出力信号を評価信号として信号の測定を行った。

◆その結果、マイクセンサによって、心臓の活動に関する信号を取得することができたが、ECG 信号とマイクセンサ出力信号との間に遅延が生じた。その後、マイクセンサの設置位置を変えた追実験を行い、遅延が心臓からの距離に依存するものであることがわかった。

◆以上より、提案手法によって、心臓の活動に関する信号を測定可能であることが示唆できた。

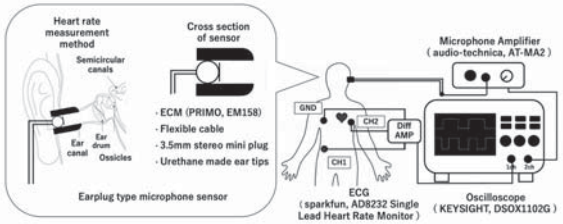


Fig.1: Heart Rate Measurement System

2-1Q-2

2-1Q-2　小腸の内容物が蠕動音に及ぼす影響の検討

Effects of small intestine contents on bowel sounds

☆齊藤慎之介，大塚翔，△前佛聡樹(千葉大)，
△堀創史，△本多通孝(福島県立医大)，中川誠司(千葉大)

◆腸蠕動音は腸機能の診断に用いられているが，その基準は専ら経験的に構築されたものである．我々は客観的診断基準の確立を目指して，蠕動音の発生メカニズム解明に取り組んで来た．前報では，腸内容物の動きが音の発生源となる可能性が示されている．

◆本研究では，空腹時，飲水後，および炭酸水摂取後において蠕動音を録音し，蠕動音発生に寄与する腸内容物の特定を試みた．

◆健常者 8 人を対象に各条件における 15 分間の腹部音計測を実施し，蠕動音の程度を表す２つの指標（RMS，蠕動音数）を算出した (Fig. 1)．また，併せて被験者の腹部エコー画像を取得した (Fig. 2)．

◆炭酸水摂取後に蠕動音の RMS，蠕動音数が有意に増加した．また，小腸内にガスの発生が確認された．小腸における蠕動音の発生には，腸内ガスの寄与が大きい可能性を示している．

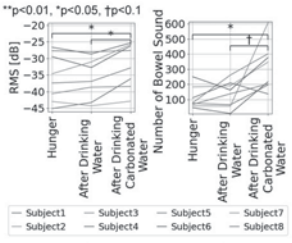


Fig. 1　Degree of bowel sounds in each state: (a) RMS, (b) Number of bowel sounds

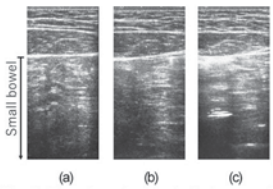


Fig. 2 Examples of abdominal ultra-sound images: (a) Hunger, (b) After drinking water, (c) After drinking carbonated water

2-1Q-3

2-1Q-3 DNN を用いた最小分散ビームフォーマの音源の動きに対する頑健性：音源追跡とエリア収音に基づくアプローチの比較

Comparative Study on DNN-based Minimum Variance Beamforming Robust to Small Movements of Sound Sources

☆西城耕平(早大),藤枝大，片桐一浩(沖電気),小林哲則,小川哲司(早大)

◆概要

- 音源を追跡して強調する手法と目的音源を含むエリアを強調する手法の比較
- ３種類の最小分散ビームフォーマを DNN と結合学習して評価
  1. **L - MVDR** : DNN と統合することで音源を追跡しながら強調
  2. **MC - MVDR** : 複数方向の無歪の拘束によりエリアを強調
  3. **RMC - MV** : 2.の拘束を緩和してより高精度にエリアを強調

◆学習・評価データ

- 学習：目的音源をマイクロホン正面方向に配置したデータ
- 評価：学習時と同様のデータ(Matched)，音源位置が変化するデータ (Mismatched) の２種類で音源移動への頑健性を検証

◆結果

- 全ての BF が音源位置変化に頑健
- 音源を追跡しながら強調する L - MVDR が最良の性能

Table 1 : SDR improvement and STOI improvement of L-MVDR, MC-MVDR, RMC-MV under matched and mismatched conditions

	Matched		Mismatched	
Beamformer	SDRi [dB]	STOIi	SDRi [dB]	STOIi
L - MVDR	7.94	0.139	**7.86**	**0.145**
MC - MVDR	4.84	0.127	4.89	0.134
RMC - MV	7.50	0.125	7.48	0.132

2-1Q-4

2-1Q-4　振幅・周波数同期変調方式に基づく集束超音波を用いた空中ハプティクスの基礎的検討

Fundamental Study on Mid-air Haptics Using Focused Ultrasound Based on Synchronized Amplitude-and-Frequency Modulation

☆上田明日斗，中山雅人，高橋徹 (阪産大院)

◆従来の集束超音波を利用した空中ハプティクスでは，人に圧力を知覚させやすくするため，振幅変調を用いて圧力を振動させていた．しかし，これまでの空中ハプティクスの研究は触覚提示位置の制御や，視覚情報との組み合わせが主であり，変調方式の影響については，検討の余地が残されている．

◆一方，我々はパラメトリックスピーカにおいて，振幅・周波数同期変調方式に基づく復調音の強調手法を検討してきた．本稿では，振幅・周波数同期変調方式を空中ハプティクスに用いることを検討する．従来の可聴音に対する振幅・周波数同期変調方式では，振幅変調と周波数変調の位相を $-\pi/2$ シフトしていたが，これは周波数によって適切な位相が異なることが知られている．そこで，本稿では集束超音波を用いた空中ハプティクスにおいて適切な位相シフト量について評価した．

◆提案手法の有効性を確認するために圧力の評価を行った．Fig. 1 に圧力測定の実験結果を示す．振幅変調と周波数変調で $+\pi/2$ シフトさせる条件で振幅変調より大きい圧力を発生できることを確認した．

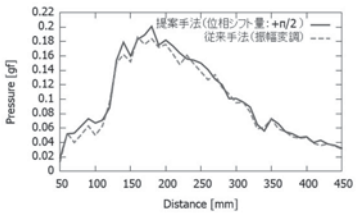


Fig.1:圧力測定の実験結果

2-1Q-5

2-1Q-5　音分類課題において有効な位相情報の表現に関する検討

An Investigation of Effective Representation of Phase Information for Audio Classification

☆日髙駿介，若宮幸平，鏑木時彦（九州大）

◆音声強調などの波形(または等価な複素スペクトログラム)から波形を生成する課題では，入力特徴量の位相情報が有用であることが確かめられている。本研究では，音声や環境音など複数の分類課題を対象に，対数振幅メルスペクトログラムの類にどのように位相情報を付与するのが有効か検討する。

◆対数振幅メルスペクトログラムのように周波数が圧縮された表現を複素数の状態で求めるために，学習可能な Gabor フィルタバンクを使用し，波形と一次元畳み込みを行なった。位相情報として，位相のフェーザ表示，瞬時周波数と群遅延を比較した。また，位相情報とパワーの要素ごとの積も検討対象とした。

◆時間周波数表現から特徴抽出，クラス予測を行う CNN として EfficientNetB0 を使用して，4 つの音分類課題で比較実験を行なった。音声の感情認識と単語認識課題において，位相情報，特に群遅延が有効性を示した。その一方で，環境音分類や鳥の存在判定課題では，位相情報の付与は，効果がない，あるいはかえって性能低下に繋がった。今後は，より多くの音分類課題を対象とし，位相情報が有効な範囲を調査することが必要となる。

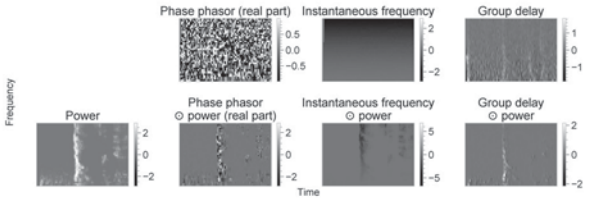


Fig.1：Heatmaps of power and phase features (⊙ is element-wise product)

2-1Q-6

2-1Q-6　ベイズ情報量規準に基づく音信号の区間分割に関する検討

An investigation on segmentation of sound signal based on Bayesian information criterion

○安倍幸治，小田原乃斗，西口正之，高根昭一，渡邉貫治(秋田県立大)

◆見守りシステムを目標とした，音信号ベースの教師無し異常検出システムの構築を検討
- オーソドックスな混合ガウスモデルに基づく日常音の確率モデルからの乖離の度合いを指標として異常を検出
- 計算量低減及び検出性能の向上を目的とした音信号の自動分割システムの構築・評価を実施

◆検討内容
- ベイズ情報量基準（BIC）に基づく音信号の自動分割とそれによる代表音響特徴量ベクトルの正規性の維持に関する検討
- 本分割手法の導入による固定長分割からの異常検出性能の比較

◆結果・考察
- BIC 分割によって分割区間の音響特徴量群の正規性は向上した
- BIC 分割を行っても異常検出性能は向上しなかった

Table 1 Number of sections and normality evaluation of each segmentation method

	BIC 分割	3 秒分割
区間数	2738	2738
MFCC1 次の正規性割合	0.99	0.78
パワーの正規性割合	0.99	0.92

Table 2 Confusion matrix of each segmentation method

BIC 分割	異常判定	正常判定
異常	84	7
正常	219	745

Recall: 0.92, Precision: 0.28

3 秒分割	異常判定	正常判定
異常	64	0
正常	124	490

Recall: 1.00, Precision: 0.34

2-1Q-7

2-1Q-7　車載動物避け用超指向性スピーカー出力音の走行時の到達範囲に関する一考察

A study on an arrival range of output sound of a super-directional speaker for avoiding animals mounted on a moving vehicle

○松﨑博季，△真田博文，△和田直史(北科大・工)

車両とエゾシカの衝突回避用車載動物避けスピーカーが発する超指向性音が、車両の走行音に打ち消されることなく車両の走行方向に届くのかを確認する音響測定実験を行った結果について報告する。

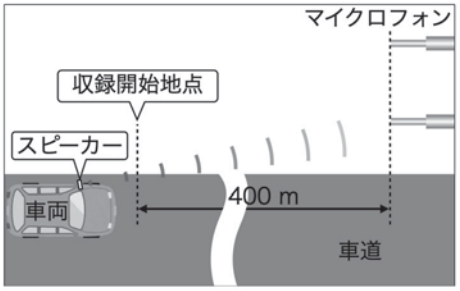


Fig.1：schematic diagram of the experimental set-up.

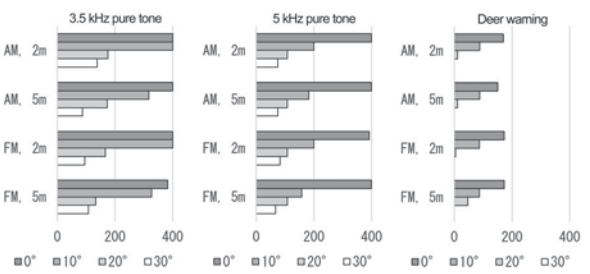


Fig.2：Maximum distance from the microphones where the sound could be confirmed.

2-1Q-8

2-1Q-8　劣決定音源分離のための冪乗相乗平均マスク

Exponentiated geometric mean mask for underdetermined speech source separation.

