3-1-1

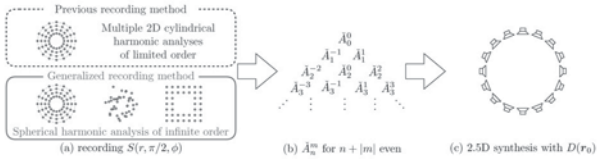
3-1-1　平面分散マイクロホンを用いた水平面3次元音場収録

Horizontal 3D sound field recording using planarly distributed microphones

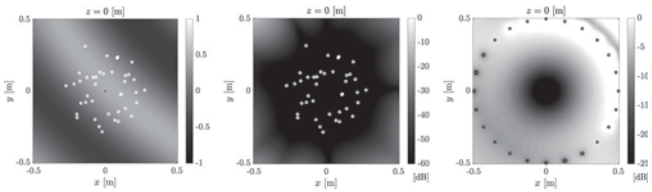
○岡本拓磨 (NICT)

Q：「2次元平面配置の分散マイクロホンで収録した3次元音場を円形スピーカアレイで再生することってできるんですか??」
オカモト：「それならこれで!!」

(a)：3次元音場を任意の平面配置マイクロホンアレイで収録
(b)：無限次元調和解析で球面調和スペクトルの偶数成分を推定
(c)：推定した音場を2.5次元高次アンビソニクスで再生



・従来法(Okamoto, ICASSP 2019)は複数円形アレイのみ対応
・無限次元調和解析(Ueno, IEEE SPL 2018)を導入し任意の平面配置に拡張



3-1-2

3-1-2　鋭指向性マイクアレイによる高次アンビソニックス収録法の検討

A study of higher-order Ambisonics encoding using narrow directional microphone array

☆柏﨑紘，岩見貴弘，尾本章（九州大･芸工）

◆鋭指向性マイクアレイを使って音場再生システムの性能向上を試みている。低周波数域ではマイクの指向性が緩いため，これを高次アンビソニックスやビームフォーミングを使って補正したい。
◆高次音圧勾配を使って任意の軸対称指向性をモデル化し，球面調和関数展開する。
◆マイクの指向性が鋭いほど，ロバスト性は向上するが，推定誤差は増加するというシミュレーション結果を得た。実際の鋭指向性マイクアレイへのフィッティングも行い，高次音圧勾配を用いることでモデル化誤差が減少することを確認した。

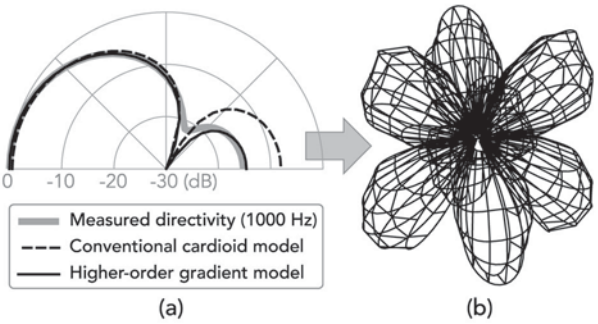


Fig.1 : (a) Directivity model fitted to measured values, and (b) example of reconstructed directivity.

3-1-3

3-1-3　HATS の周波数特性補正に関する一考察 －音響トラップ付双指向性マイクによる補正－

Consideration concerning frequency response correction of HATS -by bi-directional microphone with sound trap-

○稲永潔文（サザン音響）、櫻井一夫（ヤシマ電気）

◆バイノーラル収音/再生は、手軽に 3D 音場再現が可能な、古くから知られている優れた 2ch.の収音/再生方式である。
◆しかし、HATS 頭部や耳殻形状、マイク設置条件等で一義的に定まる、ハイ上がりの いわゆるハスキーボイスになってしまう欠点があった。
◆そこでこの独特な凸型周波数特性を補正するため、双指向性マイク振動板背部に音響トラップ回路を組み込むことにより、特性補正のための電気回路を用いる事無く、マイク単体でその補正を効果的に行う事を試みた。
◆その結果、従来無指向性マイクと入れ替えるだけで、HATS 固有の凸型特性が容易に補正され、拡散音場に於いても HRTF をほぼフラットに補正可能なことが明らかになった。
◆このHATS で収録された音は、無指向性マイクを用いた従来マイク収録時と比較して、立体音場感を損なうことなく、高域強調の無い、トーンバランスの自然な音で収音出来ることが確認された。
◆

Fig.6 SAMREC Type 2500

Fig.7 Mic. Arrangement

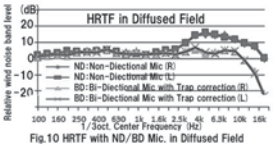


Fig.10 HRTF with ND/BD Mic. in Diffused Field

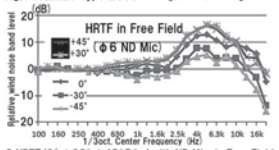


Fig.8 HRTF (0°,±30°,±45°@1m) with ND Mic. in Free Field

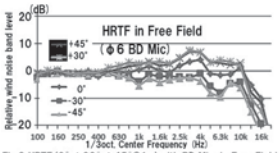


Fig.9 HRTF (0°,±30°,±45°@1m) with BD Mic. in Free Field

3-1-4

3-1-4　水平面に配置したスピーカにおける帯域卓越処理による上方音像制御

Control of a sound image to the upward direction using band-boost filters and horizontally located loudspeakers

☆西晴貴(千葉工大･院)，飯田一博（千葉工大･先進工）

◆水平面に配置したスピーカによって 22.2ch の正中面上のスピーカ (TpFC, TpC, TpBC)を代替することを目的とし，上方の方向決定帯域を卓越して音像定位実験を行い以下のことを示した.
◆各目標方向への知覚率は Fig.1 に示す．各目標方向への最大の知覚率は TpFC で 78%，TpC で 72%，TpBC で 50%であった.

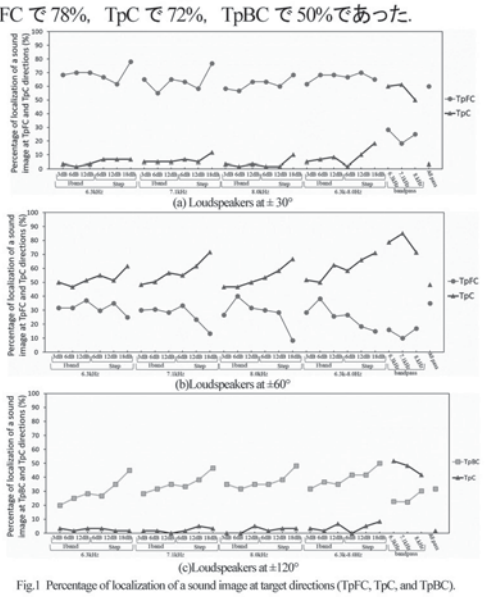


Fig.1 Percentage of localization of a sound image at target directions (TpFC, TpC, and TpBC).

3-1-5

3-1-5　オーディオ用インシュレータのホーン形状が置載機器の振動特性に与える効果

Effects on vibration characteristics of loaded equipment due to horn shape of audio insulator

○喜多雅英，西村公伸（近畿大）

◆オーディオ用インシュレータの働きや効果は明確ではないが，使用方法から置載機器の振動状態に影響を与えると推察される。

◆インシュレータ形状の影響について，円柱，円錐，指数関数形状のホーンを採り上げ（Fig.1），振動伝達特性と置載機器の振動状態の変化を実験的に評価した。

◆ホーンの振動減衰量を測定した結果（Fig.2(a)~(c)）は，ホーンの加振点からみた駆動点インピーダンスの数値計算結果と同様の傾向を示した。

◆“Mouth”側を加振側（機器側）した場合，交差周波数以上の帯域で置載機器の振動低減効果が得られる。

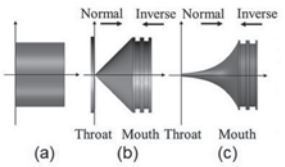
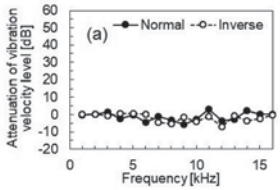
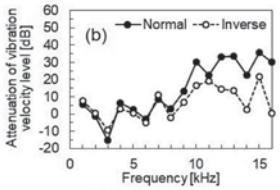


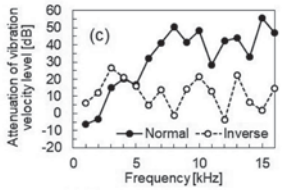
Fig.1：Side view of audio insulators.



(a) Column shape.



(b) Conical horn shape.



(c) Exponential horn shape.

Fig.2：Experimental results of vibration transmission characteristics related with transmitting direction.

3-1-6

3-1-6　ブラックボックスな聴感評点を最大化するためのDNN 音声強調の安定学習法

Differentiable Approximation of Subjective Sound Quality for Stable Training of DNN-based Speech Enhancement

☆河中昌樹（徳山高専），小泉悠馬（NTT），
宮崎亮一（徳山高専），矢田部浩平（早稲田）

◆雑音混入信号から強調した出力音声の主観品質を向上させることは音声強調において重要である.

◆主観品質の改善のために主観評価を模擬した客観評価指標（聴感評点）を最大化するようにDNN を学習する方式が検討されているが，学習が不安定である.

◆本研究では，関数近似に基づくDNN 音声強調の安定化学習法として，新たな目的関数の導入と機械学習の安定化技術を利用する方式を提案する.

◆聴感評点としてPESQ を用いた評価実験より，

(i) パラメータ更新回数に伴ってPESQ を安定して向上させ(Fig.1)，

(ii) public dataset において高い主観品質を実現した（Fig.2）.

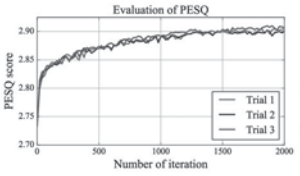


Fig.1：Relationship between number of iterations and PESQ score on test-dataset.

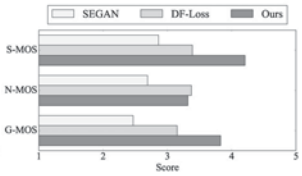


Fig.2：Result of subjective evaluation.

3-1-7

3-1-7　サンプリング周波数ミスマッチのブラインド補償に基づく音響オブジェクトキャンセラー

Acoustic Object Canceller using Blind Compensation of Sampling Frequency Mismatch

☆河村隆生（徳山高専），小野順貴，シャイブラーロビン，
若林佑幸（首都大），宮崎亮一（徳山高専）

◆一般にモノラル録音から非定常な雑音を除去することは非常に困難である．ただし，携帯電話の通知音，ラジオやテレビ放送，CD やストリーミングされた音楽など既知の音であれば，その信号を簡単に取得可能である.

◆本研究では，このような雑音を「音響オブジェクト」と定義し，モノラル録音から既知の雑音を除去する方法を提案する．録音のサンプリング周波数と利用可能な音響オブジェクトのサンプリング周波数は一致しないため，サンプリング周波数ミスマッチのブラインド補償を行い，補助関数法を用いた最尤推定により音響オブジェクトを録音から除去する．実験により提案手法の有効性を確認した (Fig.1 参照) .

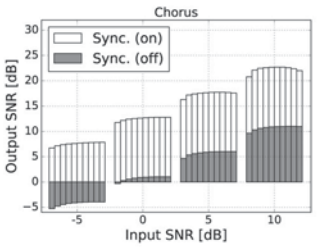


Fig.1：Output SNR with recorded data.

3-1-8

3-1-8　軽量な RNN を用いた音声強調

Speech enhancement with small RNN

◎竹内大起，矢田部浩平（早大理工），小泉悠馬，原田登（NTT），及川靖広（早大理工）

◆DNN を用いた時間周波数マスキングによる音声強調

➢ 再帰的ニューラルネットワーク（RNN）が広く適用

◆RNN は時系列データのモデリングが可能な一方で，誤差逆伝播時に勾配の発散や消失が伴い，学習が不安定

◆LSTM はRNN の勾配の発散を緩和し安定した学習が可能

➢ 一方で，３つのゲート構造によりパラメータ数は増加

◆本稿ではEquilibriated RNN を時間周波数マスク推定に適用

➢ BLSTM の 1/9 以下のパラメータ数で同等以上の性能を実現

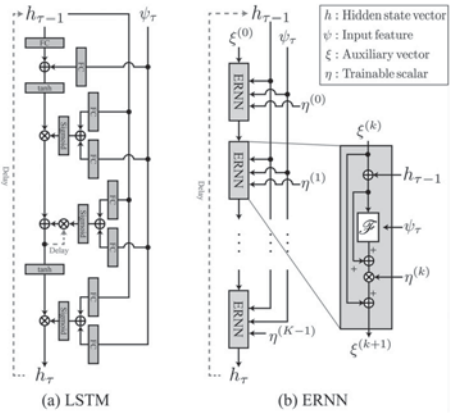


Fig.1：Block diagram of LSTM and Equilibriated RNN

3-1-9

3-1-9　DNN マスク推定に基づく畳み込みビームフォーマによる音源分離・残響除去・雑音除去の同時実現

Simultaneous realization of denoising, dereverberation, and source separation, using DNN-supported mask-based convolutional beamforming

☆高橋理希（筑波大），中谷智広，落合翼，木下慶介，池下林太郎，
Marc Delcroix，荒木章子（NTT），牧野昭二（筑波大）

◆近年、マスク推定に基づくビームフォーマ技術が盛んに研究されている。これらの技術では、ニューラルネットワーク(NN)や空間クラスタリング（SC）を用いてマスクを推定する。しかし、雑音や残響がある環境で収録した複数話者の混合音に対しては、その推定精度が低下するという問題があった。

◆本稿では、これを解決するために、NN に基づくマスク推定（パーミュテーション不変学習 uPIT）、SC に基づくマスク推定（雑音重畳型複素ガウス混合モデル noisyCGMM）、および、雑音除去・残響除去・音源分離を同時実現する畳み込みビームフォーマ（WPD）の３つの技術を単一の最適化フレームワークに統合する方法を提案する。

◆実験により、雑音残響下複数音源環境で各目的音声を推定する課題に対し、提案法が極めて有効であることを示す。

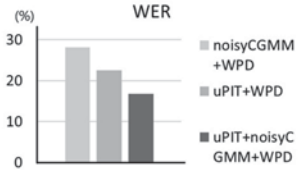


Fig. 1: Results of speech enhancement.

3-1-10

3-1-10　遠隔収録音声からの雑音残響同時抑圧のための最尤畳み込みビームフォーマ

ML convolutional beamformer for denoising and dereverberation

○中谷 智広（NTT），△Christoph Boeddeker（Paderborn 大），
木下 慶介（NTT），△Reinhold Haeb-Umbach（Paderborn 大）

◆話者から離れた位置に置かれたマイク（遠隔マイク）で収録した音声から，最尤法に基づき，雑音と残響を同時に抑圧する最尤畳み込みビームフォーマ (BF)を紹介する．

◆REVERB チャレンジや CHiME-3/4/5 チャレンジなど，近年の遠隔発話音声認識チャレンジでは，Weighted Prediction Error 最小化 (WPE) 法に基づく残響抑圧と最小分散無歪応答 (MVDR) BF や最小パワー無歪応答 (MPDR) BF などの雑音抑圧とをカスケード接続した音声強調前処理が，state-of-the-art 手法として広く利用されている．

◆しかし，これらの手法では，残響抑圧と雑音抑圧は個別に最適化されるため，処理の最適性は保証されていなかった．

◆これに対し，本稿では，主に，以下の３点を示す．

1. 最尤法に基づき全体最適化を行う最尤畳み込みビームフォーマは，従来のカスケード接続を凌ぐ性能を達成できる．
2. 最尤畳み込みビームフォーマは，WPE 法と weighted MPDR (wMPDR) BF とのカスケード接続に厳密に分解できる．
3. 従来のカスケード接続に対する最尤畳み込みビームフォーマの優位性は，MVDR BF や MPDR BF に対する wMPDR BF の優位性に由来する．

3-1-11

3-1-11　三重対角型周波数共分散行列を用いた独立半正定値テンソル分析によるブラインド音源分離

Independent positive semidefinite tensor analysis using tridiagonal frequency-covariance matrix for blind source separation

◎近藤樹（東大），高宗典玄（東大），北村大地（香川高専），猿渡洋（東大），
池下林太郎（NTT），中谷智広（NTT）

◆独立半正定値テンソル分析（IPSDTA）において、従来は計算コストの観点から周波数共分散行列に対してブロック対角制約を課していた。しかし、異なるブロックの周波数との相関を考慮できないという問題点があった。本稿では、周波数共分散行列を三重対角行列とすることで全ての隣接周波数間の相関を考慮することを考える。

◆補助関数法を用いてコスト関数の単調非増加性が保証された更新式を導出する。

◆実験により性能を評価し、その性質について考察する。

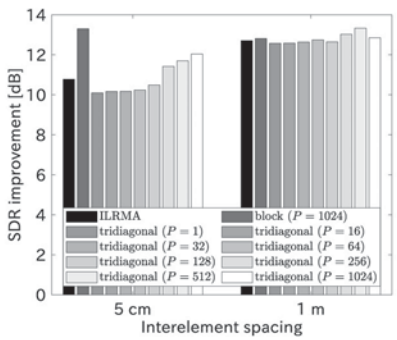


Fig.1 : SDR improvement of ILRMA, block-IPSDTA and tridiagonal-IPSDTA. P is number of blocks for 2049×2049 covariance matrix.

