1-1-1

1-1-1　オブジェクトベース音響とマルチレイヤ符号化の組み合わせによるカスタマイズ可能な放送サービスに向けたコンテンツ制作

Content production toward customizable broadcasting services using combination of object-based audio and multilayer coding of video signal

〇杉本岳大，久保弘樹，△根本慎平（NHK）

◆地上放送の高度化に向けて，オブジェクトベース音響（OBA）による放送サービスを検討してきている。

◆過去に試験的に制作した OBA コンテンツでは，音声オブジェクトが差し替わっても映像は切り替わらないという制約があった。そこで映像のマルチレイヤ符号化と OBA を組み合わせることで，音声と映像の協調動作が可能な放送システムを開発した。

◆カスタマイズ可能な放送サービスを想定した音楽コンテンツとして，従来通りに視聴する「通常モード」と，出演者のダンスの振り付けを練習できる「お手本モード」を制作した。

◆図左は「通常モード」で，コンテンツのベースとなる音楽およびベース映像による構成である。視聴者が「お手本モード」を選択すると（図右），音声はベース音楽にお手本解説音声が重畳される。映像は，ベース映像に加えてお手本ダンス・お手本解説字幕を表示する。



図　音楽コンテンツのデザイン例。左：通常モード，右：お手本モード

1-1-2

1-1-2　自己の録音音声が複数話者存在下での音声了解度に与える影響の分析

Influence of one's own recorded voice on speech intelligibility in the presence of a competing speech

◎柳田耀，井島勇祐，俵直弘（NTT）

◆家族や友人などの親しみのある声は複数話者存在下での音声了解度を高めることが先行研究において示されている [Johnsrude+, 2013][Domingo+, 2020].

◆一方，前報 [Yanagida+, 2023]では自己の録音音声に対する親近感が他人の声よりも高く評価されることが示されている．しかし，自己の録音音声に関する先行研究では自己の録音音声が複数話者存在下での音声了解度を高めるか明らかになっていない．

◆本稿では，複数話者存在下における音声了解度試験に用いられている Coordinate Response Measure (CRM) の日本語版 [Furukawa+, 2012] による実験を行い，自己の録音音声と他者の音声の音声了解度を比較した．

◆分析の結果，自己の録音音声は他者の録音音声よりも了解度が高いことが示された．



Fig.1：Response screen in the CRM task.

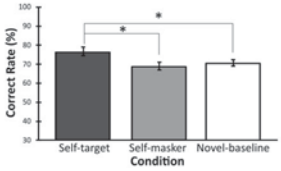


Fig.2：Correct rate of all trials

1-1-3

1-1-3　異なる向きで発話された音声の知覚 ―聴取者に対する発話者の向きの識別―

Perception of speech uttered as speaker faces different directions: Identification of speaker's facing direction from the listener

☆香嶋陽菜，荒井隆行（上智大），杉本岳大，木下光太郎，中山靖茂（NHK）

◆本研究では発話された角度に対するヒトの知覚能力を調べることを目的とした。その際に，ラウドネスを手掛かりにして話者の向きを認識できる可能性が考えられたため，話者の向きによるラウドネス変化を保った刺激と，話者の向きによらずラウドネスを一定に保ってラウドネスのキューを排除した刺激による，２つの実験（実験Ａ・実験Ｂ）を行った。

◆実験Ａと実験Ｂのそれぞれで，正面を０度とし，45 度ごと５方向（0 度，45 度，90 度，135 度，180 度）を向いて発話された音声を聞き比べ，予め開示してある放射方向のうち，どの角度の音声かを回答してもらう実験を行ったところ Fig１のような結果となった。

◆実験Ａのほうが実験Ｂよりも正答率が良いことがわかった。このことより，ラウドネスが手がかり（キュー）として発話者の向きの認識に寄与していたと考えられる。また，ラウドネスのキューが使えない場合，スペクトルの変化をキューとしていた可能性が考えられる。

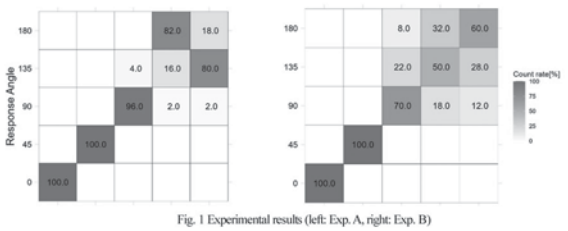


Fig. 1 Experimental results (left: Exp. A, right: Exp. B)

1-1-4

1-1-4　ポップアウトボイス生成時の調音運動のカイモグラフによる分析

Kymograph analysis of articulatory movements during pop-out voice production

☆相馬巧海，△深澤祐樹，竹本浩典（千葉工大），
北村達也（甲南大），天野成昭（愛知淑徳大）

◆背景雑音などの妨害音がある環境下でも聞き取りやすい音声をポップアウトボイスと呼ぶ。本研究では 3 名の日本語母語話者（M1, F1,F3）がポップアウトする声（PV）としない声（NV）で文章を読み上げた際の調音運動をリアルタイム MRI で記録し，カイモグラフで分析した。

◆カイモグラフから読み上げ文の PV と NV で対応する母音音素フレームごとに硬口蓋と舌の距離を求めたところ（Fig. 1），PV の/a/や/o/では口腔容積が増大し，それ以外の母音では口腔容積が減少したが（Table 1），全体として口腔容積が増大した。

◆それらの先行子音は個人差が大きかったが，共通して先行子音が/m/で後続母音が/a/，/o/の場合に口腔容積が顕著に増大し，逆に先行子音が/b/, /n/, /k/で後続母音が/i/, /u/, /e/の場合に口腔容積が減少した。

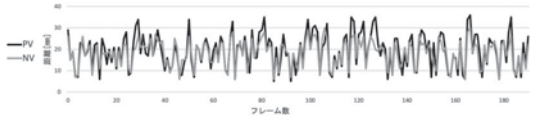


Fig 1 M1 の硬口蓋と舌の距離

Table 1 PV と NV の硬口蓋と舌の距離の差（mm）の平均。括弧の中の数値は出現回数

	/a/(73)	/i/(46)	/u/(13)	/e/(13)	/o/(40)
M1	1.38	-0.98	-1.53	-0.83	1.37
F1	1.05	-0.31	-1.75	-0.03	0.44
F3	1.54	-1.13	-0.63	-2.43	0.90

1-1-5

1-1-5　声道模型による音声のダイナミックスを再現する試み

PC-controlled vocal-tract models for speech dynamics

〇荒井隆行・鈴木良平・△EARP, Chandler・辻慎也（上智大・理工）

◆音声ダイナミックスを実現する３つのPC制御型システムを検討した.
◆梅田・寺西式声道模型によるシステム（Fig. 1）:
- 角棒の位置をアクチュエータによって制御
  → 日本語の「おはよう」を実現

◆スライド式声道模型によるシステム（Fig. 2）:
- スライド部の丸棒がパイプ中をアクチュエータによって移動
  → 英語の “How are you?” を実現

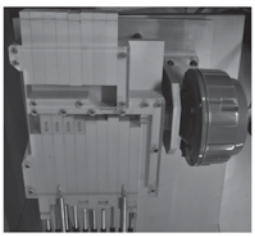


Fig. 1: PC-controlled VTM-UT system.

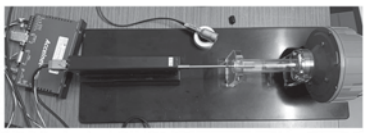


Fig. 2: PC-controlled VTM-S24 system.

1-1-6

1-1-6　日本人ドイツ語学習者の弱化母音の生成
—その他外国語学習経験がドイツ語学習に与える影響—

Production of vowel reduction by Japanese German learners
- The impact of English on the learning of German -

〇粕谷麻里乃（東邦音大・音楽），荒井隆行（上智大・理工）

◆背景：多くの日本語母語話者の L3 学習に英語の影響は排除できない．それがドイツ語の場合，両外国語は語族も言語リズムも同じだが，弱化の出現，疑問・肯定文での F0 遷移は大きく異なる[2].
◆目的：日本人 L3 ドイツ語学習者の発話にみられる英語の影響を定量的に評価するための予備実験として実施した.
◆結果：F0 持続時間，変化率，変動率すべてにおいて英語母語話者に類似した．F1, F2 は母語の影響が強くみられた.

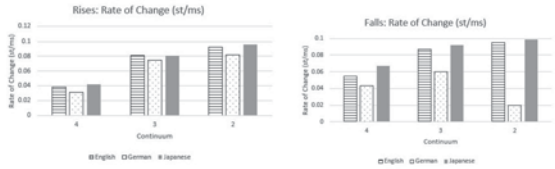


Fig.3: Rate of change (st/ms) for rises and falls.
(x: Number of Syllables. The three bars are English, German, and Japanese from left to right.)

◆考察：**F0:** ドイツ語母語話者は疑問文の時 compression 傾向を示した (Grabe[2] 支持)．日本語母語話者は疑問文・肯定文共に英語母語話者とほぼ等しく compression 傾向を示した．**F1/F2:** 英語母語話者はドイツ語母語話者の音韻的許容を超えて母音を弱化させた．日本語母語話者は大半を完全母音で発話した一方，英語に類似する語は英語のように発音していた.

1-1-7

1-1-7　音響的に多様な英語音声に対する聴取の円滑さに関する多様性分析

Analysis of diversity in learners' listening accuracy for acoustically diverse English speech

☆ 高英翔, 冨田優, 孫海童, 峯松信明, 齋藤大輔（東大）

◆ 英語教育の現場では、**綺麗な**英語音声が多い

➢実際に使う場合に、**環境音や訛り**は必ずある

◆ 東大の新たな英語授業で、多様な英語音声をシャドーイング課題として提供

➢雑音・歪み・変形、世界諸英語、感情音声

◆ 学習者の「聴取崩れ」を計測

➢日本語を母語とする学習者に見られる聴取の円滑さの**多様性**について分析

◇ **全体的な聴取崩れの傾向**

◇ **聴取流暢性に見られる学習者間の多様性**

◇ **クラスタリングによる聴取多様性の可視化**

1-1-8

1-1-8　日本人小学生による英語暗唱音声における語彙強勢位置の自動推定と母語話者評価

Automatic estimation and native speakers' evaluation of lexical stress positions in English recitation speech produced by Japanese elementary school children

☆木村 竜也, Sylvain Coulange, 加藤 恒夫（同志社大）

◆日本人小学生による英語暗唱音声を対象に，最新の音声認識技術を用いた語彙強勢位置自動推定パイプラインを適用するとともに英語母語話者に強勢位置の判定をしてもらい，両者の結果を比較分析した.
◆強勢誤り検出率に着目し比較した結果，パイプライン処理による検出率の平均は 34.3%で，母語話者評価の 11.8%よりも高い結果となった.
◆母語話者評価においては語彙強勢の知覚判断に評価者間の大きなばらつきがみられた.
◆パイプライン処理による第 1 強勢コントラストと母語話者評価による語彙強勢の正解判定の割合の間で Pearson の積率相関係数を測ると，0.29 と弱い正の相関がみられた.

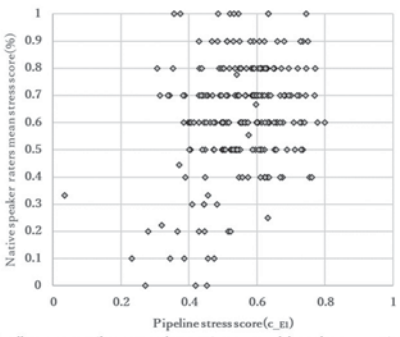


Fig.1: Pipeline vs. native speaker raters word-level mean stress score

1-1-9

講演取消

1-1-10

1-1-10　なぜパルス列音源では自然な音声を合成できないのか？

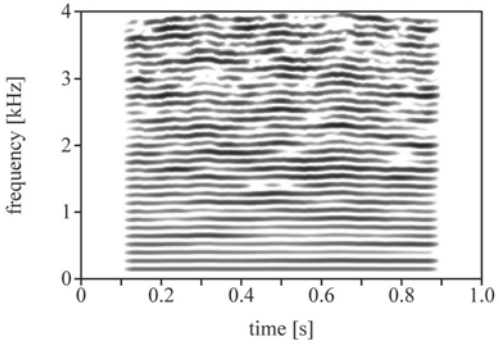
Why cannot pulse train source signals generate natural-sounding speech?

○青木 直史（北大）

■ ソースフィルタ理論にもとづく音声合成では，有声音の場合，パルス列を入力信号とすることが第一近似とされている。

■ しかし，このままでは自然性が低下することが指摘されており，さまざまな改善策がこれまで盛んに検討されてきた。

■ 本研究は，パルス列音源にかわる入力信号を考案し，音の艶ともいえる肉声感を再現する方法について検討している。



提案法による音源信号のスペクトログラム

1-1-11

1-1-11　リアルタイム MRI を用いたオペラ歌唱における音高と頸椎後弯の度合いの分析

Analysis of relationship between pitches and bending of cervical spines in operatic singing using real-time MRI

◎戸田菜月，竹本浩典（千葉工大），高橋純（大阪芸大）

◆オペラ歌手は，声区融合と呼ばれる技術などを用いて，高音でも厚みのある歌声を実現しているが，その制御は明らかになっていない。

◆また先行研究で，高音発声時に頸椎が後弯する傾向が見られてきたが，その度合いと音高の関係は明らかでない。

◆本研究では，テノール歌手１名が音高を変化させて歌唱する際の声道形状と頸椎の後弯の度合いを主成分分析した。

◆その結果，頸椎の後弯の度合いは音高と相関しており（Fig. 1），この後弯を利用して，音高を制御するメカニズムが想定された。

◆また声道形状の分析結果より，このメカニズムは中音域の上行，高音域の上行・下行の双方で用いていると推定された。

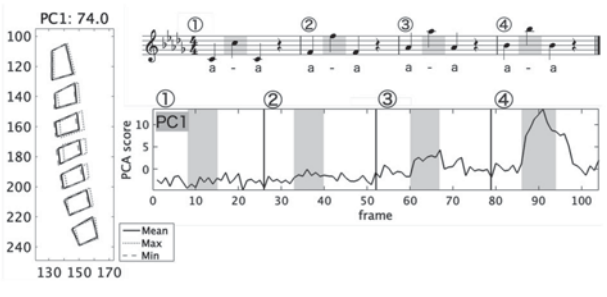


Fig.1：頸椎の主成分分析の結果（左）とその主成分スコア（右）

1-1-12

1-1-12　アカゲザルの声帯物理モデルを用いた声帯膜振動に関する研究

Study on vocal membrane oscillation using a vocal fold physical model of rhesus macaque (*Macaca mulatta*)

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

◆サル類の摘出喉頭実験において，声帯の冠状面における傾きを大きくした際，ヒト以外の霊長類に特有の“声帯膜”のみの振動を確認した。

◆アカゲザルの摘出喉頭のマイクロCT画像から作製された物理モデルを使用し，声帯の冠状面での傾きがアカゲザルの発声に及ぼす影響について調べた。

◆声帯の角度が異なる物理モデル（0° ～ 15°）を作製し，物理モデルに与える流量を0～80 L/min まで 15 秒間かけて増加させ，状態の変化を調べた。

◆同様の条件でヒト型声帯物理モデルを作製し，両モデルの比較を行った。

◆声帯の傾きが大きい条件で声帯膜のみの振動を観測し，流量を増加させると，振動の複雑化や振動モードの変化がみられた。また，ヒト型声帯ではこれらが起こらないことを確認した。

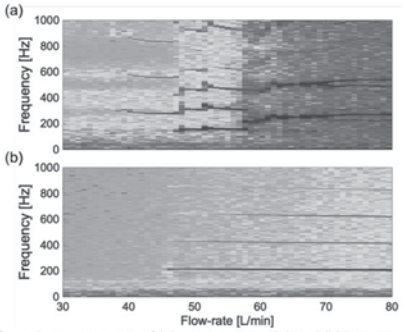


Fig.1：Sound spectrograms of (a) macaque model and (b) human model.

1-1-13

1-1-13　ウェアラブルな構音障害支援デバイスにおける制御インタフェースと音声補完の検討

Study on control interface and voice augmentation in wearable articulation disorder assisting device.

〇藪謙一郎(東大)，田中敏明(東大･北科大)，伊福部達(東大)

◆筆者らは、構音機能について、指の位置や動きで声を楽器のように声を操作できる音声生成器を開発し、残存する発声機能を用いて、リアルタイムにホルマント付加し音韻を補完する試作器を示してきた。

◆本発表では、音韻の補完に加え、話速を補正することにより操作性を高めるシステムを開発し、動作を検証したので報告する。

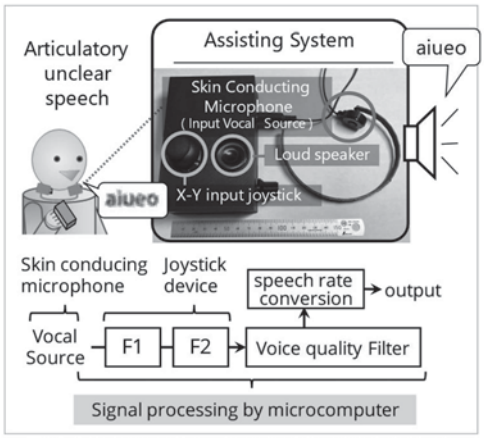


Fig.1 : A concept and a prototype of the assisting system

1-1-14

1-1-14　言語聴覚士のコミュニケーションのための音環境調査

A survey of noise environment in communication for speech therapists

〇竹内京子(順天堂大)，青木直史(北大)，荒井隆行(上智大)，△鈴木恵子(北里大)，世木秀明(千葉工大)，秦若菜(北里大)，安啓一(筑波技術大)

◆ 言語聴覚士はことばのリハビリを行う職業。養成校では，音響学・聴覚心理学が必修科目であり，学生にとって最も苦手で嫌いな科目。

◆ 改善策として，現役言語聴覚士対象に「ＳＴのための音響学」という講習会を開催。音声の音響分析に必要な録音環境の調査も行ってきた。

◆ 「音環境とその聴こえ」も，重要な学習項目で，一般的には補聴器や人工内耳を使用時の聴こえと結びつけられる。しかし，臨床のコミュニケーションも音環境によって変化するとしたら，今後，学ぶ必要があるだろう。

◆ 第 23 回の講習会で，上野佳奈子先生（明治大学）による招待講演：ST 室の音響環境を考えよう！を開催。その際に，参加者に ST の臨床現場の音環境調査を行った。本発表はその結果の報告である。

◆ その結果，ST 室の音環境が気になっている者は多く，改善策を探しているが，なかなか具体的に実現できない現状があった。

◆ 講習会で，今まで，気づかなかった音環境に対する視点，部屋の反響をなくすことが，コミュニケーションに有効でありそうだという情報を得られたことは大きい。

◆ 今後は，具体的な実践が待たれるとともに，養成校の音響学の授業でもこの話題によって，学生が臨床への応用をイメージできるように努めたい。

1-1-15

1-1-15　鼻中隔切除が鼻音スペクトルに与える影響

Effects of nasal septum resection on spectrum of nasals.

☆福島侑希，竹本浩典(千葉工大)，北村達也(甲南大)，
△鴻信義，△武田鉄平，△今川記恵，△小黒亮史，△大村和弘(慈恵医大)

◆鼻副鼻腔の悪性腫瘍切除手術では，鼻中隔切除と同時に周囲の副鼻腔の開放が行われる。しかし，これらの形態変化が鼻音スペクトルにどのような影響を与えるのか明らかでない。

◆本研究では，健常者２名（NF1，NF2）の CT 画像から鼻副鼻腔形状の抽出と模擬手術を行った。抽出のみを行ったモデル（ORG），鼻中隔を切除したモデル(SP)，篩骨洞を開放したモデル(ES)の３種類，計６つのモデルで音響シミュレーションを行い，鼻中隔切除と篩骨洞の開放が鼻音スペクトルに及ぼす影響を検討した。

◆SP の伝達関数では，両被験者共に第１，２ピーク（P1，P2）で音圧レベルが上昇した。これは鼻腔容積が増大したためと思われる。

◆ES の伝達関数では，両被験者共に音圧レベルが下降し，P1 と P2 の間にディップを生じた。瞬時音圧分布の分析結果によれば，これらのディップは上顎洞や前頭洞の共鳴に由来していた。

◆同様のディップは実際の手術患者の音声スペクトルでも観測される。本研究の結果から，これらは篩骨洞の開放によって上顎洞や前頭洞の共鳴が増大して生じたと考えられる。

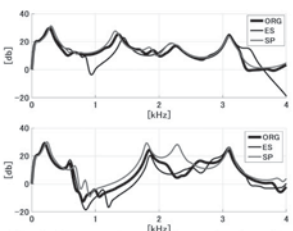


Fig.1 : Comparison of transfer functions (NF1:upper, NF2:lower)

1-1-16

1-1-16 音響管を用いた代用発声における音源位置によるフォルマント変化のシミュレーション

Simulation of formant change with sound source position in substitutional speech using acoustic tubes

☆小野寺悦希（筑波大・院・シス情工），若槻尚斗，
海老原格，△前田祐佳（筑波大・シス情系）

著者らは，音響管を口にくわえ音源を声道に伝達する発声法に注目し，声道モデルを用いた計測により音韻が保持される音源位置が硬口蓋付近であると考察した (Fig.1)。
しかし本考察は使用したモデルに依存している可能性がある。
・音源位置とフォルマントの一般的な関係の調査
・提案手法が通常の音韻・声質を保ちうるのかどうかの検証

(i) **シミュレーション**
声道を模した閉管内の音源位置と放射音の関係性を調査

(ii) **提案手法による発声と通常の発声の録音**
母音ごとのフォルマントの比較，発声しやすさの評価

結果・考察

(i) 定在波の音圧の腹節が，放射音の音圧の谷に一致 (Fig.2)

(ii) 提案手法では，硬口蓋辺りを狭める発声を共鳴させにくい
全フォルマントの保持，音響管をくわえることの不都合が
現状の課題だと明らかになった。

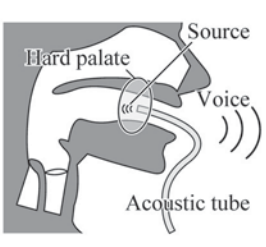


Fig.1 : Substitutional speech using acoustic tube.

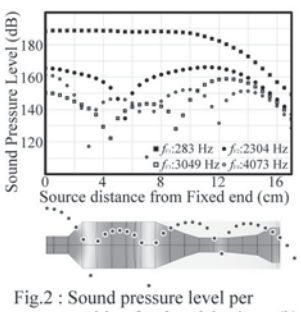


Fig.2 : Sound pressure level per source position for $f_{F1} - f_{F4}$ in shape /i/.

1-1-17

1-1-17　一側性難聴者の残響下での音声聴取に関わる要因の検討

An investigation of what factors affect speech recognition under reverberation for individuals with unilateral hearing loss

○辻慎也，荒井隆行（上智大・理工）

◆一側性難聴では，以下の場面が困難になることが指摘されてきた．
１）難聴側聴取，２）騒音下聴取，３）音源定位

◆我々が過去に行った聴取実験の結果，特に一側性難聴の発症直後，残響時間の長い環境において音声や音楽の聞こえが影響されることが新たに明らかになった．

◆本稿では，一側性難聴者の残響下聴取に関わる要因を検討した結果，次の結果を得た (Fig. 1).

A) 無響下では有意でなかった一方で，残響下で特に音声が患耳側にあるとき，年齢を重ねるほど音声明瞭度が低く，方向性マスキング解除（SRM）の程度が低い傾向がみられた．

B) ゴールドスミス音楽洗練指標で得た音楽的能力のスコアが高いほど音声明瞭度・SRM の程度が高くなる傾向がみられた．残響下では音楽的能力が一側性難聴者の聞こえに寄与し得ると示唆された．

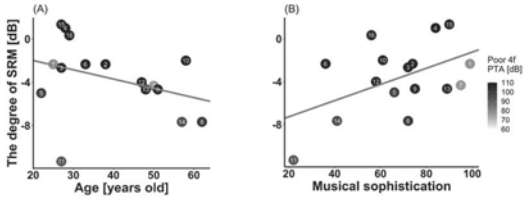


Fig. 1 Association between spatial release from masking (SRM) and participants' demographics under reverberation (contralateral condition: target speech was located on the poor ear side). Shading of the individual data points indicates four frequency pure-tone averages (4fPTA) of the participants' poor ear.

1-2-1

1-2-1　LLM によるテキスト生成を用いた音声合成による音声認識のドメイン適応

Domain adaptation for speech recognition based on speech synthesis using LLM

◎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，選択するテキストデータの質がが音声認識の性能改善に必要である．

◆ 本研究では，LLM を用いてテキスト生成を行い，ドメイン適応を行う．

◆ LLM への入力には，LibriSpeech のテキストデータ，変換を行うためのプロンプトを用い，それらに加えて，変換元のテキストデータのドメイン情報 (読み上げ)，変換後のテキストデータのドメイン情報 (lecture-style) を入力する．

◆ 実験結果により，LLM によるテキスト拡張は，LibriSpeech の言語モデル学習データから抽出したテキストや，単語を置き換えるテキスト拡張手法と比較して，性能が改善した．

◆ ドメイン指示による性能の改善が見られなかったが，変換後のテキストデータのドメイン指示を変更することにより性能が変動した．

Table 1： ASR performance (WER (%)) for TED-LIUM release-2 testset. We used 100-hour paired data from LibriSpeech train-clean-100.

Model	dev	test
No augmented text (real: 28.1k sent., generated: 281k sent.)	2.90	7.14
Augmented model (real: 100h, generated: 281k + 562k sent.)		
PPDB + RoBERTa	18.32	18.37
Official LM data in LibriSpeech	18.70	18.97
Basic Prompt	17.24	**17.10**
+ Source information	16.96	17.28
+ Target information	17.24	17.40
+ Both	**16.72**	17.40

1-2-2

多言語 SSL モデルに基づく音声認識の大規模音声データのみを用いたドメイン適応

○木内 貴浩(豊橋技科大) 小川 厚徳(NTT) 若林 佑幸(豊橋技科大)
太田 健吾(阿南高専) 北岡 教英(豊橋技科大)

目的

◆ ASR モデルの精度は，学習データのドメインに大きく依存

◆ドメイン内音声とその書き起こしデータの準備は高コスト

◆ドメイン内の**音声データのみ**であれば低コストで収集可能

手法

◆自己教師あり学習モデルをドメイン内音声で追加事前学習し，ドメイン外のパラレルデータで Fine-tuning する事によりドメイン内のパラレルデータを一切使わずにドメイン適応

◆先行研究 [1] の wav2vec 2.0 Large を用いた実験に加えて，日本語も含まれた多言語音声で事前学習されたモデル XLS-R で実験

結果

◆XLS-R を用いた場合，ドメイン内コーパスを用いない場合(Baseline)と比べて，提案手法（TgtAdap）は CER において 3.9 pts 精度改善

Table 1: Character Error Rates (%) of baseline, comparison, and proposed domain-adapted models, where Prop. and W2V2-L denotes proposed method and wav2vec 2.0 large, respectively.

Model name	Encoder model	Further training (speech only)	CTC fine-tuning (speech & text)	CSJ eval1	CSJ eval2	CSJ eval3	LTVS dev
Ideal	W2V2-L	(N/A)	LTVS	15.6	21.8	24.0	13.8
Baseline	W2V2-L	(N/A)	CSJ	6.0	4.4	5.1	27.4
SrcAdap	W2V2-L	CSJ	CSJ	4.9	3.8	4.4	30.3
TgtAdap (Prop.)	W2V2-L	LTVS	CSJ	5.2	4.1	4.6	**19.8**
Ideal	XLS-R	(N/A)	LTVS	21.1	21.1	14.0	10.7
Baseline	XLS-R	(N/A)	CSJ	4.8	3.7	4.2	24.1
SrcAdap	XLS-R	CSJ	CSJ	5.1	3.9	4.3	28.8
TgtAdap (Prop.)	XLS-R	LTVS	CSJ	4.4	3.5	3.6	**20.2**

[1] 木内貴浩, 小川厚徳, 若林佑幸, 北岡教英, "SSL に基づく音声データのみを用いた音声認識モデルのドメイン適応, " 日本音響学会秋季研究発表 会, pp. 2-Q-9, Sep. 2023.

1-2-3

1-2-3　Investigation of Cross-Lingual Mismatch in Low-resource ASR for Indonesian Ethnic Languages

インドネシアの民族語に向けた低資源 ASR における言語間不一致の調査

Sakriani Sakti (JAIST), ○ Benita Angela Titalim (JAIST)

◆Introduce Indonesian ethnic languages dataset for ML-SUPERB Challenge on the New-language track.

◆Investigate various mismatch conditions in low-resource speech recognition tasks for those languages within the context of pretrained self-supervised learning (SSL) models.

◆The results highlight the performance of the pretrained SSL model, demonstrating that as the data becomes more similar, the performance of the model improves.

Test set / Model	eng [2]	ind [2]	ind (ours)	ban (ours)	btk (ours)	jav [2]	jav (ours)	sun [2]	sun (ours)
eng [2]	42.8	63.8	60.5	58.7	51.7	66.5	60.1	59.4	56.9
ind [2]	58.5	23.9	25.2	22.3	25.0	33.5	28.6	29.7	23.9
ind (ours)	59.4	24.2	17.0	22.4	25.8	34.5	30.4	29.6	24.7
ban (ours)	60.8	32.4	22.3	11.1	25.0	33.8	23.0	29.3	20.6
btk (ours)	62.5	38.8	30.9	30.9	14.1	44.8	38.9	38.7	28.6
jav [2]	60.7	29.8	25.7	22.2	30.9	25.1	20.0	33.4	27.0
jav (ours)	61.6	30.5	27.1	24.2	32.4	26.4	13.8	35.8	27.8
sun [2]	67.0	44.5	39.3	34.1	38.3	48.4	42.5	29.4	35.2
sun (ours)	60.4	29.8	34.2	18.4	23.5	37.4	22.6	24.3	15.3
multi	5.9	60.7	57.6	58.6	61.8	25.7	27.8	58.1	61.2

Table 1:Cross- lingual evaluation of each ethnic languages

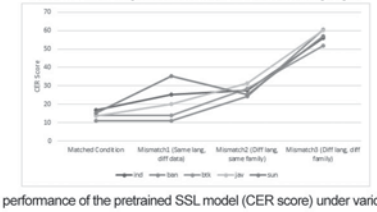


Fig.1:The performance of the pretrained SSL model (CER score) under various conditions

1-2-4

1-2-4　EARS: 日本人超高齢者音声コーパスの構築

EARS: Construction of the speech corpus of Japanese Super-elderlies
○福田芽衣子(豊橋技科大)，西村良太(徳島大)，
入部百合絵(愛知県立大)，山本一公(中部大)，北岡教英(豊橋技科大)

◆超高齢化社会の現在、介護ロボットの開発・導入が急務で、これには超高齢者音声認識技術の向上が不可欠
◆　しかし、日本人超高齢者音声の大規模コーパスは無く、若年層の音声と比べ十分な認識精度が得られていない
◆　123 名分(13.4 時間)の読み上げ音声とその書き起こしテキスト整備をし一般公開
◆　今後も収録を重ねデータ量を拡大する予定
◆　コーパスの詳細と超高齢者音声認識への効果を紹介

Table 6　Results of speech recognition experiments

model	Trainig data	CER(%)
ESPnet2	BL1	8.6
	BL+EARS	**8.3**
NeMo	BL2	11.56
(LSTM-CTC)	FT	10.52
NeMo	BL2	10.50
(Conformer-CTC)	FT	9.06
Whisper	BL3	18.66
(Large model)	FT	11.33
Whisper-v03	BL3	10.91
(Large model)	FT	11.70

BL1: trained using only JNAS, S-JNAS and CSJ speech data
BL2: trained using only JNAS, CSJ, CEJC and LaboroTV speech data
BL3: provided by OpenAI
FT: Fine tuning using EARS

1-2-5

1-2-5　年齢推定タスクの併用による高齢者音声認識の高精度化

Improving Accuracy of Speech Recognition for the Elderly by Combining Age Estimation Tasks

☆新惠拓実，木内貴浩，若林佑幸，北岡教英（豊橋技科大）

目的

◆ 高齢者の音声は一般成人と比較すると，認識を高精度に行うのが難しく，データを大量に集めることも難しい．
◆ 高齢者音声に対し高精度な認識が可能になれば，電子カルテ自動入力システムや認知症予防対話ロボットなど，多くの応用が期待できる．

モデル化

◆ E2E マルチタスク学習モデル（Enc2age,Dec2age）：音声認識と年齢推定の二つのタスクを学習し，年齢情報を加味した音声認識モデルの構築を行い，精度の向上を目指す手法である．
◆ E2E 敵対的マルチタスク学習モデル（Enc2discr）：敵対的学習を用いて年齢推定に必要な情報を取り除き年齢不変な特徴を抽出し音声認識を行い，精度の向上を図る．

結果

◆ 結果から，敵対的マルチタスク学習モデルを用いることで一般成人音声に対しては認識精度を保ちつつ，高齢者音声に対する認識精度は向上した．

Table 4: Comparison of CER among baseline and three multi-task learning models.

Model	CSJ	EARS
Baseline	4.5	8.3
Enc2age	4.9	8.0
Dec2age	5.2	8.0
Enc2discr	4.5	**7.2**

1-2-6

1-2-6　表現力豊かな音声合成に向けたテキストからのスタイルトークン埋め込み予測の強調

Enhancement of text-predicted style token with generative adversarial network for expressive speech synthesis.

◎金川裕紀、井島勇祐（NTT）

◆ グローバル・スタイルトークン（GST）に基づくスタイル音声合成
- 参照音声から発話様式（スタイル）の埋め込み表現である GST を抽出
- スタイル情報として GST を用いて音声合成モデルを条件付け
  ➔ 推論時にも参照音声が必要☹
◆ テキストからのスタイル埋め込みの予測（TPGST）
- 学習時：テキストからスタイル埋め込みを予測するスタイル予測器を学習
- 推論時：予測スタイル埋め込みで GST を代用 ➔ 推論時は参照音声不要☺
- L1 損失最小化に基づくスタイル予測器の学習では、複雑なスタイルの分布をモデル化できず
  ➢ **スタイル埋め込みが過剰に平滑化** ➔ スタイル再現性が低下☹
◆ **提案法：スタイル予測器の学習に敵対的生成ネットワーク（GAN）を導入**
- L1 損失最小化に基づく従来の TPGST では獲得できない複雑なスタイルの分布をモデル化するため、 **GAN の枠組みでスタイル予測器を学習**
- 提案法のスタイル再現性は従来 TPGST より優れ、オラクルである**抽出した GST のそれと同等**

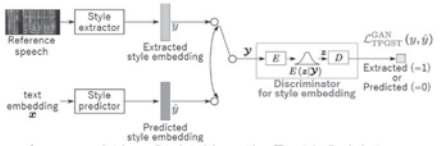


Fig. 1: A diagram of our proposed style predictor's training section. The style discriminator distinguishes the style embeddings $\mathcal{Y}$ obtained from style extractor and style predictor as either real or fake. The style predictor yields more accurate prediction than the style embedding to predicts more fluent style embeddings that the discriminator accepts as genuine.

1-2-7

1-2-7　Maintaining Personal Styles in Multilingual TTS with STEN Approach in Diffusion Framework

☆ Chung Tran (JAIST), Chi Mai Luong (IOIT), Sakriani Sakti (JAIST)

◆Introduces **Style-Enhanced Normalization TTS** (STEN-TTS), a **novel diffusion-based text-to-speech system**, that can:
- ➢ preserves individual voice characteristics
- ➢ enables cross-lingual voice cloning from just a short audio.

◆The proposed model includes three modules:
- ➢ **Style Encoder:** An audio is converted to Mel-spectrogram, and is fed into this module to extract speaker representation.
- ➢ **Synthesizer**: The input is text, language ID, and speaker representation, and the output is an auxilary Mel-spectrogram.
- ➢ **STEN:** The module executes the diffusion on the auxiliary Mel-spectrogram with conditions text encoder information and speaker representation.

◆The experimental results show that STEN-TTS achieves good performance, at 3.44 on SMOS for cross-lingual switching.

Table 1　SMOS for cross-lingual adaptation

	↑ StyleSpeech					↑ DiffSinger					↑ STEN-TTS				
	VN	JA	ZH	ID	EN	VN	JA	ZH	ID	EN	VN	JA	ZH	ID	EN
VN-to*	3.5	3.2	3.1	3.3	**3.7**	**3.8**	**3.7**	**3.9**	3.4	2.8	3.6	3.5	3.8	**3.5**	3.6
JA-to*	3.3	2.1	2.7	2.9	2.8	**3.4**	**2.3**	**3.0**	2.7	2.2	3.2	2.3	2.8	**2.9**	**3.6**
ZH-to-*	3.2	3.6	3.6	3.2	3.3	3.3	3.6	3.7	3.4	2.7	**3.4**	**3.9**	**3.8**	**3.8**	**3.9**
ID-to*	3.0	2.9	3.1	3.0	3.4	**3.2**	3.5	3.1	3.3	2.6	3.2	**3.6**	**3.7**	**3.5**	**3.4**
EN-to*	2.6	2.7	3.0	3.1	3.7	**3.4**	3.3	2.9	3.3	2.7	3.1	**3.3**	**3.3**	**3.4**	**4.0**
Agg	3.1	2.9	3.1	3.1	3.4	3.4	3.3	3.3	3.2	2.6	3.3	3.3	3.5	3.4	3.7
	3.13±0.15					3.16±0.15					3.44±0.17				

1-2-8

1-2-8 Emotion-controllable Speech Synthesis using Emotion Soft Label and Word-level Prominence

○羅 旋, 高道 慎之介, 齋藤 佑樹, 猿渡 洋(東大院･情報理工)

◆We propose a two-stage emotion-controllable TTS model that we can condition on emotion soft labels and fine-condition on word-level prominence, which conventional models cannot. Our proposed model extends the Tacotron2 model with a speech emotion recognizer (SER) and a prominence predictor (PP) to enable this dual controllability. In the first stage, we condition on emotion soft labels predicted by the SER. In the second stage, we fine-condition on the prominence predicted by the PP model, as shown in Fig. 1 (left).

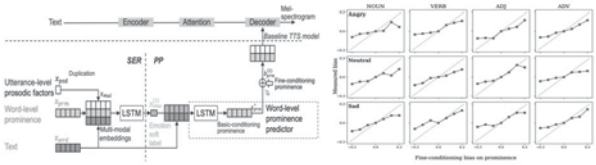


Fig. 1. Model architect (left). Ojective evaluation (center).

◆According to the objective and subjective evaluation. The experiments achieved 1) 51¥% emotion-distinguishable accuracy, and 2) 0.95 linear controllability on prominence, as shown in Fig. 1 (right). In addition, our model shows comparable audio quality performance (MOS=3.9) of synthesized speech to the conventional models that can only control emotion.

1-2-9

1-2-9 大規模な日本語笑い声コーパスを用いたテキストレス笑い声合成

Textless Laughter Synthesis with a Large-scale Japanese Laughter Corpus

☆辛 徳泰, 高道 慎之介, 森松 亜依, 猿渡 洋(東大院･情報理工)

◆笑い声は喜びなどの感情を表せ、音声の表現力を高められる重要な非言語的発声の一種です。

◆本研究はまずYouTubeから集めた大規模な日本語笑い声コーパスを提案する。

◆次に、自己教師あり学習モデルに基づいたテキストレスな笑い声合成の方法を提案する。

◆提案法は、HuBERTが出力した笑い声の連続的な特徴でK-meansクラスタリングモデルを学習させ、そのモデルを用いて笑い声を離散的な表現に変換する。

◆実験結果により、提案したテキストレスな表現は音素表現より優れたパフォーマンスを得たことを示す。

◆更に、テキストレスな笑い声表現で学習させた言語モデルをサンプリングすることにより、笑い声を生成できることを示す。

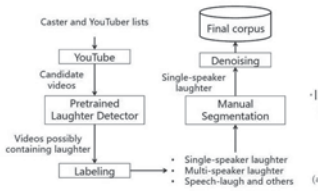


Fig.1: Data-collection process.

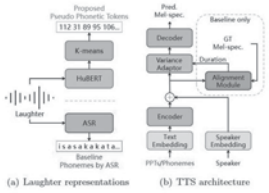


Fig.2: (a): laughter representations; (b): architecture of the TTS model.

1-2-10

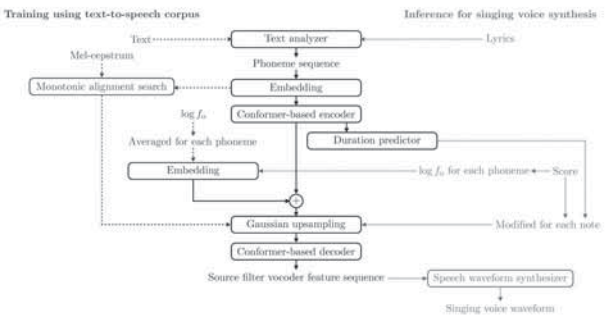
1-2-10 正弦波入力型ニューラルボコーダを用いたTTSモデルによる歌声合成

Investigation of singing voice synthesis using TTS models with sinusoidal input-based neural vocoders

☆清水聡太[1,2], 岡本拓磨[2], 高島遼一[1], 大谷大和[2], 滝口哲也[1], 戸田智基[3,2], 河井恒[1]
([1]神戸大学, [2]情報通信研究機構, [3]名古屋大学)

・テキスト音声コーパスのみを用いた歌声合成モデル
- ・TTSコーパスを用いた歌声合成音響モデルの提案
- ・基本周波数および話速を高精度に制御する必要性
  - ・正弦波入力型ニューラルボコーダの導入
    - ・Harmonic-Net+ (K. Matsubara+ IEEE/ACM TASLP 2023)
    - ・SiFi-GAN (R. Yoneyama+ ICASSP 2023)

→ 従来のHiFi-GANよりも高精度な基本周波数制御を実現



TTSコーパスを用いて学習した歌声合成音響モデル

1-2-11

1-2-11 AAS-VC:非自己回帰型系列音声変換における時間対応付け学習の頑健性

AAS-VC: On the Generalization Ability of Automatic Alignment Search based Non-autoregressive Sequence-to-sequence Voice Conversion

○ Huang Wen-Chin[1], 小林和弘[1,2], 戸田智基[1] ([1]名古屋大学, [2]TARVO)

◆非自己回帰型系列VCモデルは，韻律など話者特性を捉えつつ高速かつ安定した生成を実現することができる．

◆従来法のFS2-VCは、教師モデルとして用いる自己回帰型系列VCモデルの品質に大きく左右され、学習データ量の影響を大きく受ける．

◆Automatic alignment search（AAS; 自動アライメント探索）手法に基づくAAS-VCは，パラレルデータから直接に正解アライメントを求める．

◆実験評価によって，学習データが少ない設定においてFS2-VCより良い性能が達成できることを確認した．5分程度のパラレルデータのみを用いた場合においても、系列変換モデルの学習が可能である．

Table 1: Subjective evaluation results of different seq2seq VC systems in various training data sizes. Bold face indicates the top performing system(s) in each column (multiple bolded systems indicate that there is no significant difference in between).

System	Number of training utterances (duration)			
	80 (5 mins)		932 (1 hr)	
	Nat. ↑	Sim. ↑	Nat. ↑	Sim. ↑
GT	-		4.42±0.09	81%±3%
ANA-SYN	-		3.72±0.12	76%±4%
FS2-VC (No PT)	1.84±0.11	41%±4%	**3.62±0.11**	**65%±4%**
FS2-VC (PT)	2.34±0.11	**50%±4%**	**3.52±0.11**	**67%±4%**
AAS-VC	**2.80±0.12**	**51%±4%**	3.47±0.12	**65%±4%**

1-2-12

1-2-12　Eden-VC:音素継続長とアライメントの協調学習を用いた系列長変換型声質変換モデル

Eden-VC:sequence-to-sequence voice conversion model with collaborative duration-alignment learning

◎山下陽生(神戸大/NICT), 岡本拓磨(NICT), 高島遼一(神戸大),
大谷大和(NICT), 滝口哲也(神戸大), 戸田智基(名大), 河井恒(NICT)

◆L2 話者から L1 話者への声質変換といった複雑な問題では、従来法である Monotonic Alignment Search (MAS)では柔軟性がなく正確なアライメントが生成できず、変換精度が劣化してしまう問題がある。

◆そこで、本研究では scale-dot attention に基づくアライメントを用いた TTS モデルである EdenTTS を VC モデルに適用した Eden-VC を提案し、変換精度を MAS と比較する。

◆自然性、CER、など多くの評価指標において、Eden-VC は MAS を使ったモデルを上回り、EdenTTS にて提案されたアライメント手法の有効性を示すことが出来た。

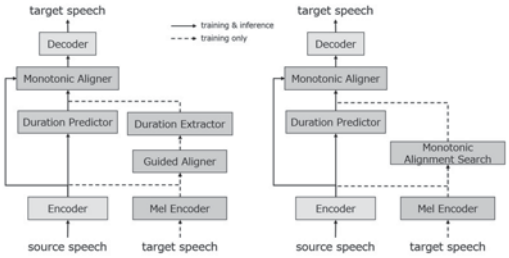


Fig.1： Network architectures of Eden-VC and Eden-VC (MAS)

1-2-13

1-2-13　Synthesis of non-native voice with native-like accent using voice conversion

○ Iago Lourenco Correa, Sei Ueno, Akinobu Lee
(Nagoya Institute of Technology)

◆Create “golden speaker” for self-imitation
- Voice conversion (VC): convert source speaker's voice onto a target speaker voice while preserving linguistic content

◆MaskCycleGAN-VC: cycle-consistent GAN model for pair-wise conversion
- Speaker embedding + native accent ID: speaker and accent information
- Discriminator conditioned to the desired accent
- Accent loss calculation

◆Experiments with Japanese and English
- Conditioned discriminator achieved improvement in most cases

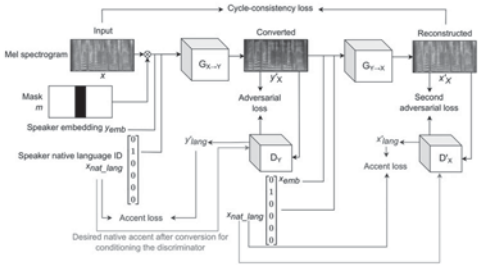


Fig.1: Illustration of one training step from *x* to *y*. Speaker embedding and native language ID, accent loss and conditioned discriminator are depicted in the figure.