○太刀岡 勇気（デンソーアイティーラボラトリ）

◆深層神経回路網(DNN)に基づくマスク推定法は、劣決定問題にもそのまま適用可能であるため、有効性が期待される

◆DNN マスク推定の問題の一つは、アクティブ・非アクティブなマスク数が不均衡なことから、マスク推定にバイアスが入ることである

◆ここでは音声とその逆(ノイズ)マスクを同時推定する方法を提案する

◆また、そのマスクを冪乗して利用する

◆マスク推定の安定化のため、音声・ノイズマスクの幾何平均から強調マスクを推定する

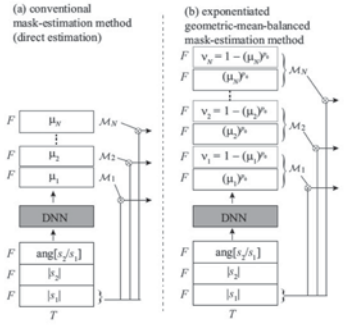


Fig.： Conventional mask estimation method (a) and our proposed method (b).

2-1Q-9

2-1Q-9　複素スペクトルの次元拡張表現の検討

Study on dimensional extension of complex spectrum.

☆樋貝満喜，島内末廣，青木茂明(金沢工大)

◆本報告では、２次元の複素平面上に新たな１次元を加えた３次元空間において複素スペクトルを俯瞰する拡張表現を提案し、その特徴を考察した結果を述べる。

◆３次元に拡張された x, y, z 軸を基準とした次元拡張位相角$\theta_x, \theta_y, \theta_z$と従来の振幅や位相との関係を確認した。

◆音声の短時間フレームの複素スペクトルから得られる特徴量として、従来の位相角θと次元拡張位相角θ_x、θ_y（Fig.1(a)）、および、従来の振幅スペクトルと次元拡張位相角θ_z（Fig.1(b)）を比較した。

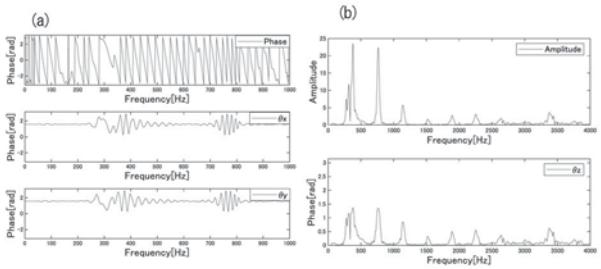


Fig.1 Properties of θ_x, θ_y, θ_z, original phase and amplitude for a speech signal.

2-1Q-10

2-1Q-10　Low-overlap window を用いたオンライン Wave-U-Net のアルゴリズム遅延の削減

Reducing algorithmic latency of online Wave-U-Net using low-overlap window

☆中岡想太郎(筑波大),李莉(NTT),
牧野昭二(筑波大，早大)，山田武志(筑波大)

◆ Wave-U-Net は end-to-end 型のシングルチャネル音源分離手法であり、時間領域で動作するため位相情報もモデル化でき、歌声分離や音声強調のタスクにおいて高い性能を示した。また、これまでに我々はそのオンライン手法を提案した。

◆ オンライン Wave-U-Net は窓処理により入力信号をフレームごとに処理するため、窓長がアルゴリズムの下限となる。

◆ 本稿では、窓の端点付近にゼロ領域をもつ窓による窓処理をオンライン Wave-U-Net に組み込む事で、アルゴリズム遅延を削減することを提案する。

◆ 実験により、提案手法は SDR の低下を 1.15dB に抑えながらアルゴリズム遅延を従来の 40%に削減し、低遅延かつ高性能の音声強調が可能であることを示した。

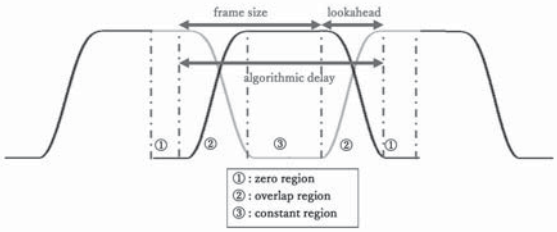


図 1: Low-overlap window の概観

2-1Q-11

2-1Q-11　ステレオ音像に着目した音楽信号の可視化の検討

Study on stereo image visualization of music signals

☆清水亮裕，島内末廣，青木茂明(金沢工大)

◆複数の楽器からなる器楽曲のステレオ音楽信号から生成される音像をステレオ空間上にリアルタイムで表現する可視化システム(Fig.1)を提案する。

◆メル尺度とラウドネス補正を適用することにより、より聴感に近い可視化が可能となることを確認した。(Fig.2)

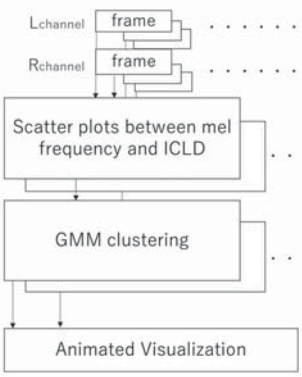


Fig.1 System of visualization.

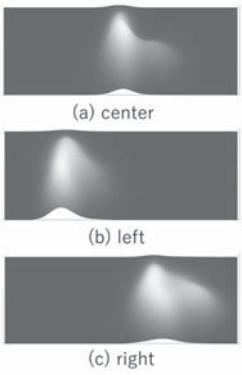


Fig.2 Examples of visualized sound images.

2-1Q-12

2-1Q-12　機械学習による瞬時スペクトル推定に基づく雑音抑圧システムの実機搭載に向けた考察

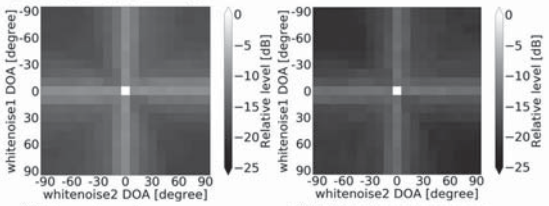
Consideration on implementation of a noise subtraction system based on instantaneous spectrum estimation by machine learning

☆大宮拓海，△野本駿典，竹中幸輝，小澤賢司 (山梨大)，
田中俊也，△鮫島健一郎，△伊藤直樹 (アルプスアルパイン(株))

◆マイクロホンアレイから得られた音圧の分布画像に対し、２次元フーリエ変換を行い、２次元スペクトルを得る。この２次元スペクトル及びニューラルネットワークを用いて雑音成分を推定し、引き去ることで雑音抑制を行う。

◆前述した雑音抑圧システムを実機搭載することを想定し、８個のマイクロホンを等間隔に配置したアレイではなく、４個のマイクロホンを不等間隔に配置し、小型化したアレイを用いた。

◆２つの白色雑音に対する抑圧量を計算した。Fig.１に示すとおり、不等間隔の４個のマイクロホンによるアレイの抑圧量は等間隔の８個のマイクロホンによるアレイの抑圧量より劣るものの、最大約−18 dB 程度の抑圧量を示した。



(a) Four-microphone array　(b) Eight-microphone array

Fig.1 : Noise suppression for two noise sources.

2-1Q-13

2-1Q-13　非定常ガウスモデルによるブラインド音源分離手法の同定可能性について

Study on Identifiability of Blind Source Separation

○池下 林太郎，中谷 智広（NTT）

研究の背景

◆ 独立成分分析（ICA）の同定可能性（モデルに従う観測から，モデルの真のパラメータを推定できるか否か）は詳しく研究されてきた．

◆ ICA の空間モデルを柔軟にした FCA や FastFCA は，研究の位置づけの意味で重要な手法であるが，同定可能性といったモデルの妥当性をはかる性質は未解明であった．

本研究の成果　以下の４つを示しました

◆ FCA も FastFCA も，一般に，同定可能でない．

・尤度を変えずに，パラメータ（実行可能解）を連続的に変形できる．

・特に，最尤推定量（最適解）が複数存在する．

◆ FastFCA は，劣決定系の BSS を壊滅的に解けない．

・尤度を変えずに，任意の音源の MMSE 推定量を 0 にできる．

◆ 決定系の BSS に対して，FastFCA の最適解の１つは ICA．

・尤度を変えずに，FastFCA の実行可能解を ICA に変形できる．

◆ パワースペクトルの真値が既知のとき，FastFCA は一般に同定可能．

・期待対数尤度を最大にするパラメータは真のパラメータただひとつ．

用語の定義

◆ FCA:　フルランクの空間相関行列に基づく BSS 手法のこと [Duong+2010]

◆ FastFCA:　FCA の近似高速化 [Ito+2018, 2021]

2-3P-1

2-3P-1　音響情報と言語情報に基づく対話エージェントの誤反応防止技術の検討

Investigation into Acoustic and Lexical Device Directed Speech Detection for Reducing Unintended Voice Agent Response

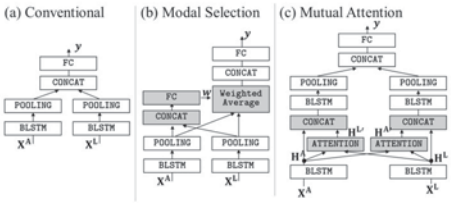
◎佐藤宏，小川厚徳（NTT）

◆概要

・音声対話エージェントに入力された音声が，ユーザーが意図して機械に話しかけた音声か否かを，音声から抽出された音響特徴系列とそれを音声認識した単語系列の両者から識別する問題を検討する

・従来法のモデル化を改良した２つの手法を提案

- Modal Selection: 音響系列と言語系列のいずれが識別に有用かを捉え，系列を重みづけて統合する手法
- Mutual Attention: 音響系列と言語系列の時間的な対応関係を考慮しながら識別を行う手法



◆評価実験 (等価エラー率 EER で評価)

モデル	EER [%]
音響系列のみ	10.76
言語系列のみ	23.77
従来法	9.68
Modal Selection（提案法）	9.52
Mutual Attention（提案法）	9.30

2-3P-2

2-3P-2　話者照合のためのステレオ音声データを用いた深層学習に基づくなりすまし検出法

Deep neural network-based replay attack detection using stereo signals for speaker verification

☆荻野天翔，今泉遼，塩田さやか，△貴家仁志（都立大）

◆本研究では話者照合のためのマルチチャネル音声を用いた深層学習に基づくなりすまし検出法を提案する。

◆深層学習モデルにマルチチャネルデータを入力する方法として、マルチチャネルデータを結合して一つのネットワークに合わせて入力する方法と、チャネル毎に分離した入力ネットワークを持ち、そのあとのネットワークを介してデータを結合する方法 (Fig.1) の２種類の手法を比較しなりすまし検出の性能を評価した。

◆ネットワークの入力がチャネル毎に分離したモデルを用いたなりすまし検出が比較的高い性能を得ること、さらに既存のステレオチャネルなりすまし検出法とスコアフュージョンすることにより性能が大幅に向上することを確認した。