3-1-12

3-1-12　同時対角化行列の事前分布を用いた高速多チャネル非負値行列因子分解によるブラインド音源分離

Fast multichannel nonnegative matrix factorization with prior distribution of joint-diagonalization matrix for blind source separation

☆加茂佳吾，久保優騎，高宗典玄(東大)，北村大地(香川高専)，
猿渡洋(東大)，高橋祐，近藤多伸(ヤマハ)

◆FastMNMF は多チャネル非負値行列因子分解 (MNMF) の空間相関行列が同時対角化可能であるという仮定を置き，計算コストを大幅に改善し，高速化した手法である．

◆本稿では，独立低ランク行列分析 (ILRMA) の分離行列と FastMNMF の同時対角化行列に密接な関係があることを示し，同時対角化行列の事前分布を導入した正則化 FastMNMF を提案する．

◆正則化によって，通常の FastMNMF で用いられている反復射影法では同時対角化行列を最適化できないため，vectorwise coordinate descent を用いた更新式を導出する．そして音楽信号を用いた音源分離実験により，提案法が従来法よりも分離性能が改善することを示す．

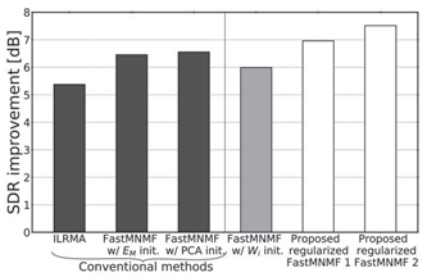


Fig.1 : Average SDR improvements of each method.

3-1-13

3-1-13　ランク制約付き空間共分散行列推定法に基づく拡散性雑音下でのブラインド複数方向性音源分離

Blind multiple directional source separation in diffuse noise environments based on rank-constrained spatial covariance matrix estimation

◎久保優騎（東大），高宗典玄（東大），北村大地（香川高専），猿渡洋（東大）

◆ 複数の方向性目的音源と全方位から到来する拡散性雑音が混在する状況を扱う．

◆ ランク制約付き空間共分散行列推定法は目的音源が１つの場合に，ランク１目的音源空間相関行列とランクが一つ落ちた雑音空間相関行列が独立低ランク行列分析で正確に推定できることに基づく音源分離手法である．独立低ランク行列分析の後処理として欠けた雑音ランク１空間成分を補い，目的音を抽出する．

◆ 本稿では方向性音源が複数個存在する場合へランク制約付き空間共分散行列推定法を拡張し，majorization-minimization (MM) アルゴリズム及び majorization-equalization (ME) アルゴリズムによりパラメタ推定を行う手法を提案する．

◆ 提案モデルは空間特性を表す行列変数をパラメタに持つ．この行列変数を ME アルゴリズムに基づき直接更新する手法を提案し，ME アルゴリズムが MM アルゴリズムより高速な推定を可能にすることの定性的な説明を与える．

◆ 拡散性を持つ人混み雑音に２つの音声信号が混在する状況を模擬した実験により，提案手法が精度の面で従来手法に優り，かつ ME アルゴリズムが MM アルゴリズムより高速な推定を可能にすることを確認する．

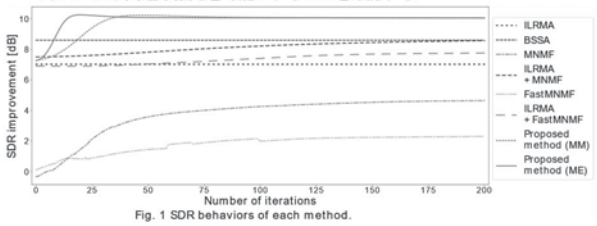


Fig. 1 SDR behaviors of each method.

3-1-14

3-1-14　所望音源の方向アトラクターに基づく時変の空間フィルタを用いた DNN 音声抽出

Deep speech extraction with time-varying spatial filtering guided by desired direction attractor

☆中込優(早大)，戸上真人(LINE)，小川哲司(早大)，小林哲則(早大)

◆ 提案方式：

➢ 所望音源の方向情報を入力の一部とした DNN と微分可能な形で表現した時変空間フィルタを，最終的な音質が最良となるように結合学習 (Fig. 1).

◆ 実験結果：

➢ 所望音源の方向情報をアトラクターとして DNN に組み込むことで，分離に用いるパラメータを効率的に推定でき，高い抽出精度を実現 (Table 1).

➢ 全音源の方向を必要とせず，高精度な音源の抽出が可能.

Table 1: Speech extraction results.

	MWF	Loss func.	SDR [dB]	SIR [dB]	PESQ
PIT w/ Oracle Permutation	time-invariant	MSA	5.22	9.31	1.43
Neural Spatial Filter	time-invariant	MSA	5.89	9.66	1.52
Prop.	time-varying	Joint training	**7.74**	**13.92**	**1.90**

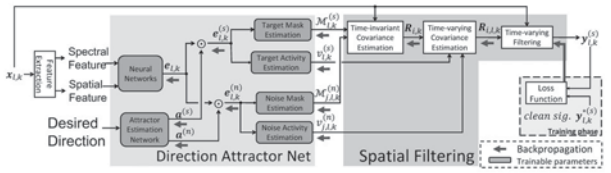


Fig.1: Overview of proposed architecture. Thick arrow represents backpropagation and DNN is jointly trained to maximize speech quality after time-varying spatial filtering.

3-1-15

3-1-15: 分離ベクトル同時更新による独立低ランク行列分析の収束性と性能向上の検討

Performance improvements for independent low-rank matrix analysis by simultaneous updates of demixing vectors

☆中嶋 大志, シャイブラー ロビン, 若林 佑幸, 小野 順貴 (首都大)

- 独立低ランク行列分析 (ILRMA) は音源間の統計的独立性と各音源の低ランク性を仮定したブラインド音源分離手法である．
- 従来の ILRMA では iterative projection (IP-1) を用いて分離行列の行ベクトル（分離ベクトル）を１個ずつ更新していた．
- 本稿では，３音源以上の ILRMA の収束性および分離性能の向上を目的として，２個以上の分離ベクトルの同時更新 (IP-2) を導入した ILRMA (ILRMA_IP-2) を提案する．
- さらに，音源モデルを固定したまま IP-1/2 を繰り返すことで，３個の分離ベクトルを同時に更新する ILRMA (ILRMA_IP-3 (1), ILRMA_IP-3 (2)) を提案する．
- 実験の結果，提案手法が従来手法よりも少ない反復回数でより高い分離性能を示した．

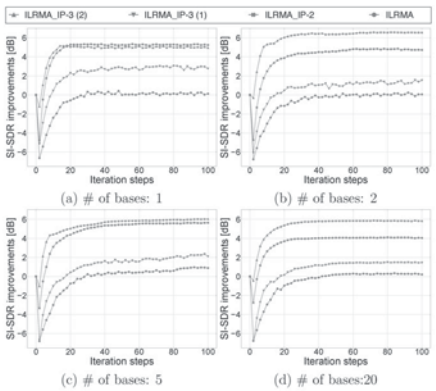


Fig. 1: Average SI-SDR improvements of each method.

3-1-16

3-1-16　調波打撃音分離の時間周波数マスクを用いた線形ブラインド音源分離

Linear blind source separation using time-frequency mask obtained by harmonic/percussive source separation

☆大藪宗一郎，北村大地（香川高専），矢田部浩平（早稲田大）

時間周波数マスキングに基づく優決定BSS（TFMBSS）は，時間周波数マスクに基づいて線形の（歪みの少ない）多チャネル音源分離が可能である．この利点を活かして，本稿では時間周波数マスクの一例として調波打撃音分離（HPSS）を用いたTFMBSSを提案する．

提案手法による線形分離化の恩恵によって，音質が向上したことを実験的に示す．

そして，各反復間のマスクが大きく変動するため信号対歪み比（SDR）の推移が安定しないという問題を適当なパラメータ設定によってスムージングすることで解決出来ることも実験的に示す．

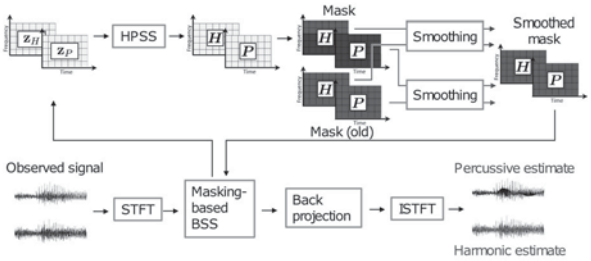


Fig. 1: Block diagram of proposed method, where $\mathbf{z}_H$ and $\mathbf{z}_P$ are parameters that corresponds to harmonic and percussive components, respectively.

3-1-17

3-1-17　局所時間周波数構造に基づく深層パーミュテーション解決法

Deep permutation solver based on local time-frequency structure

☆山地修平，北村大地（香川高専）

◆　ブラインド信号源分離の手法である周波数領域 ICA (frequency-domain ICA: FDICA)には，周波数毎に推定分離信号の順番が不定であるパーミュテーション問題が伴い，これについての高精度な解決はいまだできていない．

◆　混合信号の分離時に正解のパーミュテーションを与えた FDICA が，ブラインドな IVA や ILRMA よりも非常に高い分離精度を達成することは実験的に分かっている．

◆　本稿では，優決定条件下での複数音声の混合を対象とし，FDICA におけるパーミュテーション問題を DNN によって解決する手法を新たに提案する．

◆　実験の結果，周波数ビン毎に完全に分離されている理想信号に対して，高い精度でパーミュテーション問題を解決できることを示す．

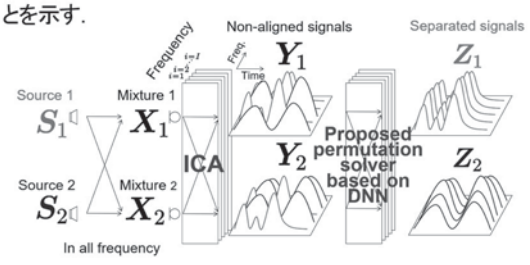


Fig.1　Permutation problem in FDICA.

3-1-18

3-1-18　独立深層学習行列分析におけるマイクロホン毎及び音源毎の座標降下法に基づく分離行列更新法の周波数別自動選択法

Automatic selection of frequency-wise demixing matrix update rule based on microphone-wise or sourcewise coordinate descent for independent deeply learned matrix analysis

◎牧島直輝，高宗典玄（東大），北村大地（香川高専），猿渡洋（東大），高橋祐，近藤多伸（ヤマハ）

◆独立深層学習行列分析（IDLMA）はブラインドな空間モデル推定と DNN による音源モデル推論を組み合わせた音源分離手法である。

◆IDLMA では，マイクロホン毎（列ベクトル毎）及び音源毎（行ベクトル毎）の座標降下法に基づく分離行列の更新法が提案されている。

◆本稿では，観測信号の尤度関数に基づいて，それらの中から適切な更新法を音源毎及び周波数毎に自動選択する手法を提案する。

◆音楽信号を用いた音源分離実験により，音源毎及び周波数毎に適切な更新法は異なっていることを確認し，提案する更新法の自動選択により分離性能が改善することを示す。

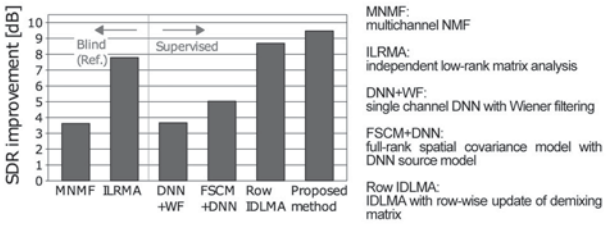


Fig.1　Average SDR improvements for 25 bass/vocal/drums songs.

3-1-19

3-1-19　リフティングスキームによる離散ウェーブレット変換を導入した深層ニューラルネットワークに基づく時間領域音源分離

Time Domain Audio Source Separation Based on Deep Neural Networks with Discrete Wavelet Transforms Using Lifting Scheme

☆小塚 詩穂里，中村 友彦，猿渡 洋（東大）

◆ 深層ニューラルネットを用いた時間領域音源分離手法である Wave-U-Net は，ダウンサンプリング層においてエリアシングや特徴量の欠落が起こりうる．

◆ これに対し，我々は以前，Haar 基底による離散ウェーブレット変換を用いたダウンサンプリング層（Haar 変換層）を設計し，これを導入した音源分離手法（多重解像度深層分析）を提案した．

◆ 本稿では，Haar 変換層を Haar 基底以外の基底にも拡張し，多重解像度深層分析の音源分離性能に対する基底の影響を調査する．

◆ 多重解像度深層分析が，基底に関わらず Wave-U-Net よりも高い分離性能を示すことを実験により確認した．

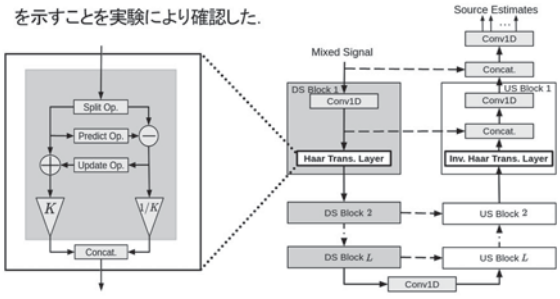


Fig. 1 : DNN architecture of multiresolution deep layered analysis. "Haar Trans. Layer" and "Inv. Haar Trans. Layer" are Haar and inverse Haar transform layers, respectively. Gray region inside Haar Trans. Layer corresponds to lifting scheme.

3-2-1

3-2-1　海外におけるヒアリングループの設置状況

Overseas state of the art in hearing loops

○中村健太郎（東工大），及川 靖広（早稲田大），小森 智康（NHK），上田 麻理（神奈川工大）

◆集団補聴支援設備の代表例であるヒアリングループの海外における設置状況について、現地で得た画像情報を中心に紹介する。

◆ヒアリングループは、かつては磁気誘導ループとも呼ばれ、音声信号を磁界として空間に放射する装置である。補聴器側では、マイクロホンの代わりにピックアップコイルで磁界を検出する（T モード）。

◆技術規格はピクトサインと共に IEC60118-4 により規定されている。

◆英国での設置率が高く、公共施設、交通機関、各種窓口などに設置されている。また、オーストラリアやニュージーランドでも設置率が高い。英国ではタクシー車内にも設置されている。

◆他の西欧諸国では国による差が大きいが、空港には設置されている割合が高い。米国では都市により、普及率の差が大きい。

◆アジア地域での設置率は一般に低い。

◆日本では、ある程度設置されているが利用率が上がっていない。別のピクトサインが用いられているなど、課題が多い。

3-2-2

3-2-2　ヒアリングループ及び聞こえ支援の最近の動向

Current situation of hearing loops and assistive listening devices

○舘野 誠(リオン株式会社)

ヒアリングループをはじめとする補聴援助システム（聞こえ支援）の最近の動向を踏まえ，それらの意義や適切な使い分け，その技術側面について述べる。

補聴器使用者にとって，音は聞こえるようになったが騒がしい場所などでことばがわからないという訴えが多く聞かれる。これは，補聴器の課題を三つに分けて整理すると解りやすい。難聴者は，補聴器を使用しないと十分な audibility が得られないために聞き取りが難しい。これが第一の課題であるが，補聴器によって聞き取りが回復し，会話における不自由が解消もしくは低減する。しかしながら，補聴器を使用していても騒音や残響のある環境では正常者よりも顕著に聞き取りが悪くなることが知られている。これが第二の課題である。この課題は，極論すれば聴覚の知見を用いなくても工学的に対処が可能であるが，補聴器だけでそれを十分に行うことは今ところ難しい。ヒアリングループ等は目下この第二の課題を解決する唯一の手段である。

ヒアリングループの最大の利点は，それが設置された場所で補聴器使用者が事前準備なしに利用可能なことである。ヒアリングループは誘導コイル（T コイル）を備えた（使用可能な状態にした）補聴器だけで受信できるため，ホールや会議室だけでなく空港の窓口カウンターや，エレベータ内の緊急用インターホン，バスやタクシーの車内等にも設けられている。このようにヒアリングループは音のバリアフリーに最も資する補聴援助システムということができる。

3-2-3

3-2-3　地域連携に基づく災害時の音情報に関する検討 -厚木市の避難所における音声聴取評価の試み-

Basic study on auditory information for emergency management and disaster based on regional cooperation.

△佐藤 登，△高橋 勝美(厚木市危機管理課)，△刈間 聡太，○上田 麻理，△田中 哲雄，
△小川 喜道，松本 一教(神奈川工科大)，中村 健太郎(東工大)

✓ 厚木市及び，神奈川工科大学では多発している自然災害においていち早く避難行動を行うために，2018 年度から地域連携に基づく災害時の音情報に関する検討を行っている．例えば，防災行政無線の聴こえに関する評価及び，周囲の騒音の影響，ノイズマップの作成である．さらに，避難所における音声をはじめとする情報提供についても 2019 年度から検討が進められている．

✓ 本稿では，厚木市の指定避難所における(特に高齢者の)音声の聴きとりの現状評価目的として，指定避難所における音声聴取評価実験を実施した（Fig.1:(左)）．さらに，実験後に市，自治体，研究者，その他のステイクホルダーによる円卓会議(Fig.1:(右))を開催し，避難所の音環境のあり方について議論したので，その進捗を報告する．

Fig.1 Example of an experiment and roundtable meeting with Atsugi city.

