1-1-1

1-1-1　効率的な異常音の正例負例分割による outlier exposure の高精度化

Outlier exposure improved by efficient division of positive and negative examples of anomalous sound

○太刀岡 勇気（デンソーアイティーラボラトリ）

◆機械音の異常音検知のための outlier exposure では対象機種の正常音からなる正例とそれ以外の機種の正常音からなる負例を分類するようにモデルを学習する

◆対象の機種以外の正常音も正例として使用できる場合があることを実験的に示し、効率的な正例・負例の分割法を提案する

◆図に提案法が仮定する埋め込みベクトルの構造を示す

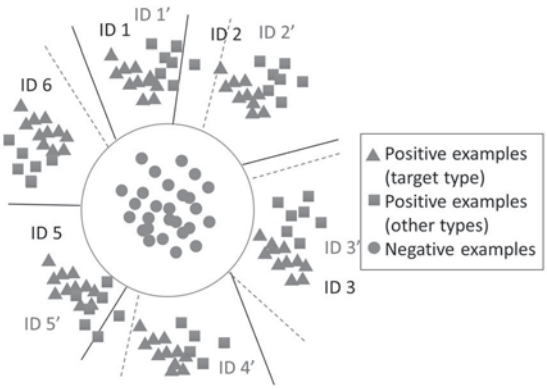


Fig.: A schematic diagram of assumption of embedded vectors obtained by the proposed OE-based model trained with two types of positive examples composed of normal data of target type and other types

1-1-2

1-1-2　音色関連特徴量に着目した教師なし機械学習による異常音検知技術の検討

Study on unsupervised deep learning for anomalous sound detection focused on acoustical features related to timbral attributes

◎小倉稜也，鵜木祐史（JAIST）

背景：音色に係る指標（音色関連特徴量）による産業機械の異常音検知システムが提案されている（Ota & Unoki, 2023）

課題：正常音のみを利用した教師なし機械学習による実現

目的：音色関連特徴量に着目した教師なし機械学習による異常音検知システムの実現を検討すること

提案法：

◆ 低次元データを取り扱い可能な識別モデルを採用

◆ 低 SNR 条件下でも機能する特徴量抽出法として，機械音区間における音色関連特徴量を抽出する機構を導入

◆ 擬似異常音データの生成とそれを利用した機械学習

評価結果：提案法が Autoencoder を上回る検出精度（AUC 評価）を達成

まとめ：提案法により，音色関連特徴量を用いた教師なし学習による異常音検知を実現

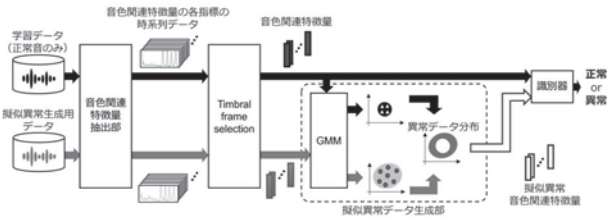


Fig.1：Overview of the proposed method

1-1-3

1-1-3　CLAP-ART: 意味情報を考慮した音響表現トークナイザを用いた音響説明文生成

CLAP-ART: Automated Audio Captioning with Semantic-rich Audio Representation Tokenizer.

◎竹内大起，グエンビンティエン，安田昌弘，大石康智，仁泉大輔，原田登(NTT)

◆音響説明文生成 (AAC)

- ➢ 一般の音からその内容の説明文を生成するタスク
- ➢ 事前学習された音響・言語モデルを接続するシステムが主流
- ➢ 最近の手法 EnCLAP では EnCodec の離散トークンを利用
  - ✧ 離散トークンで事前学習される言語モデルの追加学習に有効
  - ✧ 課題：EnCodec は音の意味情報を考慮しない

◆本研究では、事前学習された音響表現を離散トークンに変換し事前学習済み言語モデルに入力する AAC システムを提案

- ➢ 音響表現のもつ意味情報を継承した離散トークンを実現

◆実験より、従来手法と比較して説明文生成の性能改善を確認

- ➢ LLM なしで LLM ベースの手法と同程度の客観評価指標値を達成

表-1　AudioCaps での客観評価指標値の比較

	AudioCaps							
	METEOR		CIDEr		SPICE		SPIDEr	
Method	Mean ± Std	Reported*	Mean ± Std	Reported*	Mean ± Std	Reported*	Mean ± Std	Reported*
Conventional								
BART-tags [6]	–	24.1[♮]	–	75.3[♮]	–	17.6[♮]	–	46.5[♮]
AL-MixGen [7]	–	24.2[♮]	–	76.9[♮]	–	18.1[♮]	–	47.5[♮]
HTSAT-BART [8]	–	25.0[♮]	–	78.7[♮]	–	18.2[♮]	–	48.5[♮]
CNeXt-trans [9]	–	25.2[♮]	–	80.6[♮]	–	18.4[♮]	–	49.5[♮]
AutoCap (w/o text) [10]	–	25.6[♮]	–	80.4[♮]	–	19.0[♮]	–	49.7[♮]
LOWE [11]	–	**26.7**[♮]	–	81.6[♮]	–	**19.3**[♮]	–	**50.5**[♮]
Baseline								
EnCLAP-base [3]	24.2 ±0.36	24.7[♮]	74.4 ±1.43	78.0[♮]	18.1 ±0.21	18.6[♮]	46.3 ±0.76	48.3[♮]
Proposed								
CLAP-ART (BEATs)	**25.6** ±0.39	25.8[†]	**80.7** ±1.22	**82.0**[†]	**18.8** ±0.25	19.0[†]	**49.8** ±0.61	**50.5**[†]
CLAP-ART (MID-CLAP)	24.7 ±0.19	24.7[†]	76.4 ±1.40	77.7[†]	18.0 ±0.18	18.4[†]	47.2 ±0.77	48.0[†]

* この列では、♮は文献値、†は 6 回の異なるシード値による試行中の最大値を示す。

1-1-4

1-1-4　類似性尺度だけを用いた機械音の異常検知ソフトウェアの開発

Development of Anomaly Detection Software for Mechanical Sound using a New Similarity Measure

○神内教博

◆【新しい類似性尺度】を用いたパターンマッチングにより、機械音の異常を検知するソフトウェアを開発した。

◆雑音などの「ゆらぎ」の中で標準パターンと入力パターンにピークの「ずれ」が生じたとき、「ゆらぎ」を吸収しながら「ずれ」の増加とともに単調増加する距離値（Geometric Distance）を開発した。

◆このソフトウェアは、正常な機械音を学習するだけで、異常音を学習しなくても異常を検知することができる。

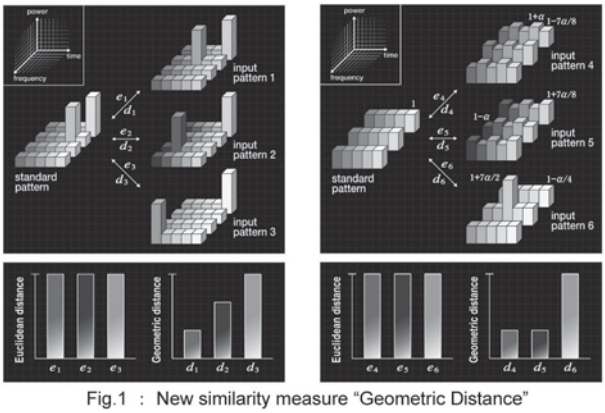


Fig.1 ： New similarity measure “Geometric Distance”

1-1-5

1-1-5　心不全診断のための音声バイオマーカー：Voice-BNP

Voice biomarker for diagnosis of heart failure: Voice-BNP

○中山仁史（広島市大院），△田村雄一（国際医療福祉大）

◆主たる音響特徴量として MFCC を採用し，さらにブラインド音声処理による音響特徴量の高度化を有した心不全診断のための音声バイオマーカーVoice-BNP を提案する．

◆Voice-BNP は血液を測定するのではなく，非侵襲な音声を入力とした信号解析技術で実現する音声バイオマーカーである．

◆心不全（入院前）>心不全（入院中）>心不全（退院後）>コントロールの関係が期待できる音響特徴量の分布と有意差を確認した．

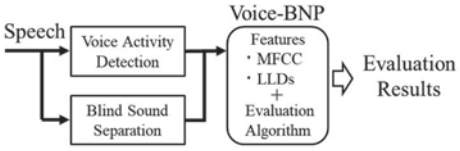


Fig. 1: Overview of Voice-BNP.

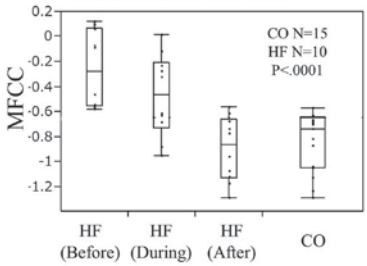


Fig. 2: Result of a MFCC vocalized a syllable.

1-1-6

1-1-6　Inverse problem processing of optically-measurement sound field based on diffusion model

◎Hao Di (Waseda Univ.), Kenji Ishikawa (NTT), Risako Tanigawa (NTT/Waseda Univ.) and Yasuhiro Oikawa (Waseda Univ.)

◆Proposal:

- We propose an inverse problem processing method based on diffusion model, including denoising, reconstruction, and extrapolation of 2D sound field images obtained by optical measurements.

◆Method:

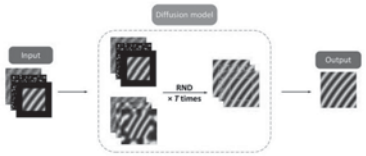


Fig.1：Inference process. The input data are processed through the diffusion model, with range-null space decomposition (RND) as the solver at each denoising step.

◆Results:

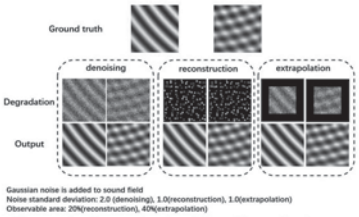


Fig.2：Sound field denoising, restoration, and extrapolation. Our method successfully addresses the inverse problems of sound fields, including denoising, reconstruction from observed noisy sound fields, and extrapolation to unknown regions.

1-1-7

1-1-7　スペクトルを適応的に抑圧する正則化項を用いた Audio Declipping

Audio declipping using regularization term that adaptively suppresses the spectrum

☆角田清香，赤石夏輝，矢田部浩平（農工大）

目的：振幅超過により歪んだ信号の復元 (audio declipping)

スパース最適化に基づく audio declipping

► スパース性を誘導することで clipping により生じた高調波成分を除去

従来手法：L1ノルムを用いた audio declipping

► 全ての成分を抑圧し，推定信号も小さくなる

提案手法：スペクトルを適応的に抑圧する正則化項を用いた audio declipping

► 振幅が最も小さい成分のみを抑圧する最小絶対値関数と，全ての成分を抑圧する L1 ノルムを補間する関数を利用

► 振幅が小さい部分を抑圧する度合いを調整可能

► 復元すべき成分の抑圧を防ぎ，clippingにより生じた歪み成分のみを抑圧(Fig.1)

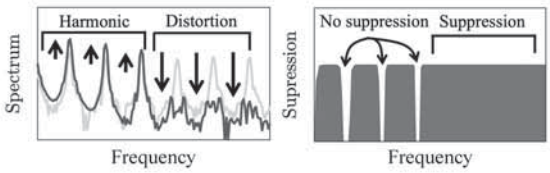


Fig. 1 An example of the spectrum of the clipped signal and the restored signal (left), and degree of suppression on each frequency bin by the proposed method (right).

1-1-8

1-1-8　東海道新幹線を対象とした空制部品におけるエア漏れ音のための音源分離に関する基礎検討

Fundamental study of blind source separation for leak sound in air-controlled equipment for Tokaido Shinkansen

○中山仁史，大島風雅，大村美結（広島市大院），村田剛基（JR 東海）

◆新幹線を対象とした空制部品におけるエア漏れ検査を実現するための基礎検討として，車両所内騒音環境下における空制部品の模擬エア漏れ音を対象とした音源分離の基礎検討を行う．

◆実験対象の編成は同社所属の N700 系とし，空制部品の詳細は保安上の理由により示さないものとする．本実験では加圧無しの車両下部付近にスピーカを設置して模擬エア漏れ音を出力する．

◆erSS-NMF では車両所内騒音のアクティベーション行列，模擬エア漏れ音の基底行列及びそのアクティベーション行列を推定する．車両所内騒音及び模擬エア漏れ音を定常音であると仮定し，各基底行列の数を１とした．

◆暗騒音に埋没した模擬エア漏れ音の音源分離が可能なことを確認した．

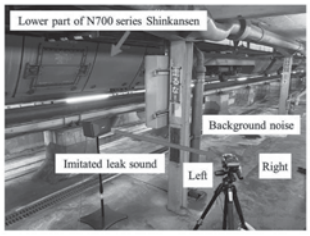


Fig. 1: Sound recording in JR Central.

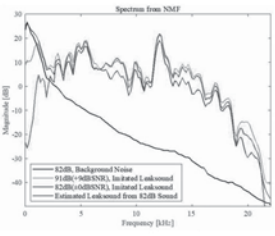


Fig. 2: Comparisons of leak sounds.

1-1-9

1-1-9　音質評価モデルの開発と車両音響性能への適用に関する検討

Investigation of the Development of a Sound Quality Evaluation Model and Its Application to Vehicle Acoustic Performance

○山中尋詞，△藤本麻由美，△若松功二（マツダ），
△青木武史，△五十嵐優司（パイオニア）

◆自動車のオーディオシステムの重要性が高まる中、音質評価の客観化と効率化が課題。

◆MAZDA CX-9 をベースに**音響リファレンスカーを構築**し、8 つの音質評価軸にて**音質を表現するモノサシを定義**。

◆車両 49 台分のバイノーラル録音データと主観評価データを収集し、**音質評価モデルを開発**。

◆音質評価モデルは、概ね±10 点以内の精度で音質評価点数を予測可能であることを確認。

◆音質評価モデルを用いて**車両音響性能への適用**を検討し、車室内に設置する吸音材/拡散材/反射材、またスピーカ角度の条件を変更することで、音質評価予測点数の変化を確認。

◆客観的なデータに基づいて車両音響性能を向上させる手段が明確になり、実際の車両開発に適用可能であることを示唆。

Fig.1：Step of Investigation

1-1-10

1-1-10　連続音場測定法を用いた音場特性の数値評価の試み

An attempt at numerical evaluation of acoustic field characteristics using the Continuous Sound Field Measurement Method

○立花杜斗，前田和昭 (TOA 株式会社)，
△小林由佳，岩田水晶，河原一彦 (九州大・芸工)

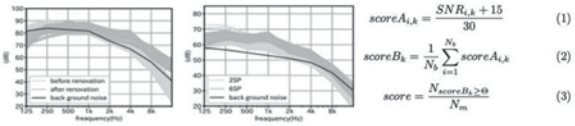
◆現場での音響測定には、測定に時間がかかること、時間の制約によって網羅的な測定が困難であること、専門的な知識が必要で誰でも簡単に実施できないという課題がある。

◆これらの課題を解決するために、M 系列信号を再生し、マイクを手に持って歩きながら測定を行う連続音場測定法を提案し、検証を重ねてきた。

◆本研究では、この連続音場測定法を用いて、音場全体における拡声音の聞き取りやすさの数値評価を試みた。

◆拡声音と暗騒音の周波数ごとの SN 比に着目し、STI の算出を参考に数値評価を行った。

◆測定数ポイントごとで聞き取りやすさの数値評価を行い、閾値を上回った割合で音場全体の評価を試みた。



$$scoreA_{i,k} = \frac{SNR_{i,k} + 15}{30} \quad (1)$$

$$scoreB_k = \frac{1}{N_b} \sum_{i=1}^{N_b} scoreA_{i,k} \quad (2)$$

$$score = \frac{N_{scoreB_k \geq \Theta}}{N_m} \quad (3)$$

Figure 1 The results of the octave band analysis for amplified sound and background noise

1-2-1

1-2-1　発話単位の潜在変数を導入した深層隠れセミマルコフモデルに基づく音声合成

Natural Speech Synthesis Based on Explicit Modeling Utterance Fluctuations in Deep Hidden Semi-Markov Models

☆三宅恭平，藤本崇人，橋本佳，南角吉彦，徳田恵一（名工大）

近年，Deep Neural Network 用いた End-to-End 音声合成の研究が進展している．特に深層隠れセミマルコフモデルは，状態継続長を明示的に制御することを可能にしながら，高品質な音声合成を実現した．しかし，この手法では学習データに含まれる，発話者や録音機材の状態によって生じる差異である発話変動を明示的に扱っていない．本稿では深層隠れセミマルコフモデルへ発話変動を表す潜在変数を導入することで，発話変動を考慮した学習を可能にする音声合成システムを提案した．

評価実験は従来手法と，本稿で提案した２種類のモデル構造に対して行った．実験の結果，発話変動を言語特徴量と音響特徴量の２つから推定する構造案２の性能が最も高く，発話変動を用いない深層隠れセミマルコフモデルより高いスコアを残すことができた．

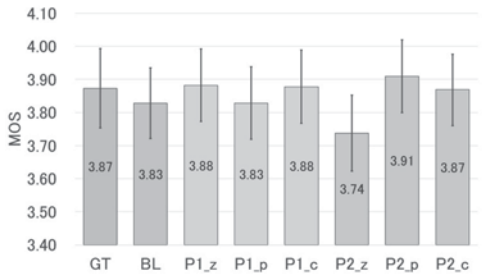


Fig.1：MOS test results of naturalness
(GT：Ground Truth, BL：Base Line, P1：Proposed1, P2：Proposed2)

1-2-2

1-2-2　Rectified Flow を組み込んだ深層隠れセミマルコフモデルに基づく音声合成

Speech Synthesis Based on a Deep Hidden Semi-Markov Model Incorporating Rectified Flow

☆淺野友紀，橋本佳，南角吉彦，徳田恵一（名工大）

◆ 深層 HSMM(DeepHSMM)と Rectified Flow を組み合わせる.

- 組み合わせによって深層 HSMM の過剰な平滑化の改善と Rectified Flow において十分に確立していなかった明示的なアライメントや継続長モデルを同時最適化する手法の実現を狙った.

◆ 提案手法：

- 深層HSMMの後段にRectified Flow を接続した音響モデルを提案し(Fig.1), その接続方法について最適な構造を調査する. また，主観評価実験によって提案手法の有効性を示した.

◆ 結果：

- 過剰な平滑化の改善と明示的なアライメントや継続長モデルの同時最適化に成功し，主観評価実験による性能向上を確認した.
- 主観評価実験(Fig.2)から深層 HSMM が 2 つの Acoustic Decoder を condition 側と開始分布側に接続し，かつ，各々が再構成誤差を通す形(**DH+RF(X0/M+Cond/M)**)が一番性能が良いことを示した.

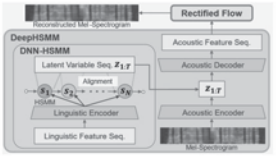


Fig.1：Proposed Model

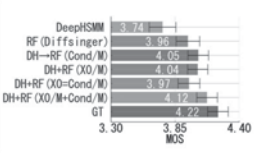


Fig.2：MOS test

1-2-3

1-2-3　双方向の自己回帰構造を導入した深層隠れセミマルコフモデルに基づく音声合成

Bidirectional-Autoregressive Deep Hidden Semi-Markov Models for Speech Synthesis

☆飯田諒，橋本佳，南角吉彦，徳田恵一（名工大）

一般的な音声合成モデルにおいては，推論時はテキストのみを入力し，合成音声を出力するが，自然性の向上には依然として課題が残る．本研究では，時刻 t の音響特徴量の生成において，言語特徴量に加えて，音響特徴量の因果的性質を考慮し，生成済みの時刻 1~t までの音響特徴量も使用し，再帰的生成を行うことで，生成に使われる情報量が向上し，音声の自然性の向上を果たした．また，ベースモデルとなる DeepHSMM では困難であったアライメントの制御性も解決に至った．

合成音声の自然性(Fig.1)について評価を行った．実験結果から，DeepHSMM と比較して，提案手法の AR_DHSMM のスコアが高い性能を示した．

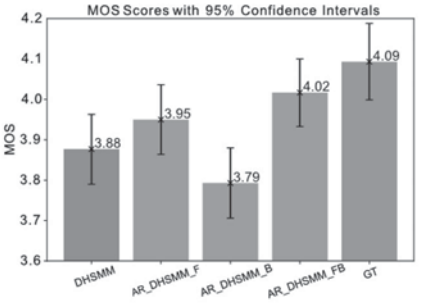


Fig.1：MOS Test

Proposed model
：AR_DHSMM

1-2-4

1-2-4　Mixture of Adapters を用いた軽量な zero-shot 音声合成

Lightweight Zero-shot Text-to-Speech by using Mixture of Adapter

◎藤田健一，芦原孝典，デルクロアマーク，井島勇祐（NTT）

◆少ないパラメータ数で効率よく話者をモデル化するために Mixture of Experts に基づく Mixture of Adapters（MoA）モジュールを利用．軽量で高速な zero-shot 音声合成を実現

◆比較の概要

- パラメータ数がそれぞれ 14M，42M の FastSpeech2（*Small*(*S*), *Medium*(*M*)），および Small に MoA モジュールを挿入した提案法の性能を比較

◆推論速度比較結果

- シングルスレッド CPU（Intel(R) Core(TM) i9-10940X CPU @ 3.30GHz）を用いて実験
- *S*, *M*, 提案法はそれぞれ real-time factor が 0.0127, 0.0286, 0.0148 であった

◆Zero-shot 音声合成における主観評価結果

- MoA モジュールを挿入することで，パラメータ数が小さくても高い自然性・話者類似性の音声合成が実現

◆40%のパラメータ数で 1.9 倍高速な音声合成を実現

Table 1：主観評価結果(話者類似性に関するXABテスト)

Model	Proposed vs *S*	Proposed vs *M*
All	**0.52** – 0.29 – 0.19	**0.36** – 0.36 – 0.28
nonpro.	**0.42** – 0.34 – 0.24	0.33 – 0.34 – 0.33
pro.	**0.62** – 0.24 – 0.14	**0.38** – 0.39 – 0.23

1-2-5

1-2-5　BERT を用いたアクセントラベル不要な日本語ニューラル TTS

Japanese text-to-speech using BERT without accentual labels

◎小椋忠志，岡本拓磨，大谷大和，Erica Cooper(NICT)，戸田智基(名大/NICT)，河井恒(NICT)

◆アクセントラベル不要な日本語ニューラルTTS用韻律予測手法の提案

◆事前学習済み日本語 BERT と強制アライメントにより抽出したモーラ単位の基本周波数の活用

◆アクセントラベルに依存せず漢字を含む単語列とカタカナ列からの韻律予測の実現

◆アクセント辞書を用いる従来手法を上回る韻律正確性と同等以上の合成音声品質の達成(T. Ogura *et al.*, ICASSP 2025)

◆エネルギーや継続長を含む複数韻律特徴量への拡張の有効性の実証

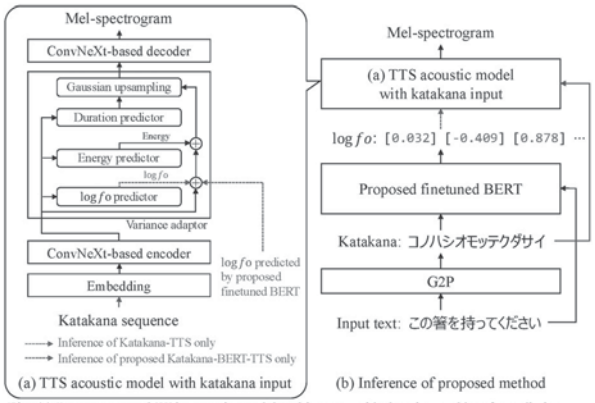


Fig. (a) Japanese neural TTS acoustic model architecture with duration and log *fo* prediction. (b) Inference procedure of the proposed method.

1-2-6

1-2-6　感情音声合成のための自己教師あり学習モデルによる音声表現抽出

Emotional Speech Synthesis Based on Speech Representation Extraction Using a Self-Supervised Learning Model

☆堀尾凌汰，橋本佳，南角吉彦，徳田恵一（名工大）

大量のデータから学習された自己教師あり学習モデルを用いて音声表現を獲得する手法は参照音声から多様な感情や僅かな感情の違いを表現できるが，それらの情報の適切な抽出手法については明らかになっていなかった．本稿では自己教師あり学習モデルによって感情表現を獲得する音声合成システムにおける音声表現抽出手法について提案する．

合成音声の自然性(Fig.1)と感情の類似性(Fig.2)について評価を行った．実験結果から自然性に関しては WeightMean が最もスコアが高いことが分かった．類似性に関しては自己教師あり学習モデルを用いた手法が One-hot ベクトルを用いた手法より高い性能を示し，多様な表現が可能であることが分かった．また，Attention を用いない WeightMean や Attention で一括で抽出する AllAttention が高い成果を残している．

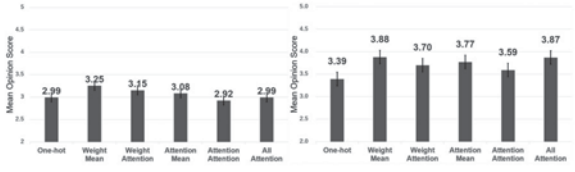


Fig.1 MOS test results of naturalness　　Fig.2 DMOS test results of similarity

1-2-7

1-2-7　アンカー話者埋め込みベクトルを用いた敵対的音声による音声プライバシー保護

Speech privacy protection with adversarial speech using anchor speaker embedding vectors.

☆勝又勇紀，中鹿亘（電通大）

◆模倣音声に対する音声プライバシー保護を目的とした、敵対的音声を生成する手法としてアンカー話者埋め込みベクトルを活用した設計を提案する。

◆実験結果から、従来手法である I-FGSM と比較して、UT-MOS と SNR がほとんど等価の条件で、Cos 類似度を大きく下げることを示した。

◆その上で、提案手法による敵対的音声から生成された模倣音声は従来手法から生成された模倣音声よりも最大で 15%優れた性能を示し、ボイスクローニングモデルに対する有効性を実証した。

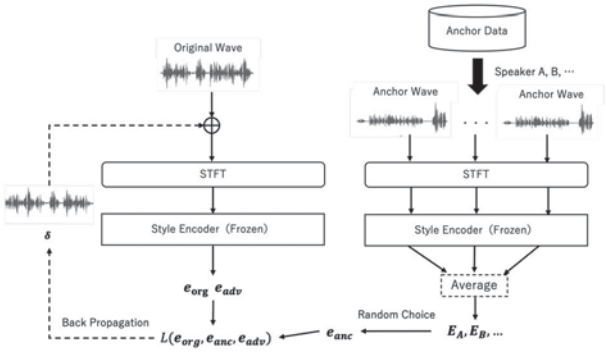


Fig. 1 Flowchart for generating adversarial speech

1-2-8

1-2-8　周期信号の位相情報を用いたフレーム駆動型ニューラルボコーダ

Frame-level neural vocoder utilizing phase information of periodic signals

☆薫田基広，藤本崇人，法野行哉，吉村建慶，橋本佳，南角吉彦，徳田恵一（名工大）

◆フレーム駆動型ニューラルボコーダ
- 特徴抽出において，アップサンプリングを行わないため高速に動作が可能だが，高音質な音声合成が難しい。

◆提案手法
- メルスペクトログラムに加え，基本周期信号の位相情報（GCI を基準とする１周期のうちの時間的な位置）を入力する。

◆結果
- 提案モデルは従来法を上回る音質を示した。
- 入力された位相情報に同期した波形が生成された。

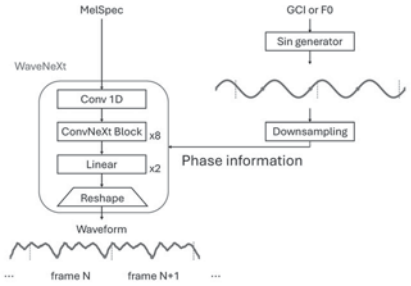


Fig.1 : Structure of the proposed model

1-2-9

Wavehax：調波信号モデルと2次元畳み込みを用いた複素スペクトログラム推定に基づくエイリアシングフリーニューラルボコーダ

Wavehax: Aliasing-Free Neural Vocoder Based on Complex Spectrogram Estimation with Harmonic Signal Modeling and 2D Convolution

◎米山怜於[1]，宮下敦志[1]，山本龍一[1,2]，戸田智基[1]([1]名大，[2]LINE ヤフー)

エイリアシングフリーボコーダのすゝめ
- 時間領域モデルは潜在空間上のエイリアシングを回避できない
- ニューラルボコーダが抱える課題の本質的な原因
  - 波形生成過程の複雑化による計算量増加
  - 汎化性能低下（特に高い基本周波数の外挿）

ボコーダの処理領域に起因する信号処理的特性を分析

処理領域	エイリアシングフリー	畳み込み次元	調波モデリング		
			雑音信号 or 入力なし	正弦波信号	調波信号
時間 Waveform	No	1D	×	○	○
時間周波数 Spectrogram	Yes	1D	×	×	×
		2D	×	×	○

エイリアシングフリー設計により軽量・高速・頑健・高品質なニューラルボコーダを実現

デモサイト
https://chomeyama.github.io/wavehax-demo/

1-2-10

1-2-10　階層的マルチタスク学習と Contextual Biasing を用いた End-to-End 音声認識

End-to-End automatic speech recognition with Hierarchical Multi-Task Learning and Contextual Biasing

◎楠奈穂美，樋口陽祐，小川哲司，小林哲則(早大)

<目的>
低頻度語の音声認識性能向上のための End-to-End モデリング
<アプローチ>
◆階層的マルチタスク学習（HMTL）
- **音声からの段階的な特徴抽出**（ex. 文字→サブワード→単語）
  - ✧ 音声から単語単位の**特徴抽出過程の強化**による低頻度語の認識強化

◆Contextual Biasing
- **Attention 機構**を用いて低頻度語にバイアスをかけ，優先して出力
  - ✧ **低頻度語のテキスト情報 (B)** による低頻度語の認識強化

<結果>
◆HMTL のみを用いたモデルと比較して，低頻度語の認識性能が向上

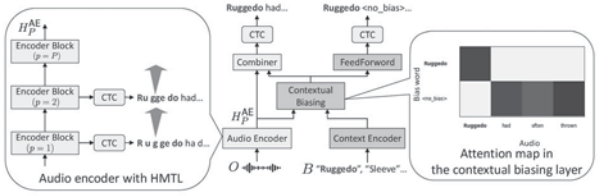


Fig. 1: The proposed model with the audio encoder based on HMTL and the contextual biasing layer.

1-2-11

1-2-11　注意機構に基づく中間特徴量を用いた CTC 音声認識の精度向上

Improving accuracy of CTC-based ASR using attention-based inter-layer features.

☆北條圭悟，若林佑幸（豊橋技科大），太田健吾（阿南高専）
小川厚徳（NTT），北岡教英（豊橋技科大）

◆本研究では，補助 CTC 損失を計算する手法を用いて，CTC に基づく音声認識モデルの推論速度を維持しつつ，精度改善を図る。
◆従来手法である attention-based CTC は，注意機構に基づき音響エンコーダの全ての中間層出力を用いて補助的な CTC 損失を計算する。
◆提案手法は，エンコーダの各層の特性を考慮し，異なる粒度の CTC 損失と注意機構の出力を推論時にも利用するモデル構造を導入することで，attention-based CTC を改良した。
◆実験の結果，提案手法は従来手法を上回る認識精度を達成した。
◆また，提案手法で用いるエンコーダの分割位置と使用する注意機構の数について分析し，中間層の最適な利用法を明らかにした。

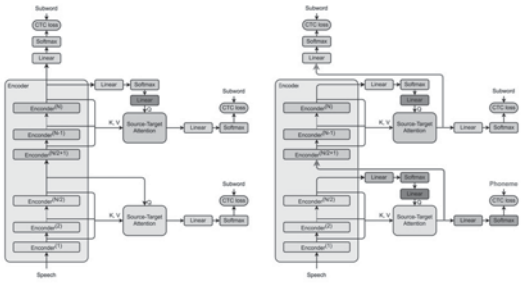


Fig.1：　Overview of the architecture of the previous method (left) and proposed method (right).

1-2-12

1-2-12　拡散モデルを用いた音声合成による音声認識のデータ拡張

Data augmentation for speech recognition using diffusion-based text-to-speech model

◎上乃聖，李晃伸（名工大）

◆ 音声合成を用いた音声認識のデータ拡張では合成音声の質が拡張性能に左右される.
◆ 本研究では，拡散モデルを用いた音声合成による拡張性能を分析する.
◆ 拡散モデルを使用時の拡張性能をより改善するために，FastSpeech 2 により生成した音声を学習初期に，その後の学習には拡散モデルで生成した音声を用いる手法 (Mix training)と，複数の事前分布を推定する手法 (Multiple mean training & sampling) を提案する.
◆ 実験結果により，FastSpeech 2 による拡張よりも，拡散モデルの拡張の方が音声認識の性能が高いことを確認した．また，Mix training，Multiple mean training & sampling の手法により test-other についてさらに改善し，両手法を同時に適用することでさらに改善が見られた.
◆ また，FastSpeech 2 の生成音声よりも拡散モデルで生成された音声の方が音響パターンが複雑になっていることを確認し，これらが音声認識のデータ拡張の性能に寄与したものと考えられる.

Table 1：ASR performance (WER (%)) for LibriSpeech testset (other). We used 100-hour paired data from LibriSpeech train-clean-100.

Model	dev	test
FastSpeech 2	14.91	14.77
Diffusion model [baseline]	13.85	14.02
Mix training	13.96	13.49
Multiple mean training & sampling	13.79	13.36
Both	**13.48**	**13.27**
Real Speech	6.68	7.06

1-2-13

1-2-13　非自己回帰型音声認識モデルにおける内部言語情報推定によるドメイン適応の評価

Domain adaptation based on internal language information estimation in non-autoregressive speech recognition models.

☆高城巽成 1，若林佑幸 1，小川厚徳 2，北岡教英 1,（1 豊橋技科大, 2 NTT）

<研究目的>
◆ 先行研究 [1,2] にて，非自己回帰型 ASR モデルにて DRA に基づく内部言語情報推定によるドメイン適応性能の向上が示されている.
◆ 使用する音声の言語や，目的ドメインの種類，学習データサイズ，ASR モデルのアーキテクチャ等の変化に対して，提案手法がどの程度頑健であるか.
<実験設定>
◆ 内部言語情報推定，アルゴリズムについては先行研究と同様.
◆ 日本語では CSJ の APS と SPS のクロスドメイン．英語ではソースを LibriSpeech，ターゲットを GigaSpeech の 6 つのサブセット.
◆ 英語では uniLSTM と Conformer の 2 つのモデルを使用.
<実験結果>
◆ 表 1 に示すように，様々なドメイン，異なる ASR アーキテクチャにおいても提案手法が頑健であり，効果があることが分かった.
◆ Conformer よりも uniLSTM の方が相対的な WER の改善が大きく，これにより 1-gram での内部言語情報推定は条件付き独立性の緩和の影響を受けることが確認できた.

Table 1　Domain adaptation performance when the source domain is LibriSpeech, and the target domain is GigaSpeech. (WER %),

Model	SubLM	AddLM	Subsets of GigaSpeech set M					
			audiobook	P-news	Y-people	Y-news	Y-science	Y-education
uniLSTM-CTC	N/A	N/A	14.4	27.8	35.0	36.0	32.4	32.5
+Shalow Fusion	N/A	4-gram	13.0	26.2	34.0	35.1	30.0	31.1
+Proposed	1-gram	4-gram	**11.8**	**23.5**	**30.7**	**31.8**	**26.8**	**28.4**
Conformer-CTC	N/A	N/A	3.69	11.6	15.4	16.4	16.3	14.6
+Shalow Fusion	N/A	4-gram	3.52	10.8	14.5	15.4	15.0	13.7
+Proposed	1-gram	4-gram	**3.37**	**10.3**	**14.0**	**15.0**	**14.5**	**13.3**

[1] 高城巽成 他, "CTC音声認識モデルにおけるビームサーチデコーディング内での暗黙的言語情報の置換", SPEASIP 2024.
[2] T. Takagi, et al. "Text-only domain adaptation for ctc-based speech recognition through substitution of implicit linguistic information in the search space", INTERSPEECH 2024.

1-2-14

1-2-14　発音プロンプトと辞書を活用した End-to-End 音声認識のキーワード認識精度改善手法

Enhancing Keyword Recognition accuracy in E2E ASR Using Pronunciation Prompting and Dictionary.

◎菅野竜雅，佐藤裕明，佐久間旭，△熊野正，
△河合吉彦（NHK），小川哲司（早大）

◆Encoder-decoder 型の音声認識モデルで、プロンプトと辞書を使って低頻度語などの認識が難しい単語の認識精度を改善する手法を提案。
◆E2E 音声認識モデルは、語彙外の単語や低頻度語で精度が著しく低下。
◆提案法では、認識が難しい単語が事前に得られると仮定。
◆得られた認識が難しい単語の発音と表記を辞書に登録。
◆モデルには辞書の発音をすべてプロンプトとして入力し、該当単語は特殊トークンで出力。
◆最後に、特殊トークンは辞書を使って正しい表記に置き換える。

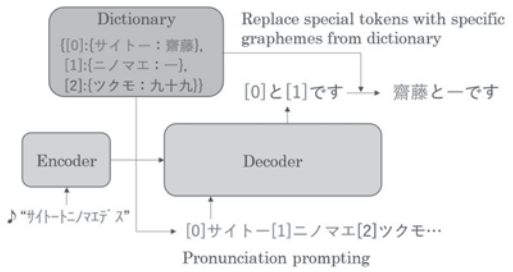


Fig.1：Proposed method

1-2-15

1-2-15 End-to-end ニューラル話者ダイアライゼーションのためのマルチチャネル話者数推定

Multi-channel speaker counting for end-to-end neural speaker diarization

◎俵直弘，安藤厚志，堀口翔太，デルクロア・マーク（NTT）

◆End-to-end neural diarization with vector clustering (EEND-VC) に基づく話者ダイアライゼーションシステムのための話者数推定法を提案

◆従来の EEND-VC システムで問題だった短時間セッションでの話者数推定性能の低下を以下の枠組みを導入することで解決（Fig.1）

- Guided source separation (GSS) の導入により，短時間チャンクから他話者や雑音の影響に対し頑健な話者埋込み抽出の実現
- 複数のチャネルから得られた話者埋込みを用いて効率的に話者数推定することで，安定した話者数推定を実現

◆提案法を NTT の CHiME-8 遠隔音声認識システムに導入することで，話者数推定精度（SCA），話者ダイアライゼーション誤り率（DER），音声認識誤り（tcpWER）の全評価においてベースラインシステムよりも大幅に高い性能を達成（Table 1）

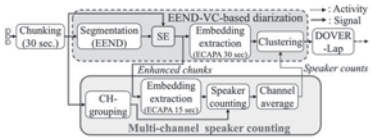


Fig. 1: EEND-VC-based speaker diarization pipeline. The red module is our proposal.

Table 1: Performance comparison with SCA [%], DER [%], and tcpWER [%]

System	SCA ↑	DER ↓	tcpWER ↓
CHiME-8 baseline	34.4	39.6	62.6
Proposed	**89.2**	**14.0**	**27.0**

1-2-16

1-2-16 年齢埋め込み特徴を用いた若年話者の年齢推定タスクについて

A study of age estimation task for young speakers using age-embedded features.

