2-1-1

2-1-1　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 107 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、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。

2-1-2

2-1-2　ITU-R 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

○大出訓史（NHK 技研）

◆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いる。

◆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として 2014 年に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のコンテンツ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。

◆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。

◆オブジェクトベース音響の導入によって，副音声サービスを効率的に放送できるほか，視聴者が番組音声をカスタマイズできるようになる。

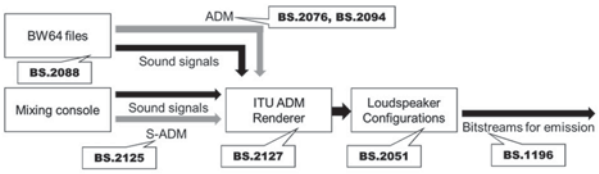


Fig.1 Recommendations related to Next Generation Audio (NGA)

2-1-3

2-1-3　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、すでに実装するメーカが現れている。

2-1-4

2-1-4　MPEG におけるオーディオ技術の標準化動向

Standardization of audio technology at MPEG

○杉本岳大（NHK）

◆本稿では，最新の MPEG 規格である MPEG-H 3D Audio (3DA)と MPEG-I Immersive Audio Coding (IAC)を中心に紹介する。

◆MPEG-H 3DA の標準化は，先進的な音響システムの符号化方式の策定を目的に 2013 年頃から開始された。2015 年に規格のベースとなる Main プロファイル，2017 年に実装用の LC プロファイルが発行された。図 1 に MPEG-H 3DA のデコーダ構成を示す。

◆MPEG-I IAC は AR や VR などの高い没入感を提供するメディアや Social VR のような空間／体験共有型メディア用のオーディオ技術として，標準化が進められている。MPEG-I IAC は AR や VR で必要とされる 6DoF（degrees of freedom）に対応可能なオーディオ技術である。図 2 に MPEG-I IAC システムの概要を示す。

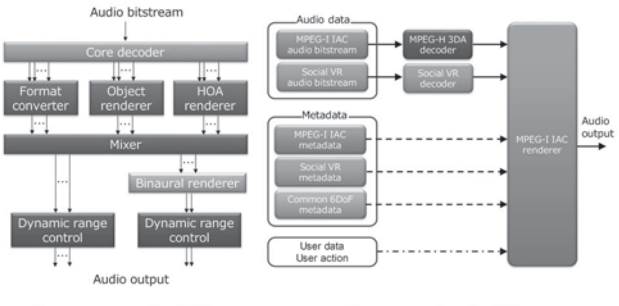


図 1 MPEG-H 3DA のデコーダ構成

図 2 MPEG-I IAC システムの概要

2-1-5

2-1-5　相反法による頭部インパルス応答の高速計測システム

Fast hrir measurement system with reciprocal method

☆ 寺島眞生，森川大輔，モクタリ パーハム，平原達也（富山県立大）

相反法による頭部インパルス応答（hrir）の高速計測システムでは、低域ほど耳栓スピーカの出力音圧レベルが低いために、計測可能な下限周波数が約 500 Hz という問題があった。

デジタル信号処理では、耳栓スピーカに用いるイヤホン用ドライバー・ユニットの周波数特性の逆特性を用いることで問題は解決するように見えるが、アナログ信号では、DA 変換器には最大出力電圧の壁があるし、計測室と計測系には暗騒音レベルの壁がある。

そこで、イヤホン用ドライバー・ユニットの共振周波数以下の S/N を確保するよう周波数ごとの振幅値を最適化した、最大振幅最適化 TSP (MAO-TSP: Maximum Amplitude Optimized-TSP) 信号を考案した。

MAO-TSP 信号を用いると、従来の全帯域で振幅包絡が同じの Log-TSP 信号を用いた場合より、低域の S/N は最大 18 dB 改善され、計測可能帯域を 200 Hz まで広げることができた。

加算平均法や引き伸ばし時間を長くして低域の S/N を稼ぐ TSP 信号を用いる方法では計測時間が長くなるが、MAO-TSP 信号を用いればワンショットで hrir の高速計測が可能である。

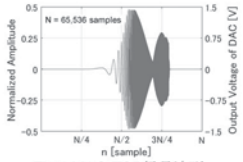


Fig.1 MAO-TSP信号波形

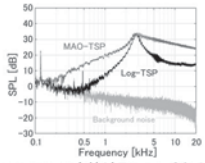


Fig.2 TSP応答波形のスペクトル

2-1-6

2-1-6　頭部伝達関数の測定距離を考慮した分散マイクロフォンアレイ信号からのバイノーラル信号生成

Generation of binaural signals from distributed microphones considering distance in HRTF measurements

☆飯島尚仁（東大），小山翔一（東大/JST さきがけ），猿渡洋（東大）

◆音場に分散配置したマイクロフォンを用いた収音信号から，HRTF を用いてその場に受聴者がいる場合のバイノーラル信号を再現する手法を提案する。

◆従来法は音場の平面波展開に基づくため，HRTF の測定距離が考慮されておらず近似誤差が大きかった。

◆本稿では，球面波展開に基づく高精度なバイノーラル信号生成手法を提案し，BEM に基づく詳細な評価を行った。

◆音像定位に重要な役割を持つ HRTF のノッチが従来法ではずれてしまうのに対し，提案法ではこれらを精度よく再現可能であった。

◆分散配置アレイを用いることで，現実的な規模のシステムで広範囲の収音領域におけるバイノーラル信号を合成できる可能性がある。

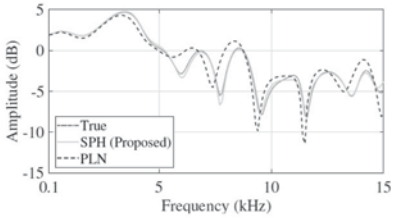


Fig.1 : Amplitude response of binaural signal of left ear for point source.

2-1-7

2-1-7　A subjective comparison of five sound field capturing methods

○ Sungyoung Kim (RIT, CET) and Otani Makoto (Kyoto Univ., Fac. Arch.)

◆The authors compared five sound field capturing methods, including a KEMAR dummy head microphone; Quad-Omni array (spaced five microphones); ViReal microphone (Higher Order Ambisonic); AMBEO (First Order Ambisonic); and H3VR (First Order Ambisonic) through a subjective evaluation.

◆Captured room impulse responses (RIRs) were binuaralized with an anechoic music recording and compared in terms of perceptual dissimilarity and overall preference.

◆All five auralized sound fields appear idiosyncratically distinct in the listeners' perceptual space (figure 1).

◆In addition, the global preference was relevant with the amount of spectral variance in the binuaralized stimuli (figure 2).

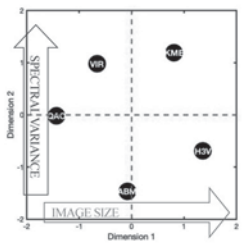


Fig.1: The perceptual space of five sound field capture methods. [KME: KEMAR, VIR: ViReal, QAO: Quad-Omni, H3V: H3VR, and AMB: AMBEO]

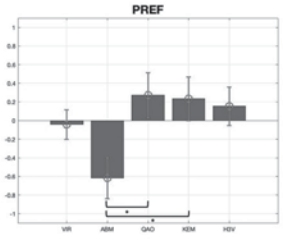


Fig.2: Mean preference ratings of 16 listeners to the five capture methods. The asterisk (*) symbol indicates a significant difference ($p < 0.05$) from a one-way ANOVA analysis

2-1-8

2-1-8　抵抗器における歪み測定方法の検討 —位相特性を考慮した評価—

Measurement of Distortion in Resistors, Evaluation Considering Phase Characteristics

◎宮岡洋平，黒澤実（東工大）

◆受動素子の代表である抵抗器に着目し，単一正弦波を用いた抵抗器における高調波歪みの測定を振幅・位相に着目し，行う.

◆抵抗器は 1kΩ, 0.25 W の炭素組成抵抗器，炭素皮膜抵抗器，金属皮膜抵抗器の 3 種を用いた.

◆440Hz の正弦波信号を用い，2 次高調波にあたる 880Hz から 10 次高調波にあたる 4400Hz までの各高調波周波数を観測した.

◆440Hz を基準周波数とし，基準周波数における位相とその高調波周波数における位相差を求めた結果を Fig. 1 に示す.

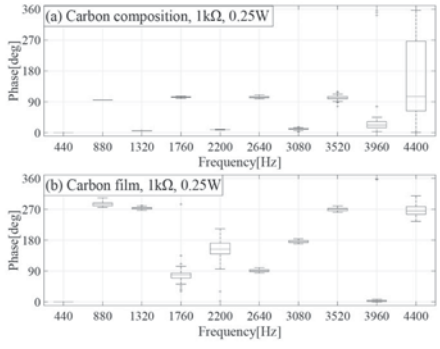


Fig.1 : Phase characteristics : (a) Carbon composition resistors 1kΩ, 0.25 W (b) Carbon film resistors 1kΩ, 0.25 W

2-1-9

2-1-9　窒化ガリウムトランジスタを用いた圧電性高分子発音体のためのスイッチングアンプの開発

Switching amplifier using a GaN field effect transistor for sound generator based on piezoelectric polymer

○杉本岳大，小野一穂（NHK）

◆家庭でマルチチャンネル音響を楽しめる小型軽量な音響機器の普及を目的に，高電圧出力可能な窒化ガリウムトランジスタ(GaN FET)を用いて圧電性高分子発音体用のスイッチングアンプを開発した。

◆図１に，開発した４チャンネル構成の GaN FET スイッチングアンプを示す。振動板の変位を周波数に反比例させる駆動特性を，カットオフ周波数 300 Hz の１次のローパスフィルタで実現した。

◆図２に，GaN FET スイッチングアンプと前報の Si FET スイッチングアンプの出力音圧レベルの差を示す。測定結果から，1 kHz 以上の周波数帯域では高域になるにつれて差が増大しており，Si FET スイッチングアンプにおける高域のロールオフ特性が，GaN FET を用いることで改善できることがわかる。

図１ 開発した GaN FET スイッチングアンプ

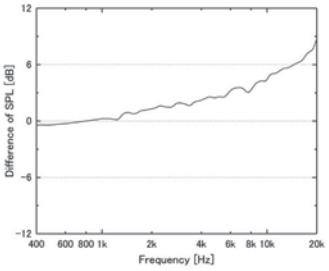


図２ GaN FET スイッチングアンプと Si FET スイッチングアンプの出力音圧レベルの差

2-1-10

2-1-10　音響的校正による導圧管周波数特性を考慮した音圧波形復元

Sound pressure waveform restoration considering frequency characteristics of pressure guide tube by acoustic calibration

○菊地勝浩，△野口雄平，△鈴木実（鉄道総研），△矢野健児（テス）

◆導圧管周波数特性で歪んだ音圧波形を復元する手法として，フーリエ級数展開法ではなく，ディジタル復元フィルタによる手法を開発した。

◆導圧管周波数特性の理論値として，Bergh らの式を採用した。導圧管の位相特性は非線形であり、そのような周波数特性によって歪んだ波形の復元フィルタによる復元はこれまで行われていない。

◆復元フィルタを設計する際に必要となるパラメータ値を音響的校正によって求めた（図１）。フィルタは FIR フィルタとして設計した。

◆開発した波形復元法の有効性を確認するため、復元処理前後の音圧波形とレファレンス音圧波形を比較した。比較の結果、復元後の音圧波形は、導圧管無し音圧波形と良く一致し、開発した波形復元法の有効性が確認された（図２）。

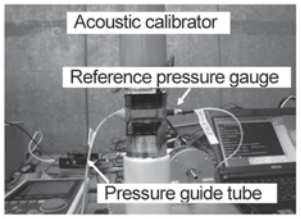


Fig. 1 Acoustic calibrator

Fig. 2 Comparison of waveforms

2-1-11

2-1-11　暗騒音下での音響計測に基づく機械からの音・振動に対する状態推定

State estimation for sound and vibration generated from machines based on sound mesurement under existence of background noise

○折本寿子，生田顯 (県立広島大)

さまざまな産業システムにおいて機械を停止させることなく，機械の状態を推定する異常診断は，近年ますますニーズが高まっている．機械の状態を推定・診断するためには，対象とする機械から発生する音および振動を正確に測定する必要がある．しかし，精密機械などでは，振動計測のためのピックアップを直接機器に設置することが困難な場合がしばしば生じる．

本研究では，非接触で計測できる騒音計により測定したデータから，暗騒音を除去した，回転機器から発生する音と振動の変動波形を推定する手法について確率のベイズ定理を用いることにより提案する．

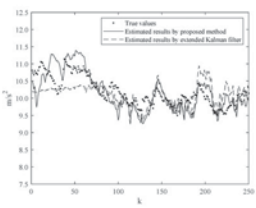


Fig. 1 Estimated results of vibration

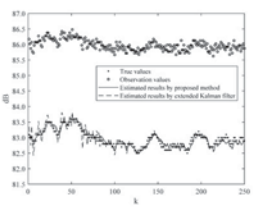


Fig. 2 Estimated results of sound

2-1-12

2-1-12　全方位画像と多チャネル音響信号を用いた自己教師あり深層音源定位

Self-supervised neural source localization based on omnidirectional image and multichannel acoustic signal

◎升山義紀（早大理工/産総研），坂東宜昭，大西正輝（産総研），矢田部浩平，及川靖広（早大理工）

◆ 目的: 画像内の音源領域を推定する DNN の自己教師あり学習

◆ 従来手法: 画像 + **単チャネル音響信号**
- 画像と音の共起関係に基づいて DNN を学習
- **音源が複数存在し，それらが視覚的に区別できないデータからの学習は困難**

◆ 提案手法: 全方位画像 + **多チャネル音響信号**
- 多チャネル音響信号を用いることで，音の空間情報が利用可能
- 音の空間情報が増えたことで，**混合音の各音源を画像内で区別しながら学習**
- 多チャネル音響信号: 混合複素ガウスモデルに基づく空間モデルで定式化
- 全方位画像: 画像中の音源位置を推論する音源定位 DNN に利用
- 「空間モデルの事後分布」と「音源定位 DNN の出力（近似事後分布）」の KL ダイバージェンス最小化（変分下限の最大化）で学習

◆ 実験結果:
- マイク４本を用いた数値実験において MUSIC 法以上の定位性能を確認

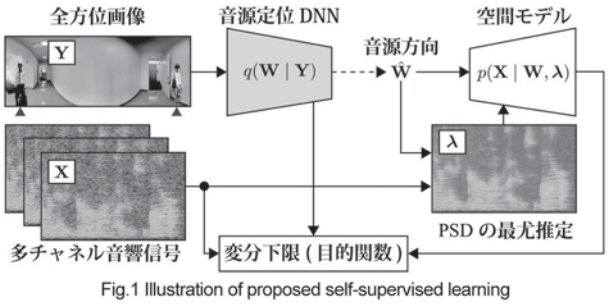


Fig.1 Illustration of proposed self-supervised learning

2-1-13

2-1-13　瞬時周波数計算のための エネルギー局在微分窓

Maximally energy-concentrated differential window function for instantaneous frequency computation

☆草野翼，矢田部浩平，及川靖広（早大理工）

◆近年，音響信号処理において短時間 Fourier 変換（STFT）の瞬時周波数が利用されている．

◆STFT の瞬時周波数推定では窓関数とその微分窓を用い，推定結果はそれらの特性に大きく影響を受ける．

◆本研究では，微分窓の周波数応答のエネルギー局在化に基づいて設計された，瞬時周波数の計算に適した窓関数を提案する．

◆ 微分窓のエネルギーを局在化することにより(Fig. 1)，瞬時周波数計算に対するサイドローブの影響を低減できることが確認された．

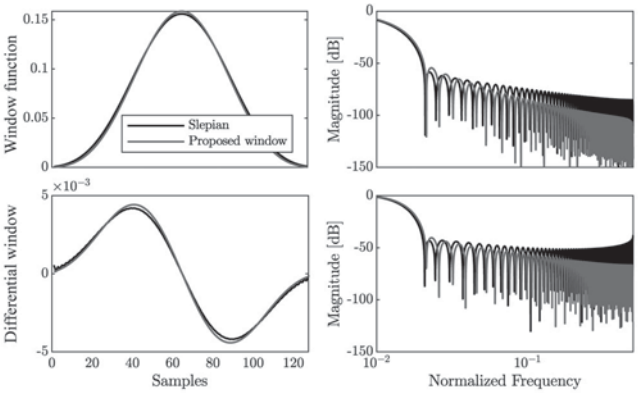


Fig. 1　Results of designed windows. The left and right columns show window shapes and their frequency responses, respectively.

2-1-14

2-1-14　補助関数法を用いた相互相関の最大化によるサンプリング周波数ミスマッチ推定

Sampling frequency mismatch estimation by maximizing the cross-correlation using the auxiliary function technique

◎山岡洸瑛，シャイブラーロビン，小野順貴，若林佑幸（首都大）

◆本研究では，２チャネルの非同期信号を対象にした相互相関最大化に基づくサンプリング周波数ミスマッチ推定法を提案する．

◆これまでに我々は補助関数法に基づき，非凸関数である相互相関を解析的に最大化する手法を提案した．

- 提案法では，相互相関が cos 関数の和の形で表されることを利用
- cos 関数は，以下の不等式を用いて下から抑えることができることから，相互相関に対する 2 次の補助関数を導出した．

$$\cos\theta \geq -\frac{1}{2}\frac{\sin\theta_0}{\theta_0}\theta^2 + \left(\cos\theta_0 + \frac{1}{2}\theta_0\sin\theta_0\right)$$

◆これを用いて Double-cross-correlation processor と呼ばれるサンプリング周波数ミスマッチ推定法の性能を向上させた（Fig. 1 参照）．

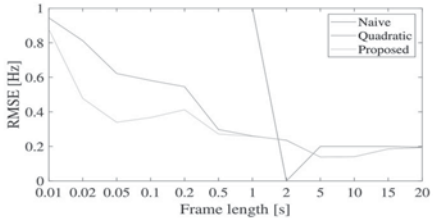


Fig. 1: Estimation error for sampling frequency mismatch over frame-length. Naive: using the discrete peak of the cross-correlation, Quadratic: parabolic interpolation (conventional method), Proposed: proposed method

2-1-15

2-1-15　音光変換デバイス「ブリンキー」を用いた DNN に基づく複数音源定位

Deep neural network based multiple source localization using blinky sound-to-light conversion sensor devices

☆堀池 大樹，石井 奏人，シャイブラー ロビン，
若林 佑幸，小野 順貴（首都大）

◆マイクロホンと LED により音光変換を行うブリンキー(Fig. 1A)を分散配置し，ビデオカメラで撮影することによって(Fig. 1B)，広範囲の音強度情報を簡単に取得できる新しい音響信号処理の枠組みを研究している。

◆これまでに提案したブリンキーの音強度分離を応用し，複数音源に対する位置推定の枠組み(Fig. 2)を提案する。具体的には，非負値行列因子分解(NMF)により推定した各音源の音強度を音響特徴量とした深層ニューラルネットワーク(DNN)に基づき音源位置を推定する。

◆計算機シミュレーションにより性能評価実験を行った。実験結果(Table 1)から，提案手法の推定性能が高いことが確認できる。

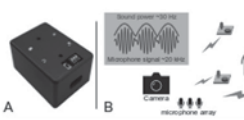


Fig.1: The sound-to-light conversion architecture

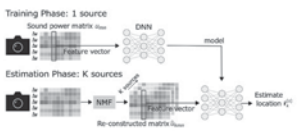


Fig.2: The system flow of proposed source localization

Table 1: The statistical data of localization error of Baseline, fully-connected NN, and ResNet

	SNR=60 dB	Baseline	全結合 NN	ResNet	SNR=10 dB	Baseline	全結合 NN	ResNet
Mean [m]		1.111	0.553	**0.590**		1.203	1.494	**0.602**
STD [m]		1.003	0.604	**0.735**		1.061	0.770	**0.717**
p_{50} [m]		0.556	0.307	**0.236**		0.584	1.473	**0.263**
p_{90} [m]		2.761	1.460	**1.828**		2.815	2.331	**1.716**

2-1-16

2-1-16　周波数領域 velvet noise を用いた音響インパルス応答計測用試験信号の拡張について

Extension of test signals for acoustic impulse response measurements based on the Frequency domain variant of Velvet noise

○河原英紀（和歌山大），榊原健一（北海道医療大），水町光徳（九工大）

◆時間伸長パルス(TSP: Time Stretched Pulse)の新しいメンバーとなる周波数領域ベルベットノイズ(FVN: Frequency domain variant of Velvet Noise)を用いて音響インパルス応答を測定することができる。

◆FVN は、高い設計の自由度を有している。そのため、個別の FVN を二値の直交系列を用いて時間軸上に配置することにより、複数の直交する計測用の信号を作ることができる。この信号を用いて、複数の音響経路の同時計測や、線形時不変成分と非線形／時変成分による影響を同時計測することができる。

◆FVN を応用した実時間で対話的な音響計測ツールを開発し、公開している。

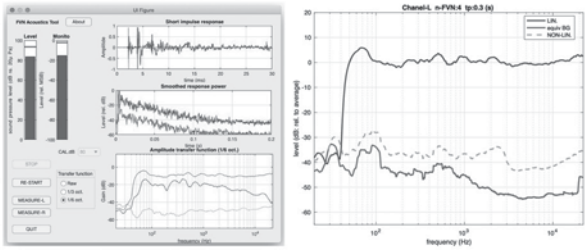


Fig.1: Interactive and real-time tool for acoustic impulse response measurement. (Left) GUI, (Right) example of simultaneous measurement

2-2-1

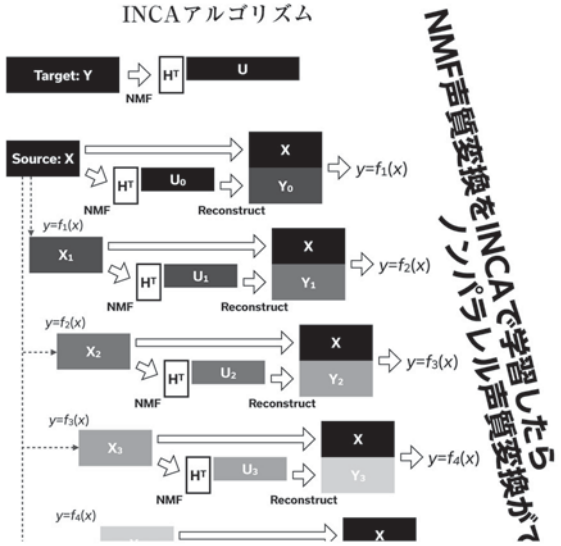
2-2-1　非負値行列因子分解による声質変換におけるINCAアルゴリズムをもとにした基底のノンパラレル学習法

Nonparallel training of exemplars in voice conversion system based on non-negative matrix factorization

☆ 須田仁志，小谷岳，齋藤大輔（東大）

キーワード：ノンパラレル声質変換　非負値行列因子分解

INCAアルゴリズム



2-2-2

2-2-2　Attention に基づく音声変換のためのアラインメント予測モデルの検討

A study of alignment prediction model for voice conversion based on attention mechanism

☆芹川 武尊(東大・工学部)，郡山 知樹，猿渡 洋（東大院・情報理工）

◆従来の Sequence-To-Sequence 音声変換は Attention 機構を利用
- ➢ Attention 機構は学習が高コストで，大量のパラレルデータが必要

◆Attention 機構の持つ時間整合の機能と特徴量変換の機能を分離し，時間伸縮曲線とカーネル回帰を組み合わせた方法を提案
- ➢ 10 分ほどのパラレルデータでの Seq-To-Seq 音声変換を実現

◆実験では以下を確認
- ➢ Encoder, Decoder には CNN ベースより Tacotron ベースのモデル構造が適している
- ➢ DTW により事前に入力・目標系列を伸長する方法と比較したとき，時間伸縮曲線を用いた方法は特に同性間変換で有効である

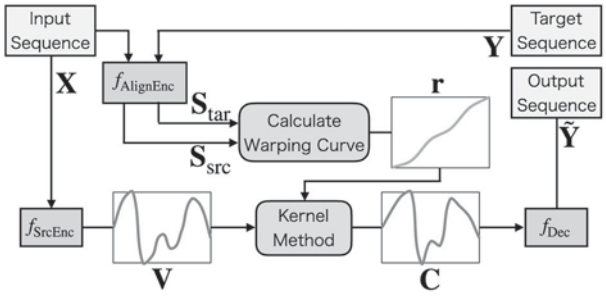


Fig.1：Overview of the proposed method during learning phase.

2-2-3

2-2-3　非負値時空間分解と深層学習を用いた英語発話リズム変換法

An English speech rhythm conversion method using non-negative temporal decomposition and deep learning

○廣谷定男(NTT CS 研)，田口史朗(九州大)

◆これまでに、非負値時空間分解法（NTD）とコードブックマッピングを用いて、日本語母語話者が発話した英語音声の発話リズムを英語母語話者のそれに変換する方法を提案してきたが、変換精度に課題があった。

◆本研究では、日本語母語話者および英語母語話者が発話した大量の英語音声から NTD を用いて抽出した時間関数を深層学習により変換することで精度の向上を図った。

◆実験の結果、従来法よりも発話リズム変換より、音素長がネイティブに近づくことが分かった。

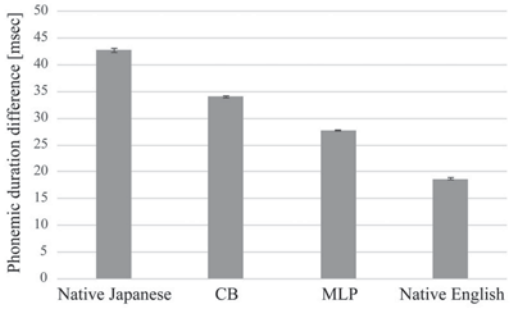


Fig.1：Phoneme duration difference by speaking rhythm conversion

2-2-4

2-2-4　声質制御のための VQ-VAE を用いた連続的な話者性の獲得

Automatic aquisition of continuous speaker identity using VQ-VAE for voicie quality control

☆竹内太法，立蔵洋介（静岡大・創造科学技術大学院）

ノンパラレルデータを用いた VAE に基づく声質制御を目的として，柔軟な話者ラベルを獲得するための声質変換モデルの概略を述べた．

連続的な話者ラベルを学習によって自動獲得することによって，自在に声質を制御することを可能にし，かつ，非話者情報を VQ-VAE に基づく表現能力の高いモデルで表現することで変換音声を高品質に保つことが期待される．

今後の課題は，提案モデルを評価することである．

2-2-5

2-2-5　差分スペクトル法に基づく広帯域声質変換のためのサブバンドリフタ学習

Sub-band lifter-training method for full-band voice conversion using spectral differentials

☆佐伯 高明，齋藤 佑樹，高道 慎之介，猿渡 洋（東大院・情報理工）

◆差分スペクトル法による広帯域声質変換
- 広帯域化により計算量が増大
- 高域の変動によるモデル化性能の低下

◆**サブバンドリフタ学習**を用いた差分スペクトル法を提案！
- サブバンド処理 & フィルタ打ち切りによるリフタ学習
- 実験的評価により，計算量の削減と有意な品質向上を確認

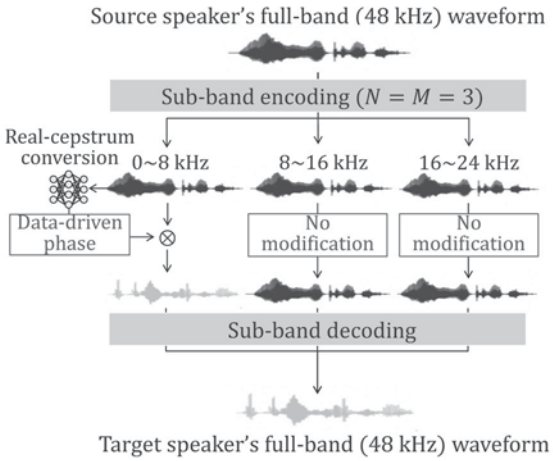


Fig.1: Procedure of proposed sub-band lifter-training method.

2-3-1

2-3-1　骨導音の同時、非同時マスキング効果の計測 － 片耳聴取と両耳聴取の比較 －

Simultaneous and non-simultaneous masking effects in bone conduction listening: monaural and dichotic

○碓井綾，岩城護（新潟大院・自然研）

◆骨導音聴取時に生じる同時、順向性、逆向性マスキング効果を片耳聴取（モノラル）と両耳聴取（ダイコティック）に対して調査した.

◆ノッチ雑音マスキング法を用いて，骨導音聴取時と気導音聴取時の純音に対するマスキング量を計測した．純音の周波数は 500，1000，2000，4000 Hz の４種類とした.

◆骨導音聴取時でもマスキング効果が確認された.

◆骨導音聴取時の方がマスキング量が少ない場合が多かった.

◆片耳聴取の方が両耳聴取よりもマスキング量が多かった.

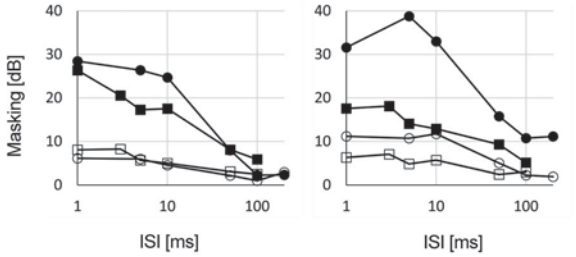


Fig.1：Masking characteristics at 1000 Hz
(●: forward masking & monaural, ■: backward masking & monaural,
○: forward masking & dichotic, □: backward masking & dichotic)

2-3-2

2-3-2 Effects of placements on bone-conducted sound transmission by measurements of hearing thresholds, ECSP and acceleration on the head

◎ Xiuyuan Qin, Sho Otsuka, Seiji Nakagawa (Chiba Univ.)

- Bone conduction (BC) is the conduction of sound to the inner ear through skin and skull, and it is much influenced by placement. Researchers have studied at the position "mastoid" for many years, but a few of them focused at the position "condyle" which is the position usually used by commercial BC headphones (Fig.1).
- Hearing thresholds, ear-canal sound pressure (ECSP) and vibration on the head produced by BC sound at the condyle and the mastoid were compared by using the standard vibrator radioear B-71. The difference between conditions with / without earplugs was also discussed.
- The threshold on the condyle was lower than it was on the mastoid (Fig. 2). As for the ECSP, the results on the condyle was higher than they were on the mastoid in both open and occluded ear canals (Fig. 3). The higher vibration at the ear canal may be the cause as shown in Fig.4.

Fig. 1 Positions of the mastoid and the condyle.

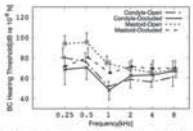


Fig. 2 BC hearing thresholds under four conditions.

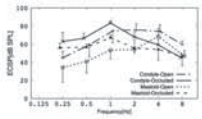


Fig. 3 Ear-canal sound pressure under four conditions.

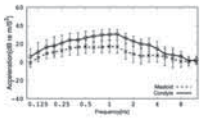


Fig.4. Vibration produced at the ear canal at the two positions.

2-3-3

2-3-3　上下前歯の歯導音純音聴力レベル測定

Measurement of air-, bone- and teeth- conducted threshold levels by pure tone audiometry

○山田朋美，△前薗葉月，△野﨑一徳，△林美加子，松井孝典，
桑野園子（大阪大学），高田正幸（九州大学）

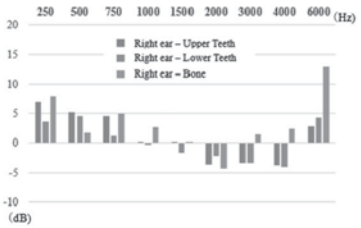
◆食べものを食べる音、歯ぎしりの音、治療中に歯を削る音等を、私たちは自分の耳からだけではなく歯からも知覚している。

◆歯と上下顎骨（歯槽骨）の間には歯根膜腔という隙間がある。下顎と頭部の間には関節腔という隙があり、骨とは結合していない。

◆歯から伝わる音に着目した研究を進める中で、歯から上下顎を伝わり知覚される音を「骨導音」と区別し「歯導音」と定義した。

◆本研究では、オージオメータを用いて気導、骨導および歯導の純音聴力レベル測定を行った。

◆気導聴力レベル（右耳）から歯導および骨導聴力レベルを差し引いた「聴力レベル差（平均値）」を下図に示す。



Average hearing level differences
between air-conduction and, teeth- and bone-conduction.

2-3-4

2-3-4　超指向性音によるエゾシカと車両の衝突回避に関する一考察

A Study of Yezo Sika Deer-Vehicle Collision Prevention by Hyper Directional Sound

○松﨑博季，△真田博文，△和田直史（北科大・工）

西興部村鹿牧場で飼育されているエゾシカに対して、音声信号で周波数変調あるいは振幅変調された 40 kHz の超音波を超音波発振子から発した超指向性音を聞かせた時のエゾシカの反応を確認する実験を行った。変調にハードウェアおよびソフトウェアを用いる二つの超指向性スピーカーを用いた。音声信号には鹿の警戒性、枯れ草を踏む音、3.5 kHz の正弦波の断続音、100 Hz から 5 kHz に変化する正弦波の断続スイープ音、ヒグマの鳴き声、雷鳴を用いた。

正弦波以外の超指向性音に対してエゾシカが反応すること、また十分な音量を得るためには超音波発振子が 100 個以上必要なことが確認できた。

2-3-5

2-3-5　オクターブバンドスペクトログラムの深層学習を用いた自然環境音からの快適音創生

Creation of comfortable sounds from various natural sounds by using deep learning of their octave band spectrograms

○星野博之，△白戸宏季（愛知工業大学）

◆目的　運転中などの人が活動する各種作業環境において、人の状態を適切かつ快適に保つための音刺激を実際の自然環境音から探索、創成することを目的としている。今回は、自然環境音とその快適さ主観評価データを用いて深層学習を行い、快適さ推定の可能性を検討した。

◆方法　自然環境音 130 音のオクターブバンドスペクトログラムと主観評価値を学習データとし、130 音中で主観評価値が高かった上位 20 音の 1kHz, 2kHz, 4kHz 帯域抽出音（計 60 音）をテストデータとして、畳み込みニューラルネットワークを用いて深層学習を行った。

◆結果　テストデータである帯域抽出音 60 音の快適さをほぼ十分な精度で推定できた(Fig. 1)。Fig. 2 に「初秋の虫の音」のオクターブバンド帯域抽出音の快適さ推定結果を示す。聴感と対応した結果であることから、本手法は快適音を生成するために有効となる可能性がある。

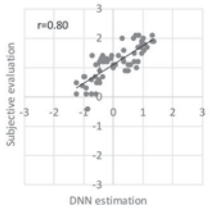


Fig.1 Relation between DNN estimation and subjective evaluation

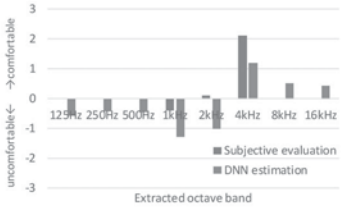


Fig.2 DNN estimation results of comfortableness of band limited natural sounds

2-3-6

2-3-6　ポテトチップスの食感に寄与する音響的要因に関する研究

Study on acoustic factors that contributes to the texture of potato chips

☆中山隆介，上田麻理（神奈川工科大学），三浦貴大（産総研）

＜実験方法＞ ポテトチップスなどのスナック菓子を題材に、音響的特徴と食感評価の関連性を調査した。50 名の参加者（男性 43 名、女性 7 名）に 10 種類のスナック菓子を食べてもらい、その際の咀嚼音を咽頭マイクと気導音マイクで測定した。その後、5 段階評価の食感評価アンケート（食感に関する擬音語、音響印象、総合評価）を答えてもらった。

＜結果＞ 周波数帯域が約 1～10 kHz の咀嚼音は、食感に関連しうる可能性が示された。また、弱い・軽いという印象を持ったスナック菓子は、「パリッ・パリパリ」といった半濁音を含む擬音語の印象が強く、硬い・強い・重いという印象を持ったスナック菓子は「ザクザク・バリバリ・ボリボリ」といった濁音を含む擬音語の印象が強かった。

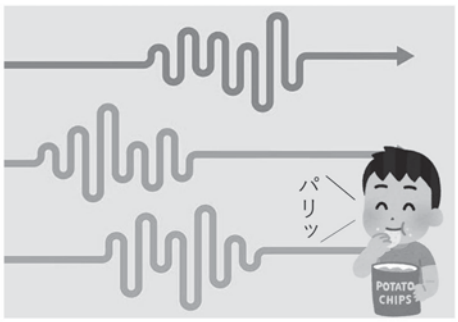


Fig.1 : Conceptual image of the experiment.