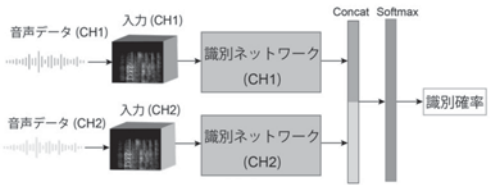


Fig.1: 各チャネルの入力特徴量を分離してネットワークに入力する場合のフロー

2-3P-3

2-3P-3　読み上げ文を対象とした End-to-End 音声点訳

A Study on End-to-End Speech to Braille Translation

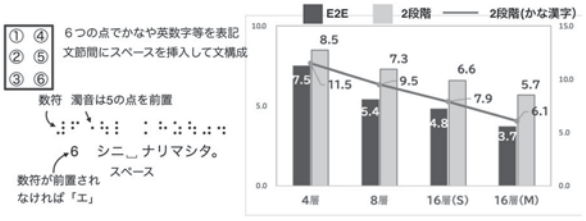
○小林 彰夫，大西 淳児（筑波技大），西崎 博光（山梨大・院），北岡 教英（豊橋技科大）

◇ 盲ろう者の情報保障

- 国内の盲ろう者（視覚・聴覚に障害のある者）は 14,000 人と推定
- 情報の受発信手段は手話や点字など障害の状況により多様
  点字・指点字による受発信が可能な者の割合は 6.5%
- 日本における点字
  – かな文字が基盤であり，音声と親和性が高い（6 点点字）
  – 音声から点字への変換（音声点訳）による情報保障の可能性

◇ End-to-End 音声点訳

- 学習データ: ASJ/JNAS および NHK ニュース (36.8 時間)
- 評価データ: NHK ニュース 742 発話 (かな漢字:39.1k 文字，点字:72.7k 文字)
- ネットワーク
  – E2E : Conformer-Transducer (16 層)
  – 2 段階点訳（認識結果 → 点訳）: Conformer/Transformer-Transducer (各 16 層)
- 実験結果
  – E2E の方が 2 段階点訳よりも性能が高い
    ⇒ かな漢字認識結果の品質の影響を受けない
  – E2E の誤り全体の 13.3%が文節区切りの誤り
    ⇒ 形態素もしくは文節相当の情報により誤り削減を期待



点字/点訳の例　　実験結果（文字誤り率）

2-3P-4

2-3P-4　誤り訂正に基づく器質性構音障害者の音声認識精度向上の検討

An investigation of error correction method
for improving speech recognition of organic dysarthria

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

◆器質性構音障害者は発音スタイルが健常者と異なるため，健常者に比べて音声認識の精度が低くなる．
◆本研究では，誤り訂正モデルを用いて，器質性構音障害者の発音スタイルに起因する音声認識誤りを補正することを検討する．
◆このとき，障害者音声データを大量に収録することは困難であるため，誤り訂正モデルを学習するための音声認識誤り文が不足する．
◆そこで提案手法では，構音障害者の誤り傾向を反映した疑似誤り文を生成することで，誤り訂正モデルの学習データを増強する．
◆特定話者音素認識実験により、提案手法の有効性を確認した．

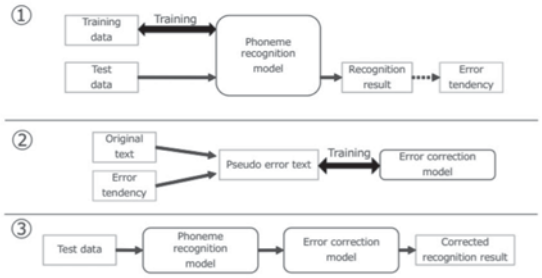


Fig.1 : Overview of our proposed method.

2-3P-5

2-3P-5　畳み込みオートエンコーダを用いた外れ値検出による叫び声検知

Shout detection based on convolutional autoencoder
with outlier detection

☆石田 泰都，福森 隆寛，山下 洋一（立命館大・情報理工）

◆研究目的

叫び声の学習データを必要としない叫び声検知の実現

◆ 研究概要

従来の深層学習を用いた叫び声検知では，モデル学習用の叫び声を事前に用意しておく必要があった．本稿では，深層学習手法の一つである畳み込みオートエンコーダ（Convolutional AutoEncoder : CAE）を用いた，モデル学習に叫び声を必要としない外れ値検出による叫び声検知の手法を提案した．評価実験の結果，信号対雑音比（Signal-to-Noise Ratio : SNR）が 20 dB の F-score において 83.44 % を示した．

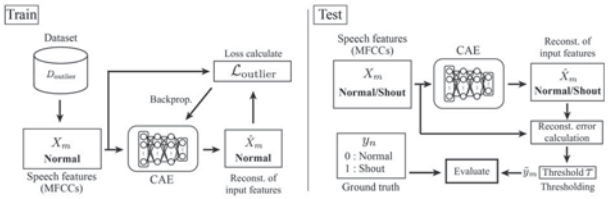


Fig.1 : Overview of Proposed Method

2-3P-6

2-3P-6　映像情報に基づく物体追跡 DNN を用いた話者音声強調の検討

A Study on Specific Speaker Speech Enhancement Based on Video Information Using DNN for Object Tracking

☆吉田敦史（立命館大院），　王浩南，　岩居健太，　西浦敬信（立命館大）

近年，計算機の性能が大幅に向上していることにより，音声と映像を同時に扱うマルチモーダルな学習が注目されている．本研究では，カメラのズームインやズームアウトを対象に映像情報に基づき画面内話者の音声を強調する手法を提案する．Fig. 1 に示すように，動画撮影において，ある話者にズームインする場合，映像に合わせてその話者の音声が強調されないという問題がある．そこで，物体追跡を用いてズームインやズームアウトなどのカメラモーションを捉えることで，映像に合わせた画面内話者の音声強調を目指す．提案システムは，Fig. 2 に示すように，音声分離，物体追跡，話者照合の 3 つのパートから構成される．評価実験の結果，提案システムの有効性が確認された．

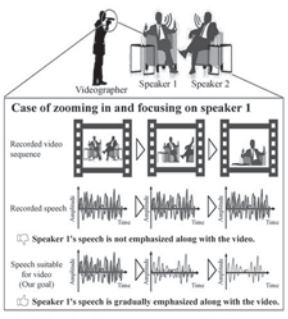


Fig.1: Video shooting scene assumed in this paper

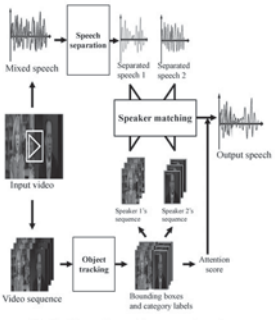


Fig.2: Overview of proposed system

2-3P-7

2-3P-7　エンドツーエンド ASR のデコーダベース言語モデル強化

Decoder-based Language Model Enhancement for End-to-end ASR

☆Zhuo Gong, Daisuke Saito, Nobuaki Minematsu (The University of Tokyo)

Nowadays, end-to-end automatic speech recognition (E2E ASR) becomes the mainstream in speech recognition. Meanwhile, independent Language models (LM) still play crucial roles to improve decoding precisions of ASR systems by introducing extra language knowledge. In an E2E ASR model, its framework is usually composed of an encoder which transforms acoustic features into intermediate vectors, and a decoder which predicts next token based on the intermediate vectors and predicting histories. In previous work, we have revealed the potential of this decoder to facilitate like an independent LM and improve the decoding precision without an external LM. This article is based on these works and proposes methods to develop the potential of a decoder as a LM in order to enhance an ASR system as a whole. We pre-trained the decoder on text corpora, then we did E2E training with the encoder on speech corpus later on. We found that even though the precision of the whole system is improved during the E2E training, the functionality as a LM of the decoder is degenerated. We propose a method to modify the objective function of E2E training to alleviate this phenomenon and improve the ASR precision. The results show that this strategy is effective. And our final goal is to supplant an extra LM in E2E ASR.

2-3P-8

2-3P-8　FFT ライブラリを対象とした実時間 Vocoder の速度比較

Processing speed comparison of the real-time vocoder in FFT library

☆堀部貴紀 (REALITY/明治大), △橋本順之 (グリー),
△白井暁彦(REALITY), 森勢将雅(明治大)

iOS デバイスや配信サービスにおけるサーバ内での声質変換のリアルタイム化を実現するにあたり，WORLD で使用されている大浦 (ooura) と Accelerate Framework, FFTW3, Math Kernel Library (MKL) のような各環境に提供されている FFT 関数の速度比較を通して，適した FFT 関数や実行環境について検討した．人工的に生成された信号音を用いた実験を通して Real Time Factor (RTF) は，MKL, FFTW3, Accelerate, ooura の順に小さくなった．大浦 FFT を FFTW3 や各環境に最適化された FFT 関数に置き換えて得られる効果は十分にある．

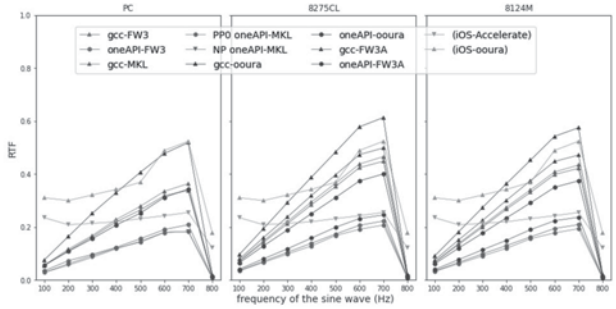


Fig. 1：周波数が100 Hzごと異なる音源を用いた各環境におけるRTFの比較
(gcc, oneAPI：使用したコンパイラ，PP0, NP：コンパイラオプションの自動並列化を有効化/無効を示す)

2-3P-9

2-3P-9　Unidirectional LSTMを用いたDNN-HSMM音声合成の検討

A study on DNN-HSMM-based speech synthesis with unidirectional LSTM

○西澤信行，△服部元(KDDI総合研究所)

DNN-HSMM音声合成において、言語情報からHSMMのパラメータを推定するDNNにRNNを用いることで、より高い品質が得られると考えられる。この際、双方向のRNNを用いると音声合成システムの応答時間が長くなるため、音声対話システム向けの音声合成システム等では、一方向のRNNを用いることが望まれる。本研究では、女声ナレータ1話者による約9.6時間の音声を学習データとして、従来のHSMM音声合成のフルコンテキストラベルに相当する言語情報から、双方向のLSTM (BLSTM)、一方向のLSTM (ULSTM)、フィードフォワード型DNN (FFNN)によるDNN-HSMM合成音声と、HSMM合成音声(HSMM)をそれぞれ作成し、さらに自然音声(NAT)と分析再合成音(RESYN)を加えた6種類、各20文の読み上げ音声に対して評定者9名によるMOS試験を行った。実験の結果から、FFNNに対してULSTMの品質が有意に優れていること、また、BLSTMに対するULSTMの品質低下は小さいことが分かった。

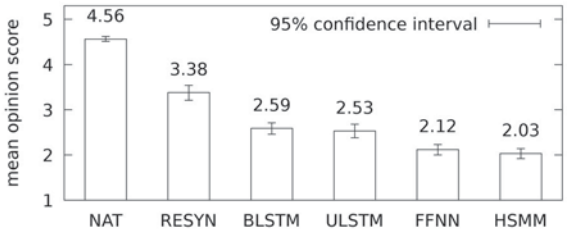


Fig. 1　Mean opinion score (MOS) of synthetic speech.