1-2-14

1-2-14　CHiME7 遠隔発話音声認識タスクのための NTT システム：全体構成

NTT system for CHiME7 distant automatic speech recognition task: Overview

○加茂直之, 俵直弘, 松浦孝平, 芦原孝典, 森谷崇史,
小川厚徳, 佐藤宏, 落合翼, 安藤厚志,
池下林太郎, 叶高朋, デルクロア・マーク,
中谷智広, 浅見太一, 荒木章子（NTT）

◆CHiME-7 チャレンジの DASR タスクに投稿したシNTT ステムの概要
- DASR は遠隔発話音声認識のタスクで、複数人が会話しているシナリオで、各話者の発話内容を推定することが目的

◆システムは Diarization、音声強調、音声認識の３つのモジュールからなり（Fig.1）、Diarization、音声認識の詳細に関しては、この後に続く NTT の発表で紹介する

◆ベースラインシステムに対して、47%の Word Error Rate の削減を達成し（Table 1）、チャレンジでは３位にランクインした



Fig.1:システムのパイプライン

Table 1:各評価セットのWER結果 [%]

システム	CHiME-7	DiPCo	Mixer6	Macro
Baseline	77.4	54.7	33.7	55.3
提案法	42.2	31.1	14.6	29.3

1-2-15

1-2-15　CHiME7 遠隔発話音声認識タスクのための NTT システム:話者ダイアライゼーション

NTT Speaker Diarization System for the DASR Task of CHiME7 challenge

○俵直弘, 安藤厚志, デルクロア・マーク, 小川厚徳（NTT）

◆ CHiME7 遠隔発話音声認識タスクに投稿した NTT システムの話者ダイアライゼーションの概要(Fig.1)
1. End-to-end neural diarization with vector clustering (EEND-VC)と ECAPA-TDNN に基づく, 話者重複に頑健なアクティビティ推定(Fig.2)
2. DOVER-Lap に基づくダイアライゼーション結果のチャネル統合
3. Self-supervised adaptation(SSA)に基づくチャネル適応

◆ 提案システムは全参加チームの中で３位の性能を達成し, クラスタリングに基づくシステムとしては１位の性能を達成

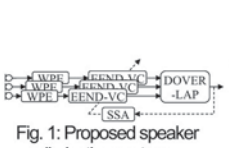


Fig. 1: Proposed speaker diarization system

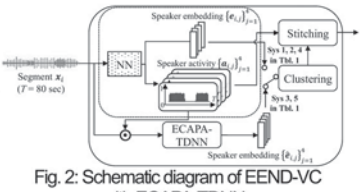


Fig. 2: Schematic diagram of EEND-VC with ECAPA-TDNN

Table: Performance comparison with top-7 participants on EVAL set

Team	Main Approach	CHiME-6 DER	DiPCO DER	Mixer 6 DER	Macro DER
USTC-NERCSLIP	TS-VAD	**25.1**	**16.4**	6.1	**15.9**
IOA-CAS-Speech	TS-VAD	27.3	22.4	7.3	19.0
NTT (Sys5)	Clustering	31.3	21.1	**5.9**	19.4
Univ. Cambridge	Clustering	48.2	25.6	10.3	28.0
Baseline	Clustering	56.3	27.9	9.3	31.2
STCON	TS-VAD	46.0	30.5	17.1	31.2
NVIDIA NeMo	Clustering	56.1	28.7	18.0	31.9
Paderborn	TS-VAD	58.7	30.0	12.4	33.7

1-2-16

1-2-16　CHiME7 遠隔発話音声認識タスクのための NTT システム：音声認識

NTT Multi-Speaker ASR System for the DASR Task of CHiME-7 Challenge

◎芦原孝典，加茂直之，松浦孝平，森谷崇史，小川厚徳，デルクロア・マーク（NTT）

◆講演番号 1-2-14 では CHiME7 と，NTT が投稿したシステムの全体像について概説した．

◆本稿では，その NTT システムの音声認識 (ASR) 部分に注目し，システムで用いた複数バリエーションの ASR モデルの概要に加え，利用したデータセット・デコーディング手法とその効果を紹介する．

◆ASR モデルについては 2 ステップからなる学習法により得られる 4 種類のモデルを組み合わせて用いた．

◆教師ありデータとして CHiME7 と LibriSpeech を学習に用いただけでなく，半教師ありデータとして VoxCeleb1+2 をデータ選択した上で学習に用いた．

◆デコード時は，上記のモデルに対し自己教師あり適応学習 (SSA) を適用し，更に Transformer LM を用いてリコアリングした (Figure 1)．

◆実験結果から，評価セットにおいて，Baseline に対し，提案システムの相対 WER 改善率は 38.6%となった．



Figure 1: Schematic diagrams of the NTT ASR system for the CHiME7 challenge.

Table 1: DA-WER on the eval set for far-field acoustic robustness task (sub-track1).

ID	Model	CHiME-6	DiPCo	Mixer6	Macro
(0)	Baseline	35.5	36.3	28.6	33.4
(7)	ASR1+SSA	24.3	23.3	14.7	20.8
(8)	ASR2+SSA	25.5	23.6	15.0	21.3
(9)	ASR4+SSA	24.4	23.2	14.6	20.7
(10)	(7)+(8)+(9)	24.0	22.8	14.6	20.5
(11)	(10)+LM rescoring	24.0	22.8	14.6	20.5

1-3-1

1-3-1　The speech perception of English consonant clusters in hearing loss simulation

○Eri Iwagami (JSPS oversea research fellow), Rosen Stuart, Tim Green (UCL)

◆ This study conducted a speech perception test in a simulated hearing loss condition to investigate the speech perception patterns of real English words and nonsense words with consonant clusters in individuals with age-related hearing loss.

◆Real words and non-words, with identical consonant clusters, such as ‘still’, and minimal pairs of real words and non-words that could exhibit consonant clusters due to the condition that weak vowels in those words tend to be deleted, such as ‘behind’, were presented in the simulation and non-simulation conditions to 22 participants .

◆Individuals' performances were significantly poorer in the simulation condition than in the non-simulation condition, suggesting that this hearing loss simulator successfully simulates age-related hearing loss.

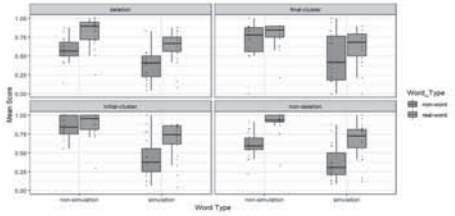


Fig.1 : Figures of Mean Scores for each condition. Figures are divided into 4 based on Cluster Conditions. The Y-axis in each figure shows HLS, and boxplots display Word Type.

1-3-2

1-3-2　ディスレクシア児の英語読み上げ音声評価に向けた特徴量の検討

Exploring features for assessing oral reading skills of dyslexic children

☆勝間田里菜，齋藤大輔，峯松信明（東大），成田あゆみ（もじこ塾）

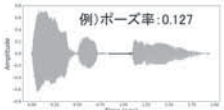
◆ **ディスレクシア**とは、知的能力や基本的な知覚能力には問題がないにもかかわらず、**読み書きに著しい困難**を示す**学習障害**である。



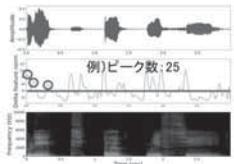
◆ 日本での発症率は低いが、英語学習を契機に、読み書きに困難を示す子どもたちがいる。

◆ ディスレクシア児童の**英語読み上げ訓練**に焦点を当て、彼らの**読み上げ音声を評価**すべく、3つの**特徴量**を検討した。

➢ ポーズ率



➢ 動的特徴量の L1 ノルムのピーク数



➢ 音声記号列の編集距離

例）　['ni', 'i', 'o', 'l', 'i', 'ð', 's', 'i'] ⟷ ['p', 'i', 'd', 'm', 'ɛ', 'x']

編集距離：9

1-3-3

1-3-3　市松音声の市松雑音マスキング：変調指数スペクトルによる分析

Masking checkerboard speech with checkerboard noise: Analysis with modulation index spectra

☆棟近光太郎（九大・芸工府），上田和夫（九大・芸工院），竹市博臣（理研・情報統合本部），蓮尾絵美, Gerard B. Remijn（九大・芸工院）

◆音声をいくつかの周波数帯域に区切り，隣り合う周波数帯間で互い違いに断続させた市松音声の無音部分を雑音（-6，0，+6 dB）で置き換え，了解度の変化を調べた（音講論2022秋，聴覚研究会2022年12月）。前報で用いた刺激について，変調指数スペクトルによる分析を行った。

◆雑音を付加しない市松音声の変調指数スペクトルには時間領域の断続（20, 80, 320 ms）による影響が見られた。一方で，-6 dB の雑音で無音部分を置き換えた市松音声は断続の影響が小さくなった。

◆+6 dB の雑音で無音部分を置き換えた場合にはマスキングにより，音声特有の振幅変調が弱められていることが確かめられた。

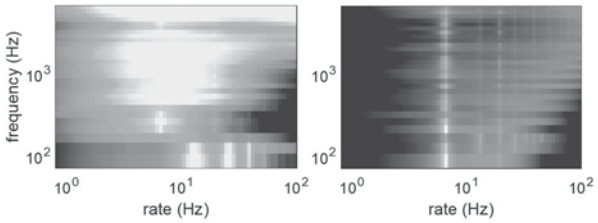


Fig. Modulation index spectra of 16-band checkerboard speech stimuli (left) and 16-band checkerboard speech stimuli in which silent parts are filled with +6 dB noise (right). Both stimuli are segmented at 80 ms.

1-3-4

1-3-4 Contributions of Instantaneous Modulation Components in Temporal Amplitude Envelope to Vocal Emotion Perception

☆Taiyang GUO, Takuto ISOYAMA, Shunsuke KIDANI, Masashi UNOKI (JAIST)

◆The temporal amplitude envelope (TAE) of speech signals is very important in vocal emotion perception. However, the contribution of instantaneous modulation components (IMCs) of TAE remain unclear.

◆Aim: To clarify whether the IMCs of TAE contribute to vocal emotion perception.

◆Method: Comparing the emotion recognition rates of NVS when its IMCs are controlled in the time domain under four conditions: no processing, stretching, compression, stretching and compression.

◆Results: Controlling the IMCs in the time domain leads to reductions in emotion recognition rates.

◆Conclusion: The IMCs of TAE contribute to vocal emotion perception.

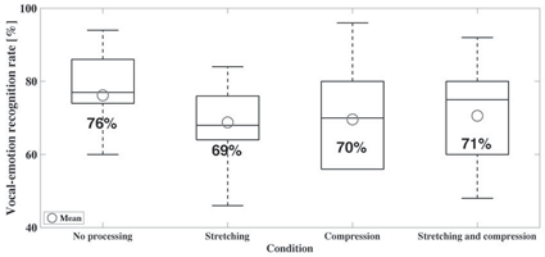


Fig.1 : Emotion recognition rates of NVS when IMCs are controlled under four conditions.

1-3-5

1-3-5 商品分類と話者属性を考慮した宣伝音声と購買意欲の関係の分析

An analysis of the relationship between advertising speech and willingness to buy considering product type and speaker attributes

◎長野瑞生，井島勇祐，廣谷定男（NTT）

◆宣伝音声が購買意欲に与える影響
- 音声印象は感情と同程度に音声の購買意欲への影響を媒介[長野他, ASJ2023a]
- 購入に伴う経済的リスク（知覚リスク）が感情・購買意欲に影響[Kim+, 2010]
- 販売員と顧客の性別等の属性が一致することで購買意欲向上[Smith+, 1998]
- 知覚リスクや話者・聴取者の属性の一致が媒介効果や購買意欲に影響する可能性がある

◆知覚リスクで分類された宣伝文による多数話者音声を用いた検証
- 話者クラスタリングにより選出した話者男女 26 名の対し、1 話者あたり 13 文章（知覚リスク低 8 文章/知覚リスク高 5 文章）を音声合成で作成
- 商品の知覚リスクが音声印象の媒介効果に及ぼす影響と話者・聴取者属性の一致が購買意欲に及ぼす影響を大規模な主観評価実験により調査

◆知覚リスク及び属性一致の影響をそれぞれ調整媒介分析、分散分析によって検証
- 知覚リスクが高い商品の場合のみ声の高さが購買意欲に影響を与える
- 話者・聴取者の属性の一致よりも音声印象や聴覚特性の影響が大きい

媒介変数	知覚リスク	Est.	S.E.	95 %CI Lower	95 %CI Upper
(a) 高い-低い	低	0.01	0.09	-0.15	0.18
	高	0.22	0.01	0.05	0.40
(f) 明るい-暗い	低	1.48	0.17	1.17	1.83
	高	1.03	0.15	0.76	1.34

Table1 : 知覚リスクで調整された音声印象の媒介効果

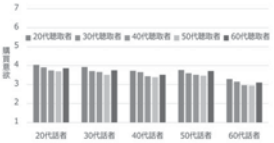


Fig.1 : 話者・聴取者年代ごとの購買意欲

1-3-6

1-3-6 スナック類の咀嚼音の音響的特徴と食感の関係その 2 ～深層学習による種類判別と美味しさの評価～

Relationship between acoustic characteristics and texture of chewing sounds of snacks Part 2 ~Type Discrimination and Evaluation of Deliciousness by Deep Learning~

☆橋本卓己，藍原直大，田中博(神奈川工科大)，三浦貴大(産総研)，上田麻理(神奈川工科大)

◆前報では，グミの咀嚼音の音響解析と食感評価項目を用いた美味しさに関わる因子分析を行い，ポテトチップスとの食感の違いを明らかにした。

◆本稿では，8 種類のグミの咀嚼音から深層学習を用いて種類判別を行い，食感評価項目からポテトチップスとグミの食感に基づく美味しさの回帰モデルの作成を試みたので報告する。

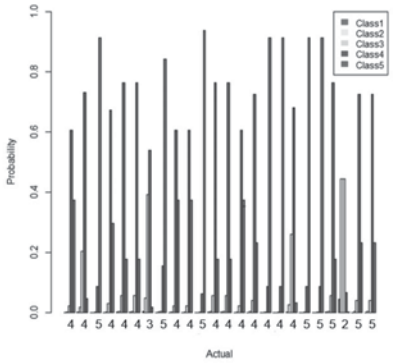


Fig.1 Predicting the Taste of Gummies by Ordinal Logistic Regression Models

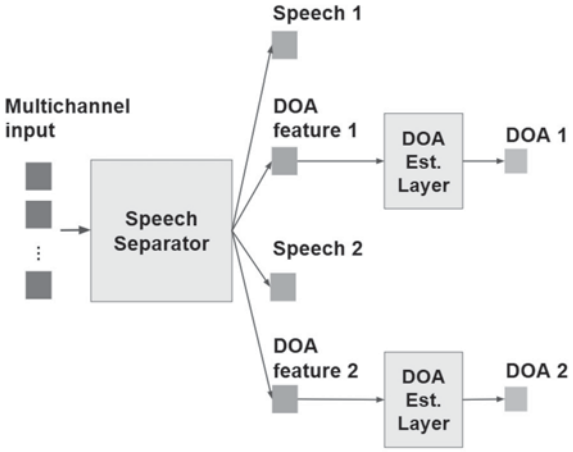
1-4-1

1-4-1 Multitask Learning of Speaker Separation and Direction-of-Arrival Estimation

☆ Roland Hartanto (Tokyo Tech), Sakriani Sakti (JAIST), Koichi Shinoda (Tokyo Tech)

◆We propose a **multitask learning** between **speaker separation and direction-of-arrival (DOA) estimation** to **enhance multi-channel speaker separation**
- We assume DOA info of each source help speech separation

◆Evaluated on the SMS-WSJ dataset, our method improves Location-Based Training performance (SI-SDR) by 6.09 points.



1-4-2

1-4-2　深層学習シングルチャネル音源分離における定常雑音の影響

Effects of stationary noise on deep-learning-based single-channel sound source separation

○古家賢一，三浦伊織（大分大）

◆本稿では，深層学習を用いたシングルチャネル音源分離の実環境下での利用を想定し，定常雑音が加わった場合の分離性能の影響を，学習データの違いにについて実験的に調査した。

◆調査に音源分離手法として用いた SepFormer では，Transformer の Multi-head self-attention 機構を利用して，音源信号の短期的および長期的な信号間の依存性を学習し，高い音源分離性能を実現している。

◆実験の結果，雑音無音声データ（WSJ0-2mix）で学習した音源分離モデルでは，定常雑音が加わることで大きく分離性能が劣化するが，雑音抑圧モデルを前処理に用いることで，雑音有音声データ(WHAMR!)で学習した音源分離モデルと同程度以上の分離性能が得られることが分かった。

Table 1 学習データの違いによる定常雑音の影響。値はSDR(dB)

		学習データ	
		WSJ0-2mix	WHAMR!
テストデータ	WSJ0-2mix	22.77	14.90
	SN比：20 dB	4.07	9.07
	SN比：10 dB	3.23	9.21
	SN比：0 dB	−2.18	9.15

Table 2 前処理に雑音抑圧を加えた場合の定常雑音の影響。値はSDR(dB)

		学習データ	
		WSJ0-2mix（音源分離）+WHAMR!（雑音抑圧）	WSJ0-2mix（音源分離）+Wienerフィルタ
テストデータ	SN比：20 dB	9.52	−0.51
	SN比：10 dB	9.56	−0.69
	SN比：0 dB	9.42	−2.28

1-4-3

1-4-3　拡散モデルに基づく音声強調の複数マイク化と評価

Multi-microphone Speech Enhancement Based on Diffusion Model

☆木村凜乃（早稲田大, NTT），中谷智広（NTT），加茂直之（NTT），
Marc Delcroix（NTT），荒木章子（NTT），牧野昭二（早稲田大）

◆本稿では，拡散モデルに基づく単一マイク音声強調手法（SGMSE）を拡張して，複数マイクの観測信号を用いて音声信号の空間情報をも高精度に回復する複数マイク音声強調手法（mSGMSE）を提案する.

◆提案法では，全マイクを同時に処理することができ，マイク数の増加による計算量の増加を最小限に抑えることができる.

◆評価実験の結果，複数マイクの観測信号に対して，SGMSE を各マイクに個別に適用する場合と比べ，提案法は同程度の音声強調性能を実現しつつ，より高精度に空間情報を回復でき，大幅に計算量を削減できることが確認された.

Table1: Accuracy of signal estimation.

		PESQ	ESTOI	SI-SDR [dB]
Validation data	SGMSE	3.07	0.83	16.4
	mSGMSE	**3.09**	**0.84**	**16.8**
Evaluation data	SGMSE	**3.21**	**0.82**	**16.4**
	mSGMSE	3.16	**0.82**	15.9

Table2: Estimation error of spatial information and processing time per utterance.

	ΔITD [sample]	ΔILD [dB]	Processing time per utterance [s]
SGMSE	0.84	0.0072	31.2
mSGMSE	**0.14**	**0.035**	**15.5**

1-4-4

1-4-4　拡散過程と敵対的学習の併用による普遍音声強調

Diffusion-based Universal Speech Enhancement with Joint Adversarial Training

○シャイブラー ロビン，藤田雄介，橘健太郎（LINE ヤフー）

◆本研究では，拡散過程に基づく普遍音声強調手法である UNIVERSE に着目し，ネットワーク構造と損失関数の改善を提案する.

◆UNIVERSE の学習効率と安定性を向上するためにネットワークに weight normalization，エイリアシング対策フィルター，入出力の正規化など，様々な改善を導入する.

◆クリーン音声の分布を予測する敵対的学習に基づく損失関数を加え，自然性の評価指標を向上する（図１）.

◆Voicebank-DEMAND データセットを用いて，性能確認実験を行った.

◆敵対的学習のみ，拡散過程のみ，また拡散過程と識別学習に基づく手法と比較し，提案法による自然性指標の向上を確認した（表１）.

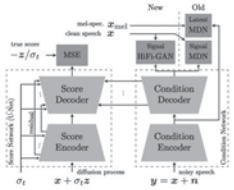


Fig.1：Network structure with old and new loss functions.

Method	SI-SDR	PESQ	ESTOI	DNSMOS
Unprocessed	8.4	1.969	0.787	2.70
HiFi++ [4]	17.9	2.90	—	3.10
SGMSE+M [8]	17.3	2.96	0.870	—
StoRM [8]	**18.8**	2.93	**0.880**	—
StoRM†	14.5	2.95	0.861	3.16
UNIVERSE [3]	17.9	2.86	0.849	3.16
+Network	18.4	2.93	0.856	3.18
+HiFi-GAN	18.3	**2.98**	0.863	**3.20**

† reproduced from https://github.com/sp-uhh/storm.

Table 1：Evaluation on Voicebank-DEMAND dataset.

1-4-5

1-4-5　オンライン補助関数型独立ベクトル分析の忘却係数の動的制御による移動音源分離

Dynamic control of forgetting factor in source separation for moving sources.

☆増子凱斗，中嶋大志，河村隆生，小野順貴（都立大）

◆背景

- 従来のオンライン音源分離では，分離行列を作成するために使用する共分散行列の推定には，忘却係数というパラメータが必要であり，固定値が使われることも多い.

◆概要

- 分離信号のスペクトルの相関を用いて逐次的に忘却係数を更新することで，忘却係数を動的に制御することの有効性を確認する.

◆実験結果

- 100 回のシミュレーション実験の結果，移動音源が存在する環境下での音源分離において従来手法より良い分離性能を示した.

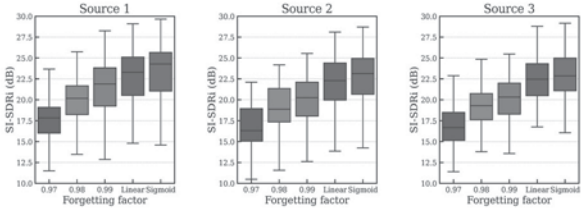


Fig.1：Boxplot of SI-SDRi for different forgetting factors. 0.97, 0.98, 0.99 are conventional methods. Linear and Sigmoid are proposed methods.

1-4-6

1-4-6　多チャネル非負値行列因子分解に基づく複数マイクアレイを用いた同時的音源定位・分離

Joint Sound Source Localization and Separation
Using Multiple Microphone Arrays
Based on Multichannel Nonnegative Matrix Factorization

☆須村允亮（理研/京大院），Aditya Arie Nugraha（理研），
△Diego Di Carlo（理研），坂東宜昭（産総研/理研），吉井和佳（京大/理研）

◆**目的：音響シーン解析の基盤技術となる音源定位・分離手法の実現**

◆**関連研究：① BSS ② 空間事前情報を用いた音源分離・定位**

① MNMF，Fast MNMF などのブラインド音源分離 (BSS) 法が存在

② 空間共分散行列に音源到来方向に基づく事前分布を導入

→ 移動音源の分離音と到来方向の同時推定を実現

◆**提案法：複数マイクアレイを用いた分離音・音源位置の同時推定**

➢ 複数マイクアレイを束ねた MNMF の確率モデルを拡張

✧ 音源位置（座標）に基づく空間事前情報を導入

➢ 空間共分散行列を MAP 推定・音源位置を経験ベイズ推定

◆**評価実験：シミュレーションで生成した観測音を用いた評価**

➢ SDR および定位誤差評価により提案法の有効性を確認

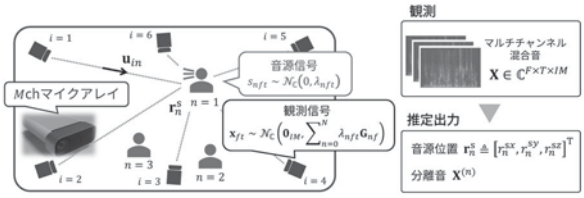


Fig.1：Overview of joint sound source localization and separation using multiple microphone arrays.

1-4-7

1-4-7　独立低ランク行列分析における反復射影法の高速化・安定化アルゴリズム

Accelerating and stabilizing iterative projection method for ILRMA

☆石川悠人，△大久保拓哉，高宗典玄（東大），北村大地（香川高専），
中村友彦（産総研），猿渡洋（東大），高橋祐，近藤多伸（ヤマハ）

◆反復更新アルゴリズムの反復射影法（IP）に対して解析的に等価な変形を繰り返すことで，数値的に安定かつ高速なアルゴリズム（Fast IP）とさらなる安定化を施したアルゴリズム（Stable IP）を提案

◆有効性の検証のため，独立低ランク行列分析（ILRMA）において適用し，処理全体に要した実行時間と数値的不安定性を比較

◆CPU 上でのシミュレーション実験により，従来アルゴリズムと比べて，提案アルゴリズムは半分以下の時間で実行され，かつ不安定環境に対して頑健であることを確認

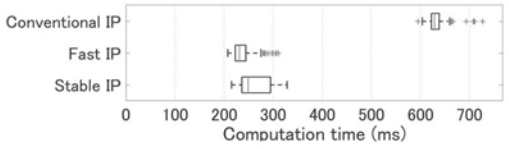


Fig.1：Boxplot of computation times for each method

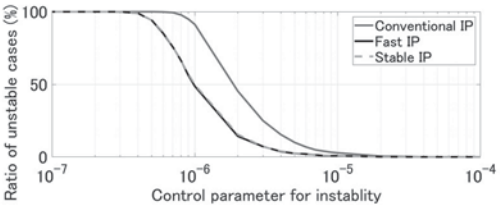


Fig.2：Ratio of unstable cases with respect to control parameter

1-4-8

1-4-8　ADMM アルゴリズムを用いた ILRMA 型ブラインド音源分離

ILRMA-like blind source separation using ADMM algorithm

◎渡會博子 (農工大)，松本和樹 (早大)，山田宏樹，矢田部浩平 (農工大)

◆ 我々はこれまで，**ADMM** アルゴリズムを用いた
優決定ブラインド音源分離 (BSS) フレームワークを提案

► **柔軟な音源モデル設計，正則化項・制約の追加が容易**

◆ 上記フレームワークに**非負値行列因子分解 (NMF)** を導入

► **ILRMA 型ブラインド音源分離フレームワークを実現**

◆ 凸である **perspective** 関数を用いた ILRMA 型 BSS を実現

► 従来法と同程度の性能，**外れ値減少により平均値向上**

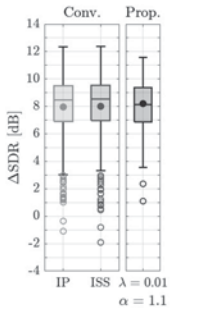


Fig.1: Comparison of separation performance after 1000 iterations

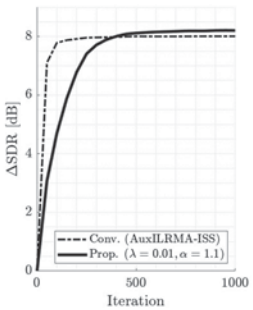


Fig.2: Separation performance per 50 iteration

1-4-9

1-4-9　振幅を柔軟に調節できる時間周波数表現の凸最適化による実現

Realizing time-frequency representations with
flexible adjustment of magnitude via convex optimization

◎新井慶大，山田宏樹，矢田部浩平（農工大）

■ 離散ガボール変換による時間周波数（T-F）表現
窓による広がりが解析に影響 ► スパースな表現が必要

従来　ℓ_1 ノルム最小化　　△ 各 T-F ビンを独立に圧縮

実際　複数の T-F ビンが関係を持つ
(e.g. 振幅スペクトログラムの低ランク性・滑らかさ)
► 振幅の構造を残しつつスパースな T-F 表現が必要

課題　一般に非凸最適化に陥る
△ 初期値が適切でないと最適解が得られない

提案　perspective 関数を用いた凸最適化手法
► 振幅を柔軟に調節できる T-F 解析を実現

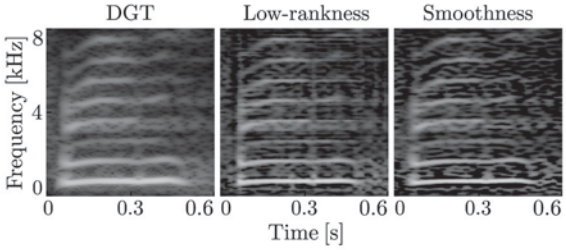


Fig. 1　Obtained time-frequency representations of a speech signal

1-4-10

1-4-10

振幅の滑らかさを考慮した凸最適化による調波打撃音分離

Anisotropic-smoothness-based harmonic/percussive source separation via convex optimization

◎赤石夏輝，山田宏樹，矢田部浩平（農工大）

背景　調波打撃音分離（HPSS）
- 音楽信号の正弦波成分と打撃音成分を分離する処理
- 振幅の構造の違いに基づく手法が提案されている

従来　振幅スペクトログラムのみを用いた凸最適化
- 位相の情報を扱えていないため**分離性能に上限**あり
- 複素に自然に拡張すると**非凸問題**となってしまう

提案　**非凸制約を回避**しつつ**複素の変数を最適化**
- 複素スペクトログラムと結びつく重みを平滑化する
- **分離性能が向上**（SDR：正弦波成分 **+1.7** dB，打撃音成分 **+1.9** dB）

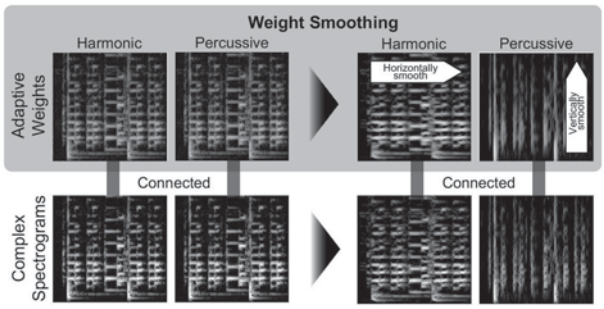


Fig. 1　Overview of the proposed method

1-4-11

1-4-11

再生信号適応型マルチスポット再生の最適化に関する検討

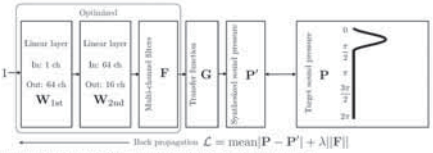
Investigating optimization methods of signal-adaptive multiple sound spot synthesis

○岡本拓磨 (情報通信研究機構)

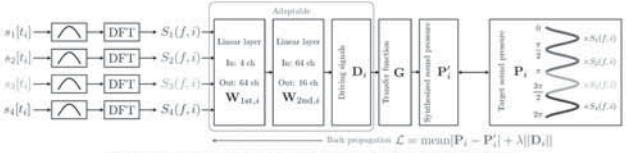
Q：「再生する音の種類によって適応的に駆動信号を最適化できるマルチスポット再生ってできるんですか?」
オカモト：「それならこれで!!」

・ニューラルネットワークの学習に用いられている誤差逆伝播を用いて再生信号に依存した駆動信号を適応的に最適化する方式を提案

・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



(a) Time-invariant multi-channel filters optimized with back propagation



(b) Time-variant adaptive driving signals optimized with back propagation

1-4-12

1-4-12

ボリュメトリック音声再生のためのバイノーラルレンダリング法の 3 次元化

Three-dimensional expansion of binaural rendering method for volumetric audio reproduction

○松井健太郎(NHK)

- NHK では，ボリュメトリックキャプチャスタジオ設備として，「メタスタジオ」の研究・開発を進めている（Fig. 1）。
- この映像のボリュメトリックキャプチャに対応する音声制作技術として，ボリュメトリック音声制作を提案している。
- 前報では，提案する制作フローのうち，ヘッドホンによるバイノーラルレンダリング法を 2 次元に限定して提案した。本報では，提案法の 3 次元への拡張について概説する。
- メタスタジオには，Fig. 1 に示すように，ステージを取り囲むようにフレームが配置されている。このフレームを形成する多面体の頂点にマイクロホンアレイをマウントし，音源オブジェクトより放射される音波を方向別に取得する。
- この音波により形成される音場を球面調和領域で外挿し，任意の聴取位置に入来する音波により形成される音場を推定する。更に，聴取位置近傍の音場から，頭部伝達関数を用いて左右耳に提示するバイノーラル信号を算出する。



Fig. 1 : Interior view of the meta studio

1-5-1

1-5-1

分類型半制限ボルツマンマシンによる全音程関係を考慮した和音認識

Chord recognition considering all interval relations using classification semi-restricted boltzmann machines

☆石川峻弥，中鹿亘(電通大)

◆目的
- 音楽音響信号において，ある時刻にどの和音が鳴っているかを認識する技術（和音認識）の認識精度の向上

◆提案手法
- 半制限ボルツマンマシンと分類型制限ボルツマンマシンを組み合わせた，分類型半制限ボルツマンマシン（ClassSRBM）を利用
- 完全 5 度の関係のみを考慮する従来手法に対し，全ての音の関係を同時に考慮可能

◆結果
- 和音認識精度の向上，およびモデル構造の効果を確認
- 音程関係の考慮に関する冗長性の削減，モデルのパラメータ数と性能の関係の分析，シフト不変性の考慮が今後の課題

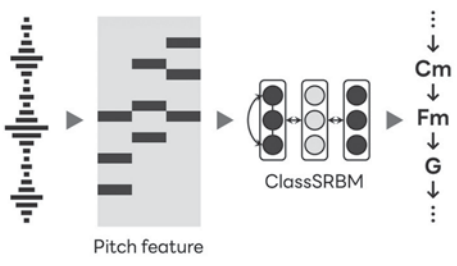


Fig.1 : Overview of the proposed method

1-5-2

1-5-2　ドラムパターンを用いた楽曲ジャンル分類に関する研究

A Study on Music Genre Classification using Drum Patterns

☆荒　尭志、青木　直史、土橋　宜典(北大)

◆研究背景
- 近年、Spotify などの音楽サービスの発展が著しい
  - ✧ 音楽の検索や推薦タスクの需要あり
  - ✧ 検索・推薦タスクには音楽ジャンルの情報が有用
- しかし音楽ジャンルの基準は曖昧
  - ✧ 人力による一意なジャンル分類は困難
- 計算機によるジャンル分類を実現させたい

◆手法
- ドラムデータをスペクトログラムに変換
- CNN による分類モデルで学習
- 学習精度を評価

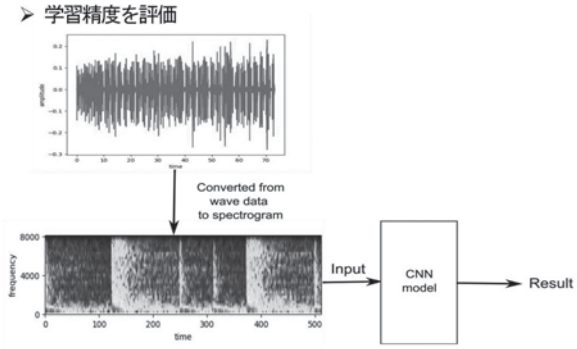


Figure: Experimental flowchart of this study

1-5-3

1-5-3　トランペット演奏を対象としたオンセット検出に基づくテンポ変化推定

Tempo Change Estimation Based on Onset Detection for Trumpet Performance

☆岡森一樹，武田一哉，戸田智基(名古屋大)

◆楽器演奏において意図しないテンポの変動は演奏のミスの誘発につながる大きな課題

◆トランペット演奏音を対象とした深層オンセット検出：トランペット演奏の学習データを MIDI で作成し，CNN を用いた手法で学習

◆オンセット列からのテンポ変化推定：オンセット間隔を観測変数，テンポ変化率を潜在変数とする線形動的システムによるモデル化

◆オンセット検出は低い deletion と insertion を達成

◆テンポ変化推定は最尤分布系列選択処理が課題

	raw sound	MIDI
n_data	204	10000
deletion	16	4
insertion	3	0

Table.1：Results of onset detection. n_data is the number of onsets for each dataset.

error	∼0.1	0.1∼0.2	0.2∼	RMSE
frequency	7830	49	70	0.044

Table.2：Distribution of errors in tempo change estimation.

1-5-4

1-5-4　蝸牛遅延特性を模擬したIIRオールパスフィルタの音楽的利用の検討

Research on musical applications of IIR all-pass filter simulating cochlear delay characteristics.

☆飯田大貴，山田宏樹，矢田部浩平(農工大)，饗庭絵里子(電通大)

● 背景
- ✓我々は以前，蝸牛の群遅延特性を模した，遅延量の調整が可能なオールパスフィルタを設計した。
- ✓蝸牛遅延を考慮した信号処理を音楽に適用することで，新たな価値を創出できる可能性がある。

● 本研究
- ✓このフィルタをスネアドラム，バスドラム，ベース，ギター音源に適用し，フィルタが各音源の聴感的特徴に及ぼす効果を主観評価実験によって調査した。

● 結果
- ✓ベース，ギターなどの弦楽器に対しては，音楽的利用価値があると考えられる。

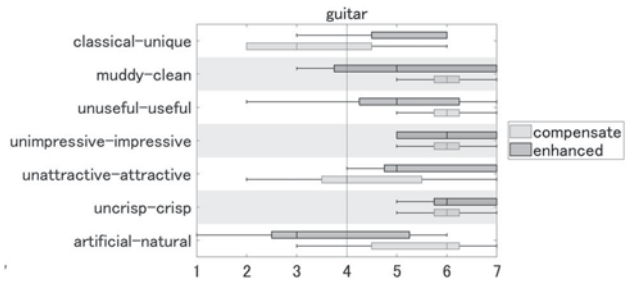


Fig.1 Results of guitar evaluations

1-5-5

1-5-5　複数のオーディオエフェクトが適用された楽音に対するエフェクトチェイン推定と原音復元

Estimation of audio effect chain and recovery of dry signal from multi-effect-applied musical signal

☆武伯寒(東大院・情報理工)，△渡邉研斗，△中塚貴之，△Cheng Tian，中野倫靖，後藤真孝(産総研)，高道慎之介，猿渡洋(東大院・情報理工)

◆複数のオーディオエフェクト（AFX）が適用された楽音からエフェクトチェインを推定し，原音を復元するタスクを提唱
- ➢ サウンドデザイン作業の効率化や，初学者のための AFX 特性理解支援への応用が期待される

◆最後に適用された AFX を推定し，その AFX を適用する前の信号を復元する DNN モデルを繰り返し適用することで，エフェクトチェインを後方から順番に推定し，原音を復元する手法を提案

◆ギターのフレーズ演奏音源に対する評価実験により，提案手法による原音復元性能が有効であり，推定したエフェクトチェインと復元した原音を用いて楽音を再現できることを確認

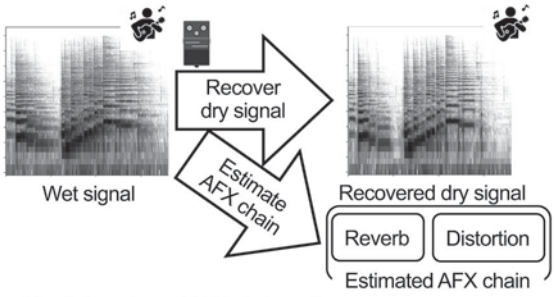


Fig. 1 Overview of AFX chain estimation and dry signal recovery tasks.

1-5-6

1-5-6　音叉の振動継続時間における打撃位置の影響：マイクロフォンおよびレーザードップラー振動計による評価

Effects of striking position on vibration duration of tuning forks: Evaluation by microphone and laser-doppler vibrometer

☆上野嵩旺，大塚翔，中川誠司（千葉大）

◆音叉では基音が長く持続し，倍音は速やかに減衰することが望ましい．

◆音叉の製作工程においては基音周波数のチューニングがなされるものの，音の持続の評価は行われていない．また，打撃位置は音の振幅に影響するが，持続時間への影響は不明である．

◆打撃位置が音の持続時間に及ぼす影響の解明を目的として，マイクロフォンとレーザードップラー振動計によって，音叉から生じる音と音叉各部の振動を評価した．

◆打撃位置を３段階で変化させたところ（脚部先端側から high, mid, low），倍音の継続時間は両計測で mid = low > high となった．これは high の位置に倍音モードの節が存在するためと思われる．

◆一方，基音の継続時間はレーザードップラー振動計では high > mid > low で，基音の腹/節の位置から予想される結果となったのに対して，マイクロフォンでは low > mid > high となった (Fig. 1)．両計測の乖離の原因は不明であり，さらなる検討が必要である．

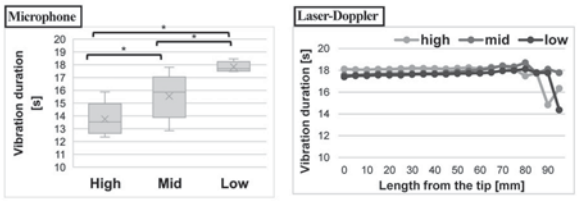


Fig.1 Vibration duration of the fundamental component measured by a microphone & a laser-doppler vibrometer.

1-5-7

1-5-7　Physics-informed Neural Networks を用いた金管楽器最適設計手法の基礎的検討

Fundamental study of design optimization for brass instruments using Physics-informed Neural Networks

◎横田和哉，△坂爪孝照，△阿部雅二朗（長岡技術科学大学）

◆楽器の設計において、音響数値解析は重要な指標の一つである。

◆近年の機械学習の進歩により、ニューラルネットワークを用いて数値解析を行う手法が提案されている。その一つである Physics-Informed Neural Networks (PINNs) は、支配方程式による拘束を損失関数に導入することで、物理的に満足する解を得る手法である。

◆本稿では、PINNs を用いた楽器の設計最適化の基礎的検討として、トランペットの寸法最適化について報告する。

◆寸法の最適値に対して 10%の誤差を初期値として与え、提案法を適用し、誤差 1%以下の精度で寸法最適化を行うことができた。

◆提案法は逆解析のための複雑なアルゴリズムを構築する必要なく設計最適化を行うことが可能であり、簡便であるという長所がある。

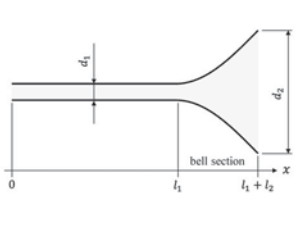


Fig.1：Trumpet shape consisting of four dimensions.

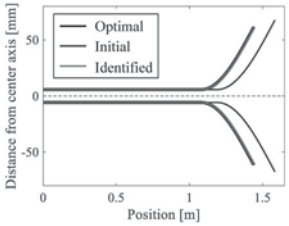


Fig.2：Initial and identified trumpet shape.

1-5-8

1-5-8　物理モデル計算を用いたピックの形態が弦振動に与える影響の検討

Investigation of the effect of pick form on string vibration using physical model calculations

☆塩屋仁菜（九大・芸工），鮫島俊哉（九大・芸工）

◆著者らは，ピックと弦を対象とした撥弦物理モデルとその解析方法を提案している。

◆本研究では，ピックの境界条件として回転バネを導入する改良を行う。この改良によって，演奏者のピックの持ち方を撥弦物理モデルに反映させることができる。

◆この撥弦物理モデルを用いて計算を行い，ピックの形態，および撥弦時の演奏動作に関するパラメータを変化させた際に，それが弦振動にどのような影響を与えるのかを検討する。

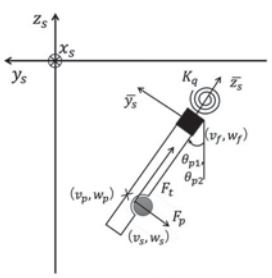


Fig.1：Collision and stick-slip between string and pick.