2-3-7

2-3-7　聴覚における内有毛細胞励起に至る音波検出モデル

A model of sound wave detection to lead to inner hair cell excitation in the hearing

○石渡信吾 (横浜国大), △田子友生, 岩原拓生 (横浜国大院)

ヒトの内耳において蓋膜の裏側にあるヘンゼン線条が内有毛細胞の感覚毛を外側に傾かせることが内有毛細胞の励起機構であることを起点に、外有毛細胞から基底膜振動までを逆にたどり、音波の新たな検出モデルを提案する。外有毛細胞の能動性に基づく応答を単安定系に帰着させた。外有毛細胞－内有毛細胞－蝸牛神経から成る回路モデルの結果を示す。1,3 段目が外有毛細胞と蝸牛神経を表す単安定回路、2 段目が内有毛細胞の積分回路である。各回路の時定数を適宜調整すると 1.9kHz で最適値な応答を示すことがわかる。音波による基底膜振動を模して正弦波を入力したときの 2 段目の積分波形の平均値を Fig.1 に示す。3 段目の回路の閾値を 1V に設定した。1.7～2.0kHz で蝸牛神経相当の単安定回路が応答する。

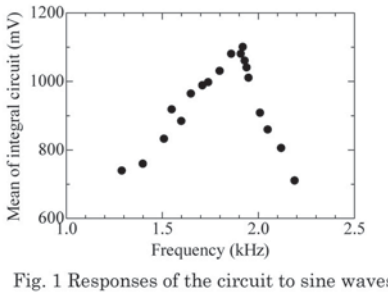


Fig. 1 Responses of the circuit to sine waves

2-4-1

2-4-1　音声認識における Encoder–Decoder モデルの注意重みを用いた CTC モデルへの知識蒸留

Distilling knowledge of attention-based encoder-decoder into CTC-based automatic speech recognition systems

◎森谷崇史，佐藤宏，田中智大，芦原孝典，増村亮，篠原雄介
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，

◆Connectionist Temporal Classification（CTC）は従来の音声認識システムの音響モデルおよび end-to-end 音声認識（E2E-ASR）モデルの両方の動作を行うことが可能であり，実用化の面で注目を浴びている
◆CTC モデルの認識性能を改善するために，従来の知識蒸留はコンパクトな CTC モデルの学習に応用可能だが，大型の CTC モデルへは応用が困難であった
◆本研究では，Encoder-Decoder モデル（S2S）の出力である注意重みを用いた大型の CTC モデルへの知識蒸留（CTC+DAW）を提案する
◆日本語話し言葉コーパスを用いた音声認識の実験において，提案手法の知識蒸留（CTC+DAW）により学習したモデルは CTC 損失のみで学習したモデルよりも認識性能が改善し，相対誤り改善率は平均で 10.2%を達成した．

Table 1 : Character error rate (CER) of each E2E-ASR system.

System (E2E)	CER [%] E1	E2	E3
S2S	10.1	8.3	9.7
CTC	11.8	9.8	12.0
CTC+DAW	**10.6**	**8.8**	**10.7**

2-4-2

2-4-2　Deep Mutual Learning に基づく End-to-End 音声認識の検討

Deep Mutual Learning for Automatic Speech Recognition

○増村 亮，庵 愛，高島 瑛彦，田中 智大，芦原 孝典，大庭 隆伸
（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）

◆本稿では，Deep Mutual Learning の End-to-End 音声認識への適用を検討した．Deep Mutual Learning では複数のモデルを協調的に学習することで，全体最適解に到達することを助けることが期待できる．評価結果から，Transformer に基づく End-to-End 音声認識において，これまでの学習テクニックと併用する場合においても有効であることが示された．

表 1：評価結果．

Label smoothing	Scheduled sampling	SpecAugment	Deep Mutual Learning	Test 1	Test 2	Test 3
-	-	-	-	8.83	6.49	7.19
√	-	-	-	8.41	6.23	6.74
-	√	-	-	8.59	6.31	6.30
-	-	√	-	7.48	5.59	5.86
√	√	√	-	7.24	5.13	5.40
-	-	-	√	8.19	5.78	6.37
√	-	-	√	8.05	5.67	6.30
-	√	-	√	7.90	5.57	5.62
-	-	√	√	7.02	4.92	5.28
√	√	√	√	**6.87**	**4.73**	**5.02**

2-4-3

2-4-3　半教師あり End-to-End 音声認識のための系列単位 Consistency Training の検討

Consistency Training for Semi-supervised Speech Recognition

○増村 亮，庵 愛，高島 瑛彦，森谷 崇史，安藤 厚志，篠原 雄介
（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）

◆本稿では,系列単位 Consistency Training に基づく半教師あり End-to-End 音声認識を提案した.提案手法の強みは,ラベルなし音声データを活用して,End-to-End 音声認識モデルを微小な変動に頑健にすることができる点であり,SpecAugment 等の強力な微小変動関数を有効活用できる．Transformer に基づく End-to-End 音声認識のモデル化を用いた評価実験において,ラベルなし音声データを活用した提案手法により,大幅な性能改善が可能であることを示した．

表 1：評価結果．

	Test 1	Test 2
Superivsed	25.88	21.88
Supervised+SpecAug	21.23	17.23
Semi-supervised (1-best)	18.35	15.03
Semi-supervised (5-best)	**17.86**	**14.62**
Ideal semi-supervised	15.60	10.13

2-4-4

2-4-4　CNN フロントエンドによる高速な End-to-End 連続 DP マッチングの実現

Efficient end-to-end continuous DP matching based on CNN frontend

☆田中智宏，篠崎隆宏（東工大）

◆2DRNN ベースの End-to-End 連続 DP マッチングはキーワードオープンな条件で埋め込み型モデルより高精度なキーワード検出が可能
◆しかし埋め込み型モデルと比較して大きな計算量が課題
◆CNN フロントエンドを組み合わせ，CNN の縮小倍率と同じ倍率で高速化する手法を提案
◆同時に並列計算のためのハードウェアコストも同じ倍率で削減可能
◆Google Speech Commands Dataset の音声コマンドと CHiME-3 のノイズデータを用いて検出精度と処理速度をそれぞれ比較
◆実験ではストライド 3 の畳み込み層 2 層を用いて検出精度を維持しながら追加の計算コストを含め約 9 倍の高速化を確認

Table.1 : Computational time of keyword detector using three minutes of data.

Model	Eval. Time (sec)
Emb	18.07
Emb+CNN	2.06
2DRNN-DP	48.75
2DRNN-DP+CNN	5.38

Table.2 : Word-level F-score of keyword detector. “Close” and “Open” means keyword close and open condition respectively.

Model	5dB Close	5dB Open	20dB Close	20dB Open
Emb	0.308	0.212	0.548	0.320
Emb+CNN	0.357	0.222	0.485	0.296
2DRNN-DP	0.388	**0.396**	0.537	**0.517**
2DRNN-DP+CNN	0.340	**0.367**	0.525	**0.518**

2-4-5

二重相続進化戦略による音声認識システムの最適化

Optimization of speech recognition system by dual inheritance evolution strategy

☆日野健人，△木村友祐，△Dong Yue，篠崎隆宏（東工大）

◆従来の進化戦略法（CMA-ES）による音声認識システムの最適化では個体の評価スコアのみが次世代へ伝えられ、それ以外の個体の学習結果はすべて捨てられる

◆生物学における二重相続理論をもとに進化戦略に知識蒸留を組み込む二重相続進化戦略(Dual Inheritance CMA-ES:DI-ES) 法を提案

◆遺伝子とともに個体が学習で得た結果を文化として次世代に継承

◆提案法は音声認識システム以外にもニューラルネットを用いた幅広いシステムに応用可能

◆音声認識ツールキット(ESPnet)の an4 レシピを対象にした進化実験において，個体数が大きいときにモデルサイズと精度が改善

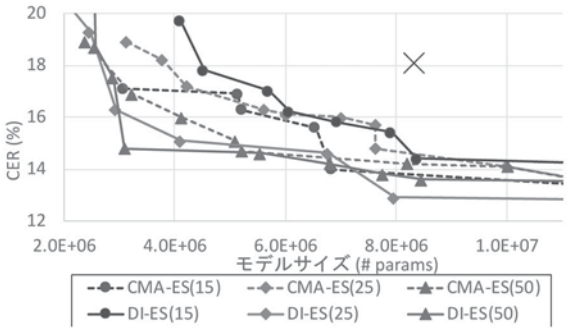


Fig.1：一定サイズ以下での最小エラー率の推移(数字は個体数を示す)
an4レシピにおけるデフォルト個体をx記号で表現

2-4-6

2-4-6　音響モデル構築用音声データのプライバシー保護

Privacy Preservation of Speech Data for Constructing Acoustic Models

〇太刀岡 勇気（デンソーアイティーラボラトリ）

◆音響モデルの学習の際にはたとえ少量でも、ドメイン内データが有効

◆ドメイン内データのプライバシー保護は、攻撃者に話者が何を話したか、話者が誰かということを知られないようにする必要がある

◆提案のプライバシーに配慮した音響モデルの学習法では、学習データを分割し、ランダムに結合することで発話内容の秘匿化を行う

◆さらに話者クラスタリングにより、話者の k-匿名化を達成する

◆CSJ の学会講演をドメイン内データ、一般講演をドメイン外データとした音声認識実験により、提案法の有効性を検証した

◆図に示すように、全話者を使って通常通り学習した場合 11.44%の単語誤り率が、全話者を秘匿した場合 12.03%、1/5 にあたる 200 名を秘匿化した場合に 11.70%(0.26%の認識率低下)となった

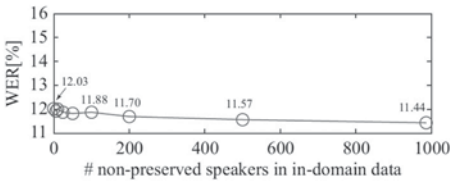


Fig.：WER[%] on the CSJ testset when partial speakers were preserved

2-4-7

2-4-7　双方向 LSTM 単語入れ換えモデルを用いた教師なしテキストスタイル変換に基づく言語モデル学習データの増強

Unsupervised text style transfer for language model data augmentation using BLSTM-based word replacing model

〇小川厚徳，俵直弘，苅田茂樹，マーク・デルクロア
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

◆ニューラルネットワークの導入により音声認識精度は飛躍的に向上したが，音声認識はドメイン依存性の高い技術であるため，自然発話やマイナー言語などの少資源ドメインでは，その精度は依然として満足できるものではない．

◆本研究では，自然言語処理分野で近年盛んに研究されている**教師なし(ペアデータを用いない)テキストスタイル変換に基づく言語モデル学習データの増強**手法を提案する．

◆テキストスタイル変換は**双方向 LSTM に基づく単語入れ換えモデル**を用いて行い，加えて変換したテキストの文法的な正さを保証するための後処理を行う．

◆CSJ 講義音声コーパスを変換元ドメインデータ，NTT ミーティング(複数人自由会話) 音声コーパスを変換先ドメインデータとして用いた音声実験により，提案手法の有効性を確認した．

Table 1. Experimental results obtained with the nine trigram language models. GenCSJxN stands for generated CSJ data that has N times larger data size.

		Development data			Evaluation data		
No.	Model	PPL	OOV	WER	PPL	OOV	WER
1.	NTT	**44.9**	1.42	21.8	44.5	1.34	21.1
2.	CSJ	128.6	2.21	22.6	122.6	2.18	21.8
3.	GenCSJx1	79.5	0.73	21.6	78.2	0.67	21.0
4.	GenCSJx10	75.4	0.56	21.2	73.9	0.51	20.6
5.	GenCSJx20	75.1	0.54	21.0	73.3	0.48	20.6
6.	GenCSJx50	74.6	0.51	20.9	72.8	0.46	20.7
7.	GenCSJx100	74.3	0.50	21.0	72.4	0.46	20.6
8.	NTT+CSJ	46.5	**0.40**	20.3	45.3	**0.40**	19.3
9.	NTT+CSJ+GenCSJx100	45.2	**0.40**	**20.0**	**44.1**	**0.40**	**19.1**

2-4-8

2-4-8　ニューラル機械翻訳による書き言葉から話し言葉へのテキスト変換を用いた言語モデル構築

Converting Written Language into Spoken Language Using Neural Machine Translation for Language Modeling

◎安藤慎太郎†，鈴木雅之‡，倉田岳人‡，伊東伸泰‡，峯松信明†
（†東大・工，‡日本IBM）

◆目的：音声認識における言語モデルの話し言葉適応

◆公共放送の番組の話し言葉音声からパラレルコーパスを作成

◇A. 音声の忠実な書き起こし → 話し言葉テキスト

◇B. Aの整形後テキスト → 書き言葉テキスト

◆収集したコーパスを用いて、書き言葉→話し言葉テキストの変換モデル（Transformer）を構築

◆変換後テキストの自然性・多様性を確保するため、推論時に top-k サンプリングを採用

◆変換後のテキストを言語モデル構築に用い、音声認識で評価

◇1. 言語モデルの単独ドメインの話し言葉適応

◇2. 汎用話し言葉言語モデル構築

◆両音声認識タスクにおいて、認識精度の向上が確認された

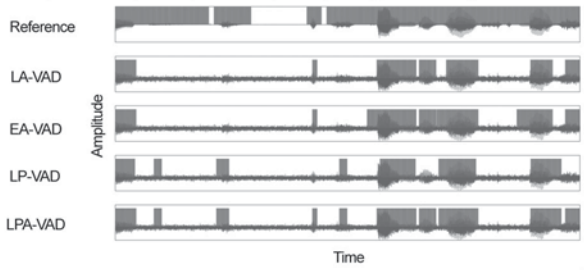
◇top-kサンプリングを用いた変換（探索）が本タスクにおいても有効であることを示した

2-4-9

2-4-9　A phase-based voice activity detection using statistical likelihood ratio of the derivative of instantaneous frequency

☆Nguyen Binh Thien (Ritsumeikan Univ.), Yukoh Wakabayashi (Tokyo Metropolitan Univ./Ritsumeikan Univ.), Takahiro Fukumori and Takanobu Nishiura (Ritsumeikan Univ.)

Conventional voice activity detection (VAD) algorithms primarily rely on the amplitude-based features. This paper introduces a new phase-based approach for VAD using the statistical likelihood ratio of the derivative of instantaneous frequency (LP-VAD). The performance of the proposed method is evaluated in comparison with a conventional amplitude-based VAD (LA-VAD) and our previous phase-based VAD using Euclidean distance (EA-VAD). The experimental results show that the proposed method achieves better performance than the VAD using Euclidean distance in most conditions and even higher accuracy than the amplitude-based VAD in the case of white noise at low signal-to-noise ratio. We also demonstrate the advantage of combining the amplitude and phase (LPA-VAD) for improving the overall performance.



An example of VAD algorithms on a speech signal impared by white noise at 5 dB SNR [(1)].

[(1)] The orange parts denote the voice activity detected by the algorithms.

2-4-10

2-4-10　クロスモーダル知識蒸留に基づく Lip reading のための教師なしドメイン適応

Unsupervised Domain Adaptation for Lip reading Based on Cross-modal Knowledge Distillation

◎高島悠樹（神戸大），相原龍（三菱電機），高島遼一，滝口哲也，有木康雄（神戸大），村山修（三菱電機）

◆音声認識モデルからリップリーディングモデルへのクロスモーダル知識蒸留に基づく教師なしドメイン適応を提案する．

◆従来の深層ドメイン適応では，各ドメインのデータが同一のクラス集合に属することを仮定しており，未知クラス (out-of-vocabulary)のデータを適応データとして使用することができない．

◆モデル構造を適切に設計することで，中間層がサブクラスを表現し，これを用いて蒸留を行うことで，教師なしドメイン適応が可能となる．

◆さらに，サブクラス表現 (sub-word)はクラスに依存しないため，本手法は未知クラスに属するデータを用いた適応を可能とする．

◆孤立単語認識実験により，提案手法は適応データのクラスが既知であっても未知であっても優れた適応性能を持つことを示す．

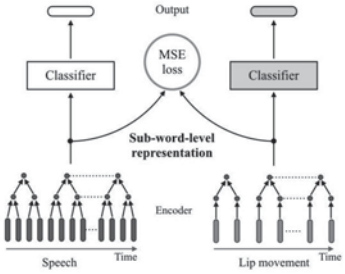


Fig.1：Overview of our proposed method.

2-5-1

2-5-1　声道形状逆推定に基づく調音運動特徴の解析

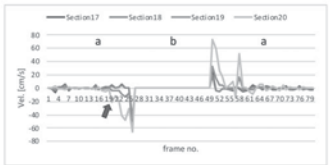
Analysis of articulatory behavior based on inverse estimation of the vocal tract shape

○緒方公一，山本健人，△伊東麻沙美（熊本大）

◆声道音響管マッピングインタフェースは母音あるいは連続母音の取り扱いを意図したものであるが，子音を含む音声についても，母音から子音に至る過渡的な部分での声道の振る舞いの逆推定に有効であり，音素環境の違いによる調音点の移動を示唆する音響管の振る舞いの把握が可能なことなどを，これまでに報告している。

◆本稿では，声道の大局的な振る舞いを捉えることのできる本逆推定手法が，声道の部位による運動タイミングの違いなどの局所的な調音運動特徴を捉えることができているかについて検討する。

◆Fig. 1 に，発話の進行に伴う，声道音響管の擬似的な速度の変化の例を示す。正門側から口唇側に向けて番号付けした音響管の内の第17,18,19，および 20 区間の音響管について速度を示している。口唇部に相当する Section 20 において，他の区間（下あごを反映）よりも早いタイミングで速度の値が大きくなり，口唇部が先行して口唇閉鎖の動作を行なっていることが推察される。



/aba/

Fig. 1 Pseudo velocity patterns of the partial vocal tract corresponding to sections 17, 18, 19, and 20 (lips) for /aba/.

2-5-2

2-5-2　壁振動を考慮した時間領域差分法による１次元の声道音響モデル

One-dimensional acoustic model of the vocal tract with yielding wall based on the finite-difference time-domain method

○竹本浩典（千葉工大），足立整治（フラウンホーファー研究所）

◆これまで，われわれはシンプルなアルゴリズムで声帯音源波形から音声波形を時間領域で生成する声道音響モデルを提案してきた。

◆しかし，このモデルでは壁振動を模擬できなかったため，第１共鳴周波数が低くなるという問題があった。

◆そこで，壁振動の影響を考察したところ，その影響は連続の式には現れるが運動方程式には現れないこと，つまり音圧の補正のみ行えばよいことが明らかになった。

◆そして，壁の変位から音圧を補正する式を導出した。

◆母音/a/の声道断面積関数（Fig. 1 上）を用いて壁振動の効果を検証したところ，第１共鳴の周波数の上昇が確認できた（Fig. 1 下）。

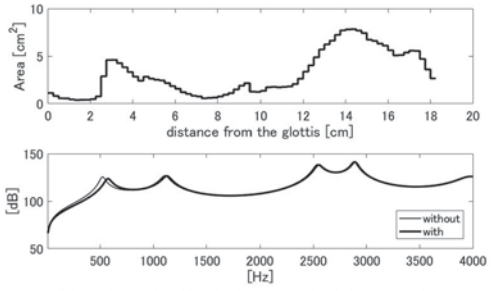


Fig.1：Area function for the vowel /a/ (top panel) and transfer functions with/without yielding wall effects (bottom panel)

2-5-3

2-5-3　rtMRI 動画から発話器官の輪郭を抽出する学習器の生成に関する検討

Examination of building learners which extract contours of the speech organs from rtMRI

☆後藤翼，竹本浩典（千葉工大），北村達也（甲南大），
能田由紀子，前川喜久雄（国語研）

◆日本語調音音声学の精緻化のため，18 名の被験者の発話運動を rtMRI 動画で記録してデータベースを構築した。

◆この動画から発話器官の輪郭を点群データとして抽出し，定量分析を行うために，機械学習を導入した。

◆本研究では，どの被験者を教師データとして選択して学習器を生成すべきか，発話器官の概形のクラスタ分析を行って検討した。

◆発話器官の概形は，Fig. 1 で示すように，発話器官が小さな女性型（F1~F2），大きな男性型（M2～M1），その他の混合型（M12～M8）の３つに分類された。

◆主に混合型と混合型から最も離れている女性被験者を教師データとして学習器を生成したとき，最も輪郭抽出の精度が高くなった。

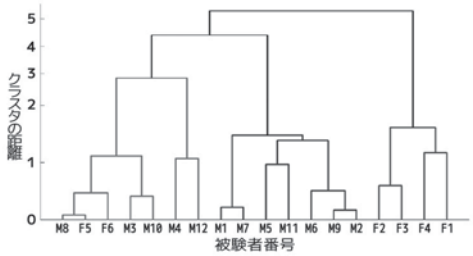


Fig.1 : Dendrogram for bounding-box of the speech organ among 18 subjects

2-5-4

2-5-4　発音診断とその確信度に基づく遠隔構音検査システムの評価

Evaluation of Remote Articulation Test System based on Pronunciation Diagnosis and its Confidence

○勝瀬郁代（近畿大・産業理工），△平島ユイ子（福岡国際医療福祉大）

◆外部専門家（言語聴覚士）が，対面での構音検査に近い方法で遠隔から構音検査ができる遠隔構音検査システムを開発してきた。

◆３つの検査方法（対面検査、遠隔構音検査システムによる遠隔検査、一般的な web 会議システムによる遠隔検査）間の診断性能を比較。

◆言語通級指導教室の児童に対し、言語聴覚士と発音指導に携わる教員が単語構音検査を実施。さらに、音素ごとに診断確信度を「あまり確信がない(1)」から「十分な確信がある(5)」までの５段階で評定。

◆異なる検査方法間並びに検査者間で、発音診断結果の AC1 統計量を算出して一致率を求めた結果、どちらも診断結果の一致率が高かった。

◆各検査方法における診断確信度の平均の差の検定を行った結果、遠隔構音検査システムを用いた検査の診断確信度は、対面検査よりも，かつ web 会議システムを使用した場合よりも有意に高かった。

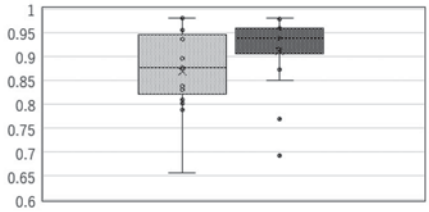


Fig.1 Observer Agreement in Pronunciation Evaluation between Inspection Methods. (: Face-to-Face vs Remote Articulation Test System, : Face-to-Face vs Teleconference System)

2-5-5

健常被験者の軟口蓋破裂音/k/産生時に生じる代償運動の観測・筋活動の推定

Observation and simulation of compensatory movements during velar plosive / k / production for normal subjects

☆古田尚久，党建武（JAIST）

◆２種類の舌の必要挙上範囲を拡大した条件（舌圧子条件，開口条件）で，健常発話者７名に /aka/～/ako/の産生を求めて EMA で調音運動を測定し，Dysarthria 患者の調音運動に見られる，舌の「引き込み」「中央寄せ」のパターンが健常発話者にもみられるかどうかを検討した．

◆通常条件に比し，７名中６名に「引き込み」を，７名中２名に「中央寄せ」を認めた．また７名中６名に左右幅の有意な拡大を認めた（いずれも重複有，table1２参照）．

◆Dang らの舌生理学的モデルを用いて，「引き込み」による代償を茎突舌筋（SG）の活動増加にて，「中央寄せ」による代償を横舌筋（T）の活動増加にて再現し，引き込み優位被験者，中央寄せ優位被験者と比較した．引き込みにおいては観測と矛盾の無い結果が得られた．また，左右幅の拡大には SG の活動増加が関わっている可能性が示された．中央寄せについては幅の変化は観測と矛盾の無い結果が得られたが，舌尖の移動方向については異なる結果が得られた．中央寄せ優位被験者では SG と T の両方が代償に関与している可能性がある．

◆Dysarthria 患者の/k/産生時の舌の「引き込み」や「中央寄せ」は，代償運動である可能性も検討する必要があることが示唆された．

table.1 : Lower gums - tongue tip

	aka	aki	aku	ake	ako
Subject A	+	+	+	+	+
Subject B	+		+		+
Subject C	+	+		+	*
Subject D	+	+	+	+	+
Subject E	+		+	+	+
Subject F				−	
Subject G	+		+	*	

table.2 : Width of tongue

	aka	aki	aku	ake	ako
Subject A	+				+
Subject B	−	+	−		−
Subject C		+			*
Subject D	−	+	−		
Subject E	+	+	+	+	+
Subject F		+		+	
Subject G				*	

"+" : Distance significantly increased
"-" : Distance significantly decreased
"*" : Missing value

2-5-6

2-5-6　チューブ発声法による音声リハビリ支援システム「スマートチューブ」の開発

A supporting system for voice rehabilitation by tube phonation exercise, Smart Tube

○川村直子（姫獨大），北村達也，濱田賢汰（甲南大）

◆音声障害患者が行う音声リハビリのチューブ発声訓練では，発声中の共鳴や顔面皮膚振動感覚が重要なフィードバックになる．

◆しかしながら，実際の発声訓練で言語聴覚士が訓練中の患者の共鳴や顔面皮膚振動の感覚を定量化してバイオフィードバックできる手段はなかった．

◆そこで，我々は先行研究でチューブ発声訓練時の顔面皮膚振動を利用したバイオフィードバックシステムを開発した（川村ら，2018）．

◆本研究では，このシステムをさらに低コスト化・小型化し，チューブ発声訓練中の振動の大きさを LED の光でフィードバックできる音声リハビリ支援システム「スマートチューブ」を開発したので報告する．

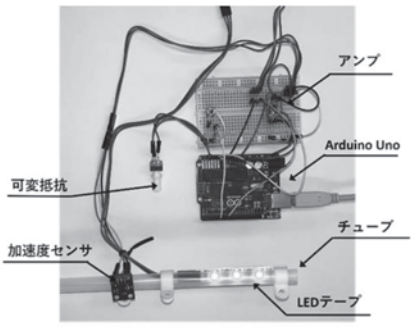


Fig. 1: Prototype of Smart Tube system.

2-5-7

2-5-7 日常生活で発音記号を使う活動

Using international phonetic alphabet in life.

○竹内京子（國學院大）

◆一般的に，音声学の授業で必ず学ぶ発音記号は何のために学ぶのだろう。もちろん，発音記号を使って音声を記述したり，辞書の表記を読んだりを念頭に学ぶのが普通である。しかしながら，音声研究を目指さない，特に音声に特に興味がない学生にとって，授業でその存在を知るだけに終わることが多い。

◆また，発音記号を知るためには，ある程度，記述のための規則と,その使い方を時間をかけて学ばなくてはいけない。「知る」という作業だけで嫌気がさしてしまう学生も多い。言語学専攻の学生でさえ，音声学，特に発音記号だけは苦手という者も少なくない。

◆なぜ，こんなに発音記号の暗記が，嫌われるのかと考えると，日常生活で全く使う場面がなく，習った知識を使う機会がほとんどないことが考えられる。今回，授業で習った発音記号の規則を練習する機会を学生が自分たちで実際に作るという活動をした。具体的には，「日本語音声学」の授業で，発音記号を使った LINE のスタンプを学生に作ってもらった。本発表は，その実践報告である。

Fig.1：スタンプの作成見本

2-6-1

2-6-1 円板付中空型ステップホーンを用いた小型空中超音波音源の検討 ー円板とホーンの寸法の検討ー

Study of small type aerial ultrasonic source using hollow stepped horn with circular plate -Examination of circular plate and horn dimensions-

○淺見拓哉，門前大樹，三浦　光（日大・理工）

◆筆者らは，小型で強力な超音波を放射する音源として，円板付中空型ステップホーンを用いた約60 kHz の音源の検討を行っている。ここでは，有限要素法(COMSOL Multiphysics 5.4)を用い，円板と中空型ステップホーンの寸法について検討を行い，実験結果と比較した。

◆Fig. 1 は，音源の先端部より放射された音波の指向性の結果である。なお，用いた音源の寸法は，有限要素法により効率よく音波を放射できる値とした。図より，距離 300 mm における測定結果の音圧は，0° 方向で最大値となり，単一の指向性となっていることがわかった。なお，半値角は約 34° で，最大音圧は約 31 Pa (約 124 dB)であった。また，測定結果の音圧の指向性は，解析結果とほぼ一致した。これより，円板付中空型ステップホーンを用いた音源の特性は，有限要素法にて推定できる可能性を示した。

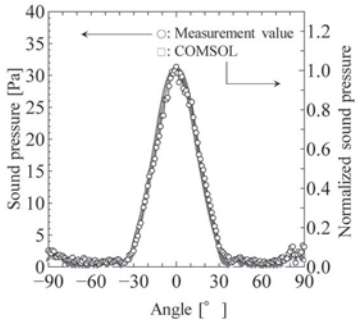


Fig. 1. Directivity of sound at the tip of the sound source.

2-6-2

2-6-2 円形たわみ振動板超音波音源の多周波駆動

Ultrasonic sound source with a circular flexural vibrating plate capable of multi-frequency drive

○伊藤洋一，大隅歩 （日大・理工）

◆円形たわみ振動板を用いた超音波音源(Fig. 1)は、振動板が同心状の節円モードで振動するため、使用する周波数帯ではただ一つの共振周波数を持ち、極めて安定した駆動が行なえる。音波の放射パターンが単純で、高効率で放射できることも大きな特徴である。

◆この音源を多周波駆動可能な音源に変革させることを考えている。

◆これを実現する方法として、まず円形振動板を半径に沿って多分割する。振動板の中心に対して対向する二つの振動片の長さを等しくする。各組の振動片の長さをわずか変化させる。これにより、各組の振動片はそれぞれ異なる周波数で共振状態となるため、音源は複数の共振周波数を持つことになる(Fig. 2)。

◆提案手法を基に、複数周波数で共振する円形たわみ振動板を用いた超音波振動源(20kHz 帯)を試作し、その可能性について検討している。

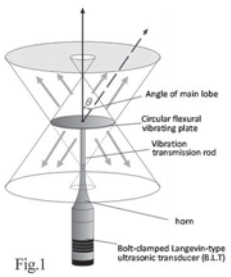


Fig.1

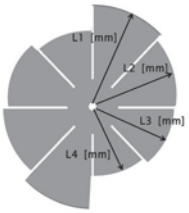


Fig. 2

2-6-3

2-6-3 Double-parabolic-reflectors ultrasonic transducers (DPLUS) for multi-modes excitation

○ Kang CHEN (The Univ. of Tokyo), Takasuke IRIE (Microsonic Co., Ltd.), Takashi IIJIMA (The National Institute of Advanced Industrial Science and Technology), Takeshi MORITA (The Univ. of Tokyo)

Double-parabolic-reflectors wave-guided ultrasonic transducers (DPLUS) to realize wideband (0 to 2.5 MHz), multi-harmonic modes excitation (over 20 modes), and large mechanical/acoustic output will be introduced in this presentation. The double parabolic reflectors mechanism serves as a horn structure at low frequencies and an acoustic focusing structure at high frequencies to enhance the energy density of the incident ultrasound. We verified the working mechanism by simulation and experiments. Superior abilities of DPLUS are very promising for examining some basic physics (such as frequency dependence) in the fields of medicine, biology, industry, etc.

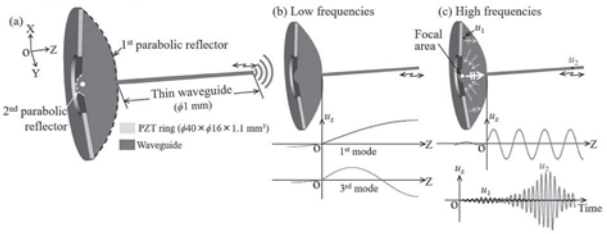


Fig.1: Working mechanisms of DPLUS. (a) Waveguide configuration. (b) At low frequencies, the 1st parabolic reflector is analogous to a horn structure, and the vibration amplitudes in the thin waveguide is enhanced. (c) At high frequencies, incident ultrasound goes through double parabolic reflections.

2-6-4

2-6-4　強力空中超音波フェイズドアレイの音波照射特性

Radiation characteristics of high-intensity aerial ultrasonic phased array

○大隅歩，清水鏡介，山田健太，伊藤洋一（日大・理工）

◆本研究では，AUPA による弾性波源走査法を提案し，空気結合超音波による非破壊検査の高速化の可能性について検討している．非破壊検査においては，対象に照射される音波のビーム径やサイドローブの影響が大きく影響する．

◆本報告では，上記の計測を実現させるための AUPA の試作，およびその音波放射特性について検証したので報告する。

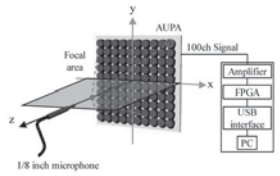


Fig.1 Experimental device and system

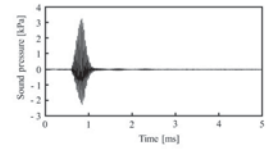


Fig.2 Wave form of sound wave at focal point

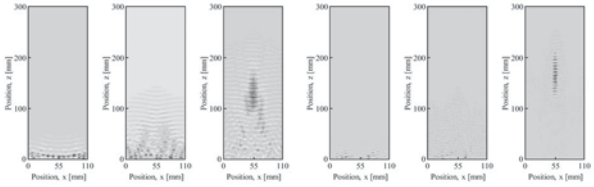


Fig.3 Sound pressure distribution in free field at fundamental frequency and third frequency.

2-6-5

2-6-5　音源位置近くの等音速層で覆われた海底における音場変化

Sound field change at the bottom covered with iso-velocity layer near sound source position

○鶴ヶ谷芳昭（山陽精工），菊池年晃（防衛大），水谷孝一（筑波大）

◆海底に音波が放射される場合、音波は海底境界面で反射するか又は海底に透過する．しかし、海底境界面で音波が臨界角を超えると音波は全反射し、海底には透過しない．つまり，水中における音場は近距離においては海底面での透過と反射から形成される．この海底境界面に堆積層が存在する場合，音場に対する堆積層の影響は強く表れる．

◆Fig. 1 に周波数 100 Hz，音源深度 10 m において，海底上の堆積層の厚さが 50 m の場合の堆積層を有する音場から堆積層のない音場を減算した結果(差音場)を示す．左図は海底層のみの差音場であり，右図は海底層を堆積層で置き換えた場合である．この両方の図を比較すると，音源位置に比較的近距離において，音場は堆積層からの全反射と堆積層上部の境界面からの透過から形成されていることが分かる．さらに堆積層の厚さを減少させていくと，層厚が約 5m において差音場は逆転し，さらに薄くなると海底層境界面の影響が増大する．

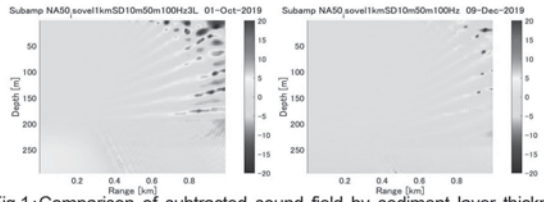


Fig.1 : Comparison of subtracted sound field by sediment layer thickness 50m　freq. 100Hz　left; subtracted bottom　right; subtracted sediment bottom

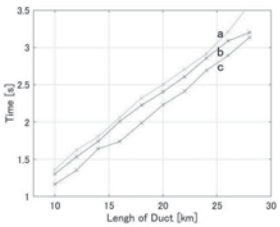
2-6-6

2-6-6　活断層モデルの内部構造と放射波

Internal structure and radiation waves of active fault model

○菊池年晃（防大），水谷孝一（筑波大）

我々は、地震の活断層内の振動構造を音響学的知見から調べている。活断層からの振動放射は方位的に双極性である。その中心軸上には活断層内の振動が反映される点が存在する。その点では、しばしば指数対数減衰型の波形が観測される。そこでその波形と活断層の内部振動を関連づけるために活断層モデルの放射シミュレーションを行った。周辺速度を 6000 m/s, 中心速度を 3000m/s とする V 型速度構造ダクトを用いる。ダクトから水平距離 30 km で受波したパルスの長さを求め、その結果を下図に示す。曲線 a, b, c は、それぞれ間隙 600, 440 と 280 m のダクトに対するものである。いずれも長さに比例してパルスが延伸している。また間隙が大きいほどパルスは延伸する。パルスの形状は指数対数減衰型であるが包含されている振動周期は間隙によって変わる。長さ 10 km のダクトで終端を超えた位置 11 km におけるパルスのパワースペクトルのピーク値は、a, b, c に対してそれぞれ 2.20, 2.88, 3.83 Hz である。即ち包含される周波数は間隙が小さいほど高くなる。このような指数対数減衰形は実際の観測波形でもしばしば観測されている。したがって、実際に観測される指数対数減衰形の波形には活断層の構造要素が含まれていると考えられる。



2-6-7

2-6-7　極浅海域における音源移動方式による高周波音波に関する実験及び解析

* Experiment and analysis of high frequency sound field with use of a moving source in extremely shallow water

△山口　拓斗，大北　和輝，△八尾　悠生，太田　和彦（金工大・工）
△小川　誠，○山本　満，△中野　正規（NEC）

◆極浅海域における高周波音波の伝搬特性把握のため，石川県穴水湾口近辺（Fig.1& Fig.2）において受波器を固定，音源を移動させて音場計測を実施した

◆下記に示す計測結果の一例（Fig.3）は音源周波数 19 kHz の場合で、ガウシアンビームに基づく音線法 Bellhop によるシミュレーション結果 (Fig.4) と比較した．この例では海底質を砂と仮定すれば両者が一致することが確認される．

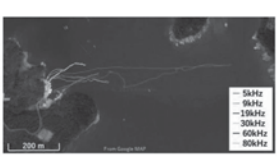


Fig.1 Experimental site

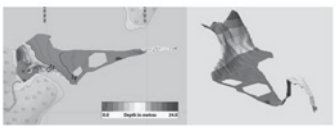


Fig.2 Measured sea bottom geometry

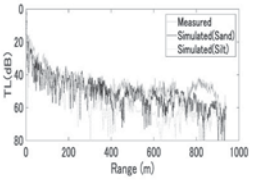


Fig.3 Comparison of measured and simulated TLs

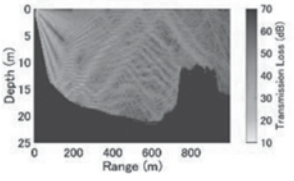


Fig.4 Simulated field using Bellhop

2-6-8

2-6-8　周囲雑音イメージングに用いる凸型音響レンズ材の音速測定：水温依存性の予備的結果

Sound speed measurements for some materials of convex acoustic lens applying to ambient noise imaging:
Preliminary results of temperature dependences

○森 和義，小笠原 英子（防衛大）

◆周囲雑音イメージングシステムを AUV 艇首に搭載するための凸型レンズに適した固体材料の音速について，その水温依存性を計測した．

◆3 種のシンタクチックフォーム TG-26/3000，TG-28/4000，TG-39/11500 について 10～30℃の水温範囲で計測したところ，いずれも水温の上昇に伴い音速が低下するという水温依存性の予備的結果を得た．(Fig. 1).