☆藤居 謙，西村 竜一（和歌山大）

◆発話信号を入力とする若年話者の年齢推定精度を向上

- 比較的小規模なデータセットの使用を想定

◆提案手法：Age-vector 抽出ネットワーク(抽出 NN)＋SGAN(識別器)

- 抽出 NN：発話音声から Age-vector を抽出
  - ✧ AgeVoxCeleb（大規模データ）で学習
- SGAN：Age-vector を入力とする識別器
  - ✧ 敵対的な学習によりデータ拡張の効果を期待

◆抽出 NN に学習データを追加、学校区分を用いた多クラス分類を実施

- 若年者・高齢者のサンプルを含む複数の学習データを追加

◆学習データの追加で Accuracy 0.09，F 値 0.08 ポイントの向上

◆全てのクラスで正解数が増加していることを確認

→ 学校区分の多クラス分類精度の向上を確認

Table1 データ追加前（上）と追加後（下）の混同行列

予測＼正解	幼稚園	小学校低学年	小学校中学年	小学校高学年	中学校	高校	大人
幼稚園	141	6	9	6	3	1	8
小学校低学年	8	162	10	10	5	6	5
小学校中学年	16	18	193	16	9	7	11
小学校高学年	10	9	10	165	10	16	10
中学校	4	2	15	6	144	11	23
高校	3	0	3	3	6	126	20
大人	15	9	12	15	62	94	756

予測＼正解	幼稚園	小学校低学年	小学校中学年	小学校高学年	中学校	高校	大人
幼稚園	179	7	14	1	3	2	5
小学校低学年	5	183	10	3	2	4	9
小学校中学年	6	11	206	15	7	4	2
小学校高学年	2	4	18	194	10	12	15
中学校	1	1	3	7	181	17	17
高校	2	1	1	3	3	176	16
大人	6	2	4	8	37	54	786

1-2-17

1-2-17 歌唱音声の特性を考慮した歌唱者照合のための頑健な特徴抽出器の構築

Differences Between Singer and Speaker Verification: Training Singer Feature Representation Extractor Utilizing Singing Voice Characteristics

☆当間佐耶佳，有賀智輝，樋口陽祐（早大），
△早坂一寿，△執行里恵（第一興商），小川哲司（早大）

< アプローチ >

◆短時間の音響変動が小さく，長時間にわたる音響変動が大きいという歌唱音声の特性を考慮し，短く分割して入力することによって歌唱者内の音響変動を正確に捉え，頑健な特徴抽出器を構築する．

Fig.1：Dividing singing audio into shorter segments for pooling allows for more accurate representation of within-speaker variations

< 結果 >

◆短いセグメントに分割して特徴抽出器を学習することで，歌唱者照合における誤棄却率を効果的に低減できることを示した．

◆学習時と照合時の入力長の違いによる影響よりもセグメント化して学習することの効果のほうが大きいことが示唆された．

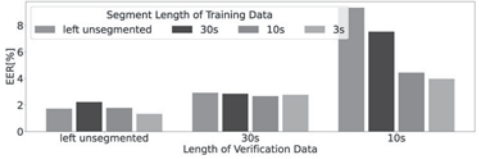


Fig.2：EER for each training data length and test data length

1-2-18

1-2-18 自己教師あり学習特徴を用いた，音声感情認識と発話区間検出の End-to-End 統合

End-to-End Integration of Speech Emotion Recognition with Voice Activity Detection using Self-Supervised Learning Features

◎山下夏生，山本正明，川口洋平（日立製作所）

◆音声感情認識（SER）は，しばしば発話区間検出（VAD）モデルによって検出された音声セグメントを処理する。

◆しかし，VAD モデルは特にノイズの多い環境で不正確な音声セグメントを出力する可能性があり，その結果，後段の SER モデルの性能が低下する。

◆そこで，自己教師あり学習（SSL）特徴を使用して VAD と SER を end-to-end に統合する手法を提案する。

◆IEMOCAP データセットを用いた評価では，提案手法がノイズの多い環境での SER 性能を大幅に向上させることを確認した。

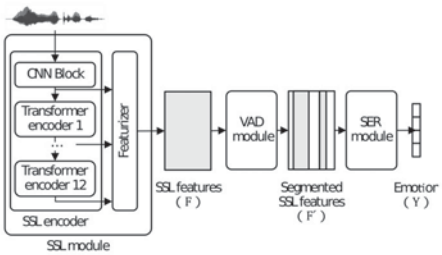


Fig.1：Overview of the proposed end-to-end approach composed of SSL, VAD, and SER modules.

1-3-1

1-3-1　オブジェクトベース音響のレンダリングにおける音声歪の防止法

Prevention method for audio clipping in rendering of object-based audio

◎久保 弘樹, 岩崎 泰士, 大出 訓史(NHK)

◆一般にディジタル音声信号は0 dBFS を超過すると音声歪（クリッピング歪）が生じる．現行の放送などでのライブ制作では，音声レベルが閾値を超過しないようにリミッタを用い音声歪を防止する．閾値は0 dBFS まで数 dB 程度の固定値のマージンを持って設定される．

◆しかしオブジェクトベース音響（Object-based audio: OBA）では視聴者が番組音声をカスタマイズして再生信号を生成（レンダリング）できるため，通常のリミッタでは音声歪を完全には防止できない．

◆OBA での音声歪の防止法を検討するため，音声信号の加算やレベル変更を伴う様々な条件でレンダリングした番組音声のサンプルピーク(SP)，トゥルーピーク(TP)，ラウドネス値の変化量を測定した．

◆その結果，番組やレンダリング条件によっては5 dB を超えるレベル変化が見られ，番組によらず固定値のマージンを設定すると番組音声のダイナミックレンジが制限されることが示唆された．

◆OBA での音声歪防止法として番組ごとにマージンを最適化することを提案し，今回測定した音源に適用した結果，固定値のマージンを用いるのに対しダイナミックレンジを2 dB程度広く保ちつつ音声歪を防止することができた．

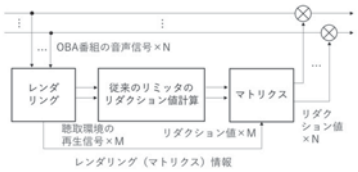


Fig.1 : Prevention method for audio clipping of OBA

1-3-2

1-3-2　音楽ライブ音源を用いたマルチチャネル曲げ波スピーカシステムの評価の試み

Evaluation of Multichannel Bending Wave Loudspeaker System Using Live Performed Music

☆橋本篤拓, 河原一彦(九州大･芸工)

◆河原により提案された Lab-made BWL は「いわゆる放射の拡散的性質」を最適化する設計法で設計されている。

◆Lab-made BWL を取り入れたマルチチャンネルスピーカシステムを利用し，市販の音楽ライブ音源を聴かせることで臨場感と包まれ感，その他の主観的要素が向上したかどうかを主観評価実験を行い検証した。

◆実験の結果から，観客の歓声や手拍子が多く含まれた音を Lab-made BWL から再生した場合に，臨場感・包まれ感・OLE・没入感が上昇する傾向がみられた

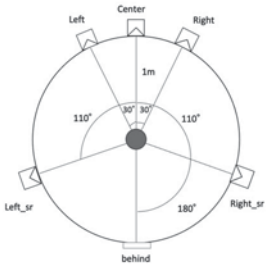


Fig.1 : Multi-channel bending wave speaker system block diagram

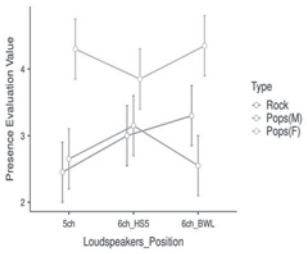


Fig2. Box-and-whisker plot of means and 95% confidence intervals foe rated value of presence

1-3-3

1-3-3　正面方向の実音源に対する開放型ヘッドホン音響透過性補正の検証

Evaluation of acoustic transparency compensation for open-back headphones with a frontal sound source.

◎渡邊悠希, 千葉大将, 野口賢一, 加古達也, 伊藤弘章, 鎌本優 (NTT)

◆近年，現実の聴覚環境を拡張するオーディオ拡張現実が実用化されている．

◆ただし，音響デバイスの装着により実世界の音の伝達特性が変化し，それに伴い実世界の音の知覚が変化する．この特性の変化はデバイス装着状態と非装着状態の伝達関数の比で表され，音響透過性と呼ばれる．

◆本研究では，スピーカ再生音源に対して音響透過性補正フィルタを設計し適用することで，音響デバイス装着者の音響透過性を補正する方法を提案する．

◆主観評価実験の結果，オーバーイヤー型ヘッドホンでは提案法の補正フィルタによって類似度とみかけの音源の幅がデバイス未装着の状態に有意に近づくことが示唆された．

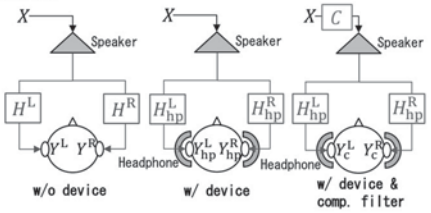


Fig.1: Audible sound without and with headphones, and proposed method.

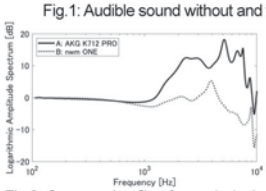


Fig.2: Compension filter for each device.

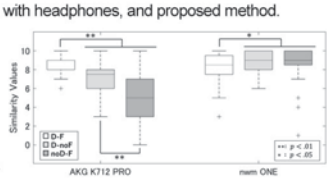


Fig.3: Similarity values for each device.

1-3-4

1-3-4　多チャンネル 3D マイクロホンアレイのための小型軽量フィールドレコーダの開発*

Development of a compact and lightweight field recorder for multi-channel 3D microphone array

☆戸谷僚, 伊勢史郎(東京電機大)

◆BoSC システム（80 ch）や HOA システム（152 ch）を用いて収録する場合には数十チャンネルを超える大規模なオーディオレコーダが必要となる。そこで，BoSC マイクを用いたフィールドレコーダを前提とした80 ch 音響信号の同期収録を可能とする小型軽量フィールドレコーダを開発した。レコーダおよび電源を Fig. 1 に示す。

◆小型軽量フィールドレコーダは 10 枚の 8 ch マイクロホンアンプ基板，音響信号をマイクロ SD カードに収録する SoC-FPGA 素子（AMD-Xilinx 社製）を搭載した基板およびそれらのインターフェース基板から構成される。

◆無指向性小型マイクロホン（PUI 社製 AOM-5024）80 個を 3D プリンタで作成した剛球（直径 23 cm）に取り付け，音響樽内に設置してレコーダと接続し TSP 信号を収録した。TSP から計算したインパルス応答の波形を Fig. 2 に示す。

Fig.1 Recorder and power supply

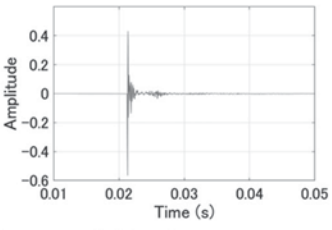


Fig.2 Impulse response

1-3-5

1-3-5　相関制御に基づくインタラクティブ仮想音源の見かけの音源の幅制御の検討

Investigation of Apparent Source Width Manipulation for Interactive Virtual Sound Sources Based on Correlation Control.

☆澤尻晃大，羽田陽一（電通大）

◆インタラクティブオーディオでは，身体動作などが音の空間的印象に影響を与える可能性がある．本研究では見かけの音源の幅に注目して調査を行った．左右・上下に配置した２つの仮想音源 (Fig.1) を用いて提示された音像を両手で上下・左右に広げる動作を行う．実験の様子を Fig.2 に示す．手の移動方向と音源の配置方向が一致または異なる場合において，見かけの音源の幅の変化への影響を明らかにする．

◆モーションキャプチャと 2 仮想音源の原信号間の相関制御によって，両手間の距離に応じて音像の見かけの幅を制御する．

◆両耳間手がかりが得やすい水平音源配置ではASWの変化が明確に知覚され，左右方向の相関を弱めると，左右の手の動きに対しては音像が左右に，上下の手の動きに対しては音像が上下・左右方向に同時に広がるように知覚された．この結果は，手の動きが音像の広がり方向に影響を及ぼす可能性を示唆している．

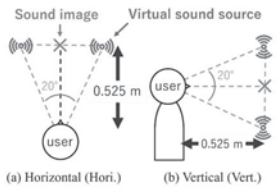


Fig.1：Source layouts

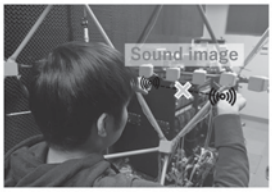


Fig.2: The experiment

1-3-6

1-3-6　直線スピーカアレイの前後を通過するインタラクティブ仮想音源の奥行き移動について

Distant movement of interactive virtual sound source passing through the front and back of a linear loudspeaker array

◎末藤立己，澤尻晃大，羽田陽一（電通大）

◆ 投げる動作に合わせて移動する仮想音源に対し，被験者の想定通りの位置に移動する場合とそうでない場合の奥行き移動感を評価した。

◆32ch 直線スピーカアレイとモーションキャプチャを用いた主観評価実験を，投げる距離を指示した仮想音源の終点位置より実際に合成した仮想音源の終点位置が手前にずれる，奥にずれる，一致するパターンの３つにわけて実施した。

◆Fig. 1 は音源として電車の音を使用した場合における，方向別の平均評価値の結果である。横軸は，（指示した終点位置，提示した終点位置）m を，縦軸は奥行き感を表す。

◆結果より，奥行き感は実際に提示した仮想音源の終点位置に影響を受ける傾向があることがわかった。

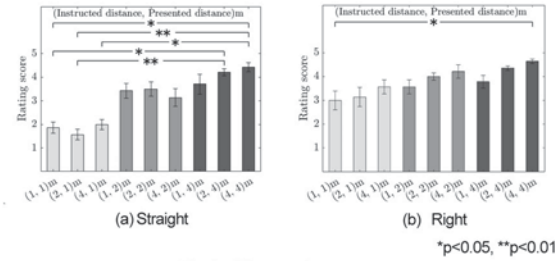


Fig.1：Mean rating score

1-3-7

1-3-7　簡易音場再生のスイートスポットに関する考察

Study on the Sweet Spot of Simplified Sound Field Reproduction, Tottedashi

☆中島佑樹，岩見貴弘，尾本章（九州大・芸工）

◆簡易音場再生(Tottedashi)は，24 ch の鋭指向性マイクアレイとスピーカアレイを用いて，簡易的に音場の方向情報を再生する方法。

◆鋭指向性マイクモデルを利用した収音・再生シミュレーションにより，再生音場を広範囲にわたって評価。(評価指標：音圧レベル [dB])

◆高次アンビソニックス(HOA)の結果から，1/3 オクターブバンド毎の評価において，音圧レベルの平均誤差が約2 dB 以下となる領域をスイートスポットと定義し，周波数毎に最大半径を算出。

◆簡易音場再生のスイートスポットは，低域から中域にかけて HOA のスイートスポットの理論値の約半分の大きさだが，マイクの指向性が鋭くなる高域において，大幅に拡大することを確認。

Fig.1：Narrow directional microphone array

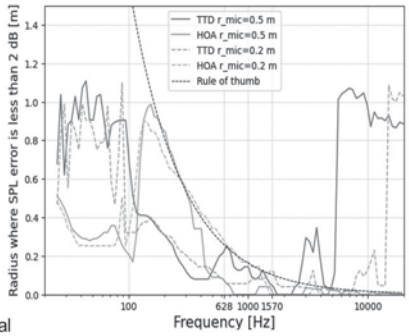


Fig.2：Sweet Spot in SPL distribution

1-3-8

1-3-8　3D オーディオにおける中層と上層の相関と聴取範囲について

The correlation between the middle and top layers and the listening area in 3D audio

○亀川徹，丸井淳史（東京芸大）

◆3D オーディオにおける上層のマイクロホンの指向性と高さによる聴取印象との関係について調査した。

◆シェッフェの一対比較（中屋の変法）の結果から，単一指向性マイクロホンを中層から離して設置した場合に空間の広がり感を感じる傾向が見られた。

◆マイクロホンの指向性や高さによって上下方向の相関値が低くなることで，中央の聴取位置と印象が変わらない聴取範囲が広くなったことから，横方向や後方の聴取位置で全指向性マイクロホンや単一指向性で高さが低い場合の印象の違いが明確に見られるようになったのではないかと考えられる。

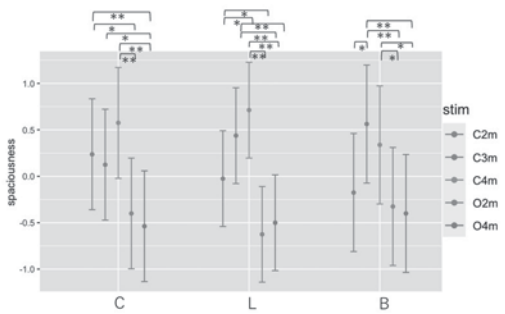


Fig.1：Means and 95% confidence intervals obtained from pairwise comparisons of the xylophone.

1-3-9

1-3-9　音場合成における波数領域音像スケーリング処理の検討

Scaling processing of virtual source in wave number domain for sound field synthesis

○佐々木陽，中山靖茂（NHK）

◆将来の音声サービスとしてボリュメリック音声制作・再生に関する研究を行っている．

- **制作技術**
  包囲型マイクロホンアレイを用いて取得した内部音源の放射特性を球面調和スペクトルとして推定し，音源オブジェクト固有の特徴と捉えてメタデータ化
- **再生技術**
  放射特性や位置などの音源オブジェクトに関するメタデータを基に音場を合成

◆放射特性として球面調和スペクトルが与えられた音源の放射場を合成する際に，仮想音源の大きさを変化させるスケーリング処理を提案．

◆数値シミュレーションにより，提案法を用いてスケーリング処理を加えることで，音源の大きさが変化したときの放射場が合成可能であることを確認．

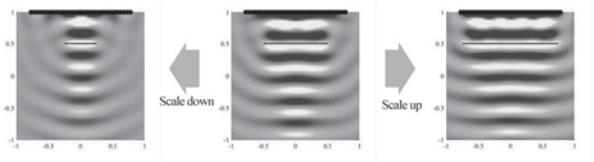


Fig. 1: synthesized sound fields (left) scaled-down field at a scale factor of 0.5 (center) original sound field (right) scaled-up field at a scale factor of 1.5

1-3-10

1-3-10　BoSC システムを用いた室形状の複雑さをパラメータとするバーチャル 3D 室内音場の実現

Virtual 3D room sound field with room geometry complexity as a parameter using BoSC system

☆大橋厚郎，伊勢史郎（東京電機大）

◆本研究では，羽入らによって提案された室内音響理論に基づくインパルス応答生成アルゴリズムを用いてバーチャルな室内音場を音響樽内に生成するシステムを開発し，このシステムを用いた演奏支援の可能性についての検討を行った。

◆バーチャル 3D 室内音場とそれと同程度の残響時間である Nuendo 付属のホールリバーブをかけて音響樽の前方中央の 12 個のスピーカーから出力した条件の 2 条件で自由に楽器演奏をさせ，被験者 10 名に対し実験を行った。楽器演奏時の音響樽内の様子を Fig 1 に示す。

◆再び演奏をするとしたらホールリバーブとバーチャル 3D のどちらが良いかという質問に対し 10 人中 8 名がバーチャル 3D の条件と回答し，室形状の複雑さのパラメータからインパルス応答を生成するアルゴリズムが BoSC 方式によるリアルタイム音場再現において有効であることを示した。

Fig.1：Example when playing an instrument

1-3-11

1-3-11　2 つの BoSC 再生室を用いた 3D オーディオ受聴システムの構築 ―聴取体験の共有と会話の効果の検討―

Construction of a 3D audio listening system using two BoSC sound field reproduction room –Effects of sharing and talking about the listening experience–

☆安念佑真，上野佳奈子（明治大），平山大祐，伊勢史郎（東京電機大）

◆聴取体験におけるコミュニケーションの効果に着目し，2 つの BoSC 再生室を用いて，3D オーディオ再生と共に，自由音場における会話を三次元的に再現するシステムを構築した（Fig. 1）。

◆友人 2 人のペア 10 組を対象に心理評価実験を行ったところ，1 人で 3D オーディオを聴くより，2 人で会話し感想を共有しながら聴く方が聴取体験を楽しめたという評価結果が得られた。

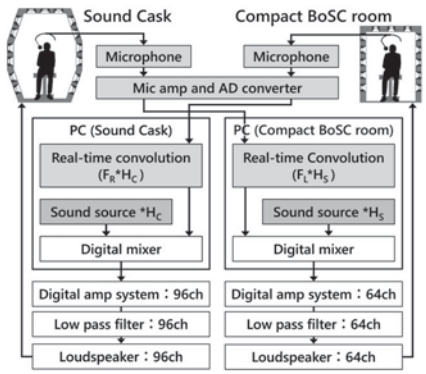


Fig.1：Sharing system about the 3D audio listening experience. (F_R / F_L：Impulse response from the conversation partner positioned to the right / left to the listener's position)

1-3-12

1-3-12　楽器の物理モデル音源は VR･AR においてどのように使われるだろうか

How do we use physical modeling sound synthesis of musical instruments in VR/AR?

○鮫島俊哉（九大･芸工）

◆楽器の物理モデルの高忠実度化，および数値計算の高効率・高精度化の過程で，楽器の発音体の部分だけではなく，奏者と楽器のインターフェース部分と演奏動作も考慮した物理モデル化を試みてきた。

◆例えば，ピアノのハンマーシャンクと鍵盤へのタッチ，ヴァイオリンの弓とその弾き方（Fig.1），シンバルのスティック/マレットとその握り方（Fig.2）などである。それらの研究事例を紹介する。

◆楽器を“真に理解”し，そして VR・AR において“真に使える”楽器を作るためには，感覚フィードバックに基づく奏者と楽器のインタラクションを考慮した物理モデルが必要であると著者は考える。紹介する研究事例は，その一歩手前にある。

◆楽器の構成要素のアレンジ（例えば Fig.3）による“新規楽器の創生”も，楽器の物理モデルの VR・AR における使いどころの一つである。

Fig.1：Effect of the bow tilt.

Fig.2：Effect of the mallet grip condition.

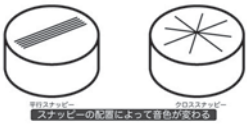


Fig.3：Alternative snare arrangement: cross arrangement.

1-3-13

1-3-13　仮想音源サンプリングによるバーチャルレコーディング

Trial of a virtual recording scheme for music production by sampling virtual sound sources of a target space.

○中原雅考（ソナ／オンフューチャー），尾本章（九大・芸工／オンフューチャー）

◆バーチャルレコーディングを実施する手法 V2MA (VSVerb Virtual Microphone Array) を紹介する。

◆V2MA は，幾何音響的に元音場の響きを再現する手法である VSVerb をベースとして，オーディオ工学的な手法で元音場の響きを創造する手法である (Fig. 1)。

◆V2MA では，元音場でサンプリングした仮想音源情報を幾何音響的な手法を用いて更新することで，任意のマイク位置や指向性，また音源位置でのリバーブの提供が可能である。

◆仮想音源の抽出は，対象空間において測定した３軸方向の音響インテンインテンシティーの空間移動速度情報を利用して実施する。この「速度検知」の手法を用いることで，後期反射音の抽出及び，それらの位相の検知を実現している。

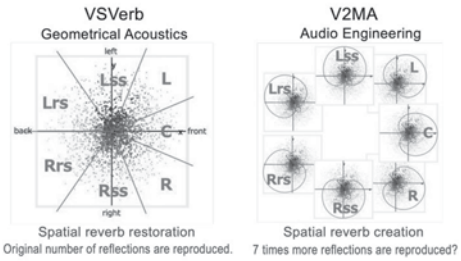


Fig.1 Detected virtual sound sources and CBA rendering; VSVerb and V2MA

1-3-14

1-3-14　建築音環境の波動音響 VR シミュレーション

Wave-based VR acoustic simulation of architectural sound environments

○奥園　健（神戸大院・工）

著者らは、室内音響設計への応用を目指し、建築の音環境を波動音響解析により高効率にモデル化する要素技術の開発に取り組んできた。近年、Unity などのゲームエンジンの普及によって、VR 環境上で現実的な建築空間を容易に構築できるようになっている。また、立体音響再現技術の進歩により、空間的な音響情報を含む高度な表現が可能となっている。

このような技術的背景を踏まえ、近年、著者らは波動音響解析技術と VR 向けの立体音響再現技術であるバイノーラルアンビソニックスを組み合わせることで、波動的な音響情報を保持しながら VR 環境内で室内音響を空間情報まで含めて可聴化する波動空間音響シミュレーション技術を開発し、その可能性を探求している。さらに、この技術を応用して、吸音設計の教育支援コンテンツや室内音響設計支援・音響材料開発支援ツールへの展開にも取り組んでいる。

本発表では、著者らの空間音響シミュレーション技術とその応用例について、関連文献を参照しつつ簡単に紹介したい。

1-3-15

1-3-15　Full Audio Bandwidth Wave-based Acoustic Simulation: State of the Art and Challenges

○Stefan Bilbao, △Jan Smits (University of Edinburgh)

Wave-based numerical simulation approaches offer, in theory, a complete solution to the problem of auralisation for a virtual acoustic space. Obvious applications are in architectural acoustics and in virtual reality. Such an auralisation requires the generation of output over the full audio bandwidth up to 20 kHz. At present, most approaches to wave-based acoustic simulation are restricted to the low frequency range, which, for auralisation purposes, can be complemented by geometrical acoustics results to cover the higher frequencies. However, even for commercial-grade hardware, computing power is considerable now, and the possibility of full-bandwidth wave-based virtual acoustic simulation over the full bandwidth is becoming less remote.

The design of wave-based simulation methods is constrained by a variety of competing requirements, including: robust behaviour, or the need to maintain numerical stability under very complex design choices with minimal user intervention; accuracy, or the minimization of solution errors, perhaps to be considered in a perceptual sense; and efficiency, or the minimization of the required runtime and memory requirements. These constraints are interlinked, and the last must be approached with great care when scaling to very large problems.

In this paper, we examine these design considerations in full, focusing on methods defined over regular grids such as the finite difference time domain method. Simulation results are presented.

1-4-1

1-4-1　音声対話を用いた多様性の相互理解と社会参加支援を目指して

Toward Mutual Understanding of Diversity and Support for Social Participation Using Spoken Dialogue

○越智　景子（京大）

◆これまで、支持的な態度で共感的に相槌を打ち、理解を示す音声対話ロボットを、スピーチや面接の練習相手、さらには人間同士をつなぐ役割を持たせていくことを目指し、社会実装を行ってきた。

◆2 人のユーザのうち片方が 3 分好きな話題について語り、その後交替してもう一人が 3 分語る実験では、1 人のユーザとロボット「きくロボ」が傾聴を行った。

◆語り手は、聞き手から相槌や質問をたくさんもらったとき、たくさん話すことができた。

◆たくさん話せた人ほどその人自身の気分の改善が見られた。

◆さらに、相手からたくさん話し手もらった人ほどその語り手に対して肯定的な印象を持った。

◆精神科領域でも、共感的な応答や興味深く質問する対話が互いの理解を深め、ロボットがそのサポートとして貢献できる可能性がある。

1-4-2

1-4-2　神経多様性としての吃音

Developmental Stuttering as a Neurodiverse Speech Style

○森浩一（国立障害者リハビリテーションセンター）

◆吃音は発話が流暢に出ない障害であるが，そのほとんどは発達性吃音であり，幼児の約 1 割に発症し，8 割程度が自然治癒するが，1%弱は青年期以降も持続し，コンプレックスとなって重症化しやすい。

◆発症原因としては遺伝要因が 8 割程度を占め，親の育て方や愛情不足，あるいは真似をするなどは原因にならないことがわかっている。

◆吃音関連として 30 以上の遺伝子変異が同定されており，このうちのいくつかが重なって発話関連の脳機能が低下すると，吃音が発症する。

◆吃音は自然治癒が多いだけでなく，治療による改善（医療モデル）もあるが，根本の遺伝子変異や脳の一部の機能不全はそのまま残る。

◆改善には個人差があり，治療でも改善しない症例や，改善後も再発する症例もあり，公平な社会参加を保証するためには，社会が障害を包容（inclusion）し，合理的配慮を提供する必要がある（社会モデル）。

◆障害者差別解消法などの制度の整備は進んでいるが，社会の理解は不足している。典型発達の能力を前提とした言動は ableism の表明になっていることがあり（例：「挨拶くらいできるでしょう？」），吃音がある者はそういう当然のことができない自分を意識させられ，心理的に傷つき，microaggression として受け取られることがある。

◆さらに，吃音を治療すべき障害ではなく，神経多様性の現れとして捉える考えがある。ヒトの多様な行動様式の 1 つとして吃音を社会が理解して受容し，適切な合理的配慮を提供できれば，吃音者が症状を隠したり修正したりする努力が不要になり empowerment につながる。

1-4-3

1-4-3　発達期における多様な音声生成

Exploring diversity in speech production of infants and toddlers

○保前文高（東京都立大・人文社会）

◆乳幼児が発する音声の多様性と発達的な変化を捉えるために、音圧変調と音節間の遷移確率を解析した。NTT 乳幼児音声データベースに収録されている、縦断的に録音された 6～15 か月児の音声を対象とした。

◆6・7 か月児では音圧の変化が低周波数ほど大きく、特定の構造が見いだされなかったが、10・11 か月児と 14・15 か月児では、2～3 Hz で極大となり、4～5 Hz で極大を示す成人の音声に近い傾向を示した。

◆全ての月齢群で、母音が[a]である音節（「あ」）が発声される割合が高く、「あ」から「あ」へ遷移する確率が高かった。6・7 か月児では「あ」と「う」を含む遷移が多く見られるが、10・11 か月児と 14・15 か月児では、それらとともに、「い」を含む音節間の遷移が増加した。

◆音圧の変動の周波数特性と、舌の位置や口腔の開き度合いを交互に変化させる発声が増加することの関係を解明する枠組みが必要である。

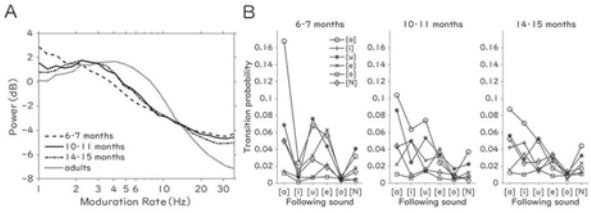


Fig. 1: A) Mean modulation spectrum of speech sounds produced by infants, toddlers, and adults. B) Transition probability between spoken syllables. Lines indicate preceding sounds in the sequence.

1-4-4

1-4-4　多文化共生社会のために音声教育でできること

The Role of Pronunciation Education in a Multicultural Society

○木下直子（早稲田大学）

日本における在留外国人数は過去最高を更新し，多文化共生社会の実現が重要な課題となっている。本研究では，「World Englishes」（Kachru 1985, Jenkins 2000）と「聞き手の国際化」（土岐 1994, 2010）の視点から，音声教育が目指す方向性として「公平な耳社会の実現」を挙げ，学校教育等で一般の日本語母語話者（NS）に対し，外国語訛りの日本語を理解する機会の必要性を述べた。具体的には，２つの調査を紹介している。調査１は，NS が日本語学習者（NNS）の音声の処理速度（Rapid Adaptation）を検討した研究である。この結果から，NS は NNS 音声の処理にかなりの時間を要していても，16 文の NNS 音声を聞くと，その処理速度が短縮されることがわかった（Kinoshita & Sheppard 2022）。調査２では，NNS の発音適応度と異文化受容態度の関連性を分析した。その結果，発音適応が早い人ほど，移民の受け入れに積極的な態度を示す傾向が明らかになった（木下・シェパード 2024）。以上２つの結果により，NS が NNS の音声を理解するための具体的な方法として，16 文を聞くタスクが有効であること，発音適応には異文化受容態度との関係があることが示唆された。先行研究には，発音適応度が高まると，その言語話者に対する偏見や態度に改善が見られたという報告（Derwing, et al.2002）もあり，多文化共生社会に向けて音声教育が担う役割は大きい。今後，学校教育への導入を目指し，現場の教員と共に実践的な教育手法の検討を進め，学習リソースの開発を行っていく必要がある。

1-4-5

1-4-5　リアルタイム MRI データからみた日本語の母音組織

Japanese Vowel System Analyzed by Real-Time MRI Data

○前川喜久雄（国語研）

◆調音音声学は長らく観察の主観性という根本問題を抱えてきた。近年、リアルタイム MRI（rtMRI）撮像技術の実用化によって音声生成時の声道正中断面全体を被爆の危険なしに動的に可視化することが可能となったことなどによって、調音音声学の知見を全面的に再検討する時期が到来している。本稿前半では、筆者らによるリアルタイム MRI データのオープンデータ化の試みと、日本語子音に関する既発表の知見をまとめる。その後、後半では母音に関する新しい知見を報告する。

◆日本語の母音組織は、1)舌体を後上方に引き上げる調音が微弱であることと、2)その代償として口唇部で精密な制御が行われていることの２点によって特徴づけられる。これらは rtMRI による観測結果を前田・本多の調音モデルに従って解釈したものであり、下図のようなコンパクトな母音空間を構成していると考えられる。

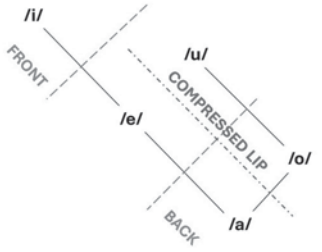


Figure: System of Japanese vowels and relevant phonological features.

1-5-1

1-5-1　クラシックとジャズで用いられる和音進行に対するプローブ音評定：ジャズ経験の有無による評定値の変化

Probe tone ratings for chord progressions used in jazz: Changes in ratings due to jazz experience

☆服部大生，松井淑恵（豊橋技科大院）

◆ジャズ経験によって音階外音の評定が変化するかを調査した．

◆プローブ音評定法で調査した．先行刺激はＩ度に終止する和音進行とし、(a) クラシック刺激，(b) クラシック+ジャズ終止音刺激，(c) ジャズ刺激の３種類とした．プローブ音は音階音の 12 音とした．

◆実験参加者は，ジャズの演奏経験がある群 15 名とジャズの演奏経験がない群 15 名とした．

◆実験の結果，参加者のジャズ経験による明確なグラフの変化は見られなかった．ジャズ演奏歴の短さが影響した可能性がある．

◆先行刺激によるグラフの変化が見られた．ジャズ刺激は音階外音を 1-3 個含むため，音階外音の適合度が上昇し，グラフが平坦化したと考えられる．

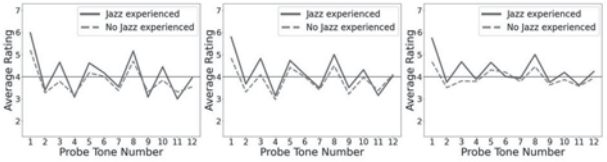


Fig. 1：Mean rating for each probe tone. The preceding stimuli were: (a) classical chord progression, (b) classical chord progression with jazz tonic, and (c) jazz chord progression. The solid line represents the jazz-experienced group, while the dashed line represents the no-jazz-experienced group.

1-5-2

1-5-2　テキストプロンプトによる楽器パート単位の編集を実現する楽曲編集 AI

Music Editing AI for Instrumental-Part Level Modifications through Text Prompts

☆池田尚騎，杉浦陽介，島村徹也（埼玉大）

◆近年，音楽分野では生成 AI を使った Text-to-Music（TTM）の技術が進展しているが，楽曲の部分的な編集を行う生成 AI はほとんど存在していない．

◆本研究では，生成過程で編集対象外の楽器パートと楽曲全体の特徴量の加重平均を取ることで，特定の楽器パートのみを変化させ，編集を行う手法を提案する．

◆提案モデル（Fig. 1）を用いて各楽器パートの編集を行い、その編集結果を既存の音楽生成 AI と評価指標を用いて，数値的に比較した．

◆これにより，特徴量を用いた編集の実現可能性を示したが，生成精度の向上や編集可能な範囲の制限といった課題が存在している．

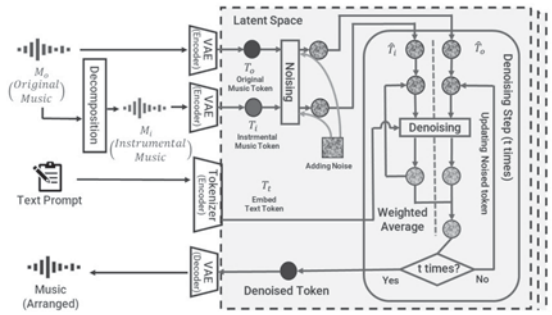


Fig.1：Architecture of the Proposed Model

1-5-3

1-5-3　小節特徴量を活用した楽曲の大局的構造を反映した自動作曲

Symbolic music composition capturing global structure of music based on bar-feature sequence modeling

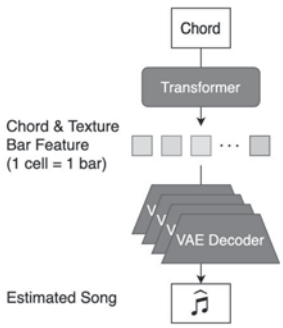
☆澤田桂都，Huang Wen-Chin，戸田智基（名大）

◆　自動作曲が抱える課題：**楽曲の大局的生成性能**

- 反復とその際の変奏の有無のバランス
- 小節や節を跨いだ楽曲構造

◆　提案手法：**小節特徴量を介した 2 段階の生成 (Fig. 1)**

- 楽曲を小節ごとに低次元の特徴量で表現
- 特徴量系列の生成 & そ hi れに基づく楽曲の生成



サンプル視聴
Webページ

Fig.1：Overview of inference architecture in proposed method

1-5-4

1-5-4　アコースティックギターのサウンドホール特性再現に基づくエレキギター音変換

Electric guitar sound conversion based on reproduction of acoustic guitar sound hole characteristics

☆古田俊樹，周桐，片岡章俊（龍谷大院・理工学研）

近年，家庭での音楽制作が普及し，アコースティックギターは多くの楽曲で使用されているが，録音時には防音環境が必要であり，コストや環境面で課題がある．一方，エレクトリックギターは雑音が少なく録音が容易であり，これにアコースティックシミュレーターを適用することで，防音設備を必要とせずにアコースティックギターの音を再現できる可能性がある．しかし，現行のシミュレーターは倍音成分の再現性が低く，実際の音響とは異なる．本研究では，アコースティックギターの空洞構造を物理モデルで再現し，音響特性を改善する手法を提案する．主観評価実験の結果，従来手法と比較して提案手法が倍音の再現性を向上させ，アコースティックギターに近い音質を実現できることを確認した．

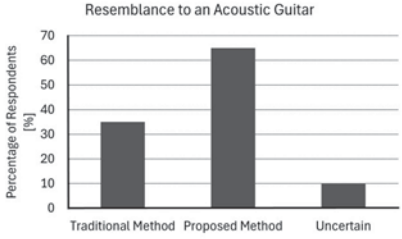


Fig.1：Comparison of resemblance to an acoustic guitar based on subjective evaluations

1-5-5

1-5-5　NMF を用いたギター音源からベース音源の生成：実楽曲を用いた実験

Generating Bass Phrase from Guitar Chord Backing with NMF: Experiments with Real Music Recordings

☆香西智雄(日大)，小口純矢(明大)，北原鉄朗（日大）

◆目的: ギター音源からベース音源の生成。

◆提案手法: VAE と NMF を用いて、ギターの特徴量からベースのアクティベーション行列を学習させる。

◆データセット: BUMP OF CHICKEN の楽曲（全 119 曲）を音源分離させたギター・ベース音源。BPM は 120 固定。

◆学習データ: 上記のデータセットのうち 106 曲。

◆テストデータ: 上記のデータセットのうち学習データ以外の 13 曲。

◆結果: 概ねギターの演奏内容に沿ったベース音源の生成は出来たが、リズムや音量のバランスを考慮できなかった。

◆今後は新たな特徴量として音量を加え、被験者実験を行いたい。

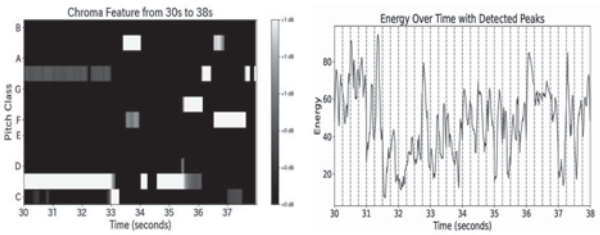


Fig1: Chromagrams of the generated bass sound (Left) and Onset_strength of the generated bass sound (Right) of "天体観測"

1-5-6

1-5-6　調波打撃音分離を介したスパース CQT 表現

Sparse CQT representations via harmonic/percussive source separation

◎ 新井慶大，矢田部浩平（農工大）

背景　CQT：中心周波数に応じて窓幅を変える変換
音楽信号の時間周波数解析に用いられることがある

- 正弦波成分：対数周波数軸上で等幅に表現
  ► 周波数の解析に適している
- 打撃音成分：低域の成分が時間方向に広がる
  ► 瞬時性が表出しにくい

△　両方の成分を含む信号の解析が難しい

提案　正弦波・打撃音成分を分離する**調波打撃音分離**を利用
► 両方の成分の解析に適した**スパース CQT 表現**

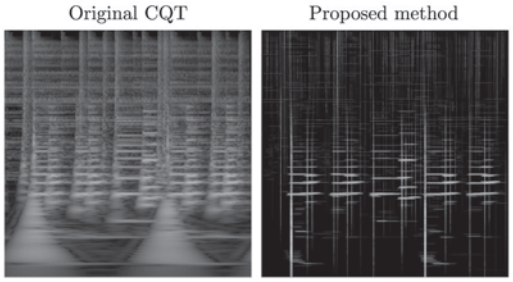


Fig. 1　Analysis results of a music signal by CQT and the proposed method

1-5-7

1-5-7　エレクトリックギター演奏におけるフィルタ処理および周波数解析処理によるフィンガリングノイズ検出法

Detection method of fingering noise in electric guitar playing by filter processing and frequency analysis