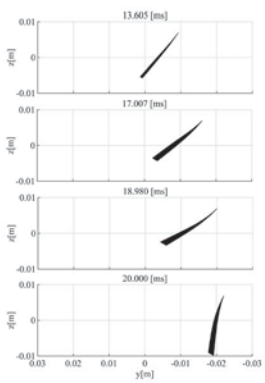


Fig.2：Time response of the vibrational displacement distribution of the pick.

1-5-9

1-5-9　有限要素法によるトロンボーンのリップ振動場モデルの高精度化

Improving the lip vibration model in trombones through the finite element method

☆竹内遥，鮫島俊哉（九大・芸工）

◆金管楽器の物理モデルの完成形ともいわれるAdach-Satoモデルでは，リップを２方向に振動する２自由度系の質点と考えている。

◆本論文では，トランペットを想定して作られているAdach-Satoモデルをトロンボーンに適用することを試みた。

◆さらに，Adach-Satoモデルを基礎として，そのリップ振動場モデルの分布定数系としての取扱い・３次元化を提案した。

◆Adach-Satoモデル，提案モデルともに，実際の演奏時よりも高い吹奏圧を加えれば，放射音を合成できることがわかった。

◆また，提案モデルではAdach-Satoモデルに比べて吹奏圧に対する音高変化の感度が低く，音高がリップの固有周波数に依存して決定されており，実際の演奏感覚と近い結果になった。

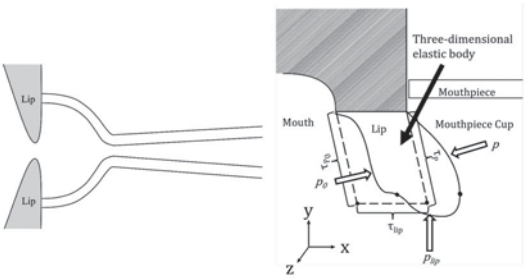


Fig.1：Schematic of a lip reed and proposed physical model of a lip reed.

1-5-10

1-5-10　有限要素モデルを用いたヴァイオリンの振動音響解析

Vibroacoustic analysis of violins using finite element models

○黒沢良夫（帝京大・理工）

◆ヴァイオリンの表板・裏板等１つ１つの部品の３D データを取得し FE モデルを作成した。また、同形状で３D プリンターを用いてヴァイオリンを作成した。音圧感度を計測し、計算による振動音響解析と比較した結果を紹介する。

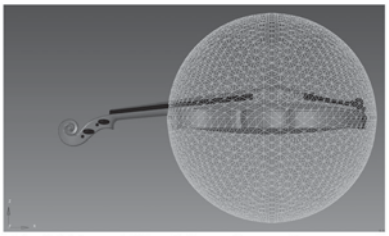


Fig.1：FE model for violin and acoustic space



Fig.2 Experimental setup

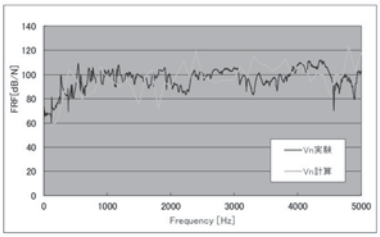


Fig.3 Comparison of experimental and calculated results of acoustic FRF

1-5-11

1-5-11　偏光高速度干渉計を用いた楽器音響等の可視化

Visualization of musical instrument sound field using high-speed polarization interferometer

☆野澤 遙（早大・基幹理工），北爪裕道（北海道教育大・教育），李 珍咏（東大院・学際情報学府），伊東 乾（東大院・情報学環），及川靖広（早大・基幹理工）

◆楽器ごとに異なる音の空間的な広がりについては、J.S.バッハの時代から「オルガン鑑定人」による調整が求められながらそれを音楽の言葉で明確に示すフレームワークは存在していない。

◆偏光高速度干渉計を用いた完全に非接触な音場可視化技術により、ヴァイオリンなどの楽器から直接放射、拡散される音波と、その結果広がってゆく音場を正確に可視化したので、その測定を概説する。

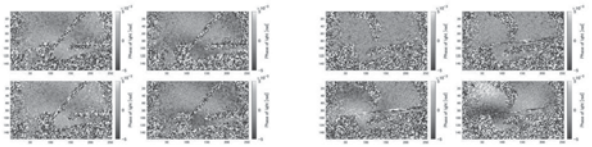


Fig.1 バイオリンの生成する音波の可視化
5-L 弓奏 5-R ピツィカート

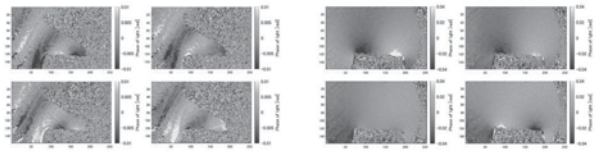


Fig.2 鑿から生成する音波の可視化
6-L 打奏 6-R コントラバス弓による弓奏

1-5-12

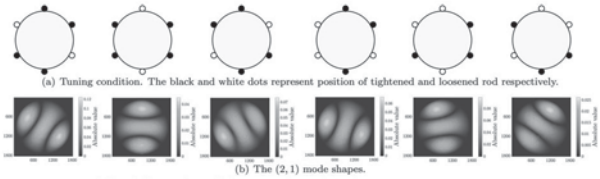
1-5-12　縞投影法による高次振動モードの可視化を通したドラムヘッドの張力分布推定

Estimation of tension distribution on drumheads by visualization of high-order modes using Fourier transform profilometry.

◎橋本涼汰，及川靖広（早大理工）

◆縞投影法を用いてドラムヘッドの(2,1)モードを可視化

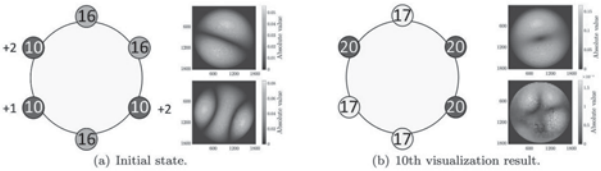
- Fig.1：チューニング条件と歪んだ(2,1)モード形状
  - ➢ ２本の弧状節線が現れる
  - ➢ ２ヶ所の小張力領域の位置を視覚的に示す
- Fig.2：(1,1)・(2,1)モードを併用した反復チューニング測定
  - ➢ モード画像から張力分布を推定し，ロッドを再調整
  - ➢ (1,1)モードだけでは困難だった初期条件を採用
  - ➢ (2,1)モードにより効率的なチューニングが実現



(a) Tuning condition. The black and white dots represent position of tightened and loosened rod respectively.

(b) The (2, 1) mode shapes.

Fig.1 Results of the rotational tuning measurements.



(a) Initial state.　(b) 10th visualization result.

Fig.2 Results of the iterative tuning measurements.

1-5-13

1-5-13　小鼓の加振位置の違いによる音の特性と膜の振動特性の評価

Evaluation of sound characteristics and membrane vibration characteristics for different excitation positions of kotsuzumi.

☆舩倉智久，Saha Chaity，松谷晃宏，田原麻梨江（東工大）

◆目的；膜の振動と胴の共鳴器としての特性の解析から、小鼓音の音響的特性について明らかにする。

◆胴の形状の音響特性の測定、膜の加振位置の違いによる振動特性の解析、膜を打つ位置の違いによる小鼓の音の解析及び、演奏家の四種類の奏法(チ、タ、プ、ポ)の音を打つ革の位置と演奏音を測定。

◆膜の振動である 192 Hz、小鼓胴の共鳴周波数である 840 Hz、胴本体の振動である 1580 Hz により小鼓の音が構成されることがわかった。

◆実際の小鼓音では、チ、タは胴の上と膜の端を打つため胴の共鳴周波数と胴本体の振動である 800 Hz、1200 Hz 付近の音、プ、ポでは膜の(0,1)モードと胴の共鳴周波数である 200 Hz、800 Hz 付近の音によって構成されていることが確認できた。

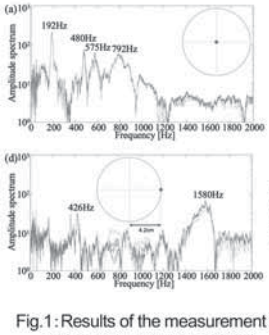


Fig.1：Results of the measurement of hitting a kotsuzumi with a ball.

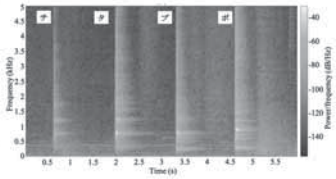


Fig.2：The spectrogram of the four types sound of kotsuzumi.

1-5-14

1-5-14

ピエゾ素子を用いた スネアドラム打面のマルチチャネル収録

Multi-channel recording of snare drum head using piezo elements

☆安藤元暉, 山田宏樹, 倉科佑太, 矢田部浩平（農工大）

背景

ピエゾ素子によるスネアドラムの収録
▸ マイクロホンに比べ他楽器の混入を抑制

▼

課題

設置位置によって振動モードが変化する
▸ 複数の位置にピエゾ素子を設置して収録

実験

スネアの音をピエゾ素子を用いマルチチャネルで収録
リングミュートを使用した場合の収録音と比較
▸ マイクロホンでの収録音と同程度のエネルギーを確認

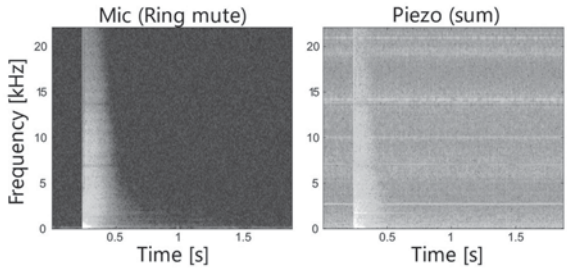


Fig. 1: Spectrogram of recorded sound with ring mute (left) and with multiple piezoelectric elements (right).

1-7-1

1-7-1

最近の室内音響設計での課題* ー実務で話題ー

* Some recent topics on room acoustical design in practice

〇石渡智秋(永田音響設計)

◆日頃の実務において感じることや課題など、音楽のためのホールの室内音響設計に関わる話題を紹介させていただく。

◆室形状の検討におけるコンピュータシミュレーションの活用・工夫、その他音響障害の検討など、いくつかの話題に触れたいと考える。

1-7-2

1-7-2

設計実務における音響性能と意匠性の融合

Integration of acoustic performance and design in architectural design practices

〇中川浩一(日建設計)

◆建築の設計は，建物を構成する機能的要素である意匠・構造・設備…といった多様な分野の検討とその融合であり，音響もその分野の一つに数えられる。本稿は，音響性能が特に重視されるホールや音楽練習室における設計実務が，何を手掛かりに始まり，どのような経過や工夫を経て音響性能と意匠性を融合させているかを事例により紹介する。

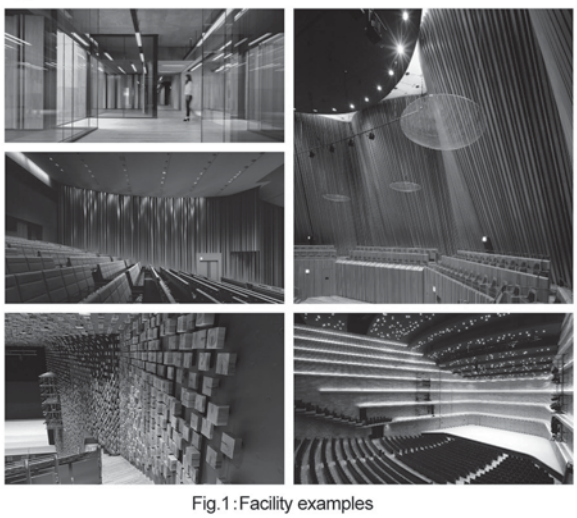


Fig.1 : Facility examples

1-7-3

1-7-3

音楽のための室内音響設計における課題と工夫

Issues and innovations in room acoustic design for music

〇荒田耕平，北村浩一(NHK-TECH)

＊近年のトピック
- 天井構造と空調方式
- 拡散体
- 客席椅子と残響調整

＊事例紹介
- 拡散性を抑えたケース (Fig.1)
- 舞台と容積の大型化 (Fig.2)
- マルチレイヤー反射壁 (Fig.3)
- 部分的デッドエンド (Fig.4)
- 軽鉄レス化粧木仕上げ (Fig.5)
- 音楽アリーナでの側方反射音 (Fig.6)
- 拡散反射仕様の後壁 (Fig.7)

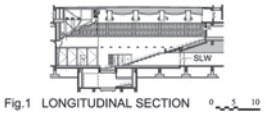


Fig.1 LONGITUDINAL SECTION



Fig.2 LONGITUDINAL SECTION

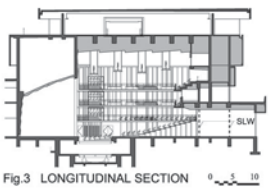


Fig.3 LONGITUDINAL SECTION

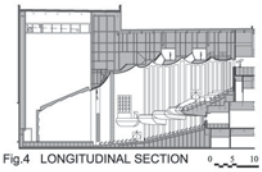


Fig.4 LONGITUDINAL SECTION

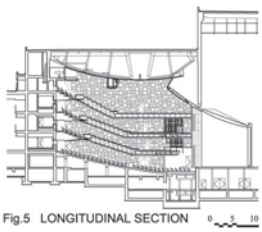


Fig.5 LONGITUDINAL SECTION

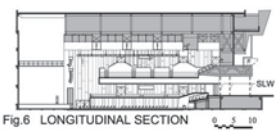


Fig.6 LONGITUDINAL SECTION



Fig.7 LONGITUDINAL SECTION

1-7-4

1-7-4　初期反射音制御に着目したホール音響設計事例の紹介

Examples of the Acoustical Design focused on the control of lateral reflections

○宮崎秀生（ヤマハ）

◆ホールや劇場の室内音響検討を進める上で重要なこととして、室の規模や用途に応じた残響時間の最適設計に加え、聴感上の空間印象への影響が大きい初期反射音制御がある。

◆筆者らは初期反射音の中でも特に側方反射音に着目し、拡がり感との関連性が高い式音響指標値 LF を用いた設計検討を当初から行っており、設計段階での音響シミュレーションによる検討や、双指向性と全指向性のマイクロフォンを組み合わせた自作の測定器を用いて実測での評価を行ってきている。

◆近年のプロジェクトの中から、それぞれ異なる視点から側方反射音に着目して、設計段階のシミュレーションや、現場段階での実験などを通して音響検討を行った実例を紹介する。

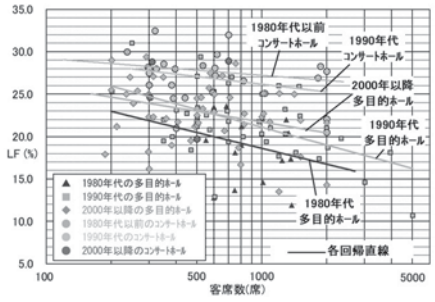


Fig. 2　LF 値（主階席）の時代変化

1-7-5

1-7-5　（招待講演）コントロールルームの音響設計と関連技術に関して

Some remarks on the acoustic design of a control room and their related acoustical technologies

○中原雅考（ソナ）

音響コンテンツ制作を行うスタジオは，音素材を録音するためのレコーディングブースと，それらをスピーカーから再生しながらミキシング等の作業を行うコントロールルームから構成される。本報では，後者の音響設計とその関連技術に関して考察する。

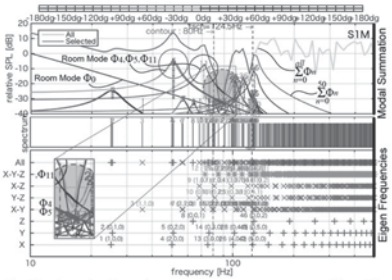


Fig.1 Acoustic Design for low frequency response and the Room Modes.



Fig.2 Acoustic Design for Ultra Early Reflections and the VSV analysis.

1-7-6

1-7-6　カーディオイドマイクを用いた音圧と粒子速度の解析手法

Method for Analyzing Sound Pressure and Sound Particle Velocity Using Cardioid Microphones

○羽入敏樹，星和磨（日大･短大）

◆カーディオイドマイクの応答を音場の音圧と粒子速度を用いて新たに定義し，これを基に c-c 法の測定原理を定式化した。

◆カーディオイドマイクの応答の新定義

$$\begin{aligned} M(\omega) &= \alpha(\omega)P(\omega) - \beta(\omega)\rho c U(\omega)\cdot \mathbf{m} \\ &= \alpha(\omega)P(\omega) - \beta(\omega)\rho c \left\| \mathbf{U}(\omega) \right\| \cos\theta_m \\ &= \alpha(\omega)P(\omega) - \beta(\omega)\rho c U_m(\omega) \end{aligned}$$

ωは角周波数[rad/s]，$M(\omega)$はマイク出力電圧の複素振幅[V]，$P(\omega)$はマイク位置における音圧の複素振幅[Pa]，$\mathbf{U}(\omega)$はマイク位置における粒子速度の複素振幅の空間ベクトル[m/s]，$\mathbf{m}$ はマイクの向き（感度最大の向き）の単位ベクトル，θ_m は$\mathbf{U}(\omega)$と $\mathbf{m}$ とのなす角[rad]，$\|\mathbf{U}(\omega)\|$ は粒子速度ベクトルのノルム，$U_m(\omega)$は$\mathbf{U}(\omega)$と $\mathbf{m}$ の内積であり $\mathbf{U}(\omega)$のマイク向きへの射影としての粒子速度複素振幅[m/s]，$\alpha(\omega)$は音圧に関する無指向性成分のマイクロホン感度[V/Pa]，$\beta(\omega)$は粒子速度に関する両指向性成分$\rho c U_m(\omega)$のマイクロホン感度[V/Pa]，ρc は媒質の特性インピーダンス[kg/m²s]である。$\alpha(\omega)$と$\beta(\omega)$は複素振幅で，使用するマイクロホン固有の値である。

上式は平面波を前提としない。近距離音場，遠距離音場，干渉音場などあらゆる音場を想定することができる。

◆2 本のマイクロホンを用いる c-c 法

音場のある点 g における音圧$P(\omega)$と x 方向の粒子速度成分$U_x(\omega)$を測定することを考える。2 本のカーディオイドマイク M_1、M_2 を点 g において それぞれ x 軸の正と負に向けたときの応答を$M_1(\omega)$，$M_2(\omega)$とする。

2 本のマイクの$\alpha_1(\omega)$，$\beta_1(\omega)$，$\alpha_2(\omega)$，$\beta_2(\omega)$が既知であれば 2 本のマイクを 180 度対向または背向させることによって音場の音圧と粒子速度を求めることが出来る。

$$P(\omega) = \frac{\beta_2(\omega)M_1(\omega) + \beta_1(\omega)M_2(\omega)}{\alpha_1(\omega)\beta_2(\omega) + \alpha_2(\omega)\beta_1(\omega)}$$

$$U_x(\omega) = \frac{\alpha_1(\omega)M_2(\omega) - \alpha_2(\omega)M_1(\omega)}{\left[\alpha_1(\omega)\beta_2(\omega) + \alpha_2(\omega)\beta_1(\omega)\right]\rho c}$$

1-7-7

1-7-7　アンサンブル平均を用いた吸音特性の現場測定法 −試料特性の処理過程について−

An in-situ measurement method using ensemble averaging technique for sound absorption of materials
- Data processing procedure -

○大鶴 徹，富来礼次，岡本則子（大分大･理工）

◆アンサンブル平均を用いる材の吸音特性測定法（EA_{pp} 法）の平均操作など、データ処理過程について検討した

◆EA_{pu} 法と EA_{pp} 法の 2 種の手法で人工芝の吸音特性を残響室で試行的に測定した結果を対象に考察した

◆単一の受音点に関する表面インピーダンスと吸音率の周波数特性には、250 Hz 以下の周波数領域で、2 種の手法による値にばらつきが認められる：Fig.1 参照

◆インピーダンスに関し、実部・虚部ともに、250 Hz ~ 2.8 kHz の周波数領域における 2 種の手法による測定値は概ね近似する

◆吸音率に関しては、250 Hz ~ 1 kHz での近似は良好である

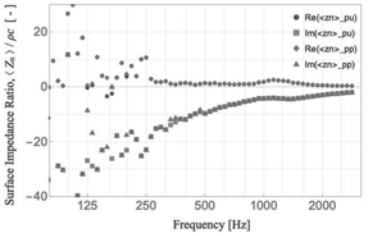


Fig.1: One-twelfth octave-band mean values of surface normal impedance ratio $\langle Z_n \rangle / \rho c$ of artificial grass measured by EA_{pu} and EA_{pp} methods.

1-7-8

1-7-8　C-C 法を用いた音響透過損失測定の試み

Attempt to measure sound transmission loss using the C-C method

◎佐藤 菖，石塚 崇（清水建設技研），星 和磨，鈴木 諒一，羽入 敏樹（日大・短大）

◆ 建築部材の音響透過損失を実験室で測定する方法として、平均音圧レベル差法 (JIS A 1416:2000) と音響インテンシティ法 (JIS A 1441-1) がある

◆ 本報では、音響インテンシティ測定に C-C 法を用いて、試験体の音響透過損を音響インテンシティ法により測定した結果を報告する

◆ C-C 法はカーディオイドマイクロホンを用いて音圧、粒子速度、音響インテンシティ等を測定する方法であり、安価で気温や湿度の影響を受けにくいといった特徴を持つ

◆ 平均音圧レベル差法と C-C 法を用いた音響インテンシティ法の 2 つの手法で二重サッシを含む壁面の音響透過損失を測定した結果、250 Hz 以上の周波数帯域では、ほぼ同等の値が得られた

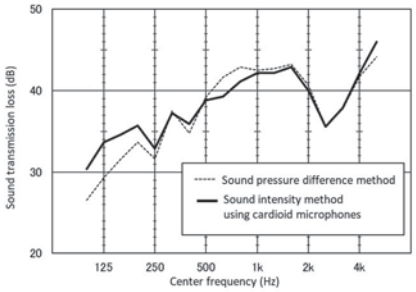


Fig. 1 Measured sound transmission losses using two different methods

1-7-9

1-7-9　衝撃源の違いによる乾式二重床構造の重量床衝撃音レベル低減量の変化に関する数値解析的検討

Numerical analysis on changes of improvement of impact sound insulation of dry-type double floor system due to different heavy impact sources

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

◆JIS A 1440 - 2 に準拠した二重床試験体の数値解析モデルを構築し，二重床支持脚の非線形モデルが ΔL_H の変化に及ぼす影響を検証した。

◆床スラブと床面材，支持脚，床下空気層をそれぞれ厚板要素，Kelvin-Voigt モデル，音響要素によりモデル化した。

◆時刻歴応答解析による音響振動連成解析。

◆防振ゴムの変形の程度による支持脚ばね定数の変化は ΔL_H の変化に影響はあるが，決定的な原因ではないと考えられる。

◆支持脚の引張時におけるばねの作用が弱いことが ΔL_H の変化の主因であることを確認した。

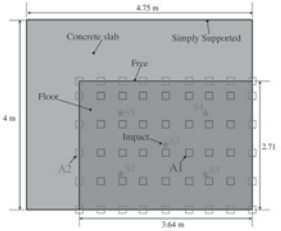


Fig. 1 Numerical model.

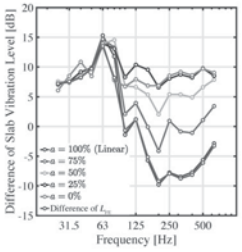


Fig. 2 Simulated difference between L_V using tyre and L_V using ball when including Factor 1.

1-7-10

1-7-10　タイヤの衝撃力周波数特性をゴムボールにより再現するための防振材の選定と計測結果

Selection of vibration isolators and measurement results for reproducing the impact force frequency characteristics of the car-tire with the rubber ball

○冨田隆太（日大・理工），△奥山雄太（日大院・理工）

◆本研究は，文献[1]の D.3 に示されている衝撃力特性（Ⅱ）に相当するタイヤ 20 cm 落下時の衝撃力特性を，JIS A 1418-2:2019 の衝撃力特性（2）のゴムボール落下により再現できる方法を提案することを目的とする。

◆Fig. 1 を見ると，タイヤ 20 cm と防振材 A の 37.5 mm 厚（実験 No. 9）が良い対応を示していることがわかる。絶対値で見ると，タイヤ 20 cm に比べて，ゴムボール 100 cm＋防振材 A（37.5 mm）は，31.5～250 Hz 帯域で 0.9 dB～1.8 dB 小さいという結果であり，ほぼ同様の衝撃力周波数特性が実現できたと考えられる。

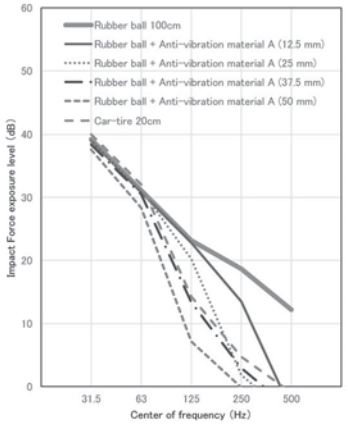


Fig.1 : Impact force frequency characteristics of different thicknesses of vibration isolators

1-7-11

1-7-11　乾式二重床の床衝撃音レベル低減量を対象とした標準衝撃源の互換性に関する検証

Verification of the compatibility of standard impact sources for reduction of transmitted impact sound of dry-type floating floors

○冨田隆太（日大・理工）

◆本報では，既報[1~10]で提案した乾式二重床の床衝撃音レベル低減量の推定方法について，既報[1~10]の 3 種類とは別の乾式二重床を対象として，検証した結果を報告する。

◆Fig. 1 を見ると，タッピングマシンからゴムボールの推定値とゴムボールの実測値について，推定した 500～2000 Hz の結果は 2.5 dB 以内の差であった。タイヤからゴムボールの推定値とゴムボールの実測値について，推定した 125～500 Hz の結果は 2.7 dB 以内の差であった。

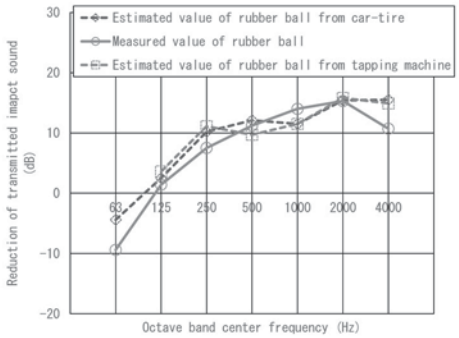


Fig.1 : Measured and estimated Reduction of transmitted floor impact sound level of rubber ball for dry-type floating floors (1st.)

1-8-1

1-8-1　オープンプラン型オフィスにおける音環境が作業種ごとの印象評価に与える影響についての基礎的検討

A Basic Study on the Influence of the Sound Environment on the Impression Evaluation of Each Work Type in an Open-Plan Office

☆中橋 樹香，荒木 拓己，△山野 拓哉，菅原 彬子，△長澤 康弘，平栗 靖浩（近畿大），△冨樫 建五（大和ハウス工業），原田 和典（岡山県立大），△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、作業種の違いによる検討を行った。

◆複数人作業者は個人作業者に比べ、音環境に対してより寛容な傾向が見られた。

◆作業人数別に異なる作業種間の比較を行った結果においても、同一の音環境下であっても評価に明確な違いが見られ、特に個人作業においては、高集中を要する作業は有意に音環境の影響を受けることが明らかとなった。（Fig.1）。

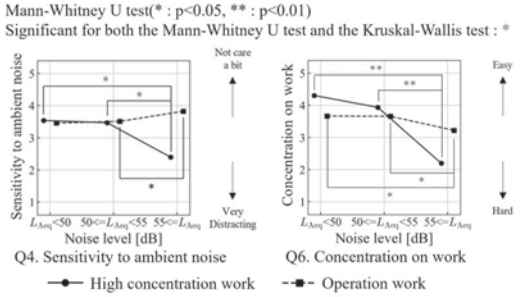


Fig. 1：Comparison of the relationship between impression evaluations and noise levels for different work types in solo work

1-8-2

1-8-2　オープンプラン型オフィスにおける騒音レベル分布の計測手法に関する検討

Study on measurement methods of distribution of noise levels in an open-office

○平栗靖浩，△山野拓哉，中橋樹香，荒木拓己，菅原彬子，△長澤康弘（近畿大）
△冨樫建五（大和ハウス工業），原田和典（岡山県立大），△岩切幸伸（コクヨ）

◆オープンプラン型オフィスにおける騒音レベル分布の計測とその評価方法について検討を行った。

◆6 日間にわたり実在するオープンプラン型オフィス内の 15 地点で騒音レベルを計測し、そのデータを基に補間手法による騒音レベル分布の推定を試みた。

◆オフィス内の騒音レベル分布を推定するための補間手法として、普通クリギングが最も高い補間精度を示した（Table 1）。

◆計測点位置に関する検討では、計測から排除しても補間により精度良く推定できる地点があることがわかった。

Table 1 Root mean square error of measured and interpolated noise levels at 15 measurement points

	Averaging time length for L_{Aeq} [min.]				
	5	10	20	30	60
NN	2.9	2.8	2.6	2.6	2.5
IDW	2.4	2.3	2.2	2.1	2.0
OK	2.3	2.2	2.1	2.0	1.9
RBF	2.5	2.4	2.3	2.2	2.1

1-8-3

1-8-3　オージオメータによる耳栓フィットテストの提案

Proposed fit test for sound attenuation of earplugs applying audiometer

○横山 栄，小林知尋（小林理研），武田 葵，中市健志（リオン）

昨年、騒音障害防止のためのガイドラインが改訂され、各騒音職場でJIS による遮音値を参考に適切な聴覚保護具を選定し、着用することが義務付けられたが、特に耳栓については実際の装着状態によって遮音性能は著しく異なり、個人ごとの遮音性能を把握することが重要である。そこで筆者らは騒音職場の健康診断でも用いられるオージオメータ（RION, AA-K1A；耳覆い型イヤホン）によるフィットテストを提案した。併せて、簡易測定器（3M™ E-A-Rfit™）による挿入損失測定、国際的に標準となっている REAT 測定も 2 種類の耳栓を対象に、各 16 名の協力を得て実施した。その結果、提案法の遮音値（Fig.1）のうち、1000 Hz単独あるいは4000 Hzを加えた2周波数による*SNR*値は全帯域の*SNR*値と相関係数が 0.8 以上で、その結果で耳栓を選定しても REAT 試験による*SNR*値を用いた場合と結果が顕著に異ならない（ただし判定不可を含む）ことがわかった。さらに騒音職場での検証も実施する予定である。

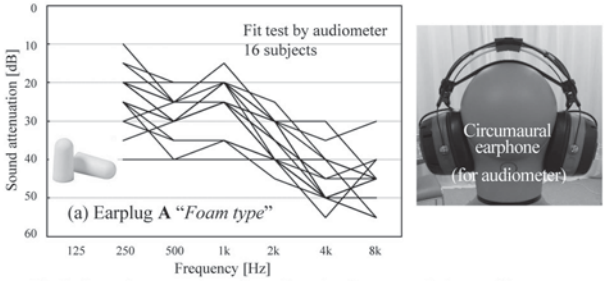


Fig.1：Sound attenuation of earplugs by fit test applying audiometer.

1-8-4

1-8-4　耳道構造を考慮した HATS を用いた聴覚保護具の遮音性能特性測定の試み

Attempt of sound attenuation measurement of hearing protection earplugs using HATS with considered ear canal structure

○稲永 潔文（サザン音響）

◆作業環境の騒音暴露による聴力損失を防止するため、様々な耳栓や聴覚保護具が用いられている。そしてその騒音低減量等の性能特性評価には REAT と呼ばれる主観評価法が、一方ノイズキャンセル型ヘッドホン等のそれは JEITA RC-8142A として物理評価法が規定されている。

◆従来、官能検査で行われてきた聴覚保護具の遮音性能特性測定を、耳道構造（構成素材）を考慮した HATS を用いた物理測定で行うことを試みた。

◆被測定イヤーマフには 3M 社“3MTM H9A（H520A）”、耳栓には同“E-A-R TM”318-4001（紐付き）”等を、HATS にはサザン音響 “SAMAR Type 4500”、また測定用音場には RC-8142A の規定を満たすサザン音響“小型疑似拡散 Box Type1020”を用いた。



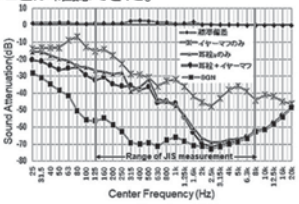
Externals of HATS used for measurement with hearing protector & with earplugs

◆イヤーマフ測定の結果、耳道素材が変化しても遮音特性や分散値も大きく変わらず、従来主観評価値と近似した結果となった。

◆耳栓測定では、何れの周波数でも遮音特性は向上し 標準偏差も大きく減少したが、主観評価値とは異なる結果となった。これは、耳道部と耳栓の密着性の向上が要因で、実耳では実現し難い遮音特性と思われる。

◆イヤーマフと耳栓を併用した場合、各周波数で遮音量が大きい方の遮音特性に依存し、また併用効果もあることが確認できた。

◆主観評価法では、ヒトの耳道に耳栓を強く挿入する結果、生理的雑音によるマスキングと耳道変形が、物理評価法の結果と異なる要因と思われる。



Measured background noise and sound attenuation and SD of Earplugs,Earmuffs and using it together by JEITA RC-3142A

1-8-5

1-8-5　一般道路の排水性舗装における自動車走行騒音パワーレベル －全国の直轄国道での測定結果－

The sound power level of road vehicles on a general road with porous asphalt pavement　- Measurements taken on national roads　-

○澤田 泰征，△橋本 浩良(国総研 道路環境研)，△大城 温(国総研 企画部)

◆2020～2022 年度に全国の直轄国道で，一般道路の排水性舗装における自動車走行騒音パワーレベルを測定した結果を報告する。

◆舗設後 5 年程度では，密粒舗装と比べて低い傾向で騒音低減効果がみられた。舗設後 15 年程度では箇所により騒音低減効果に差がみられた。

◆今後も測定を継続し，一般道路の排水性舗装におけるパワーレベルの実態を明らかにしていく予定である。

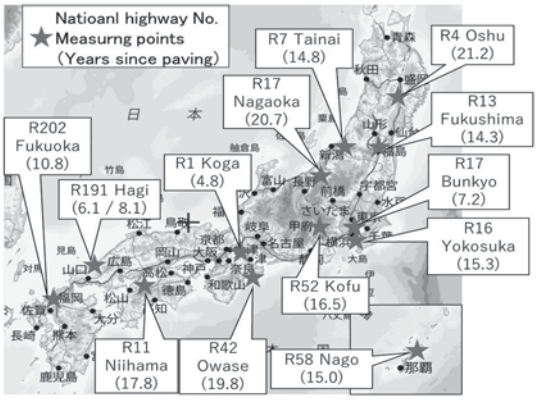


Fig.1：Measuring Points

1-8-6

1-8-6　走行速度による非定常走行区間の設定が道路交通騒音の推計結果に及ぼす影響

Influence of the setting of unsteady driving section by driving speed on the estimation results of road traffic noise

◎原田和典(岡山県立大)，△小村風我，平栗靖浩(近畿大)，大嶋拓也(新潟大)，齊藤由典，△跡部哲士(日本ミシュランタイヤ)

◆道路交通ノイズマップ推計において，速度を条件に，一定の速度以下の場合に非定常走行区間とする推計手法について検討した.

◆推計結果より，非定常走行区間を考慮した計算結果は全体的にRMSE が改善した.

◆また，非定常走行区間を 20 km/h 以下とした場合に最も良い結果が得られたが，走行状態以外の誤差要因についても改善を進めていく必要があることが示された.

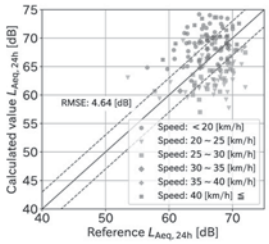


Fig.1：Comparison Reference vs. calculation value (all section calculated as stationary running state)

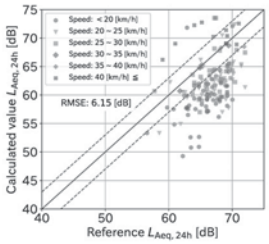


Fig.2：Comparison Reference vs. calculation value (non-stationary running state under 20km/h driving speed)

1-8-7

1-8-7　高い位置にある音源による建物群の挿入損失に関する縮尺模型実験

Scale model experiment on insertion loss of buildings in the case of high-position sound sources

○坂本 慎一(東大生研)，永田 茉莉咲，王 麒源(東大大学院)，矢野 博夫(東大生研)，穴井 謙(福岡大)

◆建物群による挿入損失に関して，藤本らの考え方を踏襲し高い音源位置の場合への拡張を目指し，縮尺模型実験を用いた検討を行っている。

◆音源が高い位置にある場合，上方からの直達音や浅い回折音の寄与が優勢になる場合がある。その場合，音源列と受音点で規定される斜平面を計算領域とし，その平面と建物の相貫面をASJ RTN-Model 2018 における建物の平面と見做して計算することによって計算精度が改善した。

◆今後，さらに多くのケースについて検討を続ける。挿入損失データは周波数毎に収拾し，道路交通騒音とは異なる周波数特性をもつ音源にも適用可能となるように整理と検討を進める予定である。

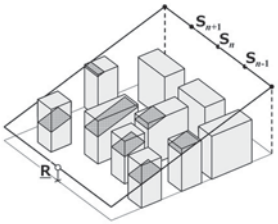


Fig. 1 Revised calculation method

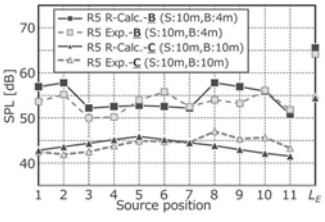


Fig. 2 An example of comparison result

1-8-8

1-8-8　ヘリコプタ動作音の特徴に基づくドローン衝突回避 －第１報 振幅変調音に基づく音源探知－

Drone collision avoidance based on the characteristics of helicopter operation sound - First report: Sound source detection based on amplitude modulated sound -

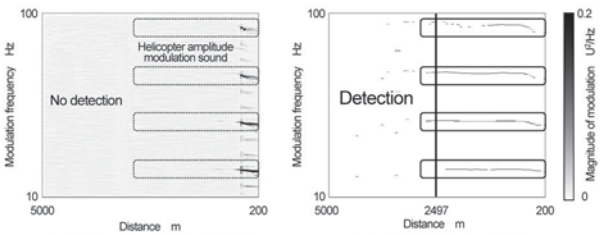
☆星野瑛彗(中央大院)，△田辺総一郎，戸井武司(中央大)

◆ ドローンの普及に伴い，高度 150 m 未満を飛行するヘリコプタとの衝突の危険性が高まり，ドローンに搭載するヘリコプタ探知の小型軽量かつ低消費電力センサが求められている。

◆本研究では，ヘリコプタ振幅変調音に基づく音源探知を行う。ヘリコプタ振幅変調音が存在する特定の変調周波数帯に着目することで，ドローン動作音が混在する中からヘリコプタの音源探知が可能となる。

◆マイクロホンアレイを用いたヘリコプタ動作音の強調処理と，ヘリコプタ振幅変調音の探知アルゴリズムより，ヘリコプタ以外に起因する振幅変調を抑制し，Fig. 1 に示すヘリコプタ振幅変調音の特徴である回転次数成分のみを強調できる。

◆実測音の変調周波数より，提案する探知アルゴリズムを適用することで，ドローンから約 2.5 km のヘリコプタを明確に探知できた。



(a) Before applying detection algorithm　(b) After applying detection algorithm

Fig. 1　Relation between helicopter and drone separation distance and modulation frequency

1-8-9

1-8-9　ヘリコプタ動作音の特徴に基づくドローン衝突回避 －第２報 振幅変調音に基づく方向推定－

Drone collision avoidance based on the characteristics of helicopter operation sound - Second report: Direction estimation based on amplitude modulated sound -

☆星野瑛彗(中央大院)，△田辺総一郎，戸井武司(中央大)

1. ヘリコプタ振幅変調音に基づくドローン衝突回避のためのヘリコプタの音源探知を行う先行研究では，ヘリコプタ振幅変調音の有無を検出するのみで，その方向推定は検討されていない。
2. 本研究では，マイクロホンアレイの振幅変調に基づく指向特性を用いて，ヘリコプタ振幅変調音の方向推定を行う。
3. Fig. 1 に示す３組のマイクロホンアレイを 60 °毎に組み合わせ，固定遅延和ビームフォーマを適用し，Fig. 2 に示す振幅変調に基づく指向特性が得られる。
4. シミュレーションによりヘリコプタが 60 °方向に存在するときに，Fig. 3 に示す Array 3 の 30 °~ 90 °方向に振幅変調に基づく指向特性の最大値を得られ，ヘリコプタが接近する方向推定が正しくできた。

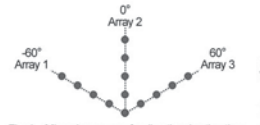


Fig. 1　Microphone array for directional estimation

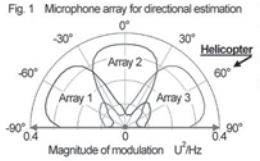


Fig. 2　Directivity characteristics based on amplitude modulation

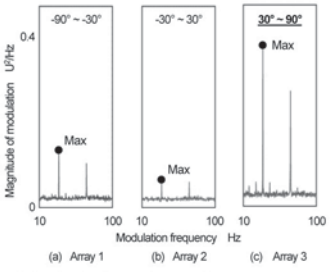


Fig. 3　Directional estimation results for each array when the helicopter operating sound direction is 60°

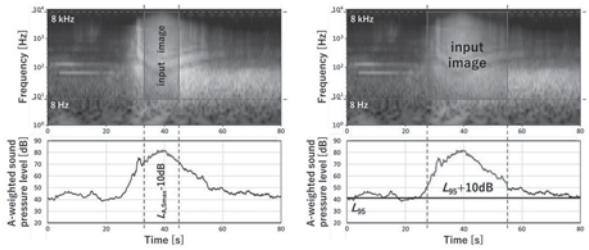
1-8-10

1-8-10 航空機騒音の AI 識別：妨害音重畳判定の分析手順に関する一検討

Artificial Intelligence identification for aircraft noise: A study on the analysis procedure in judgment of overlapped other noise

☆小泊楓奈，森淳一，土屋健伸（神奈川大），森長誠（大同大），山元一平（防衛基盤整備協会），野澤竜祐，花香和之（成田国際空港振興協会），川瀬康彰（成田国際空港振興協会/小林理研）

- ◆航空機騒音に係る調査において有人で行われている単発騒音の状態の観測を自動化するために、音響情報を用いた CNN による航空機騒音識別を行っている
- ◆前報では、”妨害音が重畳する航空機騒音”のカテゴリー追加による識別精度の低下、ならびにこれが単発騒音のスペクトログラムを描画する時間範囲に起因する可能性が示唆された
- ◆本報では、航空機の単発騒音を切り出す時間範囲（Fig. 1）が CNN の識別精度に与える影響を検討した
- ◆一度に全カテゴリーを分類した場合の精度はわずかに低下した
- ◆しかし２段階で分類を行った場合、精度は若干向上した



(a) $L_{A,Smax}$-10 dB　(b) L_{95}+10 dB

Fig. 1 Sample of the time and frequency ranges of for the input image

1-8-11

1-8-11　DTW を用いた飛行経路の分類による航空機騒音の推計

Estimation of aviation noise with classification of flightpath by DTW

☆小野弘郁(近畿大院)，原田和典(岡山県立大)，菅原彬子，平栗靖浩(近畿大)

- ◆航空機騒音の予測には、代表的飛行経路を設定することが一般的であるが、確立された設定方法は無い。
- ◆本研究では、代表的飛行経路の設定方法として、**3** 次元の **DTW** (Dynamic Time Warping，動的時間伸縮法)を用いた手法を提案する。
- ◆2023 年 4 月 18 日の大阪国際空港の離陸機 62 便の経路を検証する。
- ◆従来の代表的経路の設定方法である主観による代表的経路(fp_sub)と、2 次元平面上での DTW を用いた代表的飛行経路(fp_dtw2d)に基づく騒音レベル、3 次元の DTW による代表的飛行経路(fp_dtw3d)を、実測経路(fp_raw)に基づく騒音レベルと比較することにより妥当性を検証する。
- ◆2 次元平面上で DTW の類似度に基づき 5 種類のクラスターに分類し、クラスターごとに、飛行距離と高度のデータで同様に 2 種類のクラスターに分類する。クラスターごとに DBA(DTW Barycenter Averaging)により、**10** 種類の **3** 次元の代表的飛行経路を設定する。

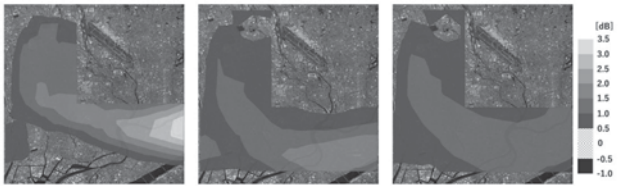


Fig.1：fp_sub-fp_raw　Fig.2：fp_dtw2d-fp_raw　Fig.3：fp_dtw3d-fp_raw

1-8-12

1-8-12　指向性を有する移動音源の２次元 FDTD シミュレーション

Two dimensional FDTD simulation of moving source with directivity

○土屋 隆生 (同志社大・理工)
Takao Tsuchiya (Doshisha Univ.)

