

2-1-1

2-1-1 歌声の声種判定における
深層学習と音響特徴量との比較Comparison of Singing Voice Type Determination
Using Deep Learning and Acoustic Features

○矢萩量子(東京理科大), 上原寛(東京藝大・扇精光)

- ◆複数の機械学習モデルを作成し、短い歌声(1.0s)での声種判定を実施
 - 女声 16 人(Soprano 10 人, Alto 6 人)、男声 15 人(Tenor 6 人, Bass 9 人)の声楽歌手が被験者として参加、1 分間の収録発声を収集
 - 音響特徴量(Formant tuning, SF 中心周波数)による声種分類に準拠
 - 4 種類の予測モデル(2 種類の kNN 法と NN, CNN)を作成
 - 音声学や歌唱音響学で用いられる音響特徴量計 255 種や音声サンプルによる学習を実施し性能を比較

◆結果:

- CNN 以外の Accuracy は 0.6 ~ 0.7 ほど、男声はある程度判定可能
- 女声は声帯基本周波数の影響を大きく受け、判定ができていない

ボイス種別	女声 (F0 平均: 42.6)				男声 (F0 平均: 66.5)				全体的平均
	入力音声	予測された声種	精度 (%)	標準偏差 (%)	入力音声	予測された声種	精度 (%)	標準偏差 (%)	
女声	ソプラノ	ソプラノ	89.9	2.7	7.4	ソプラノ	81.4	17.1	1.4
女声	ソプラノ	アルト	4.3	90.4	5.3	ソプラノ	11.2	82.4	6.4
女声	ソプラノ	バス	2.1	24.8	73.0	ソプラノ	1.4	30.5	68.1
男声	テナー	テナー	83.5	13.3	3.2	テナー	76.6	11.7	4.8
男声	テナー	バス	61.8	28.6	0.7	テナー	72.1	26.4	0.7
男声	テナー	ソプラノ	38.5	59.4	2.1	テナー	35.3	2.1	42.8
男声	バス	バス	5.7	0.7	11.3	バス	82.3	5.7	80.0

Fig. 1 Result of voice type determination by machine learning

- ◆まとめ: 男声の声種判定については機械学習で実施できることを確認
既存の判別特徴量である Singer's formant 中心周波数を用いた分析が十分に実施できない短い歌声でも、男声の声種判定が行えると示唆

2-1-3

2-1-3 金管楽器のロングトーン演奏を対象とした
自己評価と他者評価との比較Comparison of evaluations between self and others in long-tone
performance on brass instrument

☆田鎖佑弥, 小口純矢, 森勢将雅(明治大)

- ◆目標: 金管楽器演奏における演奏者の評価と第三者の評価を比較
 - 演奏後に自身の演奏をリッカート尺度で 7 段階にスコアづけ
 - 第三者は異なるスコア間でサーストンの一対比較法を実施

◆実験刺激について

- Go/NoGo 課題により演奏者が予測しないタイミングで収録
- 奏法はロングトーン演奏(F4)で統一
- 一対比較をするペアは 168 組

◆結果(Fig. 1)

- 尺度値範囲は 0.08 から 0.11
- 演奏者スコアが 1~3 のものは第三者からも低評価
- 検定から第三者と演奏者の評価は一致するとは言えない結果

◆今後の展望

- 第三者の音楽経験の有無で評価を比較
- 演奏者自身に対してサーストンの一対比較法を実施



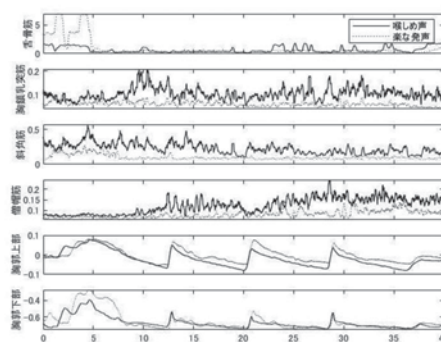
Fig.1: サーストンの一対比較法による主観評価結果

2-1-2

2-1-2 異なる発声方法における歌唱時の筋活動の
観察及び音響特性との関係Observation of muscle activity during singing in different vocalisation
methods and its relationship with acoustic characteristics.

☆宮川桃, 饗庭絵里子(電通大)

- ◆本研究は喉しめ声による歌唱聴取時の聴取者の筋活動を評価基準とする新たな歌唱評価指標の開発を目的とする。
- ◆プロ歌手 3 名に対して楽な発声での歌唱時と喉しめ声で歌唱時の筋活動及び呼吸活動の差異を計測した。
- ◆被験者によって筋活動の差異は異なり、誰しも喉しめ声によって同部位の筋活動が活性化するわけではないことが分かった。
- ◆一方で同一被験者についてはいずれの歌唱課題に対しても一貫して同様の筋活動の傾向が見られた。



喉しめ声と楽な発声での歌唱時の筋活動及び呼吸活動の差異

2-1-4

2-1-4 和音進行聴取時の期待からの逸脱
に対する印象の違いImpression differences of deviations from expectations
for chord progressions

○末崎真実(電通大), △高橋範行(愛知県立大), 森本智志(慶應義塾大), 饗庭絵里子(電通大)

- ◆多くの西洋音楽は、機能と和声のような規則によって構成されているため、次の和音がおおよそ予測可能となり、音楽的期待が生じる。また、音楽的期待から逸脱した音を聴取すると mismatch negativity (MMN) と呼ばれる脳波の振幅等に変化が現れることが報告されている。
- ◆しかしながら、音楽的期待には、典型的な和音進行から逸脱したことによる「違和感」だけでなく、予想していなかったような「よい意外性」をもつ進行も存在する。
- ◆本研究では、様々な和音進行刺激を用いて印象の異なる期待感について主観評価実験を行った。加えて、予備的ではあるが脳波機能計測を行い、主観評価との対応関係を検証した。
- ◆その結果、主観評価と脳波との間に相関がある可能性が示された。

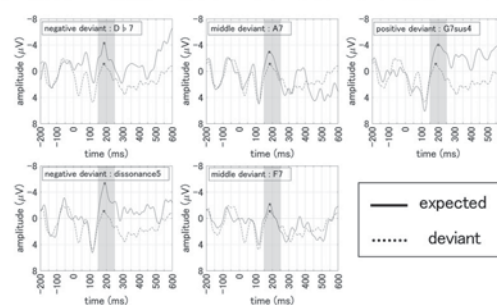


Fig.1: ERP Waveform for various chord progression

2-1-5

2-1-5 聴覚フィードバックが十分ではないとき
第2言語の発話が受ける影響

—雑音環境下における舌の位置を中心に—

Effects of insufficient auditory feedback
on second language utterances:

Focusing on tongue positions in noisy environment

○荒井隆行, 加藤明日香(上智大), 溝口愛(前橋工科大),
守本真帆(上智大・学振), 李瑋昱(上智大)

- ◆第1言語(L1)/第2言語(L2)の発話に対し、聴覚フィードバックの有無の影響を調査
- ◆雑音(パブルノイズ)の提示によって、聴覚フィードバックを制御
- ◆超音波診断装置を用いて舌形状を観察



- ◆英語初級者1名に対して、L2よりもL1において舌形状にばらつき(Fig. 1)
- ◆英語の習熟度があまり高くない場合に、聴覚フィードバックが発話にそれほど関与していない可能性を示唆
- ◆英語の習熟度が上がると、むしろ聴覚フィードバックが発話に強く影響してくることが考えられる

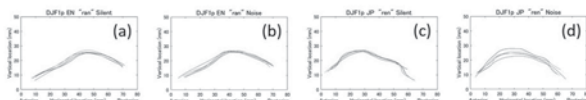


Fig. 1: Traces of the tongue shape for 5 repetitions.