☆齋藤太陽，及川靖広(早大理工)

◆背景：アコースティックシミュレータには、フィンガリングノイズ(FN)再現に関して課題がある。フィンガリングノイズ再現の第一段階として、エレキギター演奏におけるフィンガリングノイズ検出手法が必要である。

◆提案手法：ドライ音に対し、マイク音の周波数特性を模倣させるための IIR フィルタ処理を施し、変換音を生成する。マイク音、変換音の FFT 比より判定値を算出し、その大小比較により検知を行う。

◆結果：検知率 91%、誤検知率 6.6%となり、提案手法は一定の成果をなしている。

◆課題：より高い実用性の実現のため、実験環境の工夫などにより、フィンガリングノイズ以外のノイズ検知の減少を模索する必要がある。

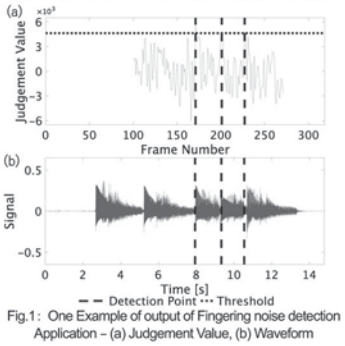


Fig.1：One Example of output of Fingering noise detection Application – (a) Judgement Value, (b) Waveform

Table 1 : The result of Fingering Noise Detection Process

Real FN	190
All Detection	243
FN Detection	173
Other Noise	54
False Detection	16
Detection Rate	91 %
False Rate	6.6 %

1-6-1

1-6-1　音響特性と主観評価に基づくターゲットサウンド手法の提案

Proposal of a target sound method based on acoustic characteristics and subjective evaluation

☆河原怜雄(中央大院)，戸井武司(中央大)

◆本論文では，目的とする機能音を制作する際に，デザイン性が変化しても機能性が担保されるよう，デザイナーに提示する音響特性と主観評価に基づく条件（以下，機能性条件と呼ぶ）の抽出方法を提案する。また，実際に抽出した機能性条件を付与してデザイナーに制作された音と，条件を付与せずに制作された音を用いて主観評価実験を行い，機能性条件の抽出方法の妥当性を検証する。

◆主観評価実験から，時間帯に応じた音の感じ方に影響を及ぼす音響特性である「パターン数」，「パターン内音符数」，「音の調子」，主観評価の条件は 6 形容詞対の評価得点とした。制作音は，イメージさせる「時間帯」と「メッセージ」のみの機能性条件なしと条件ありで制作し，主観評価実験を行った。

◆Fig. 1 に示す主観評価より，(a)目標とした機能音と(b)機能性条件を付加した制作音が朝，昼および夜の割合で同じ時間帯の推移をすることを把握した。したがって，音響特性と主観評価に基づいた機能性条件により，目標の「時間帯」イメージの音が制作できたことから，デザイン性が変化しても，機能性が担保できることを明らかにした。

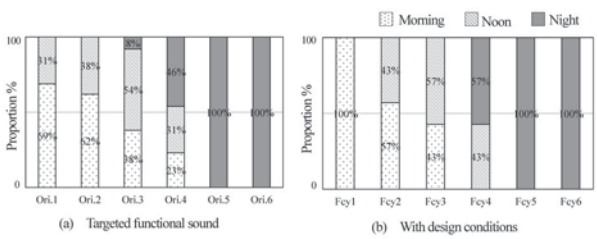


Fig. 1　Time zone evaluations

1-6-2

1-6-2　純音成分の認知を考慮した回転次数音の快音設計

Comfortable Sound Design of Rotational Order Sound Considered Perception of Pure Tone Components

☆花井奏太(中央大院)，△河野篤史，△寺内昇平，△西晃佳(コマツ)，田辺総一郎，戸井武司(中央大)

◆心理音響メトリクスのうち，トナリティとラウドネスに着目し，純音と広帯域音からなる音源に対する純音認知度合いを定量的に評価し，回転次数音において純音が不快となる閾値を明らかにする。

◆臨界帯域の中心周波数ごとに行った純音認知評価より，周波数によってトナリティとラウドネスの純音認知度合いへの寄与率が異なり，ラウドネス変化によって純音の聴こえ方が異なることがわかった。

◆Fig. 1 に示すように，純音が少し気になるが許容とされるトナリティ許容閾値をラウドネス毎に求めた。

◆回転次数音にブラウンノイズを付加し，Fig. 2 に示すようにトナリティを低下させることで，純音認知度合いが低下し，快適感が向上した。

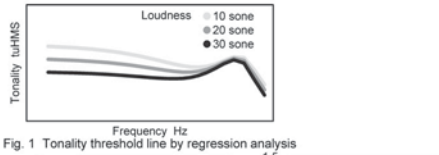
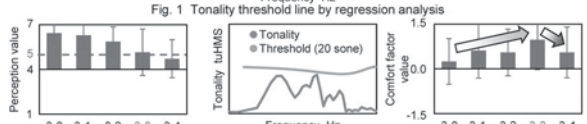


Fig. 1 Tonality threshold line by regression analysis



(a) Pure tone components perception value (b) Tonality analysis result of sound 3-3 (c) Comfort factor value

Fig. 2 Validation of Tonality tolerance threshold line

1-6-3

1-6-3　複数回転次数音の不協和度に着目した快音設計

Comfortable Sound Design of Multiple Rotational Order Sounds Focused on Dissonance

☆花井奏太，△麻生海(中央大院)，田辺総一郎，戸井武司(中央大)

◆複数回転次数音の組み合わせにより音質は顕著に変化する。そこで回転次数の１次同士の不協和度に着目し，周波数比を適切に設定することによる快音化を目的とする。

◆不協和度とトナリティに着目した二純音での純音認知評価より，周波数によって不協和度とトナリティの純音認知度合いへの寄与率が異なり，不協和度が 0.24 付近のときに純音認知度合いが低下することがわかった。

◆Fig. 1 に示すように，純音が少し気になるが許容されるトナリティ許容閾値を不協和度毎に求めた。低周波数では，単一純音よりもトナリティ許容閾値が上昇した。

◆複数回転次数音で不協和度とトナリティの印象評価を行い，Fig. 2 に示すように純音認知度合いが低かった不協和度が 0.24 付近のときに最も快適感が向上した。

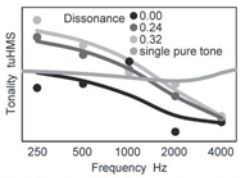


Fig. 1 Scatter plot and regression analysis line of Tonality tolerance threshold

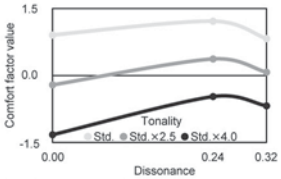


Fig. 2 Predicted comfort factor value of Tonality and Dissonance by regression analysis

1-6-4

1-6-4　音響マルチパラメータを用いた機械学習による自動車走行音の特徴量抽出 －第１報 ニューラルネットワークを用いた回転次数成分と暗騒音分離による特徴量抽出の精度向上－

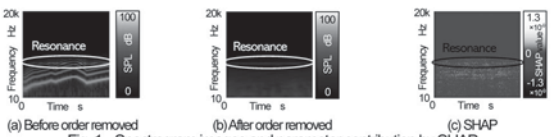
Feature Extraction of Car Driving Sound Using Machine Learning with Acoustic Multi-Parameters -First Report : Improvement of Feature Extraction Accuracy by Separation of Rotation Order Components and Background Noise Using Neural Networks-

☆大島遥汰（中央大院），田辺総一郎，戸井武司（中央大）

◆画像分類 AI を用いて走行音の特徴量を抽出するにあたり，回転次数成分以外の特徴量を解釈しやすくするため，機械学習を用いてスペクトログラム画像から回転次数成分を除去し，画像分類 AI が特徴量を抽出しやすい画像生成手法を検討する。そして，得られた画像を含む音響マルチパラメータ機械学習モデルの性能検証を実施する。

◆U-Net を用いて回転次数成分を除去することで，Fig. 1 (a)から，Fig. 1 (b)に示すように，暗騒音と共振成分の特徴量を抽出することができた。

◆特定のパラメータが特徴的な音を用いて，上記画像を含む音響マルチパラメータ機械学習モデルの性能検証を行った。Order-removed spectrogram において，Fig. 1 (c)に示すように，共振成分と暗騒音の貢献度が高いことから，音の特徴量を正確に抽出できたと考えられる。



(a) Before order removed (b) After order removed (c) SHAP

Fig. 1 Spectrogram images and parameter contribution by SHAP

1-6-5

1-6-5　音響マルチパラメータを用いた機械学習による自動車走行音の特徴量抽出 －第２報 ニューラルネットワークを用いた車種間における共通特徴量抽出－

Feature Extraction of Car Driving Sound Using Machine Learning with Acoustic Multi-Parameters -Second Report : Extraction of Common Features among Car Models Using Neural Networks-

☆大島遥汰（中央大院），田辺総一郎，戸井武司（中央大）

◆第１報にて性能検証した機械学習モデルを用いて，自動車走行音の分類を行い，主成分分析による特徴量の次元削減に基づくサウンドマップの作成および車種間において相違する特徴量を抽出する。

◆本研究では，ディーゼルエンジン車の加速音を，車格ごと（L_truck, M_truck, H_truck, Sedan）に分類する。その後，機械学習の分類貢献度可視化手法である SHAP により得られた自動車走行音の特徴量に対し，主成分分析を行うことで，第１主成分と第２主成分による特徴量マップとして各車格の音響マルチパラメータに特徴量を明示する。

◆Fig. 1 に示す SHAP 主成分分析による特徴量マップより，L_truck と H_truck には第１主成分得点に大きな差があることが分かった。また，Fig. 2 に示す２車格の Order-removed spectrogram の比較により，第１主成分の主成分負荷量の差が，高周波数の広帯域音の音圧レベルによって生じていると認められた。

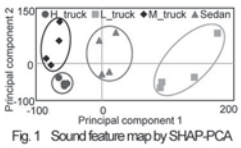
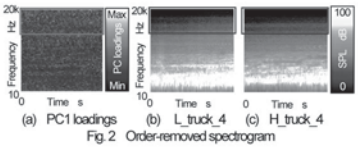


Fig. 1 Sound feature map by SHAP-PCA



(a) PC1 loadings (b) L_truck_4 (c) H_truck_4

Fig. 2 Order-removed spectrogram

1-6-6

講演取消

1-6-7

1-6-7　音量の異なる聴覚のデュアルタスクにおける脳波を用いた認知負荷の評価

Evaluation of Cognitive Load Using EEG in Auditory Dual-Task with Different Sound Volume

◎宮本 佳奈，矢澤 櫻子，伊藤 弘章，野口 賢一，鎌本 優 (NTT)

◆日常生活では，特定の音に集中する主タスクと周囲の他の音にも注意を払う副タスクを同時に行うことが求められる場合がある

◆主タスクと副タスクを同時に実施し，タスクの成績と脳波解析から，両タスクを理解しやすい音量バランスを調査した

- ➢ 主タスク：音声 N-back タスク
  - ● N 個前と同じ数字が聞こえたら反応
  - ● 1-back タスクと 2-back タスクを用意
- ➢ 副タスク：単語聞き取りタスク
  - ● N-back タスク中に 4 モーラ音の単語の聞き取り
  - ● 提示音量は N-back タスクに対して-6dB または 6dB で設定

◆主タスクの認知負荷が高い場合に副タスクの音量が大きいと，主タスクの成績低下や脳波のα波成分の低下が確認された

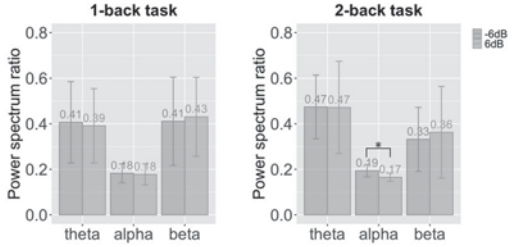


Fig.1 Power spectrum ration of EEG during an auditory dual-task (* p< .05)

1-6-8

1-6-8　「音空間デザイン」の提案 –本校図書館での恒常的な展開のための検討–

Proposal of “Sound Space Design”

- Consideration for permanent development in our school library -

☆宮本岬，石川あゆみ（岐阜高専）

◆今後、本校図書館における音空間デザインを恒常的に展開するための知見を得るために、本校図書館における音空間デザインの実践および音環境に関する評価の確認を行った。

◆音空間デザイン実践時と非実践時に図書館の音環境に対する印象を問うアンケート調査（SD 法 7 段階尺度および「今日の音環境で良いと思った点」と「今日の音環境で悪いと思った点」の自由記述）を行った。

◆Fig.2 に図書館の SD 法による音環境の評価結果（全回答者の評価値の平均値）を示す。

◆実践時の自由記述では、「緊張感が和らいだ」などの回答が得られた。非実践時の自由記述では、「一つ一つの行動に慎重になるため居心地が悪かった」などの回答が得られた。

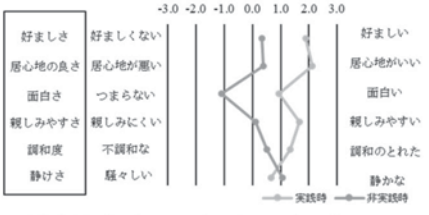


Fig.2 Evaluation results of sound environment

1-6-9

1-6-9　“これから自動車”のための サイン音デザイン

Sound Design for Future Mobilities

○山内勝也（九州大・芸工），中 貴一（関西学院大），田上宣昭（パイオニア）

自動運転が日常となり，

車の使い方が現在とは全く違ったものになった未来，車に求められる価値や意味も大きく変容しているでしょう。

そんな車で使うサイン音は，どんなデザインでしょうか？

1-6-10

1-6-10　特定条件下の完全自動運転に向けた人の安全と安心を調和する音のデザイン

Sound design for harmonizing human safety and security for fully automated driving under specific conditions

○有光哲彦（フィート）

◆自動運転のレベルが高まると共に，人に対する音の安全支援の上に機能的役割も変わる。聴覚や視覚の他，障害の有無や年齢によらないユニバーサル・モビリティを実現するために，自律神経の回復などの機能的な空間としての役割を有した車室内外共に安心できる “ゆとり” 音環境の創出 (Fig.1) が期待される。

◆身近な自動運転事例として，自動バレーパーキング，ゴルフ用カート，およびグリーンスローモビリティ等がある。音響空間は，限定空間と混在空間に分けられる。人の感性を考慮した異常音等のモニタリング，創造的コミュニケーション環境，人の健康をつくる音環境，マルチモダリティの構築等が期待される。

◆運行データや車内データがクラウド上に収集され，包括的な大規模演算によりデータの活用とシステムの更新が行われる。受動的な音のデザインは，人の感性と先進システムとの協調により能動的かつ双方向性の安全・安心を調和するユニバーサル・モビリティのためのソフトウエア中心のデザインへ移行する。

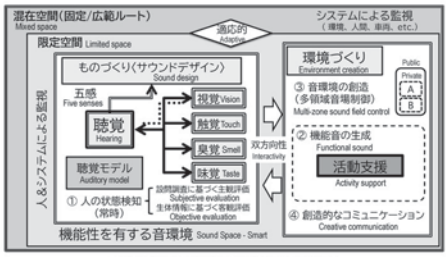


Fig. 1 The next concept of functional sound space

1-6-11

1-6-11　［招待講演］自動運転車向け車外報知音のデザインプロセス

Design process for Level 4 automated vehicle audio signals

中西宣人(フェリス女学院大学)，△小泉静香(スズキ株式会社)，
△林佳奈(スズキ株式会社)，△林秋妤(スズキ株式会社)，
△木田正吾(スズキ株式会社)，川上央(日本大学芸術学部)

◆レベル４自動運転車にはドライバが乗車しないため、車両の動作や状態を歩行者や自転車などの低速移動者に適切に情報を伝達する車外向けHMI（Human Machine Interface）が必要になる。情報伝達の手法としてはディスプレイ等の視覚表示を利用することが考えられるが、視認していなくとも情報伝達が可能な車外向け報知音が必要と考える。

◆スズキ株式会社横浜研究所およびフェリス女学院大学の共同研究として自動運転車の車外向け報知音デザインを実施し、①報知音のデザイン要件定義、②リサーチ、③モデルの作成とメソッド化、④プロトタイプ、⑤評価という流れで実施した。結果として開発者、ユーザ両者から一定の評価が得られる報知音が作成された。

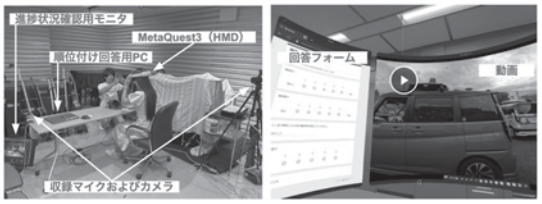


Fig.1：実験風景およびHMD内映像

1-6-12

1-6-12　自動運転時の権限移譲を伝えるサウンド UI の評価

Evaluation of sound UI to inform takeover requests during automated driving

○浅川香，粟野智治（三菱電機），山内勝也（九州大・芸工）

◆システムが運転操作を人間のドライバへ引き継ぐ権限移譲（Takeover Request: TOR）シーンに適切なサウンド UI を明らかにするため実施した一連の実験を紹介する。

◆実験１：音刺激 30 種を聴取後，TOR 場面に使用する音としての適切性などについて 7 段階で評価した。適切性が高いと評価された音は，緊急感・不快感・覚醒感が中程度の音が多かった。

◆実験２：TOR シーンを模擬した課題中に音刺激 7 種を呈示し、音が行動や印象に与える影響を評価した。その結果、緊急感の高い音の適切性は実験１よりも低下する傾向がみられた（Fig.1）。

◆実験３：実験１の印象評価値を目的変数、音響特徴量を説明変数とした重回帰モデルの検証実験を行い、推定値との有意な正の相関を確認した。

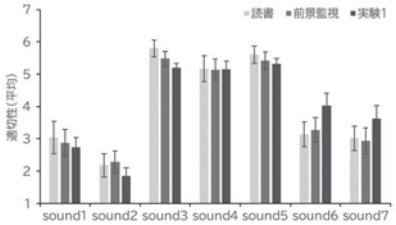


Fig.1：Mean subjective rating score for adequacy in each condition (light gray: reading, medium gray: supervising, and dark gray: Experiment 1) by sound stimuli. Error bars indicate SEM.

1-6-13

1-6-13　車室内音開発の変遷　聴覚から多感覚へ

The transition of in-vehicle sound development: From Auditory to Multi-modality

○長江新平，榎本俊夫（日産自動車株式会社）

車室内には，走行に伴う騒音，ウインカーの作動音，ナビゲーションシステムの音声案内など多種多様な音が存在しており，これらは運転者や同乗者の快適性や安全性に直接的な影響を及ぼす重要な要素である．近年，パワートレイン技術の進化，とりわけ電動化の発展に伴い走行音の音量や音質が大きく変化し，従来の内燃機関車両とは異なる特性を持つようになってきている．また自動運転技術の進化により，車両と乗員の関係性はこれまで以上に多様化し，従来の「運転者」としての役割から解放されることで車室内の乗員の行動や心理的なニーズにも大きな変化が見られるようになってきた．このような状況において，車室内の音響環境を再評価し，新たな考え方に基づいて設計する必要性が高まっている．

その一方で，車両が増々高機能化，複雑化する中，音を含む様々な性能要件を同時に満たす設計解を見つけ出すことは非常に困難な課題となりつつある．本稿では，この課題に対処するための一例として，ヒトの認知処理特性を利用した車両開発の事例を取り上げる．この事例では，実際にはエンジン騒音レベルが増大しているにもかかわらず，人間側の特性を利用しそれをうるさく感じさせないことで両立を図っている．さらに，その取り組みの中で開発した実車サウンドシミュレータについても紹介する．本シミュレータは上記のような “うるさく感じにくいシーン” を数値化する上で有用な他，広く音響設計の現場でヒトの認知処理特性を理解し，それをモデル化する上でも活用が期待できる．

1-6-14

1-6-14　感性サウンドマップに基づく自動車走行音の特徴分析手法の開発

Development of a method for analyzing vehicle noise characteristics based on a sensitivity sound map

☆中里崚人（中央大院），加曽利拓真，田辺総一郎，戸井武司（中央大）

◆加速音の印象と心理音響メトリクスとの関係を検討し，感性サウンドマップを作成することで，人間の感性の可視化を行う。

◆模擬的に作成した自動車走行音に対する印象評価を行い，感性サウンドマップを構築し，特徴の可視化を行う。

◆自動車走行音において，ICE と EV では得られる印象が異なり，ICE は「躍動感」，EV は「静穏感」に属し，また，高周波の回転次数成分を追加することで「未来感」が向上することが明らかになった。

◆サポートベクターマシン+Lasso 回帰を使用することで，Fig. 1 に示すように時間および周波数におけるスペクトルとトナリティに特徴的な領域を把握した。

◆本手法より「躍動感」，「静穏感」，「未来感」は，それぞれ変動強度，ラウドネス，トナリティに寄与していることを示した。

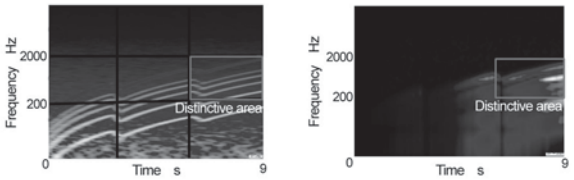


Fig. 1　Spectrogram and Tonality of ICE2

1-6-15

1-6-15　音響特徴量に基づく自動車の走行サウンドデザインの構築

Construction of automobile driving sound design based on acoustic features

○田辺総一郎（中央大），中里崚人（中央大院），加曽利拓真，戸井武司（中央大）

◆自動車の走行音を対象とし，感性サウンドマップや音響特徴量に基づく走行サウンドデザインの構築を目指す。

◆感性サウンドマップをもとに未来感と静穏感を持つEV2 を対象とし，目標値を設定した。

◆音響特徴量の分析により，未来感はトナリティ，静穏感はラウドネスの寄与が高いことが明らかとなり，実験計画法(L9)によるパラメータスタディを実施し，要因効果図をそれぞれ Fig. 1，Fig. 2 に示す。

◆Fig. 3 のように感性サウンドマップを作成したところ，適切な条件の No.4 の評価音で目標としていた未来感と静穏感を向上させることが確認できた。

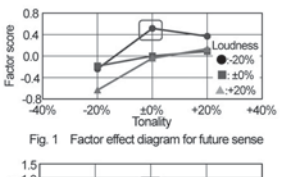


Fig. 1　Factor effect diagram for future sense

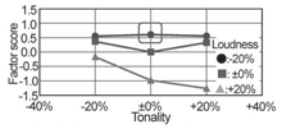


Fig. 2　Factor effect diagram for calm sense

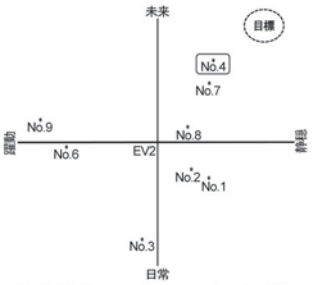


Fig. 3　Relative sensibility sound map based on EV2

1-6-16

1-6-16　主観および客観評価に基づく快適感と覚醒感を両立する覚醒サウンドの創生

Creation of an arousal sound that balances valence and arousal based on subjective and objective evaluations

☆青柳洸希（中央大院），△金堂雅彦，山口雅夫，戸井武司（中央大）

◆ 自動運転中は覚醒度が低下しやすく，漫然運転のリスクが懸念される。本研究では，快適感と覚醒感を両立する覚醒サウンドを提示することで，漫然運転のリスクを低減し，運転精度の向上を目指す。

◆ 実験手順は，まず快適かつ覚醒となる付加音を 125 Hz，250 Hz，500 Hz の組み合わせにより作成し，主観評価を実施する。次に，主観評価で最高得点となった覚醒サウンドを，運転開始 600 s 後に断続的に 5 回提示し，その提示前後の脳波と車線逸脱量の変化率を比較することで，覚醒効果を明らかにする。

◆ 実験結果は，覚醒サウンドを提示することで，主観および客観評価において覚醒効果がみられ，車線逸脱量は減少した。

◆ 適切な音圧と純音の周波数を組み合わせることで，効果的な覚醒サウンドを見出し，運転精度の向上が明らかになった。

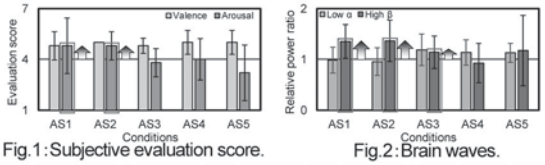


Fig.1：Subjective evaluation score.

Fig.2：Brain waves.

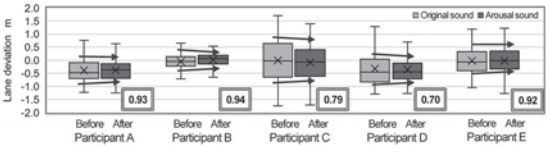


Fig.3：Changes in lane deviation with arousal sounds.

1-6-17

1-6-17　自動運転下での車内音デザインとその評価

Vehicle Interior Sound Design and Evaluation under Autonomous Driving

○石光俊介（広島市立大）

◆これまでエンジン音を中心に車内音デザインの検討を行ってきた。

◆エンジン音は加速度合いを知るフィードバック情報でもあったが，完全自動運転では搭乗者にとっては無用の情報？

◆加速走行騒音規制（R51-03）フェーズ３を受けて EV 化が進む昨今では加速時には人工音が生成される。

◆これまでの検討の中で完全自動運転でも再び採用できるトピックの洗い直しとして，本セッションを利用して整理整頓させていただく。

- エンジン音デザインは加速を知らせる情報という意味で車内生成音として残存。
- オーディオ集中するのは全体の 5%ぐらい（Automotive audio, AES, 2017）。
- 自動運転下での最も重要なコンテンツは“会話”。
- 注意が不要なときは順応*に任せ，注意が必要なときには報知音または警報音で注意をひく。

◆そこで，以下の２つのサウンドデザインとその評価を取り上げて紹介。

- 不快レベルからデザインする車内警報音
- 車内会話了解度からデザインする吸音材

◆今回は糠床おこしであったが，今後これらをベースに更なる発展検討を行う予定。

*亀山他，音論集（春），1-6-19，2025.

1-6-18

1-6-18　車内搭乗者条件の変化による空間の音伝達関数への影響検討

Investigation on the effect of changes in passenger conditions on the acoustic transfer function in the vehicle cabin

○曺 浣豪，△Jiho Chang（KRISS），△ Sung-Hwan Shin（Kookmin University）

◆本研究では搭乗者の配置と姿勢が車室内の音伝達関数に及ぼす影響について定量的な測定基づいて検討を行った。

◆測定結果から姿勢による両耳応答への影響は比較的に大きく，搭乗者数による変化は限定的である傾向が観察された。これは伝達関数自体の変化より位置による差の方が大きいということを示す。

◆本研究での結果は一つの車両に限定されたものではあるが，車内空間の類似性に基づいて，他の車両でも似た傾向があると考えられる。

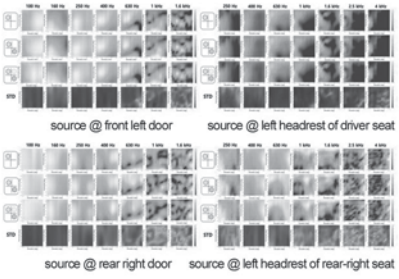


Fig.1：Effect of passenger layout on the acoustic response around the driver head

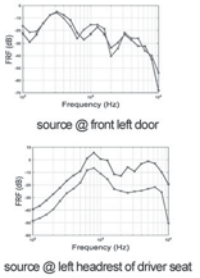


Fig.2：Effect of posture on the binaural response between the driver seat and source positions

1-6-19

1-6-19　自動運転環境を考慮した走行音に対する注意機能の評価

Evaluation of attentional mechanisms in response to vehicle sounds considering autonomous driving environments

☆亀山勇希，村上寛名，石光俊介（広島市立大院）

◆自動運転の普及に伴い，自動車車室内の音環境に求められる価値も変化していると考えられる．その一つのアプローチとして，注意を向きにくくして順応の崩れを抑える音環境の提供が有効である．

◆本検討では，音の大きさの変動に着目して，脳波の事象関連電位(ERP)を用いて注意の向きやすい走行音の特性を明らかにすることを目的とした．それより，注意の向きやすい音の特性を把握することで順応を考慮したサウンドデザインに資する知見を提供できると考える．

◆その結果，音の大きさの変動により生じた ERP 反応の傾向は刺激特性のみでなく，順応からの逸脱による影響が含まれていることが示唆され，その影響が大きいほど注意が向いていることが確認できた．

◆さらに，認知処理の複雑化によって情報処理過程にかかる時間が延長され，自動運転環境下での外部騒音の制御の必要性が示唆された．

◆走行音の低周波成分より高周波成分(1 kHz 付近)の方がより注意機能が働きやすいことが確認でき、感情価との関連性について考察した．

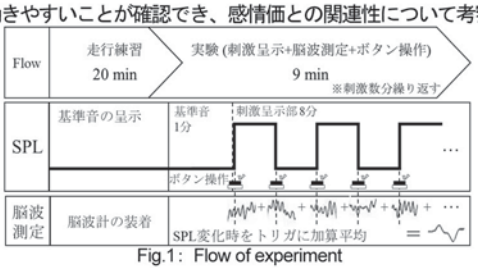


Fig.1：Flow of experiment

1-6-20

1-6-20　生体情報に基づくEV 走行音と運転動作による疲労感の評価手法

A method for evaluating fatigue caused by EV driving sounds and driving operation based on physiological information

☆加曽利拓真（中央大），中里崚人（中央大院），山口雅夫，戸井武司（中央大）

◆従来の疲労感の評価は，主観や心電による客観評価などが行われていたが，主観評価は疲労を自覚しない場合があり，一方，心電は呼吸等の自律神経活動の影響を受けやすい欠点がある。そこで本研究では，脳波を用いた客観的な評価手法を検討した。

◆ドライビングシミュレータを用いた運転タスク中に，3 種のカラーノイズを提示し，心電を用いた疲労感の評価結果と脳波との相関を分析した。そして，脳波の High α波は疲労感との相関が高いことを明らかにした。また，ブラウンノイズが最も疲労し難いことがわかった。

◆運転タスク中に，EV 走行音を模擬した評価音を提示し，心電を用いた疲労感の評価結果と脳波との相関を分析した。そして，脳波の High α波は疲労感との相関が高く，脳波で疲労感を評価できる可能性があることを明らかにした。また，EV 走行音を模擬した評価音の中では，周波数勾配が大きいほど疲労し難いことがわかった。

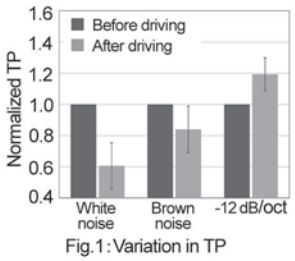


Fig.1：Variation in TP

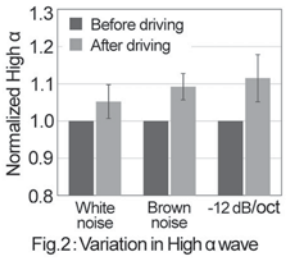


Fig.2：Variation in High α wave

1-6-21

1-6-21　年齢と眠気度が覚醒低下防止の警報音の効果に及ぼす影響の検討

The Impact of Age and Sleepiness Levels on the Effectiveness of Warning Sounds in Maintaining Arousal

☆星野慧（鉄道総研・早大理工），△鈴木綾子（鉄道総研），及川靖広（早大理工）

◆鉄道走行中の環境を模擬して走行音あり・単調課題を行っている状況で、警報音を提示して音による覚醒効果を検討した。

◆２時間の睡眠制限のうえ、実験課題中の単調作業中の眠気度が比較的強くなったタイミングで警報音を提示した。

◆年齢が及ぼす効果の差に着目し、20 代～30 代 12 名と 50 代～60 代 12 名を比較した。

◆覚醒効果を反応時間の短縮秒数で捉えた場合、50 代～60 代の短縮秒数が 20 代～30 代に比べて有意に長くなるケースが確認された。

◆事後アンケートによる、音の提示時の眠気度と覚醒効果については、20 代～30 代の年齢の方で、眠気度が高い時に覚醒効果が小さい割合が大きかった。

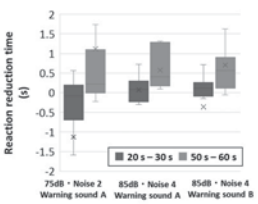


Fig.1：Reaction reduction time comparing between young group and elderly group

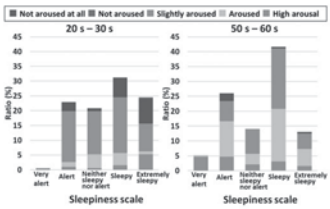


Fig.2：Relationship between sleepiness level and arousal function of warning sounds

1-8-1

1-8-1　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(3)－

Study on the noncontact acoustic inspection method for non-destructive concrete inspection

－Efficiency improvement of outer wall inspection using sound source mounted type UAV (3)－

○上地 樹, 杉本 恒美, 杉本 和子, 中川 裕 (桐蔭横浜大院)

◆ 現在我々は、音源搭載型 UAV (Unmanned aerial vehicle)を用いた音響加振およびスキャニング振動計(SLDV)を用いた，非接触による非破壊検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、外壁タイルが施工された実構造物を対象に、フラットスピーカ (FPS1030M3F1R, FPS Inc.)を搭載した UAV (MATRICE 600 pro, DJI Co., Ltd.)および SLDV (PSV Qtec, Polytec GmbH)を用いて検証実験を実施した。

◆ 高さ約 8.2 m の位置で UAV をホバリングさせ、12.1 m 程度離れた位置から SLDV により計測した。

◆ 音響加振には、700-4100 Hz のマルチトーンバースト波形を使用。

◆ 実験の結果、欠陥部と推測される応答を捉えたと考えられる。

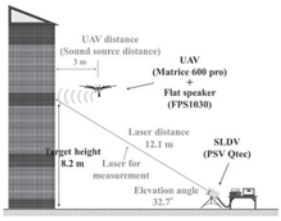


Fig. 1 Experimental setup (Side view)

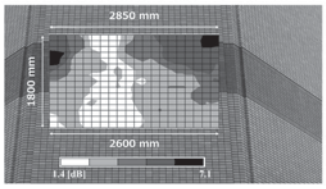


Fig. 2 Vibration energy ratio distribution (700-4100 Hz)

1-8-2

1-8-2　超音波探触子の高周波駆動に関する機械的と電気的検討

Mechanical and electrical considerations for high-frequency drive of ultrasonic probes

○田中雄介,　△小倉幸夫(ジャパンプローブ)

◆超音波振動子に固い物質が接触していると高周波になり、柔らかい物質が接触していると低周波になる。

◆超音波振動子の静電容量により振動子へ印加される電圧に遅れ時間が発生し、静電容量が大きいと低周波になる。

◆超音波振動子に直列にコンデンサを接続すると遅れ時間が下がる。

◆超音波振動子前面にガラスの接触や低誘電率の振動子選択、コンデンサの直列接続により高周波駆動が可能になる。

◆コンデンサの直列接続時は分圧により印加電圧が低下する。

Table 1 Materials in contact with the piezoelectric transducer and response frequency.

Materials	Response frequency [MHz]
Water	20.0
Polystyrene	19.2
Glass	23.8
Silicone rubber	6.8

Table 2 Materials in contact with the piezoelectric transducer and response frequency.

Connected capacitors [pF]	Response frequency [MHz]
Nothing	6.3
1500	7.6
1000	8.5
680	10.0
470	10.0
220	14.7
100	17.2
47	21.7
30	26.3
22	26.3
10	26.3
5	29.4

1-8-3

1-8-3　非線形空中超音波フェーズドアレイ波源走査法によるモルタル表面を伝搬する波動のパルス圧縮

Pulse compression of waves propagating on mortar surface using nonlinear airborne ultrasound phased array source scanning method

◎清水 鏡介(愛媛大院)，神谷 大樹 (日大院理工),
大隅 歩，伊藤 洋一(日大理工)

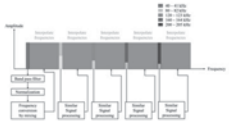
◆　非線形チャープ信号とミキシングによる周波数補間を用いたパルス圧縮方法について研究を行っている.

◆　数値解析を用いて取得したモルタルの表面波伝搬像に対して提案手法を適用し，その有効性について検証を行った.

◆　Figure 1 に提案手法の概要図を，Fig. 2 に解析結果を示す.

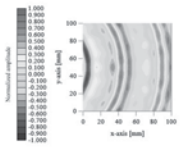
◆　提案手法によって極めて短パルスな波動伝搬が画像化されていることが確認でき，提案手法の有効性が確認された.

(a). Phase compensation.

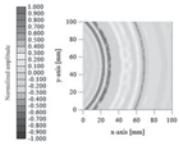


(b). Interpolate discrete frequency components.

Fig. 1 Proposed method.



(a). Before applying pulse compression.



(b). After applying pulse compression.

Fig. 2 Analysis result.

1-8-4

1-8-4　空中超音波の非線形性を利用した金属板減肉部の位相画像形成

Forming phase images of metal plate thinning area using high-intensity airborne ultrasound with nonlinearity

☆石川周男(日大･理工),神谷大樹(日大院･理工),清水鏡介(愛媛大院),
伊藤洋一,大隅歩(日大･理工)

◆空中超音波の非線形性を利用した高調波イメージング(Harmonic Imaging:HI)を提案し，研究を行っている.

◆各高調波成分の位相差に着目し，金属薄板内減肉欠陥の HI を行った.

◆その結果，高調波の位相差画像より減肉部の精細な映像を得ることが確認できた.

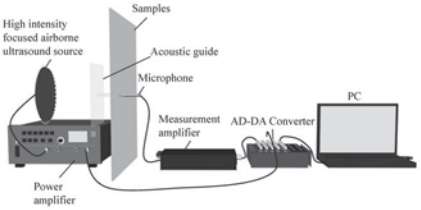
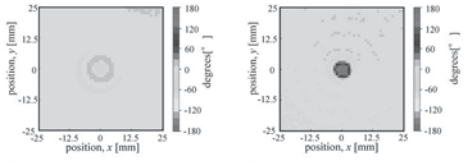


Fig.1 Experiment devices.



(a) The fundamental wave.　(b) The 3rd harmonic wave.

Fig.2 Phase image.

1-8-5

1-8-5　熱弾性効果増強基板を用いた光学式固有音響インピーダンス計測法の開発

Development of a substrate enhancing the thermoelastic effect for optical specific acoustic impedance measurement

○田村和輝，△大川晋平（浜松医大 光医学総合研究所）

◆レーザー超音波法を用いて光学的にサンプルの固有音響インピーダンスを計測する手法の開発に取り組んでいる．

◆顕微鏡視野内にナノ秒パルスレーザー光をドーナツ型に空間ホログラム化した光を照射して加振光とし，試料を支持する基板の試料に接さない面に結像させた．同じ面内で加振光の中心部にレーザードップラ計の計測光を結像させた．パルスレーザー照射の時間から 2 μs 間の基板表面の振動波形を計測した．

◆試料として基板上に空気，水，飽和食塩水を置いた場合の試料・基板境界由来の縦波の反射信号を収集し，振幅を計測した．

◆ポリスチレン基板に対する反射率と同じ大小関係の反射波振幅が得られた．

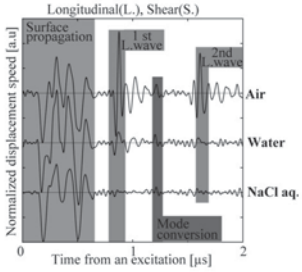


Fig.1: The time waveform of the out-of-plane displacement velocity of air, water, and saturated saline solution are used as samples. The amplitude of the longitudinal wave reflection was air: 1.236, water: 0.216 (17% compared to air), and saturated saline solution: 0.049 (4.0% compared to air)

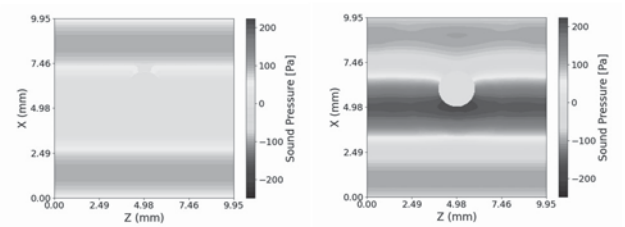
1-8-6

1-8-6　Python を用いた複雑形状物体の音響浮揚音場解析

Acoustic levitation sound field analysis of complex-shaped objects using Python.