2-3P-10

2-3P-10　敵対的生成ネットワークを用いた非並列歌声変換の検討

Examination of non-parallel singing voice conversion using generative adversarial networks

☆堀幸央，小坂哲夫 （山形大）

◆敵対的生成ネットワーク(GAN: Generative adversarial networks)を用いた日本語歌唱の歌声変換の検討を行った．

◆以前の実験では，同一の歌唱を利用する並列学習と比べ，異なる歌唱を利用する非並列学習では性能が劣化するという結果が得られた．しかしデータの関係で同一ジャンル間での比較は行っていなかった．

◆並列・非並列の違いとジャンルの違いのどちらがより影響が大きいか検討を行った．

◆Fig.1 より同ジャンル間では並列学習と非並列学習に大きな性能差は見られなかった．また Fig.2 より異なるジャンル間での変換は性能が劣化することが分かった．テンポの違いの影響などが考えられる．

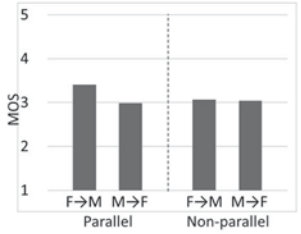


Fig 1：MOS results of naturalness for child songs (F: female, M: male)

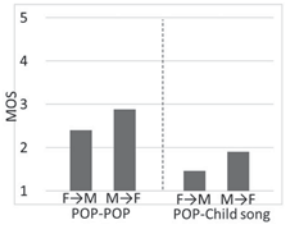


Fig.2：MOS results of naturalness for non-parallel training

2-3P-11

2-3P-11　超高齢者音声コーパス EARS における超高齢者の音響的特徴

Acoustic characteristics of the super-elderlies in the super-elderly speech corpus EARS

☆福田芽衣子，西村良太(徳島大)，
西崎博光(山梨大)，入部百合絵(愛知県立大)，山本一公(中部大)，
北岡教英(豊橋技科大)

◆　高齢者音声コーパスの S-JNAS（60~94 歳）と超高齢者音声コーパス EARS の話者（70~98 歳）について話速・母音継続長を測定した。

◆　その結果、男女ともに加齢に伴う有意な話速の低下と母音継続長の増加が認められた。

◆　高齢者音声認識の精度向上への EARS の効果を簡便な方法で検討した。

◆　具体的には、EARS に加工を施さずそのまま複数回繰り返し使用し、他の成人音声コーパスと共に音響モデルを作成した。

- DNN-HMM の音声認識では、baseline の WER18.7%が、EARS を 2 回用いたモデルで 8.9%に改善。
- Transformer を用いた音声認識では、baseline の CER12.5%が、EARS を 5 回繰り返したモデルで 10.8%に改善。

◆　EARS を繰り返し用いる効果がみられた。

2-3P-12

2-3P-12　歌声合成のための双方向 LSTM によるビブラート表現方式の検討

Study of Expression of Vibrato for Singing Voice Synthesis Using Bidirectional LSTM

☆金子隼人，原直，阿部匡伸（岡山大・HS 統合科学研）

◆本稿では，我々が提案した従来方式の主観評価実験をおこなうとともに，従来方式では不十分であったビブラート生成方式を新たに提案し，提案方式について客観評価実験をおこなう．

◆図 1 に新たに追加 したビブラート生成方式を示す．ビブラート特徴量は，ビブラートの振動と深さを表す vibrato rate と vibrato extent であり，楽譜情報 と音素情報を入力とし，BLSTM で推定する．

◆主観評価の結果，推定誤差に F0 の値を含めてスペクトルパラメータを推定することの有効性を示した．

◆ビブラート特徴量の再現度合について 9 段階交差検証を用いた客観評価実験により，ビブラートの変化を学習できていることを示した．

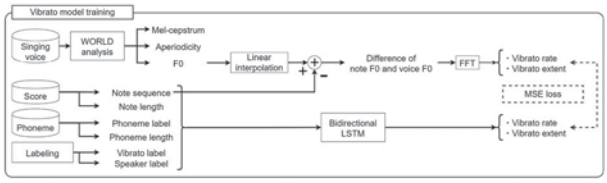


Fig.1 : Newly added vibrato generation method.

2-3P-13

2-3P-13　携帯端末を用いた口唇認識による発声支援の検討

Speech Enhancement System Using Lip-reading with mobile devices

☆江口文耀，松井謙二（阪工大 R&D 工学部），
中藤良久（九工大），加藤弓子（聖マリアンナ医科大学）

◆本研究では，先行研究で作成していた PC アプリを携帯型端末アプリとして実装し，認識性能の検証を行った．また，並行して実験対象者や語彙数による認識精度の違いについても実験を行った．

◆携帯型端末への実装は Android Studio を用いてアプリ開発を行った．携帯型端末で行った認識結果は先行研究の PC を用いた場合より低い結果となった．しかし，第６候補までを含めると 95%まで認識出来ているので候補選択機能を検討する．被験者の発声変動に関しては“母音の口形に注意”，“一定のテンポでの発話”，などについて注意することにより表１に見られるように認識率のバラツキを低減できた．

表 1 単語認識率（発話の注意後）

	認識率						
	h2	n2	o2	y7	k1	s1	t1
第一候補	17.3%	32.0%	25.6%	28.8%	26.2%	14.7%	21.1%
第二候補	27.5%	41.0%	39.7%	39.7%	43.5%	24.3%	33.3%
第三候補	35.8%	45.5%	52.5%	44.8%	54.4%	35.2%	40.3%
第四候補	46.1%	51.2%	57.0%	50.6%	58.3%	39.7%	43.5%
第五候補	51.9%	53.8%	61.5%	55.1%	60.2%	48.7%	48.0%
第六候補	60.2%	55.7%	64.7%	59.6%	65.3%	55.7%	51.2%

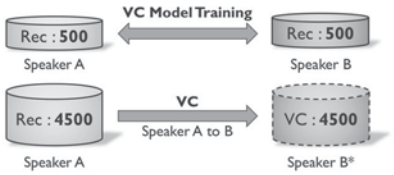
2-3P-14

2-3P-14　テキスト音声合成のための CycleGAN 声質変換を用いたデータ拡張の検討

An Investigation of Data Augmentation Using CycleGAN Voice Conversion for Text-to-Speech Synthesis

◎寺島涼，山本龍一，橘健太郎 (LINE 株式会社)

- 対象話者(話者 B)の少量データを利用した TTS モデル構築を目的として，他話者(話者 A)からの声質変換を用いたデータ拡張を行った．



- 主観評価の結果，対象話者 5,000 発話で学習した TTS モデルに対し，自然性評価では，対象話者 500 発話を用いたデータ拡張 TTS モデルが匹敵し，話者類似性評価では，対象話者 2,500 発話を用いたデータ拡張 TTS モデルが匹敵する結果となった．

MOS results on Naturalness and Similarity

Model	Type	Naturalness [1～5]		Similarity [1～4]	
		Female	Male	Female	Male
Natural	Recorded audio	4.56	4.28	-	-
TTS:500	TTS	3.30	3.38	3.12	2.94
TTS:5000		4.22	4.04	3.70	3.71
TTS:500 + VC4500	TTS with VC data augmentation	4.15	3.99	3.46	2.92
TTS:2500 + VC2500		4.15	4.08	3.66	3.70

2-3Q-1

2-3Q-1　短時間話者埋込みに基づくストリーミング型 End-to-End 話者ダイアライゼーション

Streaming End-to-End Neural Diarization based on short-term speaker embeddings

○安藤厚志，森谷崇史，増村亮，浅見太一（NTT）

◆ストリーミング処理が可能な End-to-End 話者ダイアライゼーション (End-to-End Neural Diarization: EEND)の実現に向け，短時間話者埋込みに基づく EEND (SSE-EEND)と，SSE-EEND の利点を生かした話者損失に基づくモデル学習手法の提案を行う．

◆提案手法では，混合信号の各フレームに対応する短時間話者表現と，短時間話者表現の累積表現という二種類の中間表現に基づいて時刻ごとの各話者の発話有無を推定する．

◆模擬混合信号を用いた評価実験により提案手法の有効性を示す．

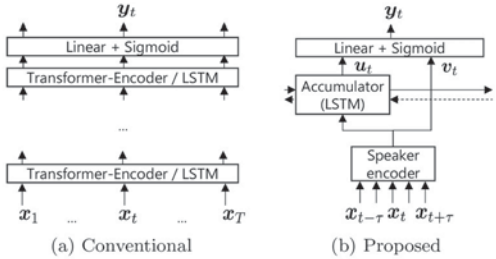


Fig.1 : Structures of the conventional and the proposed methods.

Table 1: Diarization Error Rate (DER) on the simulated mixtures.

			DER [%]				
			overall	1spk	2spk	3spk	4spk
Batch	Baseline	EEND-Transformer	7.3	1.3	4.3	7.2	10.8
		EEND-BLSTM	18.1	1.5	11.1	19.1	25.8
	Proposed	SSE-EEND-BLSTM	9.8	1.4	6.5	9.5	14.3
		SSE-EEND-BLSTM w/SpkLoss	9.0	1.2	5.7	9.4	12.7
Streaming	Baseline	EEND-ULSTM	29.0	5.8	19.7	30.4	39.4
	Proposed	SSE-EEND-ULSTM	19.1	2.0	13.4	19.7	26.7
		SSE-EEND-ULSTM w/SpkLoss	15.2	2.3	9.9	15.3	21.7

2-3Q-2

2-3Q-2　BERT に基づく Embedding を用いた言語モデルの検討

A Study on Language Modeling with BERT-based Word Embedding

○小椋忠志，藤本雅清，沈鵬, Lu Xugang，河井恒(NICT)

◆長い音声入力に対応するため、広い範囲の文脈を考慮可能な BERT に着目

◆BERT の中間層の隠れ状態が入力単語列に対する Embedding であるとみなし，従来の RNN 言語モデルの入力とすることで，より文脈を考慮した言語モデルを構築する手法を検討

◆BERT の中間層のうち、入力に近い層の Short term context と出力に近い層の Long term context を抽出し、RNN に基づく言語モデルの入力に用いる

◆日本語話し言葉コーパスを対象に、複数の Embedding の使用方法を検討し、BERT の活用法の一例としての結果示す

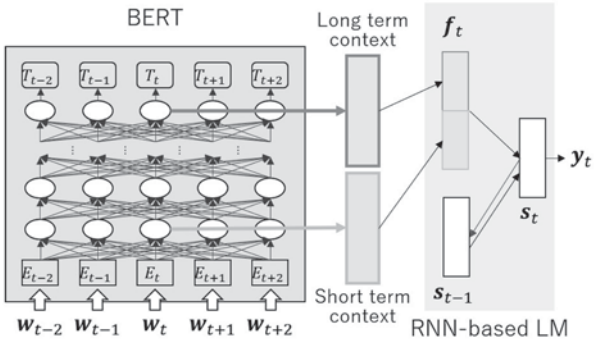


Fig.1：BERTに基づくEmbeddingを用いた言語モデル

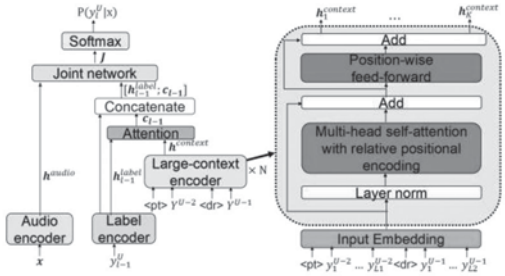
2-3Q-3

RNN-T と Transformer による長距離文脈情報を用いた音声認識の検討

A study on long-context automatic speech recognition using RNN-T and Transformer

○小島淳嗣（アドバンスト・メディア）

- 多くの音声認識は、発話単位で処理がされる
- 先行研究において、会話の音声認識を対象に、長距離文脈情報 (e.g., 過去の発話で生成された仮説) を利用する Attention -based Encoder Decoder が提案されている
- RNN-T に Large-context encoder を導入することで、長距離文脈情報を用いた音声認識を実現