2-1-7

2-1-7 残響の時間微細構造の違いが
室内での「話しやすさ」に与える影響の調査A Study of the Effect of Different Time Microstructures of Reverberation on
"Comfortability of Speaking" in Rooms.©松延皓大(日本音響エンジニアリング/九州大・芸工), 河原一彦(九州大・芸工),
大山宏, △根木健太, 山下晃一, 大橋心耳(日本音響エンジニアリング)

- ◆適切な音環境の確保には、室内の「話しやすさ」の調査も重要だが、音楽鑑賞や演奏、音声聴取に比べると十分な調査は行われていない。
- ◆「話しやすさ」に関して、残響時間やC値に関する先行研究や遅延時間の大きい反射音に関する先行研究はあるが、より快適な環境実現には、残響波形の時間微細構造や拡散性に着目した調査も必要である。
- ◆本研究では、被験者の声をマイクで収録しリアルタイムで残響音を付加した刺激を返すシステムを用いて、被験者に短い文章を読ませて「話しやすさ」を評価させる印象評価実験を行い、残響音特性の違いによる「話しやすさ」への影響を調査した。なお、この際、残響音のレベルや種類の違いに関する実験と、拡散性に関する実験を行った。
- ◆1つ目の実験より、適切なC値を持つ残響を、適切なエネルギーレベルで付加した方が話しやすいと評価された。また、その条件では残響音を付加しない場合よりも、付加した場合の方が話しやすいと評価された。
- ◆2つ目の実験より、散乱構造を持つ柱状拡散体の台数を増やすと、時系列ばらつき度の値は小さくなり、室内の拡散性が高くなっていることがわかった。また、柱状拡散体を配置しない条件に比べて、配置した条件の方が話しやすいと評価された。これより、拡散性の違いは「話しやすさ」に影響を与えることが示唆された。

2-1-6

2-1-6 雑音提示時の母音発話における聴覚
フィードバックが発話に与える影響の検討An investigation of the effect of auditory feedback on speech production
in vowel utterances during noise presentation

○上江洲安史, 赤木正人, 鶴木祐史(JAIST)

- ◆ヒトは音声コミュニケーションにおいて様々な感覚フィードバック情報に基づいて発話を制御している。雑音下での発話では、聴覚フィードバックした音声背景雑音でマスクされること避けるために、ロンパード効果などの聴覚フィードバックに基づいた発話運動制御が行なわれていると考えられる。
- ◆雑音下での単独話者の母音発話実験を通して、提示雑音のスペクトル形状や提示音圧レベルの違いが、発話音声の音圧レベルや母音フォルマントの振幅と周波数にどのような影響を与えるのかを検討した。
- ◆多くの雑音下で、発話音圧レベルの増加とともにフォルマント振幅の増加が見られ、雑音のレベルが高いほど第1フォルマント周波数のシフトは高域ヘシフトした。第2フォルマントは母音によって周波数シフトの方向が異なり、また母音によって雑音のスペクトル形状やレベルによる影響も異なることがわかった。

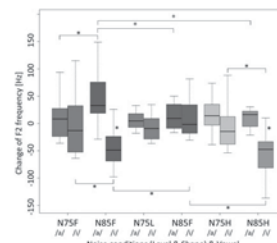


Fig. 1: Amount of F2 frequency change for each noise condition

2-1-8

2-1-8 発話と演奏における感覚フィードバック
メカニズムの同異Differences in sensory feedback mechanism between speaking
and playing musical instruments

○廣谷定男(NTT CS 研)

- ◆人間は、発話、歌唱や楽器演奏などで音を操ることにより、円滑なコミュニケーションを実現したり、心を豊かにしている。
- ◆このような音の生成過程では、脳でイメージした音が、調音器官、指や腕などの動きを通じて音として生成され(フィードフォワード制御)、聴覚、体性感覚(触覚も含む)あるいは視覚などの感覚フィードバックによってリアルタイムにモニタリングが行われている。
- ◆これらの研究は、一般的には auditory-motor interactions という分野に包含されるが、これまで、調音器官を制御する発話と、指や腕を制御する楽器演奏は別々に研究が行われてきた。
- ◆しかし近年、発話と楽器演奏における感覚フィードバックメカニズムの共通点を調べる研究が行われている。
- ◆本発表では、発話と楽器演奏における感覚フィードバックメカニズムの研究について概説し、近年のこれらの同異について紹介する。

	発話	楽器演奏
聴覚	○	○
体性感覚	○	○
視覚	×	○

Table.1: Sensory feedback used in speaking and playing musical instruments.

2-1-9

2-1-9 音楽家の卓越した技能を支える
感覚運動統合機能

Sensorimotor functions subserving musicians' expertise

○古屋晋一(ソニーCSL, NeuroPiano, 上智大学)

音楽家は聴覚情報と運動情報を統合する特殊な神経機構の柔軟性と頑健性を巧みに利用し、様々な環境で高精度な演奏を奏でる。一方、コンサートやコンクールといった舞台上での演奏における心理不安や、過剰な練習に伴い発症する局所性ジストニアをはじめとする脳神経疾患により、パフォーマンスの失調が見られるが、その背後には聴覚や体性感覚と運動を統合する働きの異常が知られている。本講演は、可変感覚フィードバックシステムを用いた心理物理実験や、経頭蓋時期刺激をはじめとする非侵襲脳刺激および脳波を用いた神経生理実験、ロボティクス技術や機械学習を活用して明らかになった音楽家の卓越した技能を支える感覚運動統合機能の働きについて紹介する。

2-1-10

2-1-10 管楽器およびソプラノ歌唱における声道制御と
音響フィードバック

On the vocal-tract control and acoustic feedback in wind instruments and soprano singing

○鎌木時彦(九州大)

- 音声と管楽器の音響メカニズムは非常に似ており、いずれも肺からの呼吸をエネルギー源として声帯やリードなどの弾性体の準周期的な振動が生じ、呼吸の流路を開閉することによって、音源波である体積速度波形が生じる。この音源波が声道や楽器本体を伝搬し、開口部から音が放射される。
- 音声生成の基本モデルは「音源-フィルタ理論」だが、管楽器は管体や声道の音響特性が音源機構に影響する「音源-フィルタ相互作用」である。Benade(1985)は、管体と奏者の声道の入力インピーダンスが直列に接続された状態で、リードに音響負荷がかかることを示した。本研究では、プロのトランペット奏者とサクソフォン奏者について、楽器演奏時の声道形状をMRIにより3次元測定し、声道の入力インピーダンスを求めた。その結果から、演奏モードの切り替えが声道調節によってなされることを示した。
- 一方、ソプラノ歌唱においては、フォルマントチューニングが生じることが知られているが、音源-フィルタ相互作用の観点からは、発声の基本周波数がフォルマントに近づく、発声が不安定になってしまう可能性がある。本研究では、アマチュアソプラノ歌手1名について同様にMRI測定を行い、声道の入力インピーダンスを求めた。A3、E4の音高で母音/i/を歌唱発声した時に比べて、D5、A5では舌が後方に移動し、明確な違いが見られた。このとき、声道の入力インピーダンスの第1ピークの周波数は上昇、第2、第3ピークは下降し、さらに第1ピークの振幅が音高とともに大きくなることがわかった。

2-1-11

2-1-11 モーションキャプチャを用いた楽器演奏に
おける感情表現と動作の関連について*

*Relationship between emotional expression and motion by playing musical instruments using optical motion capture system.