厚木市及び，神奈川工科大学では，「地域の住民の方が"自分たちの課題を解決する"上で研究者と共に取り組む」研究プロセスを作っており，地域連携型研究としてこのような試みを今後も積極的に進めていく予定である．

3-2-4

3-2-4　騒音の主観的なうるささに対する個人内変動の影響：経験サンプリング法による評価

Subjective noise annoyance on some circumference and intraindividual variation: Evaluation by experience sampling method

◎三浦 貴大（産総研），刈間 聡太（神奈川工大），
藪 謙一郎（東大 IOG），上田 麻理（神奈川工大）

◆【目的】アノイアンスは非音響学的な要因でも変化するため，日常生活下での計測がより重要である．このため，日常生活下で騒音の客観量と主観量を同時計測するアプリケーションを開発の上で，騒音と騒音量に対する主観量の関係の一端を明らかにすることを目的とした．

◆【システム】以下の図のようなデータ記録用スマートフォン用アプリを開発した．本アプリは，経験サンプリング法に則って，日に数回ほどアラームでスマートフォン所持者に入力を促すものである．

◆【方法】このシステムを音響技術者 8 名に 1~2 週間ほど使ってもらい，個人宅・オフィス・屋外などでのデータを記録してもらった．

◆【結果】1/3 オクターブバンド周波数帯域における L_A に対する相関が，感情状態で 2.5 kHz 以上，ストレスで 5 kHz 以上，疲労状況とは 12.5 kHz 以上で確認された．また，クラスタ分析の結果より，日常生活下での騒音の感じ方は 5 クラスタに分類できた．

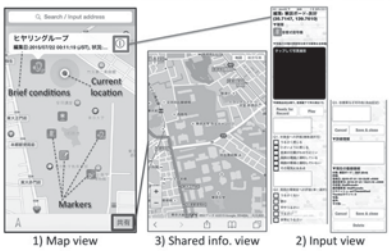


Fig. 1 User interfaces of our applications to input and share real-world environmental noise conditions.

3-2-5

3-2-5　障碍音声健常化のための母音音韻性制御手法に関する検討

A study on vowel phonological control method for normalization of dysarthric speech.

○戸次幸徳，坂田聡，上田裕市（熊本大院）

◆実ホルマント空間において話者性を正規化し、音韻性を分離する正規化構音空間(hv 平面)で障碍音声の母音音韻性の健常化を行う手法を提案する。

◆障碍話者はその障碍の種類・程度により、話者毎に異なる母音 hv 分布を有しており、規格化空間を介することにより話者毎に異なる様々な母音 hv 分布を規格化することができる。規格化空間から健常者である目標話者 hv 平面へ写像することにより音韻性の健常化を行う。

◆提案手法について原話者および目標話者を設定し、実際に健常化処理を行う。実ホルマント空間から hv 変換により得られる hv 軌跡を入力として、原話者 hv 平面、規格化空間、目標話者 hv 平面の対応関係や hv 点の遷移について確認する。

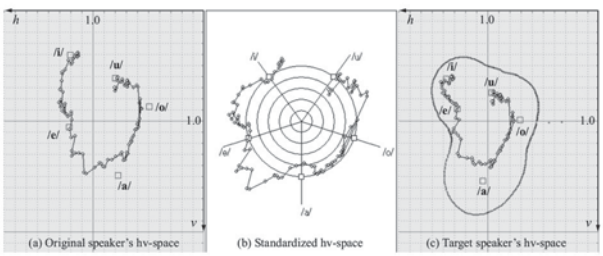


Fig.1 : Application example of vowel phonological control method using a normalized articulatory space.

3-2-6

3-2-6　聴覚障がい者の支援を目的とした拡張現実による環境音の可視化手法に関する基礎検討

Basic study on visualization of environmental sound using augmented reality to support hearing impaired persons

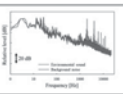
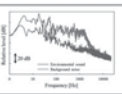
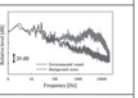
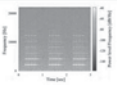
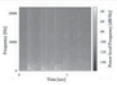
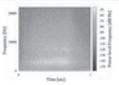
☆中屋雅文，朝倉巧（東京理科大）

◆聴覚障がい者を環境音検知の観点から支援することを目的とし、日常生活上で検出される環境音を、機械学習によって識別させる。

◆機械学習でにおける教師用音データを２種類の住居で収録した。日常生活で発生しうる環境音 15 種類を意図的に生じさせ、発生源と同室内での測定、および遠くの環境音でも判別させることが可能か検討するために別室から伝搬する環境音の測定を併せて行った。

◆本検討では機械学習で識別を行う際に環境音をいくつかのカテゴリに分類する必要があると考え、周波数特性とスペクトログラムの観点から環境音を３種類に分類し、これらの正答率を考察した。

Table 1 Sound type and graph of FFT and spectrogram

音の種類	純音性	過渡音	定常音
周波数特性			
スペクトログラム			
環境音 (★は表示した音)	アラーム音★ インターホン音など	ドアのノック音★ ドアの開閉音など	カーテンの開音★ 流水音など

3-3-1

3-3-1　頭部伝達関数の帯域分割ノッチ・ピークモデル － 分割帯域幅と音像定位精度の関係 －

Band-divided notch-peak model for head-related transfer function: Relation between band width and accuracy of sound image localization in the median plane

☆西山織絵，相崎翼，飯田一博（千葉工大・先進工）

耳介形状から個人の HRTF を生成することを目的とした HRTF の帯域分割ノッチ・ピークモデルを提案し，分割帯域幅と音像定位精度の関係を検討した．その結果，以下の知見を得た．

1) 上昇角知覚の手掛かりとなるN1, N2, P1, P2 を再現するためには，分割帯域幅は 1/12 oct.もしくは 1/6 oct.以下とする必要がある．

2) 実測 HRTF と有意差のない定位精度を与える帯域幅の上限は被験者により異なり，1/12 oct., 1/6 oct., 1/3 oct., 臨界帯域幅であった．

以上より，分割帯域幅は 1/12 oct.または 1/6 oct.が妥当と考えられる．

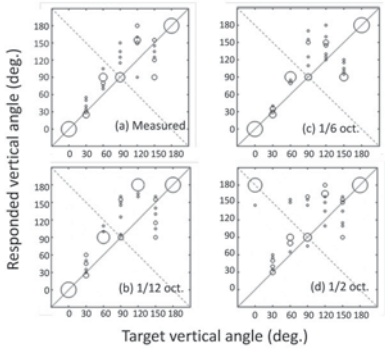


Fig. 1 Responses to (a) measured HRTFs and (b) – (d) modeled HRTFs.

3-3-2

3-3-2　信号オンセットの変更による音像定位に及ぼす加齢の影響

Influence of Aging on Auditory Lateralization Ability by Changing Signal Onset

☆白木萌子(中央大院)，△森田和元，黒瀬和希，戸井武司(中央大)

高齢者の音像定位能力の把握を目的とした一連の実験より，高齢者は若年者と比較して音源の冒頭(オンセット)部分の認知が低下することが推測された。本研究では，左右時間差をつけたもの(ITD 音源)，左右時間差をつけオンセット部分に音があるもの(Lead 音源)，左右時間差をつけず片側のオンセット部分のみに無音があるもの(Space 音源)の３種類を用いて，被験者に左右差のない音源と比較して音の方向を答えてもらうことで左右の弁別能力を測る実験を行った。

1 kHz の純音を用いた際の若年者の「同じ」の回答率を Fig. 1 に，高齢者の「同じ」の回答率を Fig. 2 に示す。Space 音源について，ITD 時間が増えると若年者の回答率は減少しているのに対して，高齢者は ITD 時間の増加に関わらず常に一定となっている。このことから，高齢者は若年者ほど音源の冒頭に挿入した無音時間を聞き取ることができないことを確認した。

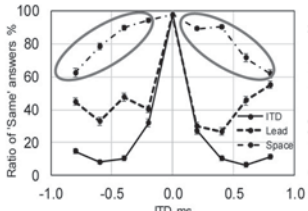


Fig. 1　Ratio of 'Same' answers of younger participants (1 kHz)

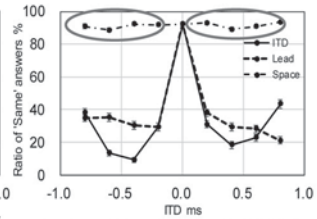


Fig. 2 Ratio of 'Same' answers of elderly participants (1 kHz)

3-3-3

3-3-3　単一スピーカから複数同時に発せられた音声の了解度に及ぼすキャリアフレーズの影響 －ターゲットとする話者の特徴情報の直前聴取－

Effects of carrier phrase on intelligibility of multiple speech signals simultaneously radiated from a single loudspeaker: Pre-listening of target talkers' individualities

○佐藤逸人，森本政之（神戸大），飯田一博（千葉工大），佐藤洋（産総研）

◆多言語一斉放送の音声了解度の向上を目的として，放送の直前にターゲットとなる話者の音声をキャリアフレーズとして再生して聴取者に学習させる方法の有効性を単語了解度試験により検討した。

◆統計的に有意ではなかったが，キャリアフレーズに含まれる話者の音声の特徴情報の量や提示順序がターゲットの単語了解度に影響することが示唆された。

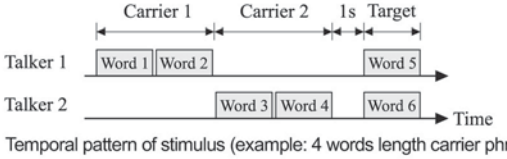


Fig.1: Temporal pattern of stimulus (example: 4 words length carrier phrase).

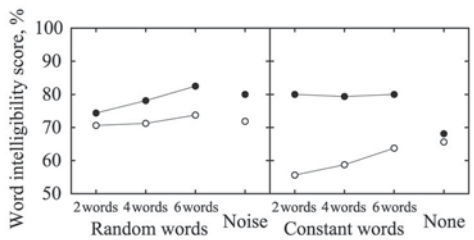


Fig.2: Intelligibility scores of words spoken by two different female talkers for each carrier phrase/sound condition (○: Talker 1, ●: Talker 2).

3-3-4

3-3-4　2 個のスピーカから発せられた音声の了解度に及ぼす提示方向の影響

-日本語単語と韓国語単語もしくは 2 つの日本語単語の同時提示-

☆高橋仰一（千葉工大・院），菊地勇成，林翔太，飯田一博（千葉工大・先進工）
佐藤逸人（神戸大），佐藤洋（産総研），森本政之（神戸大）

◆ 多言語同時提示において音声情報伝達を可能とするスピーカ配置を検討する第 1 ステップとして，以下の 2 つの目的で日本語と韓国語および日本語同士の 2 つの単語の同時提示実験を行った．

1） スピーカの開き角が音声了解度に及ぼす影響を明らかにする．

2） 単語の親密度の影響を明らかにし，外国語の代わりに低親密度日本語単語を用いて高親密度日本語単語の了解度を評価できるか否かを検証する．

◆ その結果，以下のことが明らかになった．

1） 韓国語単語と同時提示した高親密度日本語単語の了解度はスピーカの開き角に依存せず，88.0%(0°)から 91.5%(24°)であった．

2） 高親密度日本語単語の了解度は，同時に提示される単語が韓国語（H-K）の場合が最も高く，次いで高親密度日本語（H-H）であり，低親密度日本語（H-L）の場合が最も低くなった．H-K と H-L 間，H-K と H-H 間では統計的有意な差が認められた．したがって，高親密度日本語単語の音声了解度評価において外国語の代わりに低親密度日本語単語を用いることはできない．

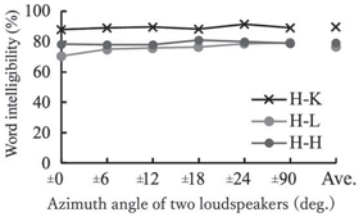


Fig. 1 Relation between azimuth angle and word intelligibility

3-3-5

3-3-5　音声の奇数倍音と偶数倍音を用いた音脈モワレ表現における群化特性

Grouping Characteristics in Moire Representation Using Odd and Even Harmonics of Speech

○河野有美，△檜野幸志郎，小坂直敏（東京電機大）

◆マルチチャネル上での音響空間エフェクトでの基本的表現として音脈モワレ表現についての受聴位置の群化特性を検討した。

◆正弦波モデルにより原音を分解し，ステレオスピーカから別々に再生された音声を複数位置で受聴した際の群化度を評価した。

◆左に偶数倍音，右に奇数倍音を再生した際，主観実験により 5 段階の群化度評価をした。(Fig. 1)

◆一方で，Fig. 1 と同様な再生環境下で被験者が移動して分離，群化の境界点を調べた。(Fig. 2)

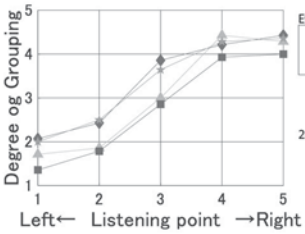


Fig.1 : Degree of grouping for decomposed speech in stereophonic listening. (Voices ◆: Female 1, ▲: Female 2, ★: Male 1, ■: Male 2)

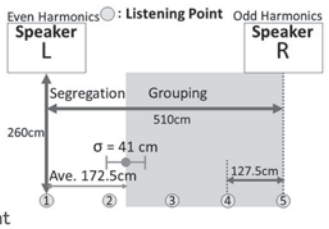


Fig.2 : Border line of segregation and grouping

3-4-1

3-4-1　音声対話システムのためのターンテイキングのタイミング評価

Evaluation of turn-taking timing for spoken dialog systems.

○藤江真也（千葉工大），小林哲則（早大）

◆音声対話システムの発話開始タイミングを決定する問題に対して，どのように評価するべきかを検討する．

◆システムがターンをとるべきか否かが識別できても，実際にどの程度の間をおいて発話を開始すべきかが明確ではなかった．

◆人同士の対話音声における交替潜時を人工的に操作したサンプルを多数用意し，それを聴取した被験者が後続発話開始タイミングの早さ／遅さを評価する実験を行った．

◆実験の結果（Fig. 1），従来の研究と同様に 900ms ほどの遅れまでが許容されることが分かった．

◆更に，詳細な分析を行ったことで，広い範囲で交替潜時の適切さを評価できる可能性を確認した．

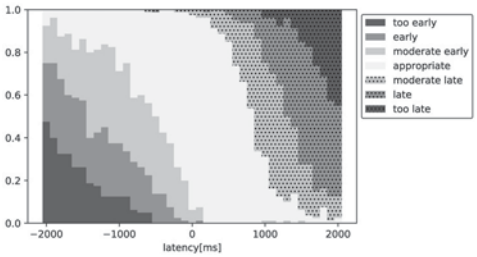


Fig. 1：聴取実験の結果

3-4-2

3-4-2　傾聴対話のための音声対話ロボットの開発と評価

Development and evaluation of conversation robot for active listening.