- 臨床会話を用いた実験によって、 large-context encoder の有効性を確認

Method	relative character error rate reduction
RNN-T	0.0
+ large-context encoder	11.9

2-3Q-4

2-3Q-4　ROS を利用したリアルタイム制御が可能な音声対話システムの構築

Developing a Spoken Dialog System with Real-time Control Using ROS

○西村良太，△森貴大(徳島大)，△北岡教英(豊橋技科大)

◆本研究での提案システム：
- 従来システムでは不可能であった相槌などの割り込み応答に対応
- ROS アーキテクチャを導入したリアルタイム音声対話システム

◆ROS アーキテクチャ導入の利点：
- デバッグが容易になる（特にマルチモーダル環境に有効）
- ロボットやその他の環境との連携が容易になる

◆提案システムの動作検証：
- 相槌を含む応答タイミング制御モジュールを容易に構築
- 実際にあいづちを出力できることを確認
- 同時に，実験中のシステム内部のデータを rosbag コマンドで記録

◆ダッシュボードによる各種システム状態の表示：

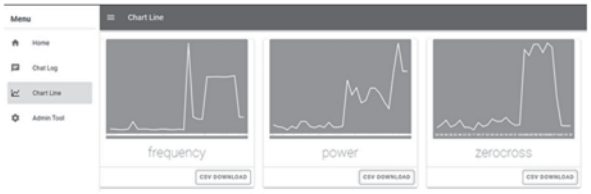


Fig.1：Dashboard view: Graph of acoustic analysis results

2-3Q-5

2-3Q-5　Conformer CPC と Deep Cluster を用いたゼロリソース言語のための表現学習

Speech Representation Learning Combining Conformer CPC with Deep Cluster for Zero-Resource Language

◎前角高史，藤田悠哉(ヤフー)，Li-Wei Chen，Xuankai Chang，渡部晋治，Alexander Rudnicky(カーネギーメロン大)

◆タスク：Zero Resource Speech チャレンジ 2021
- テキスト情報を用いずに音声データから構築した言語モデルの言語理解の性能を評価。

◆ベースライン
- CPC の特徴量を k-means クラスタリングし、離散系列に変換して BERT 言語モデルを学習。

◆提案手法：
- Deep Cluster を適用し、クラスタリング結果から得られた疑似ラベルを用いた識別タスクを追加（提案法 1）。
- CPC の系列予測用ネットワークを Transformer から Conformer に置き換え（提案法 2）。

◆実験
- 音素、語彙、文法、意味の各観点に基づいた 4 つの指標で評価。
- 3 つの指標で提案法がベースラインの性能を上回った。

	ベースライン	提案法1	提案法2	提案法1+2
音素識別スコア(↓)	6.19	**3.93**	5.4	4.05
語彙識別スコア（↑）	66.21	64.89	66.1	**67.21**
文法誤り識別スコア(↑)	52.79	52.75	**53.39**	53.38
意味的類似度相関スコア(↑)	4.92	**8.89**	5.17	7.07

Fig. 1 Overall performance of the baseline and the proposed systems

2-3Q-6

2-3Q-6　「平静」と「怒り」の感情の特性を考慮した音声感情認識のための label smoothing

Label Smoothing for Speech Emotion Recognition based on Emotional Characteristics of Neutral and Anger

☆永瀬 亮太郎，福森 隆寛，山下 洋一（立命館大）

【概要】

深層学習に基づく音声感情認識のための label smoothing の提案

【目的】

発話に含まれる感情が一つではない可能性を考慮した認識の実現

【提案手法】

先行研究が示す「平静」と「怒り」の感情の特性を考慮し，正解の感情ごとに異なる label smoothing を適応する

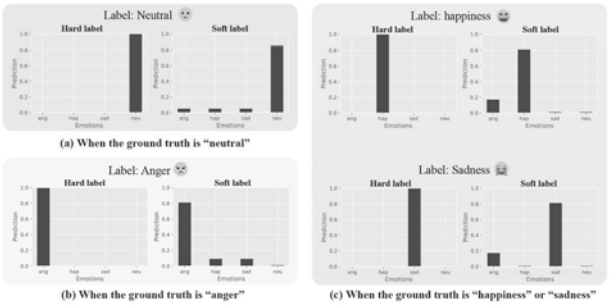


Fig. 1: An example of label smoothing in the proposed method

【結果】

One-hot 表現の正解ラベル，又は，従来法の label-smoothing を適応したラベルと比較すると，提案手法は，IEMOCAP に対する認識率向上に有効であることが分かった

2-3Q-7

2-3Q-7　音響イベント検出における GAN を用いた弱ラベルデータ生成による半教師あり学習

Semi-supervised learning using weakly labeled data generated by GAN in sound event detection

☆合馬一弥（筑波大），山田武志（筑波大），牧野昭二（早稲田大/筑波大）

◆音響イベント検出（SED）は音響データ内の発生イベントの種類，開始・終了時刻を特定する技術であり，深層学習によって実現される．

◆SED の学習には強ラベル（音響イベントクラス，開始・終了時刻からなるラベル）付きデータが必要であるが，この強ラベルの準備にかかる膨大な人的・時間的コストが課題となっている．

◆解決策として準備コストの少ない弱ラベル（音響イベントクラスのみからなるラベル）付きデータやラベルなしデータの活用が研究されている．

◆これまでに我々は GAN を用いた半教師あり学習を SED に適用し，ラベルなしデータを活用して検出精度を向上させる手法を提案した．

◆本稿ではさらに GAN によって弱ラベル付きデータを生成し，SED の学習に直接用いる手法を提案し，その有効性を示した．

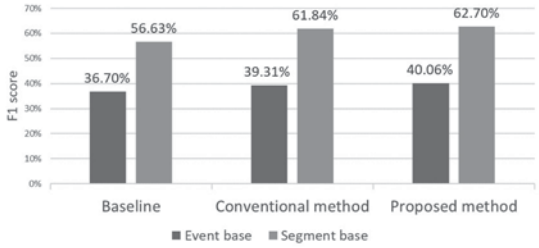


Fig. 1：F1 score in each method

2-3Q-8

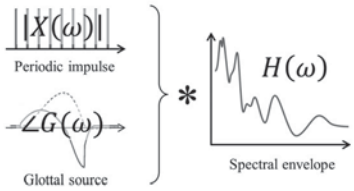
2-3Q-8　声帯音源の群遅延特性に基づく特徴量を用いた励振源による有声音波形生成法の提案と評価

Waveform generation of voiced sound considering group delay characteristics based on glottal source

☆小口純矢，森勢将雅（明治大）

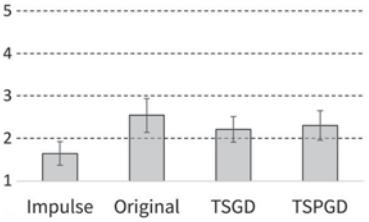
◆　声帯音源の群遅延特性を励振源に付与して波形生成

- 目的：最小位相フィルタの buzzy 感の改善



◆　以下の手法による合成音を MOS で主観評価

- Impulse：通常の周期インパルス列
- Original：声帯音源の群遅延
- TSGD：時間領域平滑化群遅延
- TSPGD：時間領域平滑化パワースペクトル重み付き群遅延



2-3Q-9

2-3Q-9　Emotional Voice Conversion by Learning Disentangled Representations with Spectrum and Prosody Features

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

◆Emotional Voice Conversion (EVC) is a voice conversion technique for converting prosody in speech, which can represent different emotions while retaining the linguistic information.

◆In this study, we present a non-parallel EVC method by learning disentangled representations with spectrum and prosody features. Our proposed framework is based on GAN which is composed of a generator and a discriminator typically. The proposed framework is shown in Fig. 1.

◆The generator is an encoder-decoder module, which is made up entirely of convolution layers in our work. The generator consists of three modules: a content encoder, a style encoder and a decoder.

◆The experimental results show the effectiveness of our proposed method.

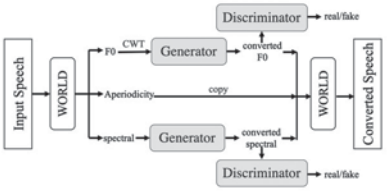


Fig. 1：Framework overview.

2-3Q-10

2-3Q-10 声質変換を用いた目標歌手の声質を再現する歌声合成システムの提案と声質変換手法による再現性の違いの分析

Proposal of a singing voice synthesis system that reproduces the voice quality of a target singer by using voice conversion and analysis of the difference in reproducibility by the voice conversion method

☆井沼望，坂野秀樹，旭健作（名城大）

◆歌声合成システムに声質変換を適用すれば，既存の歌声ライブラリを使用しながら話者性の異なる歌声が得られることが期待される．本発表では声質変換を取り入れた歌声合成システムを提案する．

◆本実験では，変換元の歌声の合成に，ニューラルネットワークを用いた歌声合成システム NEUTRINO を，声質変換システムには GMM 最尤推定による変換を行う sprocket を使用した．

◆声質の客観評価にはメルケプストラム歪み MCD を用いた．

◆学習データ数を増やすと MCD は減少し，学習データ数 28 の時，MCD は 6.09 [dB]と最小になった．（Fig.1）

◆GMM 混合数が多いと声質の不安定化が起こりうることを確認した．

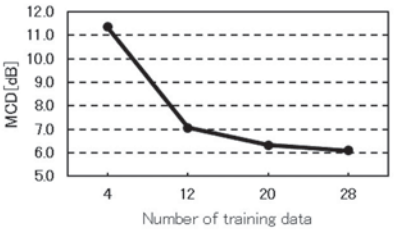


Fig.1：MCD between the coverted voice and the target voice versus the number of training data

2-3Q-11

2-3Q-11 歌唱音声における自然性評価のためのスペクトルの微細変動減衰処理によるビブラート音声の作成

Study on synthesis of vibrato voice by decrease processing for fine fluctuation of spectrum for evaluation of naturalness in singing voice.

◎鈴木千文，坂野秀樹，旭 健作(名城大)，森勢将雅(明治大)

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

◆歌声データ：ビブラート音声(目標歌手：Gackt)

◆メルケプストラム歪みの算出(Fig.1)

- 基準フレームをビブラート開始区間のフレーム(Fig.1(左))と、ビブラート終了区間のフレーム(Fig.1(右))とした場合の歪みを算出
- Fig.1(右)の方が、Fig.1(左)に比べてメルケプストラム歪みの値が全体的に大きく、基準フレーム間における差も大きい
- ビブラートの終了区間では、スペクトルの形状が大きく変化していることが考えられる

◆スペクトルの各ビンの時系列に含まれる微細変動の減衰処理によるビブラート合成音の作成

- 10Hz のハイパスフィルタにより微細変動を抽出し、重みをかけて減衰処理を行う

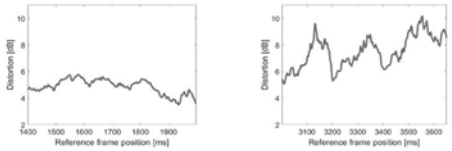


Fig.1 Mel-cepstrum distortion.