○三戸勇気(日大・芸術)、松崎桃子、山口達也(日大・理工)、

駒澤大介、川上央(日大・芸術)、三浦雅展(国立音大)、篠田之孝(日大・理工)

我々は、研鑽過程の音大生にも、わかりやすくレクチャーできる手法を見出すことができなかつたかと考えた。そこで、モーションキャプチャシステムを用いて実際の演奏動作のデータを取得し、細かな差異を具体的かつ正確に音大生に教示するための教育支援システムを作成することとした。

本稿では、鍵盤楽器(グランドピアノ、電子ピアノ、チェンバロ)の動作解析の結果と、その演奏動作のモーションキャプチャデータに対するアンケートをピアノの教員に行い、レッスンで教育支援システムを用いるために必要な情報について回答してもらった。今回は、そのアンケート結果も報告する。

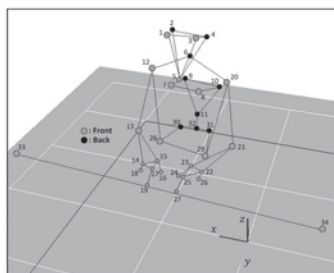


Fig. 1 Marker adhesion position

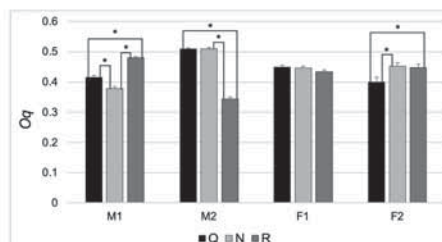
2-1-12

2-1-12 雑音下と残響下での発話における
EGG 解析による声帯振動様式の比較

Comparison of vocal fold vibration patterns of speech under noise and reverberation in EGG analyses

☆小松悠人、程島奈緒(東海大)、榎原健一(北医療大)、水町光徳(九工大)

- ◆ ロンバード効果および残響下での発話は明瞭度に個人差がある
- ◆ 明瞭度の高い話者の特徴を検討するため、ロンバード効果および残響下での発話によって変化する声帯振動様式と発声努力の関連性を男性話者2名(M1, M2)、女性話者2名(F1, F2)から調査
- ◆ EGG から発声努力の緊張-弛緩を示す声門開放時間率(O_q)を求め、静音下(Q)・雑音下(N)・残響下(R)で O_q の変化を話者毎に比較(Fig.1)
- ☆M1: Q>N, R>Q ☆M2: Q>R ☆F2: N>Q, R>Q
- ⇒ 静音下と比べ雑音下・残響下ではロンバード効果および残響下の発話により発声努力が変化しと考えられる
- ◆ すべての発話環境間で話者によって声帯の緊張度が異なった
- ⇒ 静音下での発話を基準にした雑音/残響下での発話における基本周波数の変化には個人性があるように、発声努力にも個人性がある可能性

Fig.1: O_q for each speaking condition and speaker (* $p<0.01$)

2-1-13

2-1-13 三線音における演奏者と鑑賞者の位置による音色の違いの評価

Evaluation of the difference on timbre in *sanshin* sound depending on the position of the performer and the audience

☆山田大裕, 西宮康治朗, △長秀雄(青山学院大学)

- ◆背景: 沖縄伝統の弦楽器である三線の職人や演奏家は、自身で弾いて聞いた音をフィードバックしながら三線を製作したり、演奏の制御を行う。しかし、楽器は聞く位置により音色が変化することが知られている。
- ◆目的: 弦楽器において胴は音色を決定する重要な要素であることから、胴の共振特性に注目し、胴の音と C3(130 Hz)~C4(260 Hz)の演奏者方向、鑑賞者方向における音色(倍音分布)との関連性を評価した。
- ◆結果1: 胴の共振周波数である 520 Hz と 816 Hz に着目した結果、胴の打撃試験による膜からの放射音の指向性と、これらの周波数に一致している弾弦時の倍音の指向性が同じ傾向になった(Fig. 1)。
- ◆結果2: 演奏者方向、鑑賞者方向で C3~C4 の音を弾弦した結果(Fig. 2)、演奏者は主に 1000 Hz 付近の音が大きく聞こえているのに対し、鑑賞者は 500 Hz 付近の音が主に聞こえていることが分かった。500 Hz 付近は胴の共振特性により、鑑賞者方向に振幅が大きくなっていると考えられる。以上より、演奏者は鑑賞者よりも高い音が聞こえていることが分かった。

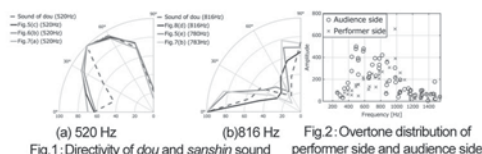


Fig. 1: Directivity of dou and sanshin sound

2-1-15

2-1-15 弦の物理特性がヴァイオリンの音色に与える影響

Effect of physical characteristics of strings on the sound of the violin

☆掛川慶人, 朝倉巧(東京理科大)

- ◆弦楽器の弦には様々な種類が存在し、素材や物性値が異なる弦であれば擦弦時の弦の振動が異なり、楽器音に影響を与える。
- ◆本研究では、弦を擦弦した際の楽器音を周波数解析し、各弦の素材や物性値の差異による倍音構造への影響、および倍音構造と主観的印象の関係性について検討した。
- ◆楽器音の1倍音と強い相関関係のある項目は確認されなかったが、2, 3倍音に関して、エネルギー割合と相関性のある主観的印象の項目があることが確認された。

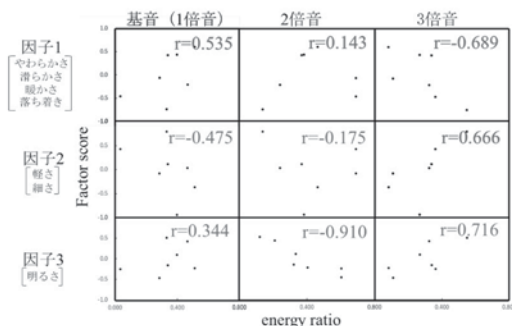


Fig. 1 Correlation coefficients between Factor and Overtone

2-1-14

2-1-14 擦弦情報の記録と奏者への提示を目的とする弓圧及び摩擦力の計測

Measurement of bow pressure and friction force for the purpose of recording bowing information and presenting it to the player

◎猪崎遼大(筑波大・院・シス工), 若槻尚斗, 水谷孝一, 善甫啓一, 海老原格(筑波大・シス系)

- ◆楽器練習において、自身の状態のフィードバックは主に指導者のご指摘や録音・録画の利用であり、楽器に与えたパラメータを提示するものは少ない。
- ◆擦弦楽器は、弓を用いるため、複数のパラメータを同時に操作する必要があり、演奏が難しい。
- ◆楽器本体に搭載したカメラモジュールにより、擦弦情報を記録、及び提示を行う。弦及び弓に非接触のため、演奏動作への影響が少ないことが期待される。
- ◆本稿では、ウルフトーン発生時における弓圧及びそのフィードバックである摩擦力の計測を行い、2つの関係性を確認した。

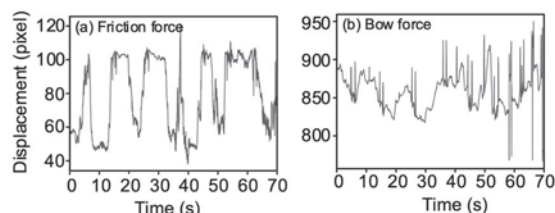


図1 ウルフトーン発生時における弓圧

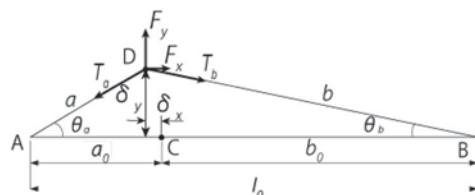
2-1-16

2-1-16 新型クラヴィコードにおける撞弦方法に関する検討

Examination of string striking method in the new Clavichord

☆田中稚洋, △鴨秀樹, △緑川結也, 西口磯春(神工大)

- ◆報告者らのグループではクラヴィコードの発音原理を用いた電気鍵盤楽器の開発を行なっている。
- ◆個々のキーのピッチ(音高)を独立して変化できることを特長としており、キーの押し込み量とピッチ変化の関係、横剛性を設計パラメータとする設計法を採用しており、これまでに、この楽器を使用した音楽CDも発売されている。
- ◆本報告では、演奏家の要望に基づき、押弦時のピッチ変化を更に大きくするための撞弦方法(弦とキーの配置および撞弦位置)の検討を行った。
- ◆検討の結果、弦とキーを同一直線上に配置し、タンジェントと弦の接触点の両側の張力が常に一致する撞弦方法を提案した。



A string deformed by the action of the tangent

2-1-17

2-1-17 三線における弦の歌口での支持条件が弦振動に与える影響の評価

Evaluation of the effect of string support conditions on string vibration in Sanshin.