☆伊島翔大（千葉工大），△関根みくり，藤江真也

◆高齢者施設などで行われている傾聴対話に着目し，専門のボランティアの代わりを担うロボットの構築を目的とする．

◆傾聴対話において基本的な機能を実装したロボットの開発を行った（Fig. 1）．

◆ 複数の音響特徴量を用いて時間構造を考慮しながら発話継続 / 終了の識別を行う．

◆ 質問文とユーザの応答文を入力し，ユーザの発話内容に応じたリアクション生成を行う．

◆ ユーザ応答の終了時，意外性に応じた「共感」や「驚き」などのリアクションの生成を行う．

◆ 傾聴対話ロボットを用いた印象評価実験を行った結果，発話継続／終了判定では「快適さ」付属質問生成では「元気さ」「楽しさ」リアクション生成では「元気さ」をユーザに与えるという効果があることがわかった．

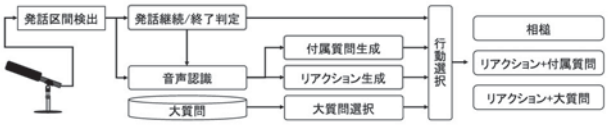


Fig. 1：System architecture

3-4-3

3-4-3　自律型アンドロイド ERICA による就職面接対話

A job interview dialogue by the autonomous android ERICA

◎井上昂治，原康平，Divesh Lala，山本賢太，中村静，
高梨克也，河原達也（京都大・情報学）

■ アンドロイド ERICA の面接対話システム

- 就職活動や入試の対面による面接の練習を想定
- 既存の面接対話システムは典型的な質問を尋ねるのみ
- 人間どうしのような緊張感があるリアルな面接にしたい
- 掘り下げ質問生成を提案
  - ① 充足判定に基づく掘り下げ質問
    例：その会社でなければならない理由が不足
    → 「他にも同じようなことができる会社はあると思うのですがなぜ当社を選ばれたのでしょうか」
  - ② キーワード抽出に基づく掘り下げ質問
    例：志願者「私は音声対話システムの研究をしています」
    →「では，音声対話システムについて詳しく教えてください」

■対話実験

- 大学生 22 名の被験者
- 掘り下げ質問の効果を確認
  - ✓「緊張感」や「リアルさ」といった面接自体に対する印象や「質問の質」の評定値が有意に向上
  - ✓ さらに，面接官の存在感も有意に向上

3-4-4

3-4-4　漸進的な音声認識・機械翻訳・テキスト音声合成に基づく音声から音声への同時翻訳

○中村 哲，Sashi Novitasari，△帖佐 克己，柳田 智也
△二又 航介，須藤 克仁，Sakriani Sakti（奈良先端大）

◆漸進的な音声言語処理に基づくインクリメンタルに動作する同時音声翻訳システム

◆インクリメンタル音声認識：文全体を入力して注視するモデルを教師(teacher)とし漸進的処理のために短いセグメント単位で注視を行うモデルを生徒(student)として学習，インクリメンタルに音声認識

◆インクリメンタル機械翻訳：音声認識の出力を受けながら，訳語選択に必要な入力が得られていない場合に適応的に入力を待つ機械翻訳

◆インクリメンタル音声合成：単語やアクセント句を単位として入力テキストをセグメントに分割し，セグメントごとに音響パラメータ（スペクトログラム）やセグメント終端を予測

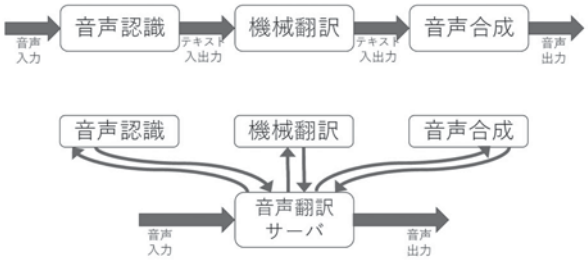


Fig.1：Incremental Simultaneous Speech-to-speech Translation System

3-4-5

講演取消

3-5-1

3-5-1　ICTを活用した音育スキルの可能性 −iPad を用いた幼稚園での実践活動−

Possibility of sound education skills using ICT : Practice activities in kindergarten using iPad

○小松正史（京都精華大学）△ニフュユス美冬（株式会社スマートエデュケーション）

◆音育のアナログ技法に**ICTによる「デジタル技法」を加えることで、新たな音育実践が生まれる**可能性がみられる事例を紹介する。

◆幼児でも iPad 上のアプリで、手軽に使えるインターフェース「おとねんど」を開発した。

◆目に見えない音を、ねんどに見立てた編集画面を指で操作することにより、イメージする響きに仕立てることができる。

◆音育活動中の試行錯誤が多く確認され、音に対する注意喚起力が高まり、**背景音への意識が持続している**様子もうかがえた。

◆アナログスキルの音育の良さを生かしながら、**デジタルスキルによって編集・再生を行う作業環境が補える**ことが確認された。

Fig. Sound editing using Otonendo

3-5-2

3-5-2 この音何デシベル?-大学生の音の大きさに対する理解度をまずは測る試みその 3-

What is this sound in dB?
We tried to determine student's comprehension of sound levels Part3.

○上田麻理，西口磯春（神奈川工科大学）

＜あらまし＞

本稿では，大学生の音の大きさとその単位のデシベル(dB)に対する理解度を測る試みとして，自作した音の大きさチャート・キットを用いて理解の具合を測ってみた．今回は高専生 74 名を対象として実施した．

(a) Sheet of sound level chart

Fig. Detail of the sound level chart kit. (a) sheet and (b) stickers of sound types (18 types).

3-5-3

3-5-3 騒音実務者のための簡易視聴デモンストレーション

Simple audiovisual demonstration for engineers engaged in noise control

○岡田恭明（名城大・理工）

実社会における騒音の問題を扱う場で，音によるデモンストレーションを行うと，当該分野ならではの視点から「この程度の大きさの音が聞こえてくるのか」などの捉え方をされる場合がある。また，実験室などの環境でない限り，不特定多数の聴講者に向けて，実際の音量に調整することも難しい。そのため，このような機会の際には，対比する音源，すなわち聴き比べできるデモになるように心掛けている。

今回，音響の知識を有する実務者に向けて，防災無線放送の屋外伝搬実験を行う際の諸課題，また集合住宅用の排水設備から発生する騒音の対策効果（Fig.1）について行った簡易な視聴デモの例を紹介する。

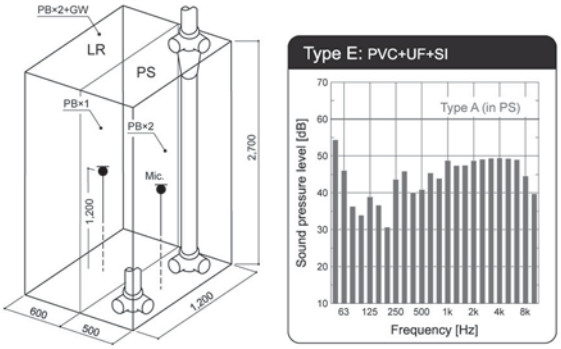


Fig.1 : Example of simple audiovisual demonstrations for reduction effect against drainage noise in residential buildings.

3-5-4

3-5-4 家庭用品で作る簡易声道教材

Educational tool in acoustics of speech production made of household commodities

★兒玉明日夏、浜田史楓、笠原美左和、真志取秀人、高橋義典
（都立産技高専）

身近にある家庭用品を用いた工作によって,参加人数の多い科学イベントでも利用可能な簡易声道教材を提案する(Fig. 1).簡易声道教材の構造は柔らかいボトル(100 円ショップなどで入手可能なフレンチボトル)で作られた人工声道の部分と,ストローに切り取った OHP シートをリードとして貼り付けた人工声帯の部分で構成されている.

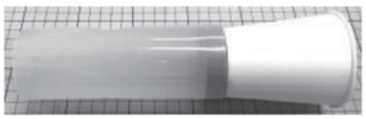


Fig. 1: Artificial vocal cords created

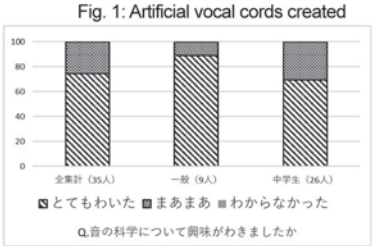


Fig. 2 : Questionnaire results

中学生26人および大人9人を対象にイベントを実施したところ,概ね15~20分程度で組み立てられ,全員が音を出すことができた.イベント後に実施したアンケートの結果から,参加者全員が音の科学に興味を持たせることが確認できた(Fig. 2).

3-5-5

3-5-5 横隔膜の陰圧による吸気を模擬する模型の製作

Making a model that reproduces the intake caused by negative pressure of diaphragm

☆増田稜平，平山亮（大阪工大）

◆音声教育及び発話ロボット開発に向けて、当研究室で行われてきたこれまでの音声器官の模型作成の研究を進展させ、声帯模型に空気を送り込む肺模型の試作を行った。陰圧肺呼吸模型とは、容器内部と外部の圧力の違いを利用した模型であり、内部の圧力が外部の圧力より小さくなるよう設定し、その状態で外部の空気を取り込むと容器内にある風船が大きく膨らむ仕組みとなっている(Fig. 1)。陰圧肺呼吸模型では、実際の横隔膜に似た動きが再現でき、その作用で肺にあたるゴム風船が膨張することに成功した。今後、これまでに製作された模型を組み合わせて発声、発話模型の製作へと進めていく。

Fig.1 Negative pressure lung respiration model

3-6-1

3-6-1　びまん性疾患の肝臓における組織構造と音速および音響インピーダンスの関係性の評価

Evaluation of relationship between structural characterization, speed of sound and acoustic impedance of diffuse liver diseases

☆橋本諒哉（千葉大・院融合），伊藤一陽（Singapore Eye Research Institute），吉田憲司，山口匡（千葉大・CFME）

◆肝細胞癌患者から摘出された肝臓の生試料および薄切試料に対し，超高周波超音波を用いて観察し，微視的な音響インピーダンスおよび音速を評価した．

◆観察後の薄切試料から作成した病理像から得られる組織構造に基づいて音響特性を評価する手法を提案し，肝実質部と線維部それぞれの音速を評価した　(Fig. 1(a)-(c))．

◆正常肝に比して肝線維症の肝臓では音響インピーダンスが高く，線維部の音速は実質部より高い傾向を示した．また，肝線維症では肝実質部の音速が低いことが確認された（Fig. 1(d), (e))．

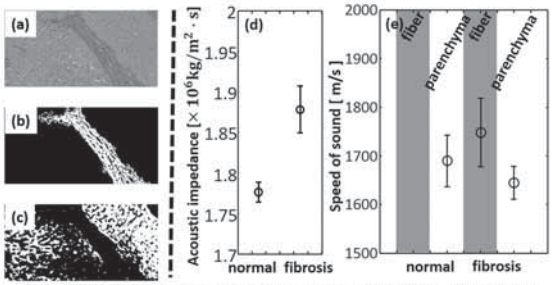


Fig.1 : Left, Original histological image (a), extracted fiber structure image (b) and extracted parenchyma image (c). Right, average value and standard deviation of acoustic impedance (d) and speed of sound (e).

3-6-2

3-6-2　超音波によるヒト胸椎描出を目指した面構造の反射特性と点構造の散乱特性の差異に関する検討

A study on the difference between reflection characteristics of the surface structure and scattering characteristics of the point structure for depiction of surface structure for human thoracic vertebral by ultrasound

☆橋本拓実，森 翔平，荒川元孝，△大西詠子，△山内正憲，金井 浩（東北大学）

◆硬膜外麻酔には，「穿刺位置の正確な特定が難しい」という課題がある．麻酔の補助として用いられる超音波診断装置によっても，現状の描出法ではその位置を容易には特定できない．この問題を解決するため，超音波プローブの各素子で得られる素子データを用いた鮮明な胸椎描出法を目指している．

◆基礎検討として，超音波診断装置を用いて，筋組織のような散乱体(ワイヤ)と，骨のような面反射体(アクリル板)からの受信信号を計測し，散乱特性と反射特性の差異について検討した．散乱特性の計測値と理論散乱特性を比較した．

◆散乱特性と反射特性には，遅延時間分布と振幅に違いがあることが確認できた．包絡線振幅の最大値から求めた散乱特性と反射特性を Fig. 1 に示す．散乱特性の計測値が理論値に完全には一致しなかったため，理論式についてさらに検討が必要である．今後は，散乱特性と反射特性の違いを利用した胸椎描出法を検討していく．

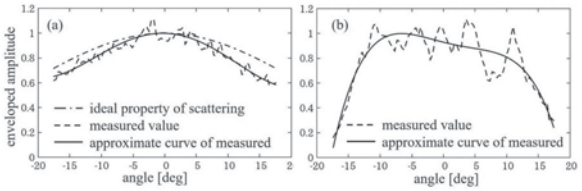


Fig. 1 : Measurement results.
((a) scattering property, (b) reflection property)

3-6-3

3-6-3　ファントム実験系を用いた動脈壁弾性率の超音波計測に関する基礎的検討

Basic study in ultrasonic elasticity measurement of arterial wall using a phantom experimental system

☆秋山星来，森　翔平，荒川　元孝，金井　浩（東北大）

動脈硬化症の早期発見のため、我々は動脈壁の弾性率の超音波計測法を開発してきた。しかし、生体内は真値が不明であるため、超音波計測の精度検証に課題がある。本報告では、ファントム実験において、レーザー計測による値の比較により、超音波計測法の精度を検証した。計測結果を Table 1 にまとめる。外半径変化より求めた弾性率は、レーザー計測と超音波計測で一致した。一方、超音波計測において、外半径変化より求めた弾性率と厚さ変化より求めた弾性率は一致しなかった。この原因として、位相差トラッキング法による厚さ変化の計測精度と弾性率導出のために異なる式を用いていることが考えられる。

Table1 : Elastic modulus calculated from wall thickness change and elastic modulus calculated from change in outer radius.

	by laser measurement	by ultrasonic measurement	
elastic modulus calculated from wall thickness change [kPa]	—	anterior wall 367	posterior wall 397
elastic modulus calculated from change in outer radius [kPa]	352	352	

3-6-4

3-6-4　分解能の影響を考慮した二成分マルチレイリーモデルによる肝線維化パラメータ推定の検討

Effect of ultrasonic beam width for evaluation of parameters of liver fibrosis based on multi-Rayleigh model with two components

☆ 張闖，平田慎之介，蜂屋弘之（東工大）

超音波画像を用いた肝線維化定量評価を目的とし，線維性肝組織の超音波画像の振幅確率分布モデルであるマルチレイリーモデルに基づく肝線維化パラメータ推定手法を提案している．先行研究では，分解能の影響を考慮せずに，モーメントの次数とモデルパラメータ推定精度の関係を検討した．しかし，臨床画像では分解能の影響を受けたエコー信号を利用するため，分解能を考慮した検討が必要である．本報告では，分解能の影響を考慮した画像を用いて，線維定量化に与える影響を検討した．散乱体モデル (Fig.1(a)) から得たB-mode画像を　Fig.1(b) に示す．また，B モード画像の振幅値を深度方向に平均した結果をFig. 1(c) 示す．　また，線維量と線維分散比の推定結果を Fig. 1(d) に示す．低振幅部と高振幅部の境界 (正常組織と線維組織の境界) においてエコー振幅値の変化がなだらかになり，線維推定値は分散比が減少しながら線維量が多く評価されることがわかる．

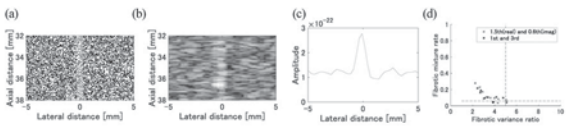


Fig. 1. (a) Model. (b) B-mode images. (c) Average of amplitude in axial direction. (d) Estimated results. Fiber widths are 0.6 mm.

3-6-5

3-6-5　皮膚組織の後方散乱係数解析における音響特性・組織構造の関連性評価

Evaluation of relation of acoustic and histopathological properties on backscatter coefficient analysis of skin tissue

○大村眞朗（千葉大・院融合），吉田憲司（千葉大・CFME），
秋田新介（千葉大・院医），山口匡（千葉大・CFME）

◆ 正常・リンパ浮腫（LE）真皮の後方散乱波に音響特性と組織構造がどのような関連を示すかについて，後方散乱係数を指標として評価した．

◆ 実計測の後方散乱波（15 MHz），音響インピーダンス分布（80 MHz），病理像（解像度 1 μm）から後方散乱係数（BSC，BSC_Z，BSC_{His}）をそれぞれ算出し，線維組織の散乱体サイズ・音響濃度を推定した．

◆ BSCはリフレクター法により，BSC_Z，BSC_{His}は音響インピーダンス比の二次元フーリエ変換に基づく理論式から算出した．

◆ 各症例の異なる空間情報から得た散乱体サイズ・音響濃度推定結果（Fig. 1）において，後方散乱波の音響濃度は音響インピーダンスおよび病理像の濃度推定結果と特に関連性が強い．

◆ 音響濃度の違いはタンパク由来の浮腫の特性に起因していると考えられ，正常に対して LE 組織は低密度と想定される．

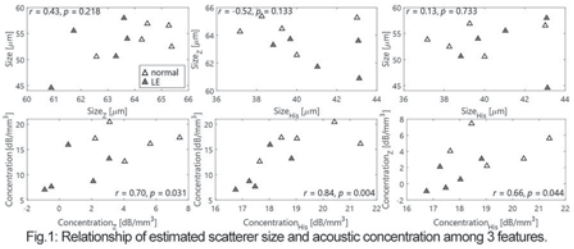


Fig.1: Relationship of estimated scatterer size and acoustic concentration among 3 features.

3-6-6

3-6-6　心腔内血流信号強調のための２次元点拡がり関数の推定に関する検討

Investigation on estimation of 2D point spread function for enhancement of intracardiac blood flow signals

○茂澄倫也，長岡　亮，長谷川英之（富山大）

◆ 心腔内血流からの超音波信号に含まれる不要エコー成分を抑圧するために，畳み込みフィルタを適用した．

◆ 畳み込み用フィルタとして，生体内からの受信信号から送受信システムの空間応答を推定し，これを元々の受信信号に畳み込んだ．

◆ Fig. 1(a)および 1(b)に推定した２次元点拡がり関数，およびその周波数スペクトルを掲載する．Fig. 1(a)の空間応答を元々の受信信号に畳み込んだ結果を Fig. 2 に掲載する．フィルタ後の画像（Fig. 2 画像右）では，フィルタ前（画像左）に比べ背景領域におけるノイズ的成分が抑圧されていることが確認できる．

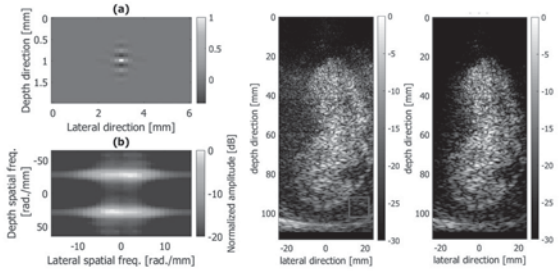


Fig.1 : (a) Estimated PSF (point spread function). (b) Frequency spectrum of spatial response in Fig. 1(a).