☆黒川陸，伏見龍樹（筑波大）

本研究では、Python を用いて複雑形状物体の音響浮揚音場を解析するシミュレーション手法を開発する。音波の非線形伝播過程を FDTD 法で計算し、音場中の複数の非線形音波の重ね合わせや、物体からの反射波の伝播を考慮した精密な解析を可能にした(図 1)。これにより、音響浮揚現象の理解を深化させ、新たな応用分野の開拓に寄与することを目指す。さらに、オープンソースとして公開することで、ユーザーが解析ニーズに合わせて柔軟に修正・拡張できるツールを提供する。今後は、提案した音響放射力の計算手法の実装と検証を進め、既存の商用ソフトウェアや理論解析との比較を通じて妥当性を確立し、計算精度の向上を図る予定である。



(a) Before interference of sound waves and objects.　(b) After interference of sound waves and objects.

Fig.1 : Sound pressure heat map based on nonlinear sound wave propagation calculations.

1-8-7

1-8-7　超音波照射された小径穴付き円筒と空隙を介して対抗する平面に働く力の検討

Examination of forces action on a plane opposing a cylinder with a small-diameter hole irradiated by ultrasonic waves through a gap

☆藤岡夕大[1]，王伊萌[1]，田村英樹[2]，青柳学[1]（[1]室蘭工大・院，[2]東北工大）

◆円筒底部に僅かな窪みを設けた円筒を用いて，円筒下部に配置したアクリル板が円筒に引き付けられるかどうかを有限要素解析および実験により調査した。

◆Fig. 1 は円筒とアクリル板間のギャップhを変化させたときのアクリル板に作用する力の解析結果である。音響放射力F_{rad0}はhによらず負の値を取り円筒へ引き付ける力として作用するが，音響流による力F_{jet}はhによらず正の値を取り，円筒から引き離す力として作用する。また，合力F_Tはhが小さいときは引き離す力として，hが大きいときは引き付ける力として作用することが判明した。

◆Fig. 2 はアクリル板を円筒の下から近づけた際に，円筒にアクリル板が引き付けられた様子を撮影したものである。空隙hが存在していることから，円筒に接触せずにアクリル板が浮揚していることがわかる。

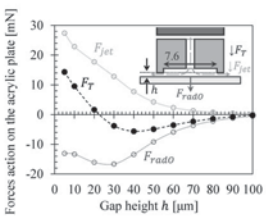


Fig. 1 Forces vs. gap height h.

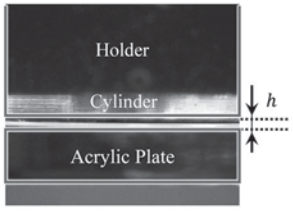


Fig. 2 Observations of attraction and levitation.

1-8-8

1-8-8　3音源を用いた合成音場中内で剛性球に作用する音響放射力の評価

Evaluation of acoustic radiation force acting on rigid spheres in wave fields formed by three sound sources

☆丸目勝斗（愛工大），畑中信一（宇都宮大），
鎌倉友男（電通大），小塚晃透（愛工大）

◆超音波振動子 36 個よりなる集束音源を製作し、３音源を逆正三角形の頂点に配置して形成した合成音場中（Fig. 1）で、超音波浮揚の実験を行った。

◆合成音場中に 5 mm の鉄球を糸で吊して投入し、その荷重の変化を測定した。Fig. 2 は、上方２音源と下方音源の周波数に 0.05 Hz の差を与えて、位相を 20 s に１周期の割合で連続的に変化した結果である。60 s（３周期分）の変化を示している。

◆音圧の節が上方にある場合は軽くなり、下方にある場合は重くなるが、音圧分布がハニカム状になるため、上下の音圧の節のみならず、近隣の音圧の節の影響を受けてこのような変化を示すと考えられる。

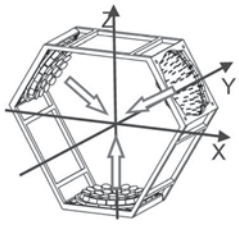


Fig. 1. Experimental apparatus.

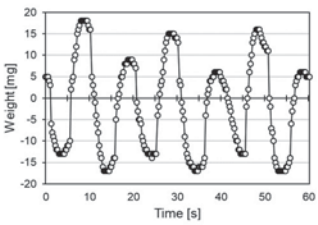


Fig. 2. Weight changes of iron ball as function of the phase difference.

1-8-9

1-8-9　音響定在波中に浮揚する液滴の表面張力と共振特性の関係

Relationship between surface tension and resonance characteristics of droplets levitated in an acoustic standing wave

☆平山喬也，○小山大介（同志社大）

◆**超音波浮揚法**では物体を非接触で捕捉することが可能であり，**物体の搬送や物性計測**での応用が期待されている．

◆**円形振動板に４つの BLT を接続し**，同相連続正弦波信号を印加することで振動板・反射板間に**音響定在波**を形成し (Fig. 1(a))，音圧節部に液滴を浮揚させた．

◆BLT に **AM 変調信号**を印加することで浮揚液滴に振動を誘起し，その振動特性を評価した．

◆**表面張力を変化**させることにより，同体積であるにもかかわらず浮揚液滴の**共振周波数が異なる**ことがわかり (Fig. 1(b))，物性計測への応用が示唆された．

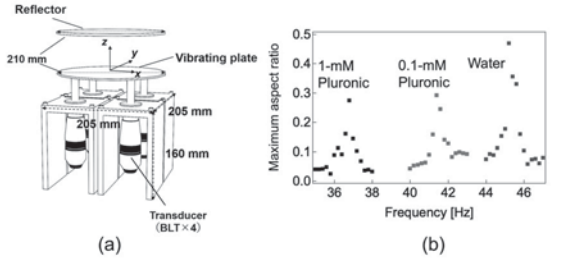


Fig.1 : (a) Configuration of the ultrasound vibrator and the reflector. An acoustic standing wave can be generated between them and (b) Relationship between the signal frequency and the aspect ratio for the water, 0.1-mM Pluronic, and 1-mM Pluronic droplets.

1-8-10

1-8-10　超音波による液面変形を用いた動的コースティックパターンの生成

Dynamic Caustics by Ultrasonically Modulated Liquid Surface.

◎永倉昂暉，伏見龍樹，筒井彩華，落合陽一(筑波大学)

超音波を用いた新しい動的なコースティックパターンの生成手法を提案しています。従来の方法では困難であった、時間的に変化するコースティックパターンを、フェーズドアレイ超音波トランスデューサ (PAT) を用いて液体の表面形状を非接触かつ動的に制御することで実現しました。

図１は、生成されたコースティックアニメーションのフレームを示しています。各フレームは、９フレーム分の音圧を重ね合わせ、さらに実験で得られたコースティクス画像をフィードバックとして利用するデジタルツイン技術によって処理されています。これにより、シミュレーションだけでは考慮しきれない物理現象を反映させ、コースティクスの精度と品質を向上させています。

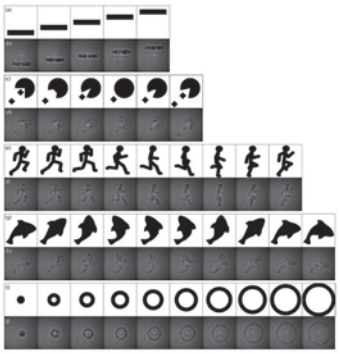


Fig.1：Comparison of target images used for animation and the generated caustics, presented frame by frame. . Each frame was generated by superimposing the sound pressure of 9 frames and then processed using the Digital Twin.

1-8-11

1-8-11　位相シグネチャーを加えた場合における平衡位置の逆転現象

Equilibrium Position Reversal: Case Study with Phase Signatures

◎頃安祐輔(筑波大院)　伏見龍樹(筑波大)

◆空中超音波を用いた非接触操作では、音源に近い領域の平面内で音響放射力が**高圧領域から物体を押し出す**一方、遠方では力の向きが反転し、**高圧領域へ物体を引き寄せる**ように作用する。

◆本研究では、この逆転現象が単純な集束ビームに限らず、ツイントラップやボルテックストラップなどの位相シグネチャーを付与したビーム、さらにはダマングレーティングを用いた場合でも発生することを明らかにした。

◆ツイン・ボルテックストラップでは、音源からの距離が伸びるに従って平面内の捕捉剛性が減少し、やがて負に転じるため、**安定捕捉が可能な範囲が音源近傍に制約される。**

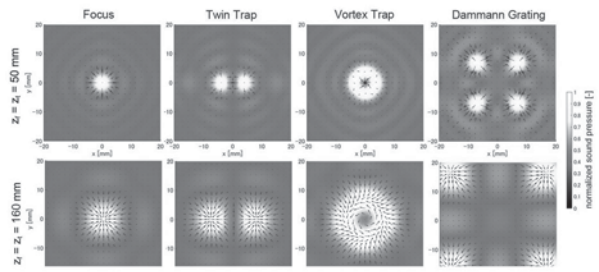


Fig.1 : Normalized sound pressure profiles and acoustic radiation forces (indicated by arrows) at distances of 50 mm (upper) and 160 mm (lower)

1-8-12

1-8-12　撥水面における液滴操作を用いた高重量物体の超音波搬送

Ultrasonic transportation of objects using droplet manipulation on a hydrophobic surface

◎伏見龍樹　頃安祐輔　△落合陽一（筑波大学）

本研究では、撥水加工を施した金属メッシュ上の液滴を搬送体として用い、高重量物体の超音波操作を試みた。OpenMPD 型フェーズドアレイ振動子と撥水メッシュを組み合わせ、液滴上に設置した軽量アイコンの移動を実現した。アイコンの直径および液滴の体積を変化させ、操作可能な条件を調査した結果、特定の条件下で液滴を利用した安定した搬送が可能であることを示した。本手法は、従来の音響浮揚では困難であった高重量物体の操作に応用可能性を持ち、さらなる発展が期待される。

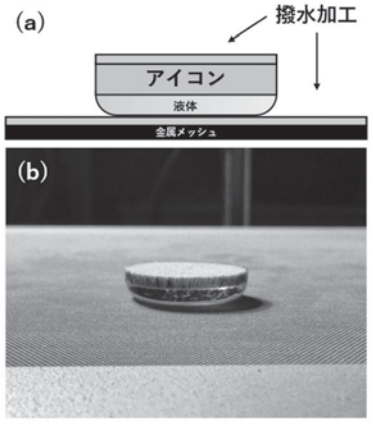


Fig.1: 水滴を用いたアイコン搬送の実現

1-8-13

1-8-13　液滴の超音波浮揚における移動速度の影響

Effect of movement speed on ultrasonic levitation of droplets

☆成田憲一，井上雄大(愛工大)，鎌倉友男(電通大)，畑中信一(宇都宮大)，小塚晃透(愛工大)

◆超音波振動子を上下に各 36 個配置した定在波音場装置を使って液滴の超音波浮揚の実験をおこなった（Fig.1）.

◆ファンクションジェネレータをプログラム（Fig.2）によって操作して位相差を作り，音場を連続的に変化させた.

◆定在波音場装置下側の周波数を下げて音場を上方へ移動させると浮揚物体は保持時間が 100ms では追従できなかった.

◆また定在波音場装置下側の周波数を上げて音場を下方へ移動させると浮揚物体は保持時間を短くしても追従した.

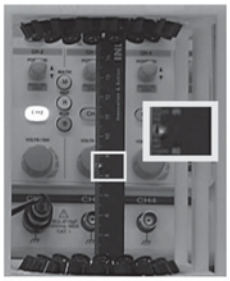


Fig.1：Fig.1 Captured droplet.

Fig.2 Operation screen of the function generator control program.

1-9-1

1-9-1　6ch 音場再現システムにおける音声明瞭度の再現性に関する検討

Study on the reproducibility of speech intelligibility in a 6-channel sound field reproduction system

☆陳科吉，佐久間哲哉（東大・工）

◆ 本研究は幾何音響解析に基づく 6ch 音場再現システムにおける音源分配方法を修正し、受音指向性の再現性を理論的に確認した。

◆ 再現音場における音声レベル、クラリティ C_{50} および STI の実測値と幾何音響解析からの解析値の対応を検討し、音声音量および音声明瞭度の再現性の向上を確認した。

◆ 以前の直接法では直接音の音量が低下し、高吸音条件では明瞭度指標の弁別域を超える低下が生じた。

◆ 直接音のみを分離再生する手法は必ずしも明瞭度指標の再現性向上につながらなかった。

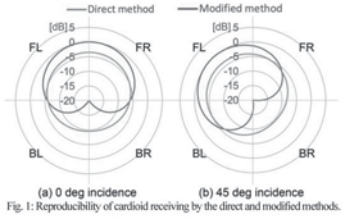


Fig. 1: Reproducibility of cardioid receiving by the direct and modified methods.

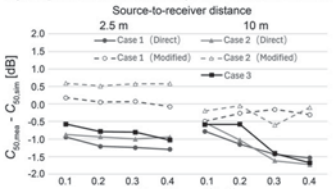


Fig. 2: Difference between measured and simulated C_{50}.

1-9-2

1-9-2　Sabine の残響理論における室内音響エネルギー収支を表す微分方程式の修正

Revision of the differential equation for the room acoustic energy balance in Sabine's reverberation theory

〇羽入敏樹(日大・短大)

◆室内の音響エネルギー収支を表す微分方程式から Sabine 理論を導出できる。本報では，室内音響エネルギー収支の微分方程式を修正し，そこから修正残響理論を導出できることを示す。

◆従来の微分方程式と Sabine 理論

$$V\frac{dE}{dt}=W-\frac{cES\bar{\alpha}}{4} \quad \Rightarrow \quad E=\frac{4W}{cS\bar{\alpha}}\exp\left(-\frac{cS\bar{\alpha}}{4V}t\right)$$

室内音場のエネルギー収支に基づく従来の微分方程式　　Sabine 理論における定常状態からの残響減衰

◆修正微分方程式と修正残響理論

$$\frac{dE}{dt}=\frac{W}{V}-\lambda E \qquad \frac{dE}{dt}=\frac{W}{V}-\frac{c}{\bar{\ell}_a}E$$

音響エネルギー密度に着目した修正微分方程式

※λは単位時間あたりの音響エネルギー密度の減衰率，$\bar{\ell}_a$は平均吸音自由行程

$$\bar{\ell}_a=\left(\frac{1}{\bar{\alpha}}-\frac{1}{e}\right)\bar{\ell}=\left(\frac{1}{\bar{\alpha}}-\frac{1}{e}\right)\frac{4V}{S}$$

修正残響理論における平均吸音自由行程

$$E=\frac{4W}{cS}\left(\frac{1}{\bar{\alpha}}-\frac{1}{e}\right)\exp\left[-\left(\frac{1}{\bar{\alpha}}-\frac{1}{e}\right)^{-1}\frac{cS}{4V}t\right]$$

修正残響理論における定常状態からの残響減衰

1-9-3

1-9-3　矩形室内の吸音と拡散の配置が残響時間に及ぼす影響の音線法による検討

Study on the effect of the arrangement of sound absorption and diffusion in a rectangular room on the reverberation time using the sound ray tracing method

〇羽入敏樹，鈴木諒一，星和磨(日大・短大)

◆吸音材と拡散体の配置が残響時間に及ぼす影響について定量的に評価するため「吸音効率」を定義した。そして音線法を用い，矩形室における吸音と拡散の配置の違いによる吸音効率の変化を検討した。

◆吸音効率 η の定義

$$\eta=\frac{A_{meas}}{A_{design}}$$

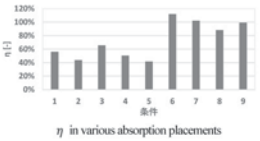
A_{design}は設計時の室の等価吸音面積，A_{meas}は実測もしくはシミュレーションによる減衰曲線から読み取った残響時間から逆算した等価吸音面積とする。逆算には拡散音場を前提とした残響時間の公式を用いる。

$$A_{meas}=\left[\left(\frac{KV}{T_{meas}}-4mV\right)^{-1}+(eS)^{-1}\right]^{-1}$$

◆吸音配置の影響

Table 1 Absorption placement conditions

	壁面積	設計時の等価吸音面積 A_design [m^2]								
壁面	S[m^2]	条件1	条件2	条件3	条件4	条件5	条件6	条件7	条件8	条件9
x+	18			15			10	5	2.5	7.9
x-	18			15				5	2.5	7.9
y+	30	30			15		10	5	4.2	4.7
y-	30				15			5	4.2	4.7
z+	60		30			15	10	5	8.3	2.4
z-	60					15		5	8.3	2.4
合計	216	30	30	30	30	30	30	30	30	30

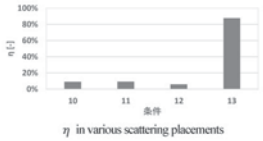


η in various absorption placements

◆拡散体配置の影響

Table 2 Scattering placement conditions

	壁面積	A_design	散乱面積 $\beta \sim \beta S$ [m^2]			
壁面	S[m^2]	[m^2]	条件10	条件11	条件12	条件13
x+	18		15			15
x-	18		15			
y+	30			15		15
y-	30			15		
z+	60	30			15	
z-	60				15	
合計	216	30	30	30	30	30



η in various scattering placements

1-9-4

1-9-4　小学校の普通教室における吸音材設置の効果に関する検討

Investigation on the effects of sound absorbing materials in an ordinary classroom at a primary school in Japan

〇Wang Yutan，上野　佳奈子（明治大学），
エバンズ　直子（TOA（株）/大阪大学），△清野　健（大阪大学）

◆小学校の普通教室（片廊下型）において、天井に吸音材を設置し（Fig.1）、室内音響性能の改善効果を調査した。

◆空室状態にて、空調・換気扇の稼働有無の２条件について、吸音材設置前後の残響時間および STIPA を測定した。その結果、吸音材による音声伝送性能の向上が確認された。

◆吸音材設置および撤去前後における、教室使用時の騒音レベルを測定し、給食および掃除時間中の変化に注目して分析した。また児童を対象に、アンケートによる意識調査をした。前者の分析結果（Fig.2）と、後者のうち「授業中の教室内の騒がしさ」の回答の結果から、教室内の吸音材の設置によって騒音レベル及び騒がしさが抑制される傾向が確認された。

Fig.1：The classroom with sound absorbing materials

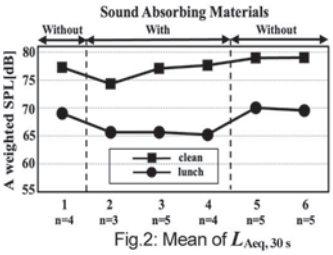


Fig.2: Mean of $L_{Aeq,30\,s}$

1-9-5

1-9-5　屋外における拡声音の長距離伝搬に風が及ぼす影響ー屈折を考慮した音線法による検討ー

The effects of wind on long-distance propagation of amplified sound outdoors: A study using the ray-tracing method considering refraction

〇佐藤逸人（神戸大院・工学研）

◆本稿では音線法を用いたシミュレーションにより，風が防災用屋外拡声システムの拡声音に及ぼす影響について基礎的な検討を行った。

◆風による屈折は，光線方程式（ray equation）を用いて音速が空間勾配を持つ場における音線軌跡を求める方法により考慮した。

◆シミュレーションは，風速，音源の高さ，地表面の傾きをパラメータとして行った。

◆逆風による上向きの屈折によって音圧レベルが低下し始める音源からの距離は，風速によらず音源位置が高くなるほど長い（Fig.1）。つまり，音源位置を高くすることで風の影響を軽減できる。

◆本研究の範囲では，地表面の傾きよりも屈折による音線の上昇の傾きが大きく，地表面の傾きの影響はわずかであった。

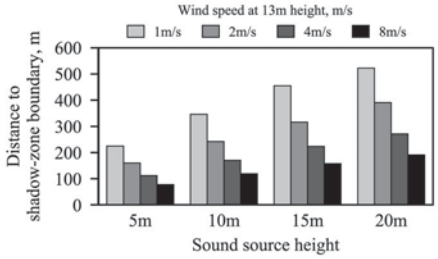


Fig.1：The distance to the shadow-zone boundary, i.e., the distance where SPL reaches −3 dB relative to the no-wind condition, for each wind speed.

1-9-6

1-9-6　パノラマ写真によりVR表示された建築空間の視覚情報に調和する残響時間の提案

−先行研究で提案されたモデルに新たな対象空間を追加した検討 その 2−

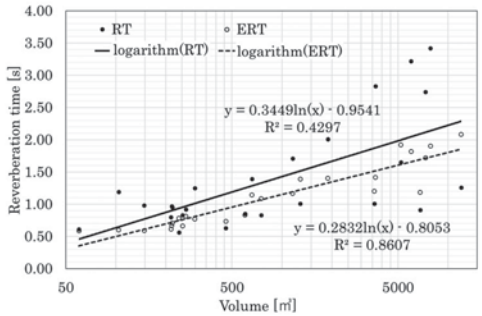
Proposal of reverberation time to match the visual aspect of architectural space displayed in VR using panoramic photographs

☆岩佐茉音歌，石川あゆみ（岐阜高専）

◆これまでの視覚印象の主観評価実験と予想残響時間 (ERT) の同定実験の結果や、それを基に作成された、パノラマ写真によりVR 表示された建築空間の視覚情報に調和する残響時間を求めるモデル（以下モデル）は、１つにまとめられた実験結果やモデルに含まれる被験者や対象空間の数が少ない。

◆関連する一連の研究で蓄積した VR 空間全ての、RT・ERT・広さに関する視覚情報の数値データ・視覚印象の尺度データの関連性を分析して散布図を検討し、モデルとして適切なものを提案する。

◆以下に、本報で提案するモデルを示す。



Model using volume and ERT with RT

1-9-7

1-9-7　簡易バイノーラル可聴化における音源方向定位に関する聴感実験

Auditory experiments on directional localization of sound source in a simple binaural auralization system

〇小松大介，篠原雄一郎，和田晋一，松原玄彦，△米倉勲
（TOPPAN ホールディングス），大林紅音，兪ハニ，佐久間哲哉（東大・工）

◆室内空間の VR 体験システムへの実装を想定した簡易的なバイノーラル可聴化手法について研究している。

◆直接音と反射音を分けて可聴化する手法に関して聴覚刺激のみの聴感実験を実施し、提案手法による静止・移動音源に対する直接音の方向定位感の再現性について検証した。

◆静止音源の方向定位では前後誤判定がかなり多く、特に後方に定位が偏る傾向にあった。

◆移動音源に対しては 2 割の被験者で方向定位が有意に認められたものの、全体的には前後判定は難しいことが明らかとなった。

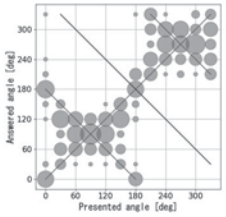


Fig.1：Result of static experiments

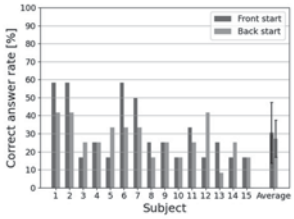


Fig.2：Result of dynamic experiments

1-9-8

1-9-8　比音響インピーダンスのアクティブ制御による室内音場調整の試み

An Attempt at Room Sound Field Adjustment Through Active Control of Specific Acoustic Impedance

◎久代連太，尾本章（九大・芸工）

◆比音響インピーダンス制御は，定在波や固有モードによる部屋の音響特性の不均一さを抑制することを目的とし，比音響インピーダンス制御を導入する。

◆一般的なスピーカを 2 次音源として用いたアクティブ制御によって室内音場を調整する手法を提案する。

◆通常のアクティブ制御とは異なり，制御対象を音圧と粒子速度の比である比音響インピーダンスとし，間接的に音場の等化と音波の進行方向の制御を試みた。

◆下図の様な仮想境界面での吸音率を最大化させるような処理でこの制御は実現できる。

◆結果より，吸音材料でのパッシブ制御が難しい低周波域での音場調整手法としての有効性が示された。

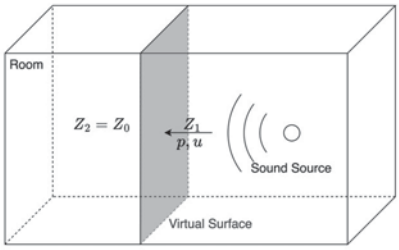


Fig.1 ： Conceptual diagram of specific acoustic impedance control

1-9-9

1-9-9　残響付加システムのモデルベースド制御に関する研究－室内音響指標に基づく自動調整－

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

☆河野光貴（東大・エ），渡辺隆行，大木大夢（ヤマハ），佐久間哲哉（東大・エ）

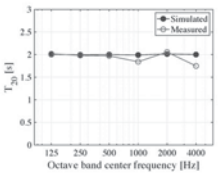
◆筆者らは、電気音響による残響付加システムの自動調整に向けて、制御用スピーカとマイク間のインパルス応答計測に基づくモデルベースド制御の検討を進めている。

◆YAMAHA 製 Active Field Control Enhance (AFC)を使用し、既報では AFC 作動時の評価用 RIR のシミュレーションモデルを構築した。

◆本報ではシミュレーションモデルを使用し、ハウリング・カラーレーション対策と室内音響指標に基づく最適調整の 2 点を目標として、EQ 調整手法について検討した。

◆ループゲインの平坦化、ループゲイン行列の固有値に基づくハウリング対策、カラーレーションの有無の判別がモデルベースド制御によって可能であることが確認できた。

◆T_{20} と ST_{Early} の目標値を定め、調整パラメータを勾配法によって反復計算することで最適調整を行った。T_{20} に関して誤差は 0.2 s 程度、ST_{Early} に関して誤差は 1 dB 程度となり、良い対応が確認された。



	T_{20}						ST_{Early}
Frequency	125	250	500	1k	2k	4k	
Target Value	2.00	2.00	2.00	2.00	2.00	2.00	-13.0
Simulated Value	2.00	2.00	2.00	1.99	2.01	2.00	-13.1
Measured Value	2.02	1.98	1.97	1.84	2.06	1.75	-14.1

Fig.1 : Reproducibility of room acoustic parameters.

1-9-10

1-9-10　木造建築物の床衝撃音遮断性能の現状と課題

Status and challenges of floor impact sound insulation performance in timber construction buildings

○平光厚雄（建研）

◆2010 年制定の「公共建築物等における木材の利用の促進に関する法律（現、脱炭素社会の実現に資する等のための建築物等における木材の利用の促進に関する法律）」により，木造建築物が増加している。

◆木造建築物の重量床衝撃音遮断性能は低くなるため，木造建築物の普及阻害の要因の一つとなっている。

◆重量床衝撃音対策の基本は、⓪音源室と受音室の配置計画を考慮，①床への衝撃入力の低減，②床躯体構造による低減，③天井での遮音，④受音室内での制御 と纏められる。

◆重量床衝撃音対策として床仕上げ構造に乾式二重床構造，天井構造を独立天井とした木造小学校では，重量床衝撃音（タイヤ衝撃源）で L_r-60（日本建築学会遮音性能基準の適用等級 2 級）程度の性能であった（Fig）。

◆現状の課題としては，①性能の向上，②音環境分野の地位向上，③木造床の重量床衝撃音遮断性能の表記方法のルール化，④正しい情報発信 などが挙げられる。

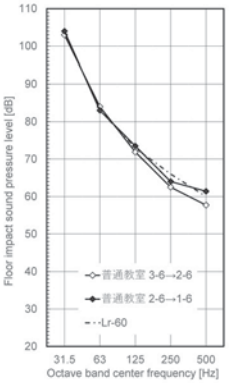


Fig. Measurement example of heavy-weight floor impact sound insulation performance at elementary school

1-9-11

1-9-11　共同住宅の居住者を対象とした音環境調査と床衝撃音に関する生活実感

Sound environment survey and living experience regarding floor impact sound for residents of apartment buildings

○冨田隆太（日大・理工），△阿部今日子（日大・芸術）

◆共同住宅は，上下左右を別住戸と接しているため，界壁の透過損失，界床の床衝撃音対策が重要となる。

◆しかしながら，居住者に隣，上下階の住戸から伝わる音は，界壁，界床の遮音性能，床衝撃音遮断性能のみで決まるものではなく，他住戸から伝搬する音の発生強度も関係する。

◆また，居住者の音に対する反応は，時代やその背景によっても変化するものと考えられ，定期的に調査していくことが重要であると考えられる。

◆本報では，1970 年代から約 50 年間について，共同住宅の居住者を対象とした音環境調査について考察した。

◆上階からの騒音を考察すると，給排水音などはかなり改善された。

◆また，ピアノ・楽器の音は居住者の配慮により指摘が少なくなったと考えられる。

◆一方で，床衝撃音については，気になる等の指摘率は下がっていると考えられるが，比較すると，気になる音の上位に現在でも位置していると言える。

◆共同住宅の居住者を対象とした音環境に関する生活実感の表現内容は重要であると考えられる。

◆そこで，本報では，共同住宅の居住者からの指摘が多いとされる床衝撃音遮断性能と生活実感の研究例を述べた。

1-9-12

1-9-12　非 RC 造の実建物を対象とした標準重量衝撃源(ゴムボール)による重量床衝撃音遮断性能の調査

Report on Heavy-Weight Floor Impact Sound Insulation Performance of Actual and Test Non-Reinforced Concrete Buildings Using the Standardized Heavy-Weight Floor Impact Source

○平川侑（国総研）

◆国土交通省国土技術政策総合研究所では、国土交通省総合技術開発プロジェクト「社会環境の変化に対応した住宅・建築物の性能評価技術の開発（R4～R8 年度）」を実施している。本プロジェクトにおいて、住宅性能表示制度における「音環境に関すること」、特に重量床衝撃音対策に関する合理化について検討している。

◆この取り組みの一環として、一般社団法人住宅生産団体連合会の住宅性能向上委員会 WG に参画する事業者に対し、非鉄筋コンクリート造住宅における実建物の重量床衝撃音に関する測定データを提供いただいた。

◆本論については、国土技術政策総合研究所資料 1263 号[1]として取りまとめ公開しているが、ここでは、当該資料の背景と趣旨、そして今後の取り組みを紹介する。

1-9-13

1-9-13　木質パネル接着工法床の床衝撃音測定結果の例

Example of floor impact sound results in wooden panel adhesion system

○渡辺大助(ミサワホーム総研),

◆筆者らは，実際に運用している標準的な断面仕様の木質パネル接着工法の床（床パネル）でも、床衝撃音遮断性能 L_{rH}-45～55 が達成できることを報告してきた。今回、木質パネル接着工法の実物件を対象として，二層二重床を用いた L_{rH}-55 仕様の性能検証を実施する機会を得た。

◆その結果，重量床衝撃音レベルの設計性能 L_{rH}-55 の実物件において所要の性能が確保できることを確認した。軽量はタイルカーペットを採用したことによって L_{rL}-45 という高い性能が得られた。

◆重量の結果に関して、タイヤとゴムボールを比較すると，衝撃力の周波数特性の差がそのまま現れたのは 63Hz 帯域のみで，125Hz 帯域以上では衝撃力の大きいタイヤ加振の方が大きな値を示した。これは，衝撃力の大きいタイヤ加振で建具などで二次的な振動が生じたためと考えられる。

◆重量の結果に関して，L 数と A 特性音圧レベルを比較すると，両者はほぼ同等の値となる傾向にあったが，同程度の L 数を示す周波数帯域が 3 つ程度ある場合，A 特性音圧レベルの方が大きな値となった。

1-9-14

1-9-14　木質系及び鉄骨系工業化住宅における床衝撃音の事例に関する考察

Consideration of case studies on floor impact sound in wooden and steel frame industrialized housing

○渡邉　将平(大和ハウス総技研)

◆本報告では木質系及び鉄骨系工業化住宅において床衝撃音の対策事例に着目し考察を行った。木質系住宅では床・天井面対策により LH-70 程度の性能となることがわかった。一方，鉄骨系工業化住宅では床材・天井材の防振支持や天井の質量を付加することで ALC を用いない比較的軽量な構造でも LH-65 の性能を確保できることが明らかにした。

◆床板と構造躯体が切り離された鉄骨系工業化住宅では加振点直下に配置される床パネルからの放射の影響が支配的であるため、パネルの大きさが放射面積に直結し、床衝撃音レベルの大小を決めている一因となっていると考えられる。

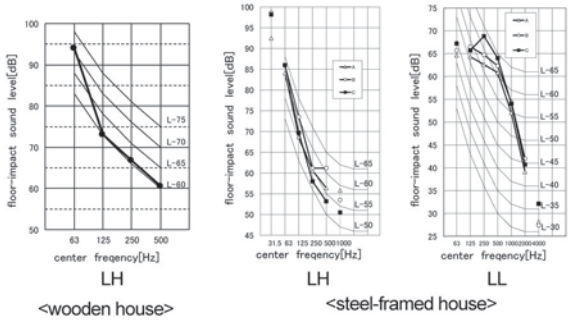


Fig.1 : Floor impact sound level in wooden and steel-framed house

1-9-15

1-9-15　鉄骨共同住宅における重量床衝撃音対策

Countermeasures for heavy-weight floor impact sound in steel-framed prefabricated apartment buildings.

○永松英夫（積水ハウス総合住宅研究所）

◆鉄骨造共同住宅における重量床衝撃音対策として，当社の床仕様の変遷および仕様選定の考え方を示し，床衝撃音性能の測定事例を通じて，各床仕様における測定結果のばらつきや，ボール衝撃源に関する課題を紹介した。

◆高遮音床の開発においては，床の剛性化・重量化が有効である。しかし，軽量化や低コストが求められるプレハブ住宅では，防振材や動吸振器などの防振技術を活用する必要がある。

◆ボールを衝撃源とした **BA 数**とタイヤの **L 数**との関係を考察し，木造および鉄骨造における**床衝撃音の合理的な評価方法**を検討する。

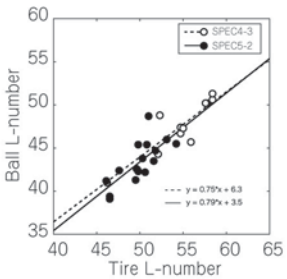


Fig.1 : Relationship between tire L-number and ball L-number.

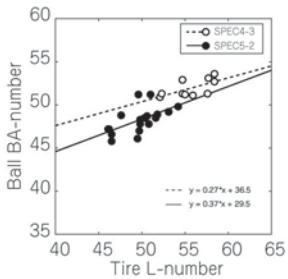


Fig.2 : Relationship between tire L-number and ball BA-number.

1-9-16

1-9-16 有限要素法による二重天井の重量床衝撃音解析に関する基礎的検討

Basic study on heavy-weight impact sound analysis of double ceiling by finite element method

○曹達（東大・エ），會田祐（長谷工技研），
井上尚久（九大・芸工），佐久間哲哉（東大・エ）

◆有限要素法を用いて，二重天井の重量床衝撃音低減性能における天井下地構造の影響を解析した。

◆野縁や吊りボルトなどの構成要素を段階的に追加した数値モデルを作成し，その性能を評価した。

◆下地構造の複雑化により低周波数帯域（例：63 Hz）での低減効果が向上する一方，高周波数帯域では性能が低下する傾向が示された。

◆二重天井の数値解析には，天井下地材の適切なモデル化が重要であることが明らかになった。

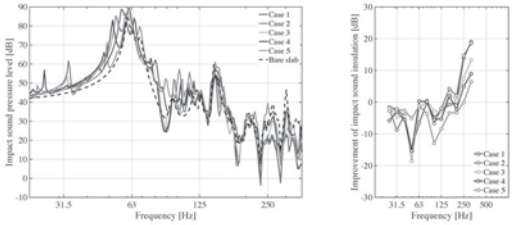


Fig. 1: Simulated results using different numerical models.

1-9-17

1-9-17 RC 造建物における床加振時の構造体振動および二重床振動低減量に関する実験

Experiments on the structural vibration and the vibration reduction of double floor system by floor excitation in an RC building

○會田祐，△室裕希（長谷工技研），曹達（東大・エ），
井上尚久（九大・芸工），佐久間哲哉（東大・エ）

◆RC 造建物の床スラブ振動の数値解析を行う場合，周辺構造を含むモデル化範囲や境界条件設定が解析結果に影響を及ぼす可能性がある。

◆RC 造建物の重量床衝撃音解析に向けた検討として，実験建物において，床スラブ加振時の加振スラブ，隣接スラブ，RC 壁等への伝達モビリティを測定し，構造体の振動性状を把握した。

◆さらに同建物において，二重床による床スラブ振動の低減量を各標準衝撃源について求めた。

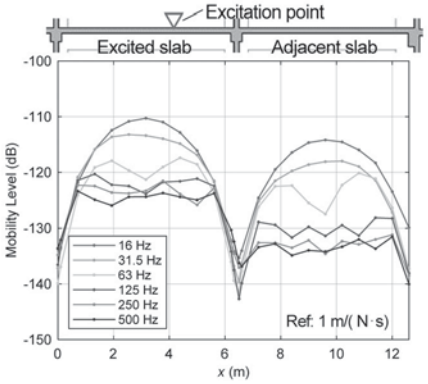


Fig.1：Distribution of the mobility levels measured on the excited slab and the adjacent slab

1-9-18

1-9-18 曲面外装の遮音性能に関する実験的検討

Experimental Study on Sound Insulation Performance of Curved Glass

◎佐藤暢、宮島徹、石塚崇（清水建設技研）

◆ヨーロッパを中心に建物外装に曲面ガラスを採用する例が増えている。一方で、外装の遮音設計に必要な遮音性能のデータはない。

◆本報では、平板と曲率の異なる２種類のガラス試験体の音響透過損失とインパクトハンマー加振による振動性状測定を行った。

◆測定の結果、曲面ガラスでは、平板と比較して低域に音響透過損失の落ち込みが現れることが分かった。振動測定の結果とあわせて考えると、曲げによって一次固有振動数が上昇し、共鳴領域が測定範囲に現れたためと推測される。

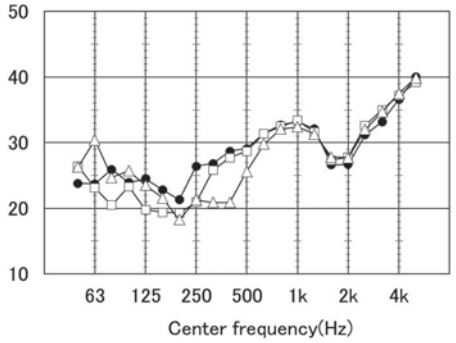


Fig.1：Sound transmission loss of three specimens with different bending.
(Participants ●:Flat, □:R3,400, △:R2,000)

1-9-19

1-9-19 微小粘性空気層を介して連結する積層板材の音響透過損失の数値解析的検討

Numerical Consideration of Sound Transmission Loss of Laminated Plates Connected via a Thin Viscous Air Layer

◎米澤 美桜，井上 尚久（九大・芸工）

微小空気層を介して連結する積層板材において，空気層の減衰を考慮した斜入射音響透過問題を取り扱う。音響振動連成系に拡張した粘性・熱伝導境界層の影響を考慮する二つのモデル（境界条件モデル，ナビエ–ストークス方程式に基づく詳細モデル）を構築し，両モデルを用いた解析を通じて現象の考察を行う。

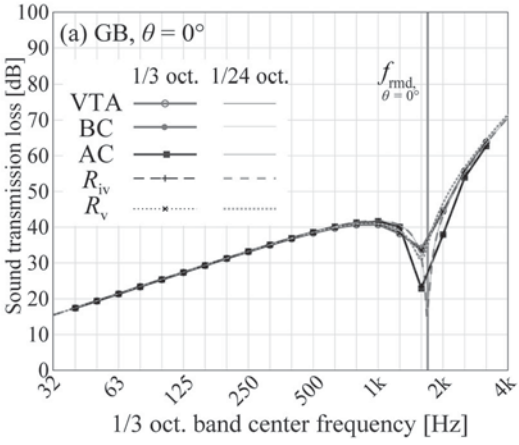


Fig.1：Sound transmission losses calculated by using VTA, BC, AC, R_{iv}, and R_v models.