☆並木達郎, 西宮康治朗, △長秀雄(青山学院大)

- ◆沖縄の民族楽器である三線は、棹が音色に最も影響を与えると言われており、様々な型の棹が存在する。棹の剛性が音色に与える影響を検討するために、弦と棹の接点である歌口での弾性支持条件と弦振動の関係性を評価する。歌口での弾性支持条件はばね定数を用いて評価した。
- ◆三線棹の剛性を調べるため、棹のばね定数を測定すると 5.23 N/mm であった。他の弦楽器の棹の剛性を比較するため、ギターについても同様にネックのばね定数の測定を行うと 13.0 N/mm であった。
- ◆ばね定数が $k=3.72, 8.82, 17.2$ N/mm である3種類の薄板ばねで弦の片端を支持した実験装置で弦振動を測定し、時間波形のスペクトルの第 11 次倍音の周波数からインハーモニシティを求めた結果を Fig.1 に示す。ばね定数が 8.82 N/mm 以下ではインハーモニシティの変化が大きいが、ばね定数が 8.82 N/mm より大きい領域ではインハーモニシティの変化は僅かになっており、棹のばね定数は弦振動のインハーモニシティが大きく変化する領域に存在することがわかった。棹の違いでインハーモニシティも大きく変わることが予想され、それが三線の棹が音色に影響を与える要因の一つになっていると考えられる。

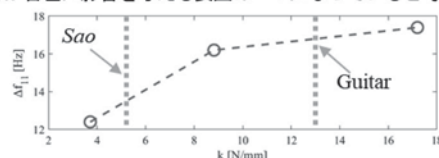


Fig.1: 11th inharmonicity of experiment with flat spring.

2-2-1

2-2-1 End-to-end Automatic Speech Recognition with Independent Vector Analysis Frontend

©Robin Scheibler (LINE), Wangyou Zhang (SJTU), Xuankai Chang (CMU), Shinji Watanabe (CMU), Yanmin Qian (SJTU)

- ◆We propose **end-to-end training** of ASR with an independent vector analysis (IVA) joint separation and dereverberation frontend.
- ◆We extend **time-decorrelation iterative source steering** (T-ISS) algorithm for IVA to the overdetermined case (channels > sources).
- ◆We demonstrate **robustness** to mismatch in training and testing data. For example, when training on clean data and testing on noisy data, T-ISS does not degrade significantly (see Table 1). On the contrary, a conventional neural beamformer (WPD) requires retraining on noisy data to maintain performance.
- ◆We further confirm by introduced a mismatched noisy dataset (TUT) for testing. Here as well T-ISS performs well without retraining.
- ◆Unlike neural beamforming, T-ISS can **separate more speakers** without retraining. We verify this on mixtures of 3 and 4 speakers.

Table 1 Results on in- and out-of-domain data.

Test set	Algo.	Train set	Domain	WER	SIR	SDR	PESQ	STOI
clean ($\mathcal{D}_1$)	WPD	$\mathcal{D}_1$	In	9.57	13.9	6.9	1.88	0.855
	T-ISS	$\mathcal{D}_1$	In	9.16	16.8	3.7	1.78	0.830
	WPD	$\mathcal{D}_2$	Out	17.12	12.3	8.7	1.70	0.890
	T-ISS	$\mathcal{D}_2$	Out	12.48	15.6	6.2	1.86	0.913
noisy ($\mathcal{D}_2$)	WPD	$\mathcal{D}_1 \cup \mathcal{D}_2$	In	11.40	14.7	10.8	1.79	0.918
	T-ISS	$\mathcal{D}_1 \cup \mathcal{D}_2$	In	11.80	14.4	7.4	1.78	0.924
TUT ($\mathcal{D}_3$)	WPD	$\mathcal{D}_1$	Out	31.36	6.3	2.7	1.41	0.744
	T-ISS	$\mathcal{D}_1$	Out	14.55	13.7	2.1	1.45	0.787
	WPD	$\mathcal{D}_2$	Out	15.17	10.0	5.2	1.57	0.816
	T-ISS	$\mathcal{D}_2$	Out	14.75	12.3	2.1	1.43	0.772

2-1-18

2-1-18 ベトナム民族楽器ダンバウの自動採譜に関する研究

A research on Automatic Scoring of Vietnamese Folk Instrument Dan Bau

☆竹内健次郎(八工大), 小坂谷壽一(八工大)

- ◆民族楽器の多くが口伝により奏法が受け継がれ、譜面の無い曲や若年層の担い手不足と様々な問題が挙げられる。
- ◆従来の研究では伝統音楽の(津軽三味線)の保存を目的とした自動採譜を行い、その採譜精度は95%以上に達している。
- ◆ベトナム民謡の採譜を行い、津軽三味線用自動採譜装置の他の楽器への応用展開として検証と解析を行った。
- ◆津軽三味線用自動採譜装置のパラメータを変更し、ダンバウ専用疑似自動採譜装置とし、採譜の際はダンバウシミュレータで録音を行った。
- ◆ベトナム民謡である「美しい竹 (Cây Trúe Xanh)」を含めた5曲の採譜を行い、採譜精度 90%以上に達し、津軽三味線用自動採譜装置は他の楽器への応用が可能であることが判明した。

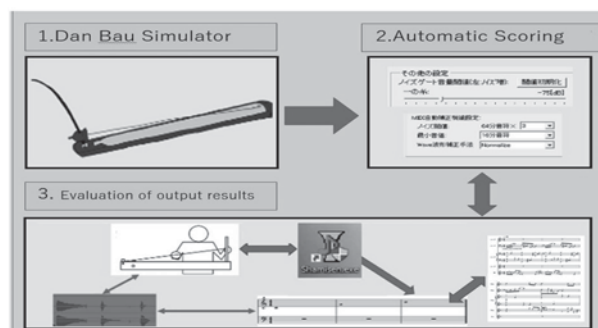


Fig.1: Device configuration

2-2-2

2-2-2 FastMVAE 法におけるブロックパーミュテーションを軽減する音源モデル学習

Source model learning to reduce block permutations in the FastMVAE method

©近藤祐斗(NTT), 李莉(NTT, 名大), 関翔悟(NTT), 亀岡弘和(NTT)

- ◆ブロックパーミュテーション (BP) 整合とは、時間周波数領域音源分離手法において同一音源に由来する分離信号が広帯域周波数ブロック単位で異なる分離信号と入れ替わる BP を軽減することである。
- ◆従来の BP 整合手法である Hungarian block permutation 法では BP を軽減するように補完する帯域をランダムに選ぶため、分離信号に生じている BP が必ずしも軽減されるとは限らない。
- ◆本研究では3つの新たなアプローチに基づく FastMVAE2 法に対する BP 整合手法の検討を行う。
- ◆実験において、音源数が 9 の場合に全ての提案法が従来の FastMVAE2 法より高い分離性能を示すことと、音源数が 3 の場合に全ての提案法が BP を解消することを確認した。

Table 1: SDR improvements [dB] for each method.