任意の指向性を有する移動音源について検討した。音源が移動する場合の放射音を理論的に導出したところ，振幅項だけでなく位相項にも音源移動の影響があることが示された。つぎに，2次元 FDTD 法による数値実験を実施した。2音源対とリニアアレイについて，移動した場合の指向性を計算したところ，理論と良好に一致し本検討の妥当性が示された。

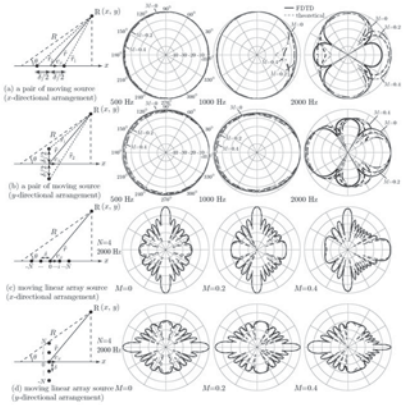


Fig.: Directivity of moving sources.

1-9-1

1-9-1　正弦加音による変動感を活用したポンプ動作音の異常判定

Abnormality judgment of pump sound
using fluctuation feeling by adding sine wave

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

◆ ポンプのような回転機器の異常を音で診断するには，分析器などのコストが必要になる。本研究では異常時に音量が増大した純音成分（以下，異常純音）に正弦波を干渉させ，人が聴くことで異常と判定することを目指す。

◆ 変動を感じる低速変調音，粗さを感じる高速変調音，不協和音の警告感に関する５段階の相対評価により，異常判定には変動感を与える低速変調音が有効であることを把握し，異常純音に対する正弦加音の周波数差は４Hz と定めた。

◆ 変調度の異なる低速変調音に対して同様に評価し，2 純音の変調度と音圧レベル差の関係において異常判定の閾値とする変調度を把握することで，異常純音に対する正弦加音の音圧レベル差を定めた。

◆ 上記の条件で定めた正弦加音を正常機と異常機に適用して印象評価を行い，警告感の評価値を Fig. 1 に示す。これより，正弦加音による変動感を想起させる異常判定の有効性を示した。

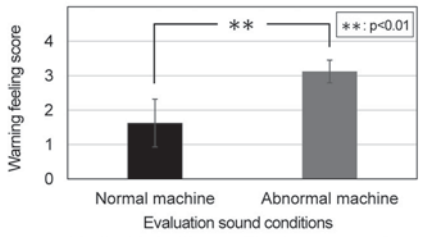


Fig.1　Evaluation results of pump sound

1-9-2

1-9-2　定常純音に着目した協和音付加によるポンプ快音設計

Comfortable sound design for pump by adding consonance
focused on steady pure tones

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

◆ ポンプ動作音が有する純音成分は不快感の原因となるため，定常純音に着目して協和音付加による快音設計を目指す。

◆ トナリティが大きい顕著な定常純音に対して，協和音付加前後で快適感に関する相対評価を行った。250 Hz，1000 Hz のような低～中周波の顕著な定常純音に対してはトナリティとシャープネスが低い低周波側への協和音付加により快適感が向上した。また，複数の顕著な定常純音に対しては，トナリティが異なる場合はトナリティが大きい純音へ，トナリティが同程度の場合は高周波の純音へ協和音付加することで，快適感が向上することを確認した。

◆ Fig. 1 に示す顕著な定常純音を有するポンプ動作音へ協和音付加を行った。周波数f_{pro}の顕著な定常純音に対し，これと周波数比が 2:3 となる純音f_{P5}を付加し，無加工の音 Ori.P と協和音付加後の音 Con.P に対して SD 法による絶対評価を行った。評価結果に対する因子分析より，Fig. 2 に示す通り顕著な定常純音を有するポンプ動作音などへ協和音付加による快音設計の可能性を示した。

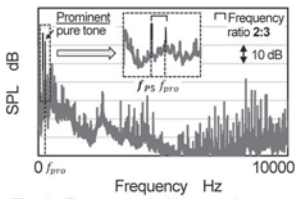


Fig. 1　Pump sound with a prominent steady pure tone

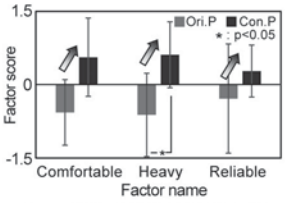


Fig. 2　Evaluation results for pump sound

1-9-3

1-9-3　振幅変調純音に着目した協和音付加によるポンプ快音設計

Comfortable sound design for pump by adding consonance
focused on amplitude modulated pure tones

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

◆ ポンプ動作音が有する純音成分同士が干渉して生じる振幅変調純音は不快感の原因となるため，振幅変調純音に着目して協和音付加による快音設計を目指す。

◆ トナリティが大きく振幅変調する顕著な振幅変調純音に対して，協和音付加前後で快適感に関する相対評価を行った。250 Hz，1000 Hz のような低～中周波の顕著な振幅変調純音に対して協和音付加により快適感が向上する可能性を示した。また，顕著な定常純音と振幅変調純音のトナリティが同程度の場合，振幅変調純音への協和音付加で快適感が向上することがわかった。

◆ Fig. 1 に示す顕著な振幅変調純音を有するポンプ動作音へ協和音付加を行った。トナリティが大きい顕著な純音成分f_{pro_m1}，f_{pro_m2}，f_{pro_m3}のうち最も低周波の振幅変調するf_{pro_m1}に周波数比 2:3 の協和音f_{P5L}:f_{P5H}を付加し，無加工の音 Ori.P と協和音付加後の音 Con.P に対して SD 法による絶対評価を行った。評価結果に対する因子分析より，Fig. 2 に示す通り顕著な振幅変調純音を有するポンプ動作音などへ協和音付加による快音設計の可能性を示した。

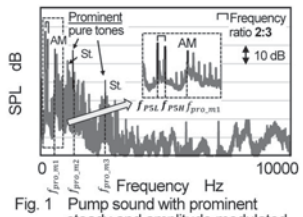


Fig. 1　Pump sound with prominent steady and amplitude modulated pure tones

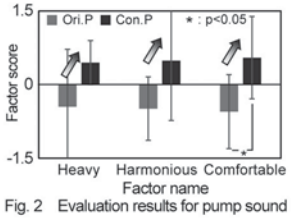


Fig. 2　Evaluation results for pump sound

1-9-4

1-9-4　スイッチ動作における聴覚および触覚による感覚統合

In switch operation creation of complex senses
through touch and hearing

☆渡瀬由人（中央大院），△大友貴史，△下村尚登，
△坂本秀樹（アルプスアルパイン），戸井武司（中央大）

◆本論文では，プッシュスイッチを動作対象として，聴覚の印象と触覚の印象がどのように統合されて複合感覚として感じるかを，それぞれの感覚の寄与率から把握する。

◆単感覚における印象評価を行い，重回帰分析より計算された聴覚における物理量ごとの標準偏回帰係数を Table 1 に示す。表より重い印象は周波数，硬い印象は立ち上がり時間など，印象によって有意にはたらく物理量が変化していることが明らかになった。

◆複合感覚における印象評価を行い，重回帰分析を用いて聴覚と触覚の標準偏回帰係数より寄与率を算出し，Fig. 1 に示す。図より重い印象は聴覚と触覚で五分五分，硬い印象は触覚の影響が大きいなど，印象によって聴覚と触覚の寄与率が異なることを把握した。

Table 1　Multiple regression analysis in hearing

目的変数	説明変数			自由度調整済み決定係数
形容詞	a周波数	b立ち上がり時間	c減衰時間	
重い	-0.80*	-0.14	-0.14	0.62
硬い	0.08	-0.83*	-0.20	0.66
はっきりした	0.65*	0.09	0.32*	0.39
好ましい	-0.49*	-0.43*	0.17	0.35

*: p<0.05

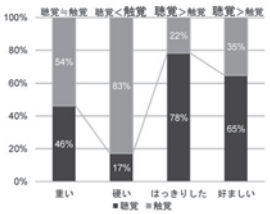


Fig. 1　Contribution per adjective

1-9-5

1-9-5 学会発表にふさわしいタイマー音をデザインして，Inter-Noise 2023 で使ってみました。

We designed a timer sound appropriate for conference presentations, and used it in Inter-Noise 2023.

○山内勝也（九州大・芸工），△清川進也（エイティーエイト），
船場ひさお（駿河台大／ADC），高橋莉紗，石田康二（Sound One），
岩宮眞一郎（日大・芸／九州大・芸工）

学会タイマー音は，なぜ「チーン」なの？



○ 決してそうでなくてもいいはずですが，
ベル音の回数で時間を知らせる方式が
広く使われています。
ただ，その音・方式がベストでしょうか？

講演の終わり時間に
「チーン！チーン！チーン！」と大きなベル音が鳴り響き，
聴講者はその音に驚き，講演者は焦り…
そのような気まずい空気は，その音のせいかもしれません。

そこで我々は，
学会発表にふさわしいタイマー音をデザインして，Inter-Noise 2023 で使ってみました。

○ デザインプロジェクトの経緯と，実践を通じて得た知見を報告します。発表では制作した音をデモンストレーションしますので聴いてみてください。

（音響学会でも使ってみませんか…？）

1-9-6

1-9-6 自動運転車内における次世代サイン音設計過程の分析と評価

Analysis and Evaluation of Design Process for Auditory Signs for Future Vehicle

☆小越安奈，星隈泉穂，山内勝也（九州大・芸工），
△田上宣昭，△磯崎賢太，△石黒留衣，△川田歩（パイオニア），
△林真琴（東北パイオニア）

◆ 新たなサイン音を設計する際の設計過程を整理・一般化することを目的とし，サイン音設計過程の分析と印象評価実験を行った。

◆ 以前の研究では，設計者と聴取者の間にサイン音の使用イメージに齟齬があったため，本研究では各実験の教示に物語型の情報提示（シナリオ）を用いた。

◆ 移入尺度や未来感などの評価が高かったシナリオを教示に採用し，音作りの経験を持つ学生にシナリオ中に登場するサイン音の設計を求め，その過程を記録・分析した。(Fig.1)

◆ 設計されたサイン音の印象評価実験では，サイン音のシナリオに対する適切性や未来感は迫力性因子の相関が高く，新しさは時間構造因子と相関が高いことが示された。

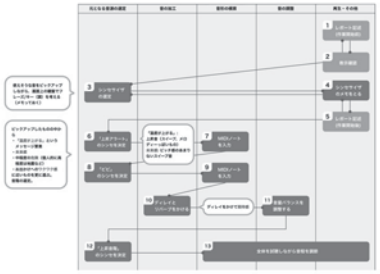


Fig. 1　Flow diagram of designing for auditory sign.

1-9-7

1-9-7　自動車に関する情報がドア閉め音の聴取印象に与える影響

Effects of information about the car on the auditory impression of door-closing sound.

☆森田洋明，丸山拓記，高田正幸（九大・芸工）

◆ ドア閉め音の評価前に自動車に関する情報を提示して，評価対象の音をイメージさせた上で，音の評価を実施

◆ 実験参加者を，「エコノミーカー」「高級車」の車格情報を伝えられた Gr.1・2 と，車格情報の提示なしの Gr.3 の 3 つのグループに分けた

◆ 実験参加者には車格情報を基にドア閉め音をイメージしてもらい，16 の評価尺度上で評価した結果，高級車はエコノミーカーより，小さい，低い，重いなどと評価された

◆ ドア閉め音 6 種類の刺激について，イメージ同様に評価を実施した結果，高級車として評価するとエコノミーカーとして評価するよりも，より閉まった感じ，高級感があると評価された

◆ イメージと音の評価結果の比較を行った結果，例えばエコノミーカー条件ではイメージよりも音の方が高級という印象で，期待不一致モデルに当てはめるとこのような場合に満足度が高まると考えられる

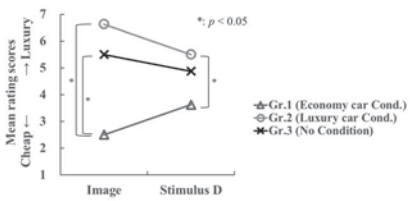


Fig. 1 Comparisons of impressions between door-closing sound image and stimulus D for each vehicle evaluation condition.

1-9-8

1-9-8　暗騒音変化に基づく変速段ごとの変速音設計による加速感向上

Improving acceleration feeling by designing gearshift sound for each gearshift stage based on changing background noise

☆五十嵐啓紘（中央大院），戸井武司（中央大）

◆ 内燃機関車（以下，ICEV）と電気自動車（以下，EV）の暗騒音変化に基づく，変速段ごとの変速音設計により加速感の向上を図る。

◆ 暗騒音の音圧は車速と共に大きくなるため，変速音の音圧を変速段ごとに大きくすることで，加速感向上が期待される。印象評価の結果，駆動システムによらず暗騒音から変速音の音圧突出変化量が変速段ごとに+2~5 dBA で加速感が向上することがわかった。

◆ エンジン／モータ音は回転数と共に高周波成分になるため，変速音の音質を変速段ごとに高周波成分にすることにより，加速感向上が期待される。印象評価の結果，駆動システムによらず暗騒音から変速音のシャープネス突出変化率が変速段ごとに+3~6 %で加速感が向上することがわかった。

◆ ドライビングシミュレータを用いて Table 1 に示す変速段ごとの変速音の音圧および音質の条件により，Fig. 1 に示す ICEV における一体感と加速感の向上を確認した。また，EV についても同様に一体感と加速感の向上を確認した。

Table 1　Good condition of gearshift sound

Driving system	Parameter	Gearshift stage		
		1→2	2→3	3→4
ICEV	Freq. range	±0 kHz	+1 kHz	+1 kHz
	SPL	±0 dB	+8 dB	+3 dB
EV	Freq. range	±0 kHz	+2 kHz	+2 kHz
	SPL	±0 dB	+10 dB	+5 dB

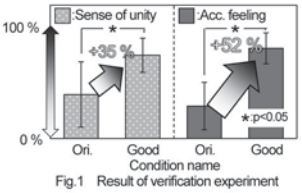


Fig.1　Result of verification experiment

1-9-9

1-9-9　変速段ごとの感性調和設計による変速操作感の創生

Creating gearshift operation feeling
by sensitive balance design for each gearshift stage

☆五十嵐啓紘（中央大院），戸井武司（中央大）

◆内燃機関車（以下，ICEV）と電気自動車（以下，EV）の変速段ごとの変速音に対応した変速ショックにより，変速操作感の創生を行う。

◆ICEV の 1 回の変速における変速音と変速ショックの感性調和設計に関する研究では，変速音の音圧 10 dBA に対して変速ショックの加速度を 0.10 G 変化させると一体感が向上することが示されている。

◆先行研究の結果より，変速段ごとの変速ショックの前後加速度が変速音の音圧変化に対して 0.005，0.01，0.02 G/dBA で変化する条件で印象評価を行うと，駆動システムによらず 0.01 G/dBA で一体感が向上することがわかった。

◆以上をもとに，変速段ごとに変化のない Ori.，変速音のみ変化する Sound，変速ショックのみ変化する Shock，変速音と変速ショックが変化する S&S で検証実験を行った。Fig. 1 より，変速ショックより変速音の方が変速操作感への寄与が大きいが，変速音と変速ショックの感性調和設計により，さらに変速操作感が向上することを把握した。

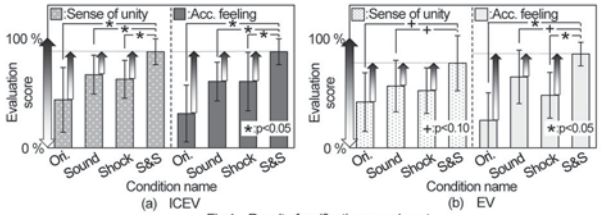


Fig.1　Result of verification experiment

1-9-10

1-9-10　コロナ禍における生活環境と音楽活動の変化

Changes in life environment and musical activities
during COVID-19 pandemic

○寺澤洋子(筑波大学)

◆2020 年のコロナ禍による移動制限などについて、その期間における生活様式や環境の変化、および環境音や音楽の認知、音楽の聴取時間などについて大規模なオンライン調査を行った。

◆調査は 2020 年 6 月に行われ、国内外の音響系のメーリングリストを通じて参加者を募ったほか、Amazon Mechanical Turk を通じて参加を募集した。620 名が参加し、男女比は 6:4，アメリカ、インド、日本などからの参加が多かった。

◆回答に対して、共分散構造分析を行った結果を図 1 に示す。これにより、ストレスやライフスタイルといった要因が、音楽活動に影響を及ぼしていたことが明らかになった。

◆目的のある音楽の利活用としては、「感情の調節」「ムードを変える」「リラックスする」といったものが多かった。

◆近隣の騒音が不快だ、という回答は、近隣の騒音の明瞭さではなく音楽や環境音への注意深さや、音楽情動と有意に相関していた。

◆すべてのグループにおいて、インターネットを通じた音楽聴取が増加していた。



図 1.　共分散構造分析の結果

1-9-11

1-9-11　オンラインライブにおけるステージや配信環境*

* Stage and streaming environment setup for live streaming concerts

○駒澤大介，三戸勇気，川上央(日大・芸術)

オンラインライブは，新型コロナウイルス感染症 (COVID-19，以下コロナ)の広まりにあわせて，本格的なものが多々実施されるようになった．2024 年現在では当たり前のように配信されるようになったオンラインライブではあるが，コロナ禍以前はクオリティーが高いものではなかった．そもそもコロナ禍以前のオンラインは補助的な扱いになっており，オンラインにあまり慣れていない状況下で日本はコロナ禍に突入することとなった．コロナ禍の影響を受けたものとして，ライブやコンサートも同様で，コロナ禍のイベントは各地で中止が発表された．そこで代替で考えられた手法の一つがオンラインライブである．本稿では，コロナ禍で本大学が行った試みを紹介し，オンラインライブの今後について考察していく．

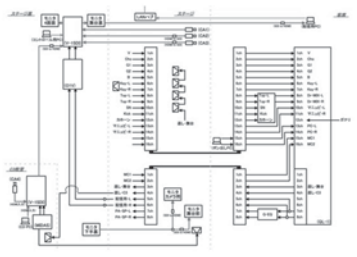


Fig.1 Line diagram for online live event

1-9-12

1-9-12　時間周波数解析に基づく機械学習による車室内音の特徴量抽出

Feature extraction of cabin sound by machine learning
based on time-frequency analysis

☆大島遥汰（中央大院），△田辺総一郎，戸井武司（中央大）

◆畳み込みニューラルネットワーク（CNN）と準再帰型ニューラルネットワーク（QRNN）を併用して，車室内音を想定した時間周波数解析の画像を分類し，勾配加重クラス活性化マッピング（Grad-CAM）に構造的類似性指数（SSIM）を適用することで，グループ間で共通する特徴量を抽出することを目的とする。

◆CNN-QRNN を用いることで，2 種類の時間周波数解析の画像において振幅変調の有無を識別することができる。また，Grad-CAM を用いることで，各時間周波数解析の画像の特徴量を抽出することができる。

◆ 各グループのすべての画像に Grad-CAM を適用しグループごとの平均値を求め，SSIM を用いることで Fig. 1 に示す音源間での共通特徴量を抽出することができた。

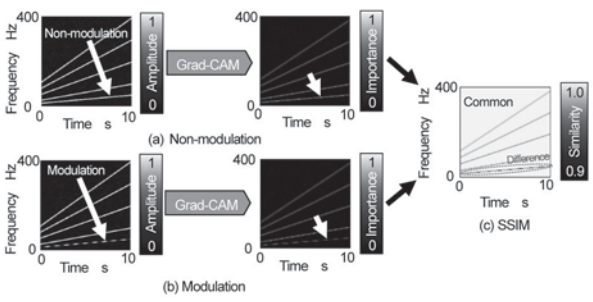


Fig. 1　Grad-CAM and SSIM of two groups of sounds

1-9-13

1-9-13　EV 加速時における加速意志に適合した車室内純音の創生

Creation of pure tone in the cabin that matches the acceleration intention during EV acceleration

☆青柳洸希，柴橋和真（中央大院），△金堂雅彦，戸井武司（中央大）

◆ 自動車走行時の運転者は，視聴覚刺激および運転動作を伴う複合刺激下にある。本研究では，緩加速および急加速の際の加速意志に適合する暗騒音からの純音突出量を解明する。

◆ 実験手順は，緩加速および急加速に対して，純音が最小可聴音圧レベルで検知される時の突出量を意味する純音突出量 Lp を用いて，Lp+0 dB から，Lp+21 dB まで 3 dB ずつ上昇させた計 8 パターンとする。評価方法は，評定尺度法による 7 段階の絶対評価で，評価項目は「加速している感覚に合わない（1 点）」から「加速している感覚に合う（7 点）」とする。

◆ 実験結果として，Fig. 1 に示すように，加速意志に適合する純音突出量は，緩加速と急加速の加速意志の差異によって 3.6 dB 差があることを見出せた。

◆ 加速意志の強い条件の方が，車両の速度変化を純音成分の周波数変化として認知し易くなるため，より純音突出量が大きい方が適切と考えられる。

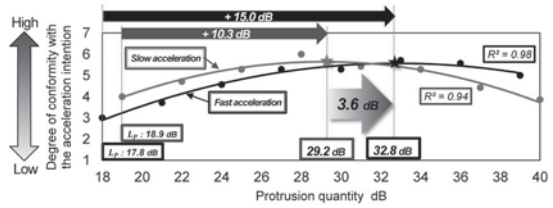


Fig. 1　Relationship between the degree of conformity with the acceleration intention and the protrusion quantity

1-9-14

1-9-14　EV 加速を想定した純音成分の臨界帯域通過速度およびラウドネス変化に基づく許容閾値

Allowable threshold based on critical band passing velocity and loudness change of pure tone components assuming EV acceleration

☆柴橋和真（中央大院），△金堂雅彦，戸井武司（中央大）

◆EV 加速時においてモータやギア次数の純音成分は，パワーユニットおよびモータ単体の共振通過時に音圧変化が強調される。

◆共振通過純音に対する不快感の印象評価を行い，感覚上の周波数掃引速度を意味する臨界帯域通過速度および感覚上の音圧変化量を意味するラウドネス変化量により，Fig. 1 に示すように，本条件における共振通過純音は約 0.70 Bark/s で最も感度の高い許容閾値となる逆放物線的特性を有することを見出した。

◆約 0.70 Bark/s で最も感度の高い許容閾値を下回る領域は不快感に有意な影響を与えない弱い不快感となるので，モータやギア次数による共振通過純音の音量設計指針として活用できる。

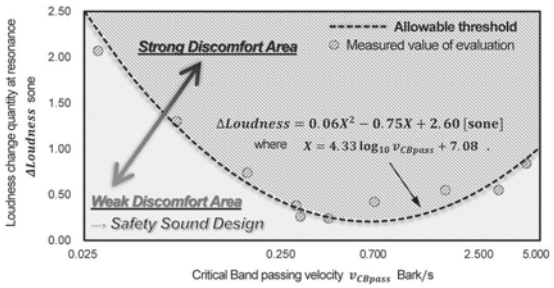


Fig. 1　Auditory characteristics for pure tone components at resonance passing

1-9-15

1-9-15　視聴覚刺激および運転動作を考慮した EV 加減速時の純音成分のマスキング特性

Masking characteristics of pure tone components during acceleration/deceleration of EV considering audio-visual stimuli and driving operation

☆柴橋和真（中央大院），△金堂雅彦，戸井武司（中央大）

◆ 車室内の運転者は，Fig. 1 に示すような聴覚+視覚+運転動作による複合感覚刺激下にある。そこで EV 加減速走行時において，視聴覚刺激および運転動作を考慮した暗騒音による純音成分のマスキング特性を解明する。

◆ マスキング評価において変化させるパラメータは，[1]:車両加減速度合いによる周波数掃引速度，[2]:車両の加減速による周波数掃引方向，[3]:提示する感覚刺激の 3 因子である。

◆ Fig. 2 は走行条件と可聴突出量 L_P を示し，感覚刺激に対する実験参加者のマスキング評価平均値と標準偏差である。本実験条件における視覚は体感音圧を 2.7 dB 低減させ，運転動作は体感音圧を 1.5 dB 低減させることを見出した。

◆ 手動運転車において，純音成分の体感音圧を低減する聴覚以外の感覚刺激によるマスキングの有効活用が考えられる。一方，報知音や快適なサウンドデザインを目的とする場合は，本研究で得られた知見に基づき複合感覚刺激下での適切な体感音圧設計が重要である。

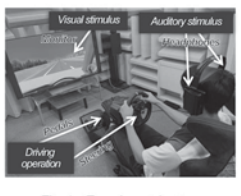


Fig. 1　Experimental setup of masking evaluation

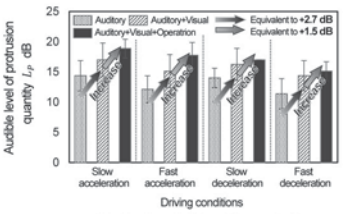


Fig. 2　Results of masking evaluation

1-9-16

1-9-16　呼吸変動音の同調による覚醒を考慮した在宅医療機器の快音化

Comfortable sound design for home medical equipment in consideration of arousal due to synchronization of respiratory fluctuation sound

☆大林葵（中央大院），△島峰徹也（東京工科大），戸井武司（中央大）

◆呼吸変動音における時間的変化，および周波数的変化に着目し，それらが非使用者の覚醒に及ぼす影響を把握する。

◆CPAP 着用時は，空気を送り込む際に気流音が発生する影響で 4000 Hz 以上の高周波数域を多く含み，呼気音と吸気音の音圧差は特に 1000 Hz 付近の低周波数域にて大きいという特徴を有する。

◆呼吸変動音の時間的変化に関して，呼吸数増加時において主観的なリラックス感の低下，および呼吸同調により呼吸数が増加することに伴い，Fig. 1 に示す覚醒度（AAC：α 波減衰係数）が上昇する傾向がみられた。また，Fig. 2 に示す呼吸変動音のシャープネス値と評定尺度法による主観評価結果の関係より，覚醒に影響するシャープネス値は 3.2 acum を境に評価傾向が異なる。中心周波数 8000 Hz の高周波数域，および呼気音と吸気音の音圧差が大きい中心周波数 1000 Hz を組み合わせて音圧低下することで眠りやすさへの音質改善につながる可能性を示唆した。

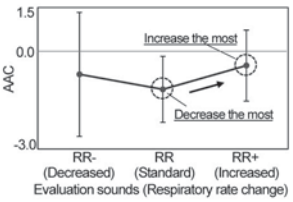


Fig. 1　Results of AAC

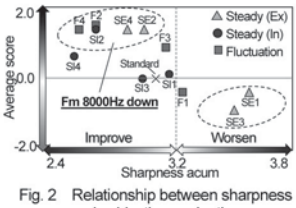


Fig. 2　Relationship between sharpness and subjective evaluation

1-9-17

1-9-17　体内時計を考慮した
高年層の日常イベントに対する生活機能音の印象評価

Impression evaluation of life function sound cues for daily events in elderly considering body clock

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

◆本論文では，体内時計を考慮して制作された８種類の生活機能音①起床，②朝食，③生活行動，④昼食，⑤身体活動，⑥夕食，⑦風呂，⑧就寝を用いた印象評価実験より，高年層が朝，昼，夜に応じた音の感じ方に影響を及ぼす要因を把握する。

◆２つの印象評価として，SD 法では 15 種の形容詞対で７段階絶対評価を行い，時間帯評価では「どの時間帯に感じたか」を朝／昼／夜の３択，「なぜそう感じたのか」を音の拍／テンポ／大きさ／数／調子の５つについて回答用紙に回答した。

◆高年層の時間帯評価を Fig. 1 に，STP(Segmentation, Targeting, Positioning)分析により抽出した時間帯評価に影響を及ぼす形容詞対の高年層の平均評価得点を Fig. 2 に示す。図より得点が４点以上に分布すると夜に，４点以下だと朝および昼に感じる傾向が見られる。しかし⑥において，若年層との傾向の違いが認められたため，さらに時間帯評価に影響を及ぼす要因を考察した。

◆重回帰分析による時間帯評価を考慮した機能音モデルを構築し，評価に影響を及ぼす音パラメータを明らかにした。目的変数を朝および昼の選択率とすると，機能音のリズムパターンを構成する要素に加え，高年層ではシャープネスの推移が影響を及ぼす要因であることがわかった。

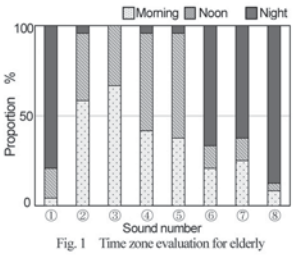


Fig. 1　Time zone evaluation for elderly

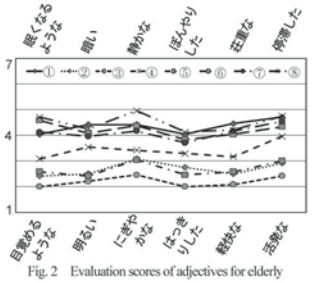


Fig. 2　Evaluation scores of adjectives for elderly

1-9-18

1-9-18　中空菓子の喫食音による快感情の創生

Creation of pleasant feelings by the sound of eating hollow snacks

☆菅田康平(中央大)，△河本政人，和食麻衣(森永製菓)，
戸井武司(中央大)

◆菓子の構造と喫食音に対する印象評価の関係性を検討した。

◆中実菓子と中空菓子では，喫食音の１噛み目の音圧の経時変化が異なる．中実菓子は１つのまとまり，中空菓子は２つのまとまりをなすことが明らかになった。

◆ヘッドホン提示での印象評価を行うため，測定した気導音の低周波成分を聴覚的印象に合うように補い，体感する喫食音を作成した。

◆因子分析の結果から，「覚醒・硬質」「快適」「繊細」の因子が得られ，ラッセルの円環モデルにプロットすることにより，菓子の喫食音から想起される感情を推定した。

◆中実菓子の喫食音は驚きや興奮をより大きく想起させ，中空菓子の喫食音は幸福や喜びをより大きく想起させることを示した。

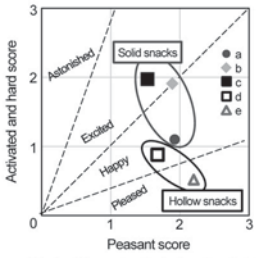


Fig. 1　Pleasant-activated and hard plots

1-9-19

1-9-19 防災放送訓練システムの効果に関する実証的研究

Empirical research on the effectiveness of disaster prevention broadcast training system

☆辻さと恵，土田義郎，高野佐代子(金沢工大)

◆筆者らの開発した防災放送のアナウンス訓練システムによる訓練効果検証を目的とした実地訓練と評価実験を行った。

◆実地訓練は地方自治体と連携し、職員が訓練に参加した。また、その訓練音声の評価実験では、住民及び訓練に参加していない職員を被験者とした。訓練する側の意識調査のため、訓練前にアンケート調査と訓練後にヒアリングを行った。

◆訓練前後の音声のどちらが避難の必要性を感じたかを２択で、聞き取りやすさと緊迫感を７段階評価で計３項目を評価してもらった。

◆評価の結果、訓練システムによる訓練後の音声の方が、逃げる気になるアナウンスになっていることが明らかになった。また、訓練によって聞き取りやすく、緊迫感のあるアナウンスができるようになることが明らかになった。

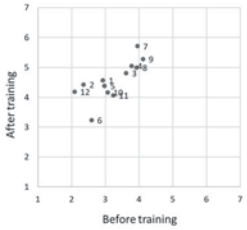


Fig.1 Ease of hearing

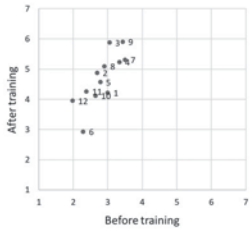


Fig.2 Sense of urgency

1-9-20

1-9-20　**快音性デザインのための感性工学的手法に基づくモータ騒音の価値構造モデル**

Hierarchical Structure Model of Motor Drive Noise based on Affective Engineering for Pleasant Sound Design

○山﨑陽一，長田典子(関西学院大・工)，
保月孝志，樋渡天次郎，佐竹彰(三菱電機・先端技術総合研)

◆本研究では，インバータ制御により生じるモータ駆動音の快音性デザインを目的として，感性工学的手法を用いて快音性の価値構造モデルを構築した．

◆構築したモデルは，価値・印象・物理層の３層からなる感性の階層構造モデルに基づき構成されており，音響特徴から快音性形成までの認知的過程に関する理解を提供し快音性デザインに資するものである．

◆さらに予測精度の観点から Aures により提案された感覚的快さのモデルよりも良い精度での快音性予測が可能となった．

➢ Aures モデル (調整済み R^2: 0.78), 構築モデル (調整済み R^2: 0.88)

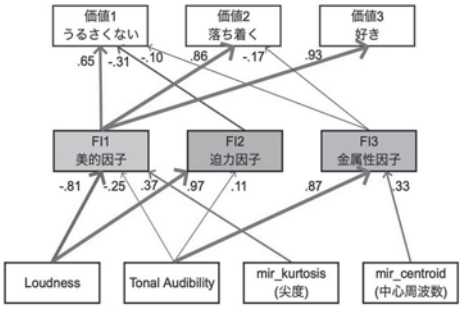


Fig.1 : Hierarchical structural model for the pleasantness of motor drive sound

1-9-21

講演取消

1-10-1

1-10-1 コンクリート非破壊計測のための非接触音響探査法に関する研究 － 移動計測時の欠陥検出性能の改善に関する検討 －

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

◆現在、我々は構造物の内部に生じる剥離や亀裂欠陥の探査を音波照射加振とレーザドップラ振動計(LDV)を用いた方法について研究を行っている。

◆音源と LDV を台車に載せ移動計測を行い、測定されたデータから欠陥位置の検出および信号処理による性能改善について検討を行った。

◆今回の実験に使用した供試体は25 mm 厚の発泡スチロールが埋設されたコンクリート供試体である。

◆相互相関処理に各種ノイズ除去処理(傾き補正、3500 Hz 以上の高周波数カット、ランダム性の高いデータの除去)を追加した結果、低速時であれば直径300 mm、深さ40、60、80、100 mm の欠陥の応答を確認する事が出来た。

(a)　(b)



Fig.1 (a) 計測領域、(b)相関エネルギー映像化例

1-10-2

1-10-2 Visualization of sound fields of cymbals

☆ Tin Ng, Haruka Nozawa, Yasuhiro Oikawa (Waseda University)

◆Research Purpose:

I. To investigate and visualize the sound fields generated by cymbals and thus understand the principles of musical instruments.

II. To investigate the relationship between sound fields and commercial music preferences.

◆Technique and technologies used:

I. Parallel Phase-Shifting Interferometry (PPSI)

II. MATLAB

◆Tools used:

I. Hi-hat

II. Microphone

◆ General Results:

It is believed that different striking methods on cymbals can bring various sound fields, thus it is essential for further researches.

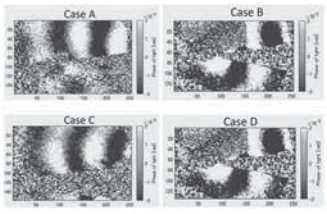


Fig.1 : Visualization of sound fields produced by hi-hat under various hitting conditions

1-10-3

1-10-3 振幅統計量コンパウンドにおけるステアリング角度と統計的ゆらぎの関係

The relationship between steering angles and statistical fluctuation in echo-envelope statistics compound

☆奈良千尋，吉田 憲司，山口 匡，平田 慎之介（千葉大）

◆超音波ビームのステアリングを行うことで，同一部位を描画した異なる複数の超音波画像を作成し，その振幅統計量をコンパウンドすることで，統計的なゆらぎを低減できる．適切なステアリング角度，間隔，コンパウンド枚数について検討を行った．

◆リニアアレイプローブを用いて±15°の範囲を 1°間隔でステアリングを行い，超音波画像とその３次モーメントマップを 31 枚作成した．

◆ステアリング角度の間隔を変えながら，３枚の３次モーメントマップをコンパウンドした（Fig. 1）．角度差が大きくなるほどモーメントの四分位範囲は小さくなったが，±10°以降はあまり変化しなかった．

◆コンパウンド枚数を変えた場合，５枚以下では四分位範囲が大きくなったが，5°間隔７枚以上ではあまり変化しなかった（Fig. 2）．

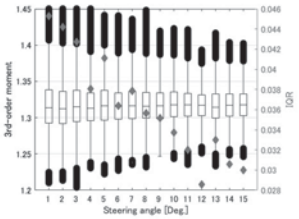


Fig.1 : Box-and-whisker plot with IQR plot of compounded 3rd-order-moments of 0° and ±α°.

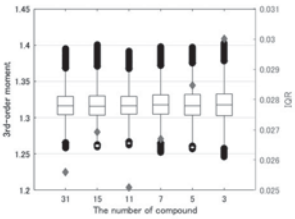


Fig.2 : Box-and-whisker plot with IQR plot of compounded 3rd-order-moments of multi angles.

1-10-4

1-10-4 Examination of quantitative evaluation method of lung ultrasound for pneumothorax detection

○森　翔平（東北大），△Khoa Tran（Carleton Univ.），
荒川元孝，金井　浩（東北大），△小野　雄（Carleton Univ.）

◆気胸の検出モニタリング法の開発を目的に，M モード信号に対するエコー包絡振幅統計解析により気胸の有無を判別する手法を検討した。

◆健常肺領域の M モード信号は時間（フレーム）方向にスペックルのような模様を呈す一方で，気胸肺領域では定常的な信号となることに着目し，スペックルを表す代表的なモデルであるレイリー分布からの逸脱度を Kullback-Leibler (KL) 情報量により評価した。

◆気胸肺領域では，胸膜（pleura）より浅部の胸壁部と同様の信号が観測されると仮定し，健常肺の胸壁部を気胸肺領域として模擬し，胸膜より深部の健常肺領域と比較した。

◆胸壁部では健常肺領域よりも KL 情報量が大きくなり，レイリー分布からの逸脱度に基づき気胸の有無を判別できる可能性が示唆された。

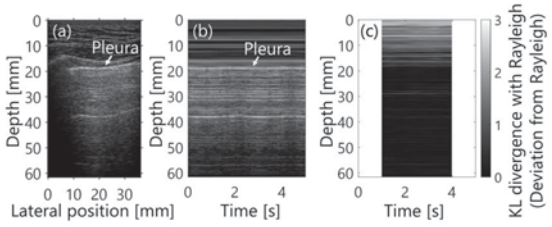


Fig.1 : (a) B-mode and (b) M-mode images of chest region.
(c) KL divergence with Rayleigh evaluated for M-mode signals.

1-10-5

1-10-5　超音波による血圧計測のための内圧・プローブ押圧と管断面形状に関するファントム実験系での検討

Phantom experiment to investigate relationship among internal pressure, pushing pressure by probe, and tube cross-sectional shape for ultrasound blood pressure measurement

☆石川諒[1]，森翔平[1]，荒川元孝[2,1]，△大庭茂男[1]，小林和人[3]，金井浩[1,2]
([1]東北大院・工)([2]東北大院・医工)([3]本多電子・研究開発部)

本研究グループでは，プローブ圧電素子の圧電効果により，超音波計測と同時に血圧を計測できる超音波プローブを開発した．しかし，プローブで計測される圧力は，プローブの押圧依存性があり，血圧を正確に計測できていない．本報告では，正確な血圧計測を目指し，ファントム実験系において，管内圧，プローブ押圧と，管断面の断面積および偏平率の関係について検討した．

血圧 $P_{\mathrm{in}}(t)$ の計測のために，皮膚に押しあてた圧力センサで計測された圧力 $P_{\mathrm{sensor}}(t)$ は血管断面形状に依存する．そこで，ファントム実験において，管断面の断面積 $S(t)$，偏平率 $\rho(t)$ と $P_{\mathrm{in}}(t)-P_{\mathrm{sensor}}(t)$ との間の関係を求めた．その結果，Fig.1 に示すように，$S(t)$ と $P_{\mathrm{in}}(t)-P_{\mathrm{sensor}}(t)$ の間には正の相関があり，$\rho(t)$ と $P_{\mathrm{in}}(t)-P_{\mathrm{sensor}}(t)$ の間には負の相関があった．これらの結果は，$S(t)$，$\rho(t)$，$P_{\mathrm{sensor}}(t)$ の計測によって $P_{\mathrm{in}}(t)$ を推定できる可能性を示す．

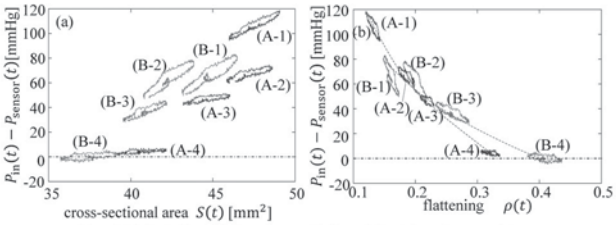


Fig. 1 (a) Relationship between $S(t)$ and $P_{\mathrm{in}}(t)-P_{\mathrm{sensor}}(t)$,
(b) relationship between $\rho(t)$ and $P_{\mathrm{in}}(t)-P_{\mathrm{sensor}}(t)$.
(A) Maximum internal pressure: 125 mmHg
and minimum internal pressure: 90 mmHg,
(B) maximum internal pressure: 85 mmHg
and minimum internal pressure: 55 mmHg.
$P_{\mathrm{surface}}(t)$: (A-1) 15 mmHg, (A-2) 38 mmHg, (A-3) 51 mmHg, and (A-4) 95 mmHg,
(B-1) 7 mmHg, (B-2) 13 mmHg, (B-3) 40 mmHg, and (B-4) 58 mmHg.

1-10-6

1-10-6　サラブレッド脚部の皮質骨における超音波音速・減衰の不均一性

Heterogeneity of ultrasonic wave velocity and attenuation in cortical bones of Thoroughbreds

☆児玉周大[1]，津幡大聖[1]，
△三田宇宙[2]，△田村周久[2]，△千葉恒[3]，松川真美[1]
（1．同志社大，　2．JRA 総研，3．長崎大）

◆競走馬への定量的超音波法（QUS）の適用を目指し，基礎研究としてサラブレッド脚部皮質骨の超音波物性評価を行った。

◆第三中足骨の近位部，骨幹部，遠位部を骨軸方向から円環状に切り出し，骨軸方向に超音波を伝搬させ**音速と減衰**を測定した。

◆Fig.1 に音速分布を示す。全ての試料において**外側，内側の音速が高く**，後方部で低めであった。遠位部，近位部では前方部の音速が低かった。部位による**多孔性の違いやトレーニングによる運動力学的な影響**が考えられる。

◆臨床的には**外側，内側での疾病が少ない**。これは組織の音速が速く硬いことと関連があると考える。

◆前方部では音速と減衰に負の相関(R = -0.5)がみられた。**前方部の減衰は骨マトリクスの物性を**，**骨孔が多い後方部では散乱**の影響を反映している可能性がある。

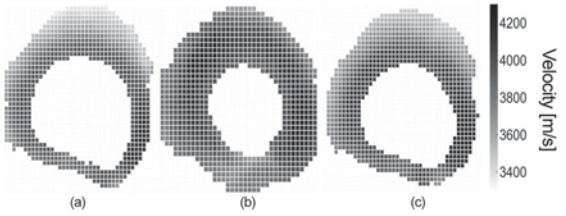


Fig.1 : Wave velocity (4 MHz), (a) proximal, (b) diaphysis and (c) distal.

1-10-7

1-10-7 チャープ信号を用いた高調波パルス圧縮の非破壊計測への適用の基礎検討

Basic study on the application of harmonic pulse compression to non-destructive measurement using chirp signals

◎清水鏡介（日大院）大隅歩，伊藤洋一（日大理工）

◆空中超音波フェーズドアレイを用いた弾性波源走査法による非破壊検査について研究を行っている．

◆本報告では，提案手法である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を非破壊検査に適用し，その有効性について検証した．

◆Figure 1 に計測試料の外観を，Fig. 2 にパルス圧縮を適用した場合の時間波形を示す．

◆その結果，提案手法の適用により急峻な立ち上がりを持つガイド波波形を取得することができ，提案手法の有効性を確認した．

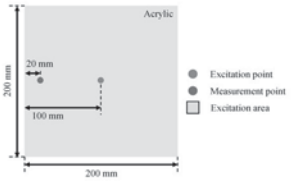


Figure 1. Measurement sample.

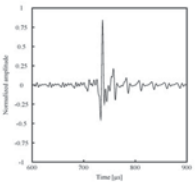


Figure 2. Measurement results of the proposed method.