Fig. 2 : Intracardiac blood flow B-mode images obtained without (left) and with (right) convolution of received echoes and PSF.

3-6-7

3-6-7　心筋の微小速度計測のための超音波ビーム送信条件依存性

Dependence of ultrasonic transmitted beam condition for measurement of myocardial minute velocity

☆菅原佳奈，森 翔平，荒川元孝，金井 浩（東北大）

◆高速で微小な運動をする心筋の診断のためには，時間的にも空間的にも高分解能な超音波計測が求められる．しかし，これらの分解能は，超音波のビーム送信条件（送信ビーム幅，フレームレート）に影響され，トレードオフの関係にある．

◆心筋の微小速度計測に最適なビーム送信条件の決定を目指し，先行研究[1]では模擬実験にて超音波による微小速度計測の精度を定量的に評価した．しかし，計測速度の再現性が不十分であり，精度が高いと予想される計測条件においても，20% 程の計測誤差が算出されてしまうという課題が残った．本報告では，解析手法の改善により評価法の精度向上を試みた．

◆心筋の微小速度波形計測を想定して構築した水槽実験系にて，計測精度が高いと予想される送信条件にて実験を行った．超音波計測による速度波形を，レーザドプラ振動計で同時刻に計測された速度波形と比較し，超音波による計測精度を定量的に評価した．

◆受信ビームの中心周波数を用いて解析することで位相差トラッキング法の追跡精度が向上した．さらに，レーザにて計測した速度波形を内挿し，レーザ計測速度波形と超音波による速度波形の位相ずれを補正し，計測誤差を 5% 程に低減させることに成功した．

◆今後は，本実験解析系を用いて最適な送信条件の決定を試みる．

[1] N. Furusawa *et al.*, JJAP, 58 (SG), SGGA08-1-SGGA08-6, 2019.

3-6-8

3-6-8　超音波位相差を用いた微小速度推定による心筋収縮応答計測の高精度化に関する検討

Improvement in measurement of contraction response of heart wall by velocity estimation using ultrasound phase difference

☆小原優，森翔平，荒川元孝，金井浩（東北大）

◆【目的】厚み変化が生じる心臓壁において，微小速度を局所的に推定することでの，心筋収縮応答計測の高精度化に関する検討を行う．

◆【原理】従来の単一周波数での位相差を用いた微小速度推定法は，安定した収縮応答計測のために空間平均を行っていた．提案法では，多周波数における位相差を用いた高精度化により，空間平均を必要としない局所的な微小速度推定を行うことができる．

◆【結果・考察】20 代健常者の心室中隔壁における，従来法と提案法による微小速度推定の結果を Fig. 1 に示す．両者の波形は異なる結果となった．厚み変化が生じる心臓壁において，従来法では深さごとに異なる微小速度を空間平均するため推定精度が低下する．一方，提案法は空間平均を行わずに連続な微小速度波形が得られるため，微小速度を正しく推定でき，収縮応答計測の高精度化に有用と考えられる．

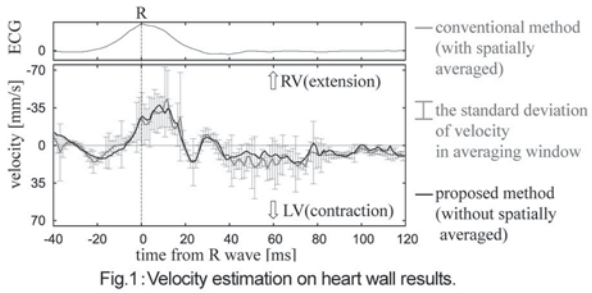


Fig.1 : Velocity estimation on heart wall results.

3-6-9

3-6-9　超音波を用いた赤血球集合度評価における駆血の安定性に関する検討

Evaluation of stability in measurement of red blood cell aggregation using ultrasound

☆深瀬晶予，△永澤幹太，森 翔平，荒川元孝（東北大），
△八代 諭，△石垣 泰（岩手医科大），金井 浩（東北大）

【背景・目的】赤血球同士が可逆的に接着する現象を赤血球集合といい，駆血など血流の低ずり速度状態で起こりやすく，炎症や高脂血症の指標として注目されている．超音波を用いた非侵襲的かつ定量的な赤血球集合度の測定法の確立を目指し，これまでに血糖値と赤血球集合度との間に正の相関があることを示した．計測の安定性を再検討するため，同一部位を対象として連続で取得した画像間の相関を調べた．

【方法】健常成人女性１名の手背静脈を対象として，中心周波数 40 MHz（波長 38 μm）の超音波を用い，3 s 間隔で 6 フレーム分の連続した計測を，非駆血時と駆血時（駆血 72 s~87 s）においてそれぞれ 5 回ずつ行った．得られた RF 信号から包絡を取って，連続したフレーム間の相関を調べた．

【結果】連続した計測の中で，超音波 B モード画像の輝度に大きな変動はなく，駆血時の計測で血流は確認されなかった．相関は，非駆血時，駆血時のいずれも計測回によって結果が異なった．非駆血時には血管内に血流があるため，相関は低くなると予想されたが，非駆血時と駆血時で目立った傾向の違いはみられなかった．

【考察】駆血時に相関が高くならなかった理由として，計測対象が生体や超音波プローブ由来のわずかなずれを考える．今後，さらに計測と駆血の安定性を検討する．

3-11-1

3-11-1　ヴァイオリンの弓の震えを考慮した擦弦振動解析手法

Vibration analysis of a bowed string involving a tremble of a violin bow

○鮫島俊哉（九大・芸工）

◆岸らにより提案された，等価回路的考察に基づく擦弦振動の解析手法に，有限要素法による弓(Fig. 1)の振動場の解析モデルを導入し，擦弦振動場と弓の振動場が適切に連成された解析手法(Fig. 2)を構築する。

◆その解析手法により，弓の振動特性が擦弦振動に与える影響，およびヴァイオリン学習者が克服する課題：「弓が踊る(震える)」現象の発生機構を解明する。

Fig. 1 : Violin bow.

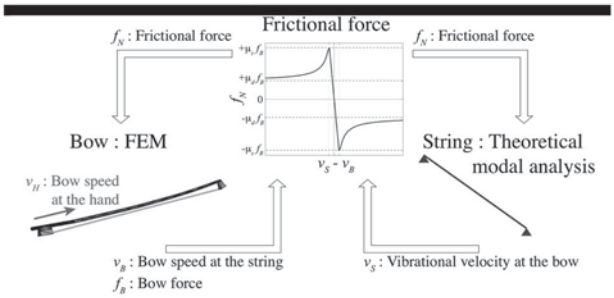


Fig. 2 : Coupled vibration of violin bow and bowed string.

3-11-2

3-11-2　擦弦楽器に用いる弓毛のヤング率の測定

Measurement of Young's modulus of bow hair used in bowed instruments

○松谷晃宏（東工大）

◆ヴァイオリンなどの擦弦楽器の演奏では，楽器と同様に弓もたいへん重要な構成要素であり，弓毛の弾力を活かすことが演奏にあたっては特に重要である。

◆筆者は，これまでに弓の顫動に関する実験結果，弓の傾斜が顫動抑制に有効であること，擦弦位置及び弓の傾斜と音色の関係について報告したが，これらの系統的実験では，弓毛の特性については詳しく確認していなかった。

◆今回は，ヴァイオリン演奏で使用される弓毛の引張試験を行い，演奏者の観点で特に興味があると思われる，弓毛の断面観察及び荷重ー伸び線図の測定，および湿度依存性の結果等について報告する

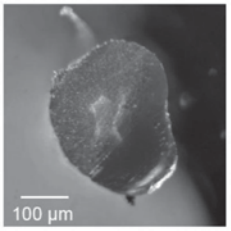


Fig.1 : Optical microscopic image of cross section of bow hair.

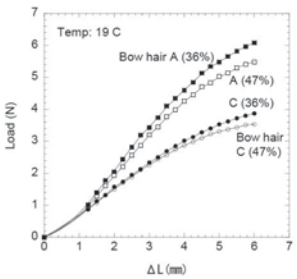


Fig.2 : Humidity dependence of relationship between load and elongation of Bow hair A and C

3-11-3

3-11-3　多孔質ポリマーフィルムエレクトレットを用いたエレクトリックアコースティックギター用センサー

Cellular polymer film electret sensors for electric acoustic guitars

○児玉秀和，安野功修（小林理研）宮田智矢、樋山邦夫、鈴木克典（ヤマハ）
△小池弘、△飯田誠一郎（ユポ）

◆合成紙に応用される多孔質ポリマーシートをコロナ放電によりエレクトレットとしギター響板の振動検出性能を評価した。

◆サドルに衝撃力を与えて出力された信号から、多孔質エレクトレットは 100 Hz～10 kHz の広帯域にわたり響板の振動を検出し、さらに信号の減衰が響板の振動減衰と対応することを見出した。

◆このように、多孔質エレクトレットは時間応答および周波数特性が優れるだけでなく、多孔質フィルムゆえに低密度かつ極めて柔らかいため、ギター響板に直接装着しても響板の振動への影響を最小限に抑えることが出来る。

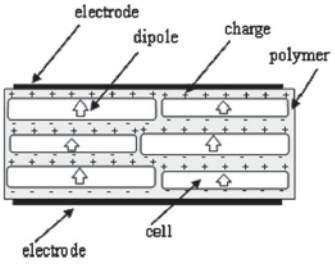


Fig. 1 Schematic diagram of the cross-section view of the cellular polymer film electret

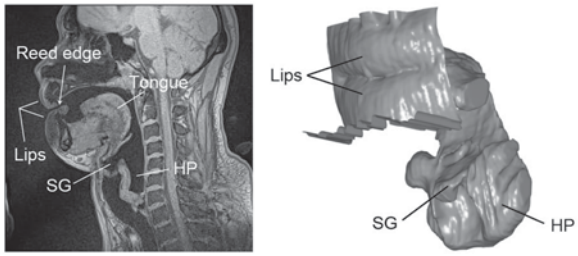
3-11-4

3-11-4 オーボエの演奏に特徴的な声道形状に関する検討

On the characteristic configuration of the vocal tract observed during the act of oboe

○鏑木時彦 田中光波（九州大）

Benade(1985)は、管楽器の演奏状態のモデルとして、管体の入力インピーダンスと奏者の声道入力インピーダンスとが直列に接続された状態で、リードに音響負荷がかかることを示した。従って、演奏中の声道形状や音響特性を明らかにすることは、楽器の演奏を理解する上で重要である。本研究では、プロの男性オーボエ奏者１名を被験者として、ダブルリード木管楽器の演奏時の声道形状を MRI(magnetic resonance imaging)を用いて3次元的に観測した。その結果、下図に示すように、食道の入口である下咽頭においてきわめて特徴的な声道形状が観察された。この下咽頭腔（図の HP）は明らかに喉頭腔（図の SG）より大きく、喉頭腔を後ろから取り巻くように生じており、音響的にも声道の主たる領域となっていることがわかった。さらに、声道のボリュームデータと時間領域有限差分法（FDTD 法）を用いて、リード端点位置から声道を見たときの入力インピーダンスを計算した。その結果、下咽頭腔は入力インピーダンスのピーク周波数を低下させ、第二ピークを強調することがわかった。



3-11-5

3-11-5 フルートの音孔条件の変化による音響特性の比較

Comparison of frequency characteristics of flute with different tone hole using a sounding device

☆斉藤猛，大田健紘（日本工大）

◆ベーム式フルートの C キーの大きさが他と比べ小さい事に注目し，この音孔条件が音響特性にどのような影響があるのかを検討した．

◆C キーを使用して発音する Cis 音・D 音を対象に，C キーの大きさや音孔位置が異なる管体を作製し計測をおこなった．

◆人間が吹奏すると吹奏条件（エアジェットの風量や空気圧）が吹奏ごとに異なるため，人工吹鳴装置を自作し一定の吹奏条件にて吹奏音を収録した．

◆**Figure 1 より，音孔条件の変化により基本周波数が異なる値となった．また，Table 1 より基音から第５倍音までの出現順を見てみると，周波数の変化や出現順の変化などがみられ音色の変化もうかがえた．**

◆**ベーム式フルートの C キーが他のキーよりも小さい理由は，操作性を重視した結果と推察した．**

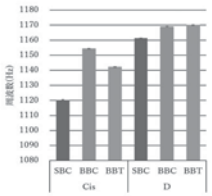


Figure 1 Average fundamental frequency of each flute with different tone hole

Table 1 Frequency from fundamental to fifth harmonic which is arranging in descending order of amplitude level of each frequency

Cis			D		
SBC	BBC	BBT	SBC	BBC	BBT
1125.3	1148.7	1137	1160.4	1172.2	1172.2
3364.1	3457.9	3422.7	3481.3	3504.8	3516.5
2238.8	2309.2	2285.7	2320.9	2344.3	2344.3
5591.2	4618.3	4571.4	4641.8	4676.9	4676.9
4477.7	5767	5708.4	5802.2	5849.1	5849.1

The first character of SBC, BBC, and BBT indicates the size of C key, the second character indicates the size of keys except for C key, and the third character indicates the position of C key. "S" is small, "B" is big, "C" is concentric with the original C key, and "T" is sharing the top position with the original C key.

3-11-6

3-11-6 付加製造法を用いた尺八の復元と音色の評価

Reproduction of shakuhachi by additive manufacturing and tone evaluation.

☆倉本 有紗，高橋 義典（都立産技高専），△水野 明哲（工学院大）

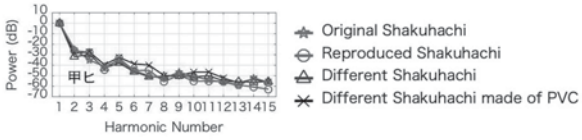
◆和楽器の一種である尺八は，竹を使用して作られていることから，一度破損・消失してしまうと全く同じ形状の復元することは困難である．

◆尺八はある程度硬化な材料であれば材質による音色の変化は見られないことから，これまで尺八を CT スキャンしたデータを元に 3D プリンタを用いて複製を作ることで，半永久的に尺八の音色・形状を残すことについて検討してきた．

◆その結果，尺八の内側に塗ってあるトノコや石膏の影響で輝度値が高くなり，オリジナルの尺八より内径が狭いことが問題であった．

◆3D モデルを作成する前に，CT 画像に内径調節を行うプロセスを導入することで，内径が実物に近い 3D モデルを作成することができた．

◆本報告では，内径調節を導入していない復元尺八に内径調節を導入した復元尺八のパーツを取り付け，音色の評価を行った．



3-P-1

3-P-1 音響的特徴量によるゲーム中の叫び声の識別

Shout discrimination using acoustic features during game-play

☆金子裕亮，有本泰子(千葉工業大)

◆ヘッドマウントディスプレイやモーションコントローラを使用した VR ゲームの普及により，ゲームの世界への没入感を高める技術が求められている。そのような技術の１つとして，笑い声や叫び声を入力とし，ゲームの世界へ適切なリアクションを返す”感情入力インタフェース”の開発を目指している。

◆ゲーム中の叫び声と叫び声以外の発話に対して音響的特徴量を抽出し,差異があるかを確認した。また，分析を基に大規模に抽出した LLD 特徴量を使用して，DNN による叫び声の識別実験を行った。

◆分析ではF_0の平均値と標準偏差，MFCC の C1 の平均値と標準偏差，継続時間長，音圧レベルの最大値に差異があることが確認され，叫び声の識別に有効であることが示唆された。

◆識別実験では評価用データ全体の正解率が 96.0%で，叫び声の再現率が 82.2%となった。

Table 1　識別性能

	正解率	適合率	再現率	*F-measure*
学習用	97.7%	87.1%	90.3%	88.7%
検証用	96.7%	84.5%	82.8%	83.7%
評価用	96.0%	80.4%	82.2%	81.3%

3-P-2

3-P-2　短発話を対象としたテキスト独立型話者認識のためのフレームレベル音素非依存特徴抽出

Frame-level phoneme-invariant speaker feature extraction for text-independent speaker recognition on extremely short utterances

○俵直弘，小川厚徳，岩田具治，
マーク・デルクロア（NTT CS 研），小川哲司（早大）

◆概要：

1. 発話単位 DNN（x-vector）とフレーム単位 DNN（d-vector）による話者特徴抽出法について発話長と話者照合精度の関係を調査した
2. 音韻情報を取り除く敵対的学習をフレーム単位 DNN に適用することで、より識別的な話者特徴が得られることを示した

◆ 評価方法：

- 様々な発話長の音声に対し、フレーム単位 DNN と発話単位 DNN を適用し得られた話者特徴を話者照合精度により評価

◆実験結果：

- 発話長が 1.4 秒未満の場合ではフレーム単位 DNN（FRM）が，それ以上の場合では発話単位 DNN（SEG）が有効であることを示した（Fig. 2）
- 単フレームの話者照合においてフレーム単位 DNN に対する敵対的学習が有効であることを示した（Table 1）

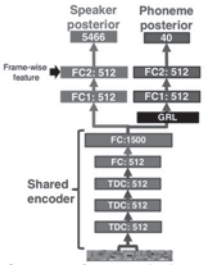


Fig1. Structure of proposed adversarial network trained on frame-wise loss

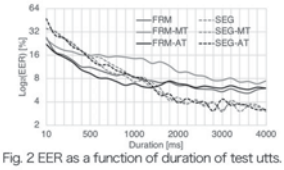


Fig. 2 EER as a function of duration of test utts.