Methods	3	6	9
Conventional FastMVAE2	31.1	19.6	18.9
FastMVAE2 with ChineseACVAE solving BP automatically	29.3	19.2	21.3
FastMVAE2 with BP frequency band estimator	31.8	22.2	18.5
FastMVAE2 with discontinuous frequency detector	32.9	26.6	26.1

Fig. 1: Example of reference and separated signals obtained by each method. The number of sources is set to 3.

2-2-3

2-2-3 混合 Differentiable Digital Signal Processing モデルによる合成パラメータ抽出のためのラウドネスの時間変動に基づくロス関数の設計

Loss Function Design Based on Time Variability of Loudness for Synthesis Parameter Extraction Using Differentiable Digital Signal Processing Mixture Model

☆川村 真也, 中村 友彦, 高宗 典玄(東大), 北村 大地(香川高専), 猿渡 洋(東大), 高橋 祐, 近藤 多伸(ヤマハ)

- ◆我々は differentiable digital signal processing に基づく混合楽器音の生成モデル(混合 DDSP モデル)を構築し, 混合楽器音から直接各音源の基本周波数, 音色特徴量, ラウドネスを抽出する方法を提案してきた。
- ◆しかし, この手法では抽出したラウドネスが過剰に時間変動する場合があった。
- ◆本稿では従来のロス関数が実演奏から乖離したラウドネスを許容しうることに着目し, ラウドネスの時間変動に関する統計量を用いたロス関数を提案する。
- ◆基本周波数, 音色特徴量, ラウドネスの抽出実験から, 提案するロス関数を用いることでラウドネスの抽出精度が向上することを確認した。

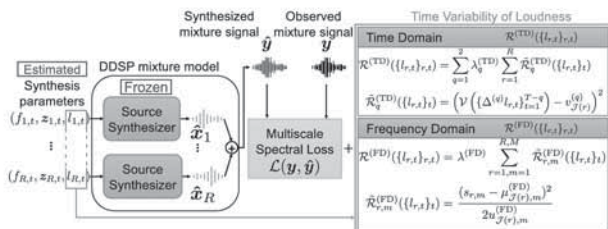


Fig.1: Overview of synthesis parameter extraction using DDSP mixture model with proposed loss function.

2-2-5

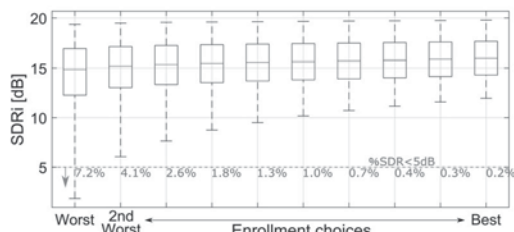
2-2-5 事前登録発話の選択に頑健な目的話者抽出の学習

Investigation into the Robustness of Target Speech Extraction to Enrollment Variations

◎佐藤宏, 落合翼, デルクロアマーク, 木下慶介, 森谷崇史, 牧島直輝, 庵愛, 田中智大, 増村亮 (NTT)

◆概要

- ・事前登録発話を用いて目的話者を, 複数話者の混合音声から抽出する目的話者抽出手法には, 事前登録発話の選び方によって性能にばらつきが生じる課題があった
- ・予備実験から確かに事前登録発話の選択は, 抽出の失敗率に大きな影響を与えることが分かった



- ・事前登録発話の選択に対する頑健性を向上させるために, 困難な事前登録発話を重視して学習する WE-SDR 学習を提案した

◆評価実験

- ・CSJ の音声データと CHiME4 の雑音を用いて生成した 2 話者音声と雑音の模擬混合音で検証
- ・提案法は, 最も困難な事前登録発話の集合に対して SDRi を 0.9 dB 改善し, 抽出の失敗率を 34 % 低減した

2-2-4

2-2-4 サンプリング周波数非依存畳み込み層のための時間領域ニューラルアナログフィルタ

Time-Domain Neural Analog Filter for Sampling-Frequency-Independent Convolutional Layer

☆今村 奏海, 中村 友彦(東大), 矢田部 浩平(農工大), 猿渡 洋(東大)

◆背景: 汎用的な音源分離の実現

- サンプリング周波数非依存 (SFI) 畳み込み層: アナログフィルタからのフィルタ設定技法を援用し, 未学習のサンプリング周波数でも頑健に動作
- ニューラルアナログフィルタ (NAF): アナログフィルタを深層ニューラルネットワークで表し, フィルタ形状も学習から決定

◆提案法: NAF を拡張した時間領域 NAF (TD-NAF)

- 時間領域フィルタ設計によるエイリアシングの低減のため, オーバーサンプリングを用いたアンチエイリアシング手法を導入
- 低サンプリング周波数での分離性能が向上

Table 1: Position of this study in SFI convolutional layer.

フィルタ設計	アナログフィルタ	
	解析的アナログフィルタ	ニューラルアナログフィルタ
周波数領域	音学シンボ 2021	音響学会 2022 春
時間領域	EUSIPCO2021	本研究

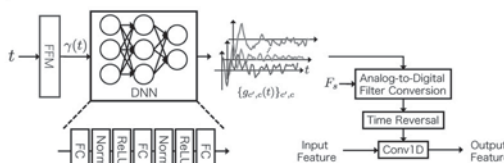


Fig.1: Architecture of SFI convolutional layer using proposed TD-NAF.

2-2-6

2-2-6 Direction-aware target speaker extraction with conditional variational autoencoders and its sensitivity to direction-of-arrival error

☆Rui Wang, Li Li, Tomoki Toda (Nagoya University)

- ◆Aim: To propose a target speaker extraction (TSE) method under underdetermined conditions.

- ◆Problems: It is hard to achieve TSE under underdetermined cases with a dual-channel system.

- ◆Solution: An iterative framework that combines the geometric constraints based on the direction of arrival (DOA) and more powerful source modeling based on conditionals variational autoencoders (CVAEs), and a nonlinear postprocessing.

- ◆Evaluation: Spatial test mixture signals are generated by using the image source method (ISM) with the database of the Wall Street Journal (WSJ0) corpus.

- ◆Summary: The proposed method exceeded all the baseline methods. But the robustness against the DOA error needs to be improved in the future.

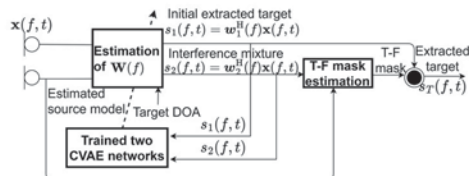


Fig.1 Scheme of the proposed framework

2-2-7

2-2-7 DNN 音声強調における Noisy-target Training の挙動分析

Behavioral analysis of Noisy-target Training for
DNN-based speech enhancement

©藤村拓弥, 戸田智基(名古屋大学)

- ◆雑音混入音声を利用する学習のターゲットとして利用する学習法の Noisy-target Training (NyTT) が提案されている。
- ◆本稿では, NyTT とその品質を改善する新しい手法 (IterNyTT) に関する挙動分析実験を行った。
- ◆IterNyTT では, NyTT により音声強調を施した雑音混入音声を利用する学習のターゲットとして用い, 繰り返し学習を行う。
- ◆実験では, (i) IterNyTT が繰り返しによって性能を改善し (Fig.1), (ii) ターゲットの雑音とミスマッチのない雑音を与えられたときクリーン音声を用いる手法 (CTT) にも匹敵することが示された (Table 1)。