2-3Q-12

2-3Q-12 音声からの声道断面積関数推定におけるスペクトル平坦化手法による推定精度の比較

Study on comparison of accuracy of vocal tract area function estimation from voice signal under different spectrum flatting conditions

☆奥田康弘，坂野秀樹，旭健作（名城大院）

◆荒川らによる帯域分割 LSP に基づく声道断面積推定手法における推定精度改善を目標としている。

◆この手法における低域の線形予測次数、高域の線形予測次数、高域スペクトルの平坦化に使用するリフタリング次数について検討した。

◆低域の線形予測次数と高域の線形予測次数を各帯域幅に基づいて決定することで声道断面積推定精度が上がることが分かった。

◆高域スペクトルの平坦化に使用するリフタリング次数を変更しても声道断面積推定精度に大きな影響は与えないことが分かった。

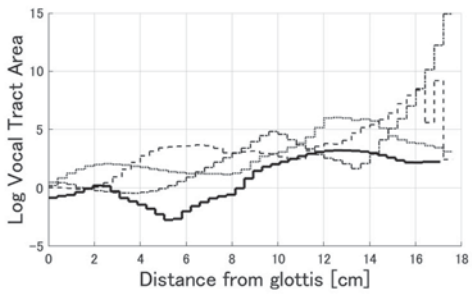


Fig. 1 Original vocal tract area function (bold line), vocal tract area function estimated by using LPC orders of low frequency band and high frequency band of 8 and 35 (one dot chain line), 10 and 33 (dot line), and 12 and 31 (dash line).

2-3Q-13

2-3Q-13 笑い声合成における音声記号表現の感情次元による制御

Generating laughter components with emotion manipulation

☆木村駿野，森大毅（宇都宮大）

◆笑い声を合成するために必要な笑い声の音声記号表現を、どのように作り出すかが課題となっている。

◆本稿では感情状態に着目し、感情次元(快-不快、覚醒-睡眠)と自然笑い声の音声記号表現との関係を機械学習によりモデル化する。

◆その結果、快-不快次元は、笑い声の長さに関連することが確認された。

◆覚醒-睡眠次元は、音声記号の出現確率に影響を与える可能性がある。

◆感情次元により、生成される音声記号表現をある程度制御できることが示された。

(4, 4)	hu hu H hu he ha hu hu hu hu
(7, 7)	hu ha a ha H hu_ h hu u hu ha H hi H hu_ H hu_ u_ H H hi he_ Na ha H hu_ H he U hu ha H hi H hu_ H hu_ u_ H

Fig.1:設定した感情次元(快不快，覚醒-睡眠)により生成された笑い声音声記号表現の例

2-4P-1

2-4P-1　コウモリのエコーロケーションにおける接近物体の認識方法の解明

How bats recognize approaching objects in echolocation

☆吉田創志[1]，長谷一磨[2,3]，小林耕太[1]，飛龍志津子[1]
（[1]同志社大院・生命医科学研，[2]名大院・環境学研，[3]学振）

◆コウモリは超音波パルスを放射し，物体から返ってきたエコーを聴取することで周囲の情報を得ている．これを**エコーロケーション**という．

◆例えばエコーが返ってくるまでの**時間差は物体までの距離**を表し，**ドップラーシフト量は相対速度**を表すことがわかっている．一方で，これまで**接近物体の認識方法**は明らかになっていなかった．

◆本研究では放射パルスを瞬時に処理し仮想的なエコーとしてプレイバックすることで，コウモリに接近する**音響的仮想物体の呈示に成功し** (Fig.1)，接近物体を認識する手がかりとしてエコー遅延の時間変化ではなく**ドップラーシフト量を用いている** (Fig.2) ことが示唆された．

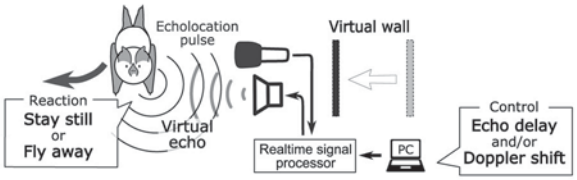


Fig.1: Experimental set up for presentation of virtual moving wall.

Fig. 2: Schematic spectrogram of pulse(black) and play backed virtual echo(gray).

2-4P-2

2-4P-2　基本周波数またはスペクトルレベルをモーフィングした感情音声に対する感情知覚

Emotional perception of speech sounds by the fundamental frequency and spectral level morphing

☆坂下尚史（豊橋技科大），河原英紀（和歌山大），松井淑恵（豊橋技科大）

◆怒り音声と平静音声のモーフィング音声に対して感情知覚実験を実施し，モーフィング時に操作する音響パラメタの影響を調べた．

◆原音声には怒りと平静の感情をこめて「はい」と発話された，男声と女声を使用した．操作するパラメタは全てのパラメタ，基本周波数（F0），スペクトルレベル（SL）の３種類とした．比較刺激のモーフィング割合は，各操作するパラメタを 0%から 100%まで 20%刻みに設定した．標準刺激は全てのパラメタを 50%の割合でモーフィングした音声とした．

◆実験はオンラインで行い，比較刺激と標準刺激のどちらが怒りの感情に聞こえるかを二肢強制選択で判断させた．操作するパラメタ３種類ごとに実験を行った．標準刺激と比べ，比較刺激の方が怒っているように感じたと回答した割合を，男声と女声ごと算出した．これらの回答結果の平均値に対し，心理測定関数として累積ガウス分布をフィッティングした．

◆F0，SL を操作した条件において，話者の性別によって丁度可知差異(JND)に 10 以上の差が存在した．全てのパラメタを操作した条件に比べ，F0，SL のみを操作した条件では標準偏差が大きくなった．怒りの感情の弁別には女性話者では F0 が、男性話者では SL がより強く影響する可能性が示唆された．

2-4P-3

2-4P-3 個人の知的作業パフォーマンスに影響する周囲環境音の特徴分析

Analysis for the characteristics of environmental sound affect individual intellectual work performance.

◎千葉 大将，伊藤 弘章，加古 達也，野口 賢一(NTT)

◆周囲環境音の種類による知的作業パフォーマンスの変化の個人差（個人性）や集団の傾向を調査するために，短期記憶課題と創造性課題を用いた被験者実験を実施した．本実験では，各被験者で同じ実験を複数回行うことで被験者ごとの個人性を評価する．

◆課題の成績や主観評価（印象）を分析したところ，課題や周囲環境音の種類による集団の傾向の違いや個人性の有無を確認した．

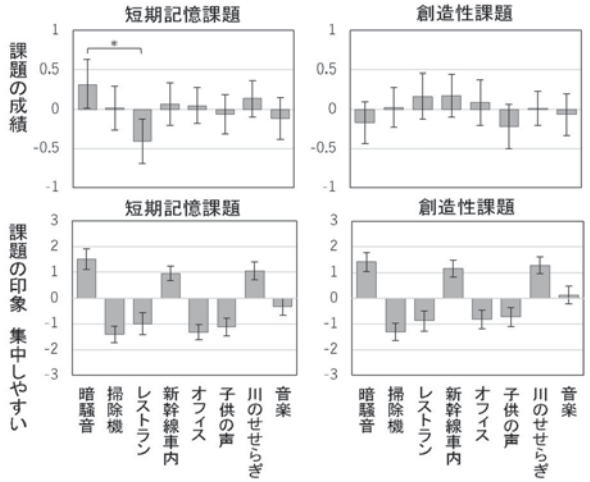


Fig.1：Average values of all subjects for each sound stimulus in task score and impression (concentration).

2-4P-4

2-4P-4　遠位呈示された骨伝導音による音声伝達性能の検討

Characteristics of speech transmission of distantly-presented bone-conducted sounds

☆石川大夢，大塚 翔，中川誠司（千葉大）

◆骨伝導デバイスが持つ装用性の問題を解決するために，体幹や上肢といった頭部から離れた部位（遠位）に呈示する方法が提案されている．

◆前報では，可聴周波数帯の遠位呈示骨伝導音の知覚特性を調査し，体内伝搬に伴う周波数及び時間情報の劣化は十分小さいことを示した．しかしながら，音声の明瞭性や音質にはいくらかの変化が生じる可能性は否定できない．

◆本報では，遠位呈示された可聴周波数帯における骨伝導音の音声伝達性能の検討を目的として，単語了解度及び単音節明瞭度試験を行った．

◆呈示部位が頭部から離れるにつれて正答率が低下する傾向が確認された（Fig. 1）．伝搬距離の増大に伴う音声信号の減衰が生じたものと考えられる．

◆異聴傾向を算出した結果，鎖骨では /tj/, /sj/, /zj/ の異聴が有意に増加していた．伝搬距離の増大に伴い高周波成分の減衰量が増加し，比較的高い周波数帯域にエネルギーが集中している音素の知覚が困難になることが示唆された．

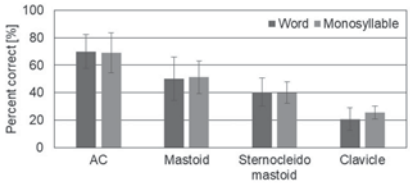


Fig.1：Scores of the word intelligibility and monosyllable articulation tests at each presentation condition and stimulus placement

2-4P-5

2-4P-5　顔面に呈示した骨伝導音による耳栓効果

An occlusion effect by bone-conducted sound presented to the face

☆美和 あす華，大塚 翔，中川 誠司(千葉大)

◆外耳道を閉塞した状態で骨伝導音を知覚すると，低域のラウドネスが増大する“耳栓効果”が生じる．

◆近年，骨伝導はスマートグラスをはじめとする様々なコミュニケーション・デバイスに利用されており，顔面への骨伝導を利用する製品も存在する．しかしながら，顔面に骨伝導呈示した際の耳栓効果の報告例は皆無である．

◆顔面に位置する頬骨と従来部位（乳様突起，顆状突起および前額部）に骨伝導呈示を行った際の聴覚閾および外耳道内音圧を計測し，それぞれの条件における耳栓効果を推定した．

◆頬骨における耳栓効果は，従来部位よりも小さくなった．

◆頬骨は従来部位と比較して外耳道までの距離が大きく，外耳道内放射成分が減少することが示唆された．

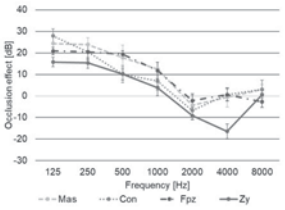


Fig.1 Occlusion effects obtained from the hearing thresholds (Mean ± S.D.).

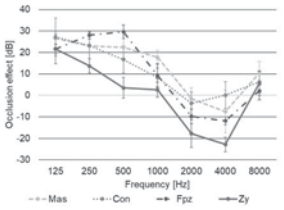


Fig.2 Occlusion effects obtained by the ear-canal sound pressure (Mean ± S.D.).