1-10-1

1-10-1　オープンプラン型オフィスの音環境とワーカーのパフォーマンスの評価構造に関する検討

A Study on the Evaluation Structure of Sound Environment and Worker Performance in an Open-Plan Office

☆中橋 樹香(近畿大院), △冨樫 建五(大和ハウス工業), 原田 和典(岡山県立大),
菅原 彬子, △長澤 康弘, 平栗 靖浩(近畿大), △岩切 幸伸(コクヨ)

◆オープンプラン型オフィスでの実地調査により実環境下での音環境と作業者の印象評価を収集し、音環境と印象評価との関係について、パス解析を用いて検討を行った。

◆個人作業の評価構造については、「音環境の快適さ」が知的生産性と大きな関連を持つ項目であり、静かな環境が知的生産性を向上させることが示唆された。(Fig.1)

◆複数人作業の評価構造においては、「会議のしやすさ」が知的生産性と比較的大きな関連を持つ項目であり、「自発音の聞こえやすさ」に対しても敏感であったことから、話し合いのしやすい環境が求められているといったことが示唆された。

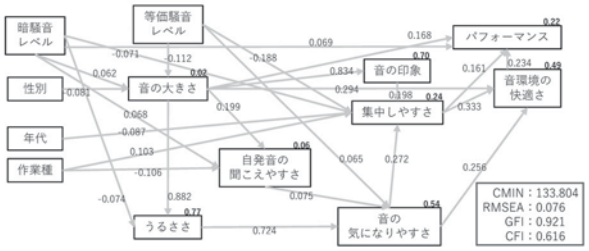


Fig. 1: path diagram (Individual work)

1-10-2

1-10-2　騒音の空間性がうるささ評価に及ぼす影響に関する実験的検討

Experimental study on the effect of the direction of noise on noisiness

☆小田切彩夏(東大大学院), 米村美紀(前橋工科大),
森長誠(大同大), 坂本慎一(東大生研)

◆既往研究では、音源の騒音の方向性や空間性によりうるささ感が異なる傾向が示唆されている。本研究では空間性条件間での定量的なうるささ評価の差を検討するため、被験者調整法による実験を行った。

◆試験音の方向条件は、Fig. 1 に示す 4 条件とした。標準刺激と比較刺激を交互に呈示し、比較刺激のうるささが標準刺激のうるささと等しいと感じるまで手元のコントローラーによって被験者自身が調整した。標準刺激としては 1ch（上）条件の音を使用した。また、標準刺激と同じ音も比較刺激に含めた。

◆周囲から音が聞こえてくる条件では、ある特定の方向から騒音にさらされる条件よりもうるささ感が増す傾向が示唆された。

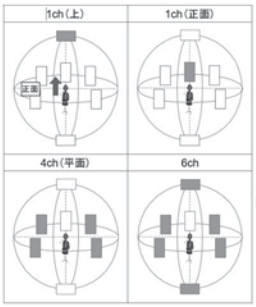


Fig.1:Direction of test sound

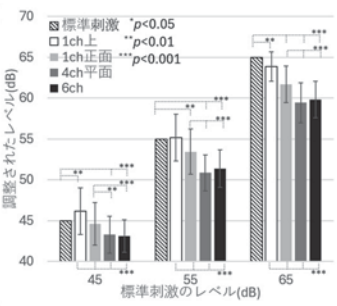


Fig.2:Standard stimulus and adjusted sound levels

1-10-3

1-10-3　多機能計測システムを用いた環境騒音のラウドネスに関する実測調査

Field survey on environmental noise loudness using a multifunctional measurement system

○菅原彬子(近畿大), 米村美紀(前橋工科大), 坂本慎一(東大生研)

◆ラウドネスレベル(*LL*)は，人間の聴覚特性を精緻に反映する指標であり主観的なラウドネスと高い相関を示す．本研究では，ISO 532-1・ISO 532-2 に基づく *LL* 測定アプリを開発した．

◆提案システムにより *LL* の手軽なリアルタイム測定が可能となった．

◆A 特性音圧レベルと *LL* を測定時間全体でエネルギー平均した L_{Aeq}, LL_{Zeq} (ISO 532-1), LL_{Meq} (ISO 532-2)はレベル・周波数依存性が異なり，同じ騒音に対し異なるラウドネス評価をしうることが示唆された．

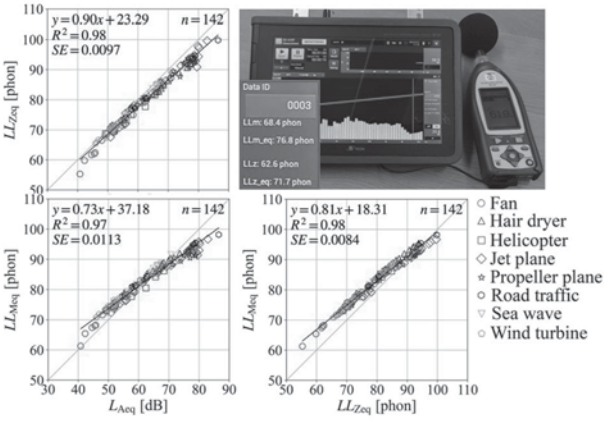


Fig.1:Relationships of L_{Aeq}, LL_{Zeq}, and LL_{Meq} for all data

1-10-4

1-10-4　時間変動のある環境騒音のうるささとラウドネス評価指標の関係

Relationships between noisiness of time-varying environmental noise and metrics for loudness

○米村美紀(前橋工科大), 矢野拓実, 菅原彬子(近畿大),
森長誠(大同大), 坂本慎一(東大生研)

◆筆者らは，種々の環境騒音の評価指標としてのラウドネスレベルの可能性を検討している。本報告では，交通騒音のような数秒～数十秒の周期で時間変動する騒音のうるささと，ラウドネス評価指標（A 特性音圧レベル，ラウドネスレベル）との関係を調べた。

◆等価騒音レベルは試験音（5 分間）全体のうるささと強い相関が認められた。ラウドネスレベル（ISO532-1，2）を 10 秒ごとに算出し時間平均した値も同等の相関の強さであった。

◆指標による差異と心理反応の関係についてはさらなる検討が求められる。

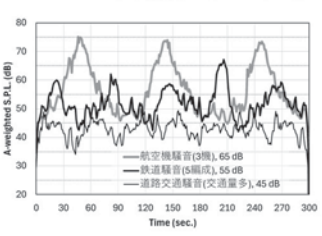


Fig.1:Time variation of levels of the test stimuli

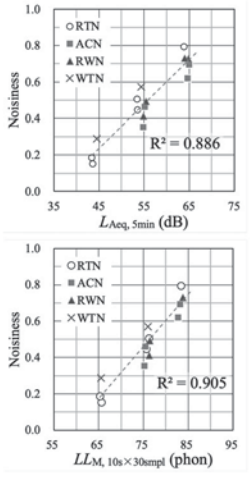


Fig.2:Relationships between noisiness and loudness metrics

1-10-5

1-10-5　個人ばく露測定によるドリルを用いた穴あけ作業音評価の試み

Evaluation of drill bit noise during drilling operations by personal noise exposure measurement

○横山 栄，小林知尋，横田考俊（小林理研），△田中典英（ミヤナガ）

2023 年、騒音性難聴防止に配慮し、騒音障害防止のためのガイドラインが改訂され、作業環境測定法として、新たに個人ばく露測定も認められた。特に手持動力工具を使用する場合、継続して等価騒音レベルが 85 dB 以上の場合、必ず聴覚保護具を使用させる必要がある。本報では、電動ドリルによる穴あけ作業を対象に、個人ばく露計による測定を実施し、労働衛生管理の観点から評価を試みた。その結果、1 条件を除き、85 dB を超過しており、振動式やハンマー式は 95 dB を超過していた。1 日 8 時間 85 dB の許容曝露レベル（PEL）に換算すると、95 dB 以上の作業環境では、許容曝露（作業）時間（PET）は 1 時間以下となる。

Table 1 Examined drills, and results of personal noise exposure levels.

no.	drill type	bit size	time [s]	$L_{Aeq,T}$ [dB]	L_{Cpeak} [dB]	PET
1	rotary/wet	18 mm	163	84.7	99.5	8h00m
2E		6 mm	85	87.6	102.3	4h23m
3	rotary/dry	6 mm	81	87.2	103.0	4h49m
4	percussion	6 mm	94	96.7	112.8	0h32m
5	hammer	18 mm	64	98.9	118.6	0h19m
5D			97	96.5	121.3	0h33m
5E			66	97.2	119.8	0h28m
6			81	99.6	117.7	0h16m
6D			111	96.0	117.1	0h38m

※ subscript for no.: D; with dust collector, E; with power supply (100 V).
※ time: operation time for drilling three holes.
※ PET: permissible exposure time per day (max.; 8h, for PEL of 85 dB).

1-10-6

1-10-6　振幅変調低周波音の感覚特性

Sensory characteristics to amplitude-modulated low-frequency sounds

○松田 礼（日大・理工），町田 信夫（日大）

◆本研究の目的は，低周波数領域の純音を搬送波とした振幅変調低周波音（以下，AM 低周波音）による感覚特性を定量的に評価する指標を確立することである。本報では，AM 低周波音を構成する物理量と心理反応量との関係を調べた結果について報告する。

◆AM 低周波音は純音の搬送波を振幅変調して作成した。AM 低周波音の音条件は，搬送波周波数，等価音圧レベル，変動周期および変調度を組み合わせて設定した。

◆AM 低周波音のラウドネスと変動感は，搬送波周波数が増加すると大きくなる傾向がみられた。定常低周波音のラウドネスと変動感は搬送波周波数 40 Hz で最も小さくなる傾向がみられた。

◆AM 低周波音の不快感，振動感，圧迫感は変調度が大きくなると増加する傾向がみられた。

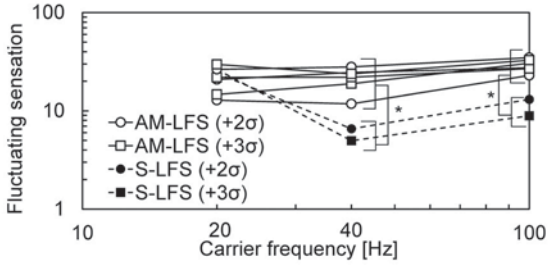


Fig.1 : Relationship between carrier frequency and fluctuating sensation (*: $p < 0.05$, Dunnet's multiple comparison test)

1-10-7

1-10-7　Determination of sound power level of aircrafts based on on-site measurement

☆ Xinyi ZHANG (Grad. Schl., Eng., The Univ. of Tokyo), Makoto MORINAGA (Daido Univ.), Shinichi SAKAMOTO (IIS, The Univ. of Tokyo)

◆To determinate the sound power level of aircrafts, this study conducted on-site measurement during the landing approach of various aircraft types, including large, medium, small, and turboprop-powered aircraft.

◆Flight paths were reconstructed using video analysis and geometric methods, incorporating parameters such as speed and altitude.

◆Based on the measured sound pressure levels at specific points, the sound power levels of different aircraft types during the landing approach were calculated.

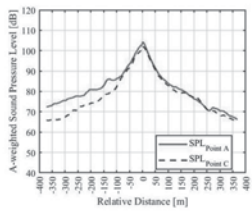


Fig.1 : Measured sound pressure levels for B738.

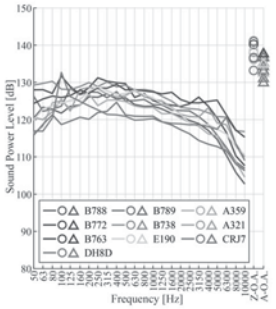


Fig.2 : Sound power levels for various aircraft types.

1-10-8

1-10-8 Experimental study on frequency-dependent prediction model of insertion loss of buildings

☆Qiyuan Wang (The Univ. of Tokyo), Ken Anai (Fukuoka Univ.), Hiroo Yano, Shinichi Sakamoto (IIS, The Univ. of Tokyo)

◆The ASJ RTN-Model provides an effective prediction model of road traffic noise (RTN) behind buildings, characterized by the insertion loss (IL, noise reduction level) of buildings. However, the model is limited to fixed frequency characteristic of the noise source (RTN).

◆In this work, we report the attempt of forming frequency-dependent prediction equations based on scale model experiments.

◆The modified prediction equations maintain the same structure: $IL = p10\log_{10}\left\{b_0 + b_1\frac{\phi}{\Phi} + b_2 10^{-0.0904\xi d_{SP}}\right\} + q$, with same geometrical variants ϕ, Φ, ξ, d_{SP} while different constants b_i.

◆We examine IL prediction model for (a) different octave bands with $b_i(f)$, and (b) noise source with arbitrary frequency characteristics using a b_i synthesis method from $b_i(f)$.

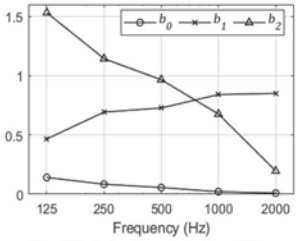


Fig. 1 Optimized constants b_i at 1/1 octave frequency bands.

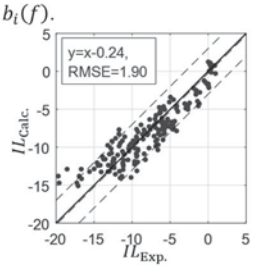


Fig. 2 Comparison of experimental and calculated $IL_{OA,RTN}$ with synthesized b_i.

1-10-9

1-10-9　一般道路における排水性舗装のパワーレベル測定・算出方法

−最大騒音レベル法と二乗積分法の算出値の比較−

Comperison of sound power levels of running vehicles on general roads by the integral method of squared sound pressure on the maximum A-weighted sound pressure level

○澤田泰征，△橋本浩良（国総研）

◆自動車走行騒音の音響パワーレベル LWA の測定において騒音レベルの測定データパワーレベルを算出する方法には最大騒音レベル法と二乗積分法がある。

◆全国 10 箇所の排水性舗装・一般道路の測定データから 2 つの算出方法の比較を車種別に行ったところ、算出値は全体としてよく一致している。

◆ただし、速度が速い場合、音源から測定点までが近い場合は、最大騒音レベル法の算出値は二乗積分法の算出値に比べて小さくなる傾向がある。

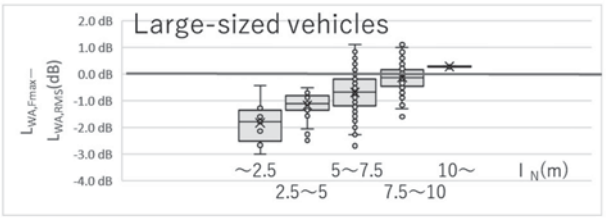


Fig.1：Relationship between horizontal distance from sound source to measurement point and level difference

1-10-10

1-10-10　自機振動の変動成分に基づく接近ドローンの音源探知手法の提案

Proposal of sound source detection method for approaching drone based on modulation components of self-vibration

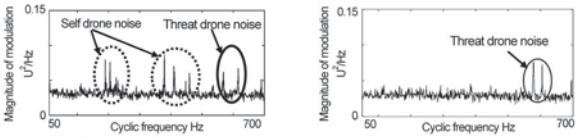
☆永田健太郎，△麻生海（中央大院），田辺総一郎，戸井武司（中央大）

◆近年ドローンが普及し，ドローン同士の衝突回避が必要となっている。そこで，小型軽量かつ低消費電力センサを用いたドローン同士の衝突回避が必要となっている。

◆本研究では，自身の機体(以下，自機)から衝突の脅威となる機体(以下，脅威機)の探知を行う際に，自機に搭載した自機の各プロペラ付近に設置した加速度計とマイクロホンを使用し，変調周波数領域に着目した処理を行うことで脅威機音の探知を行う。

◆自機音と脅威機音が混在している状態であっても，自機の各プロペラ付近に設置した加速度計から取得する振動を利用し，自機音を推定および除去することで脅威機音を推定する。

◆自機音と脅威機音が混在したFig. 1(a)に示す変調周波数実測音に対して，自機から取得した振動を利用し，変調周波数から脅威機音検知の処理を行うとFig. 1(b)となり，脅威機が自機と異機種または同機種の場合においても脅威機を検知できた。



(a) Before applying signal processing flow　(b) After applying signal processing flow

Fig. 1　Modulation frequency analysis of drone operating sound

1-10-11

1-10-11　相互相関関数を用いた音源位置推定に関する屋外実験

Outdoor experiment on sound source location estimation using cross-correlation function.

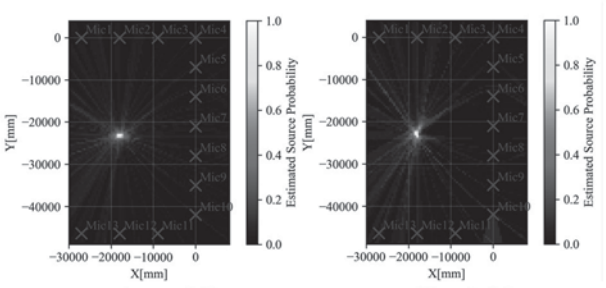
☆森本誠至（近畿大院），原田和典（岡山県立大），
菅原彬子，平栗靖浩（近畿大）

◆相互相関関数を用いた音源位置推定における屋外での音源位置推定精度の検証を行うため，屋外実験を実施した。

◆音源位置推定に用いるマイクロホンが推定対象音源を十分なSN 比で受音できる場合，低い推定誤差で音源位置推定が行われた。

◆SN 比が低下した場合，到来時間差を誤って算出したマイクロホンの組み合わせにより推定誤差が大きくなった。

◆提案するグリッドベース手法による音源位置推定(Fig.1)では，相互相関係数をグリッドに割り当てることで，SN 比の低下などにより発生する音源位置推定精度への影響を低減できることが示唆された。



(a) 高 SN 比条件　(b) 低 SN 比条件

Fig.1 グリッドベース手法による音源位置推定

1-10-12

1-10-12 電子膨張弁の異音発生メカニズム解明に向けた発音と振動特性の計測

Measurement of Sound Pressure and Vibration for Elucidation of Noise Generation Mechanism of Linear Expansion Valve

◎新井達也（三菱電機），高橋佳吾，保母陽介，榊原潤（明治大学）

◆空調冷熱機器では，電子膨張弁と呼ばれる流量が可変な膨張弁が使用されている。特定の運転条件で電子膨張弁から異音が発生することがあり，異音発生要因の解明と恒久的な対策が求められている。

◆本研究では，内部を観察可能な透明流路に電子膨張弁を組み込んだサンプルに空気を流して異音と電子膨張弁の弁振動を計測した。

◆その結果，単一周波数成分の音が発生し，電子膨張弁も同一の周波数で共振振動していることが明らかになった。電子膨張弁の共振振動に起因して音が発生しており，振動対策で異音を対策できる可能性が示唆される。

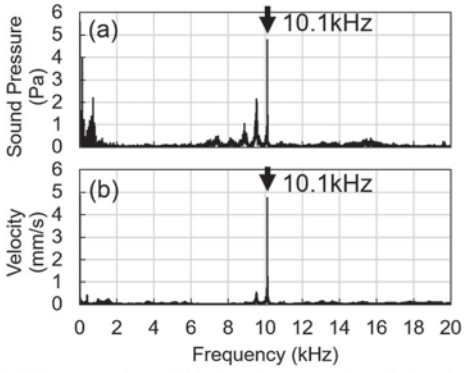


Fig.1：Measurement result (a) Sound pressure, (b) Velocity of needle tip.

1-10-13

1-10-13　骨格-膜構造型メタマテリアルを用いた遮音構造における骨格振動の影響

Effects of rib vibration on sound insulation structures using rib-membrane type metamaterial

◎宮本 光亮, 芦澤 剛(日本音響エンジニアリング),

放射面の振動そのものを抑制することによらず，その音響放射効率が低くなるような構造を設計することによって透過音を低減できる構造として著者らが提案した骨格-膜構造について，骨格が自由に振動できる場合の遮音性能に与える影響について検討した．構造全体が自由に振動できる場合においても，注目周波数領域で質量則以上の遮音性能を示すことを実測および FEM により確認したこと，および理論的な考察を報告する．

◆２種の膜を持つ骨格-膜型構造の遮音性能に関する全体振動の影響について数値解析および音響管を用いた実測結果と比較・考察した．

◆構造全体の振動が存在する場合においても，ピーク性の遮音性能を示すことを報告する．

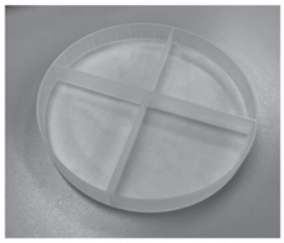


Fig. 1 3D-Printed AMM

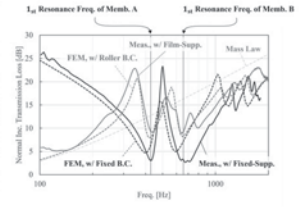


Fig. 2 Comparison of Transmission Loss by Meas. and FEM

1-10-14

1-10-14 ダミーヘッドマイクロホンによる聴覚保護具の遮音性能測定

Sound attenuation experiment of hearing protectors applying HATS.

〇横山 栄, 小林知尋(小林理研)

2023 年に、騒音作業場における難聴防止に配慮し、騒音障害防止のためのガイドラインが改訂された。等価騒音レベルが 85 dB を超える作業環境では、JIS T 8161-1 による主観的方法で測定される遮音値を参考に適切な聴覚保護具を選定し、着用することが求められ、特に激甚な騒音下では、耳栓とイヤーマフの併用が有効であると明記された。しかし、その併用効果は、SNR（single number rating）値で 5 dB 程度で、必ずしもすべての周波数帯域で効果がある訳ではなかった。本報では、耳栓、イヤーマフの単体着用時の他、併用時についてもダミーヘッドマイクロホンによる物理的方法で遮音性能の測定を試みた。その結果、耳栓とイヤーマフ併用時に一部の周波数範囲で併用効果がマイナスとなる現象が確認された (Fig. 1)。引き続き、このメカニズムの検討を進めている。

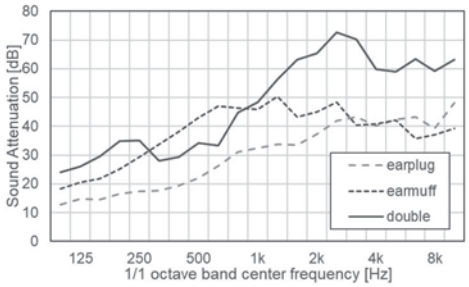


Fig.1 : Sound attenuation of hearing protector applying HATS.

1-11-1

1-11-1　操作されたピッチアクセントを持つ日本語のピッチパターンに対する妥当性の主観評価－語種間の比較－

Subjective Evaluation of the Acceptability of Pitch Pattern of Japanese Words with Manipulated Pitch Accents: A Comparative Study Across Word Types

〇勝瀬郁代(近畿大・産業理工), 白勢彩子(東京学芸大)

◆様々なピッチアクセントの特徴を持つ 4 モーラの日本語音声単語を作成し，日本語母語話者によるピッチアクセントの適切性評価をピッチアクセント特徴分布上にマッピングした。

◆同じアクセント型に分類される単語であっても，単語の語種によって適切性評定平均の分布が異なった。

◆本実験の結果は，ピッチアクセント知覚の多様性を示すとともに，言語学で報告されているアクセント変化に関する知見を音声知覚の観点から定量的に支持する。

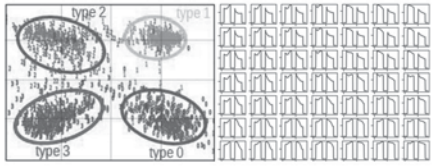


Fig.1 : (Left figure) Distribution of latent variables representing pitch accent features. (Right figure) Illustrative examples of F0 reconstructed from individual latent variables.

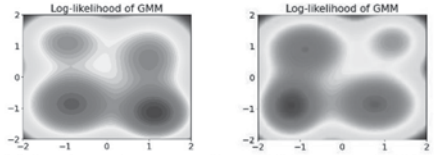


Fig.2 : Distributions of average ratings of simple verbs (left figure) and compound verbs (right figure) with accent type 0

1-11-2

1-11-2　自己の録音音声の印象評価に複数の個人特性が与える影響の分析

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

◎柳田耀, 井島勇祐, 俵直弘(NTT)

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

◆反応の違いには聞き手の個人特性（年代，性別，性格など）が影響することが示唆されており [Peng+, 2020]，前報 [Yanagida+, 2023] では，自己の録音音声の魅力や親近感等の印象に聞き手の個人特性（性格，習慣など）が影響することが示されている．しかし，各個人特性の影響を個別に検討したのみで，複数の個人特性が印象に与える影響は分析されていない．

◆本稿では，大規模な主観評価実験を実施し，自己の録音音声に対する印象に複数の個人特性が与える影響を総合的に分析する．

◆分析の結果，前報で影響が示された対人円環モデルの孤独-内向的等を含む複数の個人特性が自己の音声の印象に影響していることが示された．

Table.1 : AIC values for each combination of variables in the multiple logistic regression analysis

	No. of personal traits	Combination of personal traits	AIC value
Attractiveness	1	IPIP: Aloof-Introverted (75th)	185.36
	6	IPIP: Aloof-Introverted (75th) + BigFive: Conscientiousness (25th) + Four Factors of Happiness: Be yourself (25th) +IPIP: Warm-Agreeableness (25th) + Ten Basic Values: Self-direction (75th) + Prosocial Behavior: Empathy (25th)	178.95
Familiarity	1	IPIP: Aloof-Introverted (75th)	146.78
	3	IPIP: Aloof-Introverted (75th) + Dark Triad: Psychopathy (50th) + Frequency of listening to one's own recorded voice	142.68
Confidence	1	Approval Motivation: Conformity (50th)	189.20
	5	Approval Motivation: Conformity (50th) + IPIP: Aloof-Introverted (75th) + Dichotomous Thinking: Profit-and-Loss Thinking (25th) + Dispositional Greed (25th) + Unmitigated Communion (50th)	178.78
Intelligibility	1	IPIP: Aloof-Introverted (75th)	190.39
	9	IPIP: Aloof-Introverted (75th) + Dark Triad: Psychopathy (25th) + Unmitigated Communion (25th) + Approval Motivation: Conformity (50th) + Approval Motivation: Approval (25th) + BigFive: Conscientiousness (25th) + Four Factors of Happiness: Be yourself (25th) + Dichotomous Thinking: Profit-and-Loss Thinking (75th) + Ten Basic Values: Benevolence (75th)	179.08

1-11-3

1-11-3　多話者における直音と拗音の調音運動の分析

Analysis of articulatory movements for palatalized and non-palatalized consonants among multiple speakers

☆朝倉麗仁，△辰口隼太朗，竹本浩典（千葉工大），前川喜久雄（国語研）

◆本研究では，拗音の構造を検討するために，「リアルタイム MRI 調音運動データベース（rtMRIDB）から 11 名の話者を抽出し，対立する直音（/ka/，/ko/，/sa/，/so/）と拗音（/kja/，/kjo/，/sja/，/sjo/）の子音から母音への調音器官の変化パタンを主成分分析した。

◆その結果，直音・拗音とも第 1 主成分（PC1）の寄与率は 80%を超えることから，主要な調音運動は舌の前上方から後下方への移動であった。そして，全ての拗音で子音が硬口蓋化しており，子音と母音の形態的な相違が大きいことが明らかになった（Fig. 1）。

◆また，そのスコアの変化パタンはいずれも S 字状であったが，直音では有意に直線的（t 検定，$p<0.01$）であることから，直音と拗音では調音運動のパタンが異なることが明らかになった（Fig. 2）。

◆また，拗音のスコアの変化パタンはわたり音と同様なことから，拗音の調音運動はわたり音に近いこと，拗音は硬口蓋化した子音＋わたり音＋母音の構造を持つことが示唆された。

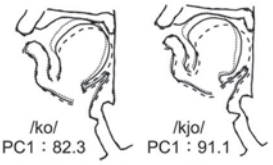


Fig 1: /ko/，/kjo/の調音運動の PC1
（数値：寄与率，点線：子音，
実線：平均，破線：母音）

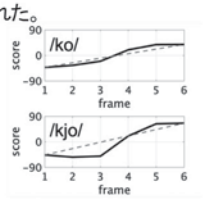


Fig 2:/ko/，/kjo/の PC1 の変化パタン
（実線：PC1，点線：PC1 を結んだ直線）

1-11-4

1-11-4　ESPnet による テーラーメイド音声合成の実践

A Practical Study of Tailor-Made Speech Synthesis Using ESPnet

○青木 直史，元由 勝人（北大）

- ALS などの病気により声を失う患者の QOL 向上を目的として，病状が進行する前に声を録音しておき，意思伝達ツールの音声読み上げに利用するしくみが広がりつつある。自分の分身ともいえる声を残すことは，言ってみれば人格を残すことに等しく，その意義は大きい。
- 本研究では，マイボイスを皮切りに，Open JTalk，そして ESPnet と，フリーの環境で実現できるテーラーメイド音声合成の方法について，これまでに調査を行ってきた。本発表では，ESPnet によるテーラーメイド音声合成の可能性について報告する。

① マイボイス

② Open JTalk

③ ESPnet（←イマココ）

1-11-5

1-11-5　Disambiguating Ambiguous Indonesian Utterances with ASR and Meaning Interpretation

Ruhiyah Faradishi Widiaputri[1], Ayu Purwarianti[2], Dessi Puji Lestari[2], Kurniawati Azizah[3], Dipta Tanaya[3], Sakriani Sakti[1]
[1]Nara Institute of Science and Technology, [2]Bandung Institute of Technology, [3]University of Indonesia

Existing ASRs are still limited to word-by-word transcription, overlooking ambiguities in the resulting text. To address this, we attempted to resolve structurally ambiguous Indonesian utterances into unambiguous texts by incorporating prosodic cues, which humans naturally use to resolve structural ambiguities. This study took a thorough approach, including corpus creation, human assessment, and the construction of the disambiguation system.

1-11-6

1-11-6　母語話者音声のみを用いた 外国語訛りに頑健な自動音声認識の実現に向けた離散トークンの活用の検討

Exploring the usage of discrete tokens for accent-robust automatic speech recognition only using native speech corpora.

☆恩田健太郎（東大院・工学系 / 産総研），深山覚（産総研），
井本桂右（同志社大・文化情報），齋藤大輔，峯松信明（東大院・工学系）

日本語のネイティブ音声 と，**英語のネイティブ音声** だけを使って，訛りのある日本人英語 の認識精度を向上！

↓ こんなことってありますよね？ ASR で再現してみました！

1-11-7

講演取消

1-11-8

音声品質と話者の声質の特徴量に基づいた Speaker Diarization Error Rate の自動推定

Automatic Estimation of Speaker Diarization Error Rate Based on Features of Audio Quality and Speaker's Voice Characteristics

○石塚賢吉, Chang Zeng, 大野正樹, 橋本泰一 (株式会社 RevComm)

本研究では、音声信号を入力として、音声信号をSpeaker Diarization(SD)した時のDiarization Error Rate(DER)を自動推定するモデルを構築する。DERの自動推定により、音声信号の推定DERが高い場合に、異なるアルゴリズムに切り替えるなどの対策を行うことが可能となる。現在のSDのアルゴリズムでは、入力された音声信号の音声品質が悪い場合や、複数の似た声の人が話している場合に、いつ誰が話しているかを高精度で推定することは難しい。そこで本研究では、音声品質の特徴量抽出部と話者の声質の特徴量抽出部と、回帰モデルから構成される手法で、DERを自動推定する。音声品質の特徴量抽出部は、音声信号から、2種類のVoice Activity Detection (VAD) アルゴリズムの結果の差異率、DNSMOS、VQScoreの3つの音声品質の特徴量を計算する。話者の声質の特徴量抽出部は、音声信号をSDして得られた話者ごとの音声区間のSpeaker Embeddingベクトルから、Silhouette Score と重心の偏差の2つの話者の声質の特徴量を計算する。回帰モデルは、5次元の特徴量と、モデリング対象の SD アルゴリズムによるDER の値との対応関係に対する重回帰分析により構築されるモデルであり、DERの推定値を出力する。

また、DER自動推定のモデリング対象をPyAnnote3.1のプリトレインドモデルとし、株式会社RevCommで行われたビデオ会議の音声データから構築したSD用のデータセットを使用して評価実験を行い、PyAnnote3.1のプリトレインドモデルによる話者ダイヤリゼーションのDERと、本研究で構築したモデルによるDERの推定値との相関関係を確かめる。

1-Q-1

1-Q-1 位相の混合微分とその強調に基づく調波打撃音分離

Harmonic and percussive source separation based on mixed partial derivative of phase and its enhancement

◎ 赤石夏輝, 矢田部浩平 (農工大)

調波打撃音分離 (HPSS)
音響信号の正弦波成分と打撃音成分を分離する処理

従来　位相の混合微分に基づく手法をこれまでに提案
分離マスクを高精度化しながら反復的に分離

課題　位相の混合微分の計算は**成分の混合に弱い**
→ 反復途中の**分離マスクの精度が低下**

提案　反復途中の**正弦波成分を強調**することで
分離マスクを高精度化

結果　打撃音成分の分離性能は同程度のまま
正弦波成分の分離性能を向上 (SDR +1.8 dB)

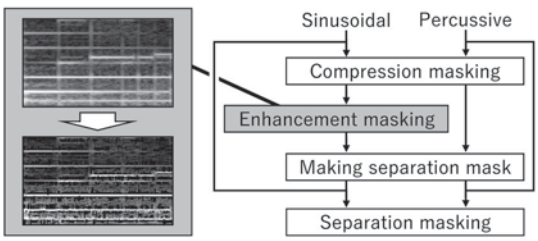


Fig. 1: Flowchart of the proposed method at each iteration. The grey box is the modification from the conventional method.

1-Q-2

1-Q-2　ディリクレ分布に基づく正則化付き非負値行列因子分解

Nonnegative matrix factorization with Dirichlet-distribution-based regularization

☆小川 遼, 北村 大地, 綾野 翔馬 (香川高専)

◆非負値行列因子分解（NMF）の基底ベクトルに**ディリクレ分布**の事前生成モデルを導入し，**補助関数法に基づく最適化法**を導出

◆基底ベクトルの**総和が１となる制約（ノルム制約）**を担保（Fig. 1）

◆ディリクレ分布の集中度母数（ハイパーパラメータ）を制御することで，**スパース正則化**と**スムース正則化**のどちらも表現可能

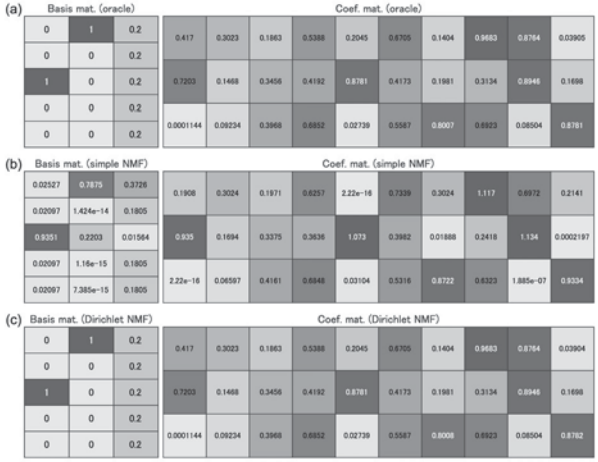


Fig.1 : (a) Oracle basis and coefficient matrices, (b) matrices estimated by simple NMF, and (c) matrices estimated by Dirichlet NMF.

1-Q-3

1-Q-3 非負値行列因子分解における主成分分析を用いた基底数推定の基礎検討

Fundamental study of number of bases estimation using principal component analysis in nonnegative matrix factorization

☆竹中遼河（広島市大），大島風雅，中山仁史（広島市大院）

◆現在の非負値行列因子分解（NMF）ではその基底数の設定が手動であり，音源の複雑さによっては設定が困難である

◆本研究では，主成分分析(PCA)によって得られる固有値と，その固有値をクラスタリングするk-means clustering を用いたNMF の基底数推定手法の提案を目的とする

◆実験から，NMF で求めた行列をPCA することで得た固有値を２群にクラスタリングすると，固有値が大きい群のクラスタ数が音源のピッチ数(最適な基底数)に合致することが明らかになった

Table 1　Number of pitches for each source

source no.	instrument	num. of pitches
no. 1	Piano	4
no. 2	Guitar	14
no. 3	Piano	4
no. 4	Guitar	8
no. 5	Piano	8

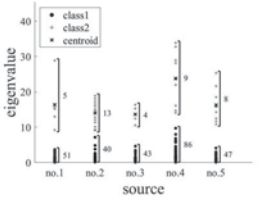


Fig. 1　The L_{max} eigenvalues of the activation vectors, and the results of k-means clustering.

1-Q-4

1-Q-4　単音節を対象とした非負値行列因子分解におけるアクティベーション行列の初期化

Initialization of activation matrix in nonnegative matrix factorization for monosyllables.

☆梨和美佑，大島風雅，中山仁史(広島市大院)，△田村雄一（国際医療福祉大）

◆ 音声認識や感情認識をベースとした音声による心不全診断技術「Voice-BNP」を提案.

◆Voice-BNP の高度化のため，モーラや音節を音素に分離し、音源や調音構造の違いから音響的特徴を抽出する方法を提案.

◆非負値行列因子分解（NMF）を用いて音声信号を周波数スペクトルとアクティベーションに分解し、単音節の音声分解を行う.

◆従来のNMF はスペクトログラムをより高く近似しようとするため、望む基底が得られない可能性がある.

◆アクティベーション行列の初期値として子音と母音の継続時間強度分布を与える.

◆従来法と比較しNMF におけるモデリングによって得られる解釈性が向上することを確認した.

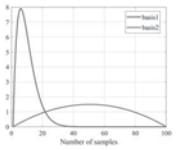


Fig.1 Duration time strength

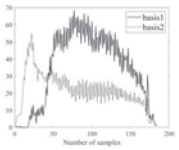


Fig.2 :Conventional

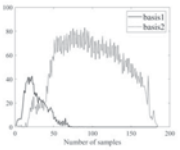


Fig.3: Proposed

1-Q-5

1-Q-5　雑音環境下における心拍音の第1ピーク間隔に同期した可視化手法

Visualization synchronized with the first peaks of heartbeat under a noisy environment.

☆松本眞太朗(阪産大院・工学研)，△塩安佳樹(はごろも内科・小児科)，△谷村旭律(阪産大)，高橋徹(阪産大院・工学研)

【背景】心拍音の可視化で高齢医師の診察支援したい．第１ピーク間隔に同期した心拍音可視化を開発してきたが，静寂環境の録音データで評価されてきた．診察室で実用的運用を始める前に雑音対策と評価が必要.

【目的】雑音対策処理の提案とその処理が可視化に与える影響を調査.

【手法】ブラインド音源分離手法(**FastICA**)で心拍音を分離抽出.

【評価】６種のノイズを４種の**SNR** で混合した音で評価.
心拍音を分離可視化しフロベニウスノルム(F-norm)を比較.

【結果】心拍音抽出処理は可視化に影響を与えず心疾患有無の判断にも影響がないことを確認できた.

Table.1：F-norm for noise levels and noise types.