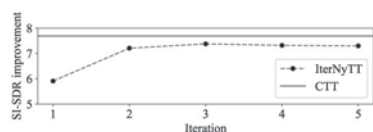


Fig. 1: Relationship between iteration of IterNyTT and its SI-SDR improvement

Table 1: Results of SI-SDR. Input noisy signals consists of CHiME

n_{add}	n_{obs}	Input	CTT	NyTT	IterNyTT
CHiME1	CHiME2		18.33	16.55	18.02
CHiME1	DEMAND2			17.03	17.80
DEMAND1	DEMAND2	10.64		14.50	16.00
DEMAND1	DCASE		16.89	15.37	14.84
DEMAND1	CHiME2			11.12	10.65

2-2-9

自由音場における CE-FDTD 法による 空間音響特性の合成と評価

Synthesis and evaluation of spatial acoustic characteristics
using CE-FDTD method in free field

©太久保翔太, 堀内俊治(KDDI 総合研究所)

- ◆CE-FDTD 法により, 境界における反射を考慮しない自由音場において, 空間インパルス応答の精度評価を行った。
- ◆周波数特性に関連する音響特徴量のうち, スペクトル重心, LSD (Log-Spectral Distance)を用いて定量評価を行った。
- ◆スペクトル重心の評価結果 (Fig. 1) から放射特性を考慮することで, スピーカ背面を除き, 実測値の分布傾向に類似する結果が得られた。
- ◆LSD の評価結果 (Fig. 2) からスピーカ正面においては放射特性の考慮の有無に関わらず精度が高く, また放射特性を考慮することによって 135 degree までの精度向上を見込める事が示唆された。

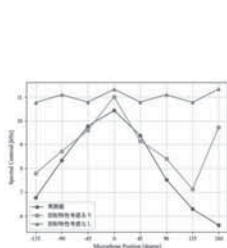


Fig.1 Spectral Centroid

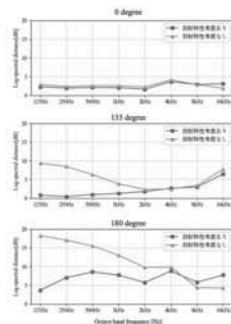


Fig.2 LSD

Figure 1 Spectral Centroid

Figure 2 LSD (Log-Spectral Distance)

2-2-8

2-2-8 騒音混入下における骨導音援用による 音声抽出法

Speech extraction method with help of bone-conducted sound
under existence of background noise

○折本寿子, 生田顯(県立広島大)

- ◆騒音混入下で実測した気導音データと骨導音データを利用し, 音声信号を推定
- ◆音声信号, 気導音, 骨導音に対する確率分布モデルを導入
 - 音声信号と気導音に対する確率分布モデル } 未知パラメータ
 - 骨導音と音声信号に対する確率分布モデル } の導入
 ベイズ定理を利用し音声信号, 未知パラメータを同時推定
- ◆ベイズ推定における観測音は気導音のみを利用し, 骨導音は予測値導出時に利用
- ◆高次相関情報を利用
 - 理論展開の簡易化と推定精度の向上を実現
 - 時間推移によるシステムモデル導入が不要
- ◆エネルギー値の S/N 比を約 1/3 にした有色雑音混入下で計測した女声信号に対し, 観測値と推定結果を示す

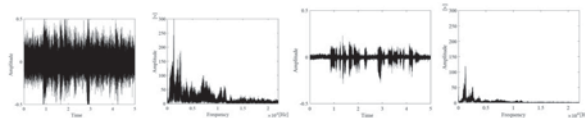


Fig. 1 観測値の波形 Fig. 2 観測値のスペクトル Fig. 3 推定値の波形 Fig. 4 推定値のスペクトル

実験は, 男声信号, 女声信号と白色雑音, 有色雑音, 機械騒音を用いて行った

2-2-10

2-2-10 バイノーラル收音用疑似頭および HATS の音響特性について

Acoustic characteristics of simple head and HATS for binaural recording

○稲永 潔文(サザン音響)

- ◆バイノーラル收音に用いられる HATS は, 近年様々な用途向けに, 頭部形状を簡略化した変形疑似頭も用いられている。

- ◆検討した疑似頭の形状は, A: IEC 60318-7 準拠の HATS, B: 左記 HATS の両耳および周辺部は同一で中央部をボール形状としたもの, C: 同中央部を $\Phi 80$ の円筒で繋いだもの (3 種) とした。



Fig.1 Type-A Head

Fig.2 Type-B Head

Fig.3 Type-C Head

- ◆各疑似頭に, 同一の IEC 60268-7 準拠の耳殻モデルと, その耳道入り口に 1/2in 無指向性マイクロホン装着し, 頭部形状以外の影響を最小とした。

- ◆自由音場における HRTF (音源 1m) の角度毎の特性は, A: HATS と比較し, 概ね 600Hz 以上の周波数で A→B→C と形状が簡略化されるほど特性変化が少なくなり, 角度毎の方向定位情報が減少している可能性がある。

- ◆この特性変化は, 顔面部における回折波や反射波が変化するためと思われる, IACC VS. 測定角度特性や拡散音場の HRFT にも現われ, 適切で適度な簡略化が必要であることを示している。



Fig.4 IACC VS. Measurement Angle of Type-A Head

Fig.5 IACC VS. Measurement Angle of Type-B Head

Fig.6 IACC VS. Measurement Angle of Type-C Head

2-2-11

2-2-11 Acoustic Contrast を目的関数とした
トランスオーラルシステムの性能評価Evaluation of transaural system using acoustic contrast
as its objective function

◎長倉知哉, 鮫島俊哉(九大・芸工)

- ◆Acoustic Contrast を目的関数としたトランスオーラルシステムは水平面前方における音像定位精度の向上が報告されている。
- ◆Fig.1 は小規模の制御器における両耳間レベル差(ILD)の再現性において当手法(C)の有効性を示すものである。
- ◆Fig.2 は再生空間伝達関数に誤差が含まれる場合の音源雑音成分の増幅(音質の劣化)における当手法(右)の有効性を示すものである。

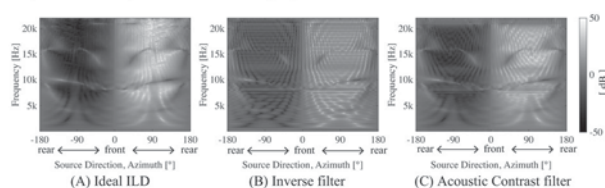


Fig.1: Interaural level difference (ILD) of reproduction signals

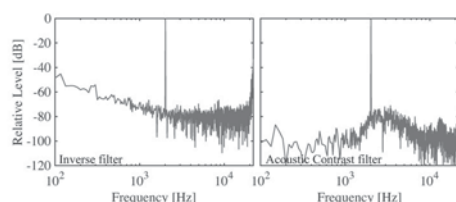


Fig.2: Frequency response of reproduction signal at direct side ear when 10% error is added into the reproduction system transfer function.