◆計測より得られた近似直線より，利用予定の水温における音速を推定して屈折率を設定した後，無収差化を施して最適化されたレンズ形状の予備的な結果を得た．ここでは，20℃において海水の音速を 1515.8 m/s，内部液体フロリナート FC-72 の音速を 512 m/s，TG-28/4000 の音速を 2589.5 m/s とした．入射角度 0, 5, 10, 15 度における音線は，それぞれ概ね一点に集束している．(Fig. 2)

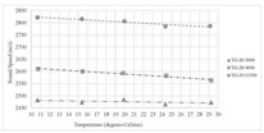


Fig. 1　Preliminary results of sound speeds for materials of solid lenses at 500 kHz

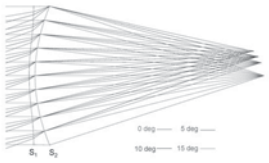


Fig. 2　Example of optimized lens shape and ray diagram for TG-28/4000

2-6-9

2-6-9　グリーン関数 PE 法を用いた海洋内音波伝搬解析

Ocean sound propagation analysis using Green's function parabolic equation method

☆甲斐　理一郎，鮫島　俊哉（九大・芸工）

◆海洋中での音波伝搬解析によく使われる手法として，放物型方程式法（Parabolic Equation Method：PE 法）があるが，PE 法には大別して，Crank-Nicholson PE (CNPE) 法と Green's function PE (GFPE) 法の 2 種類がある。

◆GFPE は CNPE に比べて有利な特徴を備えているものの，計算式の導出にやや複雑な理論が必要となり，利用者があまり多くない。本稿では，深海を想定した 3 次元空間音場において，音源周波数や音速分布を変化させながら，GFPE と CNPE の計算精度と速度の比較を行った。

◆計算時間においては，GFPE の優位性が示されたが，計算精度については，海水層と堆積層の物性値の変化が不連続の場合，GFPE と参照解の誤差は大きくなった。

◆解析音場を海水層のみに限定し，海水層と堆積層の境界をインピーダンス面として反射波を考慮した定式化・計算を行うと，精度が改善されるのではないかと考えられる。

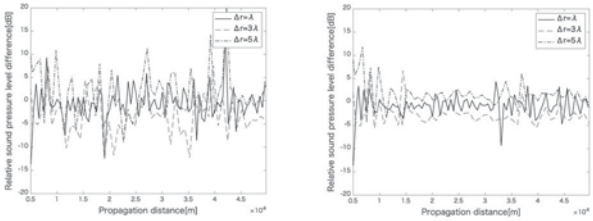


Fig.1 Difference in relative sound pressure level at sound source depth (GFPE - reference).
The sound velocity distributions of the sea water layer and the sedimentary layer are (a) Munk's profile and 1700 [m/s], (b) Munk's profile respectively.

2-6-10

2-6-10　インライン型海底地震計により推定した音源方位のずれについての考察

Study on misalignment of sound source direction estimated with in-line type ocean bottom seismometer

○岩瀬良一（JAMSTEC）

◆防災科学技術研究所により，北海道根室半島沖から千葉県房総半島沖に至る広大な海域に設置された海底ケーブル型観測システム「日本海溝海底地震津波観測網」(S-net)の海底地震計（3 成分加速度計）の波形データのパーティクル・モーションを用いて，エアガンの音源方位を求めると，その時系列において，実際の船舶方位から，観測点によっては最大で約 30 度に及ぶずれが見られる．また，水中音波の海底面への入射角についても，一部に顕著なずれが見られる観測点が存在する．こうしたずれは，音源定位に支障となる可能性がある．そこで今回は，これらのずれの原因について考察した．

◆音源方位のずれについては，岩手県釜石市沖の三陸沖システムの海底地震計を用いたエアガンの音源定位においても最大で約 9 度のずれが確認されている．これら日本海溝域に設置された観測点に共通する地形的特徴は，斜面，即ち傾斜を有する海底地形である．

◆これまでは水平境界を有する流体―固体境界への平面波入射を仮定していた．そこで，流体―固体境界が傾斜している場合の影響を検討した．その結果，海底の傾斜による影響は小さいことが分かった．

◆一方，地震計の X 成分の方位，即ち海底ケーブルの敷設方位付近では，ずれが小さくなる傾向がみられた．

◆このことから，地震計の設置方位並びに海底とのカップリング状態の影響が示唆される．

2-6-11

2-6-11　視聴覚情報の深層ベイズ学習に基づく音響シーン分析

Acoustic scene analysis based on audio-visual deep Bayesian learning

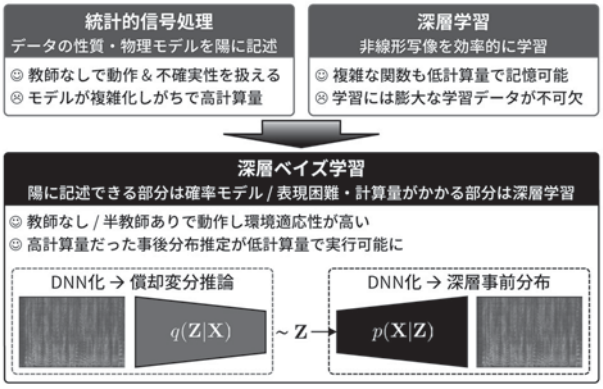
○坂東 宜昭（産業技術総合研究所）

◆目的：**画像上で音を発している領域を推定する DNN** の学習
- ➢ 従来は**単チャネル録音**と画像のペアデータから自己教師あり学習
- ➢ 類似した物体が多い画像（例：人混みの映像など）で原理的に劣化

◆キーアイディア：**多チャネル音響信号の空間情報**を利用
- ➢ 問題：**どのように空間情報を深層学習に統合する？**

◆本発表では，
- ➢ 近年の視聴覚処理による音響シーン分析の動向
- ➢ これまで我々が取り組んできた**音響信号処理の深層ベイズ学習**
- ➢ 現在取組んでいる**深層ベイズ学習に基づく視聴覚処理** (2-1-12)

について紹介する



2-6-12

2-6-12　深層学習を用いた超音波肝腫瘤像のAI 診断システム開発

AI diagnosis system developed by deep learning of ultrasound images of liver lesions

○椎名　毅，山川誠（京大・医），西田直生志，工藤正俊（近畿大・医）

近年、医用画像の深層学習による AI 診断システムの開発競争は、世界的に加速化しているが、高精度な AI システムの開発の鍵を握るのは、正確なアノテーションがなされた高品質な画像データベースである。日本医療研究開発機構（AMED）「臨床研究等 ICT 基盤構築・人工知能実装研究事業」では、病理、放射線、内視鏡、超音波など医療画像系の６学会が中心となり、国立情報学研究所と連携して、高品質な医療画像のビックデータベースの構築と、それを用いたＡ診断Ｉシステムの開発が進められている。超音波画像は、2018 年から日本超音波医学会が中心となり、21 の医療機関の協力のもと、肝臓、乳腺、循環器に関する、静止画と動画の収集と学習データの作成を開始した。1 年経過時点で、肝腫瘍 31,000 画像と乳腺腫瘍 14,000 画像が収集され、それらを活用した深層学習により、動画からの肝腫瘍の検出と、疾患タイプを鑑別をする AI 診断システムのプロトタイプを開発し、一定の成果を得ている。本講演では、これらの取り組みを紹介する。

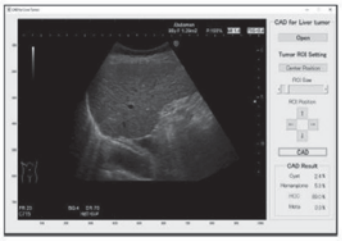


Fig.1 A prototype AI system for diagnosis of liver lesions

2-6-13

2-6-13　AI² による音響解析について

Acoustic analysis by using AI²

○井上忠宣，△Phongtharin Vinayavekhin，森國秀，
△Don Joven Agravante，立花隆輝（IBM Research）

◆我々は音響解析の要素技術的側面とその応用的側面の両面から技術開発を進めており、本発表では音源位置推定と製造業における音響解析の応用について紹介する。

◆実世界でラベル付きデータを集めるのは大変なので、ラベルに合わせて合成したシミュレーションデータと比較的集めやすいラベルなし実世界データを使った敵対的学習による手法を提案する。(Fig. 1)

- ラベル付き実世界データがなくても、シミュレーションデータで学習した場合に比べて、より高い推定性能を得られた。

◆製造業における音響解析の応用として、溶接音による溶接品質診断や打音検査システムの技術開発に取り組んでいる。

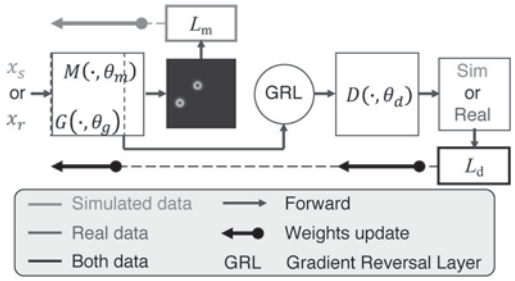


Fig. 1: Adversarial learning framework for sound source 2D localization

2-6-14

2-6-14　位相・気象・嗜好 ～波と天気と買い物のはなし～

Phase, Meteorology, and Consumer Behavior - Relationship between frequency, weather and shopping -

○吉開朋弘（一般財団法人日本気象協会）

◆気象分野におけるアコースティックイメージングの活用事例として，リモートセンシング技術を用いた**気象衛星観測**と**気象レーダー**の事例を紹介する。近年のリモートセンシング技術の発展により、衛星では **2.5 分毎に約 1km 四方の雲の分布**、気象レーダーでは **1 分毎に約 250 m 四方の雨の分布**を把握することが可能になっている。

◆また、データ活用ビジネスにおける人工知能の活用事例として、気象予測を活用した商品需要予測事例を紹介する。気温や体感気温を考慮した需要予測の結果、**冷やし中華つゆの売り上げの 97%を気象で説明する**ことに成功した。

◆実際に食品メーカーで予測情報を利用して生産量の調整を行った結果、**冷やし中華つゆの最終在庫を約 20%削減することに成功**し、近年社会問題となっている**食品ロスの削減に貢献した。**

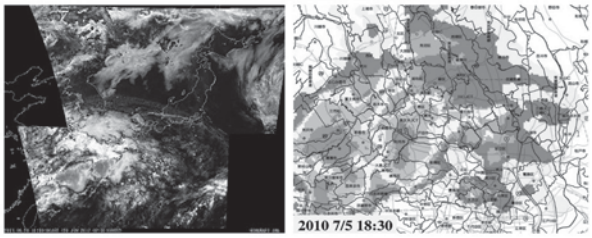


Fig.1：Example of Color Enhanced Satellite Imagery of Cloud Level from the Himawari-8 and Rainfall Intensity Distribution from X-band Radar

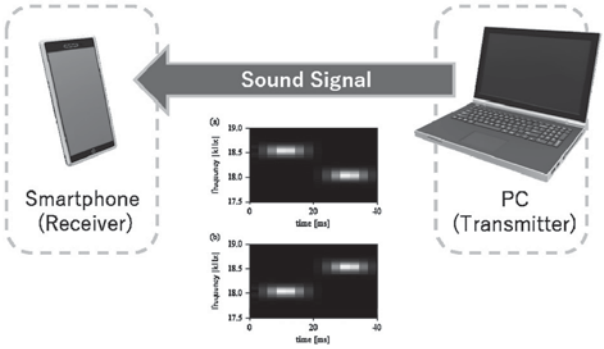
2-6-15

2-6-15　機械学習による非可聴音通信システムの精度向上に関する一検討

A Study on the Enhancement of an Inaudible Sound Communications System Based on a Machine Learning Approach

☆尾関 剛成，青木 直史，土橋 宜典（北大）

本研究では，機械学習による非可聴音通信システムの精度向上について検討している．本発表は，環境雑音，ドップラー効果，マルチパスエコーによる影響を考慮した音信号から、ニューラルネットワークを利用してバイナリ情報を抽出する識別実験の結果について報告する．



本研究では，音信号のスペクトログラムから，バイナリ情報を埋め込んだ領域を画像として抽出し，ニューラルネットワークを適用することで音信号の識別を試みている．

2-7-1

2-7-1　FDTD 法を用いた生体組織構造とせん断波伝搬速度評価精度の関係性の検証

FDTD verification of relationship between biological tissue structure and shear wave propagation velocity evaluation accuracy

☆伊藤大貴[1]，大栗拓真[1,2]，神山直久[2]，吉田憲司[3]，山口匡[3]

([1]千葉大院・融合理工，[2]GE ヘルスケア・ジャパン，[3]千葉大・CFME)

◆Shear wave elastography (SWE) の標準化アプローチのため，脂肪肝の構造や脂肪量が Shear wave velocity (SWV) に与える影響について検証するため，FDTD 法を用いて脂肪滴を含む肝臓を模擬した組織構造におけるせん断波伝搬シミュレーションを行った．

◆計算シミュレーションから得られた粒子速度分布から相互相関法により，脂肪量毎の肝臓における SWV 分布を算出した（Fig.1）．

◆脂肪量の増加により組織構造が複雑になるにつれ，　脂肪滴からの散乱によりせん断波の伝搬が複雑になることや，異なる媒質の境界付近におけるせん断波形の乱れにより，解析した SWV のばらつきが大きくなることが確認できる．

◆今後は脂肪の粘性を考慮した検討や線維も含む肝臓でのせん断波の伝搬をシミュレーションで検証していく．

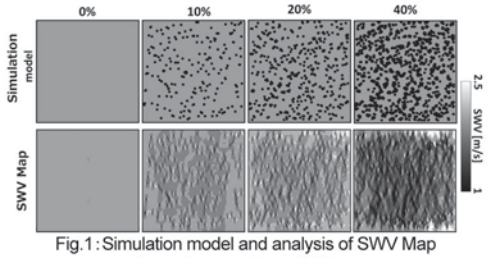


Fig.1：Simulation model and analysis of SWV Map for each proportion of fat liver

2-7-2

2-7-2　アニュラアレイでの振幅包絡特性解析におけるコヒーレンス因子の影響

Influence of coherence factor in analyzing envelope statistics using an annular array

溝口岳，大村眞朗（千葉大・院融合），
Jeffrey A. Ketterling（Riverside Research），吉田憲司，山口匡（千葉大・CFME）

◆適応ビームフォーマの一種である Coherence Factor（CF）がアニュラアレイの振幅包絡特性解析に与える影響について検討した．

◆5 つの素子で構成される中心周波数 20 MHz の高周波アニュラアレイを用いて，均質構造を有するファントムからのエコー信号を取得した．

◆CF の推定に用いる素子数を担保するため，送受信の組み合わせを考慮し，25 素子と見なして CF を推定したところ，高エコー信号以外の成分が抑制されることを確認した．（Fig. 1(a)）

◆レイリー分布を指標化可能な Signal-to-Noise Ratio（SNR）と汎用的な統計モデルである Homodyned-K 分布（HK 分布）の散乱信号のパワーを示す変数（$2\mu\sigma^2$）によりスペックル抑制能を評価した．

◆散乱体密度が高い領域では，振幅包絡特性を示す両変数が低値となり，CF によりスペックル成分が抑制されることが示された．（Fig. 1(b)）

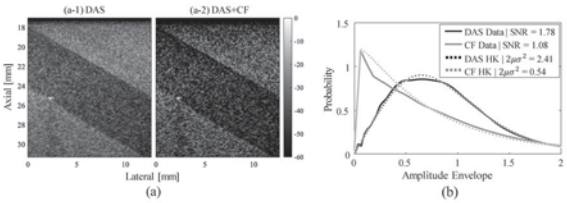


Fig.1 (a) shows B-mode images obtained with DAS (a-1) and CF weighted DAS (a-2). (b) shows probability density function of envelope and estimated parameters of high scatterer density region.

2-7-3

2-7-3　等価音源法を用いた少数マイクロホンによる一次反射音場の可視化

Visualization of sound field including primary reflections by using small number of microphones with equivalent source method.

☆津國和泉，田中琴美，池田雄介，小坂直敏（東京電機大）

◆問題点

高解像度な音場の可視化は，大規模マイクロホンアレイや長時間の測定時間が必要

◆目的

少数マイクロホンの伝達関数から高解像度な音場の可視化

◆提案手法

- 等価音源法に基づき，点音源の線形和によって伝達関数を推定
- スパース最適化によって点音源の重みを決定

◆実験

- 無響室にて反射板を用いて一次反射音を可視化

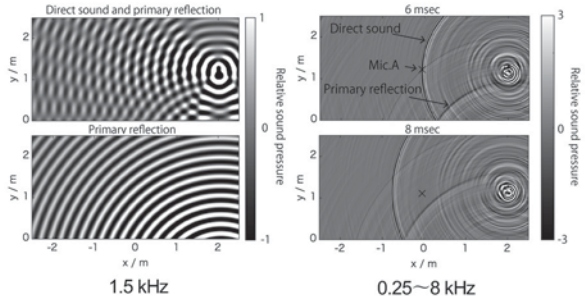


Fig.1：Visualization of direct sound and primary reflection

2-7-4

2-7-4　―室内における人の数や姿勢の推定―音響特性の計測による室内状態の検知(III)

―Estimation of the number of persons and their posture―
Detection method of room conditions by measuring acoustic characteristics (III)

☆林田耀生，△水野洋輔，中村健太郎（東工大）

◆部屋の周波数特性の変化から室内にいる人の人数を推定することを目的として，1/8 縮小モデルを用いた基礎実験を行った。

◆人の人数（0～3 人）と位置を変えて周波数特性を計測した。周波数特性を得るには、スピーカからチャープ音を出力しマイクロホンで収音，PC で FFT などを行う。

◆学習のための特徴量は、5 つの共振のピーク周波数と、600～3000 Hz における周波数特性形状の類似度である。

◆MATLAB 分類学習器・線形判別コマンドを用いて学習モデルの作成と精度の検証を行った。用意した周波数特性データは学習用として 300 個、検証用として 100 個である。その結果、97%の精度で人の人数を推定することができた。

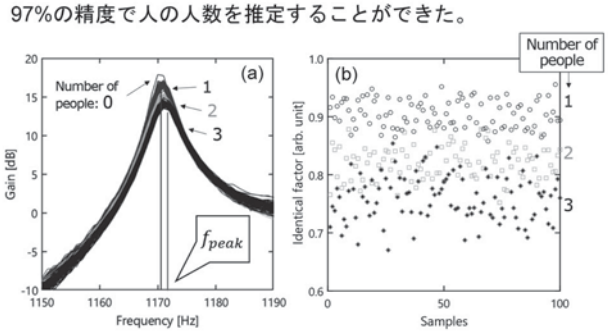


Fig. 1 Feature values used for training. (a) Peak frequencies of resonance peaks. (b) Identical factor.

2-7-5

2-7-5　２次高調波成分を用いた Continuous Shear Wave Elastography に関する基礎検討

◎ 長岡 亮(富山大), 吉澤 晋(東北大), 長谷川英之(富山大)

Continuous shear wave elastography (cSWE)は，臨床で広く使用されているSWEとは異なり，外部から機械的に振動(連続波)を加え，その伝搬速度から組織の弾性を評価する手法である．外部加振によって生体に振動を生じさせる必要があるため，適用先が比較的体表付近の組織に限られる．我々の研究グループでは，皮膚の弾性値を定量的に評価するために，伝搬によって生じる横波の２倍高調波成分に着目した推定手法を提案する．

本報告では，弾性値が既知である３種類のSWE用ファントムに対して，提案手法と基本波成分を用いた従来手法とを適用し，横波速度を推定した．また，この推定した結果を校正値より算出した横波速度とも比較した．右図に得られた結果を示す．図より明らかであるように，提案手法によって推定された横波速度の方が校正値と近い結果となった．ファントム実験結果より，２倍高調波成分に着目した提案手法の有用性が示されたと考える．

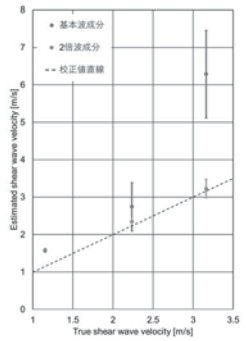


Figure: Summary of the estimated shear wave velocities using the proposed and conventional methods.

2-7-6

2-7-6　A Framework for Product Quality Assessment by Acoustic Analysis on Induced Sound

○森國 秀, △P. Vinayavekhin, △井上 忠宣, △D. J. Agravante, △A. Munawar, 立花 隆輝, △S. Wang, △B. J. Ko (IBM Research)

◆In many manufacturing applications, product quality assessment is an indispensable process, which are often expensive due to labor or equipment costs. With the aim to develop an adaptive solution, we conducted an experimental study to investigate the possibility of a proposed nondestructive testing framework, integrated with robotics and acoustic analysis.

◆To be data efficient with real-world applications, we investigated the effects of data augmentation, data collection strategy and convolutional kernel sizes.

◆We empirically tested product quality categorized in both:
- Classification: when the quality can be discretely evaluated
- Regression: when the quality is continuously evaluated

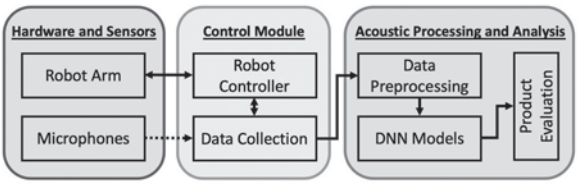


Fig.1 : Proposed framework overview

2-7-7

2-7-7　Double-Nakagami モデルによる振幅特性評価の精度検証

Accuracy verification of amplitude envelope statistical analysis by Double-Nakagami model

☆佐藤悠佑(千葉大･院融合), 田村和輝(浜松医科大), 大村眞明(千葉大･院融合), 吉田憲司, 山口匡(千葉大･CFME)

◆Double-Nakagami モデルは，正常肝の管腔構造および脂肪肝の脂肪滴からのエコー成分の足し合わせを仮定している．

◆*in vivo* ラットの正常肝（n=2），軽度（n=4）および重度（n=4）脂肪肝症例に対し，高周波リニアプローブ（5，6，7，9 MHz）を用いて多断面のエコーデータを取得し，振幅包絡分布を Nakagami および Double-Nakagami モデルで評価し，各モデルパラメータを比較した．

◆各パラメータは，脂肪沈着度に相関して平均的に高くなる傾向を示したが，特に Double-Nakagami モデルでは周波数毎の変動も確認され，正常と軽度脂肪肝の差異も大きい結果となった．（Fig. 1）

◆Double-Nakagami モデルの２つのパラメータを相互に用いることで，Nakagami モデルに比して高感度に脂肪沈着を評価可能であることが示唆された．

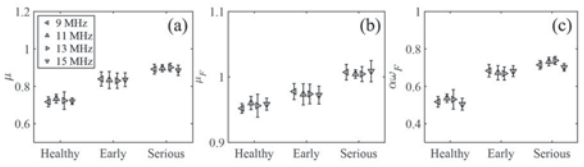


Fig.1 : Model parameters in each liver disease. (a) Nakagami parameter μ, (b) and (c) Parameter μ_F and $\alpha\omega_F$ of Double-Nakagami model.

2-7-8

2-7-8　非接触フェーズドアレイによる材料の評価

Imaging of defect in solid material by non-contact phased array

☆清水鏡介, 大隅歩, 伊藤洋一(日大理工)

◆空中超音波フェーズドアレイ(Airborne Ultrasound Phased Array:AUPA)による弾性波源走査法を利用した材料の評価方法について検討を行っている。

◆本報告では，基礎検討として強力集束超音波を発生できる AUPA を試作し，減肉擬似欠陥を有する固体材料に対する非破壊実験を試みている。

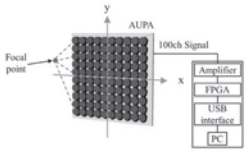


Fig.1 Schematic view of AUPA

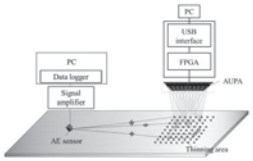
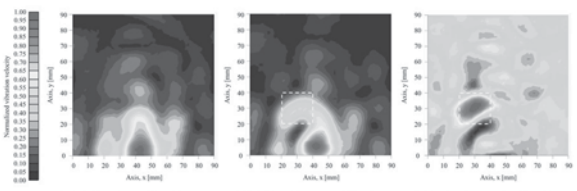


Fig.2 Experimental device



(a) Vibration distribution without defect area　(b) Vibration distribution with defect area　(c) Vibration distribution with background differencing technique

Fig.3 Measurement result

2-7-9

2-7-9　コンクリート非破壊計測のための非接触音響探査法に関する研究

-地下空洞天井部の吹付コンクリートへの適用例-

Study on noncontact acoustic inspection method for non-destructive measurement of concrete,
-Example of application to shotcrete in underground cavity ceiling-

○杉本恒美，杉本和子（桐蔭横浜大・院），
△歌川紀之，黒田千歳（佐藤工業）

◆吹付けコンクリートはトンネルや地下空洞に広く用いられているが、天井部が高い場合には、叩き点検が困難であり、遠隔から非接触で検査できる手法の開発が求められている。

◆一方で、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を用いれば、5m以上の離隔からコンクリート表層部に存在する欠陥部の検出が可能である。

◆そこで今回は、地下空洞天井部における欠陥を 20m 以上の離隔から実際に検出可能かどうかについて検証実験を行った。

◆実験結果から、適切な防音・防振対策を施すことにより、叩き点検と同程度の欠陥位置検出が可能であることが確認された。

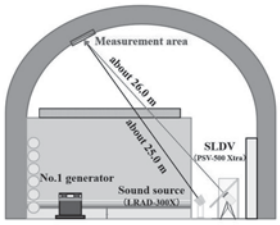


Fig.1 : Experimental Setup.

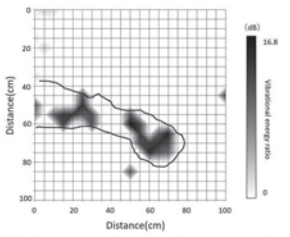


Fig.2: Distribution of vibrational energy ratio. (Black line indicates a defective portion.)

2-7-10

2-7-10　コンクリート非破壊検査のための非接触音響探査法に関する研究

－音源搭載型 UAV を用いた外壁検査の効率化に関する検討(2)－

Study on the Non-Contact Acoustic Inspection Method for Non-destructive Concrete Inspection －Efficiency improvement of outer wall inspection using sound source mounted type UAV (2)－

○上地 樹，杉本 恒美，杉本 和子，大平 武征，中川 裕(桐蔭横浜大院)，
歌川 紀之，黒田 千歳(佐藤工業(株)技術研究所)

◆ 我々は、空中放射音波および超高感度スキャニング振動計 PSV-500 Xtra (Polytec Corp.)を用いた、非接触による非破壊探査法である非接触音響探査法の検討を行っている。

◆ 今回は外壁タイル供試体を対象に、平面音源 (FPS1030M3F1R, FPS Inc.)を搭載した UAV (MATRICE 600 pro, DJI Co., Ltd.)を用いた非接触音響探査法による計測実験を実施した。

◆ 音響加振には、断続的に周波数を変調させた波を連続的に並べたマルチトーンバースト波を使用した。

◆ 実験の結果、模擬欠陥の位置と同様の場所で振動エネルギーの上昇が確認された。

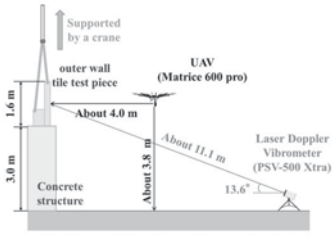


Fig.1 Experiment setup

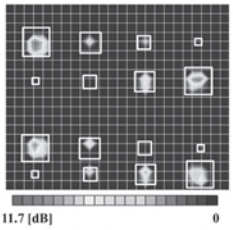


Fig.2 Vibration energy ratio (800-3400 Hz)

2-7-11

2-7-11　熱音響モード同期レーザーの検討

Study on thermoacoustic mode-locked laser

○平山悠暉，兵頭弘晃，Rémi Delage，琵琶哲志（東北大学大学院・工）

◆本研究では共鳴管内に発生する熱音響自励振動系において，互いに位相が揃った複数の振動モードを同時に発振させることを試みた．目的は音響パルス波を発生する熱音響モード同期レーザーの開発である．

◆管端からスタックまでの距離 a を 120 mm に固定し，全長 L を 1.20 m から 2.50 m まで連続的に長くしながら管端で音圧 p を計測した．

◆周期に対する半値幅の比 $\Delta t/T$ は L = 1.20 m の場合 0.29 であったが，L = 2.50 m の場合 0.13 にまで小さくなった (Fig. 1)．この値は初期のモード同期レーザー並みに小さな値である．

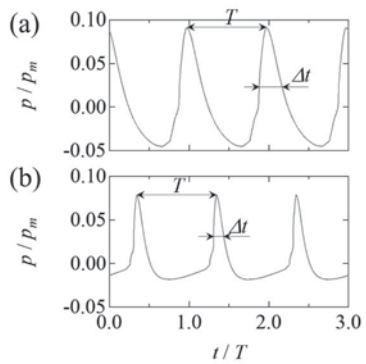


Fig. 1 Normalized acoustic pressure p/p_m as a function of time when (a) L = 1.20 m and (b) L = 2.50 m; p_m is the mean pressure. The horizontal axis is the time normalized by the fundamental period of oscillations.

2-7-12

2-7-12　電磁気−力学−音響回路の連成によるスピーカー印加電圧から音圧の予測

Prediction of the sound pressure by applied voltage to the speaker using coupling of electromagnetism - dynamics - acoustic circuit

○倉井陽太郎，上田祐樹（東京農工大）

◆熱音響冷凍機は冷却部に可動部を持たず，自然冷媒を用いることが可能なデバイスの一つである．熱音響冷凍機を設計するためには，中空の管および蓄熱器内の流体の挙動，音響入力源の力学的な運動，音響入力源の電気的な特性を同時に解く必要がある．そこで本研究では流体の挙動を熱音響理論を用いて解析し，その結果を，音響入力源を構成する可動部品の運動方程式および回路方程式と組み合わせる方法を提案する．

◆印加電圧の周波数を 1 Hz~600 Hz まで 1 Hz ずつ変えながら中空の共鳴管閉端部の圧力振幅を計算した結果を Fig.1 に示す．周波数が大きいほどピーク時の圧力振幅が大きくなることが分かった．

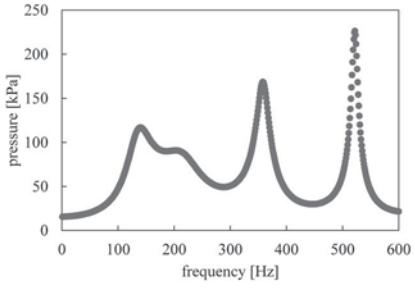


Fig.1 Pressure amplitude

2-7-13

2-7-13　音響管の急拡大部前後の圧力差の直接測定

Direct measurement of minor loss pressure difference in oscillating flow

○大橋謙一朗，上田祐樹(東京農工大)

◆熱音響デバイスは工場排熱や太陽熱などの未利用熱の有効利用手段として注目されている．熱音響デバイスは熱音響理論により解析がされているが，線形理論である熱音響理論では駆動後の出力や効率の予測は不可能である．我々は非線形ロスを熱音響理論に組み込むことで出力や効率の予測を試みた．今回は非線形ロスとして急拡大部のマイナーロスに注目し，急拡大部でのマイナーロスによる圧力差を実験式及び，実験による直接測定から得ることでそれらを比較しどれほど一致するかを確認した．

◆Fig.1 に装置内の圧力振幅を 0.2~6.3 kPa まで変化させた場合のマイナーロス前後の圧力差の実測値と算出値を示す．装置内の圧力振幅に関わらず 5%の誤差範囲内で実測値と算出値が一致することが確認できた．また，傾向としては圧力振幅が大きくなるほど実測値と算出値の差が大きくなることが確認できた．

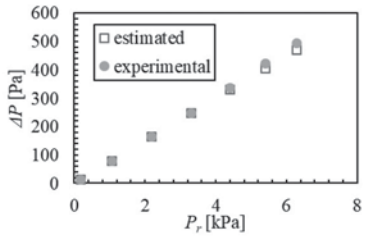


Fig. 1 Experimental and estimated pressure difference

2-7-14

2-7-14　直管型段差熱音響システムについての基礎検討 –熱緩衝管と段差形状が与える影響–

Fundamental study on straight-tube-type step-shape thermoacoustic system-Effect of thermal buffer tube and step shape-

☆熊崎雅也，坂本眞一（滋賀県立大）

◆熱音響システムでは，スタック両端に温度差をつけることで熱から音への変換が行われる．

◆スタックに入力された熱が外部方向に漏れることで温度分布が形成される区間は，熱緩衝管と呼ばれる．

◆小型熱音響システムはサイズ上放熱が難しいため，中型・大型のシステムに比べ熱緩衝管がシステム全長に対して占める割合が大きい．

◆本検討では段差形状の異なる 2 種類のシステムにおいて，熱緩衝管の長さがシステムに与える影響を実験的に検討した (Fig. 1)．

◆ヒーターを複数用いることで熱緩衝管長さを変化させ，発振温度を比較した．

◆実験結果より，システム段差形状にかかわらず熱緩衝管長さがシステム発振温度与える影響は少ないと考えられる (Fig. 2)．

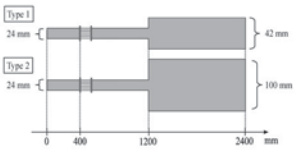


Fig. 1 Experimental systems.

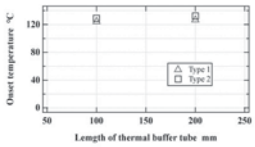


Fig. 2 Comparison of experimental results.

2-7-15

2-7-15　ループ管型熱音響システムにおける熱緩衝管に関する基礎検討 - 熱緩衝管外壁における熱移動 -

Fundamental study on the thermal buffer tube in a loop-tube-type thermoacoustic system –Heat transfer on the thermal buffer tube –

☆勝木秀和，坂本眞一（滋賀県立大）

◆熱音響システムにおける熱緩衝管の検討として，断熱材をステンレス管の間に挿入し，温度分布やシステムに与える影響について検討を行った．

◆ループ管型熱音響システムの全長は 3300 mm，管内径は 42.6 mm，スタックのセル数は 600 cell/inch^2 とし，スタック高温端から 100 mm の位置に断熱材を挿入した．断熱材として，アルミホイル 50 枚(0.85 mm)，ゴムパッキン(ニトリルゴム 1 mm)を使用した．

◆断熱材なしを Type A，アルミホイルを Type B，ゴムパッキンを Type C とし，入力電力を変更させ熱緩衝管の温度分布を測定した．

◆実験結果より，断熱材による温度変化は確認できず，管壁を軸方向に伝わる熱伝導の影響は小さいと考えられる．

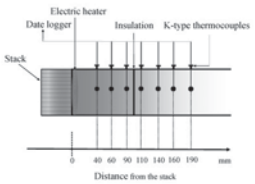


Fig.1 : Schematic illustration of thermal buffer tube.

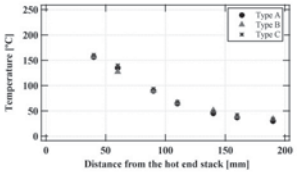


Fig.2 : Temperature distribution of the thermal buffer tube.

2-7-16

2-7-16　ループ管型熱音響システムにおける局所的な共鳴管材質変化が音場に与える影響

The effect of sound field by local material change of resonance tube in loop-tube-type thermoacoustic system

☆栗原祥太，坂本眞一（滋賀県立大）

◆熱音響システムとは，熱と音の相互エネルギー変換である熱音響現象を利用したシステムである．

◆実用化に向けての課題としてシステムの小型化・軽量化が挙げられる．

◆システムの軽量化を目的とし，システムの一部をステンレスからアクリルに変更し，実験的に検討を行った．

◆ループ管型熱音響システムの全長は 3300 mm，管内径は 42.6 mm，スタックは 600 cell/inch^2 とした．(Fig. 1)

◆共鳴管の一部をアクリルに変更した場合，音圧は約 4%大きくなることが確認できた．(Fig. 2)

◆システムの一部を軽量化することによる，性能の低下はみられなかった．

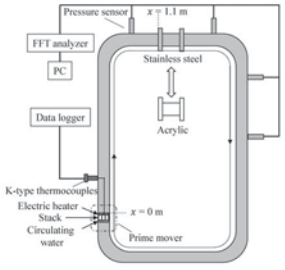


Fig. 1 Experimental system.

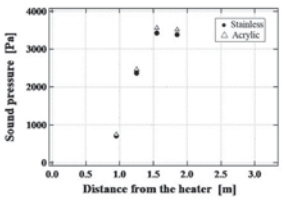


Fig. 2 Sound pressure distribution.