	Normal heartbeat				Diseased heartbeat			
Noise type	Clean	0 dB	-5 dB	-10 dB	Clean	0 dB	-5 dB	-10 dB
aircondition	2.510	2.508	2.508	2.508	2.392	2.390	2.390	2.390
clicking	2.552	2.552	2.552	2.552	2.431	2.431	2.431	2.431
rubbing	2.452	2.437	2.437	2.437	2.295	2.294	2.294	2.294
typing	2.506	2.505	2.505	2.505	2.255	2.255	2.255	2.255
crumpling	3.288	3.288	3.288	3.288	2.408	2.409	2.409	2.409
water	2.558	2.558	2.558	2.558	2.271	2.271	2.271	2.271

1-Q-6

1-Q-6　放射モードに基づくマイクロホンアレイ：残響の抑制効果

Radiation mode-based microphone arrays: Dereverberation effect

☆奥石開生、貝塚勉(工学院大学)

◆遠くから届く音を拾いにくく、近くから届く音のみを拾いやすいマイクロホンアレイに関する研究。

◆ヘッドセットへの応用を想定し、3 個の MEMS マイクロホンから構成される小型のアレイを試作した。

◆部屋の壁や床からの反射音（残響）がマイクロホンアレイの遠くから届くことに注目し、近くにある話者の口元から届く音声（直接音）を計測しつつ、残響を抑制できることを示した。

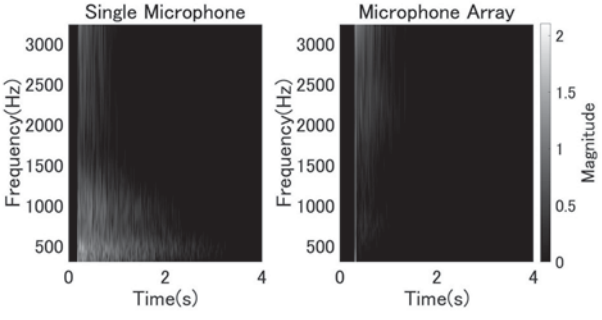


Fig.1：Scalogram of clap measured by single microphone and microphone array

1-Q-7

1-Q-7　雑音参照マイクを用いたマルチチャネル音声強調ネットワーク

Multi-channel Speech Enhancement Network Using Noise-Reference Microphone

○鈴木皓太，島村徹也，杉浦陽介（埼玉大・工）

◆大音量の騒音を抑制するため，雑音参照マイクを用いた**音声強調ネットワーク**を開発する。

◆雑音参照マイクを使用することで，**環境内の雑音情報をより積極的に取得**することが可能になる。

- 大音量の騒音環境下でもより鮮明に目的音声を判別可能に

◆従来の音声強調ネットワーク「DEMUCS」をマルチチャネルに対応させ，提案法として **Noise Fusion 構造**をもつネットワークを開発

◆従来法と提案法での比較実験を行い，機械学習における各エポックの損失の大きさ（Fig.1），また 10 エポック毎に計測した音質評価指標 PESQ の値のいずれにおいても提案法が従来法を上回る結果となった。

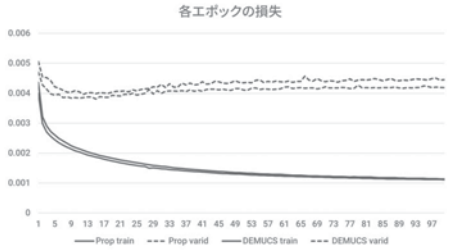


Fig.1：Learning curves in DEMUCS and proposed method

1-Q-8

1-Q-8 スマートフォンの加速度センサを用いた音声強調

Speech Enhancement for Mobile Devices Using Accelerometer Sensor

○LIN YIFENG，杉浦 陽介，島村 徹也（埼玉大院）

◆本研究では，DEMUCS をベースとし，スマートフォンの加速度センサ情報を活用することで音声強調性能をさらに向上させるマルチモーダル音声強調を提案する．具体的には，スマートフォンの加速度センサから得られた振動情報とマイクロホンから得られた音声信号をそれぞれ別のエンコーダに入力し，得られた特徴量を融合して新たな特徴量を生み出すことで，より高いロバスト性を持つ音声強調を目指す．効果的な特徴量の融合を実現するために Window-based Multi-head Cross Attention（W-MCA）を開発した．W-MCA は，Transformer に使われる Multi-head Self Attention (MSA)と比べて，データ計算量を大幅に削減しつつ，性能を向上できる．音声強調実験を通じて，STOI，PESQ の両方の指標で従来の DEMUCS モデルに比べて提案法の性能が優れていることを示す．

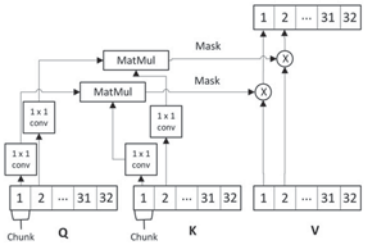


Fig.1：Window-based Multi-head Cross Attention

1-Q-9

1-Q-9　ゲーム実況音声からの BGM 抽出に向けた深層学習モデルの検討

Development of deep learning model for extracting BGM signal from game recording data

☆岩本 晃周，西村 竜一（和歌山大）

◆ゲームを記録した録音に含まれた効果音を抑圧し、BGM を抽出

◆深層学習による音源強調モデル Conv-TasNet と MRX の直列使用

- Conv-TasNet：処理した際に BGM の高音成分も減衰する傾向
- MRX：効果音の音圧は減少するが、スペクトルの形状が残る傾向

◆SI-SNR（音圧比）

(a) Conv-TasNet → MRX：平均 -3.0 dB（単体使用を含む 4 条件で最高値）、ばらつきが多い

(b) MRX → Conv-TasNet：平均 -4.3dB、効果が小さい

◆STOI（明瞭度）

(a) MRX → Conv-TasNet：MRX 単体から 1.1 の向上、Conv-TasNet 単体から 1.2 の低下

(b) Conv-TasNet → MRX：MRX 単体から 1.1 の向上、Conv-TasNet 単体から 1.2 の低下、(a)と比較して平均 0.7 の低下

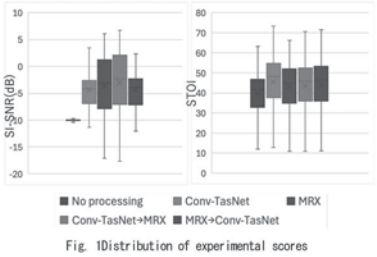


Fig. 1Distribution of experimental scores

1-Q-10

1-Q-10　珈琲焙煎における温度上昇と音響的変化についての基礎検討

Basic study on rate of rise and acoustic changes in coffee roasting

☆向井健悟（高松桜井高），大島風雅，中山仁史（広島市大院）

◆本論文では浅煎り焙煎を対象とした温度上昇と時間経過による周波数特性の関係を明らかにする．

◆周波数特性の計測はモデリング計測として erNMF を用いる．

◆erNMF（Euclidean Metric Regulated NMF）はアクティベーション行列と基底行列における強度の任意性を改善することができる．

◆温度上昇と音響的変化は常に一定ではなく，焙煎豆の状態によって得られる周波数特性が変化する．

◆焙煎終了までに 10dB 程強度が下がること，１ハゼ付近で強度の変化が大きいことを確認した．（Fig.1 及び Fig.2）

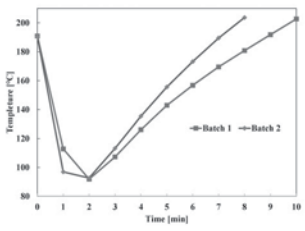


Fig.1: Temperature inside of coffee roaster

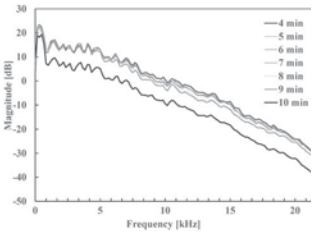


Fig.2: Frequency characteristics by roasting time

1-Q-11

1-Q-11

重複した帯域の逐次的な分離による優決定 BSS の高性能化

Improving determined BSS via sequential separation of overlapping subbands

◎ 松本和樹（早大），矢田部浩平（農工大）

研究背景：ブロックパーミュテーション問題

ある周波数を境に分離音の順序が入れ替わる現象

► 優決定 BSS 手法における**分離性能のボトルネック**

提案：重複した帯域に対し BSS 手法を逐次適用（Fig. 1）

■ 同時に分離する**帯域幅を狭める**ことで最適化問題の規模を削減 ► **分離の頑健性を向上させる**

■ ある帯域で得られた結果を次の帯域の**初期値として引き継ぐ** ► **帯域間のパーミュテーション**を揃える

► **IVA・ILRMAの分離性能を大幅に改善**し，収束に必要な**計算時間を削減**

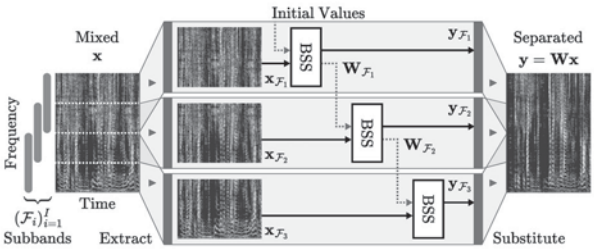


Fig. 1: Proposed technique named *subband splitting*

1-Q-12

1-Q-12 クラスベース目的音抽出における抽出スケール制御手法の検討

Investigation of Extraction Scale Control Methods for Class-Based Target Sound Extraction

◎佐藤僚，春田智穂，昼間信彦（リオン），井本桂右（同志社大）

◆目的音抽出：様々な種類の音源を含む混合音から、特定の音源を抽出

- 各音源の抽出スケールを独立に制御するには、各抽出対象クラスを個別に抽出してスケーリング・再混合するポスト処理が必要
- 計算量の増大や、誤抽出による品質低下が課題

◆各抽出対象クラスに対して目標抽出スケールを独立に制御可能とする学習手法 Target Scale Training (TaST) を提案

- クエリによって、各クラスの目標抽出スケールを条件付け
- 一度の推論で各クラスの抽出スケールを独立に制御を可能
- 従来法+ポスト処理と比較して、少ない計算量で優れた品質を達成

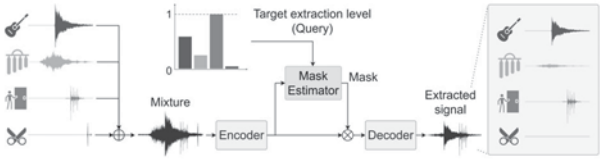


Fig.1: Overview of Target Scale Training.

Table 1: SNR and SI-SNR for three-class extraction.

Model	SNR	SI-SNR
Waveformer	6.64*	5.28*
+ TaST (Proposed)	7.17	6.12
+ TaST (Proposed) & w/o LN	**7.27**	**6.21**
pcTCN	6.42*	4.75*
+ TaST (Proposed)	**7.05**	**6.18**

1-Q-13

1-Q-13 自己教師あり学習による未知の音声データに対する音声強調性能の改善

Improving Speech Enhancement Performance for Unknown Speech Data Using Self-Supervised Learning

○半澤恭介，杉浦陽介，島村徹也（埼玉大）

◆未学習の音声に対する音声強調性能が低下するという課題を，Discriminator を導入することで雑音重畳音声からの教師なしの追加学習を行うモデルを提案する．

◆Discriminator では，音声をパッチ分割し，かつダウンサンプリングされた音声と並列に処理することで低周波数，局所的な特徴量を評価する Multi-scale Patch GAN を採用する．提案法を用いた追加学習の流れを Fig.1 に示す．

◆追加学習を通して得られた PESQ の推移を Fig.2 に示しており，音声品質の改善が確認できる．しかし，5 エポック目以降では PESQ が次第に悪化しており，過学習が発生していると推測される．

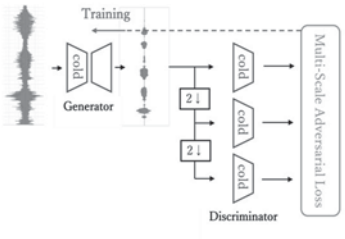


Fig. 1 Training stage 2 of the proposed method

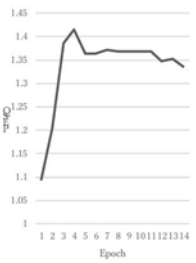


Fig. 2 Behavior of PESQ

1-Q-14

1-Q-14 制約付き補助関数法による非負値テンソル因子分解を用いたスポットフォーミングの高速化

Fast algorithm for spotforming using nonnegative tensor factorization based on constrained majorization-minimization

◎綾野翔馬（香川高専），李莉，関翔悟（サイバーエージェント），北村大地（香川高専）

◆アトラクタ正則化付き非負値テンソル因子分解

- 3 次元テンソルを 3 つの行列の積で近似
- それぞれの基底ベクトルを自動的にクラスタリング
  - ✧ 全てのチャネルに含まれる共通成分
  - ✧ 単一のチャネルにのみ含まれる固有成分

◆提案手法

- 補助関数法の設計において，各変数行列の制約を含めた制約付き補助関数を設計，Lagrange の未定乗数法を用いて求解
- 正則化項のハイパーパラメータを∞とすることで更に高速化
  - ✧ 加えてスポットフォーミングの性能も向上

◆計算時間比較

- 制約付き補助関数に基づくアルゴリズムは通常のアルゴリズムよりやや高速に動作
- 正則化項のハイパーパラメータに∞を用いることで更に高速に動作

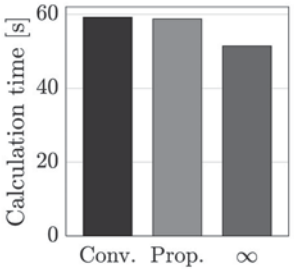


Fig. 1 Calculation time of each method.

1-Q-15

1-Q-15

凸最適化に基づく劣決定音源分離における辞書行列の適応的更新

Adaptive updating of dictionary matrix
for underdetermined source separation based on convex optimization

☆ 皆川朋樹, 矢田部浩平（農工大）

劣決定音源分離問題

音源数よりも少ない数の観測信号を分離する

従来法　空間モデルを事前に定義し，分離後にエネルギーの高い方向を推定DOAと推定音源とする
推定DOA以外の方向に分離された信号を無視する

▶ エネルギーの漏出，音質劣化の恐れがある

提案法　空間辞書の行列を適応的に更新し，推定DOAの候補を絞り込む
推定DOAのみに信号を分離する

▶ エネルギーの漏出を防ぎ，分離性能が向上した

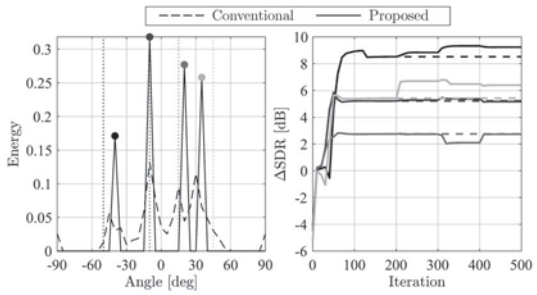


Fig. 1 Comparison of estimated source energy and separation performance per iteration

1-Q-16

1-Q-16 小型マイクロホンアレイと時空間 2 次元スペクトルを利用した周方向の音源分離に関する考察

Consideration on circumferential source separation using a small microphone array and spatiotemporal 2-D spectra

△有泉千太, 鳥谷輝樹(山梨大・工), 渡邉貫治（秋田県立大・システム科技),
坂本修一（東北大・通研), ○小澤賢司（山梨大・工）

◆スマートフォンに搭載できる小規模マイクロホンアレイ（マイクロホン数：8, 全長：14 cm）により, 周方向の複数音源を分離に挑んだ。

◆アレイからの出力を画像と見なし, 2 次元 (2D) 離散フーリエ変換を施して時空間 2D スペクトルを得た。その 2D スペクトルの特徴から, 空間の直流成分周辺の 5 空間周波数ビンのみが有効と考えた。そして, 複素スペクトルについての連立方程式を解くことで 5 音源を分離した。

◆計算機シミュレーション実験により, Fig. 1 のような到来角（DOA: Direction of arrival）の 5 音源について, 分離された信号の SDR（信号対歪）を Fig. 2 に示す。

➢ アレイ正面（D0）については, 開き角φに依らず高 SDR で分離される。

➢ 外側の DOA ほど, φの増加とともに SDR が低下する。これは, DOA が増加すると, 空間の高周波成分が優勢となり, 連立方程式の求解精度が低下するためと考察した。

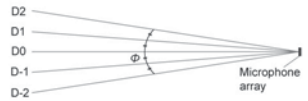


Fig. 1: Definition of DOAs and spacing.

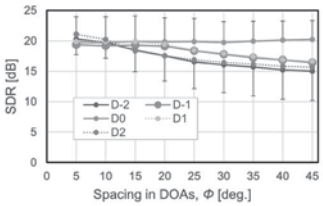


Fig. 2: SDRs of separated five sounds.

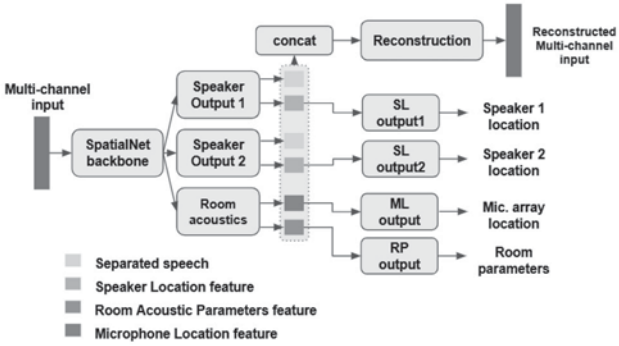
1-Q-17

1-Q-17　Multitask Training of Multi-channel Speaker Separation and Room Acoustic Parameter Estimation

☆ Roland Hartanto (Science Tokyo), Sakriani Sakti (NAIST),
Koichi Shinoda (Science Tokyo)

◆We propose a **multitask learning** of **speaker separation** and **room acoustics parameters estimation**, further leveraging room acoustic parameters information to **improve multi-channel separation in various acoustic conditions.**

◆Evaluated on the SMS-WSJ Plus dataset, our method performs slightly worse than Permutation Invariant Training (PIT) in SI-SDR by 0.4 points, while achieving better performance in terms of Word Error Rate (WER) by 0.67 points.



1-Q-18

1-Q-18　言語クエリを用いた音源分離の多チャンネル拡張の検討

Multi-channel extension of language-queried audio source separation

◎中村優希, 中嶋大志, 小野順貴(都立大)

◆様々な種類の音源が含まれた混合音から自然言語記述で任意の音源を分離する，言語クエリを用いた音源分離 (LASS) が提案されている．従来モデルの多くはシングルチャンネルの入力を仮定しており，ブラインド音源分離 (BSS) のように空間情報を考慮できない．

◆本稿では，様々な種類の音源が複数のマイクロホンで観測された場合に，混合信号に対して空間情報と言語クエリ情報の 2 種類の情報を用いて音源を分離する手法を提案する (Fig. 1 (c))．具体的には，AuxIVA の音源モデルとして LASS の分離信号を用いる．

◆シミュレーション実験によって，LASS の分離信号が AuxIVA の音源モデルとして有効であることを示した．

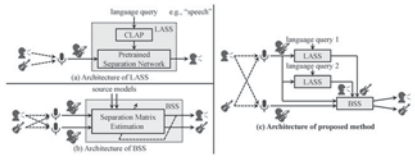


Fig. 1: Block diagram of conventional and proposed methods.

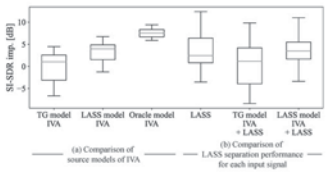


Fig. 2: Results of Experiments. TG: Time-varying Gaussian.

1-Q-19

1-Q-19　音声分離を用いた顎顔面形態の違いによる咬合状態を対象とした音響特徴分析

Acoustic feature analysis of occlusal conditions due to differences in maxillofacial morphology using speech separation

☆藤村秀弥，大島風雅（広島市大院），△村田聡，△堀畑聡，△石井かおり，△武藤佑子，△根岸慎一（日大・松戸），中山仁史（広島市大院）

◆不正咬合とは顎顔面骨格及び歯が形態的に異常をきたしている咬合状態のことである.

◆不正咬合の治療目的は咬合の改善とそれによる口腔機能の改善並びに不定愁訴のトラブルを解消することである.

◆本研究では不正咬合等が要因となる器質性構音障害の判別を非侵襲的かつ迅速に判断可能な音響分析による評価を提案する.

◆より詳細な音声分析を実現するために，NMF（Nonnegative matrix factorization）による音声分離を試みる.

	/ta/	/t/$_{BSS}$	/a/$_{BSS}$
MFCC1	**	**	**
MFCC2	ns	ns	ns
MFCC3	*	*	ns
MFCC4	ns	*	ns
MFCC5	*	ns	*
MFCC6	ns	ns	ns
MFCC7	*	ns	*
MFCC8	ns	ns	ns
MFCC9	ns	*	ns
MFCC10	ns	ns	ns
MFCC11	ns	ns	ns
MFCC12	ns	ns	ns
MFCC13	**	ns	**
MFCC14	ns	*	ns
MFCC15	ns	ns	ns
MFCC16	ns	ns	ns
MFCC17	ns	ns	ns
MFCC18	ns	ns	ns
MFCC19	ns	**	ns
MFCC20	ns	ns	ns

ns: No significant, *: $P < 0.05$, **: $P < 0.01$

Table 1: One-way ANOVA of MFCCs in /ta/, /t/$_{BSS}$ and /a/$_{BSS}$.

	/ta/		/t/$_{BSS}$		/a/$_{BSS}$	
	Open	Deep	Open	Deep	Open	Deep
MFCC1	**	**	ns	**	ns	ns
MFCC2	ns	ns	ns	ns	ns	ns
MFCC3	ns	*	**	ns	ns	ns
MFCC4	ns	ns	ns	ns	ns	ns
MFCC5	*	ns	ns	ns	*	ns
MFCC6	ns	ns	ns	ns	ns	ns
MFCC7	ns	**	ns	ns	ns	*
MFCC8	ns	ns	ns	ns	ns	ns
MFCC9	ns	ns	*	ns	ns	ns
MFCC10	ns	ns	ns	ns	ns	ns
MFCC11	ns	ns	ns	ns	ns	ns
MFCC12	ns	ns	ns	ns	ns	ns
MFCC13	**	ns	ns	ns	**	ns
MFCC14	ns	ns	*	ns	ns	ns
MFCC15	ns	ns	ns	ns	ns	ns
MFCC16	ns	ns	ns	ns	ns	ns
MFCC17	ns	ns	ns	ns	ns	ns
MFCC18	ns	ns	ns	ns	ns	ns
MFCC19	ns	**	ns	**	ns	ns
MFCC20	ns	ns	ns	ns	ns	ns

ns: No significant, *: $P < 0.05$, **: $P < 0.01$

Table 2: T-tests of MFCCs in /ta/, /t/$_{BSS}$ and /a/$_{BSS}$.

1-Q-20

1-Q-20　コード進行の近接パターン選択とメロディの自動生成に基づく歯科治療音のリアルタイム快音化システム

Real-time comfortable sound design system based on proximity pattern selection and automatic melody generation for chord progression

☆林拓哉(阪産大院)，高橋徹(阪産大)，西浦敬信(立命館大)，中山雅人(阪産大)

◆従来の歯科治療音の快音化手法では，単調なコード進行となり快音化性能が不足する問題があった．これは，コード進行パターンが１種類のみであったためだと考えた．そこで，本研究ではコード進行の近接パターン選択とメロディの自動生成に基づく歯科治療音のリアルタイム快音化システムを提案する.

◆Fig.１に提案手法の概要図を示す．提案手法では，歯科治療音のピーク周波数系列と複数のコード進行パターンを動的計画法によるパターンマッチングを行い，近似するコード進行パターンと歯科治療音の音高を用いてコード進行音とメロディ音を生成し，制御音として歯科治療音と同時に受聴することで快音化を行う.

Fig. ２に快音度の主観評価実験の結果を示す．３名による評価の結果より提案手法（**Proposed Method**）では，従来手法（**Conventional Method**）よりも **0.6** ポイント快音度が向上していることが確認できた.

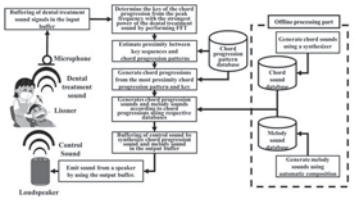


Fig.1 Overview of the proposed method

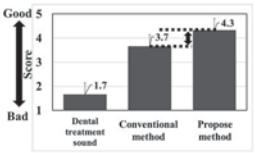


Fig.2 Mean opinion score with five grades for comfortable

1-Q-21

1-Q-21　CQT-diffusion のモデル再学習に基づくビジュアルマイクロホンの音質改善

Audio quality improvement of visual microphones based on retraining the CQT-diffusion model.

☆大野圭哉，立蔵洋介(静岡大院・総合科学技術研)

◆背景：カメラから音を抽出するビジュアルマイクロホンは抽出原理上周期的な音波の欠損が生じ，復元音の音質を低下させる.

◆目的：Audio Inpainting モデルをビジュアルマイクロホン信号に適応させ，ビジュアルマイクロホンの抽出音の欠損区間の補間を実施し，音質が改善されるか調査する.

◆方法：ビジュアルマイクロホンの抽出信号を想定し，データセットに帯域制限を行いCQT-diffusion に再学習させた後に補間信号を生成する.

◆結果：五段階評価による主観評価実験から補間による聴感上の音質の改善を確認した．しかし，既存モデルと比較した際の再学習による音質の改善は見られなかった.

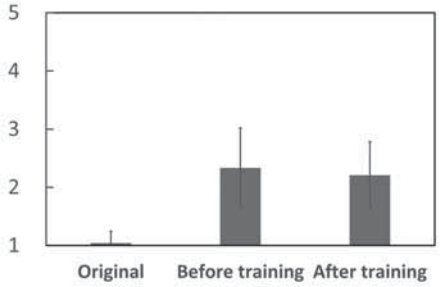


Fig.1 Comparison of subjective evaluation.

1-Q-26

1-Q-26　上下に分離した2次元スピーカーアレイを用いたスポット再生に関する検討

Study on Spot reproduction using vertically separated two-dimensional loudspeaker arrays

☆北山日向(龍谷大院・理工学研)，片岡章俊(龍谷大・先端理工)

◆多点制御法を用いて空間内の特定のスポットにのみ音を届けるスポット再生を実現するため，スピーカ配置の検討を行った.

◆４種類のスピーカ配置で計算機シミュレーションを行い，制御性能を比較した.

◆スポット再生には，２次元スピーカーアレイを上下に分離する配置が適していることが確認できた.

◆再生スポットを移動する場合，再生スポットの位置によって使用するスピーカを使い分けることも有効であることがわかった.

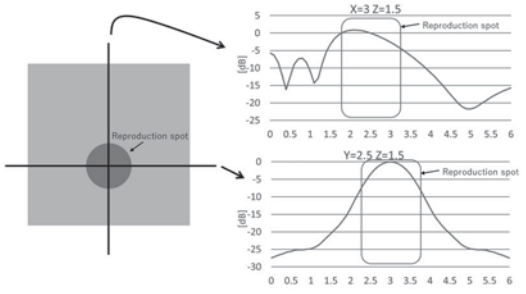


Fig.1: Sound Pressure Level (Spot reproduction using vertically separated two-dimensional loudspeaker arrays)

1-Q-27

1-Q-27　低域指向性・局所増音を実現する 3ch 音場制御技術の開発

Development of sound field control technology

○江波戸明彦，蛭間貴博，西村修（東芝）

◆生活環境において情報伝達手段として音の活用が進む一方で，音は騒音にもなり，パーソナル空間だけに音が伝わる技術が望まれる。

◆特定方向に局所増音を実現するスピーカ再生制御技術を開発した。音響パワー低減化と増音化を両立する制御則の提案により，スピーカ 3 個で急峻な増音勾配・低域指向性が実現する（Fig.1）。

◆スピーカの向きやパワー制御則・増音制御点位置の変更により，机上設置による局所増音・スポット増音・複数音声同時再生・低域指向性再生など様々な空間音場を創出できることを実験で確認した（Fig.2）。

◆本技術の一部は東芝デジタルソリューションズ(株)の音響ソフトウェア Soundimension ™ 音場制御に活用されている。

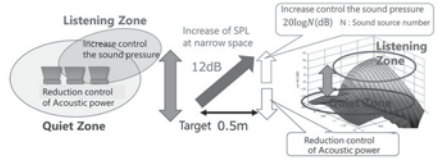


Fig.1　Basic principle

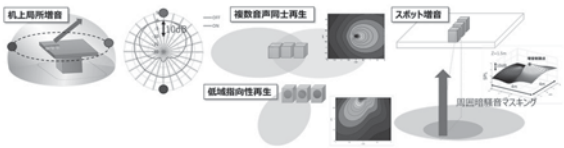


Fig.2　Reproduction scene of the sound field control

1-Q-28

1-Q-28　エリア再生による音漏れを考慮した音場制御を用いたマスキング手法に関する検討

A Study on a Masking Method Using Sound Field Control Considering Sound Leakage in Area Reproduction

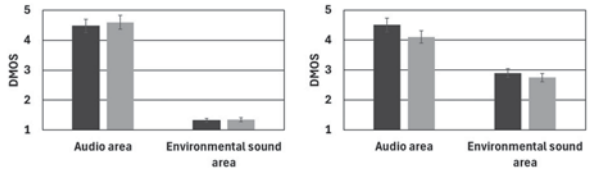
☆大橋颯士郎，周桐(龍谷大院)，安枝和哉(東京医療保健大)，
片岡章俊(龍谷大)

◆音場制御技術により特定のエリアのみ音を再生するエリア再生の研究が行われている.

◆先行研究で多点制御法(PM)を用い，再生領域内の音圧を均一化した音場を生成する研究を行った．均一化は実現できたが，抑圧性能が低下し抑圧領域に音声内容が漏洩することが挙げられる.

◆本研究では環境音によるマスキング音を抑圧領域に再生する手法として，多点制御法と音響コントラスト制御法(ACC)を比較する．2 つの手法について、環境音の再生領域と抑圧領域における音圧差の性能評価を行い、DMOS 評価法を用いて主観評価実験を行った.

◆実環境での性能評価の結果，音圧差は PM が 20.04dB，ACC が 25.2dB であった．主観評価実験結果，Pink 雑音や White 雑音では，PM に比べ ACC は不快さが減少していることが確認できた.



(a) Difficult of speech listening　　(b) Harshness

Fig.1 : White noise experiment results.
(■:ACC, ■:PM)

1-Q-29

1-Q-29　近接 4 点法のデータに基づく平面波合成を用いた音場生成

Sound field creation using plane wave synthesis based on data from closely located four point microphone method

☆江里口裕月，坂口智弘，及川靖広（早大理工）

◆背景

従来の音場再現技術は直接音と反射音の分離が困難

反射音一つ一つを厳密に再現できていなかったりする

◆提案手法

直接音や反射音を一つ一つ分離した空間情報を用いる

→反射音に対しても正確な音場合成が可能となる

仮想音源から到達する反射音は平面波として到来すると仮定

→平面波を合成によって音場を生成する

◆結果

MATLAB を用いて生成した波面の時間変化を確認した（Fig.1）

生成したインパルス応答と実測のインパルス応答の比較を行った

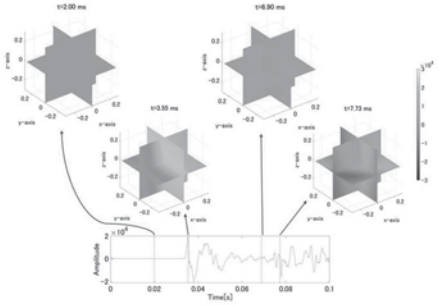


Fig.1 : Time variation of wave front

1-Q-30

1-Q-30　円調和展開における最大次数と対象音場半径と空間誤差の関係について

On the relationship between maximum order, target radius and spatial error in circular harmonic expansion

◎任逸，羽田陽一（電通大）

◆円調和関数とベッセル関数を用いた 2 次元音場展開において，数値計算上有限な次数で打ち切らなければならないため，多くの研究では経験則で最大次数 N の打ち切りを行っている。

◆本研究では，最大次数・周波数・対象領域範囲・空間平均誤差の関係性について，従来研究のセオリーと N= |kr| や N= |ekr/2| を含めた次数打ち切りの経験則を整理する。

◆また，最大次数と周波数を決めたときに，誤差がおおよそ目標値となるような領域半径を誤差曲線から探索する手法と多項式近似の解より予測する手法を提案する。

◆音場再現の計算機シミュレーションにて提案手法を検証し，探索した半径において誤差が概ね設定した目標値と一致することや，多項式から求めた半径において誤差が目標値から大きく離れる場合もあることがわかった。

計算方法	半径 r [m]	空間誤差 (r 内)	空間誤差 (r 上)
$N=15,\ \bar{\varepsilon}_N(r)=-30\,\mathrm{dB}$			
$r=N/k$	0.81	-26.40	-16.36
$r=2N/ek$	0.60	-56.22	-46.23
曲線探索	0.68	-42.91	-31.05
多項式求解	0.58	-58.70	-47.22

Table.1 : Predicted radius and actual reproduction error.

1-Q-31

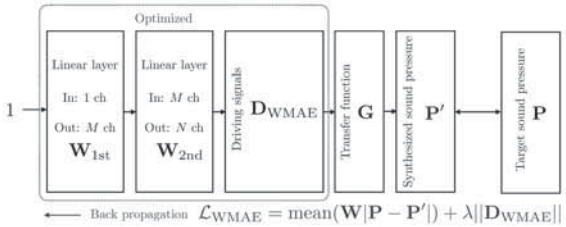
1-Q-31　誤差逆伝搬に基づく音場制御

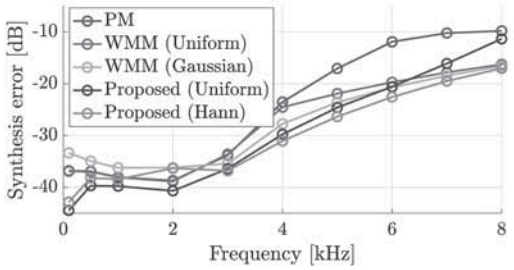
Sound field synthesis using backpropagation

○岡本拓磨 (情報通信研究機構)

「こちらオカモト，誤差逆伝搬を用いた音場制御ができたぞ，Over!!」
『Good job オカモトさん，欲しかったのは，それだけです!!』

・誤差逆伝播に基づく重み付き音圧平均絶対誤差最小化を用いた音圧制御型音場制御方式を提案
・計算機シミュレーションにおいて提案法の有効性を確認





1-Q-32

1-Q-32　偏光高速度干渉計によるインパルス応答の高分解能イメージング

High-resolution imaging of impulse response by high-speed polarization interferometer

☆齋藤結月(早大理工)，石川憲治(NTT)，
谷川理佐子(NTT/早大理工)，及川靖広(早大理工)

◆偏光高速度干渉計：時間・空間的にも高い分解能で音場を計測できる。音響機器の放射特性の評価への応用が期待される。
◆研究目的：スピーカのインパルス応答の測定に適用する。
◆提案手法：
- 偏光高速度干渉計を用いて複数方向から TSP 信号を測定。
- 取得した音場データから二次元のインパルス応答を算出。
- 物理モデルに基づく三次元インパルス応答の可視化。

◆結果：インパルス応答の空間的な挙動を詳細に観察することができた。
◆課題：SNR の向上，インパルス応答の分布の分析。

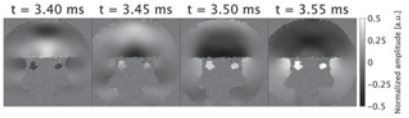


Fig.1：Visualization results of 2D impulse response. The figure shows the variation of the sound field over time from left to right.

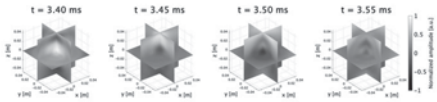


Fig.2：Visualization results of 3D impulse response. The figure shows the variation of the sound field over time from left to right.

1-Q-33

1-Q-33　非同期録音信号を用いた広範囲の定常音場推定の基礎的検討

Basic study on estimation of steady and broad sound field using asynchronous measurements

◎内田彩芽，津國和泉，池田雄介(東京電機大)，及川靖広(早大理工)

◆ 背景
- 近年、マイクロホンアレイ移動による非同期録音信号を用いた騒音源を対象とする音場推定手法が提案
- マイクロホンアレイの移動間隔が大きくなると推定精度が低下するため，依然として測定の労力削減は困難

◆ 提案
- 非同期録音信号とスパース等価音源法を用いた定常音場推定手法の基礎的検討を実施
- 騒音源位置や指向性に変化がないことを仮定し、音場をモデル化

◆ シミュレーション実験
- 2 回の移動測定で得られた非同期信号から直接音場を推定
- 提案手法により 2000 Hz において推定領域右側 (-2.0＜x＜-0.9 m) が SNR 平均 1.3 dB 改善

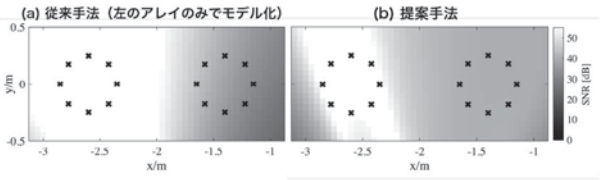


Fig. 1: SNR distribution at 2000 Hz
(The cross mark represents the microphone position.)

1-Q-34

1-Q-34　自己相関制約付き NMF による時間的周期性を持つ反射音の抽出

Extraction of reflections with temporal periodicity using autocorrelation constraint NMF

◎泉悠斗，大谷真(京都大院・工学研)

◆時間的周期性を持つ反射音成分の抽出を目的として，時間方向に自己相関制約を組み込んだ Nonnegative Matrix Factorization（NMF）を用いて反射音パターンを分析する.
◆幾何音響シミュレーションを用いた数値実験により，目標の反射音成分を分離できることを確認した（Fig. 1）.

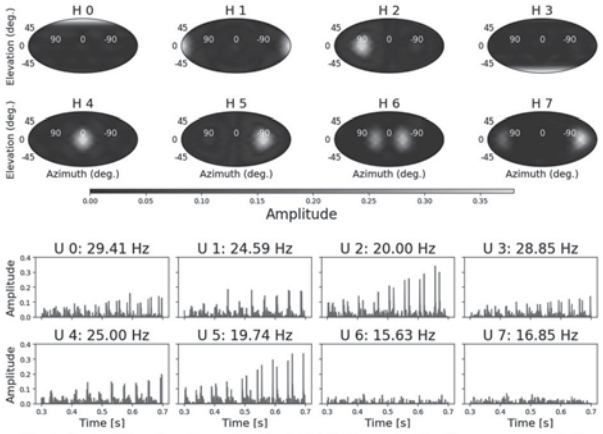


Fig.1：Part of the direction-of-arrival distribution of reflections represented by the basis H (top) and its temporal variation represented by the activation U (bottom), decomposed using autocorrelation constraint NMF.