Table 1 EER obtained by each systems of segment- and frame-wise speaker recognition task

System	All frames (avg.: 1228 frames)	Single frame
SEG	**3.07**	35.09
SEG+MT	3.14	35.88
SEG+AT	4.14	47.85
FRM	3.22	27.08
FRM+MT	3.22	36.00
FRM+AT	3.45	**22.13**

3-P-3

3-P-3　マルチモーダル音声強調における Cycle-Consistency の導入

Using Cycle-Consistency for Multimodal Speech Enhancement

☆池上凌（東京理大院・理工），△大村英史，桂田浩一（桂田研）

音声強調とは、雑音混じりの音声から目的の音声を強調して聞き取りやすくする技術を指す。この音声強調の品質を高めるために動画像を音声に加えて利用したものがマルチモーダル音声強調である。先行研究ではEnd-to-Endの深層学習器を用いることにより良好な結果を得ているが、未知の種類のノイズに対して性能が低下するという問題が残されていた。そこで本研究ではニューラルネットワーク（以下 NN）の学習の際に Cycle-Consistency を導入する手法を提案する。Cycle-Consistency によって深層学習の汎化性能が向上することから、より高品質の音声強調が実現できると考えられる。実験の結果、Table 1 に示すように SNR、PESQ 値ともに先行研究を上回る結果となった。また、提案手法で用いた損失関数$CCLoss$を改良した手法はより高い性能を示した。これらのことからマルチモーダル音声強調における Cycle-Consistency の導入と$CCLoss$の改良の有効性が確認できた。

Table 1 Experimental results

	SNR(db)	PESQ
先行研究	4.0379	1.2823
提案手法	4.4504	1.3427
提案手法(改良版)	5.5066	1.3556

3-P-4

3-P-4　対話プラットフォーム COTOBA Agent の開発とその IoT 拡張

COTOBA Agent: Dialog Platform with extension to operate IoT devices

○山上勝義，木村敏宏，土田正明，小野義博，水野孝久，松田繁樹
（コトバデザイン）

対話エージェントの開発・運用を行う既存の対話プラットフォームには，

- 記述可能な対話動作が単純なルール動作に限定される
- 対話エージェントの相互利用の機能がなく再利用性が低い
- 機器・センサとの標準化された連携の枠組みがなく個別実装対応

といった課題がある．これらの課題を解決するため，以下の特徴をもつ対話プラットフォーム COTOBA Agent を開発した．

1. 意図解釈機能付きのフルカスタマイズ可能なルールベース対話処理
2. 対話処理エージェントの相互呼び出しの枠組みによる再利用性向上
3. 機器・センサとの統合を容易にする IoT 入出力記述の枠組み提供

これらの特徴により、幅広い応用で対話エージェント開発・運用のサポートが可能である．

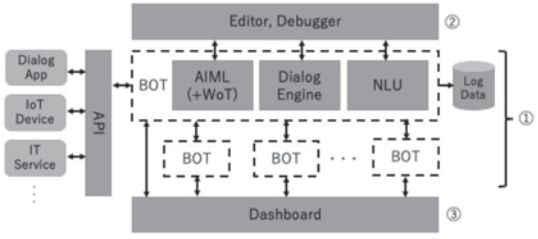


Fig.1 : Overview chart of COTOBA Agent

3-P-5

3-P-5 Dimensional Speech Emotion Recognition from Speech and Text Features using MTL

○ Bagus Tris Atmaja, Masato Akagi (JAIST)

- This study is aimed to evaluate the combination of acoustic and text features to improve the performance of dimensional speech emotion recognition (SER) using multitask learning (MTL).
- The task is to predict the degree of valence, arousal, and dominance (VAD) dimension within speech [-1, 1].
- Twenty-three acoustic features and 300 dimensions text features are fed to three LSTM layers with MTL.
- The result shows significant SER improvements, particularly on valence dimension measured in concordance correlation coefficient (CCC), see table and figure below for details.

Method	V	A	D	Total
Speech	0.110	0.494	0.352	1.056
Speech+text	0.416	0.506	0.476	1.399

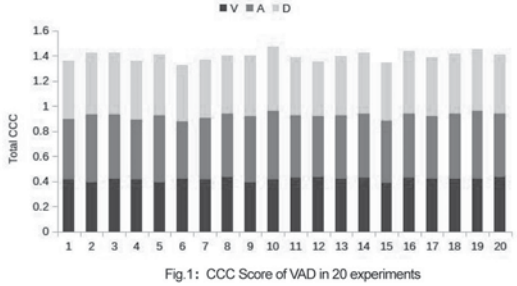


Fig.1: CCC Score of VAD in 20 experiments

3-P-6

3-P-6 ゲームプレイ中の笑い声検出モデルの構築

Modeling laughter detection in game playing

☆KHOO Yu Yen（帝京大学），有本泰子（千葉工業大）

◆ 背景と目的：

VR ゲームの普及でよりリアルなバーチャル空間が求められ、ゲーム内のアバターとの一体感の向上を狙って、笑い声や叫び声などを入力とする「感情入力インタフェース」の構築を目指す。本報告は、リアルタイムの笑声検出を目指し、SVM と DNN を用いてゲームプレイ中に表出される笑い声をフレーム単位で検出することを試みる。

◆ 方法：

フレーム単位で音響的特徴量を抽出し，SVM と DNN による３クラス（笑い声，笑いでない声，無音区間）の識別実験を行った。

◆ 結果：

SVM と DNN の笑い声の識別精度（F 値）はともに 75%前後であった。

Table 1: Results of DNN and SVM

	SVM			DNN		
label	precision	recall	f-score	precision	recall	f-score
laughter	0.84	0.71	0.77	0.79	0.74	0.76
non-laughter	0.89	0.91	0.90	0.91	0.91	0.91
silence	0.96	0.97	0.97	0.97	0.97	0.97

3-P-7

3-P-7 Analyzing Effects of Features in Automatic Fluency Detection of Spontaneous Speech

話し言葉の流暢性・非流暢性の識別における特徴量の有効性の分析

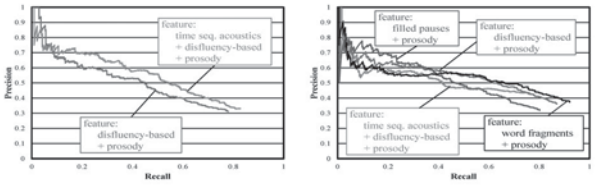
☆Huaijin Deng, △Youchao Lin, Takehito Utsuro (Univ. Tsukuba), Akio Kobayashi (Tsukuba Univ. Tech.), Hiromitsu Nishizaki (Univ. Yamanashi), △Junichi Hoshino (Univ. Tsukuba)

◆LSTM framework of integrating:

- disfluency-based features: filled pauses and word fragments
- prosodic features
- time sequential acoustic features

◆time sequential acoustic features:

- contribute to detecting *fluent* speech
- but not to detecting *disfluent* speech
-



(a) Binary classification of fluent vs. neutral-disfluent classes

(b) Binary classification of fluent-neutral vs. disfluent classes

Fig.1　Experimental results of the LSTM framework: Comparison of w/ and w/o time sequential acoustic features

3-P-8

3-P-8　平滑化調波構造パラメータマスクを用いた凸最適化に基づく雑音環境下音声強調手法の性能改善

Smoothed Harmonicity-Aware Parameter Mask for Convex-Optimization-Based Speech Enhancement

☆王浩南（立命館大院），若林佑幸（首都大/立命館大），福森隆寛，西浦敬信（立命館大）

雑音環境下の音声強調の研究分野において，振幅スペクトログラムのスパース性を用いた凸最適化に基づく手法が提案されてきた．このような手法の解は，コスト項の設計だけでなく，正則化パラメータにも影響される．本稿では，従来手法の正則化パラメータを改良する平滑化調波構造パラメータマスクを提案し，コスト項を変えずに従来手法の性能改善を図る．音質を評価する指標 PESQ (Perceptual Evaluation of Speech Quality)と音声明瞭度を評価する指標 STOI (Short-Time Objective Intelligibility)に基づく客観評価実験の結果を Table. 1 に示す．結果，従来手法の正則化パラメータに提案手法を適用することで，音声強調の性能改善を確認した．

Table. 1 The Average PESQ and STOI.

		White noise					Bubble noise					Factory noise				
		-3 dB	0 dB	3 dB	6 dB	9 dB	-3 dB	0 dB	3 dB	6 dB	9 dB	-3 dB	0 dB	3 dB	6 dB	9 dB
PESQ	Noisy	1.37	1.52	1.70	1.89	2.09	1.61	1.87	2.05	2.24	2.43	1.55	1.74	1.94	2.14	2.35
	Base.	1.83	2.07	2.22	2.45	2.67	1.65	2.04	2.13	2.35	2.56	**1.80**	**2.06**	2.28	2.48	2.68
	Prop.	**1.85**	**2.08**	**2.31**	**2.51**	**2.70**	**1.71**	**2.09**	**2.18**	**2.39**	**2.60**	1.78	2.05	**2.29**	**2.50**	**2.70**
	Prop.*	**2.12**	**2.31**	**2.48**	**2.65**	**2.80**	**1.91**	**2.11**	**2.30**	**2.50**	**2.67**	**2.06**	**2.29**	**2.44**	**2.62**	**2.79**
STOI	Noisy	0.57	0.64	0.72	0.78	0.84	**0.59**	0.65	0.72	0.78	0.83	0.57	0.64	0.71	0.78	0.84
	Base.	0.58	0.67	0.75	0.82	0.87	0.55	0.63	0.71	0.78	0.84	0.56	0.64	0.72	0.79	0.85
	Prop.	**0.61**	**0.71**	**0.78**	**0.84**	**0.88**	0.58	**0.67**	**0.74**	**0.80**	**0.85**	**0.59**	**0.68**	**0.76**	**0.82**	**0.87**
	Prop.*	**0.65**	**0.72**	**0.79**	**0.84**	**0.88**	**0.60**	**0.68**	**0.75**	**0.81**	**0.88**	**0.62**	**0.69**	**0.76**	**0.82**	**0.87**

※ **Underlined bold font** and **bold font** represent the 1st and the 2nd performed methods.

3-P-9

3-P-9　End-to-End とカスケード方式のアンサンブルによる音声翻訳の検討

A Study on speech translation systems ensembling end-to-end and cascade methods

☆民谷慎一郎，秋葉友良，塚田元（豊橋技術科学大学）

◆ 本研究では、音声認識モデルと機械翻訳モデルを用いたカスケード方式音声翻訳システムと、単一の音声翻訳モデルの End-to-End 音声翻訳システムを対象に翻訳結果のアンサンブルを行い、音声認識誤りの影響を低減することを検討する。また、カスケード方式の音声認識結果を書記素と音素の２種類を利用することでも音声認識誤りにも対処する。

◆ 英独の音声翻訳実験の結果、アンサンブルすることで単一の音声翻訳システムの性能を改善することを確認した。

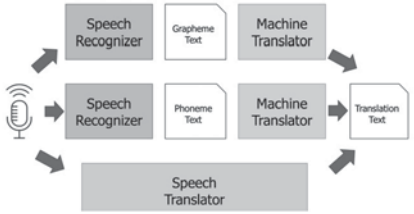


Fig.1 : ensemble speech translation systems

model	ST-BLEU
grapheme	14.98
phoeneme	15.09
End-2-End	11.71
E2E + gra	15.21
E2E + pho	15.29
E2E + gra + phone	16.96

Table.1: experiment result

3-P-10

3-P-10　音波形のビット列表現に基づく音楽・音声分類器の分析

Analysis of a Music / Speech Classifier based on Bit Sequence Representation of Audio Waveform

☆大川正暉，西崎博光（山梨大院）

◆音波形のビット列表現を用いた音楽・音声分類器の分析
- ➢ 深層学習を用いて音楽・音声分類器の学習
  - ✧ 振幅の変化とビット列の表現方法の分類精度を比較

◆実験：各種入力の分類精度の評価結果
- ➢ 使用データ：学習に GTZAN の約 70%，評価に残りの 30%
- ➢ 音波形の振幅を変化させ分類を評価
  - ✧ 振幅の変化による分類の優劣はなし
- ➢ 音波形のビット列表現を変化させ分類を評価
  - ✧ 正規化を行った 16 ビットの浮動小数点数表現の精度向上

Table 1: Classification Accuracy for Amplitude Change [%]

	振幅倍率			
	1.0	0.1	0.01	0.001
分類正解率	95.6	95.6	95.6	96.5

Table 2: Classification Accuracy for Bit Representation Method [%]

入力形式	Classification Accuracy		
	直接入力	最大値で正規化	標準化
Int16	95.6		
Float32	95.6	96.4	94.7
Float16	95.6	97.4	95.6

3-P-11

3-P-11　多人数のための音響・言語情報の重要度を考慮した応答義務推定

Estimating response obligation considered importance of acoustic and language information for multi-party conversation on smart speaker

☆柴田護，藤江真也（千葉工大）

◆スマートスピーカにおける多人数会話において，対話システムの応答義務を推定する手法を提案する．

◆話者ごとの発話を取得するために，マイクロホンアレイで取得した音声に対して音源分離を行う．

◆音響情報として，韻律情報（LLD 特徴量）と音声スペクトログラムを符号・復号化する自己符号化器の中間層の出力（AE 特徴量），言語情報として音声認識結果に対して Word2Vec モデルからの出力を抽出した．

◆ニューラルネットワークモデルを構築し，応答義務を推定する．

◆（a）提案手法，（b）重みを固定した手法，（c）各特徴量をすべて結合して入力とする手法とで比較を行った．

◆Fig. 1 に示すように，本手法により，システムの応答義務を推定することができることが確認された．

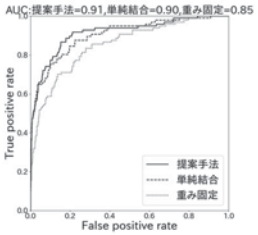


Fig. 1 : ROC curve

3-P-12

3-P-12　接客訓練のための音声対話システムの試作

Prototype Development of a Spoken Dialog System for Customer Service Training

☆佐野祐太（山梨大・工），レオ チーシャン（山梨大院），△飯田宗一郎（筑波大院），西崎博光（山梨大院），△星野准一，宇津呂武仁（筑波大）

◆接客訓練音声対話システム
- ➢ 従来の音声対話システムと異なり，ユーザと対話エージェントが逆の役割を果たす音声対話システム
- ➢ 対話の管理や接客の流れを状態遷移を用いて管理
- ➢ 訓練対話シナリオデータによる多種多様な接客訓練を実現可能

◆実験：接客訓練音声対話システムの評価実験
- ➢ 5 つの訓練対話シナリオを使用
- ➢ 6 つの評価項目について，各 5 段階で評価

◆結果：接客訓練音声対話システムの有用であると確認できた

Table 1　Experimental result

Evaluaiton item	Score
接客訓練音声対話システムの使い易さ	4.4
音声対話の使い易さ	4.0
音声対話のスムーズさ	3.4
訓練対話の自然さ	3.8
実務環境の再現度	3.4
接客訓練による接客の理解度	4.2

3-P-13

3-P-13　自由発話に対応した照応解析を備えた音声対話システム

A spoken dialogue system with anaphoric analysis for spontaneous utterances

☆森雷太，西村良太（徳島大），北岡教英（豊橋技科大）

◆対話システムでは，過去の話題が省略された発話にも対応する必要．

◆音声対話における自由発話では助詞の欠落などにも対応する必要．

◆主語や目的語が省略された自由発話に対して照応解析を行い，対話履歴を用いて補完する，以下の処理を行うシステムを提案．

Seq2Seq により自由発話を書き言葉に整形

↓

中心化理論と格フレーム辞書を用いて照応解析

↓

省略された話題を補完

◆実装したシステムが適切な返答を返すのかを検証するために，「話題を補完する対話システム」と「補完しない対話システム」での出力結果の比較を行った．(Table 1)

◆「補完しない対話システム」は，天気とは関係ない返答であったが，「補完する対話システム」では，過去の情報を加味した返答を返した．

◆発話の補完が対話システムにおいて有効であることを示した．

Table 1 : Comparison of the output results of the dialogue systems with/without complements, S and S', respectively

補完なしの入力の場合	補完ありの入力の場合
U_1：今日の天気教えて	U_1'：今日の天気を教えてください
S_1：今すぐお空に行って確認してきます	S_1'：お天気だといいですね
U_2：明日は良いの	U_2'：明日は天気が良いのですか
S_2：お願いします	S_2'：分かりません

3-P-14

3-P-14　空間特徴と音響特徴を併用する音響イベント検出の検討

Sound event detection using both spatial and acoustic features

☆陳軼夫，山田武志，牧野昭二(筑波大)

◆音響イベント検出において、ステレオ信号に含まれる空間情報を活用することにより検出精度の向上が期待できる。

◆Adavanne らは、音響特徴量として bin-mbe (binaural-log mel band energy) 、空間特徴量として GCC-PHAT (generalized cross correlation phasetransform) を入力とし、再帰型畳み込みニューラルネットワークである CRNN (convolutional recurrent neural network) を用いて音響イベント検出を行う手法を提案した。