2-7-17

2-7-17　定在波型熱音響システムにおける熱音響現象促進効果 - 重畳音波の音源設置位置が仕事流生成量に与える影響 -

Enhancement thermoacoustic phenomenon in standing-waves thermoacoustic system -Effect of installation position of superimposing sound waves on work flow generation

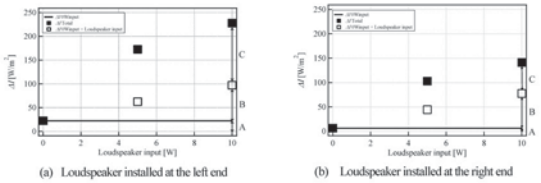
☆平松康斗（同志社大），坂本眞一（滋賀県立大），倉田侑弥（同志社大），渡辺好章（同志社大）

◆熱音響システムの実用化の課題の一つであるエネルギー変換効率の向上を目指す手法として，管内に自励的に生じる音波に外部から音波を重畳する手法が提案されている.

◆外部音源にラウドスピーカを用い，システム低温端側の管端部に設置することで仕事流生成量の増幅が確認されている.

◆本報告では，ラウドスピーカ設置位置を変更することによる仕事流生成量増幅効果について検討を行った.

◆その結果，ラウドスピーカから生じた音波がスタック内を通過する際のスタックが有する温度勾配の正負により，スタック通過後の音圧が変動し，仕事流生成量の増幅量が外部音源設置位置により異なることを確認した.



(a) Loudspeaker installed at the left end　(b) Loudspeaker installed at the right end

2-7-18

2-7-18　熱音響プライムムーバーの低温度発振に向けた研究 －3D プリンタースタックを用いた試み－

Research on low-onset temperature of thermoacoustic prime mover -Trial applying 3D printer stack-

○坂本眞一，片岡慎太朗（滋賀県立大）

◆熱音響現象を応用したシステムの実用化に向けた研究を進めている.

◆スタックとは，熱音響システムの主要部品の１つであり，非常に小さな流路を束ねたような構造をしている.

◆音と熱のエネルギー変換においてスタックはとても重要な役割を果たしており，これまで多く研究等が実施されてきている.

◆熱音響プライムムーバーのスタックに樹脂で製作された 3D プリンタースタックの利用を目指した設計を試みた.

◆スタックの流路を変更したときの発振温度の計算結果を図に示す.

◆計算結果より，流路直径を２mm とした場合において約 180℃となっている.　3D プリンタースタック作製は PLA 樹脂を約 200℃で溶かすため，この条件においての利用は可能と考えるが，実用化に向けてはさらなる低温度化が望まれる.

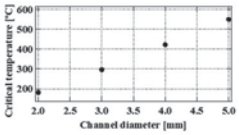


図　拡大プライムムーバー計算結果

2-9-1

2-9-1　後期残響音における反射音到来方向と室の条件との関連性について

Relationship between Direction-of-Arrival in late reverberation and room properties

☆泉 悠斗，大谷 真（京都大学大学院工学研究科）

◆後期残響音において，反射音の到来方向分布には空間的な偏りが生じることが報告されている.

◆本報では，幾何音響シミュレーションを利用し，後期残響音における反射音到来方向と室内表面の吸音率との関連性を考察する.

◆分析結果より，後期残響音における反射音到来方向分布には室の各面の吸音率と室の形状との両者が影響していることが明らかとなった.

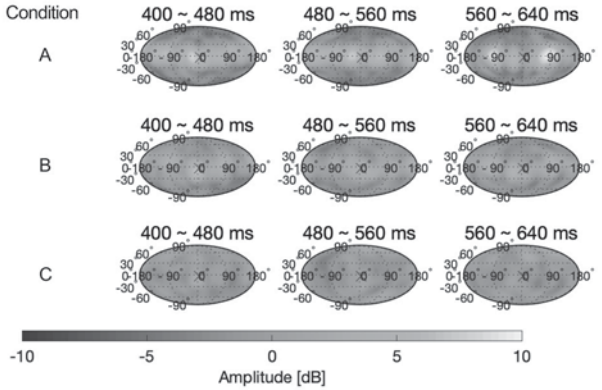


Fig.1: Direction-of-Arrival distribution 400 ~ 640 ms after the direct sound arrival on each condition. (A: Acoustic absorption coefficient for the floor surface is 0.8, and that for the other surface is 0.2. B: Acoustic absorption coefficient for all surface is 0.2. C: Acoustic absorption coefficient for all surface is 0.) Latitude and longitude correspond to elevation and azimuth, respectively.

2-9-2

2-9-2　音源の指向性と方向別インパルス応答を用いたステージ音場の評価と再生の試み

Evaluation and Reproduction of the Stage Sound Field Using Directional Characteristics of Sound Source and Directional Impulse Responses

☆岩橋佑弥，尾本章（九大芸工）

◆音響測定では一般的に全指向性音源を用いるが，本研究ではある程度の指向性を有するスピーカー (GENELEC 8050A)を様々な方向に向けながらインパルス応答(以下，方向別 IR)を測定した。

◆得られた方向別 IR から周波数特性や物理指標を算出し，音源の指向性を考慮したステージ音場の評価を行った。

◆スピーカーを仰角 $\phi = 0°$ に固定し，方位角 $\theta = 0°$，45°，90°，135°，180° に向けた時の周波数特性を Fig. 1 (直接音成分), Fig. 2 (残響成分)に示す。結果として，方向別 IR の周波数特性は残響成分に大きな違いは見られず，主に直接音成分によって決定づけられることが確認できた。

◆評価を通して，音源の指向性の影響を十分に受けることがわかった。

◆さらに方向別 IR を用いて簡易的なステージ音場の再生を試みた。

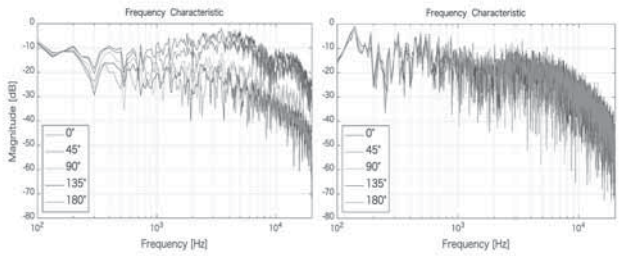


Fig. 1 : Frequecy Characteristic (direct sound)　Fig. 2 : Frequecy Characteristic (reflected sound)

2-9-3

2-9-3　音に包まれた感じの評価指標 SBTs の音響インテンシティによる測定

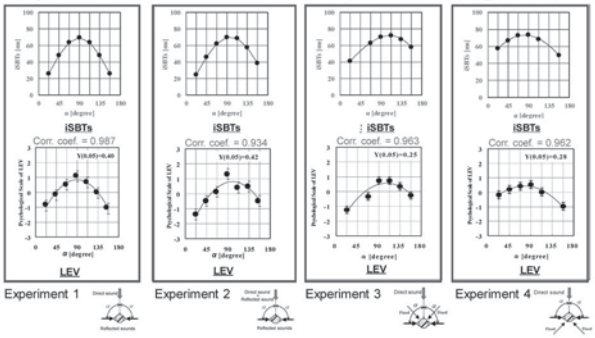
Sound intensity-based measurement of evaluation index SBTs of listener envelopment

○羽入敏樹，星和磨（日大・短大）

◆筆者らは 音に包まれた感じ(LEV)の評価指標としてSBTs を提案している。原理的にはSBTs は音場の方向別インパルス応答から算出できる。しかし実測される方向別インパルス応答は，測定に用いたマイクロホンアレーや指向性マイクロホンの指向特性によって方向間でクロストークが生じる。このクロストークは使用するマイクロホンに依存し，SBTs の値に影響を与える。

◆本報では，マイクロホンの指向特性に依存しないSBTs の測定方法を確立することを目的に，インパルス応答における瞬時インテンシティを用いて SBTs を定義すると共に，正中面内にける反射音の相互作用を修正した iSBTs を新たに提案した。

◆両指標は既往研究における LEV の心理尺度値と高い相関を示した。



2-9-4

2-9-4　仮想音環境における残響空間下での先行音効果の検討

Examination of precedence effect in reverberant space in virtual sound environment

☆武藤泰斗，安藤優佑，岩立直樹，大隅歩，伊藤洋一（日大・理工）

◆先行音効果を利用した音声避難誘導システムの研究を行っている。このシステムを実装する場合，基本音響設計に加え，現場での調整が必要となることが予想される。このような課題に対して，対象となる建物空間に本システムを設置する際の各種音響条件の設定を，近年普及している VR 技術(ヘッドマウントディスプレイ：HMD，ヘッドホン)の利用を提案している。

◆本報告では，実空間での残響を考慮し，残響を付加した場合の仮想空間における先行音効果について検討を行った。

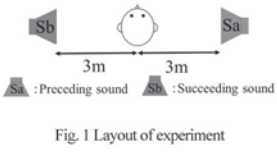


Fig. 1 Layout of experiment

Table 1 Evaluation index

評価値7：	Sa側に非常に方向感がある
評価値5：	Sa側に方向感がある
評価値3：	Sa側に方向感が少しある
評価値1：	方向感が全くない 分離していてわからない

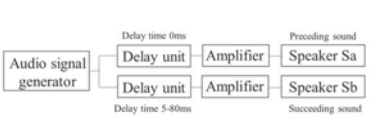


Fig. 2 Block diagram

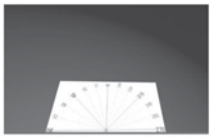


Fig. 3 Visual conditions

2-9-5

2-9-5　多群会話空間における話しやすさと距離に関する検討

A study on ease of speaking and distance in multi-group conversation space.

☆柿澤貴成，上田麻理(神奈川工科大)，Firdaus Zuhal(PENS)，
平栗靖浩(近畿大)，丸山直也，川井敬二(熊本大)

レストランやカフェなど，滞在する複数群の会話音声が混在する“多群会話空間”と定義する.

本研究では，多群会話群間の距離に着目し，話しやすい条件，話しにくい条件を明らかにすることを目的として，実際の多群会話空間である神奈川工科大学のカフェテリア(Fig.1)で音響心理実験を行い，以下のことがわかった.

◆ 背景音の大きさに応じてうるささの印象評価は変わる.

◆ 他群との距離の違いによる会話のしやすさやその他の印象に差は確認されなかった.

◆ 背景音レベル 50 dB と 60 dB では会話のしやすさに変化は無かった.

Fig.1 カフェテリアの実験風景

2-10-1

2-10-1　音源の映像情報が列車走行音の大きさ感評価に及ぼす影響

Effect of visual stimuli of sound source on loudness evaluation of train running noise

☆小野寺拓馬（東大院），三宅耀生（東京理科大院），
朝倉巧（東京理科大），李孝珍，坂本慎一（東大生研）

◆三次元映像再生にドームスクリーンを用いて，音源の映像情報が騒音の大きさ感評価に及ぼす影響を調べることを目的として，主観評価実験を行った。

◆列車走行音と映像を同時に呈示し，ME 法により試験音の大きさ感を評価させ，映像の有無と列車の色の違いによる影響を調べた。

◆映像を呈示した方が，大きさ感が小さく評価される傾向がみられたが，列車の色の条件間には統計的に有意な差は生じなかった。しかし，Red が大きく，Green と Yellow は小さく評価される傾向もみられた。

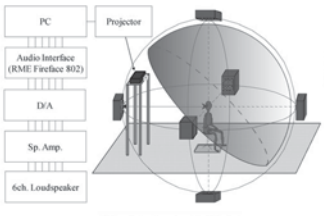


Fig.1 再生システム

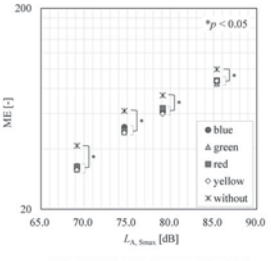


Fig.2 ME値と$L_{A,Smax}$の関係

2-10-2

2-10-2　ベトナムにおける自動車の警笛に関する意識調査

Questionnaire survey on vehicle horn use: A case study in Vietnam

○金基弘（駿河台大），△ファム・ズイクオン（駿河台大），高田正幸（九大・芸工）

◆一般的に、先進国に比べて「交通インフラ」や「交通文化」がまだ確立されていない多くの開発途上国では、交通量の多い道路などで日常的に警笛が多く使用されている。

◆本稿では、ベトナムの都市部（ハノイ市）を対象にアンケート調査を実施し、ベトナムにおける警笛の使用状況と音環境への影響を把握したうえで、アジア地域の国々との比較検討を行うことを目的とする。

◆調査の結果、ベトナムでは危険報知のために、警笛を多く使っていることが分かった。次いで、お礼や挨拶の意味で使ったり、注意を促すために使ったりする場合も多かった。また、腹が立ったためなど、感情（怒り）的な理由での警笛使用もある程度みられた。

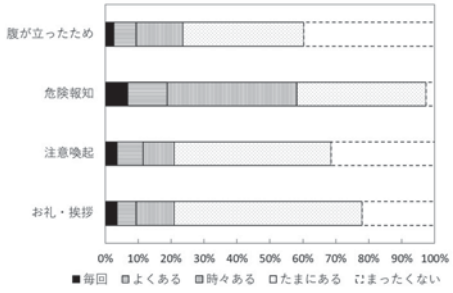


Fig.1：Driver's intention of horn use（%）

2-10-3

2-10-3　台湾と日本の都市部におけるドライバーの警笛使用に関するアンケート調査

Comparative study on driver's vehicle horn use in urban area of Taiwan and Japan.

☆橋本和真，稲田環，高田正幸（九大・芸工），
金基弘（駿河台大），△大枝良直（九大・工），
△Lin Yu Chang（台北科技大），△Lin Wei，△Huang Fang-Yi（台湾逢甲大）

◆ドライバーの警笛使用について，その使用頻度が異なる地域間で比較を行うため，台湾と日本の都市部でアンケート調査を行った。

◆日本と比べて台湾の方が高い頻度で警笛が使用されていることを示唆する結果が得られた（Fig.1）。

◆心理的要因の分析から，両地域ともに警笛を使用することが良いことであると考えているドライバーや警笛を使用することに対する心理的ハードルが低いドライバーは警笛を使用しやすいといえる。

◆環境要因の分析から，両地域ともに車両同士が急速に接近する状況で警笛の発生が増加することを示唆する結果が得られた。

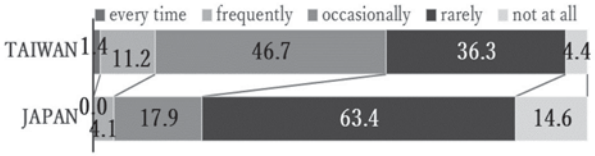


Fig.1：Response rates in questions about behavioral intention of driver(%)
(How often do you intend to honk?)

2-10-4

2-10-4　日本語形容語を用いたサウンドスケープ評価の主成分構造に関する実験

A principal components model of soundscape evaluation using Japanese adjectives

○川井敬二，△増本久美子（熊本大）

◆ISO/TS 12913-2 に記載されたサウンドスケープ主観評価形容語は、スウェーデン語で行われた主観評価実験の英訳に基づいており、これが他言語でも妥当かどうかについて検証は行われていない。

◆本研究は上記実験と基本的に同一の手法により、日本語の形容語を用いて評価実験を実施し、主成分分析によりその構造を抽出した。

◆結果として、スウェーデンの実験での第 1、第 2 主成分（pleasantness, eventfulness）と類似の軸が主成分構造に表れた。

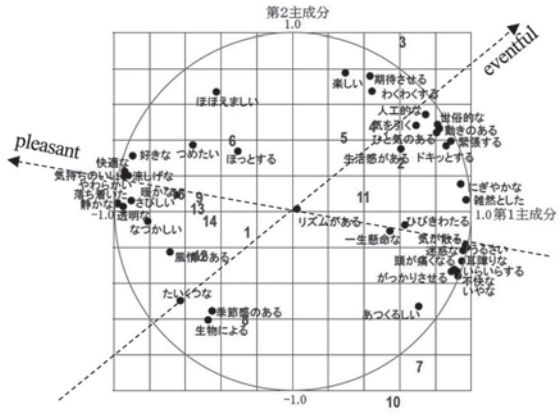


Fig.1：Principal component of soundscape evaluation

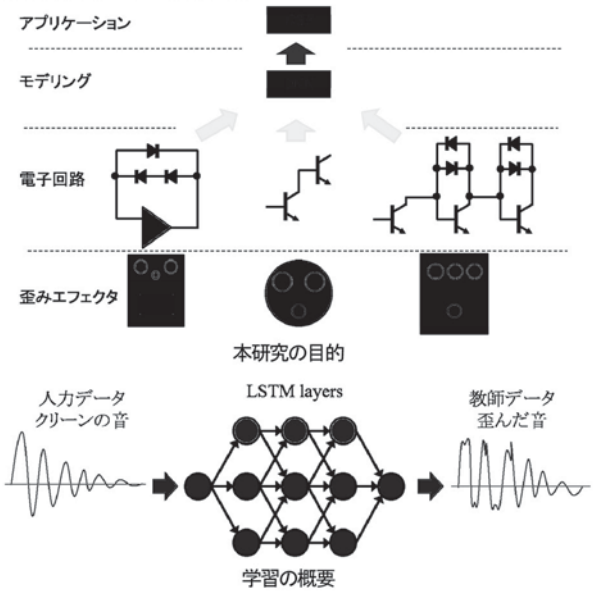
2-11-1

2-11-1　LSTM ニューラルネットワークによる歪みエフェクタのディジタルモデルング

A Digital Modeling Technique for Distortion Stomp Boxes Using LSTM Neural Networks

☆松永 悠斗，青木 直史，土橋 宜典（北大），小嶋 徹也（東京高専）

本研究は，歪みエフェクタに対する汎用的なディジタルモデリング手法の確立を目的としている．本発表では，効率的な学習モデルを作成するための手法について提案する．



2-11-2

2-11-2　音楽データの要約に基づくジャンル自動分類課題に関する一検討

A Study on Automatic Music Genre Classification Tasks Based on the Summarization of Music Data

☆菊池 雄大，青木 直史，土橋 宜典（北大）

本研究では，要約した楽曲を用いて楽曲のジャンル自動分類を行う際，その要約をどのように行えばより良い精度を得られるか，使用する特徴量および楽曲の要約手法の２つの観点から検討した．

特徴量は音楽情報検索の分野でしばしば用いられる MFCC と MFCC の導出過程で得られる MPC の２つを用い，分類の精度を比較した．

要約手法は楽曲中の一か所のみから抽出する手法と，その手法を複数回繰り返すことで楽曲中の複数個所から抽出する手法を用い，分類の精度を比較した．

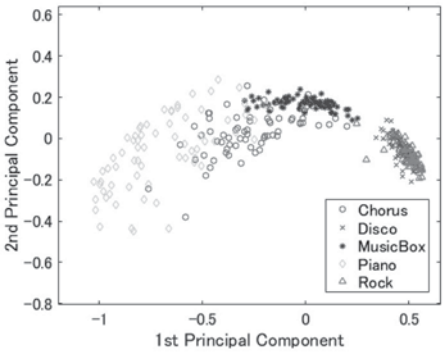


Fig.1：Scatter plot of MFCC clusters identified by using the SVM classifier. This result was obtained from the whole section of music data.

2-11-3

2-11-3　音響教育のためのプログラマブルエフェクタの開発

Development of a Programmable Stomp Box for Acoustics Education

○青木 直史（北大）

本研究では，音響教育のための教材開発を目的として，㈱北斗電子との共同研究を通して，プログラムしだいでさまざまな機能を実現できるプログラマブルエフェクタの開発を行ってきた．本稿では，これまでの開発の経緯について紹介する．

㈱北斗電子「サウンドエフェクタ」 http://bit.ly/2ZKfdtG
・C言語によるディジタルエフェクタのプログラミングが可能．
・製品には，ディレイ，リバーブ，フィルタ，ボコーダのサンプルプログラムを同梱．

ほぼあらゆる音をつくる方法 http://bit.ly/2xiU13n
・サウンドプログラミングのノウハウを一般に公開するためのウェブサイトとして開設．
・エフェクタによる音の加工について，プログラムとともにサンプルを公開．

2-11-4

2-11-4　音の広がり感に関する主観評価と物理量との関連

Relationship between subjective evaluation and physical quantity for sound spaciousness

☆前崎壮真（中央大院），戸井武司（中央大）

オーディオ機器の音質評価について，物理量に基づいた定量化は十分には行われていない。今回は音の広がり感に着目し，信号処理による物理量操作に伴う評価の変化を把握する。

広がり感評価に用いられる指標の一つに両耳間相関関数があるが，これに関わる物理量として Table 1 に示すような左右に両耳間時間差(0,20,40 ms)および両耳間強度差(0,3,6 dB)を与えた音源を用いて，それぞれの広がり感の印象変化を調査した。

評価項目のうち，今回は自然さと広がり感に注目する。Fig.1 に評価の平均得点の推移を示す。音量差に注目すると，左右差を増加させた場合，各時間差において自然さが減少し，それに伴い広がり感が減少することがわかった。一方，時間差に注目すると，時間差の増加に伴い広がり感が向上する傾向にあり，左右の音量差と時間差が広がり感の評価に影響していると推察できた。

Table 1　Combination of time difference and volume difference in evaluation sound

	0 ms	20 ms	40 ms
0 dB	Sound 1	Sound 4	Sound 7
-3 dB	Sound 2	Sound 5	Sound 8
-6 dB	Sound 3	Sound 6	Sound 9

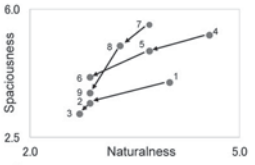


Fig.1　Average of evaluation of Naturalness and Spaciousness

2-11-5

2-11-5　Listener's experience and its impact on understanding spatial attributes associated with the 22.2-channel reproduction

○Sungyoung Kim (RIT, CET) and Shuichi Sakamoto (Tohoku Univ., RIEC)

◆A room interacts with sound sources, modulating our judgement (Toole, 2010). One's experience in music production also affects the person's performance in a listening test (Howie et al. 2019). Would room-induced impacts have an intertwined interaction with a listener's experience, esp. in the 22.2-channel reproduction?

◆A study was conducted, in which four listener groups compared four playback rooms (Kim et al, 2018 & 2019).

◆The results reveal that the degree of experience in 3D sound changed the participants' preference ratings as well as conceptualization of the sense of being enveloped by a sound image (LEV).

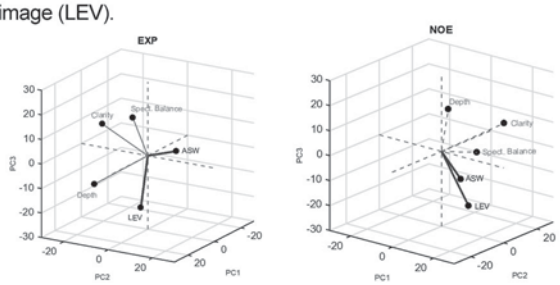


Fig.1：Different conceptualization of five attributes between the experienced group (EXP; left panel) and the inexperienced group (NOE; right panel).

2-11-6

2-11-6　（招待講演）楽器の音の計測と記述

Measurement and Analysis of Musical Instruments Sound

○及川靖広（早大理工）

◆PC の CPU のクロック数やコア数，メモリや HDD の容量，インタフェースやバスの伝送速度は桁違いのものとなっており，大量のデータを扱うことが容易な時代となった。

◆同様に計測器にも多くの変化が見られる。PC の高性能化により PC ベースで計測器を構築することが可能となり，ソフトウェアで目的にあった計測環境を容易に構築できる時代である。

◆20 年前を振り返ってみると，研究室では 4 点法を用いた世界中の音響ホールの音場計測，高速 1bit 信号の録音再生と信号処理，一般化調和解析を用いた信号処理等の研究が盛んに行われており，私もそれらに携わっていた。近接 4 点法を用いた世界中の音響ホールの音場計測はヤマハの AV アンプのデータに用いられ，高速 1bit 信号の録音再生は現在のハイレゾリューションオーディオの発展につながっており，基礎研究から実用化までを見ることができた。その後も，物理現象や信号の特徴の把握とそれに基づく応用に興味を持ち研究を行い，その対象は音響学会の中でも様々な分野にわたり，音楽音響分野でもこれまでにいくつか発表を行なってきた。

◆本稿では，近年我々が行っている音源近傍の計測，様々な信号処理の適用について紹介するとともに，今後の可能性について考えてみる。

2-11-7

2-11-7　サウンドエフェクトと電子音楽制作

Sound Effect and Electronic Music Creation

○小坂直敏（東京電機大学）

◆近年のデジタルサウンドエフェクトで、図１に示す構造的音色に属するエフェクトを中心に紹介する。

◆エフェクトの誕生には技術者側からのみならず、制作側からも提言可能である。

◆その例として擬似エコーと名づけたエフェクトを紹介する。

◆サウンドコラージュ法の応用として、音源を分解した音色を空間音響上に再生して音脈モワレ現象を起こす空間エフェクトの構想を紹介する。

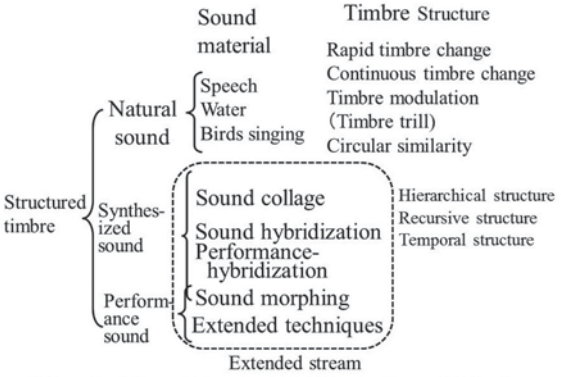


Fig. 1 Classification of Structured Timbre

2-11-8

2-11-8　平面波スピーカーによる立体音響システムの活用

Use of surround system with plane wave speaker

○古山俊一（尚美学園大学）

◆平面波による無指向性スピーカーシステム WRAPSOUN　NS-8Pro のサウンドとの出会い。

◆平面波による無指向性スピーカーシステム WRAPSOUN　NS-8Pro は立体音響に応用できるかの実験を行う。

◆「冨田勲という宇宙」というトリビュートコンサートでの実践的使用例を紹介。サラウンド音源「惑星」より「水星」再生。

◆平面波による無指向性スピーカーシステム WRAPSOUN　NS-8Pro を使っての**音場創生。実践例を紹介。サラウンド**音源「サン-サース作曲　交響曲　オルガン付き　第３楽章」再生。

2-11-9

2-11-9　複数解像度のスペクトログラムを用いた DNN 歌声分離

Singing voice separation with DNN using multiple resolution spectrograms

☆長友健人，升山義紀，竹内大起，矢田部浩平，及川靖広（早大理工）

◆従来の DNN を用いた時間周波数マスキングによる歌声分離は単一のスペクトログラムを入力 として使用 (Fig. 1)

◆スペクトログラムには時間と周波数の分解能にトレードオフの関係が存在
- ➢ 単一スペクトログラムのみでは抽出できる特徴量に制限

◆複数解像度のスペクトログラムを用いた DNN 歌声分離を提案 (Fig. 2)
- ➢ 単一スペクトログラムだけでは表現できない特徴量の抽出を期待

◆結果
- ➢ 提案手法により SAR，SIR，SDR が従来手法と比べて向上
- ➢ 複数解像度のスペクトログラムを DNN の入力特徴量として用いることの有効性を確認

Fig. 1 Conventional framework of DNN using single spectrogram

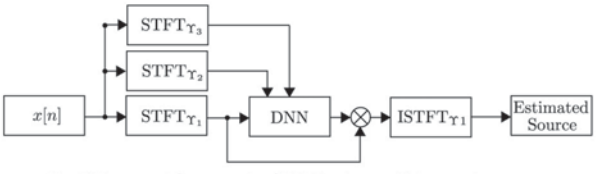


Fig. 2 Proposed framework of DNN using multiple spectrograms

2-11-10

2-11-10　マルチスピーカによるバーチャルコンサートホールの開発

Development of virtual concert hall by Multi speaker

○堀米里奈，横山真男（明星大・情）

【研究概要】

楽器演奏者は本番と練習の演奏環境の違いにより緊張する。練習での実力を発揮することができない状態になりやすい課題がある。本研究は**音響の変化**に着目して音響再現システムの開発を行う。本システムにより演奏者の**技術の向上**が期待できる。

【音響再現システムの開発】

本システムは演奏者の前に単一指向性マイク、演奏者を囲むようにスピーカを配置(Fig1)する。このマイクに向かい入力音として手を叩き、測定したインパルス応答と畳み込みによって生成された音をスピーカから出力音として再生する。

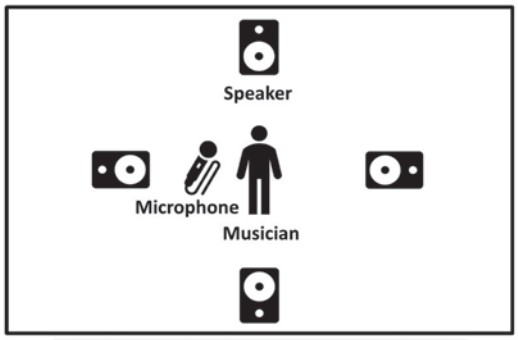


Fig1: Overview of development system

2-11-11

2-11-11 マッシュアップ生成のための楽曲ミックスの良さの DNN による評価の検討

Discussion on evaluation of music mixing for generating Mashup using DNN

☆小林史弥，シャイブラーロビン，小野順貴，若林佑幸（首都大）

◆複数の既存楽曲のパートをミックスして楽曲を作成する手法をマッシュアップという．本研究では，異なる楽曲のパートの「ミックスの良さ」を時間フレーム毎に DNN で評価することを検討した．ミックスの良さを評価できれば，動的計画法と組み合わせて適切なマッシュアップ候補を自動的に抽出できる可能性がある．本稿ではミックスの良さの評価に焦点を絞り，実験結果を報告する．

◆特徴量に 12 次元クロマグラムと 88 次元ピッチ特徴量を用意し，CNN で学習を行った．ラベルは 1 (良い)，0 (悪い)の 2 値である．

◆テストデータに対する最高正答率は 63.16%であった．各条件で，学習の反復回数 100 回のうち，バリデーションデータに対して損失関数が最小になった時点で作成されたモデルを用いてテストを行なった．実験結果の一例として，テストデータ中の 1 ミックスのスペクトログラム，正解ラベル，推定結果を Fig. 1 に示す．このミックスに対しては推定がある程度正しく行われたことが分かる．

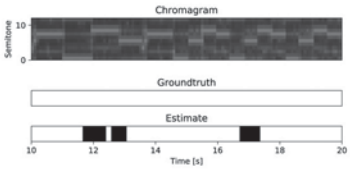


Fig. 1: Estimate result of good mix. White and black denote 1 and 0, respectively.

2-11-12

2-11-12　大学生に対する楽器演奏及び演奏練習支援に関する意識調査

Survey of attitude of university students toward musical performance and its learning support system

☆小出英範（東京情報大院），西村明（東京情報大）

◆東京情報大学の学生 251 人を対象に楽器演奏や演奏練習支援に関する意識調査を実施した。

◆意識調査の結果より、従来楽器とは異なる演奏インターフェースを持つ電子楽器とその演奏練習支援システムの必要性を調査した。

◆技術的な理由で演奏をやめたり挫折したりした演奏経験者や、楽器演奏を始めることが難しいと考える学生が多くいることがわかった。

◆自分が理想とする演奏ができると、楽器演奏が楽しいと感じる学生が多くいることがわかった。

◆楽器演奏経験者と未経験者の両方で、音の表現が制限される電子楽器や支援システムを使って、演奏をしたい学生がいることがわかった。

◆簡単に演奏できる電子楽器を演奏したい学生の中には、支援システムを使って電子楽器を演奏したい学生も多くいることがわかった。

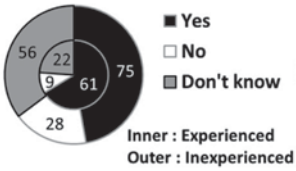


Fig.1 : Result of questionnaire : “Do you want to perform an electronic musical instrument with limited sound expression ?”.

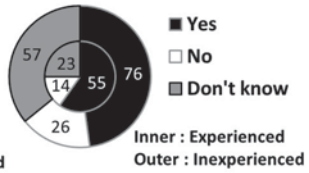


Fig.2 : Result of questionnaire : “Do you want to play a novel electronic musical instrument using a learning support system ?”.

2-11-13

2-11-13　演奏動作の軌跡長による鍵盤楽器と感情価の関連について*

Relationship between keyboard instruments and emotional valence by observing trajectory length of performance motion

○三戸勇気，川上央，駒澤大介（日大・芸術），三浦雅展（国立音楽大学），松崎桃子，山口達也，篠田之孝（日大・理工）

現在まで，我々は楽器演奏者の演奏時の動作と感情価との関連を明らかにするため，モーションキャプチャシステムを用いて検討を行ってきた．特に，鍵盤楽器であるグランドピアノ，電子ピアノ，チェンバロに着目して，研究を行ってきた．過去の研究では，それぞれ個別に検討を行ってきた．グランドピアノ，電子ピアノ，チェンバロは構造上の違いにより，音色や鍵盤の重さが異なる．そのため，鍵盤楽器の種類によって演奏動作が変化する可能性がある．そこで，実際に異なる 3 種類の鍵盤楽器における同一奏者の演奏動作をモーションキャプチャシステムにより計測し，異なる鍵盤楽器での演奏動作の違いを見出すための検討を行った．本研究では感情価に対する 3 種類の鍵盤楽器の演奏動作をモーションキャプチャシステムで計測し，演奏動作の違いを検討した．今回は，左右の肩，肘の動きに着目して分析を行った．

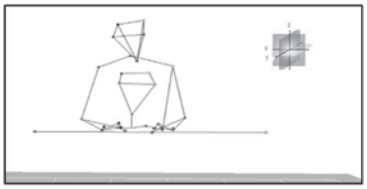


Fig. 1 Marker adhesion position

2-11-14

2-11-14 オペラ歌手の歌声におけるハイパーソニックエフェクトの存否の検証

Investigation on hyper-sonic effect for singing voice of an opera singer

☆森 亮太，山本 一公，中川 聖一（中部大学）

◆【目的】プロのオペラ歌手の歌声にも「ハイパーソニックエフェクト」が存在するのかを検証した。

◆【方法】サンプリング周波数 96 kHz，16bit の女性ボーカリストのハイレゾ音源を使用し、カットオフ周波数 20 kHz および 30 kHz の急峻なローパスフィルタをかけ、原曲・20 kHz LPF 音・30 kHz LPF 音の 3 曲を比較検討した。なお、原音には録音・AD 変換等のため 30kHz 以上で定常雑音があったため、以下の仮説を立ててスピーカー再生音の聴取実験により評価した（被験者は男性 9 名）。

評価項目は「高音感」「澄んでいる」「迫力」「広がり感」「好き/嫌い」の 5 項目である。

・**仮説 1**：各項目とも高評価は、30kHzLPF 音 ＞ 原音　(30kHz 以上に雑音が含まれるため)

・**仮説 2**：30kHzLPF 音 ＞ 20kHzLPF 音　(25kHz 付近に歌声の超高周波数成分あり)

・**仮説 3**：原音 ≧ 20kHzLPF 音　(原音は雑音が含まれるが、歌声の超高周波数成分あり)

◆【結果】　Table: Results of listening test

（○：total number of supporting subjects for hypothesis,

×　:total number of unsupporting subjects）

item / hypothesis	higher		clearer		more powerful		broader		total	
	○	×	○	×	○	×	○	×	○	×
30kHzLPFsound ＞ original	6	4	5	4	4	3	5	4	20	15
30kHzLPFsound ＞ 20kHzLPFsound	7	1	9	2	7	1	8	1	31	5
original ≧ 20kHzLPFsound	6	3	6	3	8	2	5	3	25	11
total	19	8	20	9	19	6	18	8	76	31

2-11-15

2-11-15 メロディーの輪郭の弁別特性：ミスマッチ・フィールドと弁別反応時間による検討

Discrimination characteristics of melody contour examined by mismatch field and reaction time

☆内藤柚菜（千葉大），ヤップ ゲック シエン（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）

◆メロディーの認知メカニズム解明を目指し，メロディーの輪郭形状および被験者の音楽経験の有無が，メロディーの弁別に及ぼす影響をミスマッチ・フィールド (MMF) 計測によって調べた．

◆5 音からなる音列の第 3 音の音程を上下させることで，標準音列と 5 種類の逸脱音列を作成した (Fig. 1)．全頭型脳磁界計測装置を用いて，逸脱刺激に対して誘発される MMF を観察した．

◆非音楽経験者の MMF 活動強度は，標準音列と逸脱刺激の音程差に依存して変化する傾向が認められた．一方，音楽経験者では音程差以上にメロディーの輪郭（山形 or 谷形）の影響を受ける傾向が観察された．また，音楽経験者では非経験者に比べて MMF 潜時の短縮も認められた．

◆音楽経験により，音列を群化してとらえる能力が高まったことが示唆される．

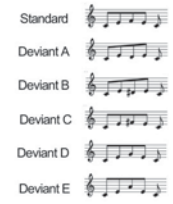



Fig.1. Tone sequences used in the experiment

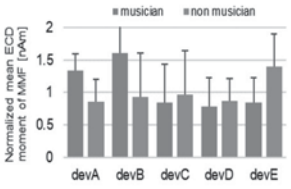


Fig.2. Comparison of ECD moment between the musician and the non-musician

2-11-16

2-11-16 外来待合室における BGM の使用と効果

Psychological effects of background music on a waiting room in a hospital or a clinic

○松本じゅん子（長野県看護大）

◆目的：病院の外来待合室で BGM を使用することで，不安定な感情に対してプラスに働きかけることが予想されるが，全ての病院やクリニックの待合室で BGM が使用されているわけではない．本研究では，病院やクリニックの待合室の BGM の使用状況について調べた．

◆方法：32 の病院及びクリニックに，調査協力を依頼した．外来待合室での BGM の使用の程度等について，質問紙を用いて尋ねた．

◆結果と考察：21 の病院またはクリニックから回答が得られた．外来待合室ではテレビの使用が多く，リラクセーション効果や退屈・疲労感の軽減等を目的としていることが示唆された（Fig1）．全体的にテレビや本によって退屈・疲労感を軽減させる一方，地域性や患者の特性により，BGM を使用しない状況も示された．

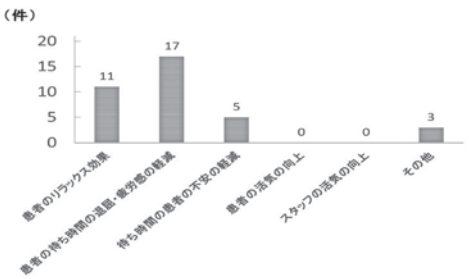


Fig.1：Aims of using background music on a waiting room in a hospital or a clinic.