1-Q-35

1-Q-35　室内 3D モデルの学習データ単純化が MESH2IR の室内インパルス応答出力に与える影響

Effects of training data simplification of indoor 3D models on room impulse response outputs in MESH2IR

☆伊藤陸人，立蔵洋介（静岡大院・総合科学技術研）

◆深層学習に基づく室内インパルス応答（以下，RIR）推定器 MESH2IR に，直方体形状の 3D モデルを学習させ，出力された RIR の特性を調査した．

◆出力された RIR の波形と，RIR を用いたマルチチャネル逆フィルタによる音場再現の結果から，初期反射部分については概ね精度よく推定できていることが確認された．

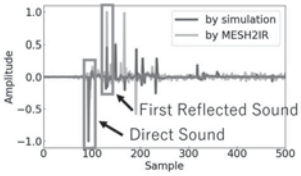


Fig.1：Waveforms of RIR output by MESH2IR and RIR generated by simulation

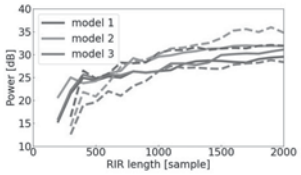


Fig.2：Changes in observed signal energy when varying the RIR signal length used for sound field reproduction

1-Q-36

1-Q-36　複数伝達経路の同時計測および信頼性のある推定値が得られる IR 推定手法（IRMOSs）の提案 —4 直交信号を用いた 2 音源 2 観測点間（IRMOSs-4）—

IR estimation method (IRMOSs-4) for simultaneous measurement between two sources and two observation points using four quadrature signals and reliable estimation of multiple transmission lines

☆徳富 響(NBU)，沖田 和久(NBU 卒(現 NJMC))，近藤 善隆(J-Tec)，
福島 学(NBU)，△松本 光雄()，柳川 博文(arsl)

◆目的：他音源・他観測点の同時計測

◆前報（音講論集 2024 年秋季 1-R-42）1 回の計測で次の 2 計測を実現
1 音源　1 観測点　：　精度計測（Fig. 1 左）（Fig. 2）
1 音源　2 観測点　：　同時計測（Fig. 1 左）

◆本報：多音源・他観測点の同時計測に拡張（Fig. 1 右）（Fig. 2）

◆検証：2 音源・2 観測点の同時計測

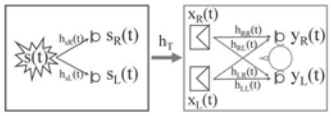


Fig.1 Signal propagation in the measurement and reproduction of acoustic phenomena

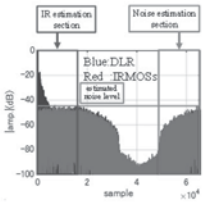


Fig.2 Comparison of noise level estimation in DLR and IRMOSs, (Blue: DLR, Red: IRMOSs)

1-Q-37

1-Q-37　開放型ヘッドホンとスピーカを併用した音像定位の検討

Study of sound source localization using open headphones and loudspeakers

◎今泉健太，△宮川和(NTT)

◆本研究では，耳を塞がない開放型ヘッドホンと受聴者前方に配置したスピーカを併用した距離提示が可能な音像定位システムを提案する．

◆定位させたい音源の直接音に相当する音声をヘッドホンから再生し，間接音に相当する音声をスピーカから再生し距離感の制御を行った．

◆被験者実験による主観評価を行い，提案手法により距離感の提示が可能であることを確認した

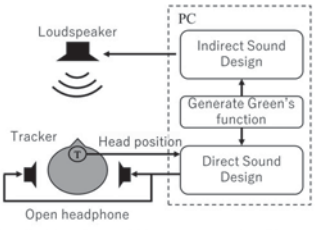


Fig.1：Structure of the proposed method

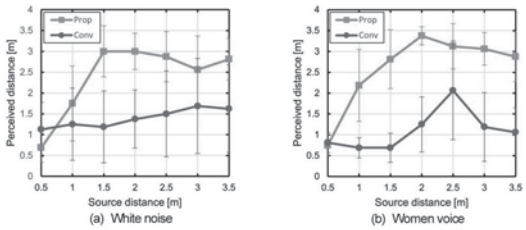


Fig.2：Results of subject experiment

1-Q-38

1-Q-38　グラフ型データベースを用いた HRTF パターンの個人適用選択手法の簡便化

Use a graph-type database to simplify the process of selecting individual HRTF patterns

○ 小塚詩穂里，伊藤弘章，鎌土記良(NTT)

◆ 人が音の方向を知覚するには個人に合わせた頭部伝達関数(HRTF)が必要

◆HRTF を計測によって得られない状況では他人の HRTF から本人に適合するものを選択

◆ 多数の中から選択するのは試聴回数の問題もあり困難

◆ データベースから HRTF データのグラフを構築し、階層的なクラスタリングとデータの類似関係を可視化する手法を提案

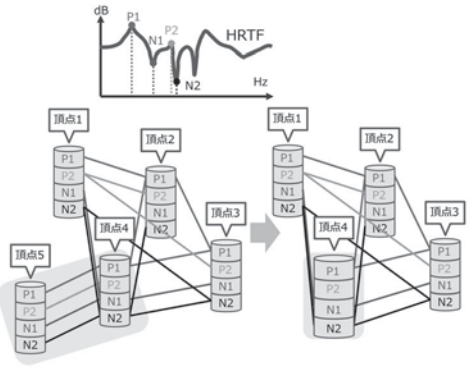


Fig.1 頂点にピークとノッチを格納するイメージと閾値以下の距離の頂点間を色分けした辺で結ぶイメージ

1-Q-39

1-Q-39　ヘッドフォン VR サウンドの再生において頭内定位を生じる要因に関する研究 —絶対判断による距離推定—

A study on the factors that cause for the lateralization in headphone-based VR sound reproduction - distance estimation by absolute judgement -.

○宇野慎太郎，工藤彰洋（苫小牧高専），武居周，坂本眞人（宮大工）

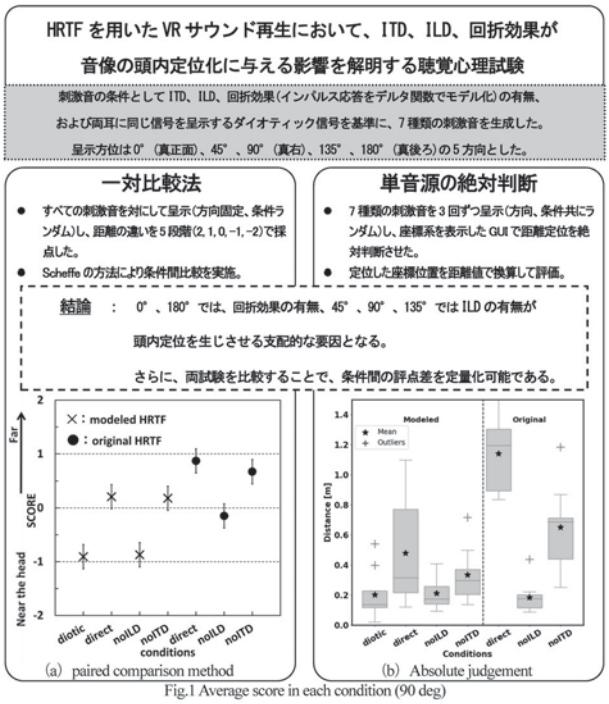


Fig.1 Average score in each condition (90 deg)

1-Q-40

1-Q-40　イヤセンタリングを適用した近距離頭部伝達関数合成における測定点数の違いが及ぼす影響

The effect of the number of head-related transfer functions on the accuracy of near-field HRTFs synthesized using ear-centering-based distance-varying filters

☆北村航太，坂本修一（東北大通研/院情科研）

◆本研究では距離変換フィルタ（DVF）による近距離 HRTF 合成法に対し，HRTF の次数下げの効果があるイヤセンタリングを適用し，合成元 HRTF の測定点数を変化させたときの合成精度の評価を行った。

◆イヤセンタリングには HRTF の測定点について，頭部中心と耳のそれぞれの位置からの見込み角を対応させ再ラベリングを行う角度補正を行った。

◆測定点数が 12，18，36 点の場合で合成精度の比較を行った結果，全ての条件において，角度補正イヤセンタリングを適用し，合成した場合の誤差が最も少なく，測定点数の変化に堅牢であった。

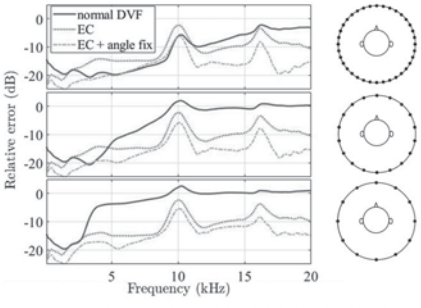


Fig.1：RMSE between synthesized HRTFs and original HRTFs (the number of HRTFs Top : 36, Middle : 18, Bottom : 12)

1-R-1

1-R-1　音韻レベルの話者情報を用いた音声認識における話者適応

Speaker adaptation in speech recognition using phonological level speaker information

☆伊藤光一，篠田浩一（東京科学大）

◆近年の深層学習ベースの音声認識は、モデルとデータの大規模化に伴い高い精度を記録するようになったが、対雑音下や複数話者条件下などで課題が残り、話者適応が重要。

◆従来は話者適応における話者情報の利用では発話平均が利用されてきたが、**話者の違いは音韻レベルにも**現れる。

◆本研究では話者情報の利用に関して、**音韻レベルの細かい単位を用いたマルチタスク学習手法**を提案する。話者情報の利用方法について複数の手法を比較検討するため実験を行った。

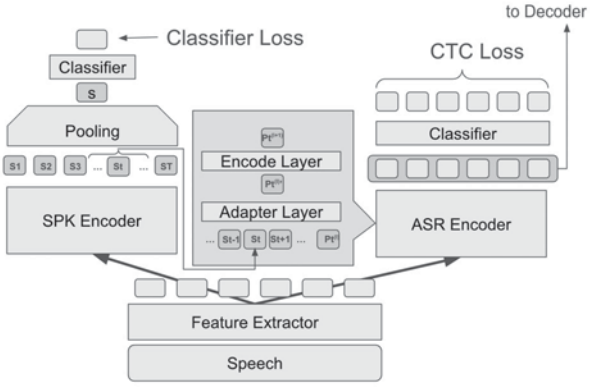


Fig.1：Proposed method overview

1-R-2

1-R-2　雑音環境下でのリアルタイム VAD レス音声認識モデルの構築と他モデルとの比較

Development of a real-time VAD-less speech recognition model in noisy environments and comparison with other models

☆江本城太郎，西村良太（徳島大），太田健吾（阿南高専），北岡教英（豊橋技科大）

◆本研究では，非発話区間の認識が可能である VAD ライクな挙動を含んだ音声認識モデルの構築を行う．

◆２音声の間と末尾に非発話区間を挿入し，雑音を重畳することで学習データを作成する．

◆非発話区間には“[雑音]”，“[無音]”の棄却タグを挿入する．

◆事前学習済み wav2vec 2.0 Base モデルを使用し，CTC によって 20 ms 毎に音声認識結果を取得可能なモデルを構築する．

◆VAD 併用時や他モデルとの精度や動作速度の比較を行う．

◆提案手法によるデータで構築されたモデルは VAD 併用時や whisper-base と比較して高精度かつ高速であった．

◆事前学習時に使用された音声の特性により，CTC におけるブランクトークンに棄却タグの役割が分散する可能性がある．

Table 1 Comparison of model performance including non-speaking interbals

		clean	20	15	10	5	0	RTF
whisper-base		31.58	28.43	31.04	32.40	46.29	93.71	0.0191
large-v3-turbo		**8.07**	**7.54**	**8.14**	**9.53**	**13.59**	**27.19**	0.0220
rinna	np	21.88	22.08	22.87	24.64	27.79	37.64	**0.0023**
	text_c	28.16	27.53	27.46	28.82	30.96	38.10	0.0054
	seg_c	33.40	32.36	33.15	33.34	35.27	41.29	0.0065
	ntag	17.81	18.02	18.53	20.10	23.54	33.86	**0.0023**
	proposed	18.36	20.16	19.62	20.14	23.40	32.00	**0.0023**
facebook	np	31.68	33.59	33.57	35.89	41.13	53.85	N/A
	text_c	35.96	35.96	36.89	38.89	42.63	53.46	
	seg_c	36.62	36.11	37.38	40.02	45.39	56.01	
	ntag	25.73	26.59	27.47	30.21	36.52	51.06	
	proposed	24.19	24.87	26.15	28.51	35.26	49.52	

1-R-3

1-R-3　多人数がいる環境での音声認識に適した話者分離技術の開発検討

Development of speaker separation technology suitable for speech recognition in environments with multiple people

○宮本正成，△浅谷尚希（パナソニックコネクト株式会社）

◆背景：LLM の進化に伴い、音声認識技術が普及しているが、そのシーンは限定的である。特にオープンなシーンでは、対象者以外の声が要因にマイクに混入するため、実用化には至っていない。

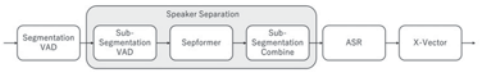
◆課題：音声認識の前処理に話者分離を検討する。分離モデルを独自で開発し、音声認識アプリへの組み込みを実施する。

◆結果１：データ拡張と多段 FT で実環境性能の高いモデルを開発した。

◆結果２：アプリに組み込み、音声認識精度の改善効果を確認した。

SI-SNR evaluation of learning models for each evaluation set

	Eval 1-3: Libri2Mix (Changing the volume ratio when mixing)			Eval 4-6: Libri2Mix (Overlap when switching speakers)			Eval 7-9: Libri2Mix (Add Noise)		
	0dB	6dB	12dB	0%	25%	50%	0dB	6dB	12dB
Baseline	19.25	19.35	18.93	33.06	26.49	22.65	5.42	8.43	12.29
Proposed	19.51	19.42	18.67	29.92	26.22	22.82	14.35	15.18	16.59



Processing pipeline for transcription app

WER for each evaluation set

	Dataset Clean	Eval dataset Add noise (SN6 dB)
Original	0.514	0.614
Proposed	0.0447	0.181

1-R-4

1-R-4　音声認識における音声コーパスに含まれる不良データ検出に関する考察

Mislabeld data detection in speech corpus for automatic speech recognition

○安井顕誠（Poetics Inc.）

❖目的
- ➢ 音声認識モデル学習を目的として構築された半自動で構築された音声コーパスに存在する音声とテキスト間の対応が不一致なデータが音声認識モデルの性能に影響するかを調べた．

❖実験
- ➢ 音声とテキスト間の対応が不一致なデータが含まれうる音声コーパスを使用して音声認識モデルを学習し，その出力結果とコーパスのテキストとを比較して，不一致が発生しているデータの候補を抽出した．
- ➢ 抽出した候補を音声コーパスから除外し，再学習した際の音声認識モデルの精度比較を行った．

❖結果
- ➢ 不一致が発生しているデータの候補として4つの類型を発見した．
- ➢ 抽出した候補を音声コーパスから除外し，再学習を行ったが，元の音声認識モデルよりWERが高い結果となった．

Table 1　データ削除前後の CER

	RS test CER	JSUT-book CER
データ削除前	7.29	27.07
データ削除後	8.31	27.66

1-R-5

1-R-5　自己回帰型 Early Exit 音声認識モデルのためのブロックリファイメント学習

Block refinement learning for autoregressive early exit speech recognition models

◎河田尚孝，折橋翔太，鈴木聡志，田中智大，庵愛，牧島直輝，山根大河，増村亮（NTT）

◆本発表では，推論高速化手法である Early Exit を用いることで，自己回帰型音声認識モデルの高速化について扱う

◆Early Exit では，図１の左側に示すモデルを用いる．このモデルからは各中間の出力層から推論結果が得られるため，浅い層で推論を適切に中断することで計算量削減⇒高速化を実現する

◆従来の Early Exit モデルの学習では，深い層の精度を維持するための学習重みが設定されている．しかし、浅い層の学習が軽視されることで十分な高速化効果が発揮できないという課題がある

◆提案手法では，浅い層の精度向上と深い層の精度維持を同時に行う再学習手法を導入することでこの課題を解決し，既存モデルの高速化が実現できることを示した

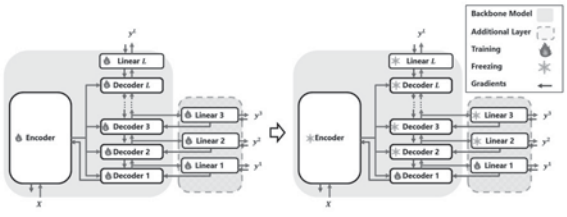


Fig.1：Overview of our proposed method. Left image shows conventional training of Early Exit. We focus on speeding up trained Early Exit models.

1-R-6

1-R-6　マスク内外同時サーチによる精度劣化のない高速な音声認識探索

Simultaneous Masked and Unmasked Decoding with Speculative Decoding Masking for Fast ASR without Accuracy Loss

○岡部浩司，山本仁（NEC）

◆ Autoregressive デコーディングの**高い音声認識精度**を保ったまま、**高速化**を実現する探索手法である**マスク内外同時サーチ（SMUD）**と **Speculative Decoding Masking** を提案します。

◆ E-branchformer などの高精度な音声認識モデルの探索に適用可能です。

Table 1: WER or CER (%) (dev set / test set)

	AISHELL-1	JSUT	LS-100	TED-LIUM2
AR	4.2 / 4.5	11.6 / 13.3	6.3 / 6.5	7.3 / 7.2
SMUD	4.2 / 4.5	11.6 / 13.2	6.3 / 6.5	7.3 / 7.2

認識精度そのまま！

Table 2: Real time factor

	AISHELL-1	JSUT	LS-100	TED-LIUM2
AR	0.67	0.67	0.47	0.59
SMUD	0.61	0.55	0.40	0.41

速い！

1-R-7

1-R-7　大規模言語モデルに基づく音声認識における湧き出し誤りデータの分析

Analysis of unrecognized error in LLM-based automatic speech recognition system

◎柳田 智也, 沈 鵬, Lu Xugang, 藤本 雅清 (NICT),
須藤 克仁 (NWU/NICT), 河井 恒 (NICT)

背景
- ◆音声の Encoder と LLM を接続した LLM-based ASR の発展
- ◆LLM-based ASR における原因不明の **湧き出し誤り**

> 正解ラベル (Common Voice 19)
> a large portion of the cylinder had been uncovered
> 認識結果
> a large portion of the cylinder had been uncovered **and it was evident that it had been used as a receptacle for liquids**

- ◆**音響的特徴**が要因？
  - ➢ 上記の場合、**長時間の無音区間**に**微弱な雑音**が重畳
- ◆長時間の無音と雑音の重畳による影響を分析

分析
- ◆ LibriSpeech で LLM-based ASR を学習
- ◆開発セットに **摂動** を付与し、湧き出し誤りをシミュレーション
  - ➢ **音声へ無音区間の挿入**・**人工的な雑音の重畳**
- ◆挿入誤り、置換え誤り、削除誤りから、湧き出し誤りを分析

結果
- ◆10 秒以上の無音区間の挿入
  - ➢ 単語や文字を繰り返す挿入誤りの増加が顕著
- ◆20 秒の無音区間の挿入
  - ➢ 少数ケースだが、正解ラベルと関係のないテキストの湧き出しを確認

1-R-8

1-R-8　YULU-CER: Yielding Usable Large Language Model Utilized Character Error Rate for Practical Evaluation

○森大輝 (Allm・DeNA), 園部良介 (DeNA), 近藤伊佐直 (Allm・DeNA)

- ◆従来の ASR 評価指標 (WER や CER)は, 表記揺れやフィラーなど非本質的な相違点を過大にカウントしてしまい, ASR モデルの実運用を念頭に置いた際の性能を正確に反映できない。
- ◆本研究では, フィラー, 言い淀み, 語断片などの表記揺れを柔軟に扱い, 実運用的に問題のある誤りのみを抽出する新しい評価指標「YULU-CER」を提案し, その有用性を示した。
- ◆Fig.1 に YULU-CER の算出するための YULU-CER 表現の作成方法を示す。
- ◆YULU-CER を用いて複数の ASR モデルを比較し, 従来の CER では捉えにくい評価ポイントを YULU-CER がより詳細に捉える可能性が示唆された。
- ◆また, LLM で自動生成された YULU-CER の精度は人手による結果と高い一致率を示し, その信頼性が確認された。

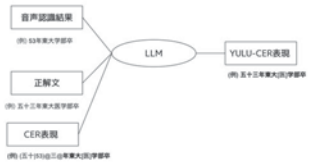


Fig.1：YULU-CER表現の作成方法

1-R-9

1-R-9　音響情報を考慮した大規模言語モデルによる音声認識の誤り訂正

Error Correction for Automatic Speech Recognition with LLM Incorporating Acoustic Information

☆鈴木萌々音, 上乃聖, 李晃伸（名工大）

- ◆LLM の入力に 10-best 仮説と音声認識モデルにおいて計算される音響スコアを使用し, 音声認識の誤り訂正を行う.
- ◆プロンプトを複数種類実行し, その違いによる誤り訂正の傾向を分析.
- ◆1-best の結果と提案法を比較すると, test_other では改善が見られた.
- ◆音響スコアを扱わない LLM を使用した実験と提案法を比較すると, test_clean および test_other の両方で性能の改善が見られた.
- ◆音響スコアが最も高い 1 文を 10-best から選択する指示をしたプロンプトでの結果が最も良いことが分かった.
- ◆誤りを訂正し, 新しい文を生成する指示を加えると新たに生成された文の割合は増加したが、同時に改悪された割合も増加した.

Table 1. Word error rate (↓, %) of error correction on the LibriSpeech. Dev_clean and dev_other were used for fine-tuning LLM.

	test_clean	test_other
1-best	2.74	6.81
Without acoustic scores	4.67	9.25
With acoustic scores		
Select one transcript with the highest score	3.01	**6.72**
Modify errors after selecting one transcript	2.97	6.79
If the highest score is less than 50, modify errors	2.94	6.74
Select one transcript or modify one transcript	2.94	6.73

1-R-10

1-R-10　中間 CTC 目標を活用した多言語 ASR におけるコードスイッチングの向上

Enhancing Code-Switching with Intermediate CTC Objectives in Multilingual ASR

☆東翔、サクティ・サクリアニ（NAIST）

- ◆コードスイッチング（CS）は、自動音声認識（ASR）システムにおいて異なる言語が混在する音声を処理する上で重大な課題である。従来の CS 対応手法は単言語モデルに比べ性能が劣り、特に言語切り替え箇所での認識精度が課題とされている。
- ◆本研究では、言語特化型アダプターと中間 CTC を組み合わせた新しいアーキテクチャを提案する。
- ◆この手法は、Transformer Code Switcher（TCS）による軽量なアダプターを使用した動的な言語切り替えと中間 CTC を活用することで、異なる言語特性を効果的に分離し、干渉を低減する。
- ◆提案手法を英語 - 中国語の CS コーパスに適用した実験では、MER が 1.40%、CER が 1.61%それぞれ改善し、最良の性能を示した。
- ◆本研究の成果は、CS ASR における新たな設計指針を提供するとともに、複数言語が混在する現実世界の音声認識タスクにおいて幅広く応用可能であることを示す。

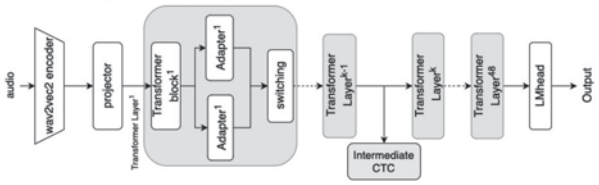


Fig.1：Architecture of TCS with Intermediate CTC.

1-R-11

1-R-11　異常検知技術に基づく畳み込み Autoencoder による非母国語話者の日本語発話誤り検出

Japanese speech error detection for non-native speakers using convolutional autoencoder based on anomaly detection technique

△加藤 動,△Zhou Qihang(愛知工科大),北岡教英(豊橋技科大),○實廣貴敏(愛知工科大)

◆非母国語話者の発話誤りを検出する研究は音声認識精度が向上してきた頃から多く，GMM-HMM を用いたもの[Witt+ 2000]などがある。

◆本研究では，異常検知技術の考え方を用いた発話誤り検出方法を提案する。以前に提案した Autoencoder (AE)によるもの[Zhou ら 2022]を発展させた，畳み込み Autoencoder (CAE)を用いたものや，LSTM を取り入れた Autoencoder (LSTM AE)を提案，検討する。

◆音声特徴量には，スペクトル包絡情報として MFCC 系を使うとともに，韻律情報として基本周波数を組み合わせて使う。中心フレームの前後フレームを連結，特徴量ごとに用いるフレーム数を最適化する。

◆新聞記事読み上げ音声コーパス JNAS で日本語母国語を学習，留学生読み上げ日本語音声データベース UME-JRF で評価を行った。

◆誤棄却率 FRR，誤受理率 FA の等誤り率では，CAE が最良の 34.0%を得た。

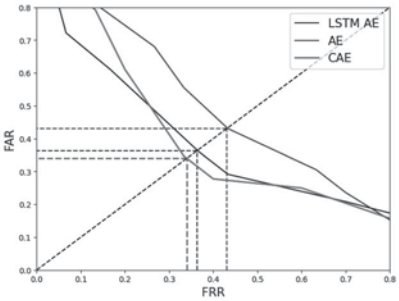


Fig.1： ROC curves by the proposed methods.

1-R-12

1-R-12　音声認識誤りが ChatGPT の翻訳に与える影響の調査

Investigating the Impact of Speech Recognition Errors on ChatGPT's Translation

△安藤宏祐(新居浜高専)，☆平野雄太，佐藤颯空，サクティサクリアニ(NAIST)

◆音声翻訳技術は社会的に重要な役割を果たしており，高精度な翻訳能力を持つ ChatGPT の応用が期待されるが，ChatGPT の音声認識誤りに対する頑健性や脆弱性は未知数である。

◆精度の異なる 5 種類の音声認識モデルを使って翻訳精度への影響を調べた結果，多少の認識誤り率の増加は翻訳の劣化に繋がりにくいことが分かった。

◆翻訳の劣化に繋がりやすい音声認識誤りは，「名詞の誤認識」と「同音異義語によるもの」の二つに大別された。このような音声認識誤りは，ChatGPT に入力される前に修正されることが求められる。

◆音声認識誤りの存在をプロンプトに含めることは，翻訳精度を表すスコアの改善に繋がらないことが分かった。

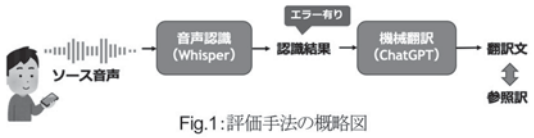


Fig.1：評価手法の概略図

	誤り例 1	誤り例 2
翻訳ソース文	リンゴを食べると故郷が思い出される。	私は紙を燃やす。
ChatGPT の翻訳結果	Eating an apple makes me think of my hometown.	I burn paper.
音声認識結果	**インゴ**を食べると故郷が思い出される。	私は**髪**を燃やす。
ChatGPT の翻訳結果	Eating **"ingo"** makes me reminded my hometown.	I'm burning my **hair**.
参照訳	When I eat apples, I am reminded of home.	I burn the paper.

Table.1：音声認識誤りに起因する翻訳誤りの例

1-R-13

1-R-13　構音障害音声を用いた wav2vec モデル学習における障害種別についての比較

A comparative study of wav2vec model training on speech from different types of speech disorders

☆大谷赳[1]，相原龍[2]，高島遼一[1]，滝口哲也[1]，△田口進也[2]
([1]神戸大学，[2]三菱電機)

◆本研究では脳性麻痺による運動障害性構音障害者，口唇口蓋裂および舌切除による器質性構音障害者を対象とした音声認識を目的とする。

◆構音障害者の音声認識モデルの学習を行う際，構音障害者の音声を大量に収集することが困難であるという課題がある。

◆本研究では，利用可能な構音障害者音声を増やすことを目的に，ラベルなし自由発話音声と，複数の構音障害者音声の効果的に活用する方法について検討する。

◆提案手法では，wav2vec2.0 の枠組みで，構音障害者音声のラベル無し音声を用いた自己教師あり学習を行う（Fig. 1）。

◆論文では，ラベルなし構音障害者音声として認識対象者の音声に限らず，認識対象者と同じ種別の障害を持つ構音障害者の音声や複数種類の構音障害者の音声を用いて実験を行い，有効性を調査した。

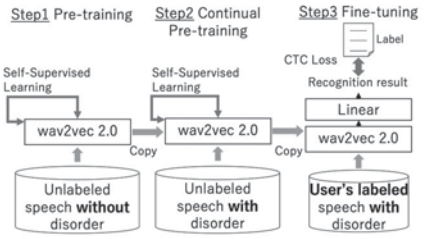


Fig.1：The proposed model training procedure

1-R-14

1-R-14　健常者音声を活用した Transformer による構音障害者音声認識

Speech Recognition for Individuals with Dysarthria Using Transformers Leveraging Non-Dysarthric Speech Data

○蓑毛悠晟，陳金輝(和歌山大)，陳訓泉(県立広島大)

◆構音障害とは、中枢神経や運動機能の障害により発話が不明瞭になる状態の総称である。構音障害者の音声認識はデータ不足が大きな課題となっている。

◆本研究では、大規模データセットである健常者の音声データを転移学習し、構音障害者の音声データをデータ拡張することでデータ不足に対処し、Transformer を用いて音声認識を行う。

◆音声認識などの分類タスクで一般的に使用される Cross-Entropy Loss に加えて、過学習が起こりやすい小規模データセットで有効な Cosine Loss を損失関数として使用し評価を行う。

◆下図は実験結果を簡易的にまとめたものである。転移学習、データ拡張の有効性を示した。また、Cosine Loss はクラス間の音声特徴が大きく重なっている音声明瞭度が「非常に低い」・「低い」グループで効果的でないことが分かった。

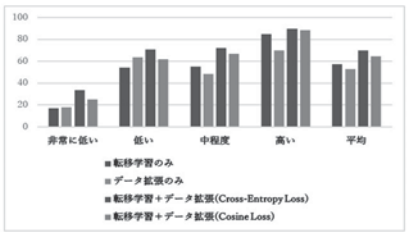


図 1:WRA を用いた評価実験結果

1-R-15

1-R-15　聴覚障害者音声における音響と言語の交互適応による音声認識の高精度化

Enhancing Speech Recognition Accuracy through Cross-Adaptation of Acoustic and Linguistic information for Japanese Deaf and Hard-of-Hearing People.

☆高橋 快斗（豊橋技科大），若林 佑幸（豊橋技科大），
太田 健吾（阿南高専），小林 彰夫（大和大），北岡 教英（豊橋技科大）

<背景>

◆以前の手法：エンコーダ層の置換より，認識精度が向上

◆少量の聴覚障害者音声による fine-tuning 時に過学習が発生

<提案手法>

◆モデルへ入力する音声に応じて学習するエンコーダの層を変更することにより，過学習を抑制し認識精度が向上

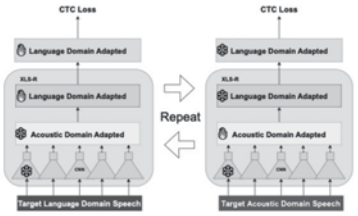


Fig. 1 Proposed alternating learning of acoustic and linguistic information.

<結果>

◆聴覚障害者音声において，エンコーダ層置換により認識精度が向上

Table 1 Comparison of ASR performance when using each method using CER

method	model	# Parameters (MB)	1st fine-tune	2nd fine-tune	JNAS CER (%)	DEAF CER (%)
Baseline	ReazonSpeech v2.0	619	DEAF	N/A	10.7	26.0
	Whisper Medium	769	DEAF	N/A	9.0	25.1
	XLS-R	319	JNAS + LTV	N/A	**7.7**	39.5
	XLS-R	319	JNAS + LTV	DEAF	8.3	23.0
	XLS-R w/ Replacement [6]	319	JNAS + LTV	DEAF	9.3	22.1
Proposed	XLS-R w/ Alternating learning	319	JNAS + LTV	JNAS + LTV ⇄ DEAF	**7.7**	**21.3**

1-R-16

1-R-16　子ども音声の音素認識における誤り傾向の調査

Error analysis in phoneme recognition of children's speech

◎堀井こはる，俵直弘，小川厚徳，荒木章子（NTT）

◆高精度な子ども音素認識を実現するためには，どの音素が誤認識されやすいかといった誤り傾向の分析が不可欠である

◆子ども音声を対象とした end-to-end 英語音素認識器を Wav2vec2.0 で構築し，その音素誤り傾向を年齢や誤りパターンの観点から分析し以下の知見を得た

- 成長に応じて音素認識誤り率が低下する
- 音素認識誤り率が改善する年齢は音素毎に異なる（Table 1）
- 特定の調音位置や様式で生成される音素（[ŋ, g]など）は，成長に伴う音素誤り率の改善が遅い（Table 1）
- 特定の音韻プロセスに起因した置換誤りが多い（Table 2）

◆ 以上の結果から，end-to-end 子ども音素認識器による主な誤り要因は，特に低年齢の子どもにおいて特定の音素が正しく発音できていないことが支配的であることが示唆された

Table 1: Age at which the phoneme error rate drops below 20% for each phoneme.

Age	5	6	7	8	11	12	14	15, adult
phoneme	ð	ʃ,ɹ,z,w, i, eɪ, k, s, h, n, aɪ	ɔɪ, f, ʊ, ɚ, l, d, p, ʌ, b	v, θ, ɑ, tʃ, t, æ, m, u	ɛ	ʒ, ɪ	ŋ	ɔ, g, dʒ, j, oʊ

Table 2: Common substitution errors made by 5-year-old.

Phonological process	Substitution pattern	Example	Substitution error rate [%]
Stopping	θ → t	**th**ank → **t**ank	9.26
Deaffrication	tʃ → ʃ	**ch**ai → **sh**ay	8.16
Fronting	θ → f	**th**ank → **f**ank	9.26

1-R-17

1-R-17　方言音声のキーワード検出における事後確率精緻化方式の提案

A Proposal for Sophisticating Posterior Probabilities in Keyword Spotting of Dialectal Speech

☆有賀智広（岩手県立大），△小嶋和徳（岩手県立大），
李時旭（産総研），伊藤慶明（岩手県立大）

◆ 遠野物語理解支援システム

- 遠野物語の理解を支援するために，語り中の音声からキーワードを自動検出し，解説を表示するシステム

◆研究目的：キーワード検出の精度向上

◆提案方式：キーワード検出に自己教師あり学習モデルを導入するとともに，複数の自己教師あり学習モデルから事後確率値を求め，局所距離計算時に複数の事後確率値を用いて精緻化する方式を提案

- 自己教師あり学習モデルは日本の標準語音声で fine-tuning

◆ ２つの照合手法に提案手法を適用した(Fig1).

- Posteriorgram 照合方式では精度が低下したが，フレームレベル系列照合方式で，これまで単体のキーワード検出精度で最も高かった検出結果と比べ，monophone で 2.19pt，syllable で 1.00pt の検出精度の向上

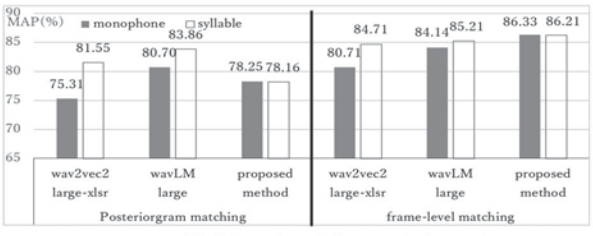


Fig1. Result of keyword detection

1-R-18

1-R-18　拡散モデルベース DNN 音声合成のバックボーンに着目した軽量化とカーネル形状変化の影響

Lightweighting Approach Focused on the Backbone of Diffusion Model-Based Text-to-speech Synthesis and the Effects of Kernel Shape Changes

☆佐藤颯空（NAIST），サクテイ サクリアニ（NAIST）

◆**背景・目的**：

近年の深層学習を用いたテキスト音声合成（TTS）モデルに関する肥大化が，リソース制約環境での利用を難しくしている．拡散モデルに基づく TTS モデルに対し，その内部に着目し、軽量化と合成音声品質を維持する手法を提案する．

◆**手法**：“ **Time - Frequency Kernel Module** ”

- 画像空間方向とチャンネル方向への畳み込み演算の分解
- 音声中間特徴量であるメルスペクトログラムのデータ構造に着目し、時間軸と周波数軸に沿ったカーネル形状の適用

◆**結果**：

- 標準的な畳み込み層を用いた場合と比較して、パラメータ数 50%程度削減と約 26%高速化を実現
- カーネル形状・サイズを変更することで合成音声の品質に影響を与えることを確認

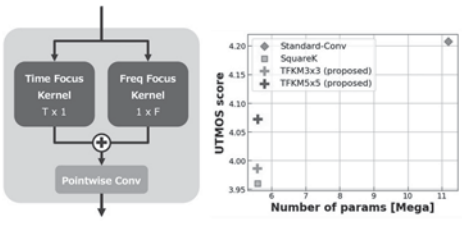


Fig.1: Time-Frequency Kernel Module (TFKM)
(left) Structure of TFKM (right) Number of parameters and UTMOS score

1-R-19

1-R-19 自然言語による声質制御のための音声・声質説明文ペアデータの作成・評価 システムの検討

A study of a system for creating and evaluating speech/description pair data for voice characteristics control using natural language

☆田牧宏都, 橋本佳, 南角吉彦, 徳田恵一（名工大）

近年，自然言語で声質制御可能な音声合成技術の研究が注目を集めている．一方，音声・声質説明文ペアデータの作成はコストが高いことが知られている．本稿では，自然言語による声質制御のための音声・声質説明文ペアデータの作成や評価の効率化に利用可能なインターフェースの検討・開発を行った．

実際にインターフェースを作成し，評価実験を行った．実験結果から，特に説明文評価機能に関して，属性ごとの評価が可能であり，説明文の改善につながる可能性があることが示唆された．

今後はインターフェースのほかの各機能に関して，実用性の評価と具体的な活用方法の提案を行っていきたいと考えている．

テキスト	オカルタはやってない．俺たちが立ってる場所に立ちたい奴は，たくさんいる			
説明文	各属性および平均評価値			
20 代くらいの男性が，低い声で，諭すように話している．	男性 5.00 ± 0.00	低い 4.63 ± 0.43	20 代くらい 3.25 ± 0.59	諭し 3.50 ± 0.77
40 代くらいの男性が，落ち着いているが，内心はすごく不安な感じで喋っている．	男性 5.00 ± 0.00	40 代 2.75 ± 0.87	落ち着いている 4.00 ± 0.77	不安な感じ 2.88 ± 0.70
若い２人男性が，低い声で，誰かをかばうようなことをしゃべっている．	男性 5.00 ± 0.00	低い 4.00 ± 0.63	若い２人 1.38 ± 0.62	誰かをかばう 3.00 ± 0.63
消極的な雰囲気の男性が，聞き取りやすいがぼそぼそとした感じの声で，アニメのような台詞口調で話している．	男性 5.00 ± 0.00	ぼそぼそとした 3.38 ± 0.77	消極的な雰囲気 3.38 ± 0.99	アニメのような台詞口調 4.38 ± 0.77

Fig.1 : Part of the evaluation for data with the same voice and different descriptions

1-R-20

1-R-20　テキスト・発話スタイル同時制御を可能とする非流暢性に着目した講演音声生成

Text- and spoken-style controllable lecture speech generation focusing on disfluency.

☆吉岡大貴 (名大), 中田優翔 (徳山高専), 安田裕介, 戸田智基 (名大)

本稿では，「書き言葉を用いた自発音声合成」という応用のため，非流暢性アノテーションを用いたテキストスタイル変換とテキスト音声合成，その組み合わせを提案する．テキスト音声合成の技術は発展し，読み上げ音声の生成においては人間に近い自然性を実現している．一方で，より人間に近い「自発音声」の生成においてはまだ発展途上である．また，既存の自発音声合成モデルの多くは，入力テキストに非流暢性などの自発的要素が含まれていることを前提としている．しかし，講演音声や説明音声を生成したい時，その元となる文章は書き言葉などの自発的要素が含まれないものである場合がほとんどである．

そこで，本稿では特に「書き言葉テキストからの講演音声合成」というタスクに焦点を置き，より自然で人間らしい講演音声の合成を目指す．具体的には，TST や TTS を訓練する際に，フィラーや言い淀みなどの流暢でない部分に対してタグを付けたり特殊な記号に変換したりする非流暢性アノテーションを用いることで，各モデルにおける非流暢性の言語的・音響的制御性を向上させる．この非流暢性アノテーションを用いた TST と TTS を組み合わせることで，講演音声生成システムを構築する．これにより，既に存在する書き言葉の資料を基に，講演音声・説明音声を合成することが可能になり，新しい原稿を作成するコストを削減することができる．非流暢性アノテーションを行わない場合と比較した客観的・主観的評価実験の結果から，本手法の有効性を示す[*1]．

[*1] 音声サンプル：https://dyoshioka-555.github.io/SponTTS-samples/audio_samples.html

1-R-21

1-R-21　ペルソナ説明文を利用した合成音声の話者性制御手法の検討

Towards speaker identity control using Persona-Sentence in Prompt-based TTS system

☆濱田 誉輝, 齋藤 佑樹, 中田 亘, 山内 一輝, 関 健太郎, 岡本 悠希, 猿渡 洋(東大院･情報理工)

◆Prompt-based TTS において，話者プロンプトと呼ばれる話者の声質を説明したプロンプトが利用されるが，先行研究で利用される話者プロンプトには，合成したい音声に対して最適な話者性を表すプロンプトを適切に言語化するのが困難であるという問題がある．

◆本研究では話者プロンプトを拡張し，合成音声の話者性の制御をより簡易にすることを目的として，**ペルソナ説明文**と呼ばれる話者自身のプロフィールに関連する自然言語記述を導入し，大規模言語モデルを利用した音声コーパスとペルソナ説明文の自動ラベリング手法を提案する．

◆提案手法により，話者プロンプトにペルソナ説明文を利用した新たな音声コーパスを構築し，分析を行なった．

ペルソナ説明文	話者プロンプト
i have a ten year old son.	very mature, adult-like
my husband is a cop.	very masculine, powerful, middle-aged, slightly strict
i live in texas.	slightly friendly, slightly relaxed very clear, very gender-neutral,, slightly mature

Table 1. ペルソナ説明文と話者プロンプトのペアデータ

1-R-22

1-R-22　データ単位前処理自動選択による音声合成コーパスのデータクレンジング

Data Cleansing for Speech Synthesis Corpora through Automatic Data-Level Preprocessing Selection

◎関 健太郎(東大), 高道 慎之介(慶大／東大), 佐伯 高明, 猿渡 洋(東大)

◆ インターネットデータ，どう前処理する？
◆ 音声強調？音声復元？そのまま使う？
◆ データごとに最適な手法が異なる…

▷ 自動選択で組み合わせる！

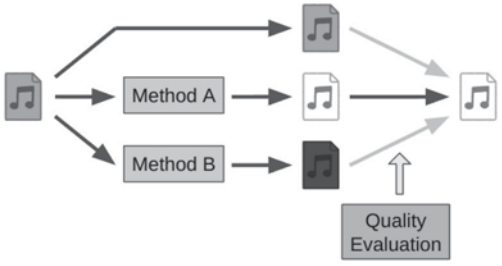


Fig.1 : Our proposed data-level preprocessing switching method.