2-2-13

2-2-13 導電性ナノシートを用いた
聴覚ピエゾ・アクチュエータの作成と
交流周波数特性のサイズ効果Fabrication of piezo actuators using conductive nanosheet and
the size effect of AC frequency characteristics

○李珍咏(東大・学府), 伊東乾(東大・学環)

マイクロ・グラビアコータ(Micro-gravure coater)を用い PEDOT:PSS (poly 3,4-ethylenedioxythiophene : polystyrene sulfonate) のナノシートをピエゾフィルム(ポリフッ化ビニリデン, PVDF)の両面に積層、サイズを変えた聴覚ピエゾ・アクチュエータを系統作成し、交流周波数特性を測定した。5V の交流電圧を印加し周波数をパラメータにした交流電流振幅を測定したところ、可聴域(20Hz~20kHz)の平均電流値は、ピエゾフィルムの延伸方向にサンプルを2倍長くすることで1.6倍、3倍のサンプルでは3.8倍と増幅した[Fig.1]。また、パーコータを用い、導電性高分子を μm オーダでパルク塗布したサンプルと比較してみたところ、ナノシート電極は約数千分の一の薄さであるにもかかわらず電流振幅は約2分の1で同じオーダに留まった。

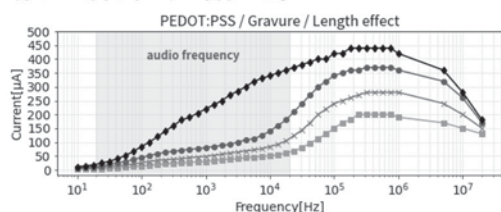


Fig.1: AC frequency characteristics of piezo actuators with different lengths in the stretching direction.

(■: Half-length, x: Basic size, ●: Double-length, ◆: Triple-length)

2-2-12

2-2-12 天頂方向のノッチ周波数の最適化による
全天空個人化頭部伝達関数の生成Generation of the individualized HRTFs in the upper hemisphere by
optimizing the notch frequency of the HRTF of the zenith direction

☆中村風香, 飯田一博(千葉工大・先進工)

- ◆3次元音像定位を実現するため、天頂方向のPNPモデルを用いたN2周波数の最適化により全天空の個人化HRTFを生成する方法を提案した。さらに、この方法を実現するツールキットを開発した(Fig.1)。
- ◆生成した個人化HRTFの音像定位実験を行った結果、水平面および横断面では良好な音像定位が可能であることが示された。正中面においては、正面、天頂、後方では良好な音像定位結果が得られたが、30, 60, 120°では改善の余地を残している(Fig.2)。

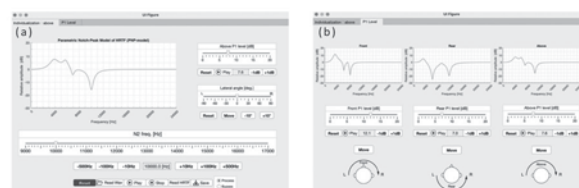


Fig.1 GUI of PNP model for individualization of HRTF for the zenith direction (a) and P1 level adjustment (b).

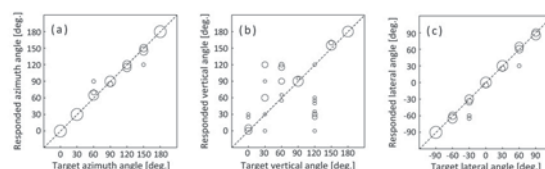


Fig.2 Responses to the individualized HRTFs in the horizontal plane (a), the median plane (b), and the transverse plane (c).

2-2-14

2-2-14 距離選択型音響レンズにおける
焦点距離可変機構の検討Variable focal length mechanism
for distance-selective acoustic lenses

☆村上龍斗(産技高専), 高橋義典(工学院大), 高野邦彦(産技高専)

- ◆本研究では、距離選択型音響レンズについて、音響ホーンの設置角度を変更できる機構を提案することで、焦点距離の可変な音響レンズを検討をした(Fig.1)。

(a) Front view (b) Back view

Fig.1 Acoustic lens with
variable focal length

- ◆音響ホーンの角度を変更することで、最適な收音距離を0.3m, 0.4mとなるように調整し、距離選択性をそれぞれ測定した(Fig.2)。

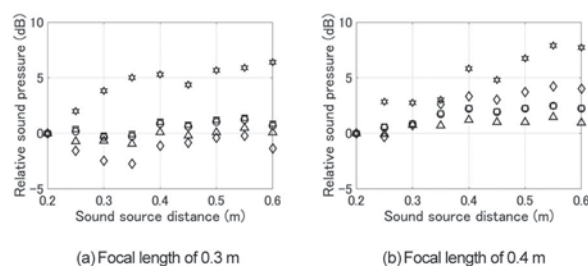


Fig.2 Observed sound pressure using acoustic lens normalized by the theoretical value of distance attenuation.

2-2-15

2-2-15

Azure Kinect を用いた発声者の指向性推定

Estimation of directionality using Azure Kinect.

☆倉本有紗, 及川靖広(早大理工)

- ◆これまで2台の Azure Kinect を用いて, 音声収録時における発声者の頭部運動測定を行うシステムが構築されてきた。
- ◆Azure Kinect に内蔵されている7マイクロホンアレイを用いて指向性推定を行なった。
- ◆指向係数 Q を算出し, 低周波帯では音が広がり, 高周波帯では指向性が鋭いことを確認した。
- ◆指向係数 Q の円は, 中心周波数によって大きさの順番が異なることから, 適切に残響音を収録することができなかったと考えた。

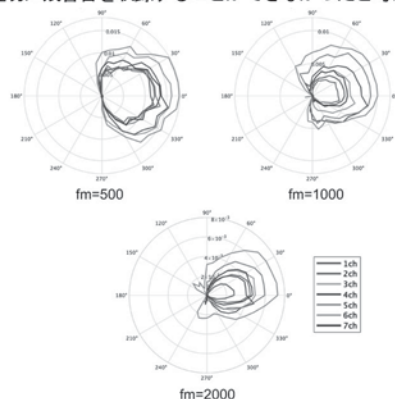


Fig.1: Calculated direction coefficient

2-2-17

2-2-17 偏光高速度干渉計を用いた
口笛演奏音収録への最小分散無歪応答法の適用

Recording of whistling sounds using parallel phase-shifting interferometry and beamforming with minimum variance distortionless response.

☆野澤遙, 草野翼, 及川靖広(早大理工)

- ◆偏光高速度干渉計による口笛演奏音の可視化結果から目的音源の収録を目的とし, ビームフォーミングによる音源強調を行った。
- ◆ビームフォーミングには遅延和 (DS) 法と最小分散無歪応答 (MVDR) 法を適用し, それらの音源強調結果を比較した。
- ◆音源強調後のパワースペクトログラムからはDS法では1377 Hz付近の第1倍音から2754 Hz付近の第2倍音までが確認でき, 計測結果から白色雑音の抑制が確認された。

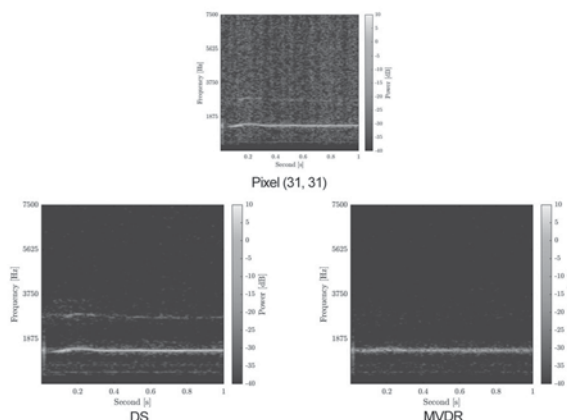


Fig.1: Power spectrograms of original signal at pixel (31,31) and after beamforming using 16 pixels by DS and MVDR

2-2-16

2-2-16 ガス管メンテナンス箇所推定システムにおけるリアルタイムマルチチャンネル処理の導入

Introduction of real-time multi-channel processing in system for estimating locations of gas pipe maintenance points.