2-P-1

2-P-1 フィードバック型能動騒音制御システムにおける事前推定適応フィルタのステップサイズ制御法に関する検討

A study on step size control method for pre-estimating adaptive filter in feedback type active noise control systems

○藤井健作（コダヴェイ研），△棟安実治（関西大），苣木禎史（千葉工大）

FB 型 ANC システムでは制御開始前に，帰還系の相殺と Filtered-x 法への適用用に帰還系を推定する処理が実行される．本検討では，この処理で生じる帰還系の推定誤差に異なる定義が求められることから，両定義に対応する推定処理が 1 つの適応フィルタで実行できる方法を提案する．係数更新にはブロック実行型学習同定法を用い，前者の相殺と後者の推定には以下のステップサイズ制御式をそれぞれ適用する．

$$\mu_n = \frac{2C_0 P_n J_0}{Q_n I + P_n C_0},\quad \hat{\mu}_n = \frac{2\hat{C}_0 \hat{g}^2 P_n J_0}{Q_n I + \hat{g}^2 P_n C_0}$$

また収束判定には異なる２つのステップサイズを適用した適応フィルタの係数の差の２乗和を以下のようにそれぞれ計算する．

$$d_n^c = \left(\hat{B}_n^1 - \hat{B}_n^2\right)^T \left(\hat{B}_n^1 - \hat{B}_n^2\right),\quad \hat{d}_n^c = \frac{\left(\hat{B}_n^1 - \hat{B}_n^2\right)^T \left(\hat{B}_n^1 - \hat{B}_n^2\right)}{\hat{g}^2}$$

以上の制御式を以下の手順に適用することで提案法が構成される．

(1) $\hat{g}^2 > 1$ならばステップサイズを$\hat{\mu}_n$とする

(2) $\hat{d}_n^c < \hat{D}_0$となった時点でμ_nに変更

(3) $\hat{d}_n^c \le \hat{D}_0/\hat{g}^2$となった時点で収束と判定

(4) $\hat{g}^2 \le 1$ならばステップサイズをμ_nとする

(5) $d_n^c < D_0$となった時点で$\hat{\mu}_n$に変更

(6) $\hat{d}_n^c \le \hat{D}_0$となった時点で収束と判定

2-P-2

2-P-2　ラウドネスと SNR に基づく雑音抑圧手法

Noise suppression method based on loudness and SNR

☆宮家一真，中島栄俊（熊本高専），藤坂洋一（リオン）

◆これまで補聴器の多くは音声明瞭度の向上やリスニングエフォートの削減といった配慮に欠けている．これらの事柄に対処する方法としてヒトの感覚量であるラウドネスに着目したノイズリダクション機能を提案してきた．

◆本報告では観測信号の sone 上での SNR(SNR_L)と観測信号のラウドネスから音声信号のラウドネスを推定し，観測信号のゲインを調節する雑音抑圧方法の提案を行う．

◆提案手法の雑音抑圧信号を評価するために，SNR_Lを推定した場合の雑音抑圧信号と，真値を用いてSNR_Lを算出した場合(SNRtrue)の雑音抑圧信号の２つを STOI 及び eSTOI の結果で比較した．

◆提案手法によって処理された音は，高い音声了解度を保持していることが期待できる．

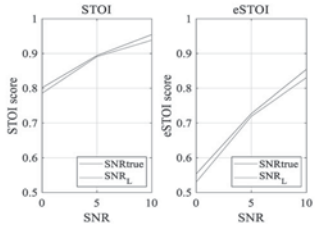


Fig.1：NR performance evaluation by using STOI and eSTOI.

2-P-3

2-P-3　End-to-end 雑音除去のためのネットワーク構造の検討

Investigation of network architecture for end-to-end denoising

☆蓮実拓也，小林哲則，小川哲司（早大）

◆概要

- 音源分離で用いられる Conv-TasNet（Fig.1）をシングルチャネル雑音除去に適用．
- CHiME-3 タスクにおいて既存のシングルチャネル雑音除去手法よりも高品質な音声を抽出．

◆実験結果

- Conv-TasNet は TCDAE-1ch と比較して，PESQ を 0.434 ポイント，単語誤り率（WER）を 45.4 ポイント改善．
- Conv-TasNet において，正解信号として目的音源と雑音を与えることで，目的音源のみを与える場合よりも音声の品質が向上．

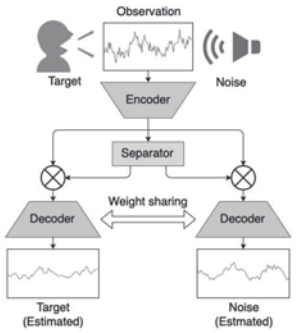


Table 1: PESQ and WER on CHiME-3 task.

Method	PESQ	WER (%)
TCDAE-1ch	1.508	95.70
Conv-TasNet -targ	1.942	50.30
Conv-TasNet -targ+dist	**2.151**	**38.63**

Fig.1: Conv-TasNet for noise reduction.

2-P-4

講演取消

2-P-5

2-P-5　光レーザマイクロホンを用いた位相情報を考慮した深層学習による高域復元

Higher frequency reconstruction based on deep learning adding phase information using optical laser microphone

☆上田奨士（立命館大院），福森隆寛，西浦敬信（立命館大）

光レーザマイクロホンは対象物体にレーザ光を照射し，その微小な振動を計測することで音を収録するマイクロホンであり，遠方音声の計測に有効である．しかし，その受音音声は振動物体の材質や計測時の環境により，Fig. 1 に示すような高域減衰や定常雑音などの音質劣化が発生する．このとき，音質改善の手法として，深層ニューラルネットワーク（DNN）が提案されている．しかし，従来の DNN では劣化音声の位相を用いるため，時間波形復元後の音質が不十分である．この問題を解決するために，本稿では Fig. 2 のように複素スペクトルを DNN の入出力に用いることにより，位相情報を考慮した音質改善手法を提案した．提案法の有効性を示すために，Fig. 1 の音質劣化を含む光レーザマイクロホンによる収録音声に従来法と提案法を適用し，Wideband PESQ による評価を実施した．Fig. 3 より，提案法は５種類の音質劣化において従来法と同等以上のスコアであることから提案法の有効性を確認した．

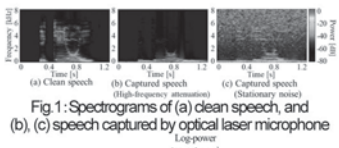


Fig.1：Spectrograms of (a) clean speech, and (b), (c) speech captured by optical laser microphone

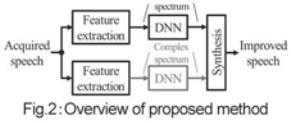


Fig.2：Overview of proposed method

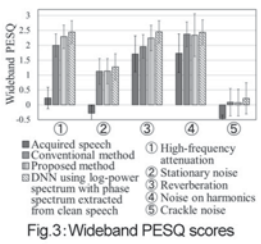


Fig.3：Wideband PESQ scores

2-P-6

2-P-6　光レーザマイクロホンのための Residual Network を用いた CNN による雑音除去

Noise reduction based on CNN with Residual Network for optical laser microphone

☆CAI Chengkai(立命館大院), 福森隆寛, 西浦敬信, 山下洋一（立命館大）

光レーザマイクロホンはレーザ光を利用し，物体の振動を計測することにより音声を収録するシステムであり，遠方の音声が収録できる．しかし，光レーザマイクロホンによる収録音声は，雑音の混入や音声構造の欠落が発生し，音質が劣化する．従来，音声強調において様々な手法が提案されているが，これらの手法は多様な歪みを含む音声において不十分である．本稿では，光レーザマイクロホンによる収録音声に対して，二段階の処理により音声強調する手法を提案する．まず，畳み込みニューラルネットワーク(CNN)により雑音除去を行う．ここで，勾配消失を防ぐため Residual Network 構造を利用する．次に，雑音除去した音声を用いて回帰型ニューラルネットワーク(RNN)により高域情報を復元する．提案法の結果を Fig. 1 に示す．提案法の有効性を確認するために，強調後の音声を PESQ，LSD，STOI の 3 つの指標による評価実験を実施した．Table 1 より全ての指標全てにおいて提案法が従来法を上回ることを確認した．

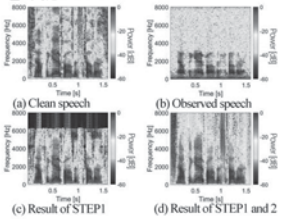


Fig. 1 Samples obtained by STEP1 and STEP2

Table 1 Evaluation results

	Observed	STEP1 only	STEP2 only	STEP1 & STEP2
PESQ	1.76	2.36	2.31	**2.41**
LSD[dB]	1.62	1.69	**1.09**	**1.09**
STOI	0.85	**0.95**	0.93	**0.95**

STEP1: Input speech pre-processing for noise reduction based on CNN
STEP2: Traditional method of speech enhancement based on RNN
STEP1 & STEP2: Proposed method of speech enhancement

2-P-7

2-P-7　区分的凸関数近似に基づく多方向ゲイン制約ビームフォーマの設計

Beamformer design methods with multiple gain constraints based on piecewise convex approximation

◎佐藤遼太郎，丹羽健太，中川朗（NTT メディアインテリジェンス研究所）

問題設定　多方向に目的音源・雑音源がある状況でのビームフォーマ

- ■ 無歪制約などの厳しい線形拘束を多数課すと，性能低下
- ■ 制約を緩和させたいが，非凸な形式に緩和させてしまうと緩和後の最適化問題を解くのが難しい

提案法

- ✓ 無歪制約の代わりに，位相の自由度を持つ不等式制約の導入

$$\mathbf{w}_f^{\mathrm{H}}\mathbf{a}_{f,d}=1 \;\Rightarrow\; |\mathbf{w}_f^{\mathrm{H}}\mathbf{a}_{f,d}|\geq 1$$

目的音無歪制約(線形，凸)　⇨　提案：目的音強調制約(非凸)

- ✓ 更に，位相の値の範囲 $[0,2\pi)$を分割し，各領域で非凸制約を凸関数で近似する（**区分的凸関数近似**）と，フィルタ設計問題は

$$\min_{\{\mathbf{c}_f,\mathbf{w}_f\}_{f=1}^{F}} \sum_{f=1}^{F} L_{\mathrm{MV}f}(\mathbf{w}_f)+\sum_{f=1}^{F}\sum_{d=1}^{D}\Lambda_{(f,d),\mathbf{c}_{f,d}}(\mathbf{w}_f^{\mathrm{H}}\mathbf{a}_{f,d})+\sum_{f=1}^{F-1}\widehat{L}_{\eta f}(\mathbf{c}_f,\mathbf{c}_{f+1})$$

出力信号分散(MV)　目的音強調制約（区分的凸関数近似）　“フィルタ応答の位相曲線の長さ”（正則化）

- ✓ これは凸最適化アルゴリズムと組合せ最適化（最短経路問題）アルゴリズムの併用により実用的な時間で解ける

適用例　LCMV では困難な状況で，複数音源強調性能改善を確認

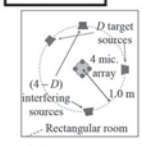


Fig. 2: Simulation setup.

Table 2: Comparison of mean signal-to-interference-and-noise ratio (SINR) improvements and mean scale-invariant signal-to-distortion ratio (SI-SDR) improvements for each target signal between conventional and proposed methods.

Criterion	SINR [dB]			SI-SDR [dB]		
# of constraints	1	2	3	1	2	3
Conv. (LCMV)	7.63	2.51	0.43	7.84	2.91	0.48
Prop. ($\eta=0$)	7.22	3.18	**1.40**	7.43	3.29	**1.14**
Prop.	**7.68**	**3.46**	1.17	**7.89**	**3.73**	1.07

2-P-8

2-P-8　動的な同期加算による微弱周期信号の強調

Enhancement of weak periodic signal by dynamic synchronous accumulation

☆住吉亨介,若林佑幸,シャイブラーロビン,小野順貴(首都大)

- ◆同期加算とは，複数の信号を，それらに含まれる目的信号に関して同期させ足し合わせる処理である．周期信号の場合，時間的に異なる周期部分を複数信号として用いることで，同期加算を行うことができる．
- ◆同期加算により得られる信号強調のゲインは同期加算をとる信号の数を増やすほど大きくなるため，長時間信号で同期加算を行えば高いゲインを得ることができ，微弱な信号も検出できる可能性がある．
- ◆しかし，機器によるサンプリング周波数の微小なミスマッチや，長時間録音によって起こりうる微小の時間的な変動が，同期加算による信号強調の性能低下の要因として考えられる．
- ◆本研究では，サンプリング周波数が時間的に微小に変動しても，動的な同期加算を行うことにより，長時間録音においても高精度の同期加算を実現する枠組みを提案する．
- ◆SN 比が低い場合でも提案法を用いることで同期加算が可能となり，対象音響信号を強調できることが確認できた．(Fig. 1)

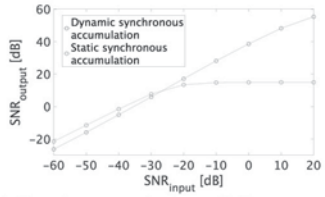


Fig.1 : Experiment results (input SNR vs output SNR)

2-P-9

2-P-9　ニューラルネットによる dequantize

Dequantize by neural network

○江村 暁†、山本 剛志‡、福井 勝宏†、原田 登†、中川 朗†
† NTTメディアインテリジェンス研究所、‡ 大分大学

センサ信号は、 A/D変換等により量子化(quantize) されてから、 デジタル信号処理される。 量子化された信号の下位ビットを拡張したいというニーズに対し、本稿ではトレーニングデータとして低ビット量子化の前と後の信号が与えられるとの想定をおき、ニューラルネットによるdequantizeを検 討した。

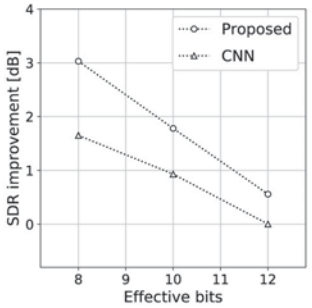


Fig. 1 入力信号の有効ビット数とsignal-to-distortion ratio(SDR)

2-P-10

2-P-10 自動車加速音における結合音適用の検討

Effect of Distortion Products in Automobile Acceleration for Sound Design

☆山際納月，石光俊介，谷本典之（広島市大院）
藤川智士，松本貢典（マツダ），村上泰樹（大島商船高専）

◆ 低周波数の強調と車室内の静粛性との両立を考慮したエンジン音のサウンドデザインとして，結合音を利用する．

◆ 加速音に結合音を適用するための基礎検討として，エンジン音に近い周波数のチャープ音を用いて音高を比較した．

◆ 実験１では，f_1 を 30~120Hz，f_2 を 40~180Hz のチャープ音（6 秒間）として 15~60Hz の結合音を生成した．MUSHRA 法により複合音の音高を比較した結果，f_2のレベルを下げると結合音の認識が示唆された．

◆ 実験２では，f_1 を 40~70Hz，f_2 を 60~105Hz のチャープ音（2 秒間）として 20~35Hz の結合音を生成し，MUSHRA 法により音高を比較した結果，f_2を65dB で呈示すると最もよく結合音が認識できることがわかった．

◆ 呈示レベルを適切に設定することで，加速音においても結合音による低周波音のデザインが可能であることが示唆された．

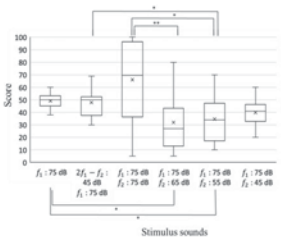


Fig.1 Experimental results 1

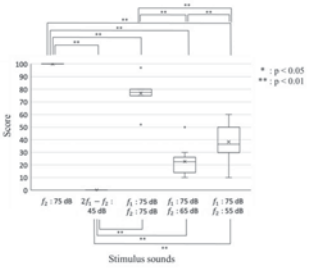


Fig.2 Experimental result 2

2-P-11

2-P-11 RU Multichannel Domestic Acoustic Scenes 2019: 複数デバイスで構成された分散マイクアレイによる音響シーン分析のための環境音データセット

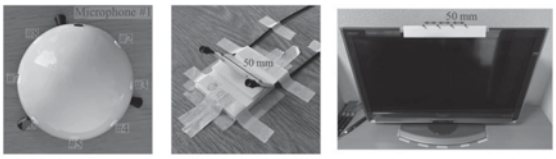
RU Multichannel Domestic Acoustic Scenes 2019: multichannel dataset for acoustic scene analysis recorded by distributed microphones with various devices

〇井本桂右（立命館大学），小野順貴（首都大学東京）

◆音響シーン分析のための新たなデータセットを構築

◆データセットは以下の特徴を持つ

- 異なる特徴を持つ複数のデバイスに搭載された 42 の非同期分散マイクロホンアレイを用いて収録を実施
- 1,996 時間（47.5 時間×42 時間）の環境音を収録
- 実際に人が住む実環境下で収録を実施
- プライベートなシーンである「toilet」「sleeping」「taking a bath/shower」を含む 21 のシーンを収録
- 空間情報を用いた音響シーン分析や，アンサンブル学習，ドメイン適応手法の評価などの利用を想定



(a) TAMAGO-03　(b) MX150B/O　(c) MX150B/O (mounted on TV)

Fig.1:　収録に用いたマイクロホンの例

2-P-12

2-P-12 ToyADMOS: 異常音検知のための おもちゃ稼働音データセット

Introduction of ToyADMOS: A Dataset for Anomalous Sound Detection

〇小泉悠馬，齊藤翔一郎，植松尚，原田登（NTT），井本桂右（立命館大）

■ 機器異常音検知用の公開データセットである ToyADMOS（anomaly detection in machine operating sounds; ADMOS）を紹介する

■ おもちゃの機械を意図的に損傷させることで，540 時間を超える正常音と，11 時間を超える異常音を収集した

■ データセットとテストコードは以下の URL からダウンロードできる

Dataset: https://zenodo.org/record/3351307#.XgRs_Ff7SUk

Python code: https://github.com/YumaKoizumi/ToyADMOS-dataset

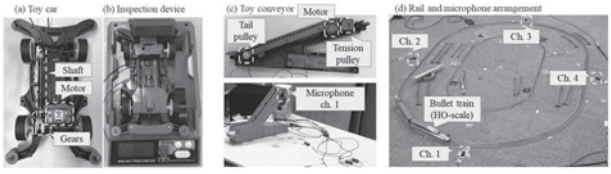


Fig. 1 Images of toys, parts, and microphone arrangement

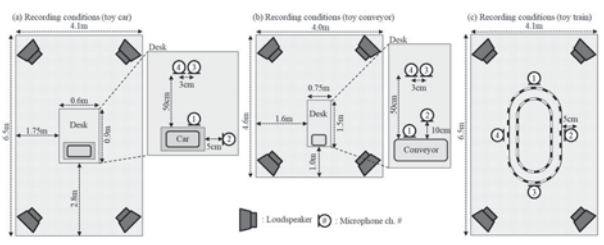


Fig. 2 Setup-positions of all equipment (top view)

2-P-13

2-P-13 類似環境音に対する識別手法の検討

Investigation of clacation methods for similar environmental sounds

☆山口彪貴（静岡大院・総合科学技術研），
竹内太法（静岡大・創造科学技術大学院），
立蔵洋介（静岡大院・総合科学技術研）

◆目的

音響的特徴が類似し，人間が聞き分けづらい音「**類似環境音**」を高精度で識別する．
→ 音による物質判定や揚げ物の揚げあがり判定などに応用可能

◆検討手法

- LFCC
- CNN, DNN

◆評価方法

- 9 クラス識別テストによる識別率の比較 (Fig. 1)
- 未知データを含めた 3 クラス識別テストによる適合率，再現率，F 値の比較

◆結果

- LFCC を入力に用いた場合，識別率は低下
- 最も高い識別結果はスペクトログラムを用いた CNN による識別
- 未知データに対する性能も CNN が比較的良好

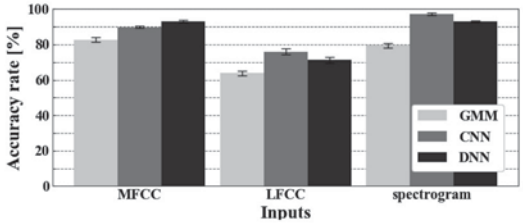


Fig. 1: Comparative results of 9-class classification

2-P-14

2-P-14　環境音とラベル情報のデュアルエンコーダに基づくスペクトログラムマスクの生成

Generating spectrogram mask for environmental sounds based on dual encoder of sounds and labels

○竹内博俊（東大院・情報理工），大石康智（NTT・CS 研），△川西隆仁（NTT・CS 研），柏野邦夫（東大院・情報理工，NTT・CS 研）

◆本研究では，複数の環境音を含む音響信号に対して特定の環境音に対応するスペクトログラムマスクを生成するシステムを提案する．

◆学習データは**複数の環境音を含む音響信号とそのラベル集合のみ**

クラス分類の問題とすると多様な環境音に対する学習が困難

→ ここでは音響信号とラベル情報のペアデータに基づく距離学習の問題として扱う

◆ラベルと音響信号それぞれにニューラルネットワークに基づくエンコーダを用意し，共通の潜在空間へ写像．潜在空間内でのペアデータの類似度を大きく，ペアでないデータの類似度をするように学習する．

- **dry source がなくても学習可能**
- 同時に音源種間の類似性を知ることができる

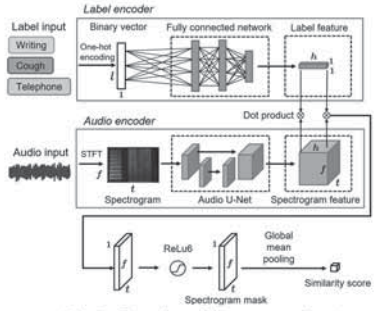


Fig.1：Overview of the proposed system

2-P-15

2-P-15　音光変換デバイス「ブリンキー」による音高のリアルタイム可視化

Pitch estimation and visualization using the sound-light conversion device blinky

☆石井奏人，堀池大樹，シャイブラーロビン，
若林佑幸，小野順貴（首都大）

◆我々は，音光強変換デバイス「ブリンキー」とビデオカメラを用いて広範囲の音強度情報を得ることによる新しい音響センシングの枠組みを研究している(Fig. 1)。

◆本研究では，新たな応用の初期検討として音高の可視化を試みた。具体的には，ブリンキーに搭載されているマイクロコントローラによりフーリエ変換を行い，クロマベクトルを計算することによって音高を推定した。

◆実環境でC，C#，…，Bの楽器音を再生し，正しく音高が推定できているかを検証した結果，音高を高い精度で推定することができた。

◆Fig. 2に再生した楽器音の音高と，ブリンキーで推定した音高を示す。

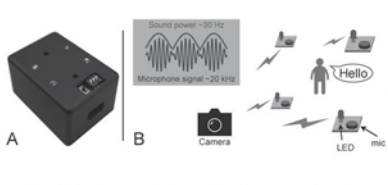


Fig. 1 The picture of blinky (A) and the sound-to-light conversion architecture (B).

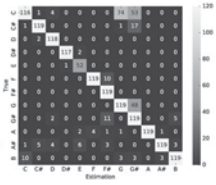


Fig. 2 The results of pitch estimation for piano tones.

2-P-16

2-P-16　豚耳標における体内伝導音からの周期音抽出による生体情報抽出に関する検討

Study on bio-information extraction by periodic sound extraction from body-conducted sound in pig ear tags

○土屋和也（広島市大），成澤健太，石光俊介，森長佳子，中山仁史（広島市大院）
三上 修，高木 道浩，井上 寛暁，石田 三佳（農研機構）

一般的に家畜は畜舎内で群飼養されることが多く，呼吸器病などの罹患発見が遅れた場合，疾病の重症化や蔓延が起こる．我々は，騒音に強い体内伝導音に着目し，個体ごとの家畜の体内伝導音による罹患の判定手法について検討している．罹患時には，心拍数や呼吸回数などの生体信号に変化があらわれる．このような変化を把握することにより，早期に罹患判定をすることが可能と考える．この中で，体内伝導音から心拍や呼吸の情報を抽出する必要性が浮上してきた．そこで，本発表では，独立成分分析(Independent Component Analysis)と適応信号処理(Adaptive Line Enhancer)を用いて体内伝導音から呼吸音と心拍を抽出することを試みた．以下にALEによる周期的な波形の抽出結果を示す．

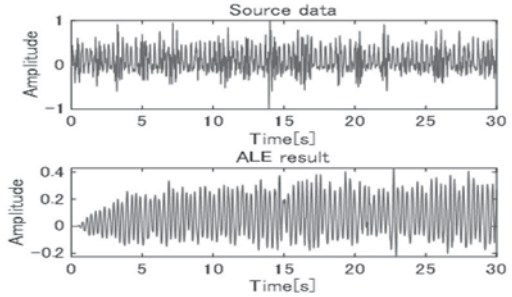


Fig.1 : Waveform after Adaptive Line Enhancer
(Upper graph : body-conducted signal(Source data),　Lower graph : Signal after ALE processing)

2-P-17

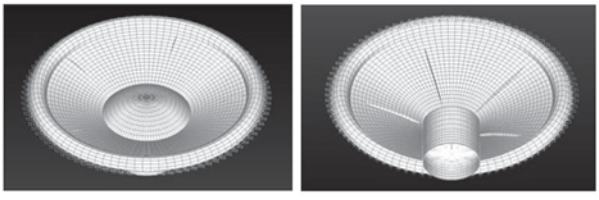
2-P-17　振動-音響解析によるスピーカ振動板形状の改善

Improvement of loud-speaker diaphragm shape based on vibro-acoustic analysis

○田中康治，△江崎雄也，△仲田剛（三菱電機）
△須田 隆亮，△近藤 誠浩（三菱電機エンジニアリング）

NCV（Nano Carbonized high Velocity）を使用した車載用スピーカの開発を進めている。NCV は、カーボンナノチューブを配合した樹脂素材であり、金属並みの伝播速度と紙並みの内部損失を有し、スピーカの振動板素材として理想的な特性を備えている。しかしながら、他のスピーカと同様に、高周波数域では分割振動による音圧周波数特性のピーク／ディップが生じ、振動板形状の最適化によって、フラットな周波数特性を実現することが必要となる。

本検討では、振動-音響解析を基にして、スピーカ振動板形状の改善を行い、その結果、元の振動板形状よりも 0.9kHz 広い周波数範囲において、ピーク／ディップの差を 12.4dB 縮小し、よりフラットな周波数特性が得られる振動板形状を導出した。



(a) 表側　　(b) 裏側

Fig.1 改善後の振動板形状

2-P-18

2-P -18　シャント音血管狭窄検出における伸縮処理を加えた特徴量算出法

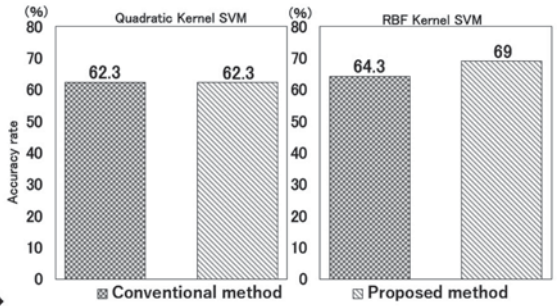
A method of feature calculation with expansion and contraction for shunt murmur angiostenosis detection

東大輔，○西島恵介，古家賢一（大分大）

◆腎不全などの患者は日々のシャント管理が求められるが、シャント機能の判断は主観に依存してしまうため、シャント機能を手軽に自動で判断するシステムが求められている。

◆本報告では、シャント音の個人差によって、データごとにバラつきが生じることで特徴量算出に影響することが考えられた。これを解決するために、伸縮処理を加えたデータを作成し、特徴量算出を行うこととした。今回は、従来手法のサンプル移動処理に加え、伸縮処理を加えたデータ拡張を行い、時間周波数解析を用いた二次元特徴量算出を提案する。

◆実験結果より、提案手法を用いることで精度の向上が確認され、従来手法と比較して最大4.7%の精度向上が確認された。



◆

Fig.1 : The result of SVM detection

2-P-19

2-P-19　複数位置シャント音特徴量を用いた狭窄の識別

Classification of stenosis using multi positions shunt sound features

○西島恵介，東大輔，古家賢一（大分大）

◆血液透析患者はシャントを管理し，狭窄を早期に発見する必要がある。

◆狭窄する位置は様々なため，狭窄位置に関わらず狭窄を識別する手法が必要である。

◆複数位置の聴診から得られたシャント音を用いて，時間周波数領域において，ある位置のパワースペクトルから残りの位置のパワースペクトル平均を減算する特徴量を提案し，狭窄の識別を試みた。

◆識別にはサポートベクターマシンを用い，最大 6 箇所で計測された 20 症例（正常 7，狭窄 13）のデータを用いた。

◆提案特徴量で狭窄位置を推定し，推定した位置での識別（従来法）と各位置での提案特徴量を用いた識別（提案法）を比較した。

◆精度の計算には 2 分割交差検証を用いた。

◆従来法の精度 45%に対し，提案法は精度 60%と 15%向上した。

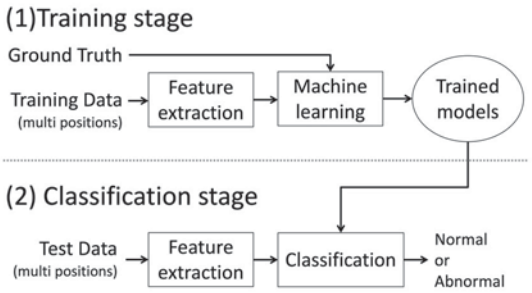


Fig. 1 Classification method

2-P-20

2-P-20　CNN による船舶航行音による船種分類に向けたデータ拡張法の検討

Investigation of data arguments to train the boat noise using convolutional network for boat type classification

☆山口晴己，武藤憲司（芝浦工大），小林洋介（室蘭工大）

◆本研究の目的は音と視覚情報の主観評価を利用した運河沿い住民への船舶騒音印象軽減のための船舶接近通知システム開発である.

◆先行研究において船種分類を行った結果，船種ごとの認識において学習量の差が識別性能に影響を及ぼしていた.

◆本研究では船種ごとの学習量を一定とすることで学習性能による認識性能への影響を調査した.

◆CNN モデルの学習より 1 時間の環境音に対し船種分類を行った結果を Fig.1 に示す．10 隻中 9 隻が所望する分類と同一の結果となった.

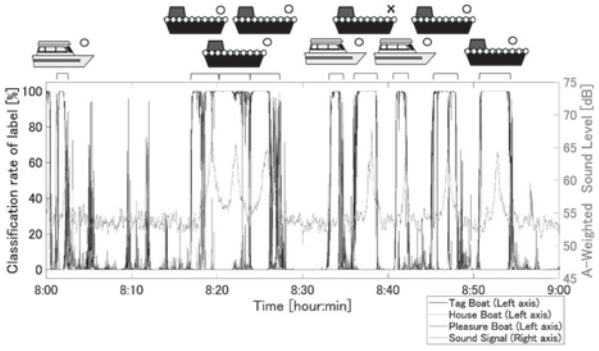


Fig.1 : Result of boat type classification by CNN model.

2-P-21

2-P-21　ニューラルネットワークを用いた走行音による交通車両の速度推定

Sound-based Speed Estimation Using a Neural Network

☆篠原拓実，若林佑幸，シャイブラーロビン，小野順貴（首都大），
△相澤範明（ネクスコ・エンジニアリング東北），△中川浩（NEXCO 東日本）

◆交通量調査には、現在使用されている車両感知器よりも設置・運用コストが安価なシステムの開発が望まれている。

◆そこで我々は、比較的低コストの交通量観測を実現する音響情報を用いた手法を開発するため、研究を行っている。

◆これまで、円滑な交通状況で収録した車両走行音を用いてニューラルネットワークによる通過台数の推定を行い、高い精度での通過台数推定が行えることを確認している。

◆本稿では、高速道路で Fig. 1 に示すようなセットアップで渋滞を含む交通状況で車両走行音の収録を行い、音響特徴量（パワーと到来時間差）を抽出し、ニューラルネットワークを用いた深層学習によって、通過車両の速度を推定する実験とその評価を行った。

◆Fig. 2 に示す推定結果より、速度の速い車両と遅い車両の区別がおおむね行えたため、渋滞の判別ができる可能性を示すことができた。

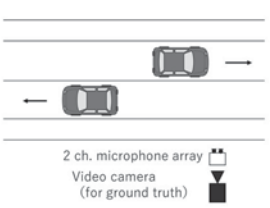


Fig.1: Setup for recording vehicle sound

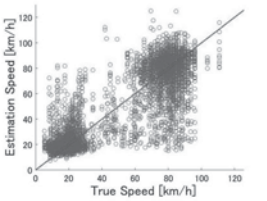


Fig. 2: Result of speed estimation

2-P-22

2-P-22　腹部多チャネル録音による腸蠕動音伝搬特性の検証

Verification of bowel sound propagation characteristics by abdominal multi-channel recording

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

◆腹部聴診により聴取される蠕動音は，腸機能の評価，診断に用いられている．

◆蠕動音の発生源推定に関する先行研究では，腹腔内を均質，かつ蠕動音が単一の局在音源から発生するというシンプルなモデルを用いている．本研究では，ヒト腹部の多チャネル録音の結果から，蠕動音の到達時間差を算出し，従来のシンプルなモデルが成り立つか否かを検証した．

◆腹部４点を同時録音（1 ch～4 ch，Fig. 1）．健常者５名の腹部聴取音を計 10 分取得

◆計測した腹部聴取音を 40 ms ごとに区切り（それぞれをブロックと呼ぶ），蠕動音の音源数が１つと判定された 61 個のブロックについて，2 ch と他チャネルの時間差を算出 (Fig. 2)．

◆音信号が 2ch，4ch，3ch, 1ch の順に到達するブロックが最も多くなった．この結果は従来の様なシンプルなモデルでは説明が困難である．例えば，複数の音源を仮定したより複雑なモデルが必要と考えられる．

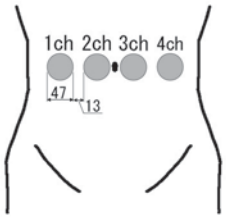


Fig. 1 Stethoscope locations on the subject's abdomen

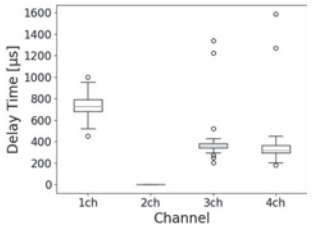


Fig. 2 Relative delay time for each channel (Ch 2 is used as a reference).

2-P-23

2-P-23　2ch マイクロホンを用いた対象物までの距離推定（音源とマイクロホンでできる線分に対して垂直方向に対象物を設置した場合）

Distance estimation from microphones to targets using 2ch microphones (in the case with the target set vertically to the line segment formed by the sound source and microphone).