2-4P-6

2-4P-6　規則性に基づく予期によってオリーブ蝸牛束反射が変動するメカニズムの解明 −聴性脳幹反応との時間的相関性に基づく検討−

Mechanisms Underlying the Enhancement of Medial Olivocochlear Reflex Associated with Rhythm-based Prediction : An investigation Based on Temporal Correlation with Auditory Brainstem Response

◎石坂勇毅，大塚翔，中川誠司（千葉大）

◆ 脳幹から外有毛細胞へは，オリーブ蝸牛束と呼ばれる遠心性神経が接続し，音に反応して，外有毛細胞の増幅を抑制することで，強大音から内耳を保護していると考えられている。この反応はオリーブ蝸牛測反射（MOCR）と呼ばれている．

◆ 著者らはこれまでに，予期に伴う MOCR 及び皮質活動の反応強度の時間変動が相関することを報告してきた．皮質からオリーブ蝸牛束へは脳幹の神経核を経由して，遠心性神経が接続している．本研究では，MOCR と皮質誘発電位に加えて，聴性脳幹反応(ABR)も同時計測し，予期によって MOCR 強度が増加する機序を検討した．

◆ 刺激が規則的な場合(Regular 条件)と不規則な場合(Irregular 条件)を交互に呈示し，2 条件間の MOCR 強度差($\Delta MOCR_{reg}$)，N1-P2 の振幅値，及び，ABR の第Ⅴ波の振幅値の時間変動を比較した．

◆ $\Delta MOCR_{reg}$と ABR 振幅の時間変動の間には相関は見られなかった．一方で，$\Delta MOCR_{reg}$と N1-P2 の振幅値の時間変動には負の相関があることが示された(Fig.1)．この結果は，脳幹反応の時間変動が，MOCR の予測的制御に関わる神経基盤に関与している可能性が低いことを示唆している．

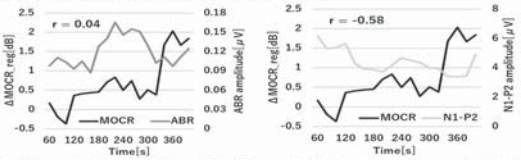


Fig. 1 A typical example of temporal variations of $\Delta MOCR_{reg}$ and ABR(left), and $\Delta MOCR_{reg}$ and N1-P2(right)

2-4P-7

2-4P-7　特定の周波数に対する選択的注意がオリーブ蝸牛束反射に及ぼす影響

Effect of selective attention to specific frequencies on medial olivocochlear reflex

☆菊池梢馬，石坂勇毅，大塚翔，中川誠司（千葉大）

◆ 脳幹から外有毛細胞へは、オリーブ蝸牛束と呼ばれる遠心性神経の束がある。オリーブ蝸牛束は音響刺激に反応し、外有毛細胞の増幅を抑制する働きがある。この反応はオリーブ蝸牛束反射（MOCR）と呼ばれている。

◆ MOCR は聴覚的注意によって周波数選択的に増強することが報告されている。しかし、その効果のチューニング特性は明らかになっていない。

◆ 本研究では、単一の周波数に注意を向けた場合と注意を向けなかった場合での MOCR の周波数特性を比較した。MOCR の周波数特性は、耳音響放射（外有毛細胞の増幅を反映）のスペクトルを分析することで評価した。

◆ 雑音中に含まれる 1500 Hz の純音を検出する課題を行うと、500 から 3000 Hz にかけての広い帯域で MOCR の強度（耳音響放射の抑制量）が増加する傾向が見られた（Fig. 1）。特定の周波数に対する注意が MOCR へ与える影響の周波数選択性は，聴覚フィルターより緩やかであることを示唆している。

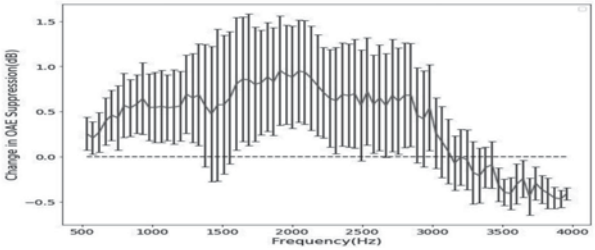


Fig.1 : Mean difference in frequency characteristics of MOCR with and without the auditory attention task. Error bars depict the standard errors.

2-4P-8

2-4P-8 ZESTによる無音検出閾値測定のシミュレーション

Computer simulations of ZEST procedure of gap detection

☆村田悠登（九州大），森本隆司（リオン），森周司（九州大）

◆無音検出閾値測定の測定時間短縮のため、適応的閾値測法の一種である ZEST を適用することをコンピュータシミュレーションによって検討した。

◆ZEST では測定開始前に実験参加者の心理測定関数と、閾値の事前推定を表す確率密度関数を設定しておく必要がある。

◆シミュレーションでは確率密度関数をガウス関数で表し、その平均値と標準偏差が測定誤差に与える影響を調べた。

◆結果として確率密度関数の標準偏差が大きいほど測定バイアスが小さくなる一方で、閾値推定値の標準偏差は大きくなるという傾向が見られた。

◆今後はシミュレーションにおいて測定誤差が小さかったパラメータ設定を使用した実際の測定を行い、その妥当性を検討する。

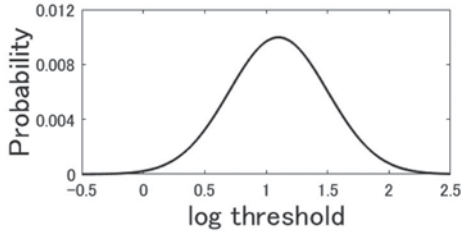


Fig 1 probability density function with small measurement error

2-5P-1

2-5P-1　山形県総合文化芸術館の音響について

The Acoustical Design of Yamagata Prefectural Arts Center

○司馬義英，井上瑞紀（日建設計），上田　泰孝，吉田　卓弥（安藤ハザマ）

山形県総合文化芸術館（やまぎん県民ホール）は山形駅西口に立地し、東北でも有数の客席数となる約 2000 人収容の大ホールを中心に、スタジオや練習室群からなる文化施設である。
1998 年の設計プロポーザル以来 20 年強を経て 2020 年にオープンした。 山形のものづくり技術と伝統を生かし山形の風土に根差した建築となっている。
大ホールは演劇・音楽・式典・講演など様々な催しに対応するよう計画された。 隣接鉄道からの固体伝搬音対策、室間遮音や静粛性の対策、各空間の室内音響・舞台音響設備計画について報告する。

2-5P-2

2-5P-2　浜松市市民音楽ホール「サーラ音楽ホール」の音響設計 —吹奏楽や合唱の大会・発表会の利用に配慮した音響調整機能を有する市民ホール—

Acoustical Design of the Hamamatsu City Civic Music Hall "SALA Music Hall"
- A Civic Hall with acoustic adjustment function mainly used for contest and recital for brass band, chorus -

○日根野翔太，高橋顕吾（ヤマハ），岸永伸二

◆本施設は、浜松市における新たな「市民の音楽文化活動や学習成果等の発表の場の創出」を目的として計画された。
◆主に学校教育団体や吹奏楽などの音楽団体をメインターゲットとしており、それら大会運用に配慮された平面計画、音楽利用を主体としながらも講演会などにも対応可能な多目的ホール、ホールと同時使用が可能な高度な遮音性能を有する多目的室などで構成されている。
◆ホールの基本形状をシューボックス型とし気積を確保することで、1 次反射音と響きを確保した。さらに合唱や吹奏楽など用途によって最適な響きを調整できるよう吸音カーテン等を配置し、響きや反射音を制御した。
◆多目的室については、大会運用に配慮し、ホールと同フロア、近距離で配置されているため、室間の遮音性能確保が要求された。そのため多目的室の RC 化やホール間に Exp.J を設け、ホール使用時にも音出し可能な遮音性能を確保した。

Fig.1 Acoustic concept

2-5P-3

2-5P-3　桃山学院聖アンデレ礼拝堂の音響設計

Acoustic Design and Characteristics of Momoyama Gakuin St.Andrew's Chapel

○相馬　真子，木村文紀，長野　武，板垣篤恵（竹中工務店）
荒木　邦彦，安井　千絵（音響デザイン研究所）

◆日本聖公会大阪聖アンデレ教会（桃山学院聖アンデレ礼拝堂）は、2020 年 8 月、学校法人桃山学院昭和町キャンパスの聖テモテ館 1 階に竣工した。本礼拝堂は、1884 年創立からの歴史ある教会として桃山学院生徒の礼拝やクラブ活動の他、日々の説教や聖歌、行事等、学外に開かれ親しまれてきた旧礼拝堂の建替えである。本報告では礼拝堂の室内音響計画について報告する。
◆「礼拝時における説教を会衆席後方まで明瞭に届けること」と「聖歌隊や楽器等の生演奏にふさわしい豊かな響きが得られる音響空間を実現すること」を目標に計画した内容を報告する。

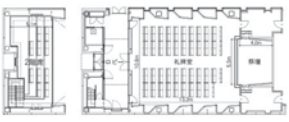


Fig.1：Horizontal Plan

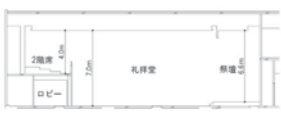


Fig.2：Vertical Plan

2-5P-4

2-5P-4　西脇市市民交流施設「オリナス」の音響設計—「穴あきブロック」で構成された側壁の音響性能把握と設計監理—

Acoustical Design of the Nishiwaki Community Hall "Orinas" -Understanding Acoustic performance of the "Perforated Block" and Design supervision-

○中川貴美子，宮崎秀生（ヤマハ）

◆「オリナスホール」の音響的な特徴の一つに、特殊な内装材を使用しているという点が挙げられる。設計当初からその音響性能の把握が課題となっており、施工段階での検討とそれを反映した最終仕様、施工状態での実測まで、一連の音響検討と結果にフォーカスして報告する。
◆ホールの内装材としては特殊な「穴あきブロック」の音響性能を把握するため、現場段階でモックアップを作成し、残響室法吸音率測定を実施した。測定結果より、ブロックの背面にボードを密着させると特定の周波数で吸音のピークを持つこと、背後空気層を設けるとピークは小さくなること、背後空気層の厚みによって吸音特性が異なることが分かった。測定結果より最終仕様を決定し、施工状態で実測したところ、周波数特性に大きなピーク・ディップは見られなかった。
◆ホールとしての各種音響特性も良好な結果が得られた

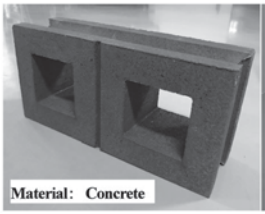


Fig.1：Perforated Block

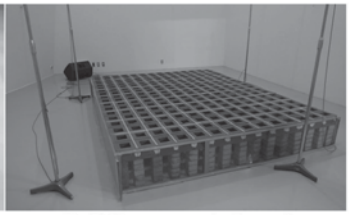


Fig.2：Measurement of sound absorption in a reverberation room

2-5P-5

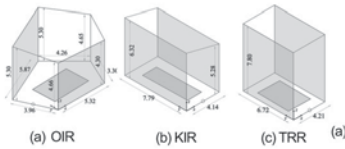
2-5P-5　残響室法吸音率の測定音場に関する有限要素解析 -壁面の傾きの検討-

Finite element sound field analysis on measurement of absorption coefficient in a reverberation room -Investigation on inclination of walls-

☆岩本駿（大分大院・工学研），富来礼次，岡本則子，
大鶴徹（大分大・理工），鈴木彰馬（大分大院・工学研）

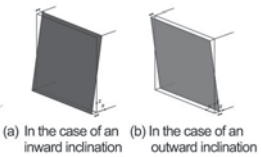
◆形状の異なる 3 つの残響室に対して，時間領域有限要素法を用いて，残響室壁面の傾きが残響室法吸音率の測定結果に及ぼす影響を検討した．

◆インパルス応答から残響時間を求め，残響室法吸音率を推定したところ，不整形七面体残響室では壁面の傾きによる推定値の大きな違いは見られなかった．一方で，平行壁をもつ不整形六面体残響室や矩形残響室では壁面の 6° 未満の傾きでも測定値の改善が見られた．



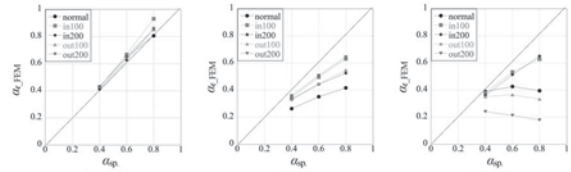
(a) OIR　(b) KIR　(c) TRR

Fig.1 : Schematic drawing of reverberation Rooms and location of specimen.