1-10-8

1-10-8　凹凸のある路面上の障害物の方位検出精度の検討

Accuracy of detecting the position of obstacles on rough road surfaces

○蜂屋 弘之（東工大）

◆不規則な凹凸路面上にある障害物の位置推定精度について検討
◆障害物からの反射波を二つの受波器で受波 (Fig.1)
◆粗面からの反射波により，受信信号振幅はライス分布で近似できる (Fig. 2 Complex plot)
◆二つの受波器の信号振幅は変動しているが，位相差・振幅分布は安定 (Fig. 2 Phase difference)
◆粗面からの反射波の影響による振幅変動が方位推定誤差の原因
◆粗面からの反射波振幅を参照すれば，位置精度を推定でき，信頼性のある測定が可能となる

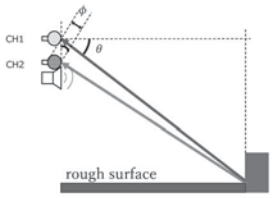


Fig.1 : Configuration of transducers and obstacle

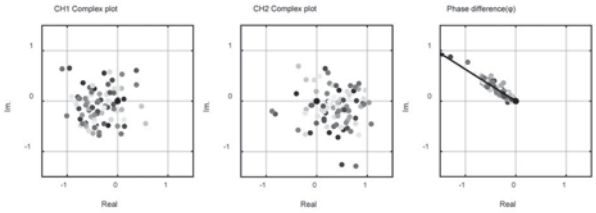


Fig.2 : Signal plot (ch1, ch2) on complex plane and difference of complex vector

1-10-9

1-10-9　三次元押し出しモデルにおけるレベルセット法最適化 —超音波振動工具のトポロジー最適化(5)—

Level-set method optimization in three-dimensional extrusion model —Topology optimization of large ultrasonic tools (5)—

○和田有司，中村健太郎（東工大・未来研）

◆超音波接合等を目的とする振動工具の作用面を平坦に振動させることは産業応用上重要であるが、経験に基づくノウハウとなっている。
◆本稿では商用有限要素法プログラムを用いて最適化を進行する手法の検討を行った。二次元レベルセット関数の進展を熱伝達ソルバーで代替し、抽出された形状を押し出して三次元モデルモデルを形成する。
◆Fig. 1 の 50 kHz で駆動される設計領域に対して、あらかじめ初期の円形穴が発生するようなレベルセット関数を与え最適化を開始する。
◆Fig. 2(a)が最適形状であり、(b)のような平坦な振動モードを得ている。

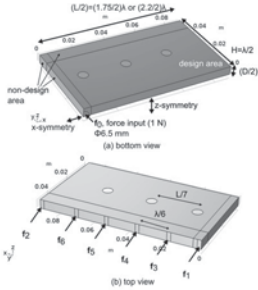


Fig.1 : Target model design area and boundary conditions with three holes by initial level-set function.

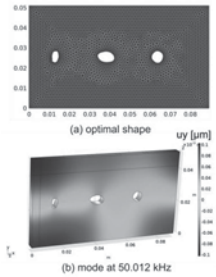


Fig.2 : (a) Optimal shape, (b) vertical displacement distribution of eigenmode at 50.012 kHz in the final optimization step.

1-10-10

1-10-10 空中強力超音波源による定在波音場内に平板を入れた場合の音圧分布

Sound pressure distribution in the case of the placed flat plate in a standing wave sound field by aerial intense ultrasonic sources

☆大和田知花，淺見拓哉，三浦 光（日大・理工）

◆筆者らは，空中強力超音波源を用いた濡れた布の乾燥について検討を行っている。本稿では試料を音圧の腹または節の位置に入れた時の音圧分布について検討を行った。
◆２台の空中超音波源の振動板面が平行かつ向かい合うように設置し，板間に強力な定在波音場を形成した。定在波音場内の音圧の腹または節の位置に濡れた布の代わりとして、板状試料のポリプロピレン(大きさ 44×84×0.5 mm)を用い，その設置位置による音圧分布を解析と測定で比較し，検討する。
◆Fig. 1 は板状試料を音圧の腹位置に振動板と平行に入れた場合の有限要素法による音圧分布の解析結果である。試料を音圧の腹の位置に設置しても定在波音場はほとんど影響を受けないことがわかる。

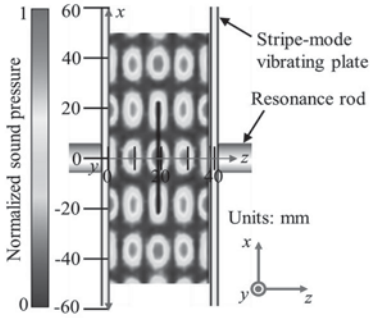


Fig. 1. Analysis results of sound pressure distribution in the case of the placed sample plate in the antinode position of the sound pressure.

1-10-11

1-10-11　PVDC フィルムを透過した空中超音波音場の評価

Evaluation of airborne ultrasound sound field through PVDC film

☆宮坂優佑（日大院・理工），大隅歩，伊藤洋一（日大・理工）

◆音波の位相を制御する方法として，放射方向変換器や電気的な制御，音響メタマテリアルなどが利用されている．これらの手法では，音源の大型化や，複雑な構造の設計が必要になるといった問題が生じる．
◆空中に放射された音波の位相を制御する方法として，ポリ塩化ビニリデン(PVDC)フィルムの利用が提案されている．
◆本報告では，PVDC フィルムを用いた位相制御の基礎検討として，PVDC フィルムを透過した空中超音波の音圧に着目して検証を行った．
◆Fig. 1 に PVDC フィルムの設置距離に対する透過音波の音圧のグラフを示す．グラフには，比較として自由空間での音圧も示している．
◆結果より，設置距離 0.10 ~ 1.10 mm の範囲では，自由空間の音圧よりも高い値を示すことを確認した．

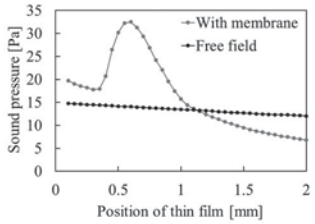


Fig. 1 : Relationship between sound pressure and position of thin film.

1-10-12

1-10-12 集束超音波が物体に与える音響放射力の平面内における分布

Acoustic radiation force profile of the focused ultrasound for the object within a plane

☆頃安祐輔(筑波大院), 星貴之(PxDT), △落合陽一(筑波大),伏見龍樹(筑波大)

光ピンセットでは物体が焦点に捕捉されるが、空気中の音響ピンセットでは、多くの物体は焦点(高圧領域)から押し出されると考えられている。よって、定在波や単一ビームによるツイン/渦トラップなどが用いられてきた。本研究では、平面内での操作においては光ピンセットと同様に、集束ビームによって物体を操作できることを示す。この場合、平面内での音響放射力は図 1 のようになり、焦点が均衡点となる場合(図 1A)と、焦点から同心円状の位置が均衡点になる場合(図 1B)に分けられる。この場合分けについて、音源から物体の距離、音源から焦点の距離、物体の大きさの 3 変数に注目して分析した。その結果、図 1A のような焦点が均衡点となる音場を得るには、音源から物体の距離は 100 mm 以上で、物体の大きさが 2~3 mm 以下であることが必要であることが分かった。

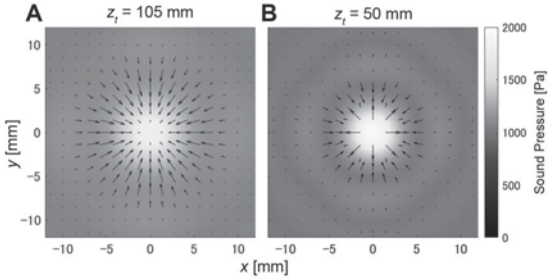


Fig. 1 Sound pressure distribution and acoustic radiation force (arrow) when the the target radius is 1 mm and the distance (z_t) between phased array transducers and target object is (A) 105 mm and (B) 50 mm.

1-10-13

1-10-13 小型音源を用いた 3 次元超音波マニピュレーション

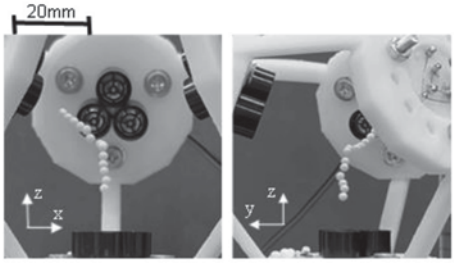
Three-Dimensional Acoustic Manipulation Using Small Sound Sources

☆橋塲望(愛工大), 畑中信一(宇都宮大), 鎌倉友男(電通大), 豊田昌弘(本多電子), 小塚晃透(愛工大)

◆超音波振動子３個配置した超音波ユニット４台を正三角錐の各頂点に置き、音軸を三角錐の中心で交差させることで３次元状に広がる定在波音場を生成する。

◆生成した音場中で粒子を捕捉させる。そして周波数にわずかに差を与えて、位相を連続的に変化させたときの粒子の移動の様子を観察し、評価を行う。

◆Fig.1 に、粒子が音軸方向に移動した様子を表した写真を示す。(a)が正面からの様子、(b)が横からの様子である。粒子が上下、奥行方向に移動する様子が分かる。



(a) Front　　(b) Side

Fig.1 Three-Dimensional Acoustic Manipulation

1-10-14

1-10-14 圧電素子の「跳躍・降下現象」の機構解明（ソフト系セラミックC-6について）

Elucidation of the mechanism of the "jumping and dropping" phenomenon of a piezoelectric element (for soft-type ceramic C-6)

○足立和成(山形大・大学院), 山吉康弘, 岸銀之条, 宮川侑也(山形大・工)

既に足立らは、ハード系圧電セラミック C-21(富士セラミックス)について、圧電素子の「跳躍・降下現象」の主たる原因が、その機械共振周波数付近で生じる内部電界の著しい乱れによる電界集中が誘起する局所的な分極の変化にあることを示唆する結果を得ている。今回ソフト系 C-6（同社）について同現象を観察した結果、印加電圧が低いと周波数領域での履歴現象すら現れず、電圧を増加させても顕著な「跳躍・降下現象」は見られなかったが、駆動周波数の掃引速度によって、その履歴現象の現れ方が著しく変化したことから、C-6 では機械共振周波数付近での電界集中による局所的分極方向の変化が、急激な分極反転ではなく、結晶粒界ごとの緩和的な回転運動によって起きている可能性があり、これは、上述の電界集中起因説を支持する新たな事実だと考えられる。

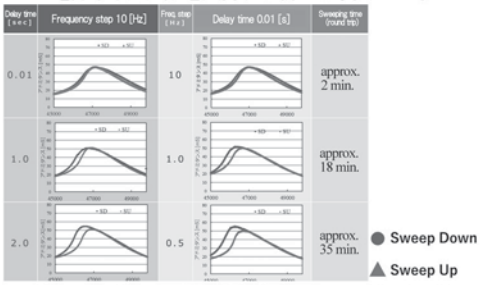


Fig.1 : Frequency-admittance characteristics measured for the 1.5mm-thick piezoelectric disk with the frequency step and the delay time for each step of the sweeping process as the parameters.

1-10-15

1-10-15 Enhancement of underwater acoustic streaming using a cylinder with a cavity

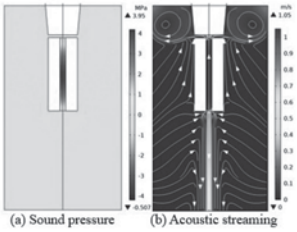
◎Yimeng Wang[1], Kohei Aono[2], Manabu Aoyagi[1] ([1]Muroran Inst. Tech.; [2]SEIDENSHA ELECTRONICS CO., LTD)

◆The effect of a cylinder with cavities of different dimensions on the underwater acoustic streaming was investigated by FEA simulation and PIV experiments.

◆The vibration source consisted of a BLT transducer and a duralumin horn, which worked in longitudinal vibration mode at the resonance frequency of 28 kHz.

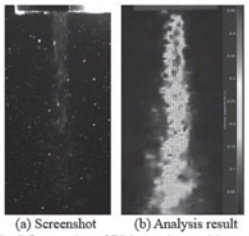
◆The diameter of the cavity (d =1, 2, 3 mm) and the gap between the vibration surface and cylinder (h=1, 2 mm) were the main parameters discussed.

◆Among the results, when d =2 mm and h =1 mm, streaming velocity was the largest. When vibration amplitude A=3.1 μm, simulation and experiment results of sound pressure and acoustic streaming distributions are shown in Fig. 1 and Fig. 2, respectively.



(a) Sound pressure　(b) Acoustic streaming

Fig. 1 Simulation results of sound pressure and acoustic streaming distributions at 28 kHz when d = 2 mm, h = 1 mm, A = 3.1 μm.



(a) Screenshot　(b) Analysis result

Fig. 2 Screenshot of PIV experiment video and analysis result of acoustic streaming at 28 kHz when d = 2 mm, h = 1 mm, A = 3.1 μm.

1-10-16

1-10-16　粘弾性ゲル膜と音響放射力を用いた超音波液体レンズのフォーカシング特性

Focusing characteristics of ultrasound liquid lens using a viscoelastic gel film and acoustic radiation force

☆三木陽斗，原田裕生，中村光佑，小山大介（同志社大）

◆超音波液体レンズは，１枚のレンズで焦点変化が可能であり，従来の超音波ゲルレンズと比較して応答時間を短縮することが可能である．

◆超音波液体レンズのフォーカシング特性を評価するために，レンズの透過光波面を解析した．

◆入力電流の増加に伴い，ゲルの表面形状は空気側に向かって平面に近づくように変形した．その際の焦点距離の変化量は液体レンズの方が11%大きかった．（Fig. 1, Fig. 2）

◆今後は本レンズの光学収差について検討を行う予定である．

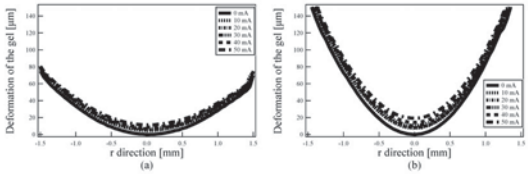


Fig.1 Surface profiles of (a) the ultrasound gel lens and (b) the liquid lens excited with several input currents.

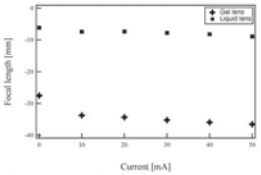


Fig. 2 Focal length vs input current.

1-10-17

1-10-17　液滴の超音波浮揚に及ぼす液物性の影響

Effect of liquid properties on ultrasonic levitation of droplets

☆井上雄大，△二神史也（愛工大），畑中信一（宇都宮大），小塚晃透（愛工大）

◆小型超音波スピーカ 36 個を凹面形状に配置した音源ユニット向かい合わせに２つ配置し，集束する定在波音場を作成する。

◆印加する電圧を変えることで音圧を変化させ，様々な種類の液滴を音圧の節に捕捉させる。電圧を上げていくと浮揚している液滴が落下もしくは弾けるが，液滴の物性によってその電圧が異なるため，実験による検証を行う。

◆以前は液滴の粘度と表面張力に注目したが，今回は液滴の接触角に注目し実験を行った。

◆接触角とは，ある固体の上に液滴を落とした際にできる，液滴のふくらみの程度を数値化したものである。

◆Fig. 1 に実験で使用した定在波音場装置を示す。Fig. 2 に浮揚液滴の印加電圧増加に伴う変形の様子の写真を示す。

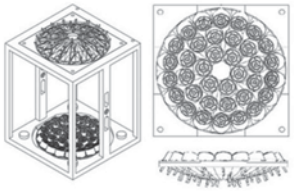


Fig. 1　Ultrasonic speaker unit

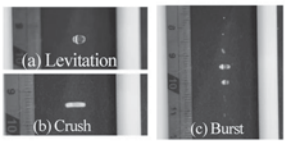


Fig. 2　Deformation of levitated droplet with increasing applied voltage

1-10-18

1-10-18　Small-sized ultrasonic linear motor utilizing traveling flexural waves

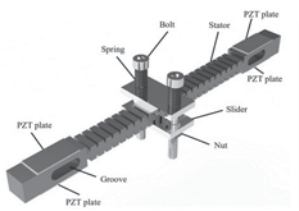
☆ Zhiyi Wen, Yuji Wada, Kentaro Nakamura
(Tokyo Institute of Technology)

◆The device utilizes bimorph transducers to induce traveling waves along a linear bar, constituting the design of a traveling wave linear ultrasonic motor.

◆The proposed methodology enabled miniaturization of linear motor down to 10 mm in diameter. A novel methodology for the miniaturization of the traveling wave linear ultrasonic motor is presented.

◆The ultrasonic motor achieves a long stroke with a maximum speed and thrust of 54 mm/s and 0.83 N, respectively.

◆The traveling wave ultrasonic linear motor achieved a long stroke and compact structure, which has promising applications in in aerospace, biomedicine, and other fields that require a miniature linear actuator.



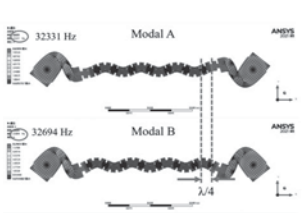


Fig. 1 Structural of the linear traveling wave ultrasonic motor.　Fig. 2 Working Modes and Principles.

1-10-19

1-10-19　Ultrasonic Levitation of Droplets in Horizontally Excited Standing Wave Sound Fields

☆AL JAMII Zahra，和田 有司，中村健太郎 (東工大)

◆ This study investigates horizontal side-by-side ultrasonic levitation with various reflector configurations.

◆ Finite element analysis was conducted for both parallel and angled cases.

◆ The reduction in peak sound pressure was small even when the reflector was rotated by 10 degrees.

◆ Sound pressure distributions were calculated for the parallel and angled cases.

◆ Figs. 1 and 2 show the measured sound pressure distributions for both parallel and angled setups.

◆ Successful droplet levitation was observed in both cases, with differences in pressure distribution and droplet volume.

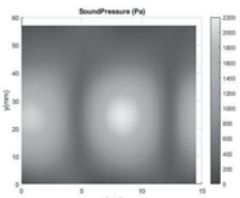


Fig. 1 Experimental sound pressure distribution for the parallel setup.

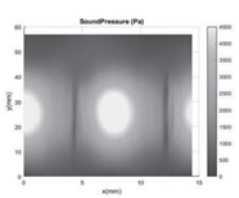


Fig. 2 Experimental sound pressure distribution for the reflector angle of 4.0°.

1-10-20

1-10-20　ダマングレーティングを活用した音響ホログラム生成法の高度化

Advancing Acoustic Hologram Generation with Dammann Grating

◎伏見龍樹(筑波大学)

音響ホログラムは、非接触触覚ディスプレイや音響浮揚などの空中超音波応用の開発に不可欠である。これまで，単一焦点の生成に限り解析解があり、複数焦点の生成には音響ホログラムの最適化アルゴリズムや機械学習モデルが必要であった。筆者の先行研究ではダマングレーティングを用いたアプローチにより複数焦点場の生成が可能であり、焦点位置の回転や平行移動が可能であることが判明している。しかし、ユーザーが望む音響場を瞬時に生成する手法には着想していない。そこで本研究では、生成手法を簡略化し、位相特異と複数焦点の関係性を解明することで算術的に音響ホログラムを生成する新たな手法を検討する。このアプローチは、音響ホログラムへの理解を深め、よりダイナミックな超音波場の構築への道を拓く。

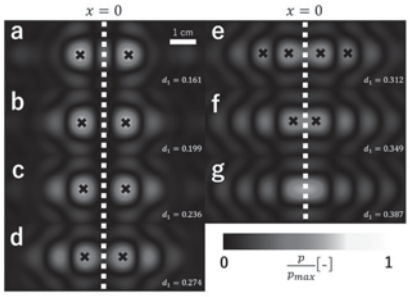


Fig.1: Adjustment of Spatial Relative Position of Focal Points using Damman Gratings

1-P-13

1-P-13　ノイズ音の周波数特性が睡眠の質に及ぼす影響の検討

Effects of frequency characteristic of acoustic noise on sleep quality

☆眞木頌太，大塚　翔，中川誠司(千葉大)

- ◆ カラードノイズを利用した非薬理学的な睡眠促進技術の開発が試みられている．しかしながら，先行報告の結果には一貫性がなく，睡眠の質の向上に関与する要因も未だに不明である．
- ◆ 前報では，ノイズに対する主観的好ましさと睡眠の質の関係を調査した．好ましさによる有意な睡眠の質の向上は観察されなかったものの，特定のノイズによる睡眠潜時の短縮が観察された．
- ◆ 本研究では5 種類のノイズ（ホワイト，ピンク，ブラウン，ブルー，パープル）を用いて，各ノイズのスペクトル特徴が睡眠の質に与える影響を検討した．
- ◆ 各ノイズ呈示条件，および無刺激条件において睡眠脳波および眼電図の計測を実施し，睡眠状態の比較を行った．
- ◆ 無刺激条件に比べ，ノイズ呈示では睡眠潜時が短縮する傾向が確認された．ノイズの高周波成分が強くなるほど睡眠潜時の短縮効果が大きかった．
- ◆ ブラウンおよびブルーノイズ呈示時には，無刺激条件に比して SWS が延長する傾向が確認された (Fig.1)．一方，ノイズの周波数特性と各睡眠段階の時間長との間には一貫した相関は観察されなかった．実験環境や実験時間の短さが結果に影響した可能性がある．

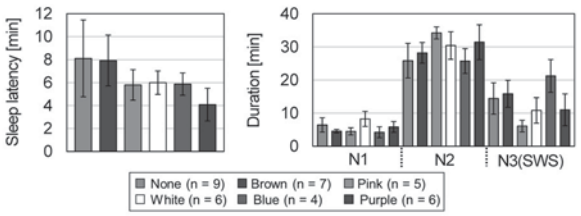


Fig.1 Sleep latency (left) and the duration of each sleep stage (right) in 6 conditions (Mean ± SD).

1-P-14

1-P-14　コード進行生成と聴覚マスキングの統合に基づく歯科治療音の快音化

Pleasant sound design based on integration of chord progression generation and auditory masking for dental treatment sounds

☆林 拓哉，高橋 徹(阪産大)，西浦 敬信(立命館大)，中山 雅人 (阪産大)

- ◆ **我々がこれまで提案してきた歯科治療音の快音化手法では，コード進行により覆い隠されない周波数に存在する騒音成分により不快感低減度が低下する問題があった．そこで，騒音対策に用いられている聴覚マスキングに着目した．最も強いパワーを持つピーク周波数にはコード進行を用いて快音化を行うことで歯科治療音の快音化を行い，その他のピーク周波数に対しては聴覚マスキングを用いて不快感を低減する手法を用いることで歯科治療音の不快感低減を行いつつ，快音化を行う手法を提案する．Fig.1 に提案手法の観測音スペクトルのイメージを示す．Fig.1 の通り，提案手法では，コード音による快音化とマスカー音による不快感低減を統合して行う．**
- ◆ **提案手法の有効性を確認するため，主観評価実験を行った．エアタービンによる歯科治療音を対象とし，快音度と不快感低減度を評価した．Fig.2 に不快感低減度の実験結果を示す．歯科治療音＋マスカー音（CMM：従来手法）、歯科治療音＋コード音（CMC：従来手法）、歯科治療音＋マスカー音＋コード音（PMCM：提案手法）を評価した．実験結果より提案手法は，従来手法と比べて不快感低減性能が向上していることを確認した．**

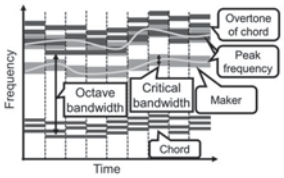


Fig.1 Image of the observed sound in the proposed method

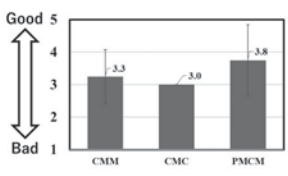


Fig.2 Discomfort reduction performance

1-P-15

1-P-15　開口量が顔面呈示骨伝導音の検出閾に与える影響

Effect of mouth opening on the hearing thresholds of bone-conducted sounds presented to the human face

☆上村 昂，大塚 翔，中川誠司(千葉大)

- ◆ 一般に骨伝導では側頭骨の乳様突起や下顎骨の顆状突起に刺激呈示されるが，近年は鼻骨，頬骨といった顔面への呈示も検討されている．
- ◆ 顔面は複雑な構造を持っており，振動子の呈示位置の変化，下顎や顔面の筋肉の動きに伴って骨伝導メカニズムも変化する可能性が高い．
- ◆ 静的な開口状態において，鼻骨，眼窩下部（上顎骨），頬骨，下顎角（下顎骨），およびオトガイ隆起（下顎骨）に呈示された骨伝導音の検出閾を調べた．また，開口量の増大に伴う検出閾の変化を検証した．
- ◆ 開口量の増大に伴う検出閾の有意な変化は確認されなかった (Fig. 1)．顔面部位における開口の影響は小さいもので，およそ安定した聴取が可能であることが示唆される．
- ◆ もっとも，下顎角およびオトガイ隆起では開口量の増大に伴い検出閾がわずかに上昇した．対して，鼻骨，眼窩下部および頬骨ではほとんど変化しなかった．開口による頭部形状や筋緊張，それに伴う伝搬経路の変化等が影響していると思われる．

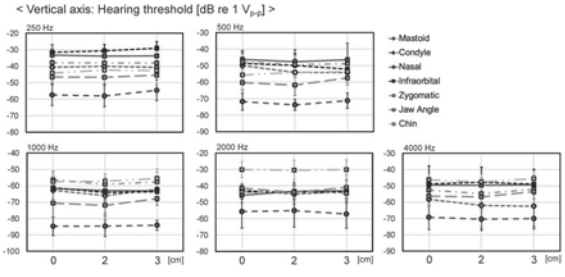


Fig. 1 Hearing thresholds as functions of amount mouth-opening at each frequency

1-P-16

1-P-16　骨伝導における耳栓効果が単音節明瞭度に及ぼす影響:騒音下での検討

Influences of the occlusion effect on monosyllable articulation of bone-conducted speech: Examinations under noises

☆美和 あす華, 大塚 翔, 中川 誠司(千葉大)

◆外耳道を閉塞した状態で骨伝導音を聴取すると，低域のラウドネスが増大する“耳栓効果（OE）”が生じる.

◆著者らは頭部の様々な部位への骨伝導刺激に対する OE を調べてきた．一部の顔面部位（鼻骨（Nasal）や眼窩下部 (Infraorbital)）は，従来部位（乳様突起（Mastoid），顆状突起（Condyle）等）を凌ぐ OE を示すことを報告している.

◆また，前報では，Mastoid, Condyle, Nasal および Infraorbital において，OE が単音節明瞭度を平均約 3％ 上昇させることを示した.

◆本研究では，ブラウンおよびパープルノイズ暴露下において OE が単音節明瞭度に及ぼす影響を評価した.

◆耳栓なし条件（Open）においては，暴露ノイズによる明瞭度低下はブラウンノイズに対して顕著であった．低域に強いパワーを持つブラウンノイズの方が音声をより強くマスキングするためと考えられる.

◆耳栓あり条件（Occluded）においては，静寂下を含む全てのノイズ条件下において，同程度の明瞭度が得られた．本実験では十分なラウドネスを持つ音声を呈示しており，耳栓を透過するノイズの影響をほとんど受けなかったものと思われる.

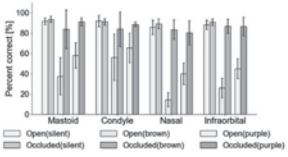


Fig. 1 Percent correct of the articulation test with female voice (Mean ± S. D.).

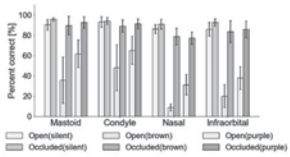


Fig. 2 Percent correct of the articulation test with male voice (Mean ± S. D.).

1-P-17

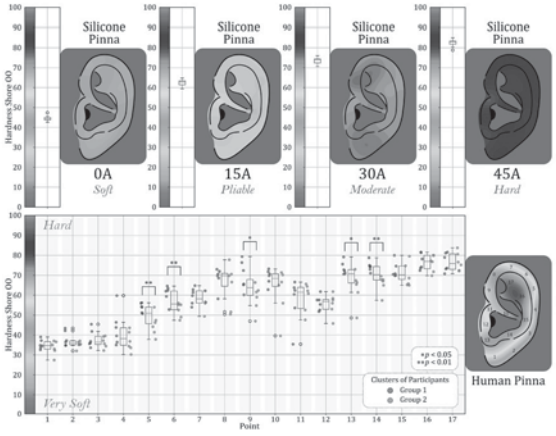
1-P-17　Evaluating Cartilage Conduction: A Comparative Study of Sound Transmission in Artificial and Human Pinnas

◎ IRWANSYAH, Sho OTSUKA, and Seiji NAKAGAWA (Chiba Univ.)

◆This study explores how cartilage-conducted (CC) sound travels in silicone versus human ears, focusing on how pinna hardness and where the vibrator is placed affect hearing.

◆The visual below shows the consistent hardness in silicone ear models versus the complex patterns found in real human ears.

◆Trends in hearing thresholds mirrored those in silicone models, yet the influence of hardness on these thresholds was less pronounced and varied by region.



Silicone pinnas vs. average human pinna: showing uniform vs. varied hardness.

1-P-18

1-P-18　刺激呈示部位の解剖学的構造が遠位呈示骨導超音波の伝搬特性に及ぼす影響：検出閾および頭部振動による検討

Effects of anatomical structure of stimulus placements on transmission of distantly-presented bone-conducted ultrasound: Assessments by hearing threshold and head vibration

☆髙橋尚也, 大塚 翔, 中川誠司（千葉大）

◆骨伝導で呈示された 20 kHz 以上の高周波音（骨導超音波）は，頭部から離れた位置に呈示（遠位呈示）された場合も明瞭に知覚可能であり，胸部や体幹などから振動子に触れた対象者に選択的に音情報を伝達する新型デバイスへの応用が図られている.

◆呈示部位やその近傍の解剖学的構造は，遠位呈示骨導超音波の知覚に大きな影響を与える可能性がある.

◆胸部・背部・上肢に骨導超音波を呈示した時の検出閾推定を行い，前報の外耳道内振動の結果と併せて，呈示部位の解剖学的構造と伝搬特性の関係を検討した.

◆背部は胸部や上肢よりも有意に検出閾が低下した（$p < 0.05$）．背部は単純な構造を有しており，振動子とのカップリングが良好で，筋の接続部が少ないことが伝搬効率の上昇に繋がった可能性が示唆される.

◆受容器までの距離に依存した検出閾変化は概して認められなかった．解剖学的構造の影響が伝搬距離の影響を相殺した可能性がある.

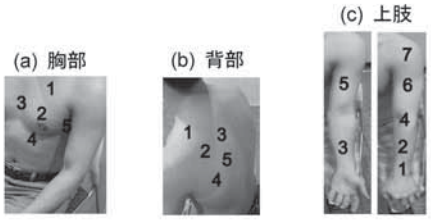


図 1 刺激呈示位置

1-P-19

1-P-19　耳栓装用時の骨伝導音知覚特性に及ぼす暴露騒音特性の影響

Effects of noise exposure on bone-conducted sound perception during earplugging

☆内田知佐, 大塚 翔, 中川誠司

◆ 骨伝導音には，外耳や中耳を介することなく直接内耳に到達する成分が存在するため，耳栓装用時でも容易に知覚される.

◆ 耳栓装用時の骨伝導音知覚特性は，周囲の騒音の時間周波数特性や耳栓の遮蔽特性，耳栓効果の影響を受けて変化するが，これらの各パラメータの影響をシステマティックに調べた例は少ない.

◆ 本研究では，暴露騒音特性が耳栓装用時の骨伝導音の検出閾および単音節明瞭度に及ぼす影響を調べた.

◆ ホワイトノイズでは，全周波数でノイズレベルの増大に伴って検出閾が上昇し，ピンクノイズでは 4 kHz 以上で検出閾の上昇が鈍化した．また，マルチトーカーノイズでは 1 kHz 以下で検出閾の上昇が顕著である一方，2 kHz 以上では検出閾の上昇が観察されなかった.

◆ また，各ノイズ曝露下の単音節明瞭度は マルチトーカー > ピンク > ホワイトとなり，先行研究と矛盾する結果となった．先行研究との言語種の違いが影響した可能性があるが，さらなる検証が求められる.

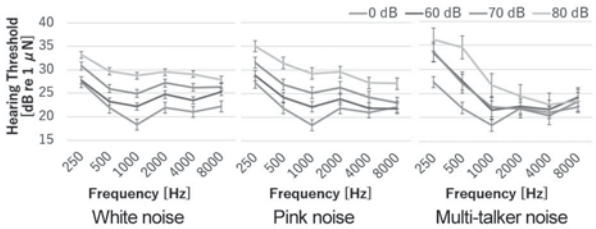


Fig.1 Hearing threshold for each noise.

1-P-20

1-P-20　Spectro-Temporal Modulation 情報を用いた背景音から目的音の検知されやすさの予測の検討

Study on prediction of detectability of target from background sound using spectro-temporal modulation information.

☆堀口遼太郎, 木谷俊介, 鵜木祐史(北陸先端大)

◆背景：聴覚探索実験では，時間方向と周波数方向が共に変動する場合の背景音中の目的音の検知しやすさについて検討されていない.

◆目的：本研究では，時間方向と周波数方向の変動を捉える方法である Spectro-temporal modulation（STM）情報を用いて，背景音中の目的音の検知されやすさが予測可能かどうかを明らかにすること.

◆方法：STM 情報をもとに背景音および目的音を作成し，聴覚探索実験を行った.

◆結果：以下の条件のときに STM 情報（Fig. 1）から目的音の検知されやすさ（Fig. 2）を予測できることが示唆された.

- 目的音と背景音の振幅包絡線が異なるとき
- 目的音と背景音の搬送波の中心周波数が異なるとき
- 目的音と背景音の搬送波が同じで振幅包絡線が異なり，SM 情報のスペクトルの広がりが狭いとき

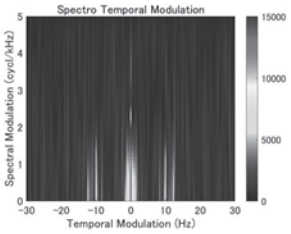


Fig.1 : Example of results of STM analysis of experimental stimuli

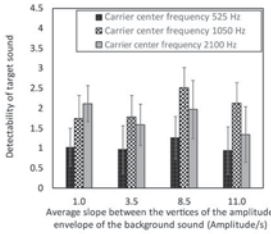


Fig.2 : Example results of auditory search experiment focusing on TM information

1-P-21

1-P-21　競合音存在下での音の出現・消失検出に変化方向が及ぼす影響

Direction dependence of detection rate of sound appearance/disappearance in noisy environments

☆久保捺姫, 坂本修一(東北大通研/院・情科研),
△Chait Maria（Ear Institute/UCL）

◆本研究では，音環境の変化が起こる方向が検出精度に影響を及ぼすかどうかを，水平面上の全方向（30°間隔，12 方向）において検討した.

◆競合音存在下で 1 音が突然出現（Change-Appear），もしくは消失する変化（Change-Disappear）を検出する課題を用い，聴取者は音の出現・消失を知覚した際に素早くボタンを押して反応した.

◆正答率と反応時間を測定し，変化が起こる方向ごとにその結果を分析した.

◆実験の結果，正答率と反応時間ともに，音が出現・消失する方向間に有意な差はなく，変化検出能力は全方位で同等であることを示唆する結果が得られた.

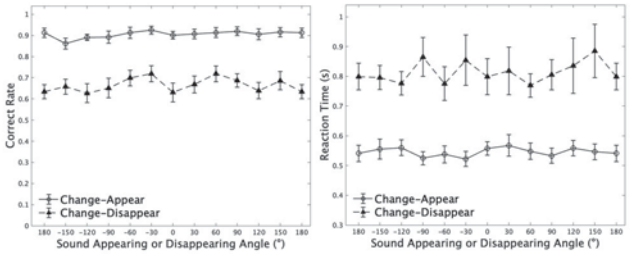


Fig.1 : Correct rate and reaction time of change detection as a function of the direction.
(○ : Change-Appear, ▲ : Change-Disappear)

1-P-22

1-P-22　聴覚刺激のモノラル/ステレオの違いが睡眠の質に及ぼす影響

Effects of monaural and stereo auditory stimuli on sleep quality

☆入交陽斗, 大塚　翔, 中川誠司(千葉大)

◆ 音楽聴取による睡眠の質の向上がいくつかの先行研究で報告されているが，その機序には依然として不明な点が多い.

◆ クラシック音楽を使用して，聴覚刺激のモノラルとステレオの違いが聴覚刺激の睡眠促進効果に及ぼす影響を検討した.

◆ (1) 無音，(2) モノラル呈示，(3) ステレオ呈示の 3 条件で睡眠時脳波の計測を行った．また，覚醒後の被験者に対し，OSA 睡眠調査票簡易 MA 版 (OSA-MA)を用いて睡眠内省の評価を行った.

◆ 楽曲の呈示により総睡眠時間と徐波睡眠時間が延長する傾向が認められたが，モノラルとステレオの違いによる睡眠促進効果の違いは確認できなかった.

◆ OSA-MA の結果は被験者による差異が大きかった．呈示楽曲に対して抱く好ましさの違いが影響している可能性がある.

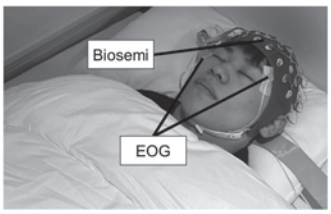


Fig.1 : Situation of experiment

1-P-23

1-P-23　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

A study on the factors that cause for the lateralization in headphone-based VR sound reproduction.

○宇野慎太郎, 工藤彰洋(苫小牧高専), 武居周(宮大工)

◆本研究では、被験者本人の HRTF を用いた VR サウンドにおいて、頭内定位を引き起こす要因を解明する聴覚心理試験を行った。

◆刺激音の条件は、人が音を定位するときの手がかりとして挙げられる ITD、ILD、回折効果の有無に加え、頭内定位の基準として、両耳に同じ信号を提示するダイオティックな信号を用いた。

◆この条件のもと 7 種類の刺激音を生成し、2 種類を対とした試験音を 0°、45°、90°、135°、180° の 5 方向について作成した。

◆定位の評価方法には、Scheffe の方法（浦の変法）を用いた 5 段階の評点を利用し、2 つの刺激音を対比較することで距離感を評価した。

結論：0°、180° では回折効果、45°、90°、135° では回折効果と ILD の有無が頭内定位の要因となる。

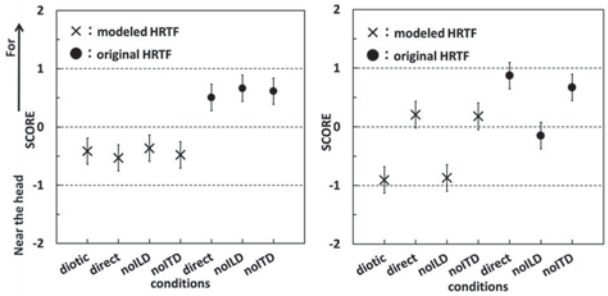


Fig.1 Average score in each condition

1-P-24

1-P-24　距離の異なる二音源の交互聴取による相対距離知覚における一次反射音の影響

Effect of the first reflected sound on relative distance perception by alternately listening to two sound sources from different distances

☆森垣詠太(阪産大院), 小原大季(阪産大),
中山雅人, 高橋徹(阪産大院)

背景　再生音の距離制御は波面合成など物理的再現が注目されてきた
感覚的距離制御によるシステムの小型化[江川+, 2021]
距離感覚制御から遠近感覚制御への転換で更なる小型化に期待

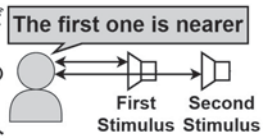


Fig.1:Experimental condition.

目的　遠近感覚の支配的要因の発見
インパルス応答（IR）のピークの役割を調査

手段　実測 IR の音源での遠近判断課題（Fig.2）
実測 IR の第三ピーク以降削除の音源での遠近判断課題（Fig.3）
（Fig.2, 3： 遠近判断正答率，横軸第一音，縦軸第二音）

結果　遠近判断には距離差 1.5m 以上必要（Fig.2）
第三ピーク以降は遠近判断に大きく寄与（Fig.3）
第二ピーク（一次反射音）は遠近判断に寄与しない（予稿参照）

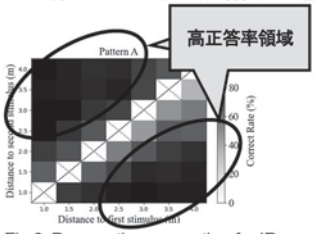


Fig.2:Perspective perception for IR.

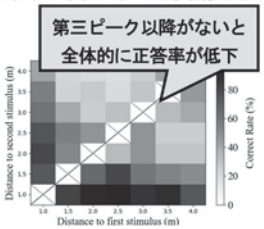


Fig.3:Perspective perception for manipulate IR.

1-P-25

1-P-25　音圧レベルと空間的位置に基づく聴覚オブジェクトの知覚に関する検討

The effect of sound pressure level and sound source position on the perception of auditory objects.

☆堀田尭八, 坂本修一(東北大通研/院情科研)

◆本研究では，音脈分凝における音圧レベルと空間的位置の要因の影響を検討した.

◆実験では，2 種類の 1/3 オクターブバンドノイズ A，B からなる A_B_B_A_B_B 音列を用い，試行ごとに B 音の音圧レベル，及び 2 音源間の角度を変化させて再生した．聴取者は，各音列が 1 つの音脈に知覚されたかどうかを回答した.

◆実験の結果，1 つの音脈として知覚される角度幅の限界（=許容限）は，2 音源の方向（front/side）に関わらず，音圧レベル差が大きくなるにつれて減少した（Fig. 1）.

◆このことから，音圧レベル差の手がかりは，空間的位置に基づく音脈の知覚に影響を与えると考えられる.

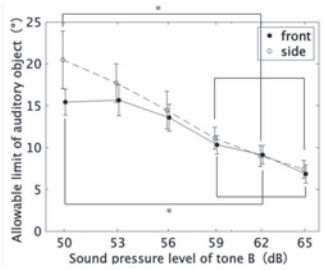


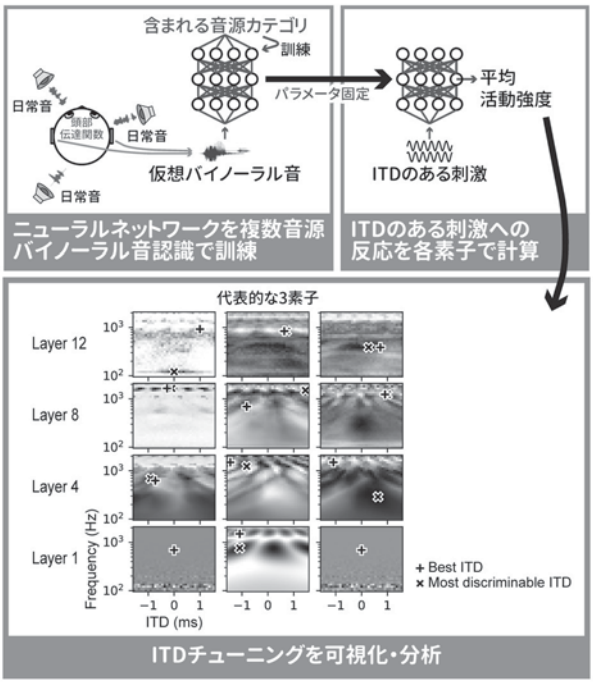
Fig. 1:Allowable limit at each sound pressure level.

1-P-26

1-P-26　複数音源認識で訓練した人工ニューラルネットワークにおける両耳間時間差チューニング

Interaural Time Difference Tuning in an Artificial Neural Network Trained for Multiple Source Sound Recognition

○上村 卓也(NTT CS 研)、寺島 裕貴(NTT CS 研)、
古川 茂人(静岡社会健康医学大院大、静岡県立総合病院、NTT CS 研)



1-P-27

1-P-27　SVM による中耳病変の分類解析

Classification of Middle Ear Pathologies based on SVM

◎周迪(金沢大学), 鳥谷輝樹(山梨大学), 村越道生(金沢大学)

- 目的：これまでに、我々が提案した周波数掃引インピーダンス（SFI）を用いた測定法は中耳病変の診断に対し、ティンパノメトリー（Tymp）より有効性があることを示してきた。将来、医療現場で応用可能な中耳病変の自動検出機能付き診断機器を開発するため、本研究では機械学習の手法（support vector machine: SVM）による耳小骨の離断（Sep）や固着（Fix）の自動検出精度について検討した。
- 方法：本研究は可視化分析及び感度（Sen)、特異度（Spe）と精度（Acc）の平均値（標準偏差）により SFI の中耳病変の自動検出性能を評価した。
- 結果：
  1. 可視化の結果より、提案した SFI が Tymp より耳小骨の離断や固着を判別しやすいことが示された。
  2. 感度、特異度と精度とも、提案した SFI が Tymp より高い性能が得られた。

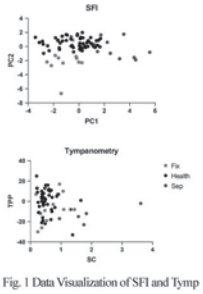


Fig. 1 Data Visualization of SFI and Tymp

Table 1 Detection accuracy of Middle Ear Pathologies

		SFI	Tymp
Sep	Sen	0.82 (0.18)	0.39 (0.28)
	Spe	0.96 (0.10)	0.94 (0.11)
	Acc	0.88 (0.10)	0.67 (0.14)
Fix	Sen	0.82 (0.20)	0.16 (0.20)
	Spe	0.96 (0.10)	0.87 (0.21)
	Acc	0.89 (0.11)	0.51 (0.10)

1-P-28

1-P-28　オリーブ蝸牛束反射と雑音下での聴取成績の関係

Relationship between medial olivocochlear reflex and listening performance in noise, by DOWAKI, Yuri, OTSUKA, Sho and NAKAGAWA, Seiji (Chiba University).