☆田井 一，篠原 寿広，上保 徹志，中迫 昇(近畿大)

◆様々な工学分野において対象物までの距離情報は重要であり，送信波と反射波の位相干渉に基づく音響測距法が提案されている．

◆従来の音響測距法では，音源，マイクロホン，対象物が一直線上に存在すると想定し，マイクロホン-対象物間距離を測定していた．

◆しかし，実環境において，対象物が一直線上に存在しない場合，正しい測距結果が得られるとは限らない．

◆本研究では，1 回の測定で距離推定を行うことを目的として，今回は，極端な場合を想定し，音源とマイクロホンでできる線分に対して垂直方向に対象物を設置し，音響測距を試みた．

◆Fig. 1 に装置の配置図を，Fig. 2 に対象物までの距離を示す距離スペクトルを示す．Fig. 2 より，真値付近にピークを確認することができた．実験結果より，提案手の有効性を確認した．

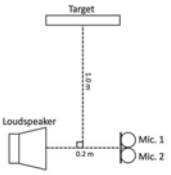


Fig. 1: Overview of acoustic distance measurement

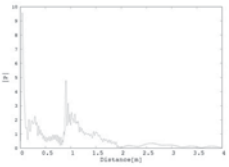


Fig. 2: Range spectrum

2-P-24

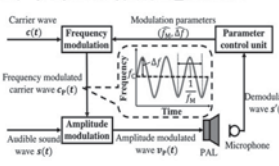
2-P-24　周波数変調キャリアを用いたパラメトリックスピーカの時間変動追従型周波数ピーク雑音低減

Spectral Peak Noise Reduction with Time Fluctuation Tracking for Parametric Array Loudspeaker Using Frequency Modulated Carrier Wave

◎森海里(立命館大院)，旭浩平（立命館大），中山雅人(阪産大/立命館大)，
西浦敬信（立命館大）

◆パラメトリックスピーカにおける従来の変調方式では超音波素子の共振周波数を振幅変調波におけるキャリア周波数として用いるが，これにより超音波素子の疲労破壊を誘発し，目的音ではない周波数ピーク雑音が発生する．この問題に対して我々はこれまで周波数変調キャリアを用いた周波数ピーク雑音低減手法を提案してきた．しかしこの手法では，事前に変調パラメタの最適化を行うため，時間経過とともに周波数ピーク雑音の低減性能が低下する．

◆そこで本稿では，Fig.1 に示すように一定時間毎に変調パラメタの更新を行うことで，経時的に周波数ピーク雑音を低減することを検討する．評価実験では，従来手法と提案手法によるキャリアを 30 分間放射した際の周波数ピーク雑音の音圧を比較した．Fig.2 に示した評価実験の結果，経時的に周波数ピーク雑音を低減可能であることを確認し，提案法の有効性を示した．



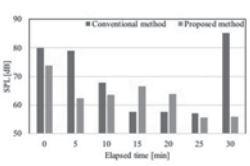


Fig.1: Overview of the proposed method.　Fig.2: Power of the spectral peak noise.

2-P-25

2-P-25　骨導パラメトリックスピーカにおける FM 変調波の位相制御に基づく復調音圧の改善

Demodulated Sound Pressure Improvement Based on the Phased Array Processing of the FM Wave for Bone-conducted Parametric Array Loudspeaker

◎藤井敏弘(立命館大院)，中山雅人(阪産大/立命館大)，西浦敬信(立命館大)

◆骨伝導で呈示された超音波（骨導超音波）であれば，音として知覚可能であることが知られている．しかし，知覚できる可聴音の音圧が小さいという課題がある．そこで本稿では，知覚できる可聴音の音圧を改善するため，骨導パラメトリックスピーカを提案する．Fig. 1 に示すように，骨導パラメトリックスピーカは受聴者の耳を囲んで複数の超音波素子を配置する．そして，骨導超音波知覚には蝸牛が大きく関与している点に着目し，骨導パラメトリックスピーカにおいて放射する FM 変調波が蝸牛の位置で同相となるように位相制御を行う．

◆提案手法の有効性を確認するため，復調音の音圧に関する評価実験を行った．本実験では，DSB，SSB，FM 変調方式において，従来の 1 素子で呈示した場合，位相制御なしで 15 素子を用いて呈示した場合，位相制御ありで 15 素子を用いて呈示した場合（提案手法）において音圧を評価した．評価実験の結果，Fig. 2 のように提案手法を用いることで復調音圧を向上可能であることを確認した．

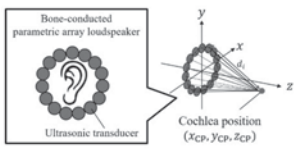


Fig.1 : Overview of the proposed bone-conducted parametric array loudspeaker.

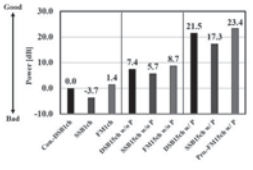


Fig.2 : Power of the demodulated audible sound.

2-P-26

2-P-26　パラメトリックスピーカを用いたミッシングファンダメンタルに基づく音楽再生音の強調

A Study on Sound Enhancement for Music Using Missing Fundamental in Parametric Array Loudspeaker

◎耿毓庭(立命館大院)，中山雅人(阪産大/立命館大)，西浦敬信(立命館大)

◆パラメトリックスピーカ(PAL)は再生の原理上，低域の再生が困難である．我々はこれまでミッシングファンダメンタルに基づき，PAL の高調波歪みを利用した音圧改善を提案してきた．そこで本稿では，その手法を音楽に拡張し，PAL による音楽再生音を強調する．提案手法では、低域抑圧を用いて復調音の音圧を改善する．PAL の周波数特性と高調波歪みの特徴を考慮して，高調波歪みを有効に利用し，調波成分を制御する．音楽の調波成分と非調波成分を分けて，それぞれ提案手法，従来手法で処理して変調し，PAL から放射する．

◆評価実験では，マルチトラックで録音した歌声のない音楽を音源とした．音圧(Z 特性，A 特性)に関する評価実験の結果は Fig.1 に示す.スペクトル歪み(0~20 kHz，1~20 kHz)に関する評価実験の結果は Fig.2 に示す.以上の結果で，歪みの増加を代償に，PAL による音楽再生音の強調が可能であることを確認した.

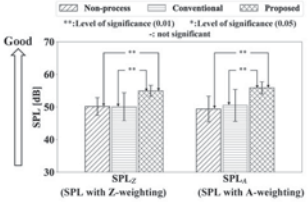


Fig.1：Experimental results for sound pressure.

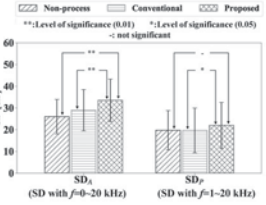


Fig.2：Experimental results for spectral distortion.

2-P-27

2-P-27　パラメトリックスピーカのための逐次音線追跡法に基づく復調音の可視化

Visualization of Demodulated Sound Based on Sequential Acoustic-ray-tracing for Parametric-array Loudspeakers

☆下方誠(立命館大院)，中山雅人(阪産大/立命館大)，西浦敬信(立命館大)

◆従来，音の可視化手法として，音線追跡法が提案されている．しかし，音線追跡法は１次音源のみを想定しているため，パラメトリックスピーカ(Parametric-array Loudspeaker: PAL)のように空気中で２次音源が生成される復調現象の表現は困難である.

◆そこで本稿では，空間と時間を離散化した逐次音線追跡法を利用して PAL における復調音の可視化手法を提案する．逐次音線により，２次音源を離散的にそれぞれ生成し，線形モデルを用いて復調現象を近似する．具体的には，音源から放射する音線の方向を PAL の指向角に限定し，音線を放射後，Fig.1 に示すように，所望の反射回数に達するまで，フレームを更新し，逐次音線を生成する．そのとき，逐次生成した音線に対して，復調音圧を算出する．本稿ではモデル化に向けて，まず実測値を用いて，２次音源を足し合わせた復調音圧をシミュレーションする．これにより，復調音の可視化を実現する.

◆Fig.2 のように，提案手法を用いることで，復調音をおおよそシミュレーションできたことより，提案手法の有効性を確認した.

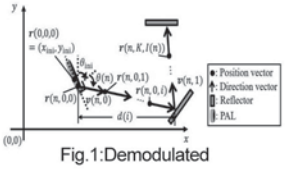


Fig.1:Demodulated sound vector update

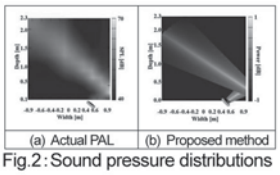


Fig.2：Sound pressure distributions

2-P-28

2-P-28 反射音環境下におけるスピーカ特性測定の検討

A study on loudspeaker characteristic measurement under reflective environment.

☆森谷晃行，金田豊 （東京電機大）

◆スピーカ周波数特性の測定には通常，無響室や無響箱を用いる．これらの設備は高価であるため，一般室内で測定を行う手段が望まれる．しかし，反射音環境下では周波数特性が劣化する(Fig.1 灰細線)．

◆反射音の影響を低減する方法として，反射音が到達するまでの直接音（先頭部分）を切り出す方法があるが，低域の特性が劣化する(Fig.1 黒破線)．

◆別の方法として，先頭部分を使って反射音を打ち消す方法がある．しかし，スピーカのインパルス応答は低域の立ち上がりが遅く，従来の先頭部分切り出しでは低域の情報が欠落するため，低域反射音の推定はうまくいかない．

◆そこで，時間－周波数分析を利用して周波数ごとに先頭部分を切り出し(Fig. 2 黒太線)，反射音打ち消しを行う手法を提案した.

◆反射音環境でインパルス応答を測定し，提案法による反射音低減を行った．Fig.1 黒実線に示した結果より，先頭部分を切り出す場合と比較して 200~600 Hz 付近の特性を回復できることを確認した.

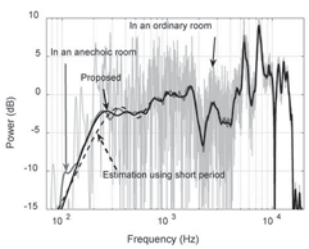


Fig.1 Frequency characteristic measurement results of a loudspeaker.

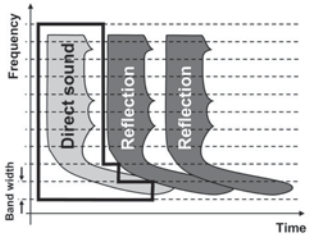


Fig.2 Extraction of direct sound by the proposed method.

2-P-29

2-P-29　最小位相を基準とした位相および群遅延特徴量の考察

Study on Phase and Group Delay Features based on Minimum Phase.

☆樋貝満喜，△矢内駿介，島内末廣，青木茂明（金沢工大），

◆本報告では、複雑な位相の形状が最小位相あるいは最大位相成分と関連が強いとの仮説のもと、位相情報あるいは群遅延の基本性質を再考した結果を述べる。

◆例えばFig. 1 は源波形を最小位相化した例を示すが、この例では (b) の場合で波形が時間反転している。この時Fig. 2 に示すように源波形とその最小位相波形との位相差を比べると時間反転が起きていない位相差は単調だが、時間反転がおきた場合の位相差は複雑な形状を持つことがわかる。

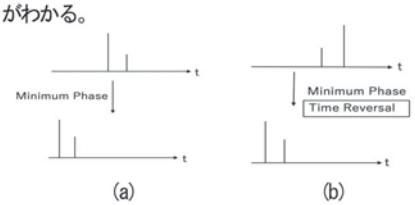


Fig.1　Phase Minimization

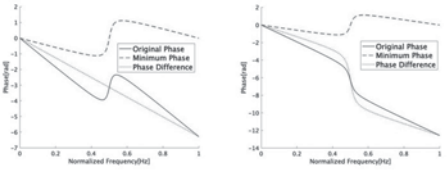


Fig.2　Phase Properties

2-P-30

2-P-30　豚の体内伝導音から抽出した心拍情報による罹患判定手法の検討

Detection Method of Infected Pigs by Heart Rate Analysis Extracted from Body-Conducted Sound

☆成澤 健太（広島市大院），土屋 和也（広島市大），石光 俊介，森長 佳子，中山 仁史，飯島 聡志（広島市大院），△三上 修，△高木 道浩，△井上 寛暁（農研機構）

◆豚の体内伝導音の解析を行い，罹患判定手法の検討を行った．

◆健常豚の体内伝導音に対して，ウェーブレット変換による心拍の抽出を行い，その心拍変動解析から豚での LF, HF 成分の周波数帯域を同定した．

◆次に，豚に PRRS*ウイルスを接種し，ウイルス接種前，接種 3, 5, 7, 10 日後に体内伝導音の収録を行い，同様に心拍変動解析を試みた．

◆スペクトル解析および変動係数による心拍変動解析の結果，罹患前後における心拍変動は確認されなかった．

* PRRS : Porcine Reproductive and Respiratory Syndrome，豚繁殖・呼吸障害症候群

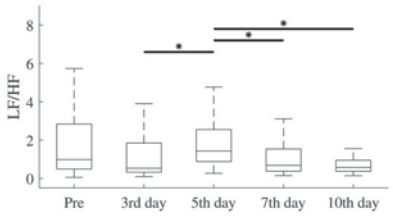


Fig. 1 ： Descriptive statistics for LF/HF value

2-P-31

2-P-31　パラメトリックスピーカとビームステアリングを用いた音響測距

Acoustic Distance Measurement Based on the Phase Interference Using Audible Sound Component with Parametric Loudspeaker and Beam Steering.

☆北野 翔子，篠原 寿広，上保 徹志，中迫 昇（近畿大）

◆超音波による指向性の鋭い周波数を搬送波として，必要な情報を変調波として音波を放射し，可聴音を再現するパラメトリックスピーカと，遅延を用いることによって任意の方向へ音声情報を伝搬するビームステアリングが知られている．

◆特定の人にだけ音声情報を伝送するためには，その人の位置・方向・距離などを知る必要がある．

◆距離を測定する方法として，可聴音域において送信波と反射波の位相干渉に基づく音響測距法が提案されている．

◆本研究では，ビームステアリングによってパラメトリックスピーカの指向性を制御し，音響測距を用いて対象物までの距離を測定する．

◆実験では，送信波を SSB 変調し指向性を制御したものを使用し測距を行った．Fig. 1 に距離スペクトル，Fig. 2 に指向性を確認した実験結果を示す．真値 1[m]付近にピークが表れており，また設定した目的方向 30 度にピークが表れている．実験結果より，提案手法の有効性を確認した．

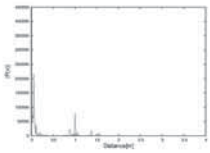


Fig.1 : Range spectrum.

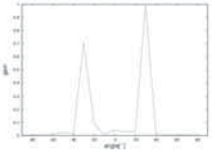


Fig.2 : Gain of Beam steering.

2-P-32

2-P-32　周波数方向冗長化 STFT に基づく位相復元の音楽信号への適用

Application of phase reconstruction with redundant short-time Fourier transform along frequency axis to music signal

◎若林佑幸，小野順貴（首都大）

◆我々はこれまでの発表において，周波数冗長度パラメータ α を導入し，周波数方向の冗長度を考慮した短時間フーリエ変換 (Short-time Fourier transform: STFT)による時間周波数表現を提案した．周波数方向冗長化STFT をよく知られた Griffin と Lim による位相復元アルゴリズムに適用し，音声信号の位相復元の性能改善を確認した．さらに，応用の一つとして音声強調を取り上げ，周波数方向の冗長化による強調性能の改善を確認した．

◆本研究では，対象を音楽信号とし周波数方向冗長化 STFT を用いて位相復元を行った．通常の音楽信号に加え，ドラムのような打楽器音を模擬したインパルス信号に対する位相復元の性能を評価した．

◆これまでの取り組みと同様，Fig. 1 のように通常の STFT（α=1）と比較して周波数冗長化 STFT（α=2,4）によって位相復元性能を改善することをスペクトル距離で示した．インパルス信号に対しても周波数方向冗長化により効果があることを確認できた．

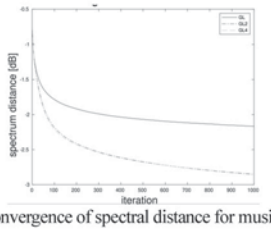


Fig. 1: Convergence of spectral distance for music signals.

2-P-33

2-P-33　船舶航行音静音化のための誤差信号の測定箇所の影響*

* Effect of measurement location of error signal for boat noise reduction

☆深津遼貴，武藤憲司（芝浦工大）

● ある運河では小型船舶が朝から夕方にかけ多く航行し，運河沿いの住宅に騒音影響をもたらしている．本研究の目的は船舶航行音を適応騒音制御し室内空間を静音化するシステム構築である．本報告では，窓スピーカを用いた単一周波数音の制御時における誤差信号が音圧の低減量および低減範囲に与える影響を示した．

● 誤差信号を測定する箇所は窓に近づくほど低減範囲が大きくなることがわかった．音より振動から取得する誤差信号を用いることで，低減範囲を大きくするとともに，安定した誤差信号の取得により，制御信号を安定して作ることができるのではないかと考える．

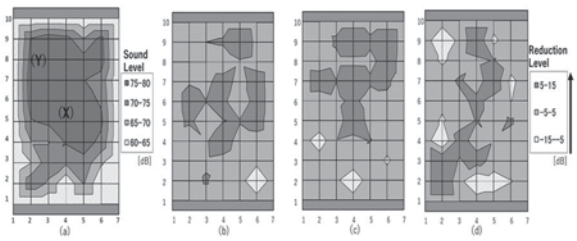


Fig. 1. Sound pressure level and reduction level of map by 120 Hz (Vibration pickup's position(X), Actuator's position(Y), (a): Sound pressure level map, (b): Reduction level map by vibrometer, (c): Microphone (50 mm), (d): Microphone (500 mm))

2-P-34

2-P-34　マイクロホンアレイを用いた低周波音源の位置推定
—MUSIC 法と遅延和法の比較実験—

The localization of the infrasound using microphone array

☆中山紬(学習院)，土肥哲也，岩永景一郎，小林知尋(小林理研)
中島康貴，青木創一朗(リオン)

◆超低周波音源を対象に屋外実験を行い，MUSIC 法と遅延和法による到来方向の推定精度を比較した。
◆スピーカから 20 Hz の帯域制限ノイズを発生させ，ラインアレイを用いて単独・複数音源の到来方向の推定を行った。(Fig. 1)
◆両方法ともに複数音源の場合に約 20 度以内で到来方向が推定できることが確認された。MUSIC 法はピークが鋭いが，遅延和法の方が推定精度が高い結果が得られた。(Fig. 2)

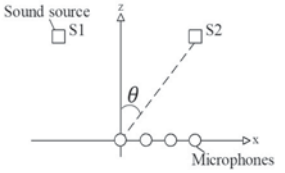


Fig.1 マイクロホンアレイと音源配置

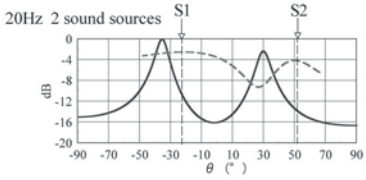


Fig.2 到来方向推定結果(実線：MUSIC スペクトル，点線：遅延和レベル)

2-P-35

2-P-35　パラメトリック/ダイナミック一体型スピーカを用いた室内インパルス応答合成に基づく仮想音源の方位・距離制御

Direction and Distance Control of Virtual Sound Source Based on Indoor Impulse Response Synthesis Using Parametric-array / Electro-dynamic Integrated Loudspeakers

○中山雅人，△江川琢真(阪産大)

◆室内インパルス応答の DRR（Direct-to-Reverberant Ratio）は，人の距離感に寄与することが知られている．本稿では， DRR が大きく異なる PAL（Parametric-Array Loudspeaker）と EDL（Electro-Dynamic Loudspeaker）を統合した PDL（Parametric-array / electro-Dynamic integrated Loudspeaker）を用いて，受聴者の両耳間レベル差と直間比を制御することで，スピーカ・受聴者間における任意の位置（方位と距離）に仮想音源を構築する新しい立体音響システムを提案する．
◆実環境において，実音源，従来手法(PAL:1, EDL:2)，提案手法(PDL:2)に対して，ダミーヘッドで室内インパルス応答を計測し，方位性能を ILD（Interaural Level Difference），距離性能を DRR で評価した．実音源・仮想音源の位置 9 点で実音源と各手法の誤差を算出し，その絶対値の平均を評価指標とした．Fig.1 に ILD，Fig.2 に DRR の実験結果を示す．実験結果より室内インパルス応答の音響パラメタをよく模擬できていることから提案手法の有効性が確認できた．

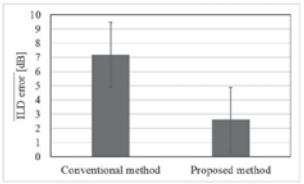


Fig.1：Experimental results for ILD

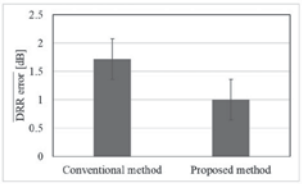


Fig.2：Experimental results for DRR

2-P-36

2-P-36　TCP ネットワーク上に構成したマイクロホンアレイによる収音について

A study of recording by microphone-array on TCP network

○高橋徹(阪産大・デザインエ)，△赤塚俊洋(阪産大・デザインエ)

◆TCP ネットワーク上に分散配置マイクロホンを構成
◆ブロードキャストパケットを用いたサンプリング開始・停止を制御
◆チャネル間時間差を評価
- 経験的に評価：50 回の試行による時間差をヒストグラム化
- 平均 311ms の遅延，平均の前後 15ms に分布

◆350ms 程度のバッファを用い，最大 30ms のばらつきを補正できればアレイ処理を適用可能なことが示唆された

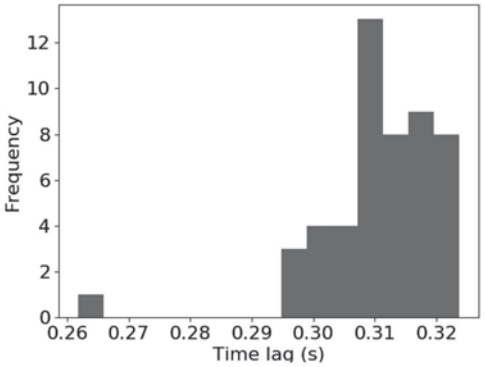


Fig1. Histogram of time lag

2-P-37

2-P-37　強調音声へのプロジェクションバックによるソースイメージ推定と分散マイクアレイによるその評価

Source image estimation with projection back onto enhanced speech signals for distributed microphone array

○荒木章子*，小野順貴**，木下慶介*，デルクロアマーク*
＊ 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
** 首都大学東京

◆広く研究されているマルチチャネル音声強調処理では、音源そのものの推定はできず、ある参照マイクにおける音源の観測信号、すなわちソースイメージを推定することが一般的である。
◆分散マイクでは、マイクごとに録音ゲインや収録音声の品質が異なるため、ソースイメージ推定の性能は、参照マイクの選択に依存してしまう。
◆そこで本稿では、参照マイクの選択に依らない音声強調として、観測信号の線形フィルタ出力である強調音声へ Projection back する音源分離手法を提案し、さらにこれを実用上容易に実現できる等価系を示す。
◆実録の複数人会話音声の音声強調で性能を評価したところ、提案法の有効性が示された。

Table: WERs (%) with projection back onto channel m (m = 1, · · · , 8) for all participants (“ch m”), with projection back onto the closest microphone to each participant (“Clo.”), and with the proposed projection back method (Prop.),

ch1	ch2	ch3	ch4	ch5	ch6	ch7	ch8	Clo.	Prop.
31.4	34.9	34.2	31.9	34.0	32.0	31.8	32.6	28.8	27.4

2-P-38

2-P-38　ユーザーからの補助情報を用いるインタラクティブ音源分離システム

Interactive audio source separation informed by user annotation

☆中野将生, 北村大地(香川高専)

優決定条件下での時間畳み込みに対処可能なブラインド音源分離手法として独立ベクトル分析，独立低ランク行列分析（ILRMA）などが提案されているが，これらの音源分離手法には周波数帯域ごと音源を間違えるブロックパーミュテーション問題と呼ばれる問題が存在する．

本稿ではユーザーとのインタラクションを利用し，ILRMAに対しアノテーションを与える事でブロックパーミュテーションを解決する音源分離システムを開発した．アノテーション作成方法には間違えた周波数帯域を直接指定するものと，沈黙区間を指定するものを採用し，それぞれ信号体歪み比（SDR）によって比較を行った．

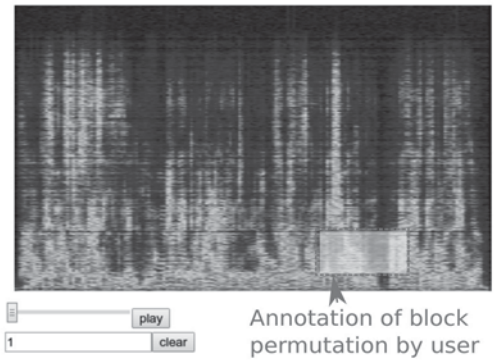


Fig. 1: User interface for annotation.

2-P-39

2-P-39　コサイン類似度罰則条件付き半教師あり非負値行列因子分解

Semi-supervised nonnegative matrix factorization with cosine similarity penalty

☆岩瀬佑太, 北村大地(香川高専)

◆ 本稿では，コサイン類似度を罰則項に用いる罰則条件付き半教師あり音源分離手法 (PSNMF) の最適化法を提案する．

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

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

◆ 楽器音の分離実験より直交化に基づく PSNMF（従来手法）とコサイン類似度に基づく PSNMF（提案手法）の SDR 値を比較し，提案手法が従来手法よりも高い SDR 値を記録したことを示す。

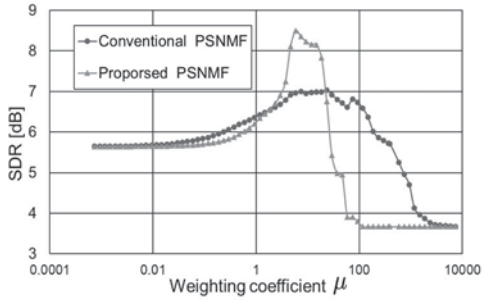


Fig.1 Example of SDR values of conventional and proposed PSNMFs for Tb.&Pf.

2-P-40

2-P-40 瞬時スペクトル減算法による同一到来方向の２雑音源の抑制に関する考察

Consideration on suppression of two noise sources in the same direction by instantaneous spectral subtraction.

○小澤賢司（山梨大・エ），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からの出力信号を 2 次元画像と見なし，2 次元 FFT により得たスペクトル上での減算を行うことで雑音抑圧を行う。この雑音スペクトルの推定を，分析の時間フレームごとに独立に行うため，瞬時スペクトル減算法と称している。目的音源と雑音源が同一方向にある場合，雑音源が 1 つならば良好な雑音抑圧が可能である。

◆ここでは，2 雑音源の場合を連立方程式として解くための定式化を行った。しかし，係数行列が悪条件であるために求解は困難であったので，2 つの雑音について独立にパワーを推定する近似解法をとった。

◆目的音源を 2 m，雑音源 #2 を 3 m に固定し，雑音源 #1 を移動させて雑音抑圧量を調べた計算機実験結果を Fig. 1 に示す。両雑音源が目的音源より遠い側にある場合には，約 20 dB の抑圧を得た。

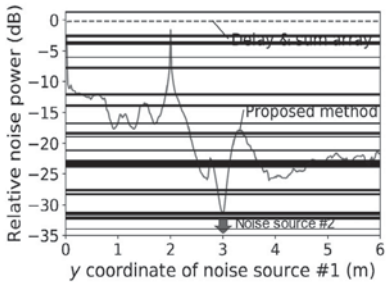


Fig.1 : Noise suppression as a function of distance of one noise source.

2-P-41

2-P-41　聴感印象推定のための脳活動特徴量抽出 —テンソル分解による機能的結合の低ランク表現の検討—

Low-rank representation of whole-head functional connectivity for prediction of subjective impressions.

☆矢野　肇(神戸大/産総研)，高島遼一，滝口哲也(神戸大)，
中川誠司(千葉大)

◆脳機能計測に基づく印象予測の精度向上のために，離れた大脳皮質間の活動の同期（機能的結合）に基づく特徴量を検討した．

◆すべての計測チャネル間の機能的結合をまとめたテンソルを Nonnegative Tensor Factorization によって低ランク近似し，皮質全体の結合パターンを表現する低次元特徴量を抽出した．

◆機能的結合 (WPLI) に基づく特徴量を用いて学習したモデルを用い，エアコン音聴取時の脳活動から一対比較判断を予測した結果，いくつかの被験者で予測精度が向上した（Fig. 1）．

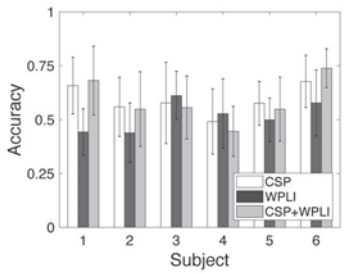


Fig. 1 Mean classification accuracies of predicted comparative judgment.

2-P-42

2-P-42　円形アレイ収音信号を用いた楕円アレイの内部音場再現

Applying Circular Microphone Array Signal for Interior Sound Field Reproduction with Elliptical Loudspeaker Array

☆任逸，羽田陽一（電通大）

◆高臨場感再生を目的とし，円形スピーカアレイを用いた音場再現法が提案されてきている。しかしながら，スピーカを円状に配置できる環境は限られるため，円形以外のアレイ形状においても，解析的な音場再現手法が望まれる。これに対し，私たちは楕円スピーカアレイを用いた解析的な音場再現手法を提案してきたが，これまで用いられてきた円形マイクロホンアレイやアンビソニックスマイクで収録された信号への対応は課題であった。

◆本研究では，マシュー関数の加法定理を用いることで，円形マイクロホンアレイの収音信号から求められる円調和展開係数をマシュー関数展開の展開係数に変換し，楕円スピーカアレイでその展開係数を合成する音場再現手法を提案する。

◆計算機シミュレーションの結果から，提案手法により楕円アレイの内側に平面波を再現できることがわかった。また，周波数が高くなるほど制御できる範囲が狭くなることがわかった。

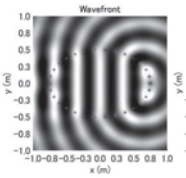


Fig. 1: Reproduced plane wave at 1000 Hz

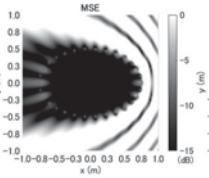


Fig. 2: Reproduction errors at 1000 Hz

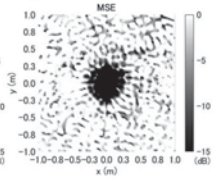


Fig. 3: Errors for sound field at 2000 Hz

2-P-43

2-P-43　複数音場の選択的合成アプリの試作

Prototyping of a selective synthesis app for multiple sound fields

○堀内俊治，△新井田統（KDDI 総合研究所）

◆提案した音場の選択的合成手法を用いて，複数音場の選択的合成を実現するインタラクティブ再生アプリ（iOS アプリケーションソフトウェア）を試作したので報告する.

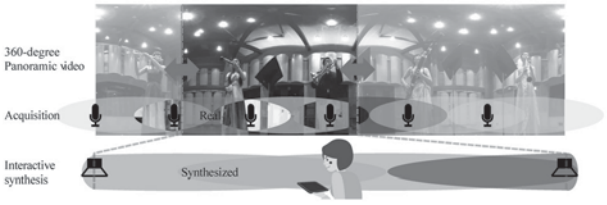


Fig. 0 Conceptual overview: Providing the freedom to choose the angle of view for both audio and video.

◆新たに試作したソフトウェアは，iOS 上で動作し，複数のオーディオストリーム入力と 1 オーディオストリームあたり最大 24 チャネルのマルチチャネル入力にそれぞれ対応させた．オーディオ処理は，短時間フーリエ変換（STFT）およびオーバラップアドを伴う逆 STFT により提案手法を実装している．これにより，一例として，2 つの音場を選択的に合成し，排他的にリアルタイム再生するなどの音響効果を実現することができた.

2-P-44

2-P-44　ミスナー法を用いた球面調和係数の推定に関する検討

Examination of estimation of spherical harmonic coefficients based on Misner's method

☆松岡 良憲，大谷 真（京都大学大学院工学研究科）

◆本稿では，Misner が提案した格子点上データからある球面上での球面調和係数を推定する手法（以下，ミスナー法）を音圧場に適用した場合について，以下の 3 つの検討を行った.

◆球面の半径のみを変更した場合：半径を変更しても全体的な精度は許容されるが，一部の 3 次，4 次成分には影響を及ぼすことが分かった.

◆展開係数の取得：ミスナー法で得られた球面調和係数から，高次アンビソニックスに必要となる展開係数を二重球アレイの考え方に基づいて正確に得ることができることが分かった．（Fig. 1）

◆放射基底関数の変更：動径方向の基底関数を球ベッセル関数，またはより単純な基底に変更してもほぼ精度に影響がないことが分かった.

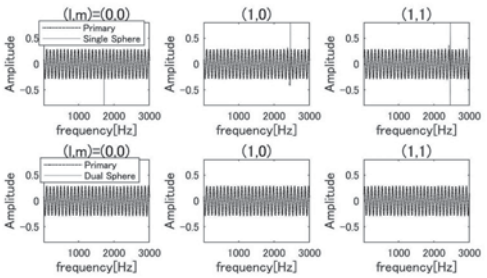


Fig.1 : Expansion coefficients of order 0 and 1, grid interval h=0.02 [m], using single and dual sphere array (real part, frequency domain)

2-P-45

2-P-45　空間標本化定理と高次のテイラー近似との融合

Spatial sampling theorem integrated with higher-order Taylor approximation

☆井上和弥，△大内翔，△鈴木遼平，島内末廣，青木茂明（金沢工大）

◆空間標本化定理によれば、マイクロホンアレイの素子間隔を波長の 1/2 以下にしなければ、正しく音場の情報を獲得することができない。

◆本報告では、空間標本化定理と高次テイラー近似を融合させたマイクロホンアレイ構成を提案し、マイクロホンアレイで実効的な素子間隔 d を 4 倍に拡大可能であることをシミュレーションにより確認した。(Fig.1)

◆テイラー近似において必要となる音圧の導関数値は差分近似により算出し精度に大きな影響がないことを確認した。

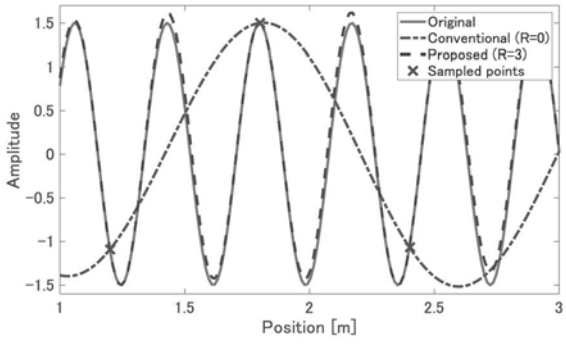


Fig.1 Instantaneous sound pressure

2-P-46

2-P-46　球面波と平面波の融合モデルを用いたインパルス応答の空間的補間

Spatial interpolation of impulse responses using hybrid model based on spherical and plane waves

☆吉野誠史，古家賢一（大分大）

◆音場の再現・制御には、部屋の音響特性としてインパルス応答を測定する必要がある。しかし、部屋全体などの大規模な領域になるとインパルス応答測定点が膨大になり、コストや時間が掛かるため、インパルス応答を空間的に補間することを考える。

◆本報告では、インパルス応答を直接波と反射波に分割してから、それぞれを補間する手法を提案する。分割することで、分割後の波に適した、仮想音源配置やモデル式が使用可能となる。

◆実環境実験において、提案手法の有効性を検証した。比較手法は音源周辺に仮想音源を追加する仮想音源プラス法である。実験結果から、直接波の補間精度は大幅に向上したが、反射波の補間精度はあまり差が見られなかった。

Table 1：直接波と反射波の補間誤差(dB)

仮想音源プラス法	直接波	−14.31
	反射波	−8.13
提案手法	直接波	−19.86
	反射波	−8.23

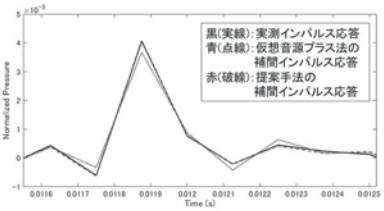


Fig.1：インパルス応答の直接波部分

2-P-47

2-P-47　円筒上に配置した複数スピーカによるトランスオーラルシステムの検討

Investigation into a transaural system using multiple loudspeakers arranged on a cylinder.

☆伊東優生，羽田陽一（電通大）

◆本稿では，スマートスピーカを用いた臨場感再生を目指し，円筒上に複数配置したスピーカを用いた Stereo Dipole における，スピーカの最適配置について検討する．

◆実測値の伝達関数を用いた条件数の計算を行い，周波数ごとに条件数が低くなる２つのスピーカを選択して使用することにより，全周波数帯域での良好なトランスオーラル制御の実現を目指す．

◆条件数の比較により，低域・中域・高域の，３つの周波数帯域のそれぞれの帯域に最も適したスピーカ配置を得た．

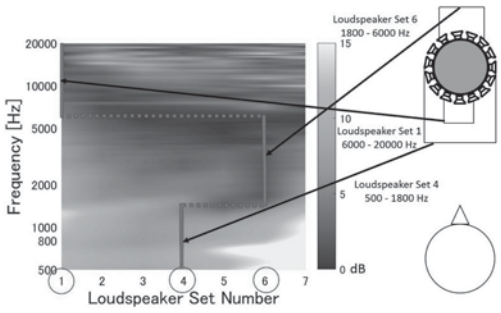


Fig1. The proposed pairing of the frequency band and loudspeaker set. The coloring denotes the condition number of plant transfer function matrix.