◆本稿では、この手法の検出性能をさらに向上させるため、音響特徴量と空間特徴量の組み合わせ方、及び空間特徴量を入力する CNN の構造の最適化を図った。

◆実験の結果、Table 1 に示すように DCASE 2017 Task 3 の第 1 位の手法よりも高い検出精度を得た。

Table 1　Error rate and F-score for each method

		Development dataset		Evaluation Dataset	
Method	Features	ER	F-score	ER	F-score
CNN for spatial feature (output size: 32 x 8)	GCC-PHAT+mbe	0.466	71.6%	0.768	43.2%
	GCC-PHAT+bin-mbe	0.496	69.0%	0.808	40.9%
DCASE 2017 Chanllenge ranking 1st	mbe	0.55	69.3%	0.791	41.7%
	bin-mbe	0.52	69.1%	0.806	42.9%

3-P-15

3-P-15　大規模感情音声データベース JTES への聞き手ラベルの付与と音声感情認識での利用

Labeling of Perceived Emotion for Large-Scale Emotional Speech Database JTES and Emotion Recognition Using Listener Label

☆山中　麻衣，能勢　隆，千葉　祐弥，伊藤　彰則(東北大学)

◆感情音声の主観的な感情強度の分析・予測や感情認識精度の向上を目的とし，大規模感情音声データベース JTES に対してクラウドソーシングを用いて聞き手が知覚した感情ラベル（聞き手ラベル）の付与を行った

◆聞き手ラベルの感情付与割合から各話者の感情毎に感情の主観的な強度（知覚感情強度）を表す知覚感情ベクトルを定義し，これを用いて深層学習による知覚感情強度の予測および感情認識実験を行った

◆実験結果から知覚感情強度と認識精度の間に正の相関が認められ，感情認識の性能が知覚感情強度に依存することを明らかにした (Fig. 1)

◆さらに，聞き手ラベルの利用により，従来の感情分類だけでなく，その強度もより高い精度で推定できることを示した (Fig. 2)

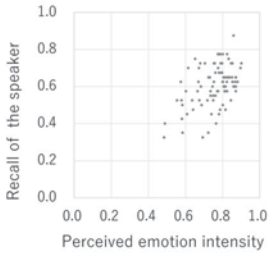


Fig.1　Relation of human recognition and machine recognition

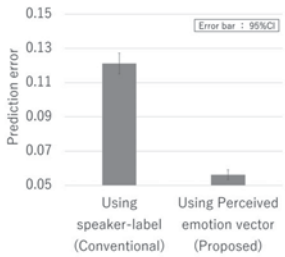


Fig. 2　Prediction Error Using Speaker's emotion expression

3-P-16

3-P-16　Improvement of x-vector for short utterance spoken language identification

○Peng Shen, Xugang Lu, Komei Sugiura, Sheng Li, Hisashi Kawai

Short utterance-based spoken language identification (LID) is one of the important tasks for real-time multilingual speech processing systems. Compared with long utterances, the representation of short utterances has large variation, that prevents the model from generalizing well. How to improve the generalization of the model on short utterances is still a challenge task. In this work, we evaluate an DNN-based language embedding approach, i.e., x-vector, for LID tasks, and propose a knowledge distillation-based training approach to improve x-vector extraction by transferring the internal representation knowledge of a long utterance-based teacher model to a short utterance-based student model.

Table. 1: Comparison of the proposed method with baseline methods

Methods	Input duration	3s	10s	30s
ResNet-GAP [6]	2-8s	11.28	5.76	3.96
CNN-BLSTM SAP [6]	2-8s	9.50	3.48	1.77
ResNet-GAP (short)	1-10s	9.18	5.14	3.71
ResNet-GAP (long)	5-10s	9.78	3.38	1.81
x-vector (short)	1-10s	**8.94**	4.12	2.73
x-vector (long)	5-10s	10.89	**3.29**	**1.25**
Proposed (0.9)	1-10s	**7.92**	**2.87**	**1.62**
Proposed (0.7)	1-10s	8.02	3.75	2.18
Proposed (0.5)	1-10s	8.62	3.61	2.46

3-P-17

3-P-17 猛禽類の鳴き声自動判別

Automatic discrimination of raptor sounds

○金寺 登，△阿知良 澪，△藤井 烈(石川高専)，△片桐寿通，△山川将径，△田屋祐樹，△辰橋浩二，△前 正人(国土開発センター)

◆深層ニューラルネットワーク(DNN)により６種類の猛禽類等（ミサゴ，ハチクマ，オオタカ，サシバ，ノスリ，ミゾゴイ）の鳴き声を自動判別した。

◆特徴量には 12 次の MFCC と対数パワーを用いた。

◆学習データの不足を補うため，ラベリングした各区間を n 等分し，n 等分点の 1 フレーム+前後 17 フレーム（計 35 フレーム）をまとめて入力特徴量とした。フレームシフトは 10 ms としたため 350 ms 程度の時間周波数情報を入力として用いたことになる。

◆ニューラルネットワークの構造は入力層，隠れ層 2，出力層の 4 層ニューラルネットワーク構造とした。隠れ層の活性化関数は sigmoid 関数，ReLU 関数もしくは tanh 関数であり，ユニット数は 2 つの隠れ層ともに 1024 ユニットとした。出力層の活性化関数は softmax 関数を使用した。また、入力は 35 フレーム分の MFCC(455 ユニット)，出力は６種の猛禽類等とその他に対応する７ユニットである。

◆Table 1 のように 90%を超える正解率が得られた。

Table 1：Recognition results

特徴量	活性化関数	正解率[%]
MFCC (C0-C12)	tanh	93.3
	ReLU	92.5
	sigmoid	92.1
MFBANK(40ch)		83.6

3-P-26

3-P-26　高齢者にとって聞き取りやすい音声合成の実現に向けた模範話者データの収集と分析

Exemplary speech data collection and analysis
toward establishing easy-to-listen speech synthesis for elderly adults

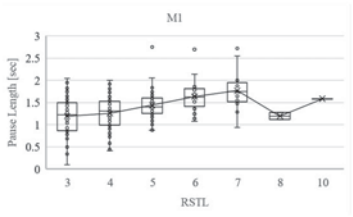
○中嶋秀治（日本電信電話株式会社 NTT コミュニケーション科学基礎研究所），
青野裕司（日本電信電話株式会社 NTT メディアインテリジェンス研究所）

◆高齢者にとって聞き取りやすい音声合成を実現するため，実際に高齢者が聞いて分かりやすいと支持した高齢者向けの模範話者の音声を収集し，既存研究の観点からの文間ポーズ長の分析を行なった

◆修辞構造理論 [Yang+ 2012]またはトピック遷移 [Smith 2004]に基づく文間距離を表わす２通りのラベルを文間に付与し，それらと文間ポーズ長との相関分析をそれぞれ行ない，以下を確認した

- 従来研究で示されたような文間距離の増加に伴うポーズ長平均値の緩やかな伸長が，25 文書*全体では*，確認できたものの，分布の重なりが大きい（本文 Fig.2（下図はその一部））
- *文書毎には*，無または負の相関の文書が半数近い（Table 2）

◆複雑ではない観点（例えば [中嶋 2015] のレイアウト）からの制御の検討を続けたい



3-P-27

3-P-27　高齢者向け発話の韻律的特徴の分析

An analysis of prosodic features of utterances for elderly person

☆岡本泰秀，水野秀之（公立諏訪東京理科大学），
中嶋秀治（NTT コミュニケーション科学基礎研究所）

◆高齢者が聞き取りやすいと評価する話者の音声(模範発話)と，一般人の読み上げ音声(学生発話)の韻律的特徴の比較分析し以下が分かった．

- ポーズ句単位の基本周波数の最大値については，平均値は同等だが模範発話の分散は３倍以上学生発話の分散より大きかった．
- ポーズ句単位の話速については，平均値及び分散ともに同等だが，学生発話の話速が遅いポーズ句では模範発話の話速が早く，学生発話の話速が速いポーズ句では模範発話では遅い傾向があった．
- ポーズ句間の基本周波数の最大値の差分を Fig.1 に示す．模範発話は学生発話より変動量が±50cent 以下のポーズ句が少なくポーズ句間で基本周波数の最大値が大きく変動する傾向があった．

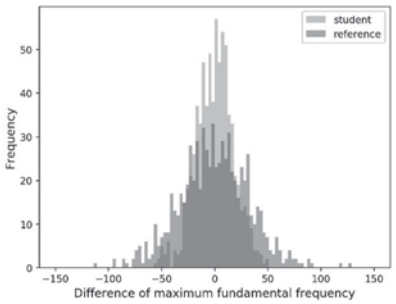


Fig.1 : Histogram of the difference of maximum fundamental frequency between pause phrase.

3-P-28

3-P-28　Investigation of Spatiotemporal Brain Network Dynamics in Perceiving Real words and Pseudo words

○Taiyang GUO[1]，Jianwu Dang[1,2] ，Bin Zhao[1,2]
[1]Japan Advanced Institute of Science and Technology, Japan
[2]Tianjin University, China

◆ This study introduced fMRI constraint on the EEG-based brain network dynamics to investigate the neural mechanism and functional integration of dynamic brain networks in speech perception and comprehension of Chinese real words and pseudo words.

◆ According to the results, we speculate that the sematic processing networks were unconsciously activated when real words can be identified, while the sematic processing networks are consciously activated with some delay to retrieve some possible meaning for the pseudo words.

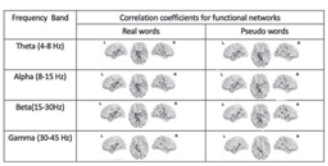


Fig. 1 Brain network activity in four frequency bands around 600ms for real words and pseudo words.

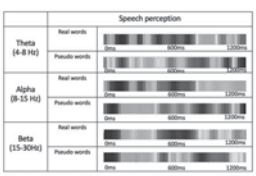


Fig. 2 Correlation coefficients of functional networks in the three frequency bands for real words and pseudo words.

3-P-29

3-P-29　韻律的特徴・ラベルを用いたDNN音響モデルに基づく英語音声中の韻律エラー検出に関する実験的検討

Experimental study on prosodic error detection in English utterances using DNN-based acoustic models trained with prosodic features and labels

○瀋陽，安藤慎太郎，峯松信明，齋藤大輔 (東大)，小橋川哲 (NTT)

◆ This paper investigate how to utilize DNN acoustic models trained with prosodic features and labels to detect prosodic error in Japanese English. There are four models trained with/without prosodic features or labels. Posteriorgram-based DTW difference between incomprehensible L2 speech and native speech is computed for each model. By comparing differences from baseline model trained without prosody and models trained with prosody, it is possible to detect prosodic error in L2 speech. The result showed that adding stressed vowels to output phoneme labels can help detect prosodic error.

3-P-30

3-P-30　高齢者の音声知覚特性に基づいた音声の明瞭化手法の検討

Study on monosyllable enhancement of synthetic speech based on speech perception characteristics of elderly people.

☆篠崎嵩大，松本悠希，朝倉巧（東京理科大）

◆高齢者においては、単音節毎で有効な明瞭化手法が健聴者と比較して異なると仮定し、有効な明瞭化手法を調査することを目的とする。

◆健聴者と高齢者の合成音声の正答率を調べるために、21 歳から 82 歳までの 25 名を対象として、原音声と明瞭化手法を施した合成音声を対象とした聴取実験を行った。聴取実験で使用する単音節は、高齢者にとって聞き取りづらいとされる/bi/、/gi/、/hi/、/ki/とし、また、明瞭化手法は子音部振幅増幅(2 倍, 4 倍)、母音のエネルギー定常部抑圧、時間領域変換、子音部時間伸張加工、フォルマント強調を使用した。

◆本研究では、聴取実験の結果を基に、それぞれの単音節毎で高齢者にとって有効な明瞭化手法を考察する。

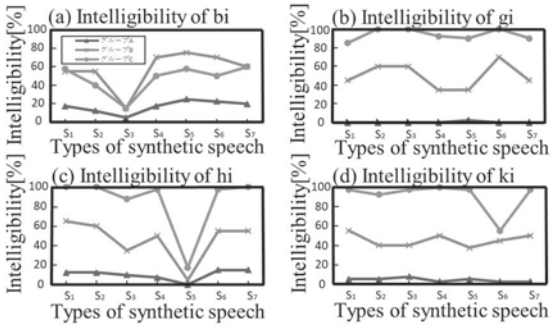


Fig.1 Intelligibility of synthetic speech

3-P-31

3-P-31　周波数伸縮に基づく話者匿名化のためのクラウドソーシングに基づくパラメータ最適化

Crowdsourcing-based parameter optimization for frequency warping-based speaker anonymization

高道 慎之介 (東大院・情報理工), ○小沼 海, 金田 卓, 金田 隆志 (HOYA), 齋藤 佑樹, 郡山 知樹, 猿渡 洋 (東大院・情報理工)

個人情報漏洩防止のために
音声の話者性を匿名化したい

でも，不自然な音声にしたくない

匿名性・自然性を両立させる
モデルパラメータを
大規模主観評価で決定

3-P-32

3-P-32　正則化線形予測に基づく時変複素音声分析とその評価

*On regularized-LP based TV-CAR speech analysis and its evaluation

○舟木 慶一(琉球大学)

解析信号に対する時変複素音声分析として、ℓ_2ノルム正則化方式をすでに提案した[1][2][3]。[1]は周波数軸でスペクトルの急激な変化を抑制するℓ_2ノルム罰則化項を導入した **RLP** 方式であり、[2]は時間軸でスペクトルの急激な変化を抑制するℓ_2ノルム罰則化項を導入した **TRLP** 方式である。[3]は **RLP** と **TRLP** のハイブリッド方式である。本稿ではハイブリッド方式をF_0推定で評価した詳細な実験結果について述べる。

[1]Keiichi Funaki, "TV-CAR speech analysis based on Regularized LP," Proc. EUSIPCO, A Coruna, Spain, Sep.2019.
[2]舟木慶一、"時間正則化線形予測に基づく時変複素音声分析," 電子情報通信学会第 34 回 SIP シンポジウム、とりぎん文化会館、鳥取、2019 年 11 月
[3]舟木慶一、"ℓ_2ノルム正則化に基づく時変複素音声分析," 電子情報通信学会音声言語シンポジウム、NHK 技研、東京、2019 年 12 月

3-P-33

3-P-33　言語情報なし感情合成音を学習に用いた CycleGAN による感情変換方式の検討

Study of emotional voice conversion by CycleGAN using emotional synthesized speech without linguistic information for training

☆松本剣斗，原直，阿部匡伸(岡山大院・HS 統合科学研)

◆ノンパラレル感情変換において，同一発話内容でない発話を学習に用いるため発話間のギャップが大きく，十分な性能が達成できていない.

◆我々はこれまでに WaveNet を用いた言語情報を含まないが感情情報は伝える，人間の発声した音声のような音(Speech-like Emotional Sound : SES)を生成する方式について検討した. [松本他, APSIPA　ASC 2019].

◆本稿では，SES を学習データとした CycleGAN による感情変換方式を検討する.

◆感情表現を一対比較法によって，音質を AB テストにより評価した.

◆感情表現に関して，変換前の音声と提案方式による変換音声の間に有意差があり，感情変換がおこなわれていることを示した.

◆Fig.1 より，生音声を学習データとした従来方式より提案方式の方が変換音声の音質が良いことが明らかとなった.