☆堂脇悠里，大塚翔，中川誠司（千葉大）

◆ 脳幹から外有毛細胞へ接続するオリーブ蝸牛束には，音響刺激に反応し，外有毛細胞の増幅を抑制する働きがある（オリーブ蝸牛束反射：MOCR）。動物モデルにおいて，雑音中のターゲットに対する神経応答がMOCRによって増大するという報告があることから，MOCRには雑音下での聴取を改善する（アンチマスキング）効果が考えられている。一方で，アンチマスキング効果のヒトにおける有効性は議論が分かれている。

◆ ヒトを対象とした先行研究においては，MOCRは聴取能力の評価とは別に，課題を行わない状態で評価されてきた。一方で，オリーブ蝸牛束には認知機能を司る皮質領域からの遠心性神経が投射しており，MOCRは聴覚的注意によって増大することが報告されている。この聴覚的注意の影響の可能性を踏まえて，本研究では雑音下での聴取課題実行中のMOCRを測定して聴取成績との関係を調べるべきではないかと考えた。

◆ 先行研究と同様に，聴取課題実行中にMOCRが増大することを確認した（Fig.1）。統計的に有意な相関は確認されなかったものの（Fig.2），聴取課題実行中のMOCR強度の方が，非注意条件と比べて聴取成績との間の相関係数が大きかった。この結果は，能動的なMOCRの制御が雑音下でのアンチマスキング効果に寄与している可能性を支持している。サンプル数を増やし，検出力を高めることで再検討したい。

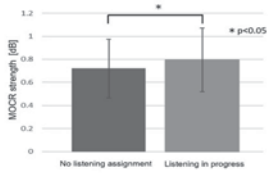


Fig.1：Comparison of MOCR strength without and with detection-in-noise task

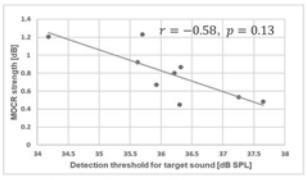


Fig.2：Relationship between listening performance and MOCR strength with detection-in-noise task

1-P-29

記憶保持課題がオリーブ蝸牛束反射に与える影響

Effects of the Memory Retention Task on Medial Olivocochlear Reflex

○高瀬　駿，大塚翔，中川誠司（千葉大）

◆ 脳幹から外有毛細胞へ投射するオリーブ蝸牛束は，音響刺激に反応し，外有毛細胞の増幅を抑制する働きがある．この作用は，オリーブ蝸牛束反射 (MOCR) と呼ばれている．

◆ 視覚的情報の記憶保持課題実行中に MOCR が増強することが報告されている．この結果は，視覚的情報の記憶を保持するのに妨げとなる聴覚からの入力を，MOCR が減少させることを示唆している．日常的なコミュニケーションにおいては，競合する音が存在する中で，聴取した内容を保持しながら会話を進めていく場面が多々ある。それゆえ，聴覚的情報の保持が MOCR に与える影響を明らかにすることは，日常のコミュニケーションにおける MOCR の役割の一端を解明することにつながると期待できる．

◆ 本研究では，イヤホンから呈示されたアルファベットを記憶する課題実行中の MOCR を測定した．記憶対象のアルファベットの文字数を 5, 6, 7 文字として難易度を変化させたところ，7 文字条件でのみ MOCR が有意に減少した (Fig. 1)．聴覚的記憶保持課題の実行中には，記憶保持の阻害となる聴覚入力を抑制する MOCR の作用が有効に働かないことが示唆された．

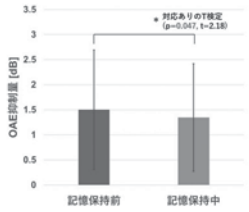


Fig.1：聴覚的記憶保持課題実行に伴う，記憶保持前後のMOCR平均値の変化 (被験者: 5名)．MOCR強度は，音響刺激に伴う耳音響放射（OAE）の抑制量を測定することで評価した．

1-P-30

1-P-30　周波数に対する注意がオリーブ蝸牛束反射に与える影響 － ターゲット周波数の経時的な変動のランダム性との関係

Effect of attention to frequencies on medial olivocochlear reflex – relationship to randomness of target frequency variation over time

☆菊池梢馬，石坂勇毅，大塚翔，中川誠司（千葉大）

◆ 脳幹から外有毛細胞へ接続するオリーブ蝸牛束には，音響刺激に反応し，外有毛細胞の増幅を抑制する働きがある（オリーブ蝸牛束反射：MOCR）。著者らは，雑音中の純音を検出する課題実行中のオリーブ蝸牛束反射を評価してきた。これまでに，特定の周波数に被験者の注意を誘導するために 2 つの注意条件を検討してきた。(1) ターゲット周波数を試行毎に変えないことでターゲット周波数に注意を誘導する条件（静的注意条件），および，(2) 試行ごとにターゲット周波数は変わるが，ターゲットと同一周波数の Cue を提示することで，ターゲット周波数に注意を誘導する条件（動的注意条件）である。

◆ 本研究では，動的注意条件においてターゲット周波数が変動する範囲を調節することで，注意に伴うMOCRの変動が，どの程度の試行毎の周波数変動まで追随できるのかを検証した。

◆ 周波数の変動条件を要因とした一元配置分散分析では有意な主効果は認められなかったが，周波数の変動範囲が最も小さい 1250~1800 Hz の場合に，周波数の変動範囲が比較的に大きい条件（750~3000 Hz，及び，1000~2250 Hz）に比べて，非注意条件に対する MOCR 強度の増加量が大きい傾向が見て取れる。この結果は，試行ごとの周波数の変動が大きくなると MOCR に対する注意の効果が追随できないことを示唆している。今後は，この可能性を，サンプル数を増やして検証していきたい。

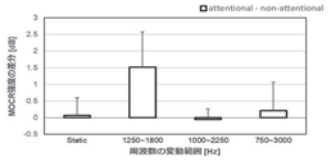


Fig.1：Mean difference of MOCR strength between non-attentional condition and statical/dynamical attentional condition. Error bars depict the standard errors.

1-P-31

1-P-31　鳴管の左右の音源を独立制御可能な鳥類の鳴き声の分析

Analysis of bird song with independent control of left and right sound sources of syrinx

☆中谷優太（農工大），松本和樹（早大），山田宏樹，矢田部浩平（農工大）

- 鳥類の発声機構（鳴管）には左右に 2 つの音源となる部分 (labia) がある
- 左右の labia の振動は互いに非線形な相互作用をもたらすため，2 つの調波音の単純な重ね合わせでは表せない複雑な鳴き声を生み出す
- labia の振動を ODE で記述したモデルが提案されている
- このモデルから得られる信号と実際の鳥の音声を比較し，類似点と相違点について考察した
  - ・類似点：スペクトル構造の急激な変化
  - ・相違点：偶数次倍音の振幅

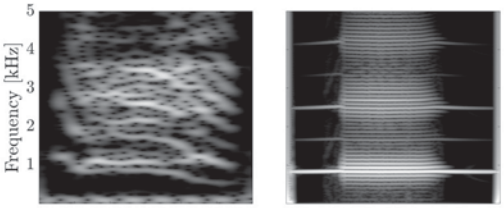


Fig.1: Spectrograms of (a) song of Zebra finch and (b) signal obtained by the model

1-P-32

1-P-32 採餌時に鶏が発する鳴き声の解析と合成に関する検討

Study on analysis and synthesis of sounds from chickens during foraging

☆照沼卓磨，矢田部浩平，新村毅，福田信二（農工大）

◆背景
- ■鶏は母子間で音声コミュニケーションをする
- ■音声を通じた母子間の関わりは雛の行動発達に良い効果をもたらす

◆目的
- ■雛の成長にとって重要な音声コミュニケーションの要素のひとつである母鶏の音声 "Food call" に着目し，その音響的特徴を解析する
- ■解析により得られた知見に基づき，Food call の合成音声を作成する

◆結果
- ■時間周波数解析により，Food call には周波数変調する成分があることが分かった
- ■変分的モード分解により，Food call の主要なモードは振幅変調していることが分かった
- ■上記の特徴を基に合成音声を作成したところ，概ね Food call の再現ができた

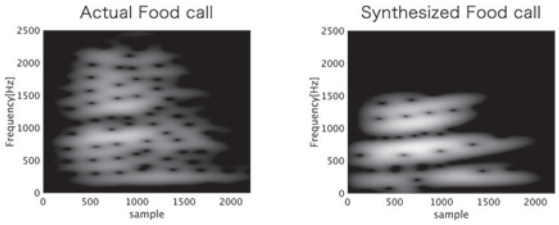


Fig.1 Spectrograms of actual and synthesized Food call

1-P-33

1-P-33 耳介型ホーンを搭載した音響ロガーによる飛行中コウモリのエコーロケーション音声の計測

Measurement of echolocation sounds of flying bats using an acoustical logger equipped with a pinna-shaped horn.

☆妹尾規良，△岡誠也，△宮崎奨功，△中村太宥，加藤めぐみ（同志社大），
長谷一磨（McMaster Univ.），手嶋優風（JAMSTEC），
土屋隆生，飛龍志津子（同志社大）

◆ユビナガコウモリは集団飛行の際，FM 音の終端周波数（Terminal Frequency, TF）を変化させる音響混信回避を行うが，飛行中のコウモリに届く音声の分析は，技術的課題からこれまで十分に行えていない．

◆そこで本報告では，コウモリ耳介付音響ロガーを開発し，ドップラー効果を考慮した FDTD シミュレーションの併用により，実測のみでは判別が難しい飛行中の他個体から自身へ届く音声を分析した．

◆その結果，Bat 1 に届く Bat 2 からの音声の TF は，すれ違いの前後で（Fig.1A），最大 3 kHz（Fig.1B, Bat 2 doppler）変動し，自他の音声間の類似度にドップラー効果が影響する可能性を見出した．

◆本提案手法はコウモリの混信回避行動の機序解明が期待できる．

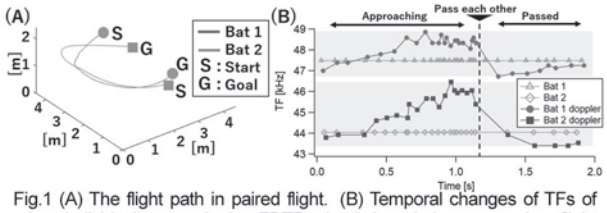


Fig.1 (A) The flight path in paired flight. (B) Temporal changes of TFs of other individual's sounds by FDTD simulation during a passing flight (Simulated sounds are Bat 1, 2 doppler).

1-Q-1

1-Q-1 ホール・ステージ上の音響特性に関する検討 ―富良野市複合庁舎サンエーホールでの実測調査―

A study on stage acoustics in a multi-purpose hall -Measurement of Furano city hall's auditorium-

○永野洋介，上野佳奈子（明治大），井上瑞紀，青木亜美（日建設計）

◆対象ホールでは，ステージ上天井反射板を固定とするために、スリットが設けられたことから，初期反射音の不足，音響特性に偏りが生じる懸念に配慮して，設計が行われた。本報では，ステージ周囲の音響反射板の効果を検証することを目的として実測調査を行い，ステージ上のエネルギ分布について分析・考察を行った。

◆分析結果（Fig. 1）より，各反射板の初期反射音エネルギへの寄与、測定点間の ST_{Early} の差は小さいことが示された。ST_{Late} の結果も含めて，ステージ上の反射音の不足，音場の不均一性は確認されなかった。

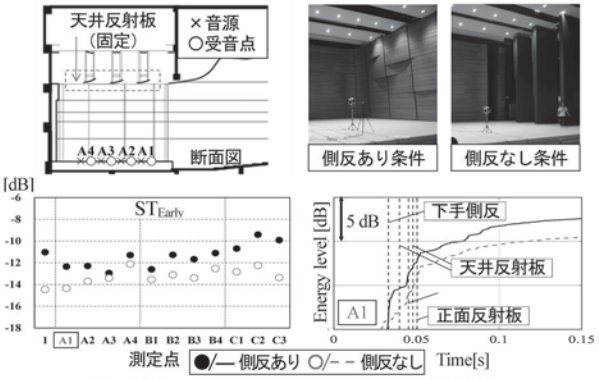


Fig. 1 Measurement points, ST_{Early} and RECC

1-Q-2

1-Q-2 波動音響 VR による会議室の音環境シミュレーション -天井の吸音体レイアウトに関する検討-

Sound environment simulation in a meeting room by wave-based acoustic VR. -Study on ceiling sound absorber layout-

○岡澤一葉，奥園健（神戸大院・工）

◆近年、天井一面を吸音するのではなく、反射面も残すことでラウドネスを確保しつつ音環境を調整する吸音レイアウトの有効性が提案されている。

◆本研究では、Fig1 に示す、等価吸音面積を一定とした様々な天井を持った会議室を対象に波動音響解析を実施し、音響指標の比較を通してその信頼性を検証した。

◆**格子状天井が吸音天井として優れるほか、半面・周縁吸音天井もコストや使用する吸音材、部屋の用途に応じて効果的な選択肢となりえることを示した。**

◆これらの音環境の違いは、当日の発表においてFig2に示すVR コンテンツを用いて体験してもらう予定である。



Fig.2：Overview of VR content

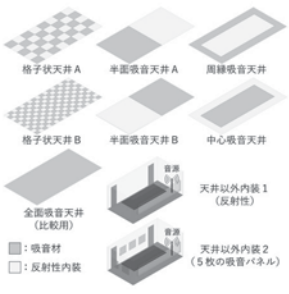


Fig.1： 7 ceiling conditions and 2 interior conditions

1-Q-3

1-Q-3　マイクロスリットを埋め込んだ音響メタサーフェスの吸音メカニズムに関する考察

Sound absorption mechanism of microslit-embedded acoustic metasurface

○奥園 健(神戸大院・工)

◆　簡単な設計手法のもとバンドパスフィルタ状の準完全吸音を実現するマイクロスリットを用いた音響メタサーフェス（MeAMS）吸音体を提案している（図１）。

◆　本発表では、基本的な設計戦略に基づく吸音特性の設計手法について、その妥当性を数値解析と実験によって検証した結果を報告する。

◆　また、MeAMS 吸音体を構成する個々の共鳴器の吸音パワーを評価することで、吸音体の吸音メカニズムを明らかにする。

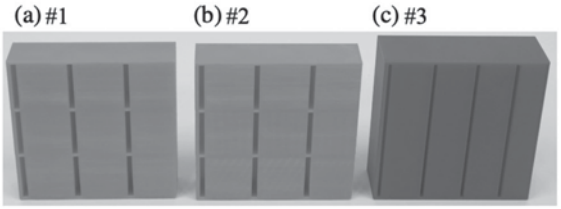


図１：3D プリンタで造形した MeAMS 吸音体の例

1-Q-4

1-Q-4　二本のカーディオイドマイクロホンと音響管による垂直入射吸音率の測定 -複数の音響管による測定結果の比較-

○星 和磨, 羽入 敏樹 (日大・短大)

一辺が 100mm の木製（MDF 板）の音響管と，直径 100mm の金属（アルミニウム）の音響管と，それぞれにカーディオイドマクロホンを挿入して吸音率を測定した．その結果，音響管によらずカーディオイドマクロホンで吸音率を捉えられることがわかった．吸音率が低い部分の測定精度についてはさらなる検討が必要である．

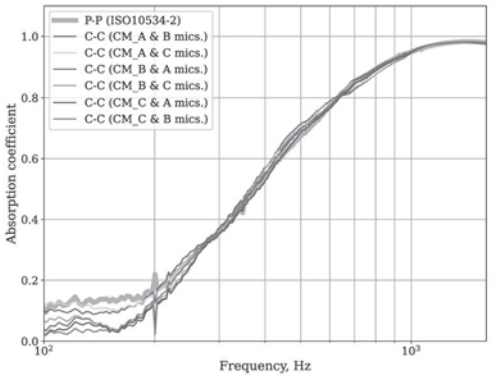


Fig. 1　Measurement results of glasswool which has 50 mm thick and 32 kg/m^3 density with no air backing using three cardioid microphones in the metal tube.

1-Q-5

1-Q-5　音響対策用床埋込式コイルバネ防振装置 開発プロジェクト ー開発の目的と実機評価試験ー

Development project of floor-embedded anti-vibration device for acoustic countermeasures using coil spring -The purpose and evaluation test-

○中来田 道, 藤田 鋭志, △榎本 祐司, △金田 慶太(ゲルブ・ジャパン)

ホールやスタジオにおける防振材として、従来ゴムやポリウレタンが広く用いられてきた。本開発プロジェクトでは、設置に必要な高さをこれらと同等以下としつつ、防振効果を向上させ遮音特性も考慮した床埋込式コイルバネ防振装置を開発した(Fig. 1)。本稿では、開発した装置を対象に振動伝達損失測定による防振性能の評価をおこない、低い固有振動数を実現しつつ金属製バネの弱点であるサージングを抑制できることを確認した(Fig. 2)。また、音響透過損失測定により装置を組み込んだ浮床が十分な遮音性能を持つことを確認し(Fig. 3)、これらの結果からコンサートホールなどの静粛性が求められる音響空間において適用可能であることが示された。

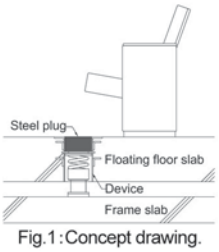


Fig.1：Concept drawing.

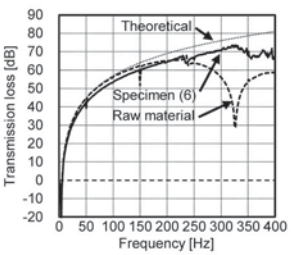


Fig.2：Results of transmission loss.

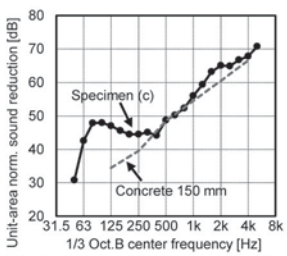


Fig.3：Results of unit-area normalized sound reduction index.

1-Q-6

1-Q-6　ホール空間に露出した空調ダクトの吸音特性に関する実験的検討

Experimental study of the absorption characteristics of air conditioning ducts exposed in a hall space

○中川 貴美子(ヤマハ)

◆近年、特定天井の耐震対策やホール建設コストの圧縮等の理由により、内装天井を設けないホールが増加している。内装天井を設けないことにより、天井に露出する設備類、中でも面積の大きい空調ダクトによる吸音過多が懸念される。

◆本報では空調ダクトの吸音特性について、測定用のダクトを作成し、残響室法吸音率を測定した結果を報告する。また、測定したデータを使用して多目的ホールの残響推定計算を行い、実測値と比較した結果を紹介する。

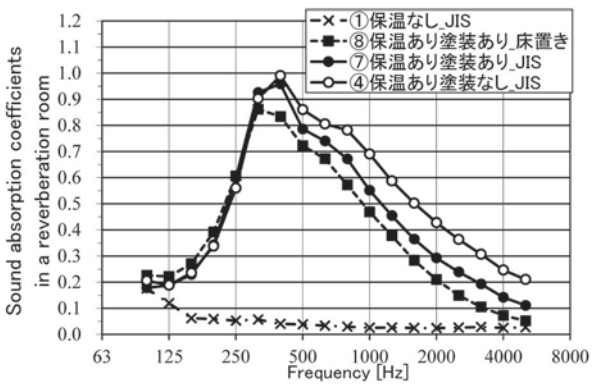


Fig.1：Sound absorption coefficients in a reverberation room of air conditioning ducts

1-Q-7

1-Q-7　複数の障害物で構成された音響ディフューザ

Acoustic diffuser composed of multiple obstacles

☆増川英寿，齋藤優真，高橋義典（工学院大）

◆従来型の音響ディフューザの凹凸を見かけ上平坦にする目的で，複数の円盤で構成された音響ディフューザを設計した(Fig. 1).

◆反射音の指向性を FDTD シミュレーションを用いて計算した.

◆円盤表面での鏡面反射が確認されたものの，対象とする周波数では反射音の拡散が確認された(Fig. 2).

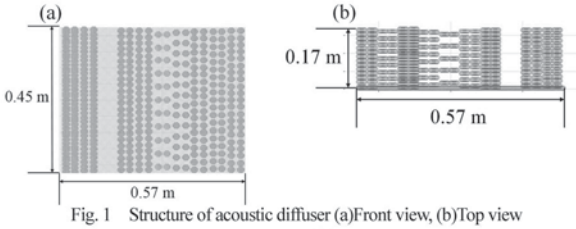


Fig. 1　Structure of acoustic diffuser (a)Front view, (b)Top view

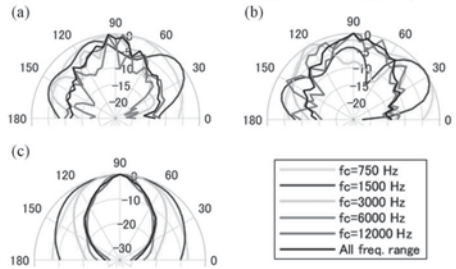


Fig. 2　Directivity of reflected sound (a)Proposed acoustic diffuser, (b)Conventional acoustic diffuser, (c)Flat board

1-Q-9

1-Q-9　2023 年に採集したミンミンゼミの高潮音のピーク周波数の統計分析

Statistical analysis of peak frequency of climax sound for robust cicada in 2023

☆御園玲央，武藤憲司（芝浦工大）

◆本研究は，ミンミンゼミの鳴き声で最も音圧レベルが大きい高潮音の 1 周期「ミ」「ン」の部分の周波数特性を分析した.

◆我々のこれまでの研究で，10 kHz まで自由音場とみなせる範囲で，ミンミンゼミの鳴き声で「ミ」と聴こえる部分の周波数特性を正確に測定してきた．本研究では 28 kHz まで自由音場の環境下で，1 匹当たりのミンミンゼミの鳴き声の周波数特性を測定した.

◆ミンミンゼミ 73 匹の高潮音「ミ」「ン」のピーク周波数を統計分析した．「ミ」のピーク周波数の平均値は 4727 Hz，標準偏差は 213 Hz であった．「ン」のピーク周波数の平均値は 15648 Hz，標準偏差は 1014 Hz であった.

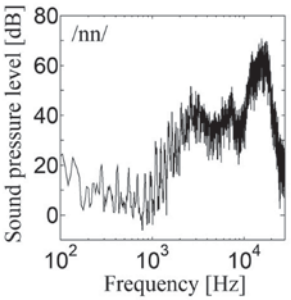


Fig.1：Spectrum in the part of /nn/ made by a robust cicada ($r_2 = 1$ m).

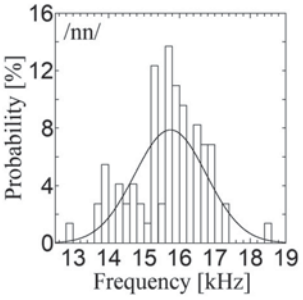


Fig.2：Histogram of peak frequency of /nn/ made by 73 robust cicadas (Mean = 15648 Hz)

1-Q-10

1-Q-10　複数マイクロホンアレイを用いた移動音源の平面内位置推定に関する実験的検討

Experimental investigation about horizontal source localization of moving noise source using multiple microphone arrays

◎星野嗣人，竹林健一，田中彩，矢入幹記（鹿島技研）

◆建設現場でのリアルタイムでの騒音源位置把握を目的とし，移動する音源の同定に対するマイクロホンアレイの適用を検討した。

◆単一の音源を屋外で位置推定を，Minimum Variance 法を用いて下記手順で実施した。

1. 複数のマイクロホンアレイを敷地周辺に設置する。
2. 各マイクロホンアレイにて任意高さにおける水平面上の等角度ごとの空間スペクトルを算出する。
3. 各マイクロホンアレイの結果を重ね合わせて音源位置を推定する。

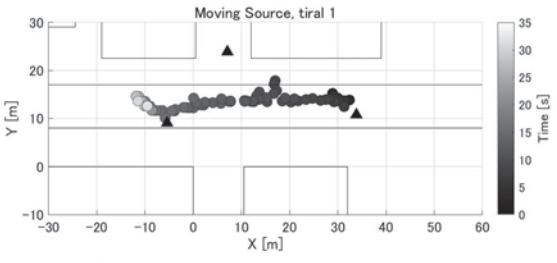


Fig.1 Localization of moving source using three microphone-arrays.
▲: Microphone array which has seven microphones.
○: Estimated source locations.

1-Q-11

1-Q-11　200Hz 以下の騒音を対象とした板振動利用 2 周波数吸音音響メタマテリアル

The evaluation of the developed two peak sound absorber using the plate vibration for under 200Hz noise.

◎後藤 達彦，江波戸 明彦，西村 修（東芝）

◆電源由来の騒音や、ファン・モーターなどの回転要素由来の騒音として、60Hz, 120Hz など、一般吸音材の吸音効果が低い周波数帯域にて、顕著にピーキーな騒音が生じることが機械騒音では多い。本稿では、提案する板振動利用 2 周波数吸音音響メタマテリアルの原理と、単体ユニットの音響管評価結果を示す。

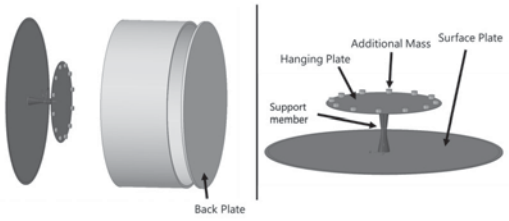


Fig.1：The structure of the proposed sound metamaterial.

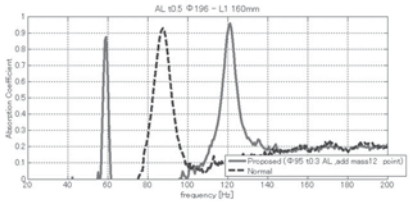


Fig.2：Measured absorption coefficient

1-Q-12

1-Q-12　MEMS 気圧センサと小型超低周波マイクロホンを用いた複合型インフラサウンド観測装置

Hybrid infrasound measurement device comprised of a MEMS pressure sensor and tiny infrasound microphones

○西村竜一（NICT），△大井拓磨（クローネ），鈴木陽一（東北文化学園大/NICT）

◆従来、インフラサウンド（超低周波音）観測には空振計や精密微気圧計が用いられてきたが、インフラサウンド周波数帯域全体を網羅できる安価な装置は存在しなかった。

◆日本音響学会 2023 年秋季研究発表会において、補聴器用マイクロホンを元にした**小型超低周波マイクロホン**※が発表された。

◆異種センサ（**MEMS 気圧センサ**および**小型超低周波マイクロホン**※）を組み合わせて複合型のインフラサウンド観測装置を構成することで、周波数および振幅のダイナミックレンジの拡張を試みた（Fig.1）。

◆2 次標準にも用いられる精密微気圧計を比較対象とした並行観測による評価の結果、平均二乗残差は 0.2 Pa 以下で、残差信号は全周波数帯域において暗騒音レベルを概ね下回ることが確認された。

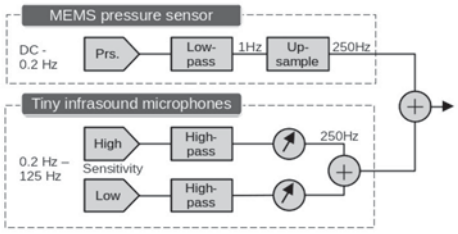


Fig.1：Blockdiagram of the hybrid infrasound measurement device.

1-Q-13

1-Q-13　音声強調のための拡散モデルにおける計算量の削減

Computational Reduction of Speech Enhancement Using Diffusion Models

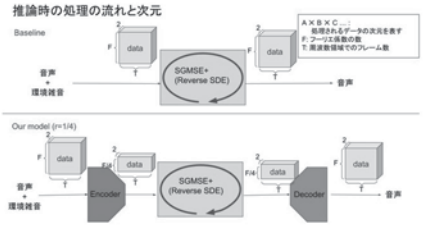
○西悠希（東工大），岩野公司（東京都市大），篠田浩一（東工大）

◆　本研究は，拡散モデルにおける生成段階の計算コストを，なるべく精度を落とさずに削減することを目指したものである。

◆　NN の生成モデルにおいて高精度を達成していた GAN と比べ、拡散モデルによる NN の生成モデルは、同程度の精度を達成しつつ訓練も安定しているという利点がある。しかし、生成段階では何度も NN にデータを通すので計算コストが大きいという問題点がある。

◆　Encoder, Decoder を用いた潜在空間を介することで，入力される音声データを圧縮し，その中で拡散モデルの生成処理プロセスを実行することで，計算コストの削減を精度を落とさずに実現した。

◆　特に圧縮率を1/4にすることで、精度を落とさずに 50% 以上の推論時の所要時間の短縮を達成した。



1-Q-14

1-Q-14　雑音・無音棄却型リアルタイム VAD レス音声認識モデルの開発

Development of Noise and Silence Rejection Real-time VAD-less Speech Recognition Model

☆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,事前処理を行うことなく,雑音や無音区間が含まれた状態の音声の入力が可能である VAD レスな音声認識モデルの開発を行う.

◆提案モデルの開発は Wav2Vec 2.0 XLSR-53 を用いて行った.

◆音声認識モデルが発話区間と非発話区間を同時に認識することが可能となるように,学習データの音声中に雑音区間と無音区間の２種類の非発話区間を挿入した.

◆各音声に対応する書き起こし中に非発話区間を表すタグを付与し,モデルの学習を行う.これにより音声認識モデルが発話区間,雑音区間,無音区間の認識を同時に行うことができることが期待される.

◆無音区間の棄却ではほぼ 100%,雑音区間の棄却では約 90%の棄却率を達成することができた.

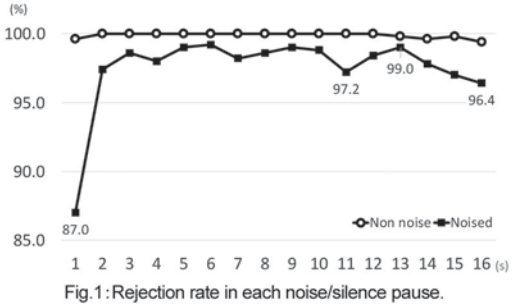


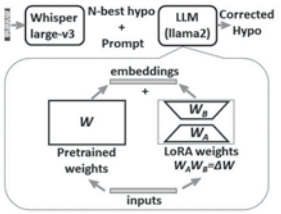
Fig.1：Rejection rate in each noise/silence pause.

1-Q-15

1-Q-15　Investigating Effective Methods for Combining Large Language Model with Speech Recognition System

◎李 勝 (NICT)，楊 正東 (京大)，△周 汪勁 (京大)，△褚 晨翠 (京大)，河井 恒 (NICT)

◆ This paper uses LLM for ASR error correction. We use second-pass rescoring in the output transcriptions generated by the ASR system (N-best decoding hypotheses).

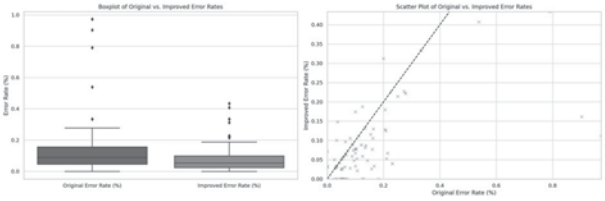


The experimental settings:

- LLM is ELYZA-japanese-Llama-2-7b.
- Whisper-Large-v3 generates 10-best outputs (1-best CER%=12.91).
- NVIDIA Tesla V100 GPU 8-bit LoRA finetune.
- 15 epochs, learning rate 1e-4, batch size 64, and LoRA rank 4.
- Japanese dataset from SPREDS-U1.
- NIST-SCTK for evaluate the Character Error Rate (CER%).

◆ The improvement from LLM model (Epoch 11) CER% is statistically significant (p-value<0.01). This paper proves combining LLM with a speech recognition system effectively improves speech recognition performance, even in low-resource settings.

Before Correction ASR CER%	After Correction CER%			
	Epoch 7	Epoch 9	Epoch 11	Epoch 15
12.91	26.20	8.24	**7.77**	16.51



1-Q-16

1-Q-16　LongHuBERT: 文書単位の音声言語処理のための自己教師あり学習

LongHuBERT: Evaluating the Importance of Leveraging Full Document Contexts in Speech Processing

ウィリアム・チェン（CMU），○ 叶高朋（NTT），小川厚徳（NTT），マーク・デルクロア（NTT），渡部晋治 （CMU）

◆文書単位の音声認識・翻訳・要約を行うために、Conformer の自己注意機構の計算量を削減する方法を提案し、自己教師あり学習モデルである HuBERT へ適用する。

◆発話単位ではなく、文書単位の End-to-end 音声言語処理では、長い文脈情報を学習することができる。

◆本研究で提案したモデルにおいては、自己注意機構の計算量の削減を実現しているが、一方でその表現力の低下が生じている。しかし、提案モデルは文書単位の学習を実現することで、結果的に発話単位のモデルよりも高い精度を示した。

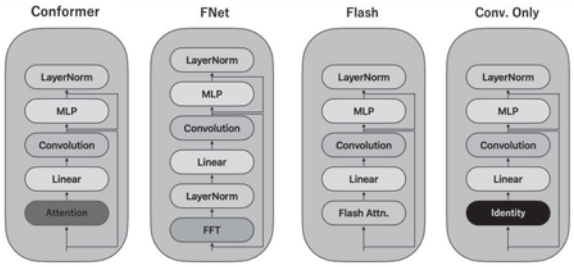


Fig.1: Overview of each LongHuBERT architecture.

1-Q-17

1-Q-17　シングルチャネルマルチトーカー重畳音声認識と発話時間推定の同時モデリング

Joint modeling of single-channel multi-talker overlapped speech recognition and utterance-level timestamp prediction

◎牧島直輝，鈴木啓太，鈴木聡志，安藤厚志，増村亮（NTT）

◆マルチトーカー重畳音声認識は，複数話者のオーバーラップを含む音声に対して各話者の発話内容を推定するタスクである．

◆従来のマルチトーカー重畳音声認識手法では，各話者の発話時間が推定されないため，実用上重要である話者交代のタイミングやオーバーラップの検知及び非オーバーラップ音声区間の推定ができないという課題が存在する．

◆本発表では，これを解決するため，マルチトーカー重畳音声認識と発話時間推定の同時モデリングを提案する．提案法では，各話者の発話内容と発話毎の離散化された発話開始時間及び発話終了時間を自己回帰モデルにより逐次的に推定する．

◆評価実験により，提案法は従来のシングルトーカー音声認識及びマルチトーカー音声認識を上回る音声認識性能を達成し，また，従来のダイアライゼーション手法を上回る発話時間推定性能を達成することを示す．

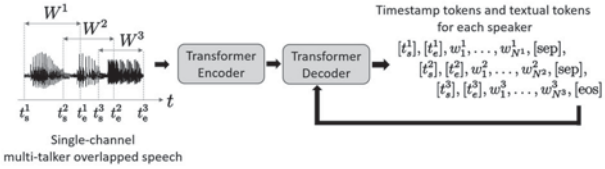


Fig.1： Overview of the proposed method.

1-Q-18

1-Q-18　話者の学年に着目した子ども音声認識の評価

Evaluation of child speech recognition focusing on speaker's grade level.

◎堀井こはる，俵直弘，小川厚徳，荒木章子（NTT）

◆子ども音声は特徴が大人と大きく異なり，また，年齢による変化が大きいことから，大人音声を対象にした場合に比べ音声認識が困難

◆話者の学年に着目し，子供の成長が音声認識精度に与える影響を調査

- 比較モデル：子ども音声で学習された GMM-HMM モデル，DNN-HMM モデル，大人音声のみで学習された end-to-end（E2E）モデル，子ども音声で適応された E2E モデルを評価
- 実験結果:すべてのモデルにおいて学年が低いほどWERが高く，年少の子どもの音声認識が困難なタスクであること，子どもの成長が音声認識精度に与える影響が大きいことが示された（Fig.1）

◆今後は子どもの成長を考慮した音声認識モデルの実現を目指す

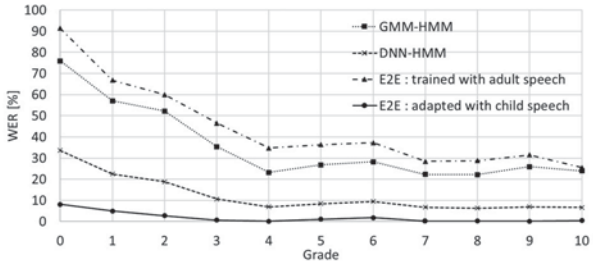


Fig.1:Child speech recognition results as a function of speaker's grade.

1-Q-19

1-Q-19　発音出力を利用した chain of thought 音声認識

Chain of thought speech recognition using pronunciation output.

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

◆Encoder-decoder 型の end-to-end 音声認識でカタカナを出力させてから、漢字仮名交じりの表記を出力させるモデルを提案。

◆実験では Base(漢字仮名交じり表記だけを出力する)、Dual(カタカナ列と漢字仮名交じりの表記列を 2 つの decoder で出力)、CoT(提案手法)を比較。

◆Encode された音響特徴量を cross attention で decoder に導入するよりも、言語的な解釈の進んだ状態での文脈も decoder に入力したほうが、同音異表記語の多い日本語では有利であると確認。

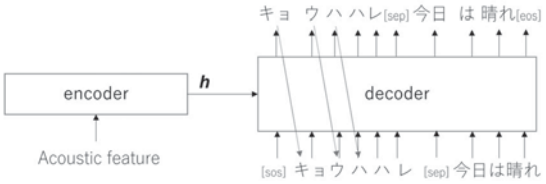


Fig.1:Proposed CoT model

Model	eval1	eval2	eval3
Base	6.1/3.3	4.3/2.3	5.4/3.1
Dual	5.8/3.1	4.1/2.3	4.9/2.7
CoT	**5.4/2.9**	**3.9/2.2**	**4.5/2.5**

Table 1 left side is Character error rate(%) and right side is substitution error rate(%)

1-Q-20

1-Q-20　wav2vec 2.0 と疑似ラベリングを活用した脳性麻痺者の音声認識

Speech recognition for persons with cerebral palsy utilizing wav2vec 2.0 and pseudo-labeling.

☆松坂勇樹(神戸大), 高島遼一(神戸大/JST さきがけ), 滝口哲也(神戸大)

◆脳性麻痺者の多くは、手足を自由に動かすことが困難であり、日常生活の支援として音声認識を用いた支援システムが期待されている。

◆健常者との発話特徴の違いのため、本人の音声で学習した音声認識システムが必要となるが、大量の音声データの収集が難しい。

◆本研究では、より多くのデータを収集するためのアプローチとして、脳性麻痺者のラベル無し音声を学習に活用することを検討する。

◆ラベル無し音声の活用法として、以前我々の先行研究で実施した wav2vec 2.0 の自己教師あり学習による活用法に加え、今回新たに疑似ラベリングによる活用法を取り入れ、2 つの活用法を併用する。

◆Fig.1 のように、2 つの方法を併用してラベル無し音声を活用することで、脳性麻痺者の音声認識における認識性能の向上を確認した。

(学習手順の詳細は講演原稿にて記載)

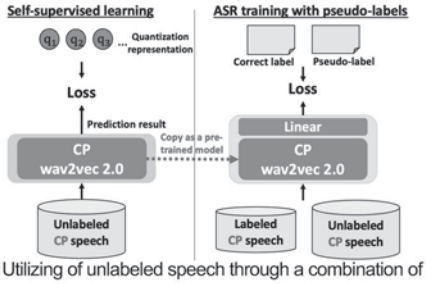


Fig.1 : Utilizing of unlabeled speech through a combination of self-supervised learning with wav2vec 2.0 and pseudo-labeling.

1-Q-21

1-Q-21　再帰的フィードバックを用いた階層的マルチタスク学習による End-to-End 音声認識

Hierarchical Multi-Task Learning with CTC and Recursive Feedback

☆楠奈穂美, 樋口陽祐, 小川哲司, 小林哲則 (早大)

◆研究目的 : End-to-End 音声認識における, 洗練メカニズムの導入による階層的マルチタスク学習(HMTL)モデルの精度向上

◆HMTL モデル（Fig. 1）

- H-CTC(Fig. 1a)…階層化し,入力から出力まで段階的に推定
- **提案の HR-CTC**(Fig. 1c) …**条件付けと再帰操作**を導入し,パラメータの増加なしに各階層での情報を洗練

◆実験：**再帰操作を導入**することで性能改善を確認

- 推論時**リピート数の増加**により,性能改善(学習時のリピート数=3)

Model	#Repeat N	WER[%]
H-CTC		12.7
HR-CTC	1	17.1
	2	12.0
	3	11.7

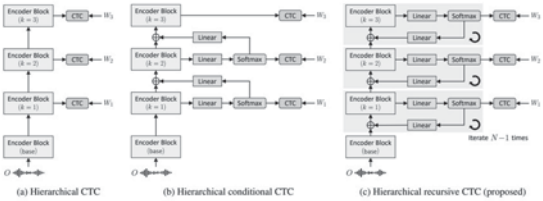


Fig.1 : Comparisons between hierarchical multi-task learning models

1-Q-22

1-Q-22　CTC を用いたストリーミング音声認識における言語情報の置換に基づくドメイン適応

Domain Adaptation Based on Language Information Substitution in Streaming Speech Recognition Using CTC.

☆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] で CTC を用いたストリーミング ASR モデルにて DRA によるドメイン適応が有効であることが示されている。

◆ この結果は ASR モデルのサイズや言語モデルのサイズ、データのサイズに依存するものであるのか。

＜ 提案手法 ＞

◆ 大規模なデータセット、より大きな n-gram オーダーの言語モデル、パラメータ数の大きなエンコーダを持つ ASR モデルを使用して DRA によるドメイン適応実験を行う。

＜ 実験結果 ＞

◆ 結果から認識精度は向上したが、先行研究とは異なる傾向が見られた。

◆ 相対的な CER の改善に関しては先行研究には及ばなかったことなども踏まえて、データやモデルの選択、デコーディング手法についての検討が重要であることがわかった。

Method	LM_{src}	LM_{tgt}	λ_{sub}	λ_{add}	CER (%)
baseline	-	-	-	-	25.375
DRA	1-gram	1-gram	1.0	1.1	**25.356**
DRA	1-gram	2-gram	0.1	0.4	25.359

[1] 高城巽成, 小川厚徳, 若林佑幸, 北岡教英 : "CTC デコーダを用いたストリーミング可能な音声認識における density ratio approach に基づくドメイン適応 ", 日本音響学会秋季研究発表会, No.3-Q-6 (2023 年 9 月).

1-Q-23

1-Q-23　聴覚障害者音声における自己教師あり学習に基づく音声認識の評価

Evaluation of Automatic Speech Recognition based on Self-supervised learning for deaf and Hard-of-Hearing people.

☆高橋 快斗, 木内 貴浩, 高城 巽成, 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・目的

◆聴覚障害者音声の音声認識の精度が低い

◆構音障害音声において自己教師あり学習が有効 [1]

◆聴覚障害者音声における自己教師あり学習の有効性を検証

手法

◆自己教師あり学習に基づく音声認識の評価

- XLS-R, Conformer-CTC, Hybrid CTC/Attention で評価実験

結果

◆ XLS-R を日本語音声で追加事前学習をし, 健聴者音声, 聴覚障害者音声と続けて二段階ファインチューニングしたモデルが健聴者音声, 聴覚障害者音声共に最も高精度

Table 2　CER of models which are finetuned by DEAF corpus after trained using normal-hearing speech, where "pretrain" means pre-training in only XLS-R, "Default" indicates commonly-used databases for training XLS-R, that is, the combination of VP-400K, MLS, CV, VL, and BBL, "→" and "+" respectively indicate the order of training and the concatenation of datasets.

model	train/pretrain	fine-tune	JNAS CER	DEAF CER
Hybrid CTC/Att	JNAS	DEAF	8.8	30.1
Conformer-CTC	CSJ + JNAS + CEJC + LTV	JNAS → DEAF	2.9	15.3
XLS-R	Default → LTV	JNAS + DEAF	4.2	18.6
XLS-R	Default → LTV	JNAS → DEAF	**1.4**	**15.1**

[1] Abner Hernandez et al. "Cross-lingual Self-Supervised Speech Representations for Improved Dysarthric Speech Recognition," 2022.