結局どの手法が有効なの？
気になる実験結果はポスターまで！

1-R-23

1-R-23　明瞭音声を学習データに用いた場合の DNN 音声合成音のポップアウトの程度の調査

Investigation of the degree of pop-out in DNN speech synthesis using training data of high-intelligibility speech.

☆上杉亮介，坂野秀樹，旭健作（名城大院）

◆背景雑音から際立って目立つ音声をポップアウトボイスと呼ぶ．本研究ではTTSにおける学習音声に通常音声・明瞭音声を用いた場合の，合成音声のポップアウトに関する音響的な特徴を調査する．

◆以下、通常音声を Org-normal，明瞭音声を Org-HI，それぞれの音声を学習音声に用いて合成した音声を Syn-normal, Syn-HI と表記する．

◆メルケプストラム距離による客観評価の結果，音声合成モデルは明瞭音声の特徴を捉え切れていない可能性があると考えられる．

Table.1：Mel-cepstral distances [dB] across four speech types.

	Org-normal	Org-HI	Syn-normal	Syn-HI
Org-normal	-	3.01	2.91	3.35
Org-HI	3.01	-	3.05	3.11
Syn-normal	2.91	3.05	-	2.35
Syn-HI	3.35	3.11	2.35	-

◆動的特徴量の平均値を比較すると，音声合成モデルが動的特徴量の変動を学習しようとしていることが確認できる．

Table.2：Averages of dynamic feature of four speech types.

	Org-normal	Org-HI	Syn-normal	Syn-HI
Avg of D_Δ	0.272	0.301	0.237	0.270

1-R-24

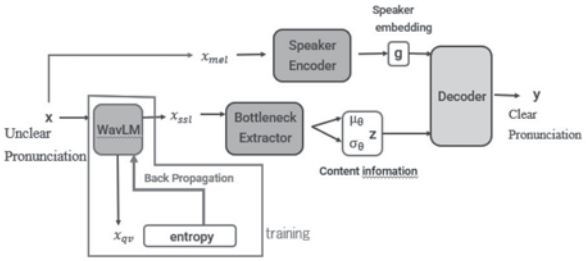
1-R-24　不明瞭音声の文字認識率向上と明瞭音声の合成

Enhancing Speech Recognition for and Synthesizing Clear Speech

○吉良青空(島村研)

◆ 本研究では、不明瞭音声の明瞭化と話者特性の保持を両立する音声変換システムを構築する。

WavLM-Large を音響特徴量抽出器とし、**CTC デコーダ** を用いて日本語音声認識精度を向上させる。さらに、**FreeVC** を活用し、認識結果を基に明瞭な音声を生成する。**JVS コーパス** を用いた学習により、CTC 損失を最適化し、日本語音声の変換精度を高める。本手法は、リアルタイム処理が可能な音声コミュニケーション支援技術としての応用が期待される。



x:source waveform , y:generated waveform , xmel:mel-spectrogram , xssl:SSL feature, xqv:predicted score , $\mu\theta$:Speaker-Independent Mean Adjustment , $\sigma\theta$:Feature Normalization for Generalization

1-R-25

1-R-25　感情音声合成における精度向上手法の検討

Proposing methods for improving accuracy in emotional speech synthesis

☆安田遥佳，志賀芳則（東京電機大院・工学研）

感情表現が可能な音声合成システムの高音質化を目指して、Self-attention (SA) または Lambda Networks (LNs) を用いた音響特徴量推定法を提案する。従来手法に比べて、SA や LNs は、より長期的な系列の依存関係をモデル化できるため、感情とその遷移がより明確に伝わる音声の合成が可能になると期待される。SA / LNs を Tacotron 2 に適用し実装した音響特徴量推定処理の構成を図 1 に示す。

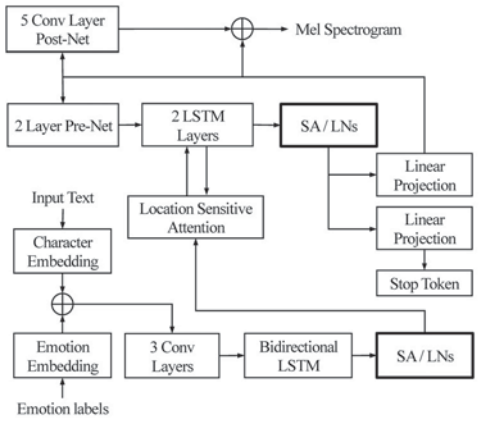


Fig. 1: Integration of SA / LNs into Tacotron 2

1-R-26

1-R-26　言語別球面座標感情ベクトルと音声トークンを使用した多言語感情音声合成に関する考察

A Study on Multilingual Emotional Speech Synthesis Using Language-Specific Spherical Vectors and Speech Tokens

☆朴 浚鎔，中村 建一（Preferred Networks）

◆本研究では、多言語感情音声合成の性能を向上させるために、新たな感情・音響統合モデルを提案する。

◆モデルは、球面感情ベクトルと離散トークン特徴を組み合わせることで、従来の感情ラベルや単一の感情ベクトルを用いたモデルに比べ、感情表現力や自然性の向上を目指している（Fig.1）。

◆感情情報を球面座標系で表現することで、感情表現の滑らかさや制御性を向上させ、SSL モデルを活用して抽出した離散トークン特徴を加えることで、音響的多様性を統合するアプローチをとる。

◆これにより、従来手法と比較して、感情の細やかな韻律や言語情報での了解性を保ちながら、自然で表現力豊かな多言語音声合成を示す。

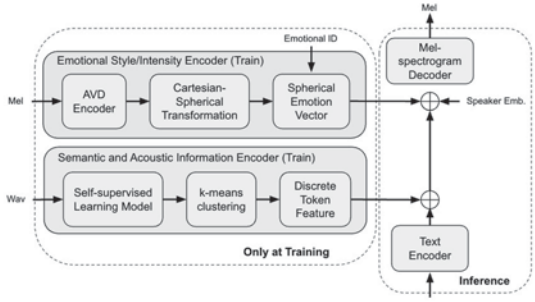


Fig.1：Model Architecture of Proposed Method

1-R-27

1-R-27　日本語多数話者音声コーパスを用いた混合感情音声合成の性能向上

Performance improvement of mixed emotion speech synthesis using a Japanese multi-speaker speech corpus

☆坂田一成，小坂哲夫(山形大院・理工学研)

◆我々はこれまで End-To-End モデルの TTS を用いて、**日本語を対象とした混合感情音声合成**について検討を行ってきた。

◆しかし、2 ヶ国語で学習された事前学習済みモデルの性能が低く、合成音声の発音が不明瞭であるなど、自然性が不十分であった。

◆本研究では、**日本語多数話者コーパスで学習された事前学習済みモデルを用いることによる合成音声の自然性向上、及び先行研究で検討されていない混合感情について評価**を行った。

◆MOS による主観評価では**提案法が従来法より高い自然性**を示した。

◆MCD による客観評価では**提案法がより Ground Truth に近い結果が得られた。**

Table.1： Results of subjective and objective evaluations for synthesized speech

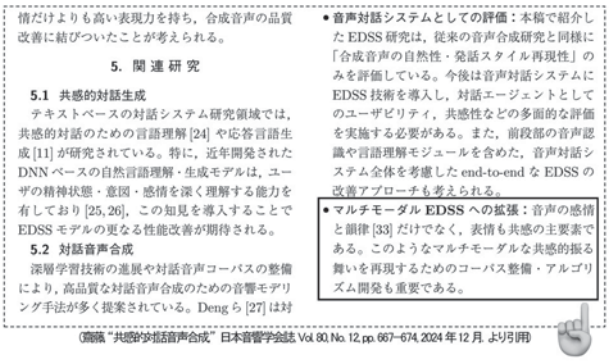
Configuration		MOS	MCD
Ground Truth		4.57	-
Baseline	Target Speech	1.68	11.04
	Speaker's Avg.	1.74	10.92
	Other's Avg.	1.57	11.58
Proposed	Target Speech	2.90	**7.69**
	Speaker's Avg.	**3.12**	7.85
	Other's Avg.	3.04	8.22

1-R-28

マルチモーダル共感的対話音声合成に向けたコーパスの構築

Corpus construction towards multimodal empathetic dialogue speech synthesis

○齋藤 佑樹*，陳 晋升*，楊 棟，丹治 尚子(東大院・情報理工)，土井 啓成，白旗 悠真，朴 炳宣，橘 健太郎(LY)，猿渡 洋(東大院・情報理工)
(*の付いた著者による equal contribution)


情だけよりも高い表現力を持ち，合成音声の品質改善に結びついたことが考えられる。

5. 関 連 研 究

5.1 共感的対話生成

テキストベースの対話システム研究領域では，共感的対話のための言語理解 [24] や応答言語生成 [11] が研究されている。特に，近年開発された DNN ベースの自然言語理解・生成モデルは，ユーザの精神状態・意図・感情を深く理解する能力を有しており [25, 26]，この知見を導入することで EDSS モデルの更なる性能改善が期待される。

5.2 対話音声合成

深層学習技術の進展や対話音声コーパスの整備により，高品質な対話音声合成のための音響モデリング手法が多く提案されている。Deng ら [27] は対

- **音声対話システムとしての評価**：本稿で紹介した EDSS 研究は，従来の音声合成研究と同様に「合成音声の自然性・発話スタイル再現性」のみを評価している。今後は音声対話システムに EDSS 技術を導入し，対話エージェントとしてのユーザビリティ，共感性などの多面的な評価を実施する必要がある。また，前段部の音声認識や言語理解モジュールを含めた，音声対話システム全体を考慮した end-to-end な EDSS の改善アプローチも考えられる。
- **マルチモーダル EDSS への拡張**：音声の感情と韻律 [33] だけでなく，表情も共感の主要素である。このようなマルチモーダルな共感的振る舞いを再現するためのコーパス整備・アルゴリズム開発も重要である。

(画像 "共感的対話音声合成" 日本音響学会誌 Vol. 80, No. 12, pp. 667–674, 2024 年 12 月. より引用)

マルチモーダル共感的対話音声合成（EDSS）の研究に使えるコーパスを作りました

- 単一話者による**約 9.5 時間**の音声 & Face Mesh 特徴量で構成
- 研究目的であれば無償で利用可能とする予定

1-R-29

1-R-29　Mobile PresenTra：スマートフォン上で高速動作可能なニューラルTTS

Mobile PresenTra: Fast neural TTS system on smartphones

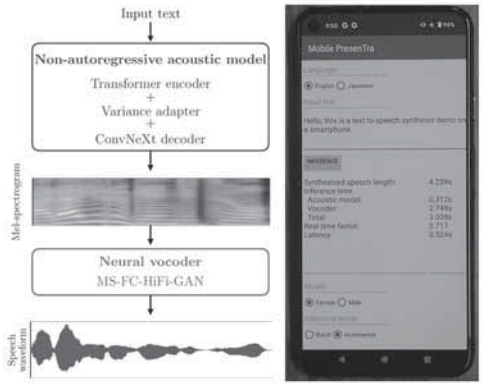
○岡本拓磨，大谷大和，河井恒
(情報通信研究機構)

・ポイント

・高品質・高速に動作するニューラルテキスト音声合成技術を開発

・CPUコアー つで1 sの音声をわずか0.1 sで高速合成(既存モデルの約8倍の速さ)することが可能

・ネットワークに接続されていないスマートフォン上でテキスト入力からわずか0.5 sの高速生成を実現



ポスター会場にてスマートフォン実機を用いた実演デモ(日英中韓)

1-R-30

1-R-30　VAE-SiFiGAN：変分自己符号化表現に基づくSiFiGAN

VAE-SiFiGAN: SiFiGAN Based on Variational Autoencoder Representations.

☆荻田健一，米山怜於，HUANG Wen-Chin，戸田智基(名大)

◆速度・品質・基本周波数（F0）制御性能を高水準で達成したニューラルボコーダ SiFiGAN の信号処理由来の音響特徴量に起因する問題

- ➢ 音声信号に含まれる揺らぎ成分の表現が限定的
- ➢ End-to-End の枠組みへの直接的な応用が難しい
- ➢ 信号処理による音響特徴量の抽出アルゴリズムは雑音耐性が低い

◆VAE に基づく学習可能な潜在表現を活用することで上記の問題を解決し，SiFiGAN の応用範囲の拡大を狙う

◆潜在表現に内在する F0 情報と SiFiGAN に別途与えられる F0 系列の不整合により F0 制御性能が低下する傾向

- ➢ 潜在表現から F0 情報を除去する機構を導入

◆実験的評価により提案手法が従来の F0 制御性能を上回ることを実証

Table 1　Experimental results.

Method	RMSE ↓	V/UV ↓	MOS ↑ w/ 95% CI
$1.0 \times F_0$			
Original	-	-	4.34 ± 0.07
SiFiGAN	0.02	5	$\mathbf{4.32 \pm 0.08}$
VAE-SiFiGAN	0.02	5	4.30 ± 0.08
w/o F_0 removal	0.02	5	4.22 ± 0.08
$0.5 \times F_0$			
SiFiGAN	**0.03**	**5**	2.48 ± 0.10
VAE-SiFiGAN	**0.03**	**5**	$\mathbf{3.04 \pm 0.11}$
w/o F_0 removal	0.04	6	1.85 ± 0.09
$2.0 \times F_0$			
SiFiGAN	0.06	9	2.46 ± 0.11
VAE-SiFiGAN	**0.04**	**8**	$\mathbf{3.25 \pm 0.12}$
w/o F_0 removal	0.10	**8**	2.65 ± 0.11

1-R-31

1-R-31　Flow Matching による周波数領域でのフローマッチングを用いた高速ボコーダー

Fast Neural Vocoder via Flow Matching in Frequency Domain

○オウ・エイケツ，サクティ・サクリアニ（NAIST）

◆ニューラルボコーダーは、音声合成システムにおいて、音声波形を生成する役割を持つ。従来の逆短時間フーリエ変換（iSTFT）を用いた手法は高速な推論ができるが、データに対するロバストネスが乏しく、不安定な学習になりやすい。その一方、拡散モデルは安定な学習と高度な生成能力を持っているが、推論速度が遅いという課題がある。

◆本研究では、iSTFT と拡散モデルに基づいた Flow Matching を組み合わせた新しいニューラルボコーダーを提案する。

◆この手法では、STFT により抽出された振幅成分と位相成分を用いて、モデルを Flow Matching で学習させることによって、推論速度と音声品質を向上させる。

◆提案手法は LibriTTS と JSUT データセットに適用した実験では、最良のロバストネスを示し、拡散モデルの Baseline より高速な合成ができることを示した。

◆本研究の成果は、ニューラルボコーダーにおける iSTFT と Flow Matching のトレードオフを取った新たな設計を提供することで、幅広い音声合成システムに応用可能であることを示す。

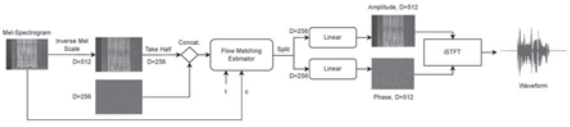


Fig.1 : model overview

1-R-32

1-R-32　ニューラルボコーダ合成音における有声区間での瞬時的振幅低下の抑制に関する初期検討

Preliminary investigation on the suppression of instantaneous amplitude decrease in voiced segments of neural vocoder synthesized speech.

○平井 俊男，スティアワン イファン（株式会社アルカディア）

◆ニューラルボコーダでは、合成音の有声区間において瞬時的に振幅が低下し、合成音の品質が劣化してしまう問題がある。

◆本稿では、End-to-End 音声合成システムである VITS の合成音における瞬時的な振幅低下（“dip”）に焦点を当て、その抑制に関する初期的検討について報告する。

◆検討の結果、潜在変数から音声波形へのアップサンプリングの初期段階におけるフレーム単位でのデータ削除等により、dip を 75％ 以上抑制できることが分かった。抑制の例を Fig.1 に示す。

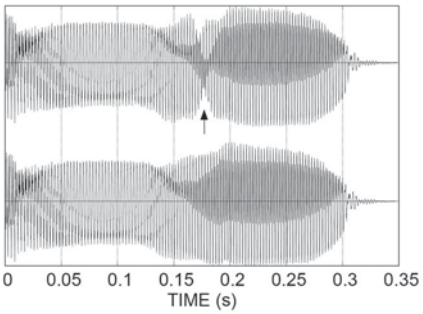


Fig.1 : VITS’“dip” example (arrow) and its suppressed waveform by {pst, del}. Japanese utterance:“(ka)iyo:”(海洋“ocean”).

1-R-35

1-R-35

基本周波数非依存なDNNによる鳥類の鳴き声のスペクトル構造の分類

Spectrum classification of bird song using pitch-independent DNN

☆ 中谷優太，矢田部浩平（農工大），Tarciso Velho，Diego A. Laplagne（Federal University of Rio Grande do Norte）

- 畳み込み層と Fullsort 層によって，入力したスペクトルから基本周波数非依存な特徴量を取り出す特徴抽出器（Fig. 1）を提案する
- それを基本周波数非依存なスペクトル分類問題である，鳥類の鳴き声のスペクトル構造分類に用い，その有効性を確認した

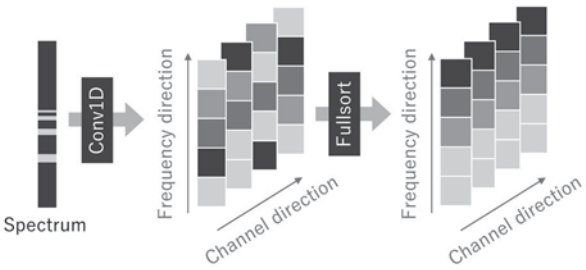


Fig. 1 Network architecture for spectrum classification

1-R-36

1-R-36

鳥類の鳴き声のスペクトル構造の分類に用いる特徴抽出器と分類手法の比較

Comparison of feature extractors and classification methods for spectrum classification of bird song

☆ 中谷優太，木住野貴宏，矢田部浩平（農工大），Tarciso Velho，Diego A. Laplagne（Federal University of Rio Grande do Norte）

- キンカチョウという鳥の発声器官には音源が 2 つある
  - ▷ 鳴き声のスペクトルには，1 つの調波で表せる構造と，2 つ以上の調波が重なったような構造が含まれる
- 鳴き声の各時刻のスペクトル構造を，Fig. 1 に示すモデルを用いて，上記 2 つの構造に無音区間を合わせた 3 つのラベルに分類した
- 2 つの特徴抽出器と，分類器に用いる 5 つの機械学習手法の組み合わせを試し，性能を比較した

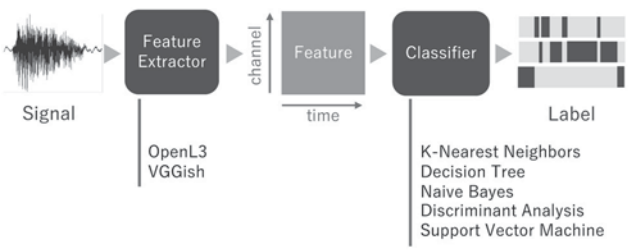


Fig. 1 A model for classifying the spectrum of bird song

1-R-37

1-R-37

動物音声用の基盤モデルを用いた鶏の鳴き声の分類の検討

Study on classification of chicken calls using foundational model for animal sounds

☆王樑, 照沼卓磨, 矢田部浩平, 新村毅, 福田信二（農工大）

◆ 背景
- 鶏の音声コミュニケーションは雛の行動発達に影響
- 特に，母鶏が雛の採餌を促す food call は重要
- しかし，food call の検出には課題がある
  - ✧ 鶏の活動により生じるノイズと似ているため，単純な処理では検出が難しい
  - ✧ 機械学習を検討したいが，ラベル付きデータが少ない

◆ 本研究
- 少量の学習データから food call を検出するため，鳥類音声に特化した基盤モデル BirdAVES を用いたモデルを提案
- Food call の周波数帯域を考慮したデータの前処理も行う

◆ 実験結果
- BirdAVES を用いたモデルとデータの前処理を組み合わせて，food call をある程度の精度で予測できた

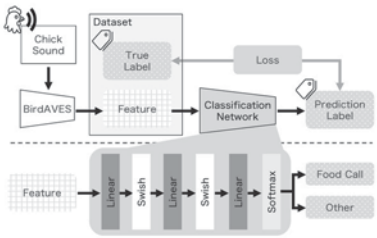


Fig.1: The structure of the classification model based on the foundational model for animals sounds

1-R-38

1-R-38 鶏の鳴き声の時間周波数的な特徴を考慮した鳴き声分類

Classification of chicken calls considering their time-frequency characteristics

☆照沼卓磨，王檬，矢田部浩平，新村毅，福田信二（農工大）

◆背景
- 鶏の疑似的な母子間音声コミュニケーションシステムの再現度向上に向け，鳴き声の解析とそれに向け高精度な検出が必要である
- 雛の鳴き声は親鶏の鳴き声に比べて特徴的であるため検出しやすい

◆目的
- 雛の鳴き声を高精度に検出
- 従来手法と位相の時間二階微分の振幅加重平均値を組み合わせた手法を提案

◆結果
- 提案手法は従来手法に比べて，F値が向上
  - 従来手法のF値　：0.648
  - 提案手法のF値　：**0.896**
- 位相の時間二階微分の性質を利用することで，鳴き声の種類の識別が可能
  - pleasure call 検出タスクにおけるF値　：0.910
  - distress call 検出タスクにおけるF値　：0.837

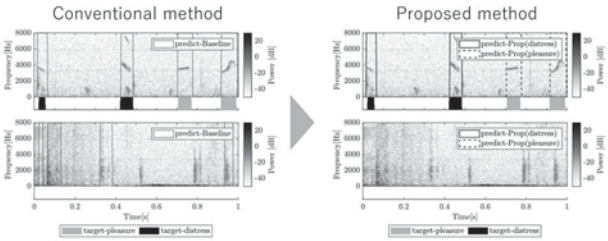


Fig.1 Chicken call detection by conventional and proposed method

1-R-39

1-R-39 事前学習済み深層学習モデルを活用した無尾類の種判別アルゴリズム

Species identification algorithm for anuran species using pre-trained deep learning models.

☆松原陽，福田信二（農工大），△中島直久（帯畜大）

◆背景
- 外来種の対策には，正確な分布域の把握に基づく生息環境評価と行動特性評価が必要である
- 北海道において国内外来種であるトノサマガエルが他の無尾両生類に与える影響が懸念されている

◆目的
- VGG16 や Vision Transformer など事前学習済み深層学習モデルの転移学習により無尾両生類の鳴音を用いて種判別を行う
- 水田圃場内における無尾両生類の対象種の分布域，活動が活発化する時間帯を推定する

◆結果
- 事前学習済み深層学習モデルによる分類により，トノサマガエルは圃場の間に多く分布域を持つことが示唆された
- トノサマガエルは24～26 時に鳴音を多く発することが分かった

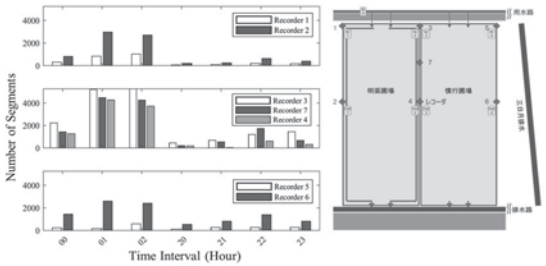


Fig. 1 Differences in the number of segments holding the calls of *Pelophylax nigromaculatus* in rice paddy fields

1-R-40

1-R-40　日本語モザイク音声の了解度に単語親密度が与える影響: 80 ミリ秒以下の処理時間窓長で低親密度単語に見られる了解度低下

The Effect of Word Familiarity on the Intelligibility of Japanese Mosaic Speech: Lower Intelligibility for Unfamiliar Words with an 80-ms or Shorter Processing-Time Window

○埋橋悠奈（九州大・芸術工），中島祥好（サウンド株式会社），蓮尾絵美，上田和夫，Gerard B. REMIJN（九州大・芸術工）

◆日本語単語のモザイク音声の聴取実験を親密度の高い単語と低い単語で行い，トップダウン的処理が音声の知覚に与える影響を検討した。

◆処理時間窓長 20, 40, 80 ms の条件で，高親密度単語のモーラ了解度が低親密度単語の了解度を有意に上回った (Fig.1)。また，モーラ誤答のうち母音正答は 37.5 %，母音誤答は 10.5 %，無回答は 52 % であった。無回答率に着目すると処理時間窓長 80 ms にて高親密度単語の方が低親密度単語に比べて無回答率が高くなった。

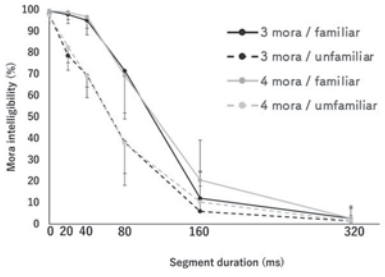


Fig.1: Relationship between segment duration (ms) and mora intelligibility (%) (n = 10). The error bars indicate standard deviations.

1-R-41

1-R-41　周波数変調刺激音の聴取に伴う発声の不随意応答：発声ピッチと刺激音の基本周波数差の影響

Involuntary vocalizations corresponding to modulated frequency: effect of fundamental frequency difference between vocal pitch and stimulus tone

☆土田晴登（豊橋技科大），河原英紀（和歌山大），松井淑恵（豊橋技科大）

◆発声する際、聴取音のピッチ変動と逆方向に発声ピッチが変動する補償応答について、予測不可能な変調刺激音を用いて調査を行った。

◆変調刺激音は、正弦波、複合音など５種類を用いた。刺激音の基本周波数は、参加者の普段の発声の基本周波数（個人条件）と性別によって一定の周波数（固定条件）の２種類を用いた。参加者は刺激音を聴取しながら刺激音と同じピッチで /a/ を 20 [s] 発声した。

◆各周波数、刺激の種類ごとの応答の大きさを比較したところ、刺激の種類による差は見られたが、周波数条件による差は見られなかった（Fig. 1）。

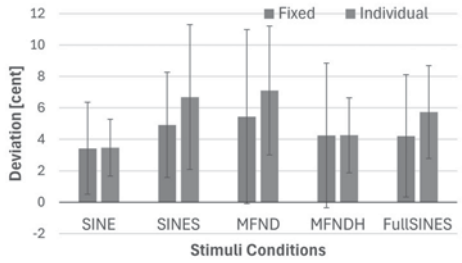


Fig. 1: Mean of pitch deviation to auditory stimuli.
Blue bar: fixed pitch condition, Orange bar: individual pitch condition.
Error bar: SD.

1-R-42

1-R-42　日本語における調音構造の変化が音声知覚へもたらす影響

The effect of articulatory structure's change on speech perception in Japanese

☆宗片亮太，△長村秀一，△鈴木開，冨岡晟多，小林耕太（同志社大院・生命医科学）

◆発話音声に変化を加えて聴覚的にフィードバックされると、その変化を補償するように発話運動を修正することが報告されている。

◆聴覚フィードバック環境下での補償応答のため調音構造が変わることにより、言語音声のカテゴリー知覚に影響が及ぶ可能性が示唆されているが、この現象が言語普遍的であるかは未だ不明である。

◆本研究では、発話音声のフォルマント周波数を変えてリアルタイムでフィードバックした後、日本語母音のカテゴリー識別を行い、補償応答後に母音カテゴリーの境界に影響が及ぶか検討した。

◆その結果、実験群ではフィードバックに対して補償するような発話運動は見られなかったが、曖昧な母音を「い」と知覚する割合が上昇した。特定の刺激を繰り返し聞くと、その項目に対する知覚感度が低下し、他の刺激の識別が行われやすくなる「曝露効果」が影響した可能性が考えられた。

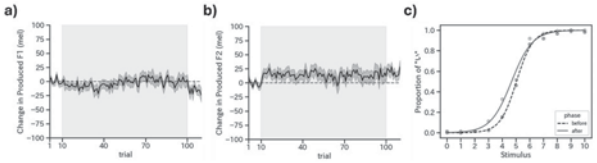


Fig. a) Change of produced pitch from Baseline in F1. b) Change of produced pitch from Baseline in F2 . c) Proportion of identification for sounds as "い". (Wilcoxon signed-rank test, $p < 0.05$).

1-R-43

1-R-43　倍音の振幅操作に基づく母音の無限音階化の検討

Generating tones of vowels with pitch circularity based on manipulation of the amplitude spectrum

☆橋本圭織，河村隆生，小野順貴（都立大），西澤佳飛，戸田智基（名古屋大）

◆目的：母音知覚と循環的なピッチ知覚を同時に生じる音の生成
- 新しい歌唱表現や音楽表現の創出，人間の聴覚メカニズムのさらなる解明に効果的であることが期待

◆提案手法：音声のスペクトログラムの振幅操作による音生成
- 基本周波数に基づいてスペクトログラムの奇数次倍音の帯域と偶数次倍音の帯域の相対振幅を操作し，知覚されるピッチを操作

◆評価実験
- 被験者に生成音のペアを呈示，母音の種類と音の高低を判断
- 母音判別の F1-score は７割以上、音高判断の結果は Fig. 1

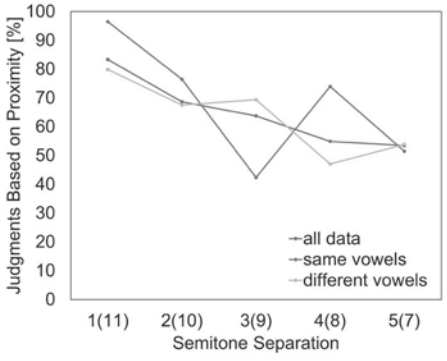


Fig.1 : Percentages of judgments based on pitch class proximity

1-R-44

1-R-44　Incorporating Hearing Loss with Binaural Speech Intelligibility Prediction for Hearing Aids

☆ Xiajie Zhou, Candy Olivia Mawalim, Masashi Unoki (JAIST)

◆ **Aim:** To investigate the integration of a hearing loss model and an Equalization-cancellation (EC) model to improve binaural speech intelligibility prediction for hearing aids.

◆ **Problem:** Individuals with hearing loss face difficulties in perceiving speech intelligibility in noisy environments due to masking effects and hearing impairment.

◆ **Solution:** The hearing loss model simulates hearing impairment for each ear based on audiograms, incorporating different parameters to model outer ear, middle ear, and cochlear damage. The EC model then processes the binaural signals using binaural cues to differentiate target speech from background noise.

◆ **Evaluation:** The prediction methods were assessed using the Clarity challenge dataset with two metrics: Pearson correlation coefficient (ρ) and root-mean-squared error (RMSE).

◆ **Summary:** Compared with HASPI, the proposed method improved ρ by 8.8% and reduced RMSE by 15.4%.

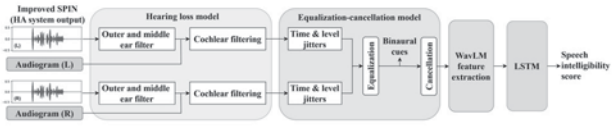


Fig.1 : Simplified overview of proposed method

1-R-45

1-R-45 Auditory numerosity perception of sound sequences with varying frequencies

○ Gerard B. REMIJN, Kana KUSUMI, Emi HASUO (Kyushu Univ.)

◆Listeners tend to underestimate the perceived number of successive, short sounds presented in rapid sequences.

◆Here we show that significant underestimation occurs even if a sequence consists of sounds with gliding or random frequencies.

◆Significant underestimation of auditory numerosity typically occurs for sequences of 8 or more sounds with a duration of 50 ms and an inter-stimulus interval of 20, 50, or 100 ms.

◆In the auditory modality, sound heterogeneity within a sequence thus does not facilitate accurate enumeration of short sound sequences.

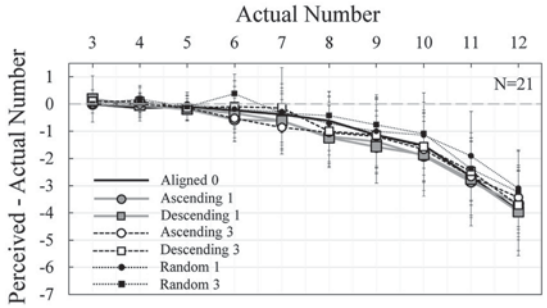


Fig.1: Mean difference between perceived number of sounds and the actual number of sounds for series with a 50-ms inter-stimulus interval. Error bars show standard deviation; 0, 1, and 3 after Aligned, Ascending, Descending, and Random represent frequency change in octaves/ sec within a series.

1-R-46

1-R-46 聴覚時間分解能検査の作成 —(8) 初期確率密度関数を二峰性とした場合の測定効率—

Tests of human auditory temporal resolution: (8) Measurement efficiency in case of using the initial probability density function of a bimodal distribution

○森本隆司(リオン)，森田健志(九州大)，山本弥生(国際医療福祉大)，小渕千絵(筑波大)，岡本康秀(済生会中央病院/慶大)，神崎晶(東京医療センター)，森周司(九州大)

◆短時間で測定可能な聴覚時間分解能測定方法の確立を目指し，ギャップ検出閾値及び振幅変調検出閾値に対して ZEST を応用する検討を進めている。

◆本報告では，初期確率密度関数に実際の閾値分布に近くなるようにするために二峰性の分布を用いた場合の測定効率について評価した。

◆分布は，健聴者および難聴者の分布をそれぞれ正規分布で表現し，それらを(9:1)もしくは(5:5)で組み合わせたもの（二峰性）を用いた。

◆若年健聴耳 30 耳，高齢健聴耳 32 耳の被験者に対し測定を行い，結果をこれまで測定してきた単峰性の測定結果と比較した。

◆若年健聴耳では，二峰性（特に(9:1)で組み合わせた条件）の方が収束は早く，効率よく測定できていた一方，高齢健聴耳では単峰性と類似もしくは収束が遅かった。

◆本実験に用いた初期確率密度関数に用いた健聴者の分布は若年者の結果であったが，本実験で測定した結果は若年健聴耳に対して高齢健聴耳の方が高かった（感度が悪かった）。

◆高齢健聴耳の結果の収束が遅かった理由の一つは，実際の分布と異なる初期確率密度関数を用いたことが考えられる。

1-R-47

1-R-47 楽音のBPMが心理的時間に及ぼす影響

The Influence of Musical BPM on Psychological Time

☆伊藤大貴，石光俊介(広島市大・情報科学)

◆人が時間をどのように感じるかは、個人や状況によって大きく異なる。その中でも、「充実時程錯覚」と呼ばれる現象は、同じ時間幅の中でも多数の出来事が生じた場合の方が実際の時間よりも長く感じられる。本研究では、楽音のBPMによる充実時程錯覚への影響を調査した。

◆本研究では、既知音源と未知音源に対する影響も調査するため、合計 4 つの実験を行った。実験方法は被験者に 2 種類の音源を聴かせ、長く感じた方を選択するサーストンの一対比較法を採用した。

◆実験 1、実験 2、実験 4 では、異なる BPM の音源を用いて、解析を行った。その結果、音楽経験が無い場合、BPM が大きいと時間が長く感じられる傾向が確認された。対して、音楽経験がある場合には BPM が小さいと時間が長く感じられる傾向が示唆された。

◆また、既知音源を扱った実験 3 では、音楽経験がある場合、元音源の BPM が最も長いと感じる傾向が確認され、BPM が元音源から離れるほど短く感じる傾向が確認された。

◆以上より、音楽経験の有無で音源の BPM によって、人間の時間感覚に影響を及ぼすことが示唆された。さらに、楽音の既知感にも時間感覚に影響を及ぼすことが示唆された。

1-R-48

1-R-48 拍の知覚の変化に必要なアクセントの検討: 3 音ごとの拍と 2 音ごとの拍の比較

Investigation of accents necessary for change in beat perception: Comparison between ternary and binary groupings

☆後藤光，松井淑恵（豊橋技科大）

◆アクセントを含まず時間的に等間隔な音は，2 音ごとのグループで拍知覚されやすいことが知られている.

◆強度アクセントによる拍知覚の変化に，2 音ごとのグループでの拍知覚への選好による影響が見られるかを調査した.

◆2 音ごと/3 音ごとに拍知覚される刺激に，拍知覚を変える強度アクセントを加えた刺激に対して，タッピングさせる実験を行った.

◆3 音ごとのグループでの拍知覚から 2 音ごとのグループでの拍知覚への変化より，2 音ごとのグループでの拍知覚から 3 音ごとのグループでの拍知覚への変化のほうが起こりやすいことが示された.

◆このことから，強度アクセントによる拍知覚の変化には，2 音ごとのグループでの拍知覚への選好による影響は見られず，むしろ 3 音ごとのグループでの拍知覚からの変化のしにくさがあることが示唆された．また，今回の実験では分析対象外の要因が拍知覚に影響することが示唆された.

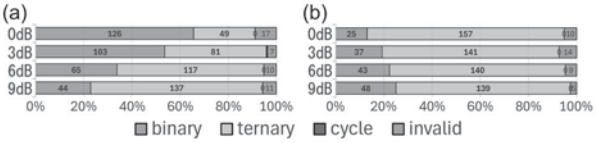


Fig. 1: Changes in beat perception when accents are added. (a) shows the results when the ternary grouping accent is added to stimuli that tend to be perceived with binary grouping. (b) shows the opposite.

1-R-49

1-R-49　テンポのずれに関する知覚実験

Perceptual experiment regarding tempo deviation

☆谷口友紀*, 藤江真也*†, 小坂直敏* 小林哲則*（*早大, †千葉工大）

◆目的：会話システムへの応用を指向して，テンポのずれに対する人の知覚特性を評価する.

◆貢献：

- 純音と発話を用いた知覚実験（Fig.1）により，音信号列の間隔の変化に対する弁別閾の調査を行なった.（Fig.2）
- 弁別閾の近似式から物理時間と心理時間との関係式を導いた.（Fig.3）

Fig.1 Perceptual experiment
(perceiving the deviation ΔT in the interval T between signals S_A and S_B)

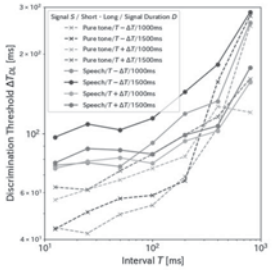


Fig.2 The Relationship Between the Interval T and the Discrimination Threshold ΔT_{DL}

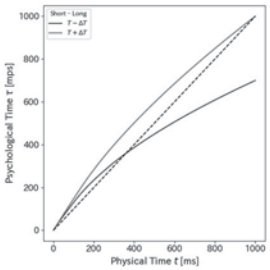


Fig.3 The Relationship Between Physical Time t and Psychological Time τ

1-R-50

1-R-50　グルーヴ知覚とリズミックな音楽における楽音の時間的包絡との関係

The Relation Between the Perception of Groove and the Temporal Envelope of Musical Tones in Rhythmical Music

○米田優一, Gerard B. REMIJN, 蓮尾絵美（九州大・芸術工）

◆グルーヴ知覚および感覚運動同期と，楽音の時間的包絡のアタック特性およびディケイ特性との関係を，主観評価と同期タッピングの実験によって調べた.

◆グルーヴの主観評価では，アタック特性に関しては 15 または 30 ms にピークを持つ逆 U 字型が表れ(Fig.1)，ディケイ特性に関しては 120 ms でピークに達しその後緩やかに減少する結果となった.

◆タッピング実験では，タップの強さはグルーヴとは逆の傾向が，安定性と正確性はグルーヴと同様な傾向が示された.

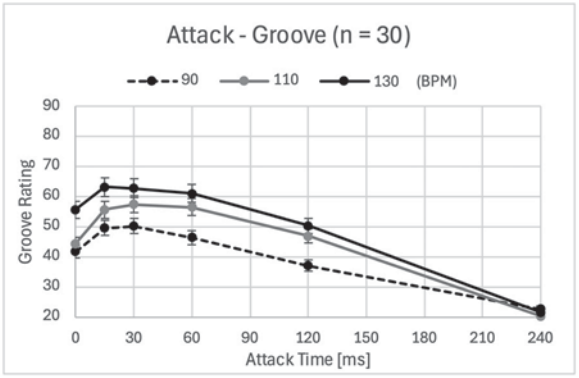


Fig.1：Relationship between attack time (ms) and groove rating (n = 30).The error bars indicate 95% confidence intervals.