(a) In the case of an inward inclination　(b) In the case of an outward inclination

Fig.2 : Pattern diagram of inclination of walls.



(a) OIR　(b) KIR　(c) TRR

Fig.3 : Relationship between α_{r_FEM} and α_{sp} for each wall inclination condition .

2-5P-6

2-5P-6　有限要素法を用いた残響室法吸音率の面積効果に関する検討

Investigation on the area effect of sound absorption coefficient measurement in a reverberation room by finite element method

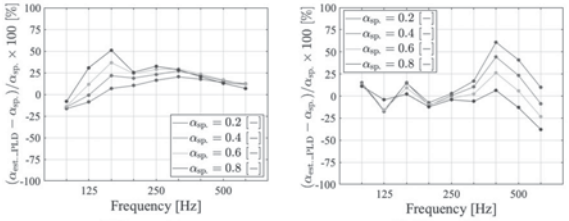
☆鈴木彰馬（大分大院・工学研），富来礼次，岡本則子，
大鶴徹（大分大・理工），星野嗣人（鹿島技研）

◆残響室法吸音率の面積効果に関し，時間領域有限要素法（TDFEM）を用いて Deep-well と測定値の関係を検討した

◆不整形七面体残響室（OIR）を対象に，TDFEM で与えた試料の統計入射吸音率 α_{sp} に対する，TDFEM で得られた減衰曲線を PLD 補正して算出した残響室法吸音率 α_{est_PLD} の相対差を示した

- ✓ Deep-well 高さ $H_{dpwl.}$ = 0.2 m の場合（Fig.1 (a)）では 200 Hz 以上の周波数で α_{sp} に対する α_{est_PLD} の相対差が小さくなる
- ✓ Deep-well 高さ $H_{dpwl.}$ = 0.8 m の場合（Fig.1 (b)）では 250 Hz 以下の周波数で α_{sp} に対する α_{est_PLD} の相対差が小さくなる

➢ 対象とする波長に応じて，有効な Deep-well の高さが変化することを確認した



(a) $H_{dpwl.}$ = 0.2 m　(b) $H_{dpwl.}$ = 0.8 m

Fig.1 Relative difference of absorption coefficient $\alpha_{est.}$ from $\alpha_{sp.}$ measured in OIR.

2-5P-7

2-5P-7　アンサンブル平均を利用した材料の吸音特性の *in-situ* 測定法に関する研究

−統計入射吸音率の算出および残響室法吸音率との比較−

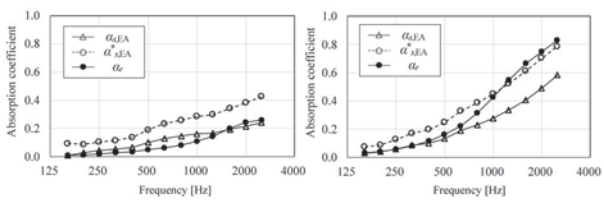
In-situ sound absorption characteristics measurement method by ensemble averaging technique
-Calculation of random incident sound absorption coefficient and comparison with sound absorption coefficient in a reverberation room-

☆田原有彩（大分大院・工学研），岡本則子，富来礼次，
大鶴徹（大分大・理工）

◆筆者らは，従来より，アンサンブル平均を利用した材料の吸音特性の *in-situ* 測定法(EA 法)を提案している．

◆本研究では，EA 法による吸音測定を行い，得られた吸音特性から統計入射吸音率を算出した．さらに，残響室法吸音率と比較し，その関係性を明らかにした．

◆その結果，吸音率が低い帯域では Paris の式で算出した統計入射吸音率 $\alpha_{s,EA}$ が，吸音率が高い帯域では，London の式で算出した統計入射吸音率 $\alpha^{*}_{s,EA}$ がそれぞれ残響室法吸音率 α_r と近似した．



(a) CT　(b) NF

Fig. 1 Comparisons among $\alpha_{s,EA}$, $\alpha^{*}_{s,EA}$, and α_r for CT, GW25, and NF.

2-5P-8

2-5P-8　複合現実技術を用いた三次元音響インテンシティの可視化システムへの PAGE 法の適用

Application of PAGE method to Visualization system for three-dimensional acoustic intensity using Mixed Reality

☆内田彩芽，大川祐貴子，渡辺靖明，
池田雄介（東京電機大），及川靖広（早大理工）

◆目的：
複合現実技術を用いた三次元音響インテンシティの可視化システムの推定する音響インテンシティの精度の改善

◆提案システム
音響インテンシティ可視化システムへの PAGE 法の適用

◆ 可視化結果
高い周波数において従来手法であるクロススペクトル法より PAGE 法の方が推定精度の改善が見られた。

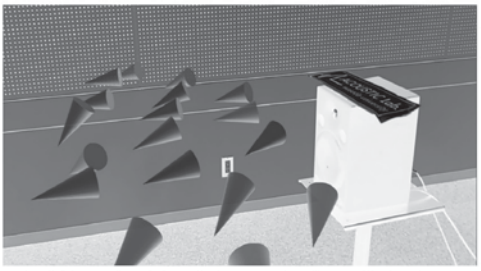


Fig.1 : Example of visualization result of sound intensity using PAGE method

2-5P-9

2-5P-9　広帯域音場に対する瞬時軸ビームフォーミング

Instantaneous Axis-Beamforming for Broadband Sound Field

◎岩見貴弘，尾本章（九大・芸工）

◆球上帯域制限空間の再生核に基づき，広帯域信号に対する単純な波数領域フィルタとして瞬時軸ビームフォーミングを実現した。

◆提案手法は瞬時の空間情報からビームフォーミングを行うため，波の到来方向の前後識別が不可能であり，故にあるステアリング軸に対する手法となる。

◆提案手法は広帯域信号に対して適用する際，狭帯域信号の結果を算出して統合する必要がない。

◆数値実験より，高周波数域になるほどステアリング軸での利得が大きくなることを確認した。

◆空間的ビームフォーミングの数値実験より，低域になるほど誤差は大きくなるが，有効範囲は広くなることが示された。

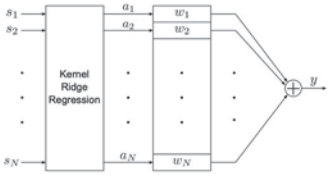


Fig.1　A block diagram of the instantaneous axis-beamforming system.

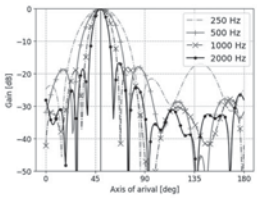


Fig.2　Gain of the proposed method with the angle for each band.

2-5P-11

2-5P-11　英語リスニングテスト実施教室の音声明瞭度と残響時間に関する研究

Speech Intelligibility and Reverberation Time in Classrooms Administering English Listening Tests

☆本野広大，△川田牧人，鶴田(濱村)真理子，長谷川光司(宇都宮大)

◆外国語学習者のリスニングテストに教室の音響環境が影響を与える可能性がある。

◆本研究では，大学で外国語の試験会場として使用されている８教室と，改装された２教室において，３種類のスピーカ (壁掛けスピーカ，ラジオカセットプレーヤ，アンプ付きスピーカ) を用いて音声明瞭度 (STI) と残響時間 (RT) を測定した。STI の測定値を Fig. 1 に示す。

◆アンプ付きスピーカが，教室の大きさなどに関わらず STI 値が高い場所が多かった。残響時間は推奨値である 0.6 秒を満たす教室が多い一方で，残響時間が１秒を超える教室もあることが確認された。

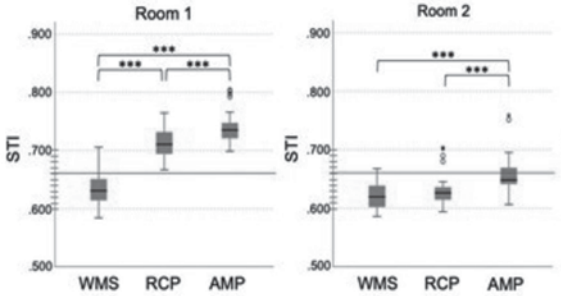


Fig.1：STI measurement results in each test room and effect of types of speaker (WMS: Wall-mounted speakers, RCP: Record and Cassette Player, AMP: Speaker with amplifier).

2-5P-10

2-5P-10　FDTD 法により複数の空間音響特性を一様化した混合音響信号の印象評価

Impression evaluation of mixed acoustic signals with equalized spatial acoustic characteristics using FDTD method

○大久保翔太，堀内俊治(KDDI 総合研究所)

◆CE-FDTD 法を用いて算出した実空間がモデルの空間インパルス応答を使用して，複数空間で収音された音響信号の空間音響特性を一様化し，知覚的印象を調査するための聴取実験を実施した(Fig.1).

◆Fig.1 の①は収録した TSP 信号を用いた音響信号であり，FDTD 法を用いた音響信号が②~④である。「好き」「自然な」「心地よい」の項目で平均値が正の値であることから実際に収音した音響信号は異なって認識される結果になったが，④においては平均値が 0.5 以下であることから周波数分布が狭く突発的な音を含まない楽器の音響信号においては聴感上自然な結果と考えられる.

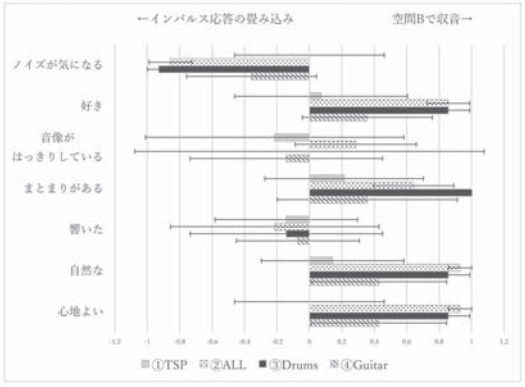


Fig.1：Result of impression evaluation about mixed acoustic signals