1-Q-24

1-Q-24　ChatGPT によるデータ拡張に基づく電子カルテ入力用に適応した音声認識モデルの構築

Construction of a Speech Recognition Model Adapted for Electronic Medical Record Entry Based on Data Augmentation by ChatGPT

☆LI CHENGFENG, 高城巽成, 若林佑幸, 北岡教英 (豊橋技術科学大学)

＜ 背景･目的 ＞

◆小規模なテキストデータから大量の模擬テキストデータを生成して，言語モデルを作成することで音声認識精度の改善に有効 [1]

◆医療専門用語を用いてテキストデータを拡張して，音声認識率を改善する

＜ 提案手法 ＞

◆テキストデータ拡張によるドメイン適応

◆テキスト･音声データ拡張によるドメイン適応

＜ 実験結果 ＞

◆提案法により専門用語に対応した言語モデルを音声認識モデルと統合して音声認識率を改善した

◆ファイチューニングした音声認識モデルには Shallow Fusion 手法を用いて，音声認識率を改善した

Table 1: CER(%) of Language Model Fusion Experiment

		SF		DRA	
Model	base	LM_{old_tgt}	LM_{new_tgt}	LM_{old_tgt}	LM_{new_tgt}
ASR_{base}	32.0	27.8	19.0	25.8	15.9
ASR_{ft}	13.1	11.6	11.6	14.3	13.6

[1] 前島亮, 森大輝, 若林佑幸, 北岡教英. 小規模学習データドメインのための文生成に基づく音声認識用言語モデルの構築. SPEASIP ワークショップ, 3 2023.

1-Q-25

1-Q-25　フレーム単位･発話単位自己教師あり学習による汎用･話者表現モデルのレイヤー別分析

What Do Self-Supervised Speech and Speaker Models Learn? New Findings From a Cross Model Layer-Wise Analysis.

◎芦原孝典, デルクロアマーク, 森谷崇史, 松浦孝平, 浅見太一, 井島勇祐 (日本電信電話株式会社)

◆音声向け自己教師あり学習 (音声 SSL) は，文脈を考慮したフレーム単位の音声表現を学習するように設計されているのに対し，近年，話者や感情などの発話単位の音声情報を捉えることに特化した話者 SSL の研究も盛んとなっている

◆しかしながら，音声 SSL とは対照的に，話者 SSL が音声のどのような要素表現を捉えているのかについては，ほとんど分かっていない

◆そこで本稿では，音声 SSL・話者 SSL・教師あり話者モデルの表現の違いを探索するために，2 種類の手法 (プロービングタスク・類似度分析) を通じて間接的・直接的に比較する (Figure 1)

◆実験から，2 つの話者モデルは特に中間層で異なる表現を示し，それが話者タスクの精度差につながっていることが示唆された．また，これらのモデルは音声 SSL に見られるような高度な言語表現は獲得していないようであった．

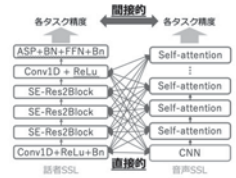


Figure 1: Schematic diagrams of the analysis methods for the direct and indirect comparison.

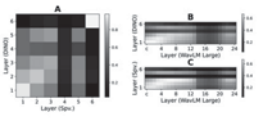


Figure 2: Similarity results of each layer comparing (A) DINO with supervised speaker model, (B) DINO with WavLM Large, and (C) supervised speaker model with WavLM

1-Q-26

1-Q-26　ニューラルネットに基づく音素アライメントのためのモデル構造・学習法および特徴量の比較評価

Comparative Evaluation of Model Structures, Training Methods, and Features for Phoneme Alignment Based on Neural Networks

○郡山知樹 (サイバーエージェント)

- **背景**
  - 近年のend-to-end音声認識・音声合成モデルの音素アライメントは不正確な音素境界を得るには不十分
- **手法**：
  - VAEによる音響・言語特徴量抽出とアライメントロスを用いた多段階学習
- **実験**：様々な実験設定によるアライメント性能の比較
  - Conformerエンコーダや畳み込みなどネットワークの構造
  - forward-sum/HMM/HSMMによる継続長制約
  - HuBERTやWhisperによる音響特徴量入力
  - 音素単位 vs 状態単位

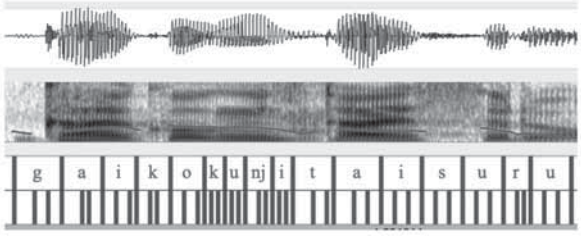


Fig 1: Alignent result using the proposed model.

1-Q-27

1-Q-27　FIRNet: F0 制御可能な有限インパルス応答に基づく高速ニューラルボコーダ

FIRNet: fast neural vocoder based on finite impulse response with F0 controllability

○大谷大和, 岡本拓磨 (NICT), 戸田智基 (NICT/名古屋大学), 河井恒 (NICT)

◆有限インパルス応答 (FIR) に基づく基本周波数 (F0) の制御および高速な波形生成が可能なニューラルボコーダ，FIRNet を提案．

◆ソースフィルタモデルに基づき，混合励振源に対して音響特徴量からニューラルネットを通じて得られた複数の FIR 係数を畳み込むことで音声波形を生成．

◆評価実験より，提案する FIRNet はニューラルボコーダとしての品質を維持しつつ，F0 制御性と従来の信号処理型ボコーダと同程度の波形生成速度を実現．

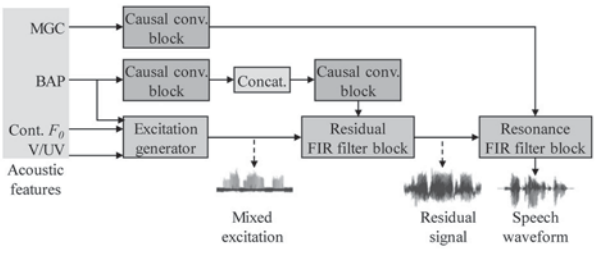


Fig.1: Architecture of FIRNet

1-Q-28

1-Q-28　テキストスタイル変換を用いた話し言葉音声合成

Spoken speech synthesis with text style transfer.

☆吉岡大貴, 安田裕介, 戸田智基 (名大)

テキスト音声合成において，話し言葉風・キャラクタ風などの自由な「スタイル」を付与した音声を合成するためには，音声だけでなく「発話するテキスト」自体のスタイルも制御可能な技術が有用であると考えられる．我々は，Conditional VAE (CVAE) に注意機構と位置埋め込みを導入し，ノンパラレルな所望のスタイルテキストデータセットを用いてテキストスタイル変換学習を行う CVAE+Attn を提案した．さらに，CVAE+Attn で生成したテキストを疑似的な正解変換文として利用して，再構成だけでなく，疑似正解変換文を元のスタイルに戻すことで変換処理も考慮した学習を可能とする CycleCVAE+Attn+CWAPE を提案し，その有効性を示した．

スタイル制御されたテキストを音声合成した際，そのスタイルを持っている自然な音声になっていること，元テキストの意味を損なわない音声であることが必要である．本稿では，テキストスタイル変換と TTS を組み合わせて上記２つの項目が達成できるか検証するため，日本語話し言葉コーパスと日本語 Wikipedia を用いて，「非流暢性の有無」「話し言葉・書き言葉」に関するテキストスタイル変換を行う．これらテキストの合成音声に対する主観評価実験結果から，テキストスタイル変換を用いた TTS の課題を明らかにする．評価に用いたテキストサンプルの一部を Table. 1に示す．

Table1. Sample text used for TTS

変換前	それから合わせて、こちらにあるビデオカメラで乳児の様子を録画しました。
変換後	えーそれから合わせて、ここちらにあるあのビデオカメラで乳児の様子を録画しました。
正解	えと合わせて、こちらにあるビデオカメラでありますけども乳児の様子を録画しました。

1-Q-29

1-Q-29　日本語音声合成における対話行為情報の利用による韻律改善

Prosody Improvement by Using Dialog Action Information in Japanese Speech Synthesis

☆佐藤匡紀, 高道慎之介, 猿渡洋 (東大院・情報理工)

◆ターンテイキングにおける韻律を表現する TTS モデルの研究は少ない．本研究ではターンテイキングに関する対話行為情報と句末境界音調を用いるテキスト音声合成モデルについて，イントネーションの再現性やターンテイキングの表現に対する有効性を検証した．

◆既存の日常会話コーパスに対して発話ターンの維持・譲渡に関するラベルを付与して学習データに用いた．

◆イントネーションの自然性に関する XAB テストと，前後の発話と共に一連の会話として合成音声を受聴してターンテイキングの自然性を評価する AB テストを実施した．

◆複数の対話行為情報の入力方法を比較した結果，対話行為情報から句末境界音調を予測しその句末境界音調を TTS モデルに入力するモデルが主観的なターンテイキングの自然性と客観指標を改善した．また，句末境界音調の入力がターンテイキングの自然性を改善することが示唆された．

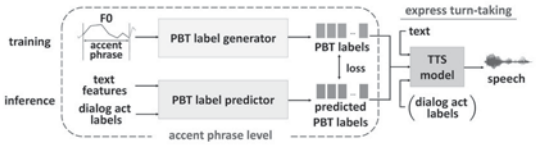


Fig.1 : Concept of proposed method.

1-Q-30

1-Q-30　暗黙的な非線形処理を導入した拡散モデルを用いた音声合成

Speech Synthesis Using Denoising Diffusion Probabilistic Model with Implicit Nonlinear Refinement

☆岡本海, 上乃聖, 李晃伸(名工大)

◆拡散モデルにおける拡散過程と逆拡散過程において非線形処理を可能にした音声合成モデルを提案する．

◆Grad-TTS のアーキテクチャをベースとし，Implicit Nonlinear Diffusion Model (INDM)の学習・生成アルゴリズムに基づいたデコーダを導入した．

◆提案手法による合成音声の品質はベースラインモデルと同等以上の性能を示し，特に拡散ステップ数を少なくした時に効果を発揮した．

Table 1. The MOS with 95% confidence intervals, MCD (Mean Cepstral Distortion), PESQ (Perceptual Evaluation of Speech Quality) and RTF.

	MOS	MCD	PESQ	RTF
Ground-Truth	4.77±0.07	-	-	-
Grad-TTS 1	2.77±0.13	7.343	1.188	0.008
Grad-TTS 5	3.86±0.13	6.976	1.193	0.019
Grad-TTS 10	3.96±0.13	6.930	1.212	0.033
Grad-TTS 100	3.90±0.13	6.917	1.193	0.292
Proposed 1	2.15±0.14	7.223	1.189	0.020
Proposed 5	**3.99±0.13**	**6.943**	1.196	0.031
Proposed 10	3.90±0.13	6.947	1.197	0.045
Proposed 100	3.80±0.13	6.934	1.205	0.306

1-Q-31

1-Q-31　日本語日常会話の潜在的な発話スタイルに基づく対話シーンに応じた音声合成

Speech Synthesis for Japanese Conversational Scenes Based on Implicit Speech Style

☆嶋崎純一, 上乃聖, 李晃伸(名工大)

◆Chiba らはCEJC の会話の書き起こしテキストを用いて因子分析を行い，７個の共通因子に分割でき，それらを日本語日常会話内の潜在的な発話スタイルとした．

◆日本語日常会話を再現するために対話シーン ID 化モデルと，日本語日常会話に存在すると示された因子の各対話シーンの因子得点を用いたモデルの ２ つの対話シーンに合わせて制御可能な音声合成モデルを提案する．

◆対話シーン ID 化モデルの合成音声の品質ではベースモデルと比べて同等以上の性能を示した．

◆因子得点を用いたモデルでは合成音声の品質では PESQ 以外でベースモデルと比べて同等以上の性能を示した．

Table 1. The MOS with 95% confidence intervals, MCD (Mean Cepstral Distortion), PESQ (Perceptual Evaluation of Speech Quality).

	MOS	MCD	PESQ
Ground Truth	4.28 ± 0.18	-	-
FastSpeech 2	2.55 ± 0.16	7.80	0.82
Based on Speech Style ID	2.26 ± 0.13	**7.45**	**0.86**
Based on Factor Score of Speech Style	**2.89 ± 0.15**	7.61	0.76

1-Q-32

1-Q-32　英語リスニング学習支援のための複数話速スタイルのモデル化と制御に関する検討

A study on modeling and controlling multiple speech-rate styles for English listening learning assistant

☆千葉理揮，能勢 隆，伊藤 彰則（東北大）

◆本研究では，英語リスニング学習支援システムのための，話速を制御可能なテキスト音声合成モデルについて検討した。

◆話速の制御法により同じ速度の音声でも音響特徴が異なり，英語音声の聞き取りやすさに影響を及ぼす。

◆英語学習支援においては，話速のみならず話速に依存して変化する特徴を考慮し，話速制御後も自然音声により近い特徴をもつ制御法が望ましい。

◆本研究では，低速・高速な音声による英語読み上げコーパスの構築とそれを用いた学習や，継続長損失関数の変更により，話速や話速に依存する特徴をより正確に再現可能な TTS を実現した。

◆また，テキストエンコーダに話速情報を追加で入力することで，特に低速音声を合成した場合に，発話速度に依存する特徴の再現性が高まることが確認できた。

Table 1: Comparison of conventional and proposed encoders by MCD (dB)

Speech rate	Conv. enc.	Prop. enc.
Slow	6.71	6.63
Fast	6.77	6.75

1-Q-33

講演取消

1-Q-34

1-Q-34　吃音者向け TTS システムのための健常者音素継続長を反映した VITS の学習手法の提案

Training of VITS model reflecting the duration of a physically unimpaired speaker for a text-to-speech system for a person with stutter

☆長久保 諒[1]，山下 陽生[1]，高島 遼一[1,2]，安井 美玲[3]，滝口 哲也[1]

([1]神戸大学，[2]JST さきがけ，[3]大阪人間科学大学)

◆本研究では，**吃音者の声質でかつ流暢な音声合成システム**の作成を目的として音声合成システムである VITS の学習手法について，テキスト整形を行う手法と健常者音素継続長を用いる手法の２手法を提案する。

◆吃音者音声合成においては，難発と呼ばれる吃音症状に伴う発話特徴が反映されることで，**「入力文章に無い休止」**や**「極端に長い休止」**が発生してしまう問題がある。

◆テキスト整形を行う手法では，**どちらの問題も改善**されることが確認したが，この手法は整形の手間がかかってしまう。

◆健常者音素継続長を用いる手法では，「極端に長い休止」の問題は改善された。一方，「入力文章に無い休止」の問題については，完全な除去には至らなかったものの，**休止を短縮**しある程度滑らかな音声が合成できることを確認した。

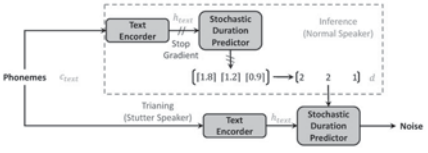


Fig.1 Proposed training procedure for the stochastic duration predictor of VITS.

1-Q-35

1-Q-35 音声合成用コーパスおよび日常会話コーパスのハイブリッドモデリングによる日常会話音声の合成

Synthesis of daily conversation speech using a hybrid model of read and conversational speech corpora

☆古川晃大，森大毅(宇都宮大)

◆ 目的

日本語日常会話コーパス (CEJC) から韻律情報を，別の高品質なコーパス (JSUT) からスペクトル情報を学習することにより，読み上げ音声の合成と同等な品質を保ちつつ，イントネーションや会話のリズムは会話音声そのものであるような音声を合成することを試みる。

◆ 手法

CEJC モデルのスペクトル予測部のみを JSUT でファインチューニングしたモデル (Hybrid) を提案する。

◆ 結果

Hybrid モデルは，日常会話音声らしさは JSUT モデルに比べ高く，CEJC モデルよりも明瞭な音声が合成できることが示された。

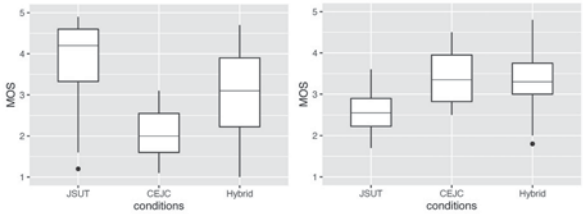


Fig.：明瞭性(左)，日常会話音声らしさ(右)の評価実験結果

1-Q-36

1-Q-36　転移学習を用いた少量データからの口唇動画音声合成

Lip-to-Speech Synthesis from Low-Resource Data Using Transfer Learning

◎藤田直明, 南汰翼, 鏑木時彦（九州大学）

◆ 口唇動画から音響特徴量を予測する口唇動画音声合成は,大規模なデータセットが整備されている英語での検討が主で,日本語等のデータセットが小規模な言語には適用が難しい。

◆ 本研究では口唇動画が少量しかない状況下で適用可能とするため,転移学習と潜在変数の離散化を検討した。TTS(Text-to-Speech)モデルのTacotron を事前学習して学習済 Decoder を獲得し,さらにその Decoder に入力する潜在変数を離散化した。

◆ 実験の結果,TTS 転移学習はいずれのデータ量においても合成音声の評価指標が向上し,より少量のデータでも適用が可能になった。

◆ 転移学習+離散化はデータ量が少ない場合(1000 文以下)では有用であるが,ある程度データがある場合には転移学習のみの方が良好だった。

	PESQ	STOI	WER(%)	PER(%)
Baseline(3000)	1.092	0.587	59.66	24.67
Baseline(2000)	1.072	0.544	70.54	33.12
転移学習 (3000)	1.099	0.615	40.19	13.22
転移学習 (2000)	1.093	0.571	46.13	15.86
転移学習 (1000)	1.059	0.467	78.29	37.28
転移学習+VQ(3000)	1.086	0.591	42.85	14.11
転移学習+VQ(2000)	1.082	0.574	46.72	16.65
転移学習+VQ(1000)	1.084	0.579	56.40	22.58
転移学習+VQ(500)	1.061	0.493	79.58	39.40

合成音声の客観的評価指標

1-Q-37

1-Q-37　NecoBERT：音声合成のために事前学習された自己教師あり学習モデル

NecoBERT: Self-Supervised Learning of Speech Representation for Speech Synthesis

◎中田 亘, 佐伯 高明, 齋藤 佑樹, 高道 慎之介, 猿渡 洋(東大院･情報理工)

Neural Audio Codec と SSL モデルを合体

● 音声再合成に優れたモデルを提案
- ➢ 話者認証でも優れた性能を発揮
- ➢ 一部条件ではメルスペクトログラムより高品質な音声再合成

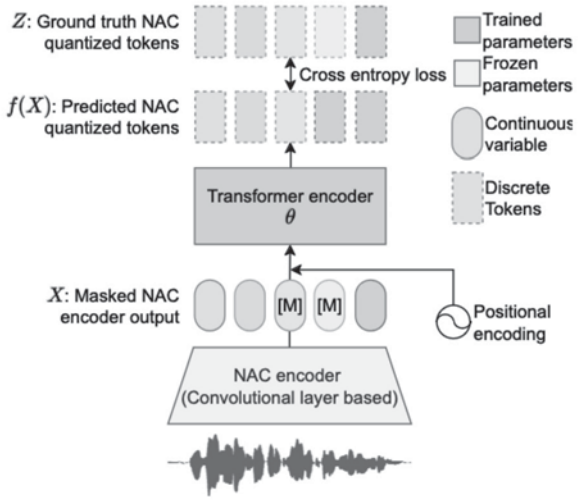


図1. 提案するNecoBERTのモデル構造及び学習の流れ. [M]で示されているのはマスクされたNeural Audio Codecエンコーダ出力である.

1-Q-38

事前学習済みモデルによる埋め込み表現を組み込んだ音声編集モデルの検討

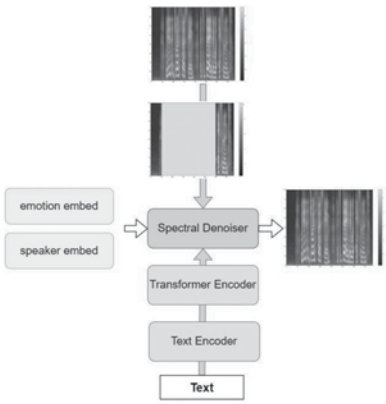
Investigation of speech editing models that incorporate embedded representations with pre-trained models.

○平本佳弘, 中鹿亘(電通大)

◆音声編集技術とは、音声の一部の言語情報を置き換え、元の音声のスタイルをなるべく保ちながら新たに音声を生成する技術である。

◆従来の音声編集は、感情音声などの独特なスタイルの音声の編集が難しいという問題があった。

◆本稿では、拡散確率型の音声編集モデルを基に感情音声に対応した日本語音声編集モデルについて検討する。



1-Q-39

講演取消

1-Q-40

講演取消

1-Q-41

1-Q-41　DNN 音声合成モデル学習のための データ水増しにおける 音声信号の加工方法に関する検討

A study on modification Methods of speech signal in data augmentation for training of DNN speech-synthesis models

○小牧昇矢，坂野秀樹，旭健作（名城大）

◆DNN 音声合成モデルの学習において，学習データの周波数特性が学習結果に大きく影響を与えるものと考え，**音量を保持したまま一部の周波数成分を喪失させる** Method 1，2，**基本周波数と倍音成分を際立たせる** Method 3，4，5，合計 5 種類の水増し手法について検討した。

◆5 種の水増し手法のうち 4 種類の手法において，学習したモデルを用いた**音声の生成に失敗しやすくなった。**特に学習時に過学習に陥っているような挙動がみられた Method 2，5 は、音声の生成に成功する割合が 50%を下回った。

◆基本周波数に着目した Method 3，4，5 は，生成される音声の声質は水増しなしに劣るが，**発音は自然であるように感じられ**，主観評価実験による**評価値が高くなった。**これらの手法のうち，**<u>最も音声の品質評価値の高い手法は、過学習に陥っていた Method 5</u>** であった。

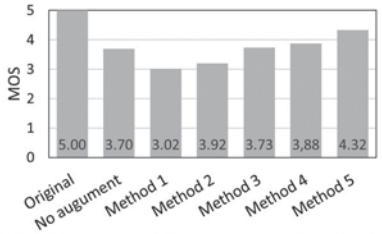


Fig. 1 The mean opinion score for each method.

1-R-1

1-R-1　音声聴取時の脳律動に聴取難易度が及ぼす影響

Effect of listening difficulty on neural oscillations during speech perception

☆齊官重樹，大塚翔，中川誠司（千葉大）

◆ 聴取困難に伴うリスニングエフォート（LE，聴取するために割く注意資源の量）の増大は，日常生活でのストレスや疲労を増加させる要因として認知されつつある．

◆ 雑音下の聴取における LE に応じて，α帯域の脳律動が増強することが報告されている．しかし，情報マスキング(IM)条件下での音声了解度との関係については知見が乏しい．本研究では，EM と IM を想定し、SN 比の異なるノイズを重畳した音声の聴取成績，及び，IM の効果が異なる条件として，妨害音声とターゲット音声が同性，異性，同一話者の場合での聴取成績を測定し，α帯域の脳律動との関係を調べた．

◆ EM 条件では SN 比が下がるにつれて，IM 条件ではターゲット音声と妨害音声の類似度が高いほど，聴取成績が低下することが確認された (Fig.1)．α帯域の強度は，EM・IM いずれの条件においても，非注意条件に比べて増加した (Fig.2)．現時点でのサンプル数 (N=10)では，SN 比，及び，妨害音声の種類に応じた統計的に有意な変動は確認されなかったが，EM 条件では SN 比が低い（LE が高い）ほど強度が増加する一方で，IM 条件ではより聴取難易度が低い（LE が低い）条件で，α帯域の強度が増加する傾向が見て取れる．この結果は，α帯域の脳律動が，EM・IM 下の聴取において異なる役割を果たしている可能性があることを示唆している．

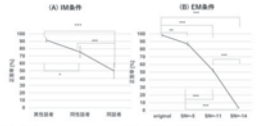


Fig.1 Correct answer rate. **p<0.001, ***p<0.001 (Corrected for multiple comparisons with Ryan's method).

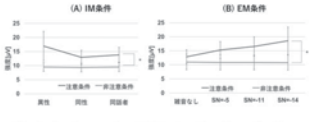


Fig. 2 α band power. * p<0.05 (main effect of attentional vs. non-attentional condition)

1-R-2

1-R-2　若年者と中高年者の音韻修復機能の比較

Comparison of phonemic restoration between young and middle-aged adults

☆湯淺　麻生，大塚 翔，中川 誠司（千葉大）

◆加齢に伴い多くの知覚・認知機能の低下が見られる一方で，音韻修復機能は高齢者と若年者で差がないことが報告されている．それゆえ，音韻修復は，加齢に伴う他の知覚・認知機能の低下を補う役割を果たしているという仮説が提案されている．

◆一方，若年者を対象とした研究では，音韻修復機能を評価する課題に別の課題を課すことで，音韻修復機能が強く働くことが示されている．

◆本研究では，若年者と中高年者における音韻修復機能について，二重課題に対する感受性も含めて比較した．音韻修復課題では，４モーラの単語の 1～4 文字目の子音音素にノイズを置換，または，重畳した．被験者は２つの音声を聴取し，弁別する課題を行なった．音韻修復が生じると弁別が困難になることから誤答率が高いほど音韻修復が生じていると解釈される

◆音韻修復課題のみの場合，中高年者と若年者の成績に差は見られなかった．

◆同時に記憶課題を課した条件では，音韻修復課題のみの場合よりも，中高年者群では共鳴音の誤答率が増加し，若年者群よりも高くなった．

◆聴取以外に注意が割かれる環境下では，加齢に伴い音韻修復が生じやすくなることを示唆している．

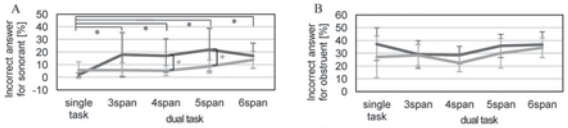


Fig. 1 Percentage of correct responses to phonemic restoration for sonorant (A) and obstruent (B) in the dual task. $^{+}p<.1$, $^{*}p<.05$

1-R-3

1-R-3　音声からの感情弁別に対する難聴の影響 -模擬難聴処理を用いた健聴者実験-

Effects of hearing loss on discrimination of emotion in speech: Experiments using simulated hearing loss sounds for normal-hearing listeners

○花谷幸歩(和歌山大院), 岸田 一馬, △内藤 朱里, 河原 英紀, 入野 俊夫(和歌山大)

◆高齢難聴者への感情音声知覚特性の解明を目指す.本研究では,健聴者を対象に模擬難聴処理を用いた聴取実験を実施.

- 音声スクリーニングを実施し,全 47 感情の音声を５段階評定したのちに主成分分析を行い,最も「喜び・悲しみ・怒り」らしい 10 単語を抽出.
- 感情間をモーフィングした音声を用いて,恒常法で喜び（悲しみ・怒り）の判断をさせ心理物理曲線から PSE/JND を算出.
- 全ての感情対において,健聴状態(Unpro)・模擬難聴状態(80yr)間で PSE/JND に有意差なし.聴覚末梢系の機能低下は感情知覚に影響しない可能性がある.

◆健聴者の結果と比較するために,高齢者を対象とした実験を実施.

- 怒り-喜び対の判断が著しく難しいことがわかった.原因究明は今後の課題である.

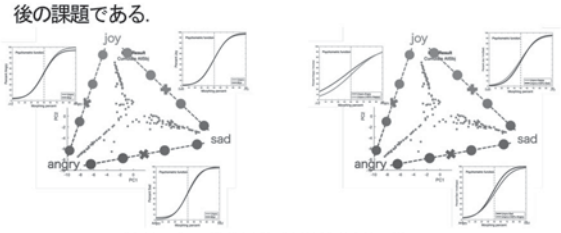


Figure: PCA result and psychometric function.
left: Results for normal hearing (gray: Unpro, purple: 80yr.)
right: Results for elderly (green: Joy, blue: Sad, red: Angry.)

1-R-4

1-R-4　大規模事前学習モデルに基づく音声認識の中間表現を用いた日本語単語了解度の推定

Estimation of Japanese word intelligibility using latent representations of speech recognition based on large-scale pre-training models

☆服部真稀, 近藤和弘（山形大院・理工学研）

◆音声了解度の客観評価手法として，事前学習モデルをクラス分類タスクとして JDRT (Japanese Diagnostic Rhyme Test) の模擬に応用

◆従来法で考慮されていなかった，JDRT 基準の子音分類能力と未知の雑音・単語に対する交差検証を反映

◆モデルはエンコーダに HuBERT 事前学習モデル，デコーダに 2 クラス分類の線形層で構成．子音特徴別に独立したモデルを作成

◆主観評価に対して，単語分割での検証(exp-word)において線形相関 0.94, RMSE 12.86．音素認識による従来法と同程度の精度を達成

◆子音特徴間のバランスと，未知の単語に対する汎化性能が課題

◆今後の拡張として, CTC 誤差とクロスエントロピー誤差によるマルチタスクモデルの実装

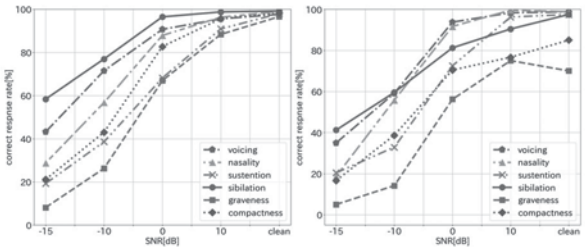


Fig.1 : Correct response rate with subjective(Kaga2006) for each attribute vs. SNR (Left)
Fig. 2: Correct response rate with HuBERT(exp-word) for each attribute vs. SNR (Right)

1-R-5

1-R-5　同一環境での合成音声と自然発話の音声了解度の違い

Comparison of speech intelligibility between synthetic speech and natural speech in the same environment.

◎小林洋介，渡邊健太郎，△梶原颯斗(室蘭工大)

◆近年，統計的合成音声の品質が劇的に向上し，MOS や DMOS の主観評価では人間の発話とほぼ等しい（高い）評価が得られている。本稿では人間の発話とほぼ等しい自然性を達成した Tactron2 を用いて親密度別単語了解度試験用音声データセット 2007 (FW07)の音声単語セットを合成し，人間の自然発話と同条件で音声了解度を比較した。

◆主観評価は発話（自然発話・合成音声），親密度（高・低），インパルス応答畳込の有無の「雑音のない条件」と，発話，親密度，インパルス応答畳込と雑音加算 (SNR7 段階) の「雑音のある条件」で行った。

◆Fig.1 に雑音のある条件かでの結果を親密度別に示す。自然発声より合成音声の方が了解度が低い傾向にあった。また，分散分析より，実験条件の単独作用・交互作用ともに有意差がみられた。

◆この結果は，今回使用した合成音声による FW07 の評価音声には，合成段階でのアクセントエラーがあり（低親密度で約 21%，高親密度で約 4%），FW07 の標準話速である約 5 mora/s より速い約 7 mora/s だったのを FW07 の標準話速に伸長したことも影響している。

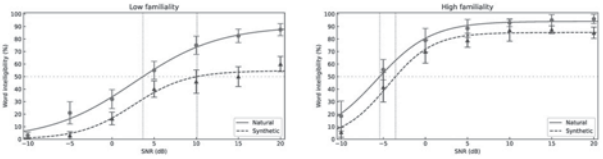


Fig.1: Results of under noise conditions.

1-R-6

1-R-6　周波数変調刺激音の聴取に対する発声の不随意応答: ノイズ付与及び発声母音による検討

Involuntary vocalizations corresponding to modulated frequency: effects of noise masking and type of vowel

☆大森善博(豊橋技科大), 河原英紀(和歌山大), 松井淑恵(豊橋技科大)

◆予測不可能な周波数変調刺激音を聴取しながら一定のピッチを発声することで得られる不随意な補償応答について，刺激音へのノイズ付与及び発声母音条件による影響を調査した。

◆刺激音には，正弦波，調波複合音など 5 種類を用意し，それぞれ 25 cents の振幅で基本周波数変調を加えた。参加者には，刺激音を聴取しながら，120 Hz のピッチで 13 s 間単独の母音を発声させた。

◆ノイズ実験（Fig. 1L）では，これらに 2 種類のノイズを付与した条件を加えた。発声母音は，ノイズ実験では/a/のみ，母音実験（Fig. 1R）では日本語５母音で行った。

◆ノイズの影響については系統的な効果が見られなかった。母音実験については，今後刺激音と参加者の母音スペクトルの関係を調査する。

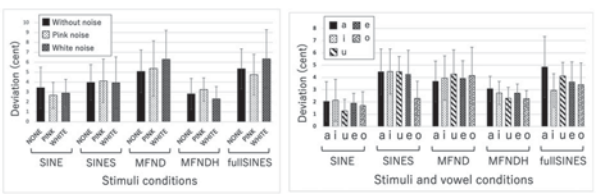


Fig.1 Mean of pitch response to auditory stimuli.
Left: Noise Exp., Right: Vowel Exp. Error bar: SD.

1-R-7

1-R-7 Conversation Scenario Classification Based on Conversation Entrainment

☆Yuning Liu (JAIST), Di Zhou (Kanazawa Univ.), Jianwu Dang (TJU), Aijun Li (CASS), Masashi Unoki (JAIST)

◆**Problem:** How interlocutors employ various strategies reflected in entrainment to guide conversations in different scenarios remains relatively underexplored.

◆**Aim:** To investigate whether entrainment exhibits different patterns in various conversation scenarios and may have the potential to enhance the accuracy of conversation scenario classification tasks.

◆**Solution:** This study proposes a linear model to describe how different dialogue strategies modulate and integrate entrainment calculated by various speech features.

◆**Evaluation:** Use a dialogue corpus containing four conversation scenarios to conduct classification experiments by using entrainment.

◆**Summary:** The entrainment exhibited distinct clusters for different conversation scenarios as shown in Fig.1 and the potential in classification task.

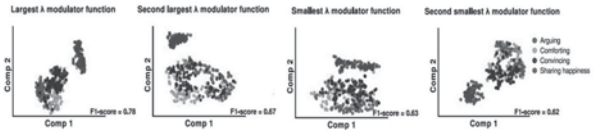


Fig.1: Modulator functions in different conversation

1-R-8

1-R-8 Leveraging Equalization-Cancellation Model in Speech Intelligibility Prediction for Hearing Aids

☆ Xiajie Zhou, Candy Olivia Mawalim, Masashi Unoki (JAIST)

◆ **Aim：** Investigating the Equalization-Cancellation (EC) model to improve speech intelligibility prediction for hearing aids.

◆ **Problem:** Individuals experiencing hearing loss face difficulties in perceiving speech intelligibility in noisy environments.

◆**Solution:** The EC model in the proposed method was employed to adjust signals from both the left and right ears using binaural cues (equalization step) and to eliminate the mutual interaction between signals (cancellation step). This enhances EC model's capability to differentiate target speech from background noise.

◆ **Evaluation:** The speech intelligibility prediction methods were assessed using the Clarity challenge dataset with three metrics: Pearson correlation coefficient (ρ), root-mean-square error (RMSE), and coefficient of determination (R^2).

◆ **Summary:** Compared with HASPI, proposed method improved ρ by 6.4%, reduced RMSE by 14.0%, and improved R^2 by 14.8%.

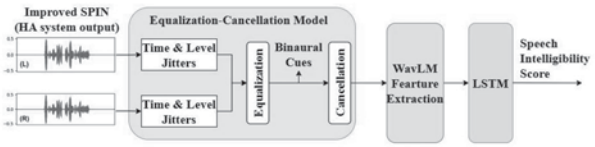


Fig.1: Simplified Overview of Proposed Method

1-R-9

1-R-9 ゲート付き活性化関数を用いた劣化音声に対する non-intrusive 了解度推定

Non-intrusive Intelligibility Estimation for Degraded Speech Using Gated Activation Functions

◎中澤和司（無所属），近藤和弘（山形大）

◆従来の了解度推定方法では聴覚フィルタによる周波数帯域分割と帯域ごとの特徴量に対する重み付け演算がされていた

◆本研究では従来法の演算を意識した Deep Neural Network (DNN)ベースの Non-intrusive（原音声を必要としない）推定モデルを提案

◆スペクトログラムを入力として，周波数軸に沿ったサイズのフィルタとゲート付き活性化関数を用いた重み付け処理を行う

◆日本語（単語了解度），中国語（文章了解度）音声のデータセットで学習し精度を線形相関と順位相関と RMSE で評価

◆既存 Non-intrusive 手法の STOI-net と SRMR と比較して高い精度で了解度推定が可能であることを示す

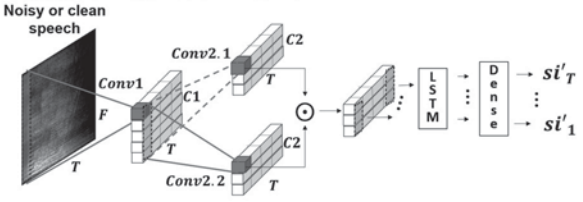


Fig.1: Overview of the proposed method for intelligibility estimation.

1-R-10

1-R-10 想起音声識別のための変分自己符号化器に基づく表現学習

Representation learning based on variational autoencoders for imagined speech classification.

◎矢野　肇，高島遼一，滝口哲也（神戸大），中川誠司（千葉大）

◆脳活動データは収集が難しく，さらに音声想起時の脳信号は信号帯雑音比が低いため，大規模な機械学習モデルの学習は難しい．

◆複数の個人やタスクの異なるデータを集め脳信号の特徴表現を学習することで想起音声の分類精度を向上できる可能性がある．

◆本発表では，変分自己符号化器（VAE）を用いて複数人の音声想起及び音声聴取時の脳磁図の特徴表現を学習し転移学習を行った．

◆VAE を用いた転移学習や VAE の各個人への適応によって，想起音声の分類精度の精度が向上した（Fig.1）．また，Conditional VAE を用いたデータ拡張によっても分類精度が向上した．

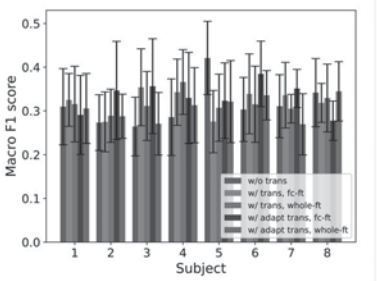


Fig. 1 Macro F1 scores of imagined speech classification with transfer learning. Each errorbar indicates the standard deviation.

1-R-11

1-R-11 フォルマント変形聴覚フィードバックにおける気導・骨導提示が発話に与える影響の検討

Study on the effects of air and bone conduction presentation on speech in formant altered auditory feedback

○上江洲安史，赤木 正人，鵜木 祐史（JAIST）

◆発話音声の聴覚フィードバック経路には気導と骨導が存在するが、聴覚-発話運動制御メカニズムの解明を目的とした研究では、気導経路でフィードバック音声を提示する場合がほとんどである。そのため、フィードバック音声の提示経路の違いが聴覚-発話運動制御にどのような影響を与えるのかは十分に明らかにされていない。

◆本研究では、聴覚フィードバック音声の気導・骨導の提示経路の違いが発話運動制御にどのように影響するのかを、フォルマント変形聴覚フィードバック(AAF)を用いて検討した。

◆日本語母語話者５名を対象に、母音/a/と/e/を発話するフォルマントAAF 発話実験を実施した結果、気導・骨導の提示経路の違いが/e/のF2 補償応答において確認された。これより、フィードバック音声の提示経路の違いが発話運動制御に影響を与えた可能性が示唆された。

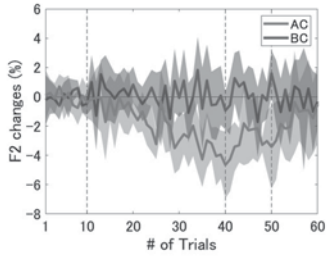


Fig.1: Patterns of change in F2 of the vowel /e/ under air and bone conduction presentation conditions under the formant altered auditory feedback.

1-R-12

1-R-12　雑音下拡声通話音評価における空間知覚と音声処理歪みの交互効果

Study on robust distorted speech quality evaluation method for hands-free communications with auditory spatial perception.

○栗原 祥子，鎌土 記良，千葉 大将，中山 彰
（日本電信電話株式会社 NTT コンピュータ＆データサイエンス研究所）

◆ 雑音環境下における拡声通話の評価法として，空間知覚特性模擬を用いた受聴試験を提案した。本稿では，発話条件を拡大し，空間知覚と音声処理歪みによる評価への影響を調査したので報告する。

◆ 発話者，音素バランスの条件を増やしたことで，信頼区間が小さくなり，条件毎の評価傾向が明確化した。雑音と通話音を分離呈示(separated)することで，カクテルパーティ効果に似た効果が得られることを確認した。空間知覚特性模擬は，拡声通話音の主観評価に少なからず影響を与えるものと考えられる。

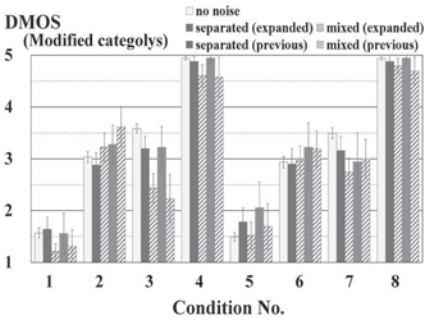


Fig.1: Results of listening tests.

1-R-13

1-R-13　不定数個の音源を含む混合音の低遅延多チャンネル収録のための雑音抑圧ビームフォーマ

Low-latency beamformer enhancing mixture of undetermined number of sources

△Ning Guo (AudioLabs/NTT), ○中谷智広，荒木章子，守谷健弘（NTT）

◆本稿では，不定数個の音源信号を未知の環境で多チャンネル(ch)収録して得られる観測信号に対し，低遅延（≦20 ms）かつ低演算量処理で，全音源信号を保持しつつ背景雑音のみを抑圧する変形パラメトリック多 ch ウィーナフィルタ (Mod-PMWF) を提案する.

◆提案法により，例えば，人数や収録状況が不明な多ch会話収録に対して，リアルタイム処理で雑音を抑圧する多入力多出力ビームフォーマ(BF)を実現できる.

◆Fig. 1 に，MUSHRA テストで，処理音の聴感品質を比較した結果を示す．図より，Mod-PMWF は，他のすべての手法と比べ，各条件下で，統計的に優位に優れた聴感品質を実現していたことが確認できる.

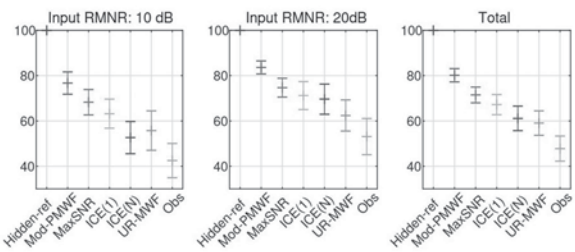


Fig. 1: MUSHRA test results. Error bars indicate 95% confidence based on t-distribution test.

1-R-14

1-R-14　非負値テンソル因子分解に基づく分散マイクロホンアレイを用いたスポットフォーミング

Spotforming using distributed microphone array based on nonnegative tensor factorization

☆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）

◆スポットフォーミング
- Fig. 1 のような特定の領域中にある目的音源を強調する技術

◆提案手法
- 複数ビームフォーマーの出力に非負値テンソル因子分解を適用，フィルタを構成し音源強調

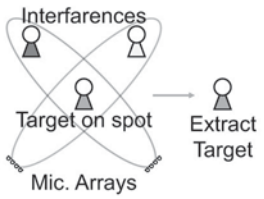


Fig. 1:Situation of spotforming.

◆実験結果
- 非負値行列因子分解を用いた従来手法および非負値テンソル因子分解を用いた提案手法について性能を比較
- SDR 中央値・最大値が高いことを確認

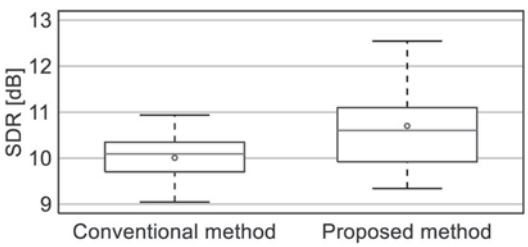


Fig. 2: Box plots of SDRs in simulation when T_{60} = 0 ms.