2-P-48

2-P-48　空間的マスキング効果に基づく3D 音響符号化に関する検討

Study on 3D Audio Coding Based on Spatial Auditory Masking

☆山中佑馬（秋田県大院），
西口正之，渡邉貫治，高根昭一，安倍幸治（秋田県大）

◆音響符号化方式の心理音響モデルに空間的マスキング効果を組み込むことにより量子化ビット数を削減できることを実験的に確認するため，聴取実験を行った．

◆マスキーとマスカーを空間的に異なる位置に配置し，マスキーの量子化ビット数を周波数マスキング効果に基づいて導出したものから段階的に減少させることで，量子化雑音が知覚される量子化ビット数の減少数を測定した．

◆マスキーと空間的に異なる位置にマスカーが存在することで量子化ビット数の許容可能な減少数が増加していることから，空間的マスキング効果による量子化ビット数の削減が可能であることを確認した．

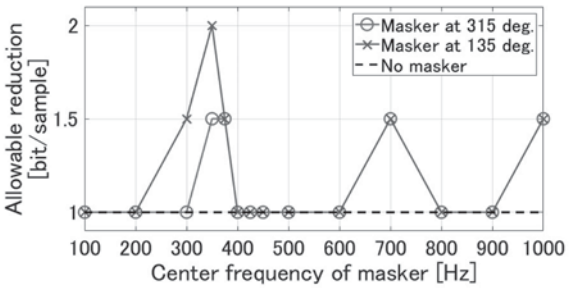


Fig.1：Allowable reduction of quantization bit rate of a maskee signal at 45 deg. in presence of a masker signal.

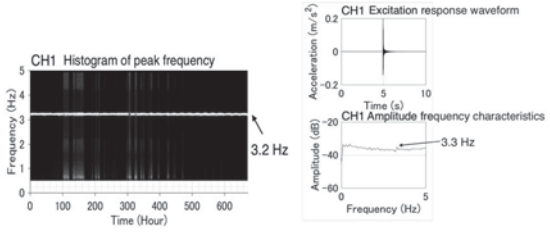
2-Q-1

2-Q-1　環境雑音を利用したポール型の屋外広告物の診断モニタリング

Structural health monitoring for pole-type outdoor advertisements utilizing environmental noise.

☆金子卓矢，高橋義典（産技高専）

建造物や屋外に設置された広告物は経年劣化によって強度が低下する．地震に備えて常にその健康状態をモニタリングする必要がある．風や交通騒音などの環境雑音によって生じる建造物の僅かな振動から，建造物の共振周波数と常時モニタリングするパッシブな診断手法が高橋等によって提案されている．本研究では，ポール型の屋外広告物の診断モニタリングを対象として，屋外に設置された街灯の共振周波数をモニタリングする実験を試みた．その上で，アクティブな加振実験と比較することで推定値の評価と比較をした．その結果，提案されているパッシブな診断手法では，相対誤差が 1%未満となりパッシブ診断手法の有効性が確認できた．



Result of resonant frequency monitoring　**Resonant frequency obtained by active measurement**

Examples of experimental results

2-Q-2

2-Q-2 変動性低周波音による心理・生理的影響に関する検討

Psychological and physiological effects of fluctuating low frequency sound

☆蛇子寛治，富田智晶（日本大院・理工研），
松田礼，町田信夫（日本大・理工）

◆周波数 100Hz 以下の低周波音は，環境問題の一要因として認知されている。しかし，風力発電用風車等から発生する，音の大きさが時間変動する変動性低周波音の人体影響は研究例も少なく，人体に及ぼす心理的，生理的な影響はよく分かっていないのが現状である。

◆本報では変動性低周波音を全身暴露した際の心理・生理反応量を測定し，音を構成する物理量（搬送波周波数.変動周期.変調度.等価音圧レベル）との関係性や、定常性低周波音と変動性低周波音で比較し関係性を調べた。

◆搬送波周波数が低いと不快感が弱まる傾向にあり、20Hz では不快感が強い傾向が見られた。音を構成する物理量と生理反応との間に傾向は見られなかった。

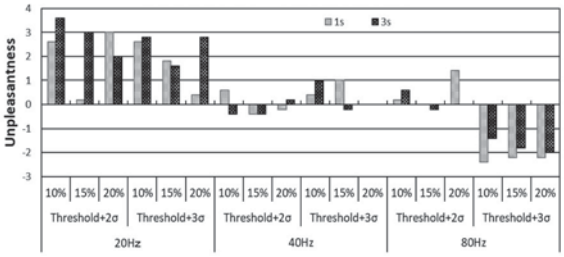


Fig.1：Discomfort with fluctuating low frequency sound when compared to steady low frequency sound

2-Q-3

2-Q-3 水平振動と音の同時暴露における優位性の検討

Experimental study of superiority in concurrent exposure of horizontal vibration and sound

☆高木耀太，並木敬祐（日本大院・理工研）
松田礼，町田信夫（日本大・理工）

◆振動環境では人間は振動に加えて音も同時に暴露される複合環境に置かれている場合が多い。しかし，乗り物を含む振動評価に用いられる全身振動の評価規格には複合環境における評価方法は定められていない。

◆本研究の目的は水平振動を全身暴露した時の振動感覚に及ぼす音の影響を明らかにすることである。本研究では，振動と音の優位性が判断できない大きさに感じる音量を調べるため，上昇系列と下降系列の２方向からの調整法により測定した。

◆振動レベルと振動と音の優位性の判断が切り替わる等価騒音レベルは比例関係にあり，優位性の判断ができない等価騒音レベルは 10 dB 程度の幅をもつ

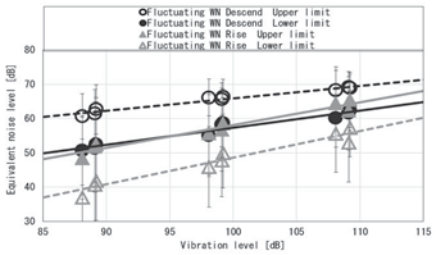


Fig.1：Relationship between equivalent sound level and vibration level (Fluctuating white noise)

2-Q-4

2-Q-4 高高度を飛行する航空機からの騒音の伝搬 －季節による伝搬性状の違い－

Sound propagation from aircraft flying at high altitude
- seasonal differences in propagation characteristics

○牧野康一（小林理研）、廻田恵司、△篠原健二（リオン）

◆高高度からの航空機騒音の伝搬に関して、これまでの夏季の測定データに加え、新たに冬季の測定を実施し、騒音伝搬性状と高層気象データについて、夏季と冬季の違いについて整理した。

◆冬季は高高度で強い西風（高度 10 km で 80 m/s）が吹いており、夏季との風速の鉛直プロファイルが大きく異なっていた。

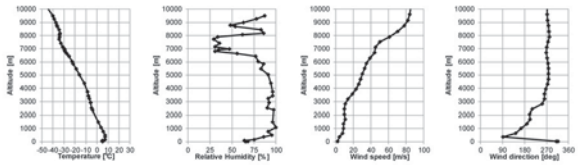


Fig. 冬季の館野高層気象観測データの例(12/30 9時)

◆短い周期のレベル変動は、今回の測定データでは夏季に比べて大きくなかった。

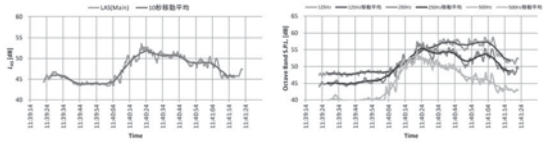


Fig. A 特性とオクターブバンドレベル波形と 10 秒間移動平均波形の例
B773 12/30 11:40 南 4km 東→西 高度 5800 m（雨、地上風 1.2 m/s、上空風 35 m/s）

2-Q-5

講演取消

2-Q-7

2-Q-7　ツイート情報を利用した言語モデルによる感情音声認識

Emotional speech recognition with language models adapted by tweets data

☆佐伯和哉，加藤正治，小坂哲夫（山形大院・理工学研）

◆感情音声認識は、感情の種類や感情強度による音響特徴の変動に加え、言語的特徴が平静音声と異なるため認識が難しい。

◆本研究では Twitter に投稿されたツイートデータが、感情音声に含まれる発話内容と言語特徴が近いと考え、適応データとして利用し、言語モデル適応が感情音声認識に有用であるか検討を行った。

◆51 日間で収集したツイートデータは2,586 万語からなる。

◆言語モデル適応と音響モデル適応を併用した場合、最終的に単語誤り率は 18.34 ポイント低下した。先行研究で行った少量の感情テキストを利用した同様の実験と比較して大幅な性能向上が得られた。

◆適応言語モデルは特に口語的な言い回しにおいて、有効である傾向が見られた。

Table.1:　Word error rates for each model adaptation (%)

	Baseline	AM adapt	AM + LM adapt (small data)	AM + LM adapt (Tweet)
Ang	40.97	28.21	28.29	20.70
Joy	41.23	32.62	24.24	21.85
Neu	23.13	19.88	18.26	12.65
Sad	39.09	26.52	25.82	15.91
Ave	36.11	26.91	24.15	17.77

2-Q-8

2-Q-8　Attention に関する損失を利用したノイズに頑健な End-to-End 音声認識

Noise-robust Attention Learning for End-to-End Speech Recognition

樋口陽祐（早大），○俵直弘，小川厚徳，岩田具治（NTT CS 研）
小林哲則，小川哲司（早大）

◆提案方式：

- Attention 機構を用いた End-to-End 音声認識において，雑音下音声に対するモデルの頑健性を向上させる学習法を提案
- 雑音下での音声入力に対して推定される Attention 重みを雑音を含まない音声入力（クリーン音声）による重みに近づけるように学習（Figure 1）.

◆評価方法：

- 以下のモデルに対し，CHiME-4 評価セットの文字誤り率を評価
  - ✧ Clean　：クリーン音声のみを用いて学習したモデル
  - ✧ Multistyle　：クリーン音声と雑音下発話音声を用いて学習したモデル
  - ✧ NRAL　：クリーン音声と雑音下発話音声を用いて，雑音に頑健な Attention 機構を学習したモデル（提案モデル）

◆実験結果（Table 1）：

- 提案する雑音下発話音声に対する Attention 重みを補正する学習は，End-to-End モデルの雑音に対する頑健性を向上できることを確認

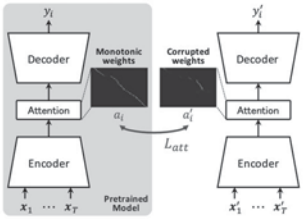


Figure 1. Proposed noise-robust attention learning using pair x and noisy speech x'.

Table 1. Character error rate (%) on CHiME-4 evaluation tasks.

Model	et05_simu	et05_real
Clean	79.1	82.3
Multistyle	29.9	35.5
NRAL (Proposed)	28.8	33.1

2-Q-9

2-Q-9 Mask-based MVDR 設計低遅延化のための、Reference Mic Selection と TF mask 予測のマルチタスク学習

Multi task learning to predict reference mic and TF mask for designing MVDR with low latency

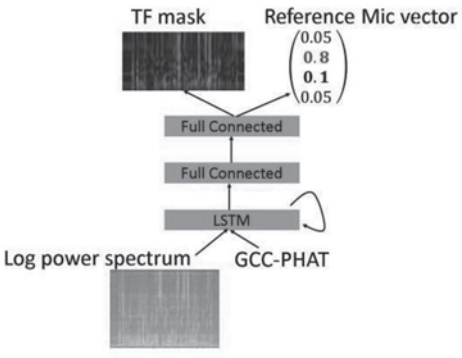
○小島淳嗣，小黒玲（アドバンスト・メディア）

◆ファーフィールド音声認識において，認識率を向上させるための音声強調が盛んに研究されており，特に Time Frequency (TF) mask に基づくビームフォーマが大きく成功

◆TF mask に基づくオンラインビームフォーミングでは，ビーム方向を決定づける one-hot vector の推定のため発話単位で SNR を計算
→ フレームごとにビームフォーマをアップデートするのが困難

◆提案：one-hot vector と TF mask のマルチタスク学習

- フレーム単位でのアップデート可能となった
- 同じ遅延時間である先行研究に比べ，CER が削減



2-Q-10

2-Q-10　会議シーンにおける End-to-End 音声認識のパフォーマンスに関する検討

The Performance of End-to-End Speech Recognition at Meeting Scenes

○長野紘之，△後藤悠斗，△加藤暁浩，△下野普也，△能勢将樹（リコー）

◆会議シーンにおける日本語話し言葉コーパス(CSJ)を訓練データに用いた End-to-End 音声認識のパフォーマンスを調査した。

◆追加の訓練データを用いずパフォーマンスを向上させるアプローチの一つとして「人間が音声認識を意識して」会議を実施した際のパフォーマンスについても調査した。

◆通常の会議では認識対象となる単独発話(非重複発話)の割合が 41.8%であること、また、音声認識を意識して会議をする事で単独発話の割合が 85.1%へと大幅に増加することを確認した。

◆一方、認識率(CER)は通常の会議、音声認識を意識した会議の間に大きな差は見られなかった。

Table 1. Performance survey results of End-to-End speech recognition at conference scenes.

	通常の会議	音声認識を意識した会議
会議時間	31m21s	30m07s
発話時間	30m36s	23m10s
総発話数	885 発話	637 発話
単独	370 発話（41.8%）	542 発話（85.1%）
重複	515 発話（58.2%）	95 発話（14.9%）
相槌数	174 発話（19.7%）	26 発話（4.1%）
笑い数	79 発話（8.9%）	23 発話（3.6%）
認識率(CER)	67.9%	65.2%

2-Q-11

2-Q-11　咽喉マイクによる多人数会話音声認識のためのデータ拡張

Data augmentation for multi-party conversation speech recognition using throat microphone

☆鈴木貴仁（静岡大），緒方淳（産総研），
△綱川隆司，西田昌史，西村雅史（静岡大）

◆咽喉マイク（Fig.1）は外部雑音の影響を受けづらいが、接話マイクとは大きく異なる音が収録されるため、通常の音声認識システムでは認識精度が著しく低下する

Fig.1：Throat microphone

◆咽喉マイクで収録された大規模音声データベースは存在しないため、咽喉マイク音声のみで音響モデルを学習するには学習データ量が不十分である

◆一方で、データ不足を補うために **Speed perturbation** や **SpecAugment** といった音声のデータ拡張手法が提案されている

◆本研究では、Speed perturbation と SpecAugment を三種類の咽喉マイク音響モデルの学習に対して適用し、比較実験・検証を行った

◆認識実験の結果、（1）咽喉マイク音声のみでの DNN-HMM の学習、（2）音響ミスマッチを抑制することを目的とした咽喉マイク特徴量空間から接話マイク特徴量空間へのマッピング用 LSTM の学習では二つのデータ拡張手法の適用によって認識精度の改善が得られた

◆しかしながら、（3）接話マイクから咽喉マイクへの特徴マッピングを CSJ（約 240 時間）に適用して得た擬似的な咽喉マイク特徴量で初期学習した DNN を実際の咽喉マイク音声で再学習する場合には、二つのデータ拡張手法の適用による性能改善は得られなかった

2-Q-12

2-Q-12　Uni-directional LSTM と Local Attention を用いたストリーミング音声認識

Streaming speech recognition with Uni-directional LSTM and local attention

☆陳家浩，西村良太(徳島大学)，北岡教英(豊橋技術科学大学)

◆本稿では、単方向 LSTM と local attention を用いたストリーミング可能な End-to-End 音声認識モデルを提案した。

◆CSJ データを用いて、文字単位出力のモデルを学習し、評価した。

◆Local attention を導入することによって認識精度の改善が確認できた。

◆CTC 出力において、local attention は高サブサンプリングレートでの改善効果が大きいことが確認された。

◆本手法は local attention の幅の遅延のみでストリーミング認識が可能である。

CNN	att_unit	att_heads	att_winsize	downsampling	avg CER
×	×	×	×	1/4	12.60%
×	×	×	×	1/6	13.67%
×	200	1	13	1/4	**10.37%**
×	200	1	13	1/6	**10.40%**
○	×	×	×	1/4	10.03%
○	×	×	×	1/6	12.27%
○	200	1	13	1/4	**9.87%**
○	200	1	13	1/6	**10.27%**
○	200	1	7	1/6	**10.23%**
○	512	5	13	1/4	**9.63%**

Fig.1：Model configurations and their character error rates

2-Q-13

2-Q-13　高齢者音声コーパス構築と音声認識への年齢・方言の影響の分析

Construction of elderly Japanese speech corpus for acoustic modeling and analysis of the influence of age and dialects on elderly speech recognition

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

◆高齢者音声認識精度の向上のために S-JNAS（平均年齢：67.6 歳）より更に高齢のコーパスが必要と考え、現在までに全国4カ所、平均年齢 79.2 歳の 211 名、音声時間 21.7 時間の音声をデータベース化した。

◆今回はベースラインの音響モデルの精度向上を考え大量の音声データ（既存のコーパス：JNAS、S-JNAS、CSJ）並びに本コーパス音声データを用いたモデル（Table 1 の‘Mixed’）を作成したところ、既存のコーパスのみのモデル（Table 1 の‘w/o’）に比べ、高齢者音声を学習モデルに用いる有用性が示唆された。また、この認識実験では高齢になるほど WER 上昇する傾向が見られた。

◆ ‘w/o’モデルを地域別音声に適応化させたところ、同じ地域同士で適応化後の WER は低下するが、他地域の音声に適応化すると WER が増加し、方言が音声認識に影響を与えることが示唆された。

Table 1　WERs (%) for speech recognition experiments'w/o elderly utterance'acoustic models and BCCWJ language model with acoustic adaptation to each dialects

	Mixed	w/o	Adp. (Nagasaki)	Adp. (Tokushima)	Adp. (Yamagata)
Nagasaki	11.7	**11.6**	**11.0**	13.7	11.6
Tokushima	30.3	**31.0**	32.2	**28.0**	36.2
Yamagata	5.5	**5.6**	6.4	7.9	**4.9**
Total	**13.4**	**14.4**		**12.4**	

2-Q-14

2-Q-14　書き言葉から話し言葉へのテキスト変換を用いた話し言葉音声認識用言語モデルの評価

Evaluation of Language Model for Spontaneous Speech Recognition Trained Using Written Language-to-Spoken Language Text Conversion

☆小橋優矢(徳島大)，西村良太（徳島大），北岡教英（豊橋技科大）

◆正確な話し言葉の音声認識には大量の話し言葉のデータが必要.

◆話し言葉テキストデータを増やすために Sequence-to-Sequence モデルを用いて書き言葉のテキストを話し言葉のテキストに変換する.

● また，変換前の書き言葉のデータで学習した言語モデルと変換後の話し言葉のデータで学習した言語モデルでは，後者の方が話し言葉に対してより良いパープレキシティを得た.

● 作成した言語モデルは，話し言葉の統計的特徴を上手くとらえていた.

Table.1：Perplexity of language models trained using each data

言語モデル	パープレキシティ
書き言葉（変換前）	152.1
話し言葉（変換後）	**140.2**

Table.2：Probability of each language model in the sentence

(例) （s）｜えー｜、｜ちょっと｜え｜、｜ ちょっと ｜待っ｜て｜（/s）

	書き言葉	話し言葉
パープレキシティ	67.3	**42.5**
P（、｜（s） えー）	0.457	**0.795**
P（ちょっと ｜ えー 、）	0.002	**0.004**
P（、｜ 、 ちょっと）	**0.047**	0.019
P（え ｜ ちょっと 、）	0.0004	**0.0006**
P（、｜ 、 え）	0.049	**0.260**
P（ちょっと ｜ え 、）	0.001	**0.004**
P（待っ ｜ 、 ちょっと）	**0.110**	0.019
P（て ｜ ちょっと 待っ）	0.917	**0.930**
P（（/s）｜ 待っ て）	0.051	**0.234**

2-Q-15

2-Q-15 双方向 LSTM における先読みフレームの効果に関する検討

A Study on Effects of Prefetched Frames for Bidirectional-LSTM-Based Low-Latency End-to-End ASR

○ 小林 彰夫 (筑波技大) 西崎 博光 (山梨大院) 宇津呂 武仁 (筑波大)

- 低遅延 end-to-end 音声認識
  - 一定の連続フレーム(セグメント)に対し LSTM を適用 [Zang, 2016]
    - 推定すべき系列の将来の特徴量を先読みして頑健性を確保
    - 日本語の場合，先読みにより認識性能はどのように変わるのか？
  - ２種類の低遅延型 LSTM による比較 (図 1)
    - LC-LSTM-A … 処理セグメント内のフレーム数固定
    - LC-LSTM-B … 発話始端付近のフレーム数を大きめに設定
    - 学習済み CTC でアラインメントを取得後，cross-entropy 基準で学習
  - 特徴量の先読みによる比較
    - prefetch …先読みした音響特徴量を用いる
    - w/o prefetch …先読みした音響特徴量は用いないが，LSTM はオーバラップ分を含めて遷移する
- 実験＆実験結果
  - ネットワークは LSTM 4 層＆文字出力，JNAS で学習/評価
  - 発話始端付近のセグメントのフレーム数を 50 フレーム，以降を 30 フレームで処理すると，LC-LSTM-A/B の結果はほぼ同等(表 2,3)．
  - LC-LSTM-A/B とも音響特徴量の先読みなしの場合，先読みありには劣るが，オフラインの単方向ネットワーク(表 1)と同等の結果

図 1 低遅延型 LSTM

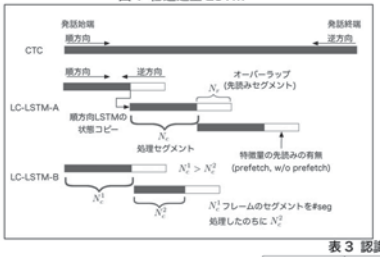


表 1 認識結果(オフライン)

	WER(%)
CTC(Bi)	**5.8**
CTC(Uni)	10.5
CE(Bi)	7.4
CE(Uni)	10.3

表 2 認識結果(LC-LSTM-A)

N_c/N_r	prefetch	w/o prefetch
30/0	18.1	-
30/10	8.9	11.1
30/20	9.0	13.0
50/0	13.9	-
50/10	8.7	**10.6**
50/20	**8.4**	**10.6**

表 3 認識結果(LC-LSTM-B)

$N_c^1/N_c^2/N_r$	prefetch		w/o prefetch	
	#seg=1	#seg=2	#seg=1	#seg=2
30/10/10	11.3	12.3	25.7	17.7
30/20/10	9.7	9.6	11.8	12.4
30/20/20	9.0	10.2	12.0	12.3
50/20/20	8.2	9.9	12.7	11.7
50/30/10	**8.1**	9.4	10.1	**9.2**
50/30/20	**8.1**	8.3	9.6	11.5

2-Q-16

2-Q-16　発話の時間変動に着目した音声認識誤り区間推定の検討

Speech recognition error detection based on the time variation of utterances

☆舒 禹清，山田 武志，牧野 昭二(筑波大)

◆自然発話の時間変動に着目し，CRNN（convolutional recurrent neural network）と変調スペクトルを用いた音声認識誤り区間推定法を提案する．

◆話し方に起因する認識誤りは発話の時間変動に関係することが多いので，Log-mel filterbank energy をベース特徴量とする変調スペクトルを用いる．

◆変調スペクトルはフレーム毎に 2 次元の特徴マップとして表されるので，従来の BLSTM よりも CRNN を用いて分析する方が適していると考えられる．

◆認識誤り区間推定の実験を行うことにより，提案手法の有効性を検証した．

◆Table 1 より，提案手法の F 値は従来の BLSTM を用いた手法，及び単語信頼度を用いた手法よりも高いことが分かった．また，提案手法においては RNN 部よりも CNN 部の方が性能改善に寄与していることを確認した．

Table 1　F-score for each method

手法	F 値
信頼度を用いた推定法	0.684
BLSTM を用いた推定法	0.672
CRNN を用いた推定法（提案手法）	**0.689**
CNN を用いた推定法	0.686

2-Q-17

2-Q-17　CNN ベース識別モデルによる F0 推定と伴奏重畳歌唱音声および雑音環境下読み上げ音声における評価

Fundamental frequency estimation by using CNN-based discriminative model and evaluation on singing voice mixed with accompaniment and read speech under noisy environment

☆川村智規(静岡大院・総合科学技術研)，甲斐充彦(静岡大)，
中川聖一(中部大)

◆本研究では，Convolutional Neural Network (CNN)ベース識別モデルによる雑音に頑健な基本周波数の推定方法を提案する．

◆自己相関法の考えをもとに自己相関近似値特徴量を獲得するための CNN ベースのモデルを構築し，全結合層により量子化された周波数帯域クラスに識別する．

◆歌唱音声と読み上げ音声を用いて実験を行うことで，どのような違いが生じるか検証した．

◆読み上げ音声は，歌唱音声に比べピッチの変動が少なく，高い正解率を得られた．歌唱音声は，入力の際にコンテキスト情報を考慮することで大きく改善がみられることがわかった．

Table1 歌唱および読み上げ音声における結果比較
（フレーム単位F0正解率）

	SNR=∞[dB]	SNR=0[dB]
歌唱音声		
RAPT法	80.42	31.78
CREPE法	-	72.25
NP+CNN(提案法)	82.28	75.39
読み上げ音声		
RAPT法	89.53	77.44
CREPE法	-	91.22
NP+CNN(提案法)	90.73	90.11

2-Q-18

2-Q-18　ボトルネック特徴量の合成に基づく音声認識のためのデータ拡張の検討

Data augmentation based on bottleneck features for End-to-End ASR

☆高木景矢，秋葉友良，塚田元(豊橋技科大)

◆本研究では、単一話者の音声合成モデルを用いてテキストのみから音声認識用のデータ拡張を行う手法を提案する。

◆そのために，音声から話者性を取り除いた特徴量 (ボトルネック特徴量) を入力とする End-to-End 音声認識を構成する。

◆データ拡張には、テキストデータから単一話者の音声を合成し、そこからボトルネック特徴量を抽出することで End-to-End 音声認識の学習データとする．

◆Librispeech を用いた評価実験の結果、テキストデータと単一話者の音声合成モデルのみを用いて音声認識性能が改善することを確認した。

◆また、同じテキストデータから言語モデルを構成し、提案手法と併用することで、さらに認識性能が改善することを確認した。

Table 1. 特徴量の違いによる音声認識性能

特徴量	ペアデータ (train_clean_100)	テキスト＋TTS (train_other_500)	言語モデル (train_other_500)	文字誤り率 (test_clean)	単語誤り率 (test_clean)
fbank＋pitch	✔			7.47	16.18
	✔	✔		7.17	16.12
	✔		✔	6.92	14.07
bottleneck (提案法)	✔			7.17	16.17
	✔	✔		**6.20**	14.53
	✔		✔	6.49	**13.90**
	✔	✔	✔	**5.55**	**12.52**

2-Q-19

2-Q-19　Hybrid CTC/attention モデルを用いた構音障害者音声認識の検討

An investigation of speech recognition using Hybrid CTC/attention model for dysarthric speakers

☆澤佑哉，高島遼一，滝口哲也，有木康雄（神戸大）

◆　構音障害者は筋肉の不随意運動により健常者とは異なるスタイルの発話をするため，従来の健常者を対象とした音声認識モデルでは認識が困難である．

◆　本研究では，構音障害者の発話を使用して End-to-End 音声認識モデルを構築し，性能を評価した．

◆　構音障害者の特定話者モデルの学習時に，不特定健常者話者モデルを用いてモデル適応を行うことにより，認識性能の向上を試みた．

◆　加えて，モデル適応に使用する発話数による認識性能の変化と，構音障害者の音素認識の誤り傾向について調査した（結果を Fig. 1 に示す）．

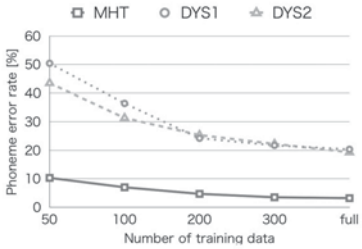


Fig.1 The correlation between error rates and the number of training data. MHT is a non-dysarthric person and the others are dysarthric persons.

2-Q-20

2-Q-20　クラウド型汎用音声認識 API とドメイン依存音声認識器の出力統合による認識精度向上

Improvement of accuracy by merging recognition results from open-domain and domain-dependent speech recognition systems

☆木村崇，千葉祐弥，能勢隆，伊藤彰則（東北大）

◆本研究では複数の音声認識結果を統合して音声認識誤りを低減する手法を提案する．先行研究の ROVER では単語の出現頻度から新たな仮説を生成するところを，BLSTM を用いて単語系列の依存関係を考慮して仮説を生成することで誤りが低減されることを期待している．

◆仮説生成は１つのドメイン依存音声認識器と３つのクラウド型汎用音声認識器から得られる 1-best の合計４つを統合して行われる．

◆認識器単独の精度と ROVER，提案手法の精度を比較した結果，大半の評価データにおいて提案手法が最も精度が高くなり，提案手法の認識仮説生成結果が単語の系列情報を用いて出力語彙を推定できていることが確認された．一方で未知語率が高いほど認識精度が低くなることも確認された．これは提案手法の対応できる語彙の種類が学習データに依存するため，認識仮説に含まれる単語そのものが未知語の場合には生成仮説の単語として利用できないことが原因である．

Table 1: Word error rate for evaluation data

Domain	Target	Open			ROVER	Proposed
System	Kaldi	Azure	Google	Watson		
Target domain	7.7	12.1	11.1	15.4	8.5	**1.2**
JTES	40.2	16.6	12.3	21.6	13.9	**1.5**
JNAS	31.4	22.9	23.0	**21.9**	22.0	34.7
CSJ (EVAL1)	18.7	26.7	24.1	24.4	17.4	**14.4**
CSJ (EVAL2)	16.1	25.3	22.7	25.4	17.3	**11.9**
CSJ (EVAL3)	17.4	20.9	18.0	19.6	**14.7**	16.1
Average	20.4	20.1	18.0	20.9	15.0	**12.2**

2-Q-21

2-Q-21　軽量・動的畳み込みを用いた end-to-end 音声認識

End-to-end speech recognition using lightweight and dynamic convolution

◎藤田悠哉(ヤフー), Aswin Shanmugam Subramanian(Johns Hopkins University), 大町基(ヤフー), 渡部晋治（Johns Hopkins University）

◆機械翻訳で提案された動的・軽量畳み込みを end-to-end 音声認識に適用した。具体的には以下３つの取り組みの結果を報告する。

- CTC(Connectionist Temporal Classification)とのマルチタスク学習
- 周波数方向の動的・軽量畳み込み
- 自己注意ネットワークとの組み合わせ

◆理論的には Transformer より学習が高速化できる。

◆CSJ, Librispeech, REVERB, CHiME4 で評価実験を実施。

◆評価セットによっては Transformer を上回る性能を達成した。

				REVERB Real Room 1		*CHiME4 dev*		*CHiME4 test*	
Model ID	H^S	K^e	K^d	*Near*	*Far*	*simu*	*real*	*simu*	*real*
RNN	N.A.	N.A	N.A	18.4	19.2	10.4	9.9	20.4	19.2
SA	N.A.	N.A	N.A	10.9	13.7	10.1	8.9	16.9	15.8
LC	4	101	71	12.7	14.0	10.5	9.1	18.4	16.5
DC	4	101	71	16.2	17.0	11.0	8.7	19.1	17.4
LC2D	16	101	71	11.0	13.1	10.1	**8.2**	17.5	16.0
DC2D	2	31	11	16.0	16.9	11.6	10.1	21.6	19.4
SA-LC	8	N.A.	31	8.5	11.8	9.6	8.6	16.3	16.1
SA-DC	8	N.A.	31	9.3	12.4	**9.3**	**8.2**	16.7	16.0
SA-LC2D	4	N.A.	11	8.0	**12.0**	9.5	8.3	16.7	15.4
SA-DC2D	4	N.A.	11	**7.5**	12.1	9.6	8.3	**16.2**	**15.3**

Table 1 : WER of REVERB and CHiME4

2-Q-22

2-Q-22　再帰型ニューラルネットワークを用いた残響下音声認識性能の劣化予測

Degradation Estimation of Speech Recognition Performance for Reverberant Speech using Recurrent Neural Network

○福森 隆寛，西浦 敬信（立命館大）

音声認識技術を搭載した情報機器を開発する中で，音声認識性能を測定する状況があるが，この測定には，利用環境で発話した膨大な音声データを準備する必要がある．音声認識性能の測定にかかるコストの削減を目指して，これまでに，我々は音声認識実験を実施せず，Multi-Layer Perceptron（MLP）を用いて，少量の残響音声から音声認識性能の劣化量を予測する手法を提案した．本稿では，残響が後続の特徴量に長く影響することに着目し，Fig. 1 に示すように時系列情報を考慮した再帰型ニューラルネットワークを用いて音声認識性能を予測する回帰モデルを構築した．音声認識性能の予測実験の結果，Long-Short Term Memory に基づいた回帰モデルを用いることで，MLP よりも高い精度（平均予測誤差：5.7 %）で音声認識性能を予測できることを確認した．

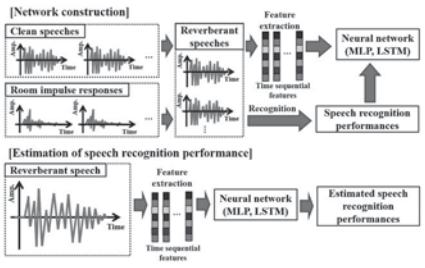


Fig. 1 Overview of performance estimation of reverberant speech recognition.

2-Q-23

2-Q-23 堺市民芸術文化ホールの音響設計

Acoustical Design and Characteristics of
Sakai Performing Arts Center

◎高橋藍子，板垣直実，北村浩一（NHK-TECH）

【大ホール】　＊吊りボルトを使用しない準構造化半球型大天井
＊躯体直貼り木仕上げによる軽鉄レス大臣柱
＊ユニットパネル化された側壁音響拡散体
【小ホール】　＊リブ仕上げ背後空間の最適音響チューニング
＊スライディングウォールを利用した舞台反射板

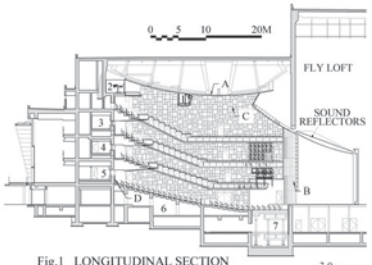
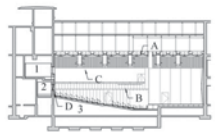


Fig.1 LONGITUDINAL SECTION OF LARGE HALL



【MAIN FINISH MATERIALS】
CEILINGS
A : FG10+8
WALLS
B : Wood Rib+GB-Ft15
C : GB-Rt15+PW19
D : Wood Rib30×45×60×90@60 +GC+GW

※GB : Gypsum Boards
PW : Plywood
FG : Fiber Reinforced Gypsum Boards
GC : Glass Cloth
GW : Glass Wool

Fig.2 LONGITUDINAL SECTION OF SMALL HALL

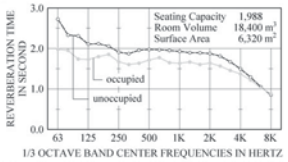


Fig.3 REVERBERATION TIME OF LARGE HALL

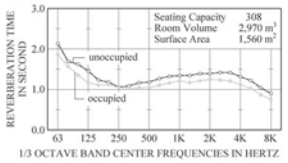


Fig.4 REVERBERATION TIME OF SMALL HALL

2-Q-24

2-Q-24 池袋西口公園「グローバルリングシアター」の音響
－野外劇場における建築および電気音響的手段による残響延長－

Acoustics of GLOBAL RING THEATER in Ikebukuro Nishiguchi Park" - Reverberation enhancement by architectural and electroacoustic technique in an open-air theater -

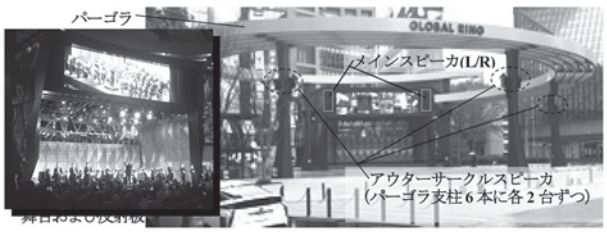
○清水寧（S/F Design Lab.）、△石川静（三菱地所設計）、
高橋顕吾、橋本悌（ヤマハ）、△河野峻也（ヤマハサウンドシステム）

◆ 池袋西口公園はクラシックコンサートやダンス、演劇に対応した野外劇場、大型ビジョン、インフォーメーション機能をあわせ持つカフェ、シンボリックなグローバルリングが整備されて 2019 年 11 月にリニューアルオープンした。

◆ 中核をなすグローバルリングシアターは、従来の軽音楽系のイベントに加えて、室内楽、オーケストラ等のクラシック音楽にも対応できるようにダブルシェル（建築）および AFC システム（電気音響）を導入することで積極的な残響（初期反射音、後期残響）の付加を可能としている。

◆ 舞台上：ダブルシェルの効果により、残響時間（250～2KHz 平均）が 1.2 秒、ST early が -10.1dB と、音楽演奏時の適度な響きと戻り音が確保されている。

◆ 客席エリア：客席中央での残響時間が AFC-on で 3.0～4.5 秒（Pattern-A～B）と、音楽演奏に十分な響きが確保されている。また、C80（楽音の明瞭度指標：目安値＝0±2dB）が AFC-off で 4.1dB、AFC-on で 1.8～-0.2dB（Pattern-A～B）と、AFC-on で残響音の増加による C80 の減少（-2～-4dB）が認められた。



2-Q-26

2-Q-26 SQ-STD のフレームレベル状態系列間照合における格助詞を用いた検索精度改善方式

A Method of Improving Retrieval Accuracy by Using Case Particles in Matching at a Frame Level for Query-by-Example Spoken Term Detection

☆丹治遥，小嶋和徳（岩手県立大），李時旭（産総研），
南條浩輝（京大・メディアセンター），伊藤慶明（岩手県立大）

◆ 目的：SQ-STD（Spoken Query - Spoken Term Detection）の検索精度改善

◆ Posteriorgram 照合：○高い検索精度，×大メモリ，×長い検索時間 ⇒ 最尤状態レベルで照合

◆ クエリの後ろには何らかの格助詞を付け話されていると仮定

- テキストレベルでは精度改善 [南條ら 2017]

◆ 格助詞拡張 SQ 法(提案方式)：SQ(音声クエリ)の末尾に格助詞のフレームを付けて検索

① SQ をフレームレベルの状態系列に変換

② 格助詞の音素列 ⇒ Triphone 系列 ⇒ 状態系列に変換

③ ②で作成した格助詞の状態系列を SQ の状態系列の末尾に付加（拡張 SQ と呼ぶ）

④ SQ と拡張 SQ の照合距離を比較し、小さい距離を採用

◆ 評価実験

- 従来の SQ：Baseline
- ９種の格助詞（で、が、から、に、の、と、を、や、より）を使用
- SQ と９種の拡張 SQ の 10 種を用いた場合は 0.3pt 低下
- SQ と「が」「に」「の」の拡張 SQ の４種を用いた格助詞拡張 SQ 法：1.5pt 改善

◆ 今後の展望：検索時間の削減手法の検討

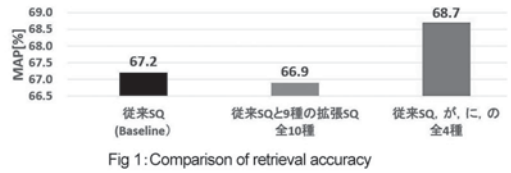


Fig 1：Comparison of retrieval accuracy

2-Q-27

2-Q-27 CTC に基づく音素コア区間検出を用いた音声中の検索語検出

Spoken Term Detection with Phonemic Core Zone Detection by CTC

☆石浦嘉晃（山梨大院），西崎博光（山梨大院），
小林彰夫（筑波技術大），宇津呂武仁（筑波大）

◆音素コア区間検出を用いた音声中の検索語検出

- ➢ CTC を用いて事前学習
- ➢ 音素コア区間検出
  - ✧ CTC によって音素と判断される部分のみを選び出す

◆実験：音素コア区間検出ありなしでの STD 性能評価

- ➢ 音素コア区間検出を用いた場合と用いない場合で比較
- ➢ 音素コア区間検出により STD 性能が向上

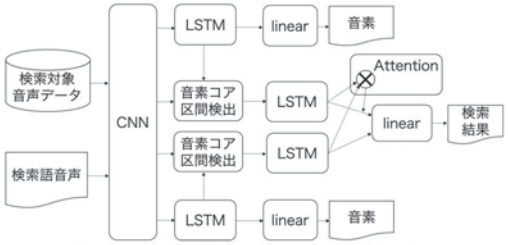


Fig 1: STD with phonemic core zone detection

Table 1: MAP

	MAP	
	全検索語	未知語
音素コア区間検出あり	0.965	0.687
音素コア区間検出なし	0.911	0.380

2-Q-28

2-Q-28　音声中の音声検索語検出のための最尤状態系列化方式の音声データの Posteriorgram 圧縮法

Reduction of Posteriorgram for speech data in Maximum likelihood state sequence in Query-by-example

☆横田平志，小嶋和徳（岩手県立大学），李時旭（産総研），
伊藤慶明（岩手県立大学）

◆SQ-STD : Posteriorgram 照合→○検索精度，×検索時間，×メモリ量
　音声クエリ最尤系列化方式→△検索精度，○検索時間，×メモリ量
◆提案手法 : 音声データの各フレームを最尤状態番号に対応。最尤状態番号が共通するフレームの事後確率ベクトルを平均
　⇒　平均事後確率ベクトル(36MB=4B×3000 次元×3000 状態)
　⇒　全フレームの事後確率ベクトルを保持せずに Posteriorgram 圧縮
◆先行研究と比較

手法	メモリ使用量,	検索精度,	検索時間
全講演の圧縮法：	**57MB**(先行研究の約 **1/2000**)，	**70.63%**,	**2.0 秒**
発話毎の圧縮法：	**58GB**(先行研究の約 **1/2**),	**73.83%**,	**2.0 秒**

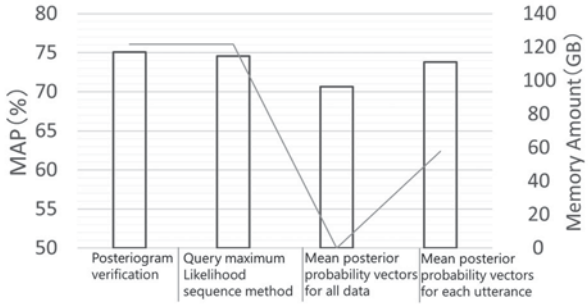


Fig.1 : Retrieval performance of each method in NTCIR-10.