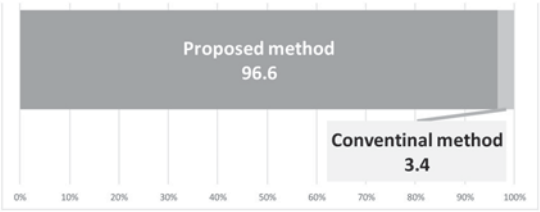


Fig.1 : The result of the AB preference test

3-P-34

3-P-34　携帯機器を用いた口唇情報利用発声支援デバイスの開発

Development of Mobile Device-based Speech Enhancement System Using Lip-reading

○濵田三弦，福山晃平，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，スマホによる低コスト，周囲の視線が気にならない外観，カメラ，ディスプレイ連携による使いやすいユーザインターフェイスなどを特徴とする読唇方式で発声支援実現のための基礎検討を行った．今回は，YOLOv3-tiny を用いた口唇認識による母音推定の認識精度の検証，使用感調査を行った．

◆学習では 6 人の実験協力者の”あ，い，う，え，お，無声音”の顔画像を合計 72,907 枚用意し，224×224 の解像度で学習を行った．認識精度の評価実験の結果，実験協力者 8 人のテスト用データ 1,440 枚に対し口唇を検出できクラスが正しく識別できたのは1,196枚(83.1%)であった．スマートフォンアプリの開発はUnity を使用し，ユーザインターフェイスのデザインと YOLOv3-tiny の実装を行った．

◆アプリケーションの使用感評価実験では，スマートフォンの発声支援アプリの有効性を確認したが，さらなる認識速度の向上が必要であることがわかった．

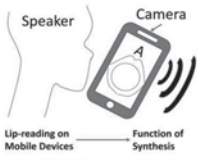


Fig. 1 発声支援システムのイメージ


Fig. 2 開発したスマートフォンアプリ

3-P-35

3-P-35　有声重子音での咽頭拡張: リアルタイム MRI の分析

Pharyngeal expansion during voiced geminates: An rt-MRI analysis

○藤本雅子（早稲田大），北村達也（甲南大），前川喜久雄（国語研）

◆日本語の標準語では有声阻害音の重子音は「バッグ(bag)」，「ベッド(bed)」等の外来語の語彙にみられるが，実際の発音では子音区間の一部または全体が無声化することが多い。

◆南琉球方言では，有声阻害音の重子音が「バッダ(脇)」など日常使用する語彙にも生じ，かつその有声性が子音区間を通じて保たれる。

◆阻害音の子音では声道内の閉鎖や狭窄のため声帯振動の維持が困難になるが，持続時間の長い重子音ではそれが顕著になると考えられる。

◆有声阻害音では無声阻害音に比べ，声道内の閉鎖や狭窄の弱化，咽頭の拡大，喉頭の下降，軟口蓋の下垂の調音特徴が生じる。これらは声帯振動を保つための方策と考えられる。

◆宮古島の池間方言話者の Real-time MRI を分析した結果，有声重子音/dd/，/zz/の咽頭径がそれぞれ/tt/，/ss/より長く，有声音での咽頭の拡大が認められた。

◆本稿では日本語標準語話者が発話した/tt/-/dd/，/kk/-/gg/での咽頭径を比較した。

◆その結果，有声の/dd/の咽頭径が無声の/tt/より大きく子音後半で拡大する傾向が認められたが，軟口蓋破裂音では/gg/と/kk/の咽頭径に一定の傾向が見られなかった。

◆これは咽頭径の拡大の傾向が調音位置により異なる可能性を示唆する結果と考えられる。

◆今後，他の調音位置の子音などについても検討が必要である。

3-P-36

3-P-36 Speech chain を模倣したボルツマンマシンによるワンショット多対多声質変換の検討

Study on one-shot many-to-many voice conversion based on speech chain Boltzmann machine

◎岸田拓也，中鹿亘（電通大）

◆ 声質変換を行うためのモデルの学習に用いられていない話者２名の少量の音声サンプルが与えられた状況で、一方の話者の発話を他方の話者が話しているかのように声質を変換することを「ワンショット多対多声質変換」と位置付ける。

◆ 我々がこれまでに提案した speech chain ボルツマンマシンを用いた手法では、未知話者の声質を学習話者の声質の足し合わせで表現することでワンショット多対多声質変換は可能である。しかし、特徴量のフレーム単位での変化の影響を受けやすく、品質が安定しないことが想定される。

◆ そこで本モデルに発話単位で抽出される話者埋め込み表現を組み合わせる手法を提案したところ、変換音声の品質が従来法から改善されることを確認できた。

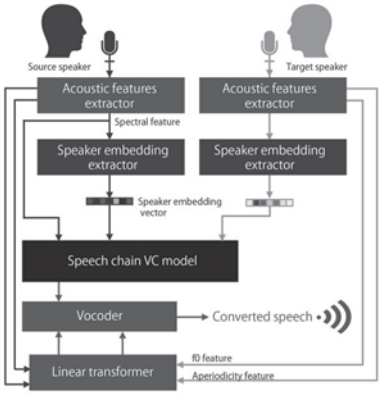


Fig.1 : Proposed one-shot many-to-many voice conversion method.

3-P-37

3-P-37　発声支援のための読唇手法の検討

Preliminary Study of Lip-reading Method for Speech Enhancement

☆福山晃平，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，PC，スマホによる低コスト，周囲の視線が気にならない外観，カメラ，ディスプレイ連携による使いやすい UI などを特徴とする読唇方式で発声支援実現のための基礎検討を行った．今回は，発声したい単語を登録することで推定でき，少量のデータを用いて使用者に最適化する手法を検討した．

◆モデルの学習には，各母音(A, I, U, E, O)と閉唇状態(X)の 6 種類の口形と口形を 2 音節ずつ組み合わせた計 36 種類の動画を使用した．単語認識精度の確認として，テスト用データ 15 種類の動画を使用した．単語認識精度の結果を Table 1 に示す．

◆提案手法では，PC やスマートフォンのカメラで撮影した動画を 30fps で画像に変換し，各画像に対して口唇領域画像の抽出と部位画像の特徴量抽出を行う．さらに各母音と閉唇状態の 6 種類の口形を推定し，生成された口形列から単語を認識する．

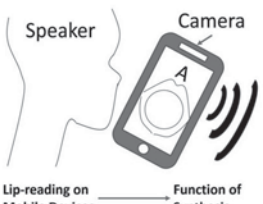


Fig. 1 Image of the speech enhancement system

Table 1 Word recognition results

	True	Prediction
1	ありがとう	○
2	いいえ	すみません
3	おはよう	○
4	おめでとう	○
5	おやすみ	おはよう
6	ごめんなさい	○
7	こんにちは	○
8	こんばんは	○
9	さようなら	○
10	すみません	○
11	どういたしまして	すみません
12	はい	○
13	はじめまして	○
14	またね	○
15	もしもし	おめでとう

3-P-38

3-P-38 End-to-End Articulatory Attribute Modeling for Low-resource Multilingual Speech Recognition

© Sheng Li, Chenchen Ding, Xugang Lu, Peng Shen, Hisashi Kawai (NICT)

◆ Many low-resource languages using other writing systems typically have more symbols than alphabets. As a result, the output nodes of the softmax layer are too large and make constructing multilingual E2E ASR system very challenging.

◆ In this study, we investigate an efficient method of encoding multilingual transcriptions with universal articulatory representations. The pronunciations from all of the existing languages can be decomposed into a set of "atom" units of the articulation.

◆ Experiments show the multilingual attribute-based model can significantly outperform other models (character, word, and globalphone-based). The proposed universal articulatory representation can share knowledge between different languages more effectively. It also provides more compact representations.

Multi-lingual model	CER%			
	Myanmar	Khmer	Sinhalese	Nepalese
#w=71,812	23.7	1.5	34.2	36.7
#c=365	26.6	**0.7**	14.5	**10.8**
#p=182	28.1	2.0	**13.3**	**10.3**
#a=23	**21.6**	2.4	**13.6**	**10.7**

The results compared to the lowest result without statistical significance (p-value < 0.05) are shown in bold fonts.

3-P-39

3-P-39 StarGAN-VC モデルにおける潜在表現への制約の有効性について

Effectiveness of constraints on latent representation in StarGAN-VC model

☆柴宮怜，△大村英史，桂田浩一(東京理科大)

本研究では StarGAN-VC の GAN 等の学習部を取り除き，声質変換を行う Encoder-Decoder 型変換器における Encoder 部分の出力である潜在表現に対し，話者非依存とする制約を導入することによって，StarGAN-VC と同等の変換性能を得ることを目標とする．これにより StarGAN-VC の学習と比べて学習を単純化しつつも性能を保つことが可能になる．8 話者間の変換器を構築して制約の有無に関する評価を行った結果，制約を与えることでメルケプストラム歪のスコアが大幅に改善することが分かった．また StarGAN-VC と比較したところ，同程度の性能を得ることができたことから，制約を与えるという簡便な方法の有効性が明らかになった．

Table 1 メルケプストラム歪みの比較

変換の種類	制約あり	制約なし	StarGAN-VC
女性→女性	6.618	8.671	7.245
女性→男性	6.839	8.910	6.997
男性→男性	6.674	7.877	6.603
男性→女性	6.845	8.683	6.948
全体	6.758	8.455	6.870

3-P-40

3-P-40 人間 GAN：人間による知覚評価に基づく敵対的生成ネットワークと音声の自然性知覚における評価

HumanGAN: generative adversarial network with human-based discriminator and its evaluation in naturalness modeling of speech

☆藤井一貴(徳山高専，東大院・情報理工)，
齋藤佑樹，高道慎之介(東大院・情報理工)，馬場雪乃(筑波大)，
猿渡洋(東大院・情報理工)

◆人間による知覚評価を用いて生成モデルを学習する敵対的生成ネットワーク(GAN)を提案する．Fig.1 に提案法と従来法の比較図を示す．

◆通常の GAN は実在データ分布を逸脱した領域の生成は不可能である．

- ➢ 一方で，人間の知覚はデータの逸脱に対して許容範囲を持つ．
- ➢ 例：人間の音声を加工し創り出した非実在音声でも人格性を容認

◆本研究では，通常の GAN の識別モデルを人間によるクラウドソーシングを用いた評価に置換することで，人間が許容できる範囲(知覚分布)を表現可能な生成モデルを学習する．

◆音声の自然性に関する主観的評価により，人間の知覚分布が通常の GAN で表現可能な分布よりも広い範囲をカバーしていることを示し，本研究の枠組みの必要性を示す．また，提案法により人間の知覚分布を表現する生成モデルを学習できることを示す．

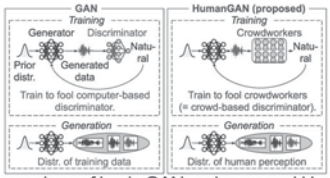


Fig.1 : Comparison of basic GAN and proposed HumanGAN.

3-P-41

3-P-41 マルチタスクモデルを用いた disentangleな学習による楽器音変換

Instrumental sound conversion using a multitask model

☆荒川賢也，岸田拓也，中鹿亘 (電通大)

◆ 近年，音楽音響信号にニューラルネットワークを適用する研究が多く発表されており，音楽の楽譜情報や楽器音などのドメインを変換する研究なども行われている．

◆ 本論文では楽譜情報を保持したまま楽器ラベルを中間特徴表現から削除することで楽器ドメイン変換を行う手法を提案する．

◆ 本論文ではオートエンコーダのエンコーダ出力に対して敵対的に学習を行うラベルの識別器，協力的に学習を行う楽譜の生成器を追加したマルチタスクモデルを提案する．

◆ 結果として提案手法は従来手法のFader Networksと比較して客観評価実験および主観評価実験で高い楽器音変換の精度を示した．

Table 1 : Result of evaluation

Method	RMSE	MOS	SD
source	0	4.57±0.10	1.48±0.18
target	359.34	4.57±0.10	3.68±0.15
conv.	323.56	2.28±0.10	**2.51**±0.19
prop.	**317.52**	**2.68**±0.11	2.26±0.22

3-P-42

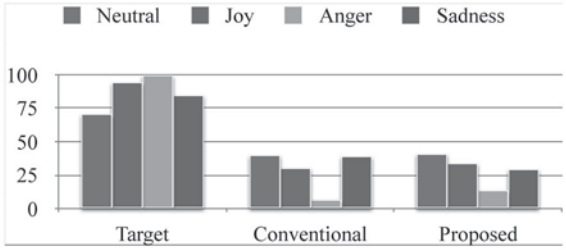
3-P-42　適応型 RBM を用いた音声情報の分離による話者と感情の同時変換

Simultaneous Conversion of Speaker and Emotion by Separating Speech Information Using Adaptive RBM

☆塚本伸，岸田拓也，中鹿亘 (電通大)

- 声質変換と感情音声変換の両方を行う場合，それら二つのモデルを別々に学習させる必要がある．
- 本稿では，適応型 RBMを拡張し，話者と感情の変換を同時に行えるモデルを提案する．また，従来の適応型 RBM を用いて感情のみを変換した音声との比較を行う．
- 主観評価実験より，話者と感情を同時に変換する提案手法が感情のみを変換する従来手法と比較し，ほぼ全ての感情について上回る結果となった．
- 提案手法は話者も同時に変換を行うので，目標話者の音声に対する回答の割合と傾向が近くなったと考えられる．

Figure. 1：Results of the subjective evaluation.
Answer rates correctly recognized as an intended emotion for each conversion condition.



3-P-43

3-P-43　中国語を母語とする日本語学習者による態度音声の韻律的特徴

Prosodic features of Japanese attitudinal speech by Mandarin Chinese learners

☆李歆玥（神戸大院），石井カルロス寿憲（ATR），林良子（神戸大）

- ◆日本語母語話者と中国語を母語とする日本語学習者が発話した態度のペアである「友好/敵対」，「丁寧/失礼」，「本気/冗談」，「賞賛/非難」から音響特徴量（F0mean，F0range，Duration，H1-A1，H1-A3，F1F3syn）を測定し，それらの態度間および言語間の違いについて検討した。
- ◆態度音声にあらわれる母語話者と学習者の句末音調（平叙文と疑問文）の特徴についても検討した。

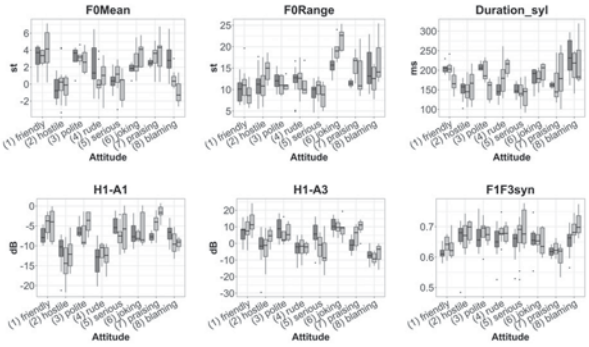


Figure 1: Distributions of prosodic and voice quality features to attitudes in all languages.
(red: Mandarin Chinese by native; yellow: Japanese by Mandarin Chinese learner; grey: Japanese by native)

3-P-44

3-P-44　日本語母語話者による英語自然発声を対象とした流暢さ(fluency)の自動スコアリングの検討

An experimental study of automatic scoring of fluency of spontaneous English utterances by Japanese learners.

☆安ヵ川彩乃，安藤慎太郎，紺野瑛介，林振超，井上雄介，齋藤大輔(東大)，斉藤一弥(UCL)，峯松信明(東大)

キーワード：言語学習　音声分析　音響特徴量　音声認識　韻律

- ◆ 語学のスピーキングテストにおける指標の一つの fluency についてその自動スコアリングと再定義を試みた。
- ◆Fig.1 の問題を用いた日本語母語話者の英語の自由発話の先行研究のデータから音響特徴量を取り出し、英語母語話者による fluency スコアとの相関を調べた。
- ◆Table.1 の Lasso 回帰分析結果を得たとともに、fluency の定義について発音や語彙サイズとの関係性を示した。

Fig.1. Questions used in the previous research

		A)	B)	C)
回帰変数	認識率	6.0	5.7	5.0
	語彙サイズ	1.9	2.1	2.6
	最大音素事後確率	0.0	0.73	0.87
	話速	2.8	2.3	2.8
決定係数 r2		0.63	0.66	0.65
相関係数		0.79	0.84	0.83

Table 1. Lasso regression analysis

3-P-45

3-P-45　2 段階声質変換を用いた狭帯域音声の広帯域化

Broadband up-sampling method from telephone band speech using pix2pix speech conversion system

◎小林洋介，△山部匠，赤泊寛和，野口啓太(室蘭工大)

- ◆様々な情報技術の開発に伴い，高品質な音声情報を利用したシステムが開発されている。一方で既設の内線電話や構内放送などは電話帯域音声を想定した設計のまま利用され続けている。特に構内配線が関わる設備は，設備更新コストが高く，既存設備を建物の耐用年数一杯まで使い続けることも想定され，低品質な機器が使われ続ける。
- ◆既存の配線設備を用いたまま，交換可能なマイクロホン及びスピーカに取り付けられる装置による高品質化として 2 段階声質変換による話者変換により電話帯域に制限された音声を元の広帯域音声に戻す声質変換とその客観品質評価を男女各２話者で行った。
- ◆実際に学内の内線電話系により帯域制限とノイズ加算された音源に対し広帯域化変換を行い，周波数領域の劣化度合いを示す log likelihood ratio (LLR)と明瞭性の指標となる short-time objective intelligibility (STOI)で評価した。結果を Table 1,2 に示す。周波数領域では品質改善したが，明瞭性の改善は話者による結果となった。

Table 1　Speech quality evaluation using LLR

Speaker	FKN	FYM	MMY	MSH
Tel. band	2.15	2.01	1.72	1.90
Wide band	**1.85**	**1.66**	**0.84**	**1.63**

Table 2　Speech quality evaluation using STOI

Speaker	FKN	FYM	MMY	MSH
Tel. band	**0.131**	0.233	**0.884**	0.213
Wide band	0.080	**0.243**	0.820	**0.229**

3-P-46

3-P-46 An Experimental Study of Sequential Annotation for Comprehensibility Based on Native Listeners' Shadowing and Reading

☆ Zhenchao Lin, Daisuke Saito, Nobuaki Minematsu
(The University of Tokyo, Graduate School of Engineering)

◆Nowadays, language teachers claim that the goal of verbal skill training should not be based on native-likeness but based on intelligibility and comprehensibility [Derwing2015+].

◆In order to automatically evaluate the utterance, the dataset including annotations, which is also called as corpus, is required for CALL (Computer-Aided Language Learning).

◆Comprehensibility-based corpora are difficult to find and the annotation is not fine-grained enough.

◆Since smooth shadowing can reflect smooth understanding [Trisitichoke+2019], the acoustic feature of native shadowing showed high correlation with perceived comprehensibility.

◆In this paper, native shadowing is viewed as sequential annotations and these annotations are mapped on each frame of a learner's utterance. It is also discussed how to acquire and map these annotations in a reliable way, and how we validate these sequential annotations.

◆As a result, correlations of machine objective score to human subjective score reach a high level.