1-R-15

1-R-15　注目すべき音を自動検出するニューラルビームフォーマの残響下での有効性評価

Evaluation of the effectiveness of neural beamformer with automatic detection of notable sounds under reverberation

☆市川 創大，山田 武志(筑波大)

一般にビームフォーマの空間フィルタ（指向特性）を決定するためには，目的音方向などの事前情報が必要となる．しかし，個々の音響シーンにおいて注目すべき音がそもそもどんな音であり，またその音がどの方向にあるのかは自明ではないため，前処理としてビームフォーマを適用することは難しいと考えられる．この問題に対してこれまで我々は，注目すべき音を自動検出するニューラルビームフォーマを提案し，無響室環境でその有効性を示した．しかし，その後の検証により，残響が存在する環境においては適切な指向特性を安定して形成できないという問題が明らかとなった．本稿では，1. 生成するフィルタ係数を時間領域から周波数領域に変更する，2. 無響室のデータで事前学習したモデルを残響環境にドメイン適用する手法を提案し，残響下においても適切な指向性を持ったフィルタを生成できることを確認した．

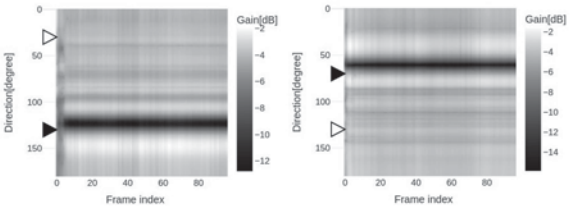


Fig.1：Two examples of directivity pattern obtained in a reverberant environment
▶: noise　▷: target

1-R-16

1-R-16　音響による接近車両検出における CNN モデルのパラメータに関する検討

A Study on Parameters on CNN model for Approaching Vehicle Detection Using Sound

☆伊藤隆佑，神谷珠緒，旭健作，坂野秀樹（名城大院）

◆出会い頭衝突事故の防止を目的に，道路上で収録した音を用いた接近車両検出を検討している．従来，入力データを中央の時刻で分割して入力する 2 入力 CNN モデルを提案した．しかし，モデルの構造は未検討で，特に全ての畳み込み層における kernel サイズを(3×3)に固定しており，他の kernel サイズについて検討していなかった．

◆本研究では，kernel サイズを変更し，従来より高い検出精度を目指す．また音データのメルスペクトログラムのメルバンド数も変更し，kernel サイズとメルバンド数が検出精度に及ぼす影響を調査する．

◆各メルバンド数・kernel サイズで評価した結果，メルバンド数 96，kernel サイズ「(9×9), (21×21)」の設定において，最接近時刻（0.0 秒）の 5.0 秒前から 3.0 秒前までの正答率が，従来より **10 ポイント以上向上**した．

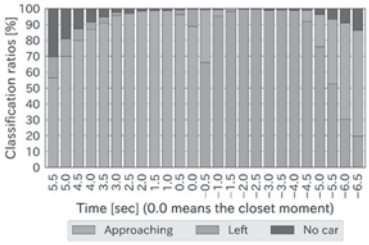


Fig.1：Classification ratios at each time in approaching and moving away situations

1-R-17

1-R-17　CNN を用いた音響による接近車両検出における複数マイクロホンデータの活用

Utilization of Multiple Microphone Data for Acoustic Approach Vehicle Detection Using CNN

☆神谷珠緒，伊藤隆佑，旭健作，坂野秀樹（名城大院）

◆我々の研究室では，CNN を用いて車両走行音から接近車両をリアルタイムに検出する検討を行っている．これまで，検出精度向上を目的とし，4 チャネルの音響データから**メルスペクトログラムとパワーの時系列**を CNN の入力とする手法を検討した．

◆しかし上述の手法は，音響データから得られる位相情報を考慮できていない．そこで本研究では，これまで使用してきた振幅情報に加え，**2 チャネルの音響データにおける位相差**を新たな特徴量として用いる手法により，先行研究からの検出精度向上を図る．

◆以上の方法により，モデル全体の判定精度は約 83.09%となり，振幅情報のみを 4 チャネル分学習した結果から**約 6.30pt** の精度向上を確認した．

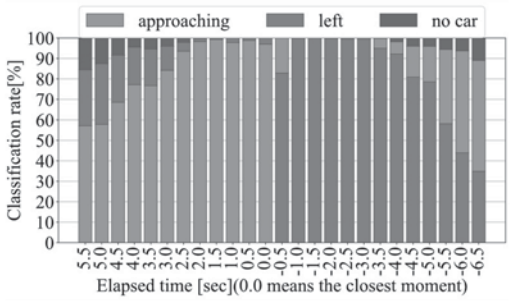


Fig.1：Classification rate at each time

1-R-18

1-R-18　蒸留を適用した MobileNet ベースの軽量なモデルによる異常音検知

Abnormal sound detection using a lightweight MobileNet-based model applying distillation

☆加藤洸人，立蔵洋介（静岡大）

◆研究目的

エッジコンピューティングのために，識別精度を高く維持した計算量の少ない軽量な異常音検知モデルの構築を目的とする．

◆提案手法

知識の蒸留により，モデルサイズの大きな教師モデル T の予測に MobileNetV2 ベースの軽量な生徒モデル S の予測を近づける．

◆実験

7 種のマシンタイプの正常な動作音を学習させた後，動作音の正常・異常を識別させる．蒸留をしない軽量なモデル S-nd を加えた 3 つのモデルで，Source と Target の 2 つのドメインデータに対する AUC (AUC-s, AUC-t)と pAUC での評価を比較して蒸留の効果を調査する．

◆結果

生徒モデル S は S-nd と比較すると，全マシンタイプに対するスコアの調和平均（All）におけるスコアを全て上回った．また，T と比較すると，モデルパラメータ数を約 85.76%削減しながら AUC-s は同等となった．

Table 1: Anomaly detection results.

Machine	Method	T	S	S-nd
All	AUC-s	60.15%	60.83%	56.15%
	AUC-t	62.90%	57.76%	54.60%
	pAUC	59.51%	57.39%	56.88%

1-R-19

1-R-19　分類器と再構成器を用いた異常音検知の有効性評価

Evaluation of the effectiveness of anomalous sound detection using classifier and reconstructor.

☆松本啓汰，山田武志（筑波大）

◆正常音データのみを用いる教師なし学習によって，未知の異常音を検出する技術が求められている．
- 正常音データの再構成に基づく手法
- 正常・異常の二値分類器と潜在空間上の距離学習を組み合わせた手法 [Kuroyanagi+ 2021]

◆これまでに我々は上記の２つの手法を統合・拡張した手法を提案した．
- 多クラス分類器（距離学習）と再構成器で構成される．
- 異常スコア：$s = \alpha(1-p) + \beta d' + \gamma r'$

p：正常確率，d'：正常クラスとの距離，r'：再構成誤差

◆異常音検知のタスクである DCASE2020 Task2 の追加データを含めて実験を行い，提案手法の有効性を確認した．

Table 1：Comparison of proposed and conventional methods (%)

machine	DCASE2020 Task2 Rank1[Giri+ 2020]		proposed	
	AUC	pAUC	AUC	pAUC
valve	96.13	90.89	**96.73**	**90.92**
slider	**97.63**	**89.73**	97.62	87.91
pump	**93.65**	81.73	90.31	**82.15**
fan	94.54	84.30	**96.13**	**88.76**
ToyCar	**94.34**	89.73	94.01	**91.98**
ToyConveyor	**91.19**	**73.34**	86.53	73.24
average	**94.58**	84.95	93.56	**85.83**

1-R-20

1-R-20　差分を考慮した自己教師あり学習を用いた音響表現の獲得

Difference-aware self-supervised learning for general purpose audio representation

◎竹内大起，安田昌弘，仁泉大輔，原田登(NTT)

◆自己教師あり学習による汎用音響信号表現の獲得
- ラベルのないデータのみから学習可能
- データ拡張による微小な変化に依存しない表現を学習
  - ✧ 大きな変化による表現の変動を考慮しない

◆本研究では、音の変動を考慮した自己教師あり学習を検討
- 音の混合を表現同士の差分で表現する学習手法を提案

◆実験より従来手法と比較して、音響分類タスクの精度を同等程度に保ちながら、音の違いに注目した音響信号検索タスクでの精度を改善

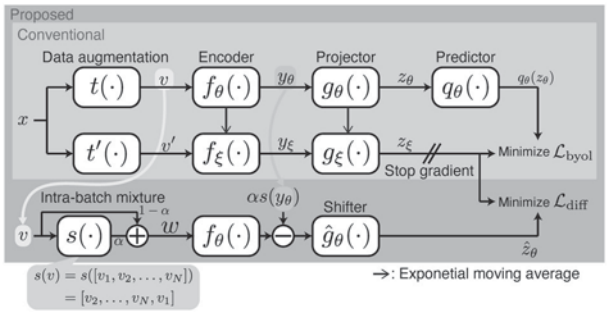


Fig.1 Illustration of proposed method

1-R-21

1-R-21　M2D for Speech: 汎用音響信号表現 M2D の音声への特化

M2D for Speech: Specializing General-Purpose Audio Representation to Speech

○仁泉大輔, 原田登, 竹内大起, 大石康智, 柏野邦夫 (NTT)

◆自己教師あり学習 M2D は多様な音のタスクで汎用的な有効性を示した一方、表現を分野に特化した最先端性能の実現が期待される。

◆本研究は M2D とその音声分野への拡張 M2D-S を用いて、競争の激しい分野に特化した表現の学習における可能性を検証した。

◆SUPERB を用いた実験では HuBERT 同等あるいは上回る最先端性能を示し、特化型表現の学習における M2D の有効性を確認した。

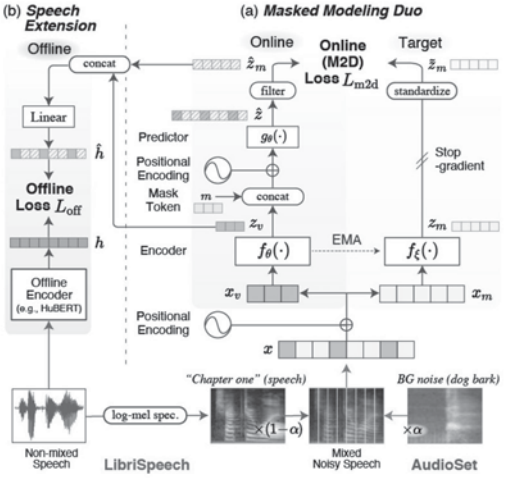


Fig. 1 M2D-S pre-training flow.

1-R-22

講演取消

1-R-23

1-R-23　音色情報と音高情報の抽出に基づく多楽器楽曲の音高推定に関する検討

Investigation for pitch estimation of multi-instrument music based on extraction of timbre information and pitch information

☆川凌司，坂野秀樹，旭健作（名城大・理工）

◆多楽器で演奏される楽曲の音高推定精度は単楽器と比べて低い
- ➢ 音色のパターン数が多いためパターンの学習が難しいことが要因

◆本研究では多楽器楽曲における音色情報を分離して音高推定を行うことで音高推定精度の向上を図る

◆提案モデルは入力スペクトログラム X から音色情報 $\hat{T}$ と音高推定結果 $\hat{Y}$ を推定し再構成するようなモデルである (Figure 1)

◆提案モデルは音の再構成を含む従来モデルより MusicNet におけるフレーム単位の音高推定精度が向上した (Table 1)

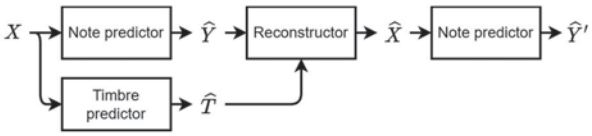


Figure 1: The overview of the proposed model.

Table 1: Frame-level transcription result for the MusicNet test set (1819, 2303, 2382).

Model	Prec. (%)	Recall (%)	F-score (%)	AP (%)
Gardner et al.	-	-	68.0	-
Cheuk et al.	-	-	66.3	71.1
Proposed	77.5	67.1	71.9	79.9

1-R-24

1-R-24　敵対的生成ネットワークを用いた狭帯域音声の帯域拡張に関する検討

Voice band enhancer of narrow band audio using generative adversarial networks

☆柴田 夏生(龍谷大院・理工学研)，片岡 章俊(龍谷大・先端理工)

◆画像処理や画像生成を目的に開発された生成モデルの一種である敵対的生成ネットワーク(GAN: Generative Adversarial Networks)をドメイン変換に向けて応用した pix2pix GAN を用いて狭帯域音声の帯域拡張手法を検討した．

◆原音声・狭帯域音声と生成器が生成した生成音声・狭帯域音声の２つの音声データの組を識別器が区別する処理を繰り返して学習させる．その後，学習済みの生成器に帯域が制限された音声を入力することで帯域拡張を行った．

◆精度向上のための追加処理として，狭帯域音声に対する全波整流処理で得られた高調波とホワイトノイズを高域部分に付加した．

◆訓練用のデータセットには単一性別話者のみと男女混合話者で学習するモデルを想定し，話者数を変化させた．

◆客観評価 PESQ では提案法の効果が見られなかったが，広帯域感に関するMOS 主観評価実験(評価値:1～5)では男女混合話者によって学習させたモデルにおいて，帯域拡張処理の有用性を確認できた(下表)．

Table：Results of subjective evaluation test

	M10	M20	M40	F10	F20	F40	M10F10	M20F20	M40F40
広帯域	4.60	4.79	4.71	4.69	4.90	4.67	4.88	4.79	4.89
全波整流	2.44	2.33	2.38	2.85	2.63	2.42	2.33	2.40	2.40
処理後	2.85	2.02	2.04	2.75	2.79	2.75	2.33	2.67	2.92

1-R-25

1-R-25　マルチチャネル参照信号予測における使用マイク数削減のためのグラフ探索手法適用に関する検討

Study on the Application of Graph Search Methods for Reducing the Number of Microphones Used in Multichannel Reference Signal Prediction

◎ 小塚詩穂里，伊藤弘章，鎌土記良(NTT)

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

◆ANC の構成要素の相対配置やアルゴリズム遅延，二次系の遅延などによる消音性能劣化を防ぐため，参照信号を予測することは有望

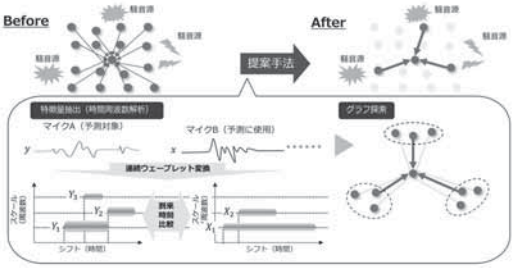


Fig.1：Overview of the proposed system

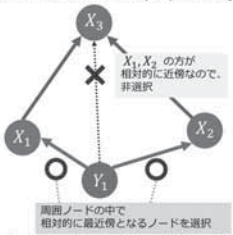


Fig.2：Procedure for the Selection Method of the Microphones to be Used

1-R-26

1-R-26　窓振動における振動誤差信号の計測位置の検討

Consideration of measurement position of vibration error signal for window vibration control.

☆中島尚哉，武藤憲司（芝工大）

◆本研究は窓振動低減システムのために，効果的な誤差検出振動ピックアップの配置を検討した。窓の振動の腹と節の中から４パターン選び出し，ANC を ON にした際の音圧レベル結果を比較した。

◆我々のこれまでの研究では，２つの ANC システムを用いた実験では，腹と節の間に置くことが有効だとわかっている．２つの振動ピックアップを用い，配置は ① - ④ とした．

◆最小低減量は- 4.25 dB が最も良い結果となったが，窓から５cm 離れた空間全体における平均値は，② - ④で大差はなかった．これより，誤差検出の位置は窓の中心から離れ，加振点に近すぎない③の位置が適しているとわかった．

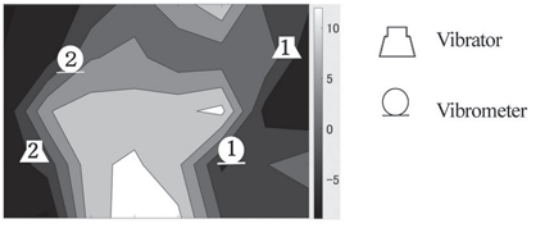


Fig.1：Distribution of attenuation level of sound pressure near window surface.(③)

1-R-27

1-R-27　複数のマイク間相関を用いたバーチャルマイクによる騒音制御の性能評価

Performance Evaluation of Noise Control using Virtual Microphones based on Cross-Correlation between adjacent microphones

☆吉松亨真, 信夫直樹(電通大), 伊藤弘章, 小塚詩穂里, 鎌土記良(NTT), 羽田陽一(電通大)

◆耳元近傍に等間隔に配置した複数のモニタリングマイクで観測した信号から各マイク間の相互相関関係を求め，この情報に基づいて耳位置の騒音信号を予測するバーチャルセンシング技術を提案してきている.

◆無響室内において頭部位置に何も配置せずに測定したインパルス応答，および疑似頭またはアクリル球を配置して測定したインパルス応答を用いて，本提案手法の能動騒音制御の性能評価を行う.

◆提案手法による騒音信号予測の有無やモニタリングマイクでの信号を用いた騒音低減性能の比較，疑似頭やアクリル球を用いた頭部からの反射による収束性能への影響を検証した結果について報告する.

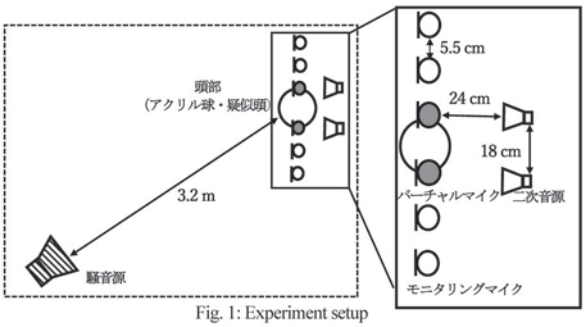


Fig. 1: Experiment setup

1-R-28

1-R-28　三つの非同期分散マイクロホンアレイによるビームフォーミングに基づく二音源の同時推定

Simultaneous localization of two sound sources based on beamforming using three asynchronous dis-tributed microphone arrays

☆上林大輝(阪産大), 奥屋夏生(阪産大), 中山雅人, 高橋徹(阪産大院)

●従来の対話活動の記録
相手の時間変化記録は不可
（発話時刻，発話内容のみ記録）

●提案：話者の発話方向も記録
津田らのシステム構想(Fig.1)
・発話対象（相手）の推定が可能に
・対話グループの推定も可能に
（発話時刻，音源定位，発話方向推定）

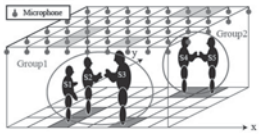


Fig.1：Monitoring system.

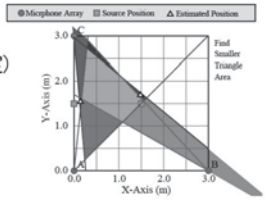


Fig.2：Sound source localization.

目的　津田らのシステム向けに矩形領域内の音源定位手法の開発
→二音源の同時推定へ対応

手段　音源到来方向を統合
・三直線の成す三角形の面積に基づく音源候補の選択(Fig. 2)

結果　二音源同時推定の平均推定誤差
MVDR：0.63 m　DSBF：0.90 m
DSBF から MVDR で改善(Fig. 3)

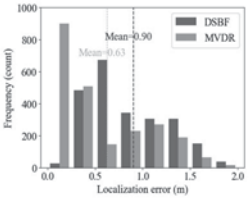


Fig.3：Histogram of estimation

1-R-29

講演取消

1-R-30

1-R-30　サーモホン搭載型超音波ヒトエコーロケーションロボットの開発 －時変 FIR フィルタによるリアルタイムオーディオピッチ変換の応用 －

Development of a thermophone-mounted ultrasonic human echolocation robot - Application of real-time audio pitch conversion using time-varying FIR filters -

☆細野太貴，△長谷川勘太，△布川貴康，山田恭史(広島大)，浅田隆昭(村田製作所)，△佐々木晋一(村田製作所)，蘆原郁(同志社大音響ナビゲーションセンター)，小林耕太，飛龍志津子(同志社大)，

◆開発した超音波ヒトエコーロケーションロボットを用いた方位と距離の知覚精度及びリアルタイムでの障害物実験を通じて，エコーロケーション能力，学習過程に向けての有用性の検討.

◆ロボットの特徴(Fig. 1)
①広帯域超音波を発信できる熱音響素子サーモホン
②頭部情報の保持が可能な人型（MDH）/ bat head
③時変 FIR フィルタによるピッチ変換
→超音波から可聴音にリアルタイムで変換可能

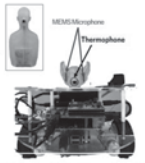


Fig.1 Robot with bat head and MDH

◆実験結果より，距離識別で直達音よりもエコーの音圧が高い種類の音の正答率が高く，音圧差で判別していることが示唆された

◆試行数を重ねるほど，障害物の有無や障害物との方向のずれを認識し，上達が確認できた(Fig.2). よって，本提案システムは，ヒトのエコーロケーションの効果的な訓練や学習過程の調査に有用と考えられる.

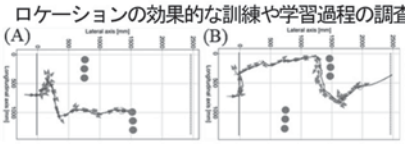


Fig. 2 Comparison of the running trajectories of the robot in the first(A) and tenth(B) trials.

1-R-31

1-R-31　音声認識における深層モデルに対するベイズ学習の適用

○田村雪菜，折本寿子，生田顕(県立広島大学)

◆ 音声認識の技術は近年多くの研究が行われており，認識率も飛躍的に向上している。

◆ 一般的な音声認識は，事前に雑音除去の処理を行い認識を行っており，雑音が多く存在する場合の音声認識は課題が残る。

◆ 音声認識技術として広く使われているDNNの基礎であるニューラルネットワークは，重みをかけて足し合わせて段階的にパラメータを推定していたため，過学習や学習不足といった問題がしばしば生じている。

◆ 本研究では，ベイズ推定により，確率的にパラメータを求めることで，過学習を防ぎ，雑音混入下でも，より精度の高い認識を実現する。

◆ ベイズ推定を行うため，近似推論の一つである期待値伝播法を学習モデルに適用した。期待値伝播法を用いることで，推定対象である値を分布で捉え，その期待値を伝播することで，精度のよい値を伝播させることができる。

Table 1 Recognition success rate

S/N	Word	Proposed method	Comparision method
1/1	あるとき	46.7%	80.0%
	北風	100.0%	93.3%
	太陽	100.0%	100.0%
	力比べ	100.0%	100.0%
	Average	**86.7%**	**93.3%**
1/2	あるとき	53.3%	46.7%
	北風	100.0%	100.0%
	太陽	100.0%	100.0%
	力比べ	13.3%	0.0%
	Average	**66.7%**	**61.7%**
1/3	あるとき	73.3%	46.7%
	北風	100.0%	60.0%
	太陽	93.3%	86.7%
	力比べ	86.7%	66.8%
	Average	**88.3%**	**78.3%**

1-R-32

1-R-32　事前学習不要な DNN 音声強調の反復処理に基づく性能改善

Iterative DNN-based speech enhancement without pre-training

☆中田優翔，大中緋慧，広政遼汰，宮崎亮一（徳山高専）

◆一般的な DNN 音声強調では収集コストの高いクリーン音声 (clean) が大量に必要であるが，Deep Audio Prior (DAP) を用いた手法 (DAP-SE) は雑音音声 (noisy) のみで音声強調が可能である．

◆本稿では DAP-SE の性能改善を目的として，DAP-SE の出力信号を次の DAP-SE の目的信号に置き換え，反復的に音声強調を行う手法を提案する (Fig.1)．

◆また，手法の改良案として複素スペクトログラムの位相修正 (Instantaneous Phase Correction: iPC) を導入する．

◆実験より，iPC を導入した反復手法 (Prop.) は DAP-SE と比較して全ての入力 SNR で各評価指標値の改善が確認できた (Fig. 2)．

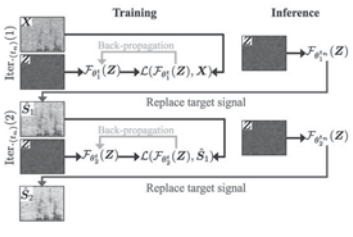


Fig.1 Overview of proposed method.

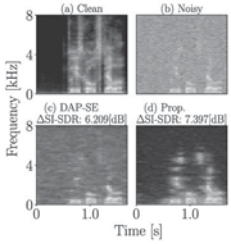


Fig.2 Spectrograms.

1-R-33

1-R-33 クリーン不要な DNN 音声雑音除去に用いる目的信号の分析

Target signal analysis in DNN-based noise reduction without clean signal

☆広政遼汰，大中緋慧，宮崎亮一（徳山高専）

◆一般的な音声雑音除去のための DNN は，クリーン音声 (clean) を目的信号として用いる Clean-target Training (CTT) と呼ばれる手法で学習しているが，clean は収集難易度が高いという問題がある．

◆clean 不要な学習手法として，雑音 (noise) が混入した音声 (noisy) を目的信号として用いる Noisy-target Training (NyTT) と呼ばれる学習手法が提案されている．

◆NyTT の分析も行われており，NyTT は noisy を表現し，追加の雑音 (add. noise) を除去する学習であることがわかっているが，まだ不明な点も多い．

◆本稿では，CTT と NyTT は学習時の目的信号が異なる点に着目し，コンピュータ上で生成できる noise や容易に収集できる noise など，様々な noise を目的信号として用いて分析を行った（Fig. 1）．

◆実験により，目的信号にバブルノイズを用いた場合に雑音除去できることがわかった（Table 1）．

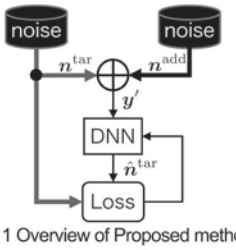


Fig. 1 Overview of Proposed method.

Table 1 Improvement in SI-SDR as a result of each dataset. DEMAND contains babble noise and others don't.

dataset ($\boldsymbol{n}$-$\boldsymbol{n}^{\mathrm{add}}$)	ΔSI-SDR [dB]
sin-TAU	−14.41
square-TAU	−28.84
white-TAU	−3.68
DEMAND-TAU	**6.67**
CHiME3-TAU	−5.75

1-R-34

1-R-34　アクティブサウンド自動珈琲焙煎システム -珈琲焙煎音の音響分析と音源分離-

Active Sound Automatic Coffee Roasting System -Acoustic analysis and sound source separation of coffee roasting sounds-

○中山仁史（広島市大院），△向井健悟（高松桜井高校）

◆珈琲焙煎音のみを得ることが可能な技術を用いて，焙煎音の音響モニタリングと，これを用いたフィードバック制御によるアクティブサウンド自動珈琲焙煎システムを提案する．

◆提案システムを実現するために，①SS-NMF による焙煎音と燃焼音の音源分離，②NMF による焙煎音の時間的変化の周波数解析を行った．

◆焙煎音は珈琲豆とロースタにより不規則かつ非定常なタイミングで発音する．そのため，発音タイミングと周波数スペクトルの行列表現が可能な NMF による音響モデリングを試みる．

◆SS-NMF において焙煎音＋燃焼音より教師信号として燃焼音を用いることで焙煎音のみの抽出が可能なことを確認した．（Fig.1）

◆フライパン焙煎，焙煎機及び自動焙煎機の３焙煎を対象に，分単位による珈琲焙煎の周波数スペクトルの変化を確認した．（Fig.2）

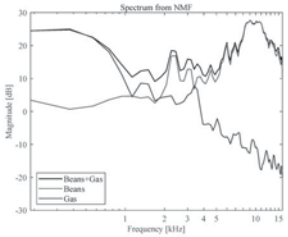


Fig.1: BSS from SS-NMF

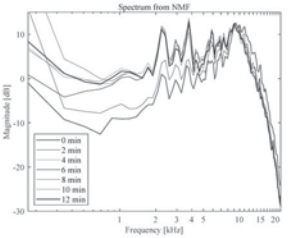


Fig.2: Spectrum of NMF during roasting

1-R-35

1-R-35　Exploring Feature Integration and Training Strategies for Enhanced Acoustic Scene Classification on SSAST

☆Shuting Hao, Nobuaki Minematsu, Daisuke Saito (UTokyo. , Fac. Eng.)

- This study addresses the task of Acoustic Scene Classification (ASC), which involves categorizing environmental audio streams into predefined semantic labels, as shown in Fig. 1.
- We introduce a novel architecture that incorporates multi-layer classifiers and direct fine-tuning, offering a fresh perspective in the field of ASC.
- The research utilizes the TAU Urban Acoustic Scenes 2022 Mobile dataset for fine-tuning and validation.
- We utilized the SSAST model, pre-trained on AudioSet and LibriSpeech datasets, and fine-tuned it on the TAU Urban Acoustic Scenes 2022 Mobile dataset with a novel approach to enhance feature learning specific to ASC.
- Our Layer wised SSAST system achieved a 52.03% accuracy and 88.17% AUC in ASC, showing an incremental yet noteworthy improvement over the baseline with absolute increases of 0.85% in accuracy and 0.36% in AUC.

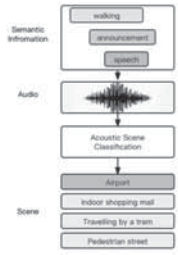


Fig.1: Overview of an ASC system.

1-R-36

1-R-36 振幅スペクトログラムの 2 次元フーリエ変換を用いた製品異音検査

Anomaly detection using 2D Fourier transform of amplitude spectrogram

○宗戸俊樹, 笠原太郎, △坂本湧暉, 伊藤秀浩(アイシン),
西浦敬信(立命館大/名産研), 耿毓庭(立命館大)

◆ 自動車部品の製造現場では熟練の官能検査員による異音の合否判定が行われているが，省人化・自動化による生産性の向上が求められており，本稿では，周期的な突発音を検出する方法を提案する.

◆ 従来は調波打楽器分離(Harmonic/percussive source separation)を用いた正常成分と異常成分の分離を行ったが，2 次元離散フーリエ変換(2 dimensional discrete Fourier transform: 2D-DFT)を用いた分離(Fig.1)による異音検出を検討する.

◆ 分離した異常成分の二乗平均平方根(Root mean square)の値を異常度とし，AUC(Area Under the Curve)による分類性能を比較する. 従来手法(AUC: 0.89)と比較すると提案手法(AUC: 0.97)の結果となり，精度向上を確認した

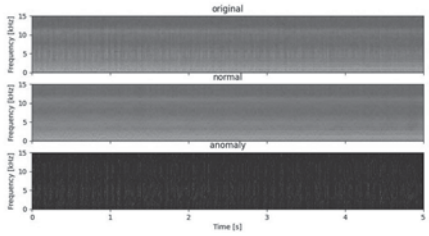


Fig.1 : Separation results with proposed method using 2D-DFT on amplitude spectrogram (top : original, middle: normal components, bottom: anomaly components)

1-R-37

1-R-37　1チャネルノイズ抑圧のための変分ベイズEMによるノイズ推定

Noise estimation by variational Bayes EM for single-channel noise reduction

○江村 暁
京都先端科学大学

本報告は「音声・ノイズの確率モデルと最尤推定のみから、ノイズ推定抑圧法を解析的に導出できるか」という問いに対し、変分EMアルゴリズムを適用するアプローチを提案する。このアプロ ーチにより、ノイズ推定はEMアルゴリズムよりもロバストになり、ノイズ推定およびノイズ抑圧後音声品質の点でstate-of-artのノイズ抑圧手法と同等の性能が得られることがシミュレーションから分かった。

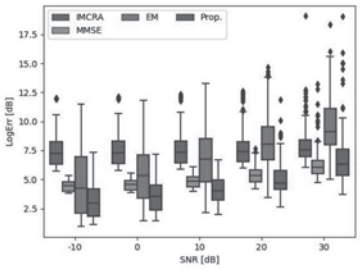


Fig. 1 各SNRにおけるノイズ推定結果のLog error: IMCRA, MMSEベース法、EMアルゴリズム、提案法

1-R-38

1-R-38　残響環境下における事前学習モデルに基づいた音声認識による音声了解度推定方法の検討

Estimating speech intelligibility by speech recognition based on pre-trained model in a reverberant environment.

☆桜井健人(山形大・工), 服部真稀, 近藤和弘(山形大院・理工学研)

◆残響劣化した音声に対する日本語了解度推定

◆wav2vec 2.0 の事前学習モデルに対して，JDRT を目的として発話内容と残響劣化環境をファインチューニング

◆モデルに未知話者かつ未知残響環境の音声を入力し推定了解度算出

◆JDRT 全単語平均における試験では，女性話者において主観了解度と推定了解度の間に高い相関

◆子音属性別では 6 属性の内，4 属性で主観値と似た傾向も，他 2 属性では残響時間の増加に伴い精度が劣化

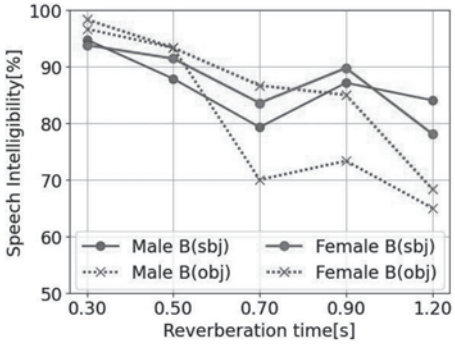


Fig.1 Subjective(sbj) and Objective(obj) intelligibility measured by JDRT on the distance is d0

1-R-39

1-R-39
スパース行列復元に基づく ブラインド残響除去におけるバイアス軽減の効果

Effect of bias reduction in blind dereverberation based on sparse matrix recovery

◎ 饒平名文希，山田宏樹，矢田部浩平（農工大）

目的：単チャネルの音響信号に含まれる残響を補助情報無しで抑圧する単チャネルブラインド残響除去

以前提案した手法：観測信号を行列リフティングによりモデル化することでスパース行列復元問題として定式化

課題：スパース正則化で用いる軟閾値関数のバイアスにより，残響成分だけでなく直接音成分まで削られる

► 硬閾値関数を近似したp-shrinkageを用いてバイアスを軽減した際の性能面における効果を検証

結果：残響時間(RT_{60})が短い信号でのバイアス軽減効果を確認

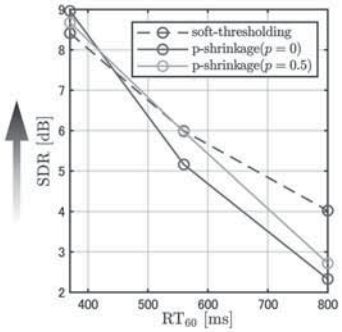


Fig. 1 Each SDR with soft-threshold function and p-shrinkage

1-R-40

1-R-40
時間周波数領域での厳密な畳み込みに用いる 変換行列の転置の導出と実装

Derivation and implementation of the transpose of the matrix used for rigorous computation of convolution in the time-frequency domain

☆松山直哉，山田宏樹，矢田部浩平（農工大）

背景

- 時間領域での畳み込みを離散 Fourier 変換した結果は，周波数領域で要素積になる
- 時間周波数領域での要素積は，時間領域での畳み込みの近似として用いられるが，場合によって近似精度が悪くなる
- 前回，時間周波数領域での厳密な畳み込みを導出し，時間領域のインパルス応答から時間周波数領域でのインパルス応答にあたる表現への順変換を提案した

課題

この変換の逆変換が実現できていない

提案

逆変換に必要な作用素を導出し，**効率的な計算法**を提案

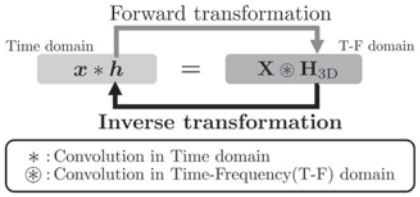


Fig. 1 Overview of the proposed method

1-R-41

1-R-41
コンプレッサー処理された信号の 復元に関する検討

Study on recovering signals processed by compressors

☆王檬，赤石夏輝（農工大），中村友彦（産総研），山田宏樹，矢田部浩平（農工大）

◆ 背景
- コンプレッサー処理された信号は圧縮されている
- 楽曲の再編集などで圧縮前の信号が必要となる場合あり
- コンプレッサーは非線形システム　→　**復元が難しい**

◆ 本研究
- 変分不等式に基づく復元アルゴリズムが存在
- コンプレッサーに適用　→　**元信号の復元を目指す**

◆ 実験
- コンプレッサーのパラメータを変えたときの復元精度を調査

◆ 結果
- 推定パラメータと真のパラメータが一致すれば**概ね復元が可能**
- 各パラメータの誤差の影響が異なる

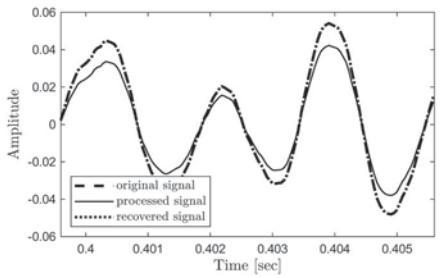


Fig.1: Comparison of original signal, compressed signal, and signal recovered by the proposed method under the best conditions.

1-R-42

1-R-42
制約集合を逐次改良する 再重み付きスパース audio declipping

Iterative refinement of a constraint set for reweighted sparse audio declipping

◎ 田中僚郎（早大理工），矢田部浩平（農工大），及川靖広（早大理工）

audio declipping ＝ 振幅が制限されて歪んだ信号の復元

□ 従来手法：振幅スペクトログラムのスパース性を誘導
△ **推定解の振幅を小さくしてしまう**性質

□ 提案手法
ℓ_1ノルムの重みを更新（従来も検討されてきた）
＋
制約集合も逐次改良
○ **振幅を積極的に大きくする**ことで振幅の減少を打ち消す

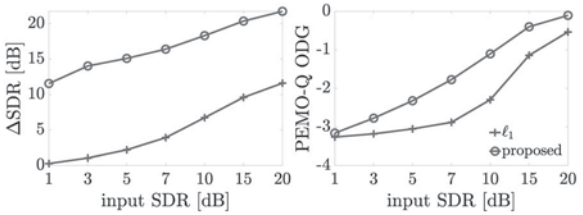


Fig. 1 Average of ΔSDR and PEMO-Q ODG vs. input SDR

□ 実験結果：単純な ℓ_1ノルム最小化に比べて、平均で 12 dB 高いΔSDR と 0.6 高い ODG を実現

□ 今後は重みの評価や作成方法について検討

1-R-43

1-R-43　周波数領域優決定音源分離の分離性能とスペクトログラム無矛盾性の度合いの関係の理論的考察と実験的評価

Theoretical Analysis and Experimental Evaluation of Relationship between Separation Performance and Spectrogram Consistency for Frequency-Domain Determined Source Separation

☆児島孝明(東大)，池下林太郎(NTT)，高宗典玄(東大)，北村大地(NTT)，猿渡洋(東大)，中谷智広(NTT)

◆本報告では，スペクトログラム無矛盾性を利用した独立低ランク行列分析 (ILRMA) である Consistent ILRMA と無矛盾正則化付き ILRMA (CR-ILRMA) の分離性能が ILRMA より向上する要因を分析する．

◆Consistent ILRMA と CR-ILRMA は観測信号のエネルギーの大きい周波数ビンのパーミュテーション正答率 (CorrPerm) を改善しており，**性能向上の要因はパーミュテーション推定の差異である**と示唆された．

◆**分離信号のスペクトログラムが無矛盾に近づきすぎると，分離性能は低下する**ことを示唆する予想が成り立つことを部分的に示した．

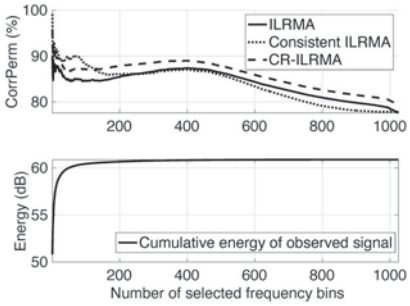


Fig.1：Correct permutation ratios (CorrPerm) and cumulative energy of observed signal as function of number of selected frequency bins.

1-R-44

1-R-44　IVA と DNN を近接平均化した優決定 BSS における DNN のリプシッツ定数に関する検討

Research on Lipschitz constants of DNNs proximal-averaged with IVA in determined BSS

◎ 松本和樹（早大），山田宏樹，矢田部浩平（農工大）

◆ 前回 **IVA と DNN の近接平均 (PA) を PDS アルゴリズムに Plug-and-Play(PnP)** した優決定 BSS フレームワーク (PA-BSS) を提案
► 双方の利点を両立し，**雑音除去 DNN を音源分離に活用できる**

◆ 今回は画像処理の PnP で重視される **DNN のリプシッツ定数**に着目
► **Spectral Normalization (SN)** に基づいた畳み込み層の正規化処理を PA-BSS に導入し，その利点と欠点を調査した

利点：分離行列の発散を防ぎ，**数値安定性に寄与する** (Fig. 1)
欠点：DNN の表現力低下を招き，**分離性能を損なう** (Fig. 2)

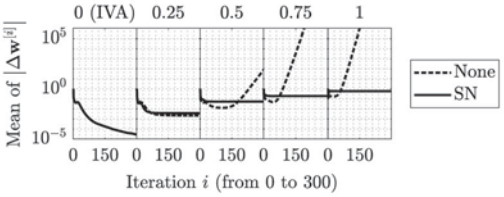


Fig. 1 Change in separation matrix at each iteration for each averaging parameter.

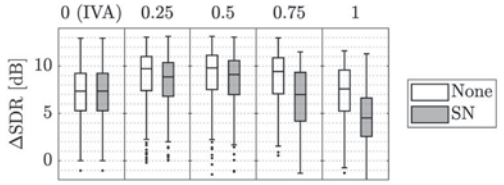


Fig. 2 Separation performance for each averaging parameter.

1-R-45

1-R-45　残響環境下におけるエイリアス成分の分離に有効なパーミュテーション解法の検討

Effective permutation solver for unaliasing based on blind source separation in reverberant environments

◎中村優希，宮崎亮一（徳山高専），小野順貴（都立大）

◆ローパスフィルタが適用できない状況において，エイリアシングが生じた観測信号に対してブラインド音源分離を適用することで，低域のエイリアス成分を分離し高域を復元するアプローチが提案されている．

◆エイリアス成分の分離には周波数領域独立成分分析が有効であるが，残響の影響を受けると，位相差モデルに基づくパーミュテーション解法ではパーミュテーション正解率が大幅に低下する．

◆本稿では，残響環境下でも有効にパーミュテーションを解決できる手法として，スペクトログラム連続性に基づく解法を提案する．

◆実験により，提案手法が残響環境下においても高い精度でパーミュテーションを解決でき (Table 1)，提案手法でパーミュテーションを解いた復元信号が，エイリアシングが生じた観測信号に対して SI-SDR や PESQ のスコアで上回ることが示された (Table 2)．

Table 1：Accuracy rate of permutations for each permutation solver (PS).

	PS	0 ms	100 ms	200 ms
speech	conv.	0.990	0.807	0.682
	prop.	**1.000**	**0.964**	**0.962**
music	conv.	0.984	0.809	0.693
	prop.	**0.986**	**0.935**	**0.934**

Table 2：SI-SDR and PESQ for each signal.

		SI-SDR			PESQ		
	signal	0 ms	100 ms	200 ms	0 ms	100 ms	200 ms
speech	Aliased observed signal	16.80	18.14	19.60	2.44	2.69	2.82
	Restored signal (conv.)	25.51	6.84	3.91	3.73	1.73	1.35
	Restored signal (prop.)	**35.27**	**24.83**	**20.56**	**4.41**	**3.67**	**3.05**
	Low-pass signal	19.72	21.04	22.49	3.76	3.95	4.00
	Restored signal (true)	35.27	25.19	20.83	4.41	3.59	3.03
music	Aliased observed signal	12.78	12.81	13.40	2.10	2.02	2.02
	Restored signal (conv.)	23.69	7.18	4.32	1.98	1.28	1.18
	Restored signal (prop.)	**29.77**	**18.26**	**17.44**	**2.86**	**2.32**	**2.20**
	Low-pass signal	15.67	15.68	16.26	2.35	2.19	2.18
	Restored signal (true)	29.87	23.36	21.39	2.85	2.35	2.24

1-R-46

1-R-46　相互相関と動的計画法を用いた欠損を含む信号のブラインド同期

Blind synchronization of signals with missing data using cross-correlation and dynamic programming

☆竹内駿人，河村隆生，植野夏樹，小野順貴(都立大)

◆目的
➢ 機器の故障や通信時のパケットロスなどにより，録音信号には欠損が生じる可能性がある．部分的な欠損がある非同期録音信号の同期手法を提案する．

◆概要
➢ 各録音信号を短時間フレーム分割し，フレーム間における相互相関関数のピークをコストとして，動的計画法を適用することで同期する．

◆実験
➢ 発話音声と，発話音声に欠損を生じさせた信号に異なるホワイトノイズを重畳した実験において，提案手法の有用性を確認した．

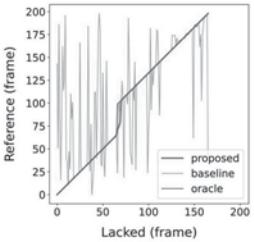


Fig.1：Alignment paths using the proposed method ("proposed") for a 3-second speech signal. The missing signal interval was from 1.0 to 1.5 seconds, with an input signal-to-noise ratio of 5 dB.