2-Q-29

2-Q-29　複数言語を学習に用いたモデルによるゼロ資源言語の音声検索の検討

Spoken Term Detection based on Acoustic Models Trained in Multiple Languages for Zero-Resource Language

☆水落 智，千葉 祐弥，能勢 隆，伊藤 彰則（東北大学）

◆本研究では，ゼロ資源言語のデータベースに対する単語検索として，学習に用いる言語と対象言語の言語的な差異による影響を低減するために，日本語と英語の音素識別器から得た Posteriorgram を連結して用いる手法を検討した．検索対象言語はカクチケル語とする．
◆対象言語が日本語や英語の場合，複数の言語の Posteriorgram を連結した特徴表現を用いても検出性能は向上しなかった．
◆対象言語がカクチケル語の場合，複数の言語の Posteriorgram を連結した特徴表現を用いると，単一の言語の Posteriorgram を用いた場合と比較して検出性能を改善することができた．
◆今後は言語の差異による影響をさらに低減するため，ボトルネック特徴量の使用を検討する予定である．

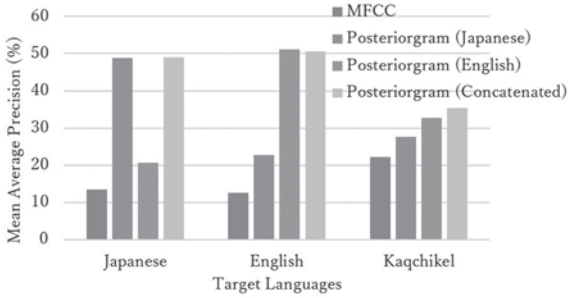


Fig.1 : Result of STD using Posteriorgram

2-Q-30

2-Q-30　JVS-MuSiC: Japanese multispeaker singing-voice corpus

JVS-MuSiC: 日本語多数歌唱者歌声コーパス

☆Hiroki Tamaru, Shinnosuke Takamichi, Hiroshi Saruwatari
(The University of Tokyo)

◆ Current Japanese singing-voice corpora
- Only consisting of single singer's voices

◆ Design of JVS-MuSiC
- 100 Japanese professional speakers' voices of common song (49 min and 23 sec) and songs different for each singer (88 min and 3 sec)
- Raw voice: recorded wav files
- Modified voices: Singing voices modified using singing-voice modification software Melodyne
- MPD file: Melodyne project documents used to modify singing voices
- Similarity and oneness matrices: results of subjective evaluations measuring similarity and oneness of unison

◆ Experimental analysis
- Similarity and oneness of unison: positive and moderate correlation
- Similarity of singing voice and similarity of speech: weak correlation

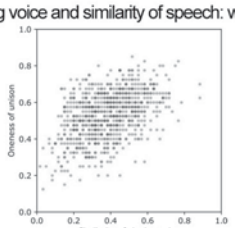


Fig.1: Scatter plot of average similarity and oneness of unison. Correlation coefficient is 0.45 and p-value is 1.4×10^{-39}.

2-Q-31

2-Q-31　ビブラートの自然さに関する知覚モデル構築に向けた実験と解析

Experiment and analysis for building a perceptual model in naturalness of vibrato.

☆宮崎嵩大(山梨大)，森勢将雅(明治大)

◆ビブラートの構成パラメータとして，
ビブラートの速さ・深さの 2 つのパラメータが存在する．
本研究はこれらのパラメータとビブラートの自然さの関係について主観評価実験を行い解析する．
◆ 評価項目では「ビブラートらしい音高の揺れの知覚」と「その揺れはビブラートとして適切か」の 2 点を用いた．
◆揺れの知覚では深さのほうが大きく影響する可能性を示唆した．
◆ユーザに関係なく，速さ・深さにおけるビブラートの自然さの範囲が存在することを示唆した．
◆今回の実験において，ビブラートの速さ・深さによる自然さの知覚モデル構築が実現可能である可能性を示唆した．

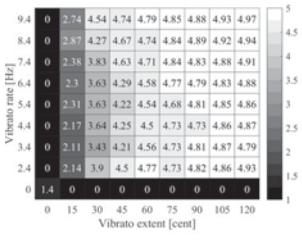


Fig. 1 ： Heat map of perception of pitch fluctuation

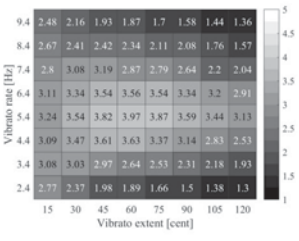


Fig. 2 ： Heat map of naturalness of vibrato

2-Q-32

2-Q-32　End-to-End 音声認識を用いた音声合成の半教師あり話者適応

Semi-supervised speaker adaptation for TTS using End-to-End ASR

☆井上勝喜，原直，阿部匡伸（岡山大・HS 統合科学研），
林知樹（名大），山本龍一（LINE），渡部晋治（Johns Hopkins University）

◆End-to-End 型のテキスト音声合成（Text-to-speech synthesis: TTS）には，数時間に及ぶ音声とテキストのペアデータが必要とされる．これを解決するために，数十分の音声データを用いた fine-tuning に基づく話者適応が提案されている．しかし，音声に対応したテキストデータを必要とし，これがボトルネックとなる．

◆本稿では，非ペア音声データを用いた End-to-End TTS の話者適応を提案する．まず，End-to-End ASR を用いて音声データに対応するテキストを生成する．続いて，生成したテキストと音声データを用いて TTS を fine-tuning する．客観評価と主観評価により提案方式（PROPOSED）と対応するテキストを用意できた条件（AD-FT-P）などを比較した．実験結果より，提案方式の使用によりペアデータを用意できた場合と同等の性能を示した．

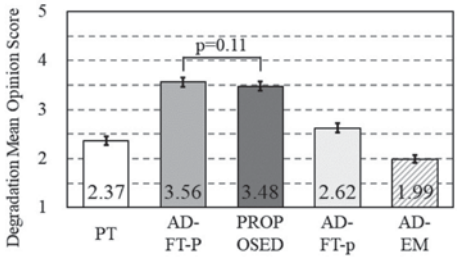


Fig.1 : The results of DMOS test with their 95% confidence interval about speaker similarity (PT: pre-trained model, AD-FT-P/p: fine-tuning based adaptation with paired data, **PROPOSED: fine-tuning based adaptation with unpaired data**, AD-EM: feature embedding based adaptation)

2-Q-33

2-Q-33　オノマトペを用いた環境音合成のためのデータセット構築とその分析

RWCP-SSD-onomatopoeia: onomatopoeic word dataset for environmental sound synthesis

☆岡本悠希，井本桂右（立命館大学），高道慎之介（東大院・情報理工），
山西良典，山下洋一（立命館大学）

◆ 本研究では RWCP-SSD 内の音響イベントに対し，オノマトペ（擬音語）が付与された環境音合成のためのデータセット構築を行った

◆ オノマトペを入力とし，環境音合成を行うことにより細やかな音の調整が可能になると考えられる

- オノマトペと音響イベント間の関係性を獲得するため，音響イベントに対しオノマトペが付与されたデータセットが必要

→ 現在そのようなデータセットは公開されていない

◆ データセットの概要

- 145,830 個のオノマトペ（105 個の音響イベント）
- オノマトペを付与した本人によるオノマトペに対する自信度
- 他者によって付与されたオノマトペに対する許容度

◆ 音響イベントにオノマトペを付与する際，複数のオノマトペが想定される場合がある → より環境音に適したオノマトペを選択するための基準として自信度と許容度を付与した

Table 1　Examples of labelled onomatopoeic word

Name	Onomatopoeic word	Confidence score	Average acceptance score	Description
whistle1	ピイイイー ピ ツィーッ	4 5 1	4.9 4.9 2.8	笛の音
shaver	ブーン ジー ブンブンブン	4 1 1	4.5 4.5 3.3	電気カミソリの動作音
file	ヒュンッ ミーッ サッー	4 5 3	3.5 1.9 3.3	金ヤスリで金属棒を擦る音
coffmill	ギュリギュリギュリ ガリガリガリ ブブブブブ	4 5 3	4 3 3.3	コーヒーミルで豆を挽く音
tear	ズシャアアア ビヤビヤ ギリャギリャ	5 3 1	4.1 2.9 2.5	コピー用紙を破る音

2-Q-34

2-Q-34　スタイル適応した DNN 音声合成における話者性の検討

A study on speaker reproducibility in style adapted DNN speech synthesis

☆蛭田宜樹（東工大），郡山知樹（東大），太刀岡勇気（デンソーIT ラボラトリ），
小林隆夫（東工大）

◆スタイル適応した話者依存の変換層（話者層）の異なる DNN による合成音声の話者性の違いからスタイル付与における適切な話者層の位置について検討した．

◆合成音声と自然音声間の音響特徴量歪と合成音声と自然読上げ音声間の i-vector の cos 類似度による客観評価結果から，話者層が出力層にある場合に最も話者性が高くなることが示唆された．

Table 1. Cosine similarities between i-vectors from synthesized speech samples and reading-style natural speech samples.

	悲しげ			楽しげ		
話者層	入力	中間	出力	入力	中間	出力
女性 1	0.63	0.52	**0.82**	0.56	0.65	**0.72**
女性 2	0.57	0.46	**0.75**	0.51	0.55	**0.60**
男性	0.72	0.76	**0.89**	0.62	0.65	**0.79**
	読上げ			ぞんざい		
女性 1	0.85	0.88	**0.90**	0.69	0.73	**0.86**
女性 2	0.81	0.85	**0.87**	0.60	0.65	**0.79**
男性	0.88	0.88	**0.91**	0.72	0.85	**0.89**

2-Q-35

2-Q-35　Joint Training End-to-End Systems for Speech and Speaker Recognition with Speaker Attributes

◎ Sheng Li, Xugang Lu, Raj Dabre, Peng Shen, Hisashi Kawai (NICT)

◆We propose to joint train the state-of-the-art transformer-based E2E ASR systems with speaker recognition task.

◆The speaker recognition task is designed to promote the ASR performance of the highly compressed transformer model.

◆Experiments (selected) show feeding bags-of-attributes of speakers as a ground-truth in training is effective for improving ASR performance, while speaker-ids leads to no significant improvement.

	CER%			
	CSJ-Eval01	CSJ-Eval02	CSJ-Eval03	Average
DNNHMM-CE 38M	9.7	7.8	8.4	8.6
TDNN-LFMMI 11M	8.3	6.6	**6.5**	7.1
BLSTM-CTC 11M	9.4	7.3	7.5	8.1
Attention/CTC 10M	8.4	6.1	6.9	7.1
Transformer compress 36M	9.5	7.3	8.3	8.4
Transformer compress, spk-id 36M	11.2	8.8	9.3	9.8
Transformer compress, sex+age+dur 36M	**7.9**	6.3	**6.5**	**6.9**

- Models are trained with CSJ training set (around 500-hour lectures).

2-Q-36

2-Q-36 カスケード構造を用いた音声パラメータ予測に基づく統計的パラメトリック音声合成

Statistical parametric speech synthesis based on acoustic parameter prediction using cascade model architecture

☆三井 健太郎 (Microsoft/東大)，法野 行哉 (Microsoft/名工大)，
△坪井 一菜，沢田 慶 (Microsoft)

◆一般的な統計的パラメトリック音声合成
- 音声パラメータである有声・無声情報，基本周波数，スペクトル包絡，非周期性指標を同時に予測
  →これらの間の関係を明示的にモデリング不可能

◆音声パラメータを段階的に予測するカスケード構造を提案 (Fig. 1)
- 音声分析では音声パラメータを段階的に抽出
- これらの知見を活用し段階的に音声パラメータを予測
- 複数の発話スタイルを考慮したモデル構造も検討

◆客観評価により従来の LSTM よりも高い音響モデリング性能を確認

◆主観評価により Seq2Seq モデルとも比較を行い有意に高い自然性

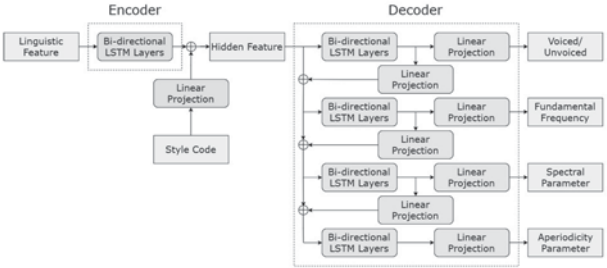


Fig.1：Block diagram of the cascade model architecture.

2-Q-37

2-Q-37 オンライン・オフライン編集機能を統合したビブラートデザインインタフェースの開発

Development of a vibrato design interface combining online and offline editors.

☆小野雄大(山梨大)，森勢将雅(明治大)

◆本研究では，歌声合成ソフトウェアにおける歌唱パラメータの調整作業（歌唱デザインと呼称）を支援するインタフェース開発を目指す．

◆従来の歌声合成ソフトウェアによるデザインでは，エディタによる編集後，歌声を合成し，結果を確認するオフライン編集を行っている．

◆著者らは，歌声を聴きながら歌唱パラメータを編集するオンライン編集機能を持つインタフェース Parrot を開発した．

◆本研究では，オンライン・オフライン編集機能を統合する手法を提案し，前報の Parrot V2 を改良した Parrot V3(Fig.1) を開発した．

◆Parrot V3 は，オンラインエディタで大まかなデザインを行い，オフラインエディタで詳細なデザインを行う作業が容易になり，作業に要する時間の短縮や作業時のストレスの軽減が期待される．

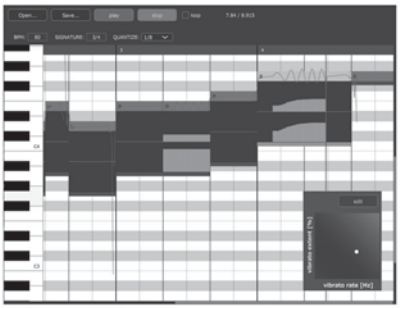


Fig.1： The Screenshot of Parrot V3

2-Q-38

2-Q-38　ビブラート歌唱音声におけるスペクトル変動消失音声の作成方法とその自然性の関係の調査

A study on relationship between how to generate singing voice without spectral fluctuation and naturalness in vibrato singing.

☆鈴木千文，坂野秀樹，旭 健作（名城大院），森勢将雅（明治大）

◆本研究ではスペクトル変動が一定であるビブラート合成音(スペクトル変動消失音声)の作成方法の検討を行った

◆歌声データ：ビブラート音声(目標歌手：Gackt)

◆スペクトル変動消失音声の作成
- ビブラート区間の基本周波数の各フレームのスペクトルの値で、ビブラート開始以降のフレームの値を置き換える

◆メルケプストラム歪みの算出(Fig.1(左))
- ビブラート音声に対し、一定にした際に用いたフレームのメルケプストラムと各フレームのメルケプストラムの歪みを算出
- Fig.1(左)より、2500[ms]付近ではケプストラム歪みが小さく、3000[ms]付近では大きい

◆エネルギー比の算出(Fig.1(右))
- 500～2000[Hz]と 2000～3500[Hz]の帯域のエネルギーを算出し、各帯域の比を算出
- Fig.1(右)より、1500[ms]付近ではエネルギー比は小さく、3000[ms]付近ではエネルギー比が大きい

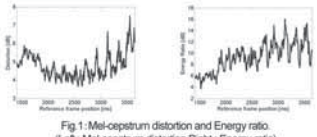


Fig.1 : Mel-cepstrum distortion and Energy ratio.
(Left : Mel-cepstrum distortion Right : Energy ratio)

2-Q-39

2-Q-39　HMM音声合成の学習時に考慮する動的特徴の追加に関する検討

A study for supplementary dynamic features for HMM-based speech synthesis

○西澤信行，△小原朋広，△服部元 (KDDI総合研究所)

HMM 音声合成では音声合成パラメータの１階差分等も HMM で同時にモデル化するが、それらは３フレームの変化といった短時間的な特徴のため、より長時間の特徴変化を適切にモデル化できない可能性がある。本研究では、音声合成に使うパラメータのストリームとは別に、音声合成パラメータの長時間変化に対応する要素を含むストリームを HMM に追加する方法について検討した。HMM 学習に 76.6 分の女声音声データを用いた実験により、最長 15 フレームから計算される特徴で構成されるストリームを HMM に追加した場合に、従来法よりも少ない HMM の状態数で、合成音声のメルケプストラムひずみが小さくなる条件があることが分かった。

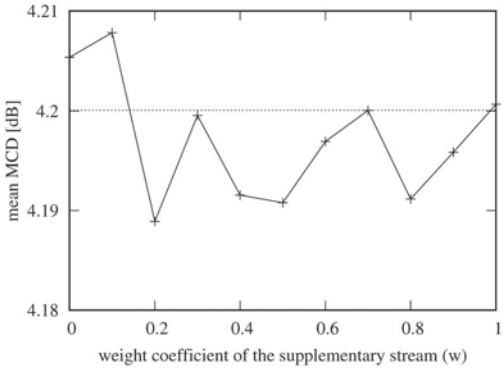


Fig. 1　Melcepstral distortion of synthetic speech sounds of 53 sentences (dashed line: conventional method).

2-Q-40

2-Q-40　指令応答型モデルを用いた句末対話韻律と与える印象の関係分析

An Analysis for Correspondense between Phrase-Final Prosody and Lexical Impression using Command-Response Model

☆高田一眞，句坂芳典（早稲田大），中嶋秀治（NTT）

◆自然な対話韻律制御に向け、発話内の語彙に即した韻律制御を目指す。

◆そこで、話者の判断を表す句末表現を用い、句中語彙が与える話者の判断に関する印象と対話韻律特徴の関係を分析した。中でも、対話において韻律変化が大きい句末モーラ F0 に着目した。

◆語彙が与える印象と句末韻律制御特性の関係を調べるため、以下の実験を行った。

(1) 指令応答型モデルを用い、句末モーラ F0 をアクセント指令で表現

(2) 同句の語彙が与える印象を SD 法で定量化

◆結果、句末モーラ指令の強度は、語彙が与える話者の判断に関する印象が強いほど負になりやすいという傾向が確認された。

◆また、この制御特性が聴覚自然性に影響することも確認された。これらの結果から、語彙が与える印象に基づく対話韻律制御が期待できる。

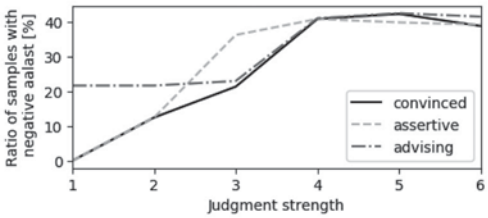


Fig.1 : Ratio of communicative speech samples whose A_{alast} are negative (A_{alast}: magnitude of phrase-final mora accent command)

2-Q-41

2-Q-41　少量データを用いた構音障害者音声合成の健常者モデルによる明瞭性改善

Speech synthesis for dysarthric people using a small amount of data and improvement of clarity using a healthy person model

☆南坂竜翔（神戸大），高島遼一，滝口哲也

脳性麻痺による構音障害者の発話は，筋肉の不随意運動により喉や口の筋肉を正常に動かすことが出来ないため，健常者と比べて不安定であり，コミュニケーションを難しくしている．また，多量の発話の収録は構音障害者に大きな負担を掛ける場合がある．そこで本研究では，脳性麻痺による構音障害者を対象としたコミュニケーション支援として構音障害者の少量データを用いた，テキスト音声合成(TTS)システムを提案する．まず健常者で音声合成モデルを生成し，そのモデルを構音障害者の少量データで再学習する．その際に以下の 2 つの学習手法で明瞭性改善を行う．

1. 2000Hz 以上の周波数帯域において健常者のスペクトルと構音障害者のスペクトルの誤差を取り，高周波成分が構音障害者に近づきすぎないように重み付けを行う

2. 健常者音声で学習された音素認識モデルに構音障害者の音響特徴量を入力し，得られた Loss を用いることで認識率があがるように学習を行う

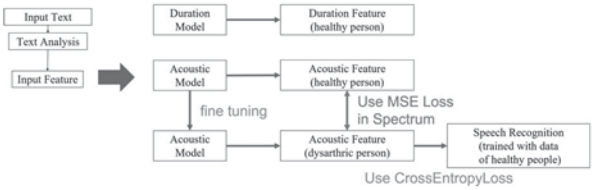


Fig1: A flow of proposal speech synthesis system

2-Q-42

2-Q-42　東北きりたん歌唱データベースを対象とした歌声の統計的解析

Statistical analysis of singing voice in Tohoku Kiritan singing database.

☆小川樹（山梨大），森勢将雅（明治大）

◆前報では，東北きりたん歌唱データベースのコンセプトや概要について述べた．本研究では，歌声を統計的に解析した．

◆解析結果

音素

歌いなれている歌を優先したため，バランスに大きな偏り

音名

最小値は F3，最大値は E5 となり，音域は約 2 オクターブ

継続長

平均値は約 0.33 秒で，0.1 秒以上 0.4 秒未満にノートが集中

◆童謡コーパスとは継続長の分布が異なる可能性があり，統計的歌声合成で合成される歌声の得意とする楽曲が異なる可能性

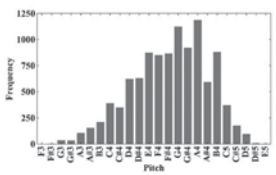


Fig 1: 音名のヒストグラム

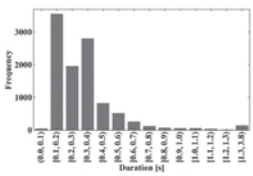


Fig 2: 各ノートの継続長のヒストグラム

2-Q-43

2-Q-43　アクセント情報付きローマ字テキストを入力とする日本語 end-to-end 音声合成の検討

A study on Japanese end-to-end speech synthesis using romanized input text with accent symbols

○平井俊男†、Tuan Vu Ho‡，スティアワン イファン†

（†アルカディア、‡北陸先端大）

◆近年、深層学習を利用した end-to-end 音声合成が盛んに研究されており、日本語への適用も検討されている。

◆高速でシンプルな日本語 end-to-end 音声合成システムの構築を目指し、単一女性話者の音声データ約 21 時間とそのアクセント情報付きローマ字テキスト、約5万対を Tacotron2 および WaveGlow で学習した。合成サンプルを Fig.1 に示す。

◆入力テキストでは、「高」になる各モーラの直後に /! を付加する。例：/iq'shu':'ka'N'ba'kari/,「一週間ばかり」

◆Real Time Factor (合成時間／音声長) は 0.194 と十分実用に耐える速度であった。

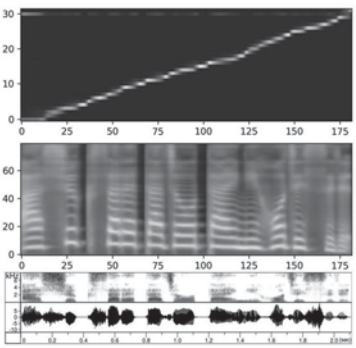


Fig.1 : From top to bottom: Alignment and mel-spectrogram inferred by Tacotron2 from a romanized Japanese input text*, and speech waveform inferred by WaveGlow with its linear-spectrogram.
(* /aqshukusaretajo:taini ho'jisuru./,「圧縮された状態に保持する。」、"hold it under compression.")

2-Q-44

2-Q-44　自由な歌唱速度の歌声の合成に関する検討

Note length prediction for singing voice synthesis with free singing speed

☆片平健太（神戸大），足立優司（メック株式会社），
田井清登（メック株式会社），高島遼一（神戸大），滝口哲也（神戸大）

◆アカペラ歌唱では伴奏が存在しないため歌唱速度は歌手に依存し，楽譜上のテンポを基準にしながら局所的な歌唱速度の変化が頻繁にみられる．

◆アカペラ歌唱音声合成における音符の発声タイミング推定では，音符継続長が大きく変化するため楽譜上の発声タイミングと実際の発声タイミングのずれが大きく推定が困難である．

◆音符継続長の伸縮率を推定することで局所的なテンポの揺らぎを推定する手法を検討する．

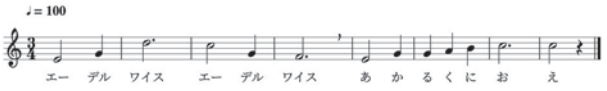



Fig.1：Musical score.

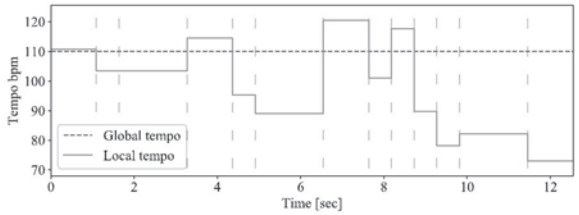


Fig.2：Comparison of Tempo changes between global and local.

2-Q-45

2-Q-45　深層ガウス過程音声合成における関数の確率微分方程式表現の利用の検討

A study of speech synthesis based on differential Gaussian process flows

○郡山知樹，猿渡 洋 (東大院・情報理工)

- **背景**：深層ガウス過程(DGP)に基づく音声合成
  - DNNにおける隠れ素子の連続的な表現と見なせる
  - DNN音声合成より高い自然性
  - 層を増やしても必ずしも性能が上がるわけではない
- **手法**：GPに基づくベクトル場を持つDiffGPを音声合成に適用
  - 確率微分方程式によって連続的な変化を表現
  - DGPの離散的な深層構造と異なりパラメータ効率が良い
- **実験**：DiffGPによる音声合成はDGP音声合成と比べて高品質な音声を合成可能

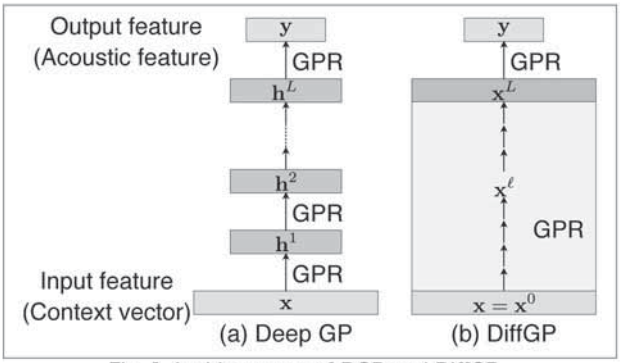


Fig. 1 Architectures of DGP and DiffGP.

2-Q-46

2-Q-46　テキストに付与する韻律記号とその音響的実現に関する日本語 end-to-end 音声合成を用いた音声教育的考察

Educational review on prosodic symbols and its acoustic realization using Japanese end-to-end speech synthesis

☆吉澤風希（東大）,熊野正（NHK技研），
峯松信明（東大）,栗原清（NHK技研）

- 音声教育において、テキストにどのように韻律記号を与えれば学習者にとって有益か考察したい。
- 日本語 end-to-end 音声合成システムを学習者に見立てることで韻律記号の指示形式について検討した。
- 指示形式ごとにモデルを学習させ、それらが合成した音声の自然性を通して指示形式について評価した。
- アクセント核を示さないと自然性が下がるが、ピッチの上がりはどのように示しても有意差が見られなかった。
- 音声合成システムが学習者のシミュレータとして利用できる可能性が示された。

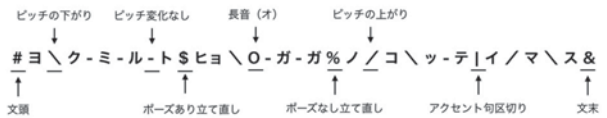


Fig. 1 Example of sentence with prosodic symbols
（よく見ると、氷河が残っています。）

2-Q-47

2-Q-47　スキャットを含む多人数歌唱データベースの収録と歌唱表現の分析

Building of singing database include scat and analysis of singing expression

☆芦田 紘輝，井本 桂右，山西 良典，山下 洋一
（立命館大・情報理工）

◆本研究では，通常の歌詞での歌唱に加えて，スキャットを含むデータベースの作成を行い，歌唱表現分析に対する有用性を示すための基礎分析を行った。

◆言語情報ありの歌唱音声を分析した際に現れる特徴は言語の特性に依存している可能性がある。

- **言語情報を持たない歌唱音声データベースが必要**

◆データセットの概要

- アマチュア歌唱者 38 人のポピュラー楽曲４曲の歌唱音声
- **同一歌唱者の通常の歌詞歌唱とスキャットによる歌唱を収録**

◆収録された歌唱音声の多様性を確認するために，代表的な歌唱表現であるビブラートに関して，歌唱データの分析を行った。

Table 1: Metadata of songs in data base

Song ID	Song Title	Range	BPM
M01	Jingle Bells	C3 ～ D4	100
M02	Stand by Me	A3 ～ A4	120
M03	世界にひとつだけの花	E3 ～ F#4	99
M04	Story	G3 ～ G4	64

2-Q-48

2-Q-48　　f_0 が高い歌声に向けた声帯音源波形と声道共鳴特性の推定

Estimation Method of Glottal Source Waveform and Vocal Tract Resonance for Singing Voice with High f_0.

☆高橋響子(JAIST)，李永偉(中国科学院)，赤木正人(JAIST)

◆目的

- **f_0 の高い歌声に対応可能な**声帯音源波形と声道形状の推定方法の提案

◆従来法が抱える問題点と提案法における解決方法

問題 1: 声帯音源波形の推定精度の低下

- 声帯音源波形の44.1 kHzサンプリングでの推定と声道共鳴特性推定のための12 kHzへのダウンサンプリング
- 声門閉鎖タイミングの全探索，その他のパラメータのGAによる最適化

問題 2: 声道共鳴特性の推定結果のばらつき

- 声帯音源波形と比べて声道共鳴特性はゆっくり変化するという仮定の設定
- 次の周期へ応答の流れ込みを考慮した最適化目的関数の設定

◆母音/a/を模擬した合成音を用いた評価

- 提案法による従来法の抱えていた二つの問題の改善を確認

◆結論: **提案法により f_0 が高い歌声を精度よく推定可能**

2-Q-49

2-Q-49　顔画像から予測される埋め込みベクトルを用いた複数話者音声合成

Multi-speaker text-to-speech synthesis using an embedding vector based on a face image

☆後藤 駿介(ディー・エヌ・エー/東大)，大西 弘太郎(ディー・エヌ・エー/電通大)，齋藤 佑樹(ディー・エヌ・エー/東大)，橘 健太郎，森 紘一郎(ディー・エヌ・エー)

◆本研究では，顔画像から推定される埋め込みベクトルを用いた DNN 複数話者音声合成モデルを提案する．提案手法は3 つのモジュール：Speech Encoder，Multi-speaker TTS，Face Encoder で構成される．

◆Speech Encoder はある話者の音声から，他話者と区別可能な埋め込みベクトルを出力し，Multi-speaker TTS はテキストと埋め込みベクトルを用いて音声を生成する．Face Encoder は顔画像から，その人物の声質に対応する埋め込みベクトルを推定する．

◆主観評価実験の結果，顔画像から推定される埋め込みベクトルを用いて生成した音声は，音声から推定される埋め込みベクトルを用いて生成した音声と比較して，顔画像との親和性，自然性のどちらにおいても同等の結果であることが示された．

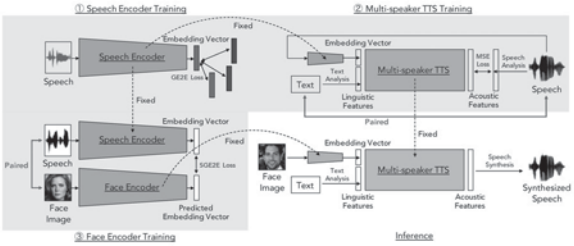


Fig.1：The overview of the proposed framework.

2-Q-50

2-Q-50　　歌声と楽曲の時間的ズレの知覚とテンポとの関係性の解析

Analysis of the effect of tempo on perception of shift length between singing and background music

☆小林礼奈(山梨大)，森勢将雅(明治大)

◆本研究では，歌声と背景楽曲の時間的ずれを知覚できる時間について，楽曲のテンポごとに差があるのかを調査するための音声評価実験を行った．

◆評価用データとして，プロ歌唱による歌唱 DB の中から，テンポの異なる 10 曲を用いた．

◆実験では，楽曲のサビ 1 フレーズ (10 s 程度)を聴き，歌声に対して知覚した違和感として評価させた．歌声に対して知覚した違和感とは，音高のズレではなく，歌声と背景音楽の時間的ずれとした．

◆結果として，歌声と背景音楽のずれている時間の許容範囲に，楽曲のテンポによる差は確認できなかった．

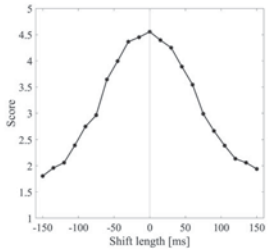


Fig.1：Result of evaluation

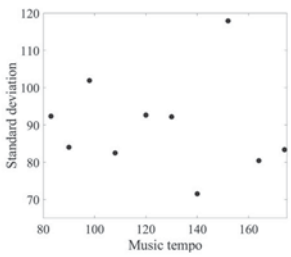


Fig.2：Relationship between standard deviation and tempo