☆黒沢琢登, 及川靖広(早大理工),
土井敏行, 中里直人, 小野寺文也(東京ガスネットワーク)

- 従来のガス管メンテナンス箇所推定システム
 - 音の強弱から磁場分布を可視化
 - 分布の対称性から磁場発源位置を自動推定
 - ◆発信受信コイル同士の向き関係によっては推定精度が低下

■提案手法

- システムの入力に3軸コイルアレイを導入
- 3軸情報を用いて磁場発源位置を推定

■実験結果

- 提案手法(Fig. 2 "all")ではコイル同士の向き関係(Fig. 1)によらず推定誤差が減少

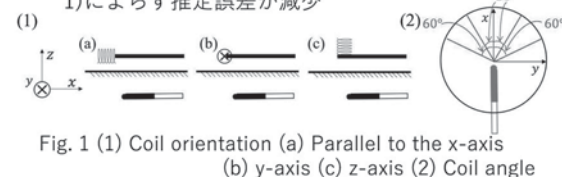


Fig. 1 (1) Coil orientation (a) Parallel to the x-axis (b) y-axis (c) z-axis (2) Coil angle

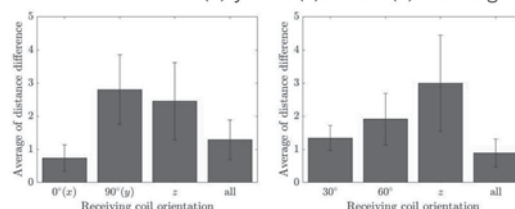


Fig. 2 Average of estimation error

2-4-1

2-4-1 「決定」を表すための操作反応音の特徴

Features of reaction sound that suitable for enter button

和氣早苗, 岸本陽(同志社女子大)

- ◆【目的】決定を表す2ピッチ音(ピボといった2つのピッチの音から成る音)の特徴について, これまでの研究を踏まえ, 選択音との関係性を含めた実験を行い, 考察を行う。
- ◆【実験の方法】テンキーから3つの数を選択し(選択音)その後決定ボタンを押す(決定音)という操作において, 選択音は単音, 決定音は2ピッチ音8種類を用意した上で, 決定音としてふさわしい2ピッチ音について絶対評価と比較評価を用いて印象評価を行う。
- ◆【結果】2ピッチの決定音について, 以下の2点が考察された。
 - (1) 2音ともが選択音と同ピッチ以上
 - (2) 少なくとも2音目が選択音と同じピッチの上行音

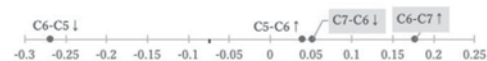


図4 心理尺度値 (C と C の組み合わせ)

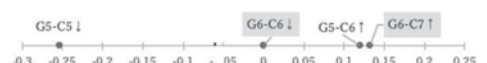


図5 心理尺度値 (G と C の組み合わせ)

2-4-2

2-4-2 和音のピッチが警告感に及ぼす影響の検討

Effects of pitch of chords auditory impression of warning

☆阿部理花子, 大塚翔, 中川誠司(千葉大)

- ◆警告効果の大きい報知音の設計を目指して、和音のピッチおよび協和性が警告感に及ぼす影響を検討した。
- ◆Scheffe の一対比較法を用いて、各刺激音の聴感印象評価を行い、警告感に及ぼす影響を調査した。
- ◆まず、構成純音の周波数比を 1:2:3 に固定した三和音の基本周波数を段階的に変化させ、Scheffe の一対比較法を用いて、警告感との関係を調べた。基本周波数、すなわちピッチが高くなるほど、警告感が有意に増大した。
- ◆次に、ピアノ音で構成される三和音を用いて、その構成音の基本周波数比を変化させることで和音の協和性を変化させ (Fig.1)、警告感との関係を調べた。周波数比が単純で協和性が高い音は警告感が小さく、周波数比が複雑で協和性が低下すると警告感が大きくなった (Fig.2)。
- ◆両実験から、ピッチが高いほど警告感が増大し、協和性が高いほど警告感が低下することが改めて確認された。前報 (阿部ら, 2021) の結果と比較すると、警告感には和音の協和性よりもピッチに強く影響される可能性が示唆される



Fig.1 Triads used in the experiment II

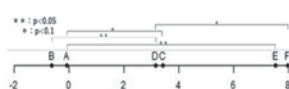


Fig.2 Mean scale values of impression of warning (experiment II)

2-4-4

2-4-4 聴覚及び触覚に着目した口紅容器開閉感の評価 —感覚モダリティごとの主観評価実験の実施—

Evaluation of opening/closing sensation of lipstick packages focusing on auditory and tactile sensation
-subjective evaluation for each sensory modality-

©岡崎龍太, △荒井観, △荒川尚美(樹資生堂 みらい開発研究所)

高橋莉紗, △竹下真, 大平菜津季, 内藤美桜, 三神圭司, 石田康二 (株)小野測器)

- ◆化粧品容器は気密性を保つことで中身を湿気や乾燥から保護するといった機能的側面のほかに、デザインや感触を通じてその製品やブランドの世界観を伝えるという感性的側面を併せ持つ。
- ◆本稿では口紅容器の開閉によって生じる音と感触 (Fig.1) に着目し、聴覚、触覚の両面から主観評価を行った結果について報告する。



Fig.1: Measurement of opening/closing tactile sensation of lipstick package

2-4-3

2-4-3 残響による音楽の拡がり感が接近音の気づきやすさに与える影響

Effect of sound image size of reverbering music on the perception of approaching vehicle's sound.

☆高倉 怜花(国立音大・院), 三浦 雅展(国立音大)

- ◆歩行者がヘッドフォンを用いて音楽を聴取する際、車やオートバイの接近音などが妨げられ、重大な事故を招く可能性がある。
- ◆本研究では従来のヘッドフォンを用いて音楽を聴取しながら、周囲の音も知覚できる新たな聴取環境を目的とし、音楽の定位と残響による拡がり感がオートバイの接近音の気づきやすさに及ぼす影響について検討した。
- ◆ヘッドフォンから再生される音楽に残響を付加し、水平面 0° と頭頂、定位させる前の元音源の 3 通りの位置から呈示し、同時に時速 30km/h で右後方から右横 (水平面 225° から 270°) を通過するオートバイの接近音を呈示した場合、音楽が頭頂に定位させた場合に、平均 14.6m, 最大 26.2m, 時間でいうと平均 1.8 秒, 最大 3.1 秒も早く知覚されることが示唆された。すなわち音楽を頭長付近に定位させることで、オートバイの接近音が過剰に知覚されることが示唆された。

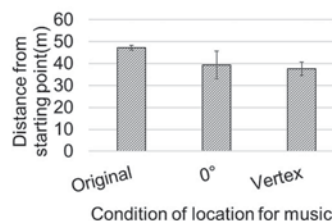


Fig.1: Average and 95%CI of the distance at which participants in the experiment perceived the sound of an approaching motorbike.

2-4-5

2-4-5 聴覚及び触覚に着目した口紅容器開閉感の評価—評価グリッド法を含む考察—

Evaluation of opening/closing sensation of lipstick packages focusing on auditory and tactile -Evaluation including the results of the Evaluation Grid Method-

○高橋莉紗, △竹下真, 大平菜津季, 内藤美桜, 三神圭司, 石田康二 (株)小野測器),

岡崎龍太, △荒井観, △荒川尚美(樹資生堂 みらい開発研究所)

- ◆口紅容器開閉感を人がどう知覚し、評価しているかを把握するため、評価グリッド法の手順により口紅容器開閉操作の評価構造を作成し、関連する音の物理量を考察した。
- ◆対象は、嵌合爪を用いた口紅 (ツメ嵌合、Sample a など) 及びキャップと本体の摩擦で茶筒のように固定する口紅 (空気嵌合、Sample e など) から選定した (Fig.1)。
- ◆「高級感」の評価には音の大きさや周波数特性など、複数の物理量が影響し、嵌合方式により結果が異なる場合もあることを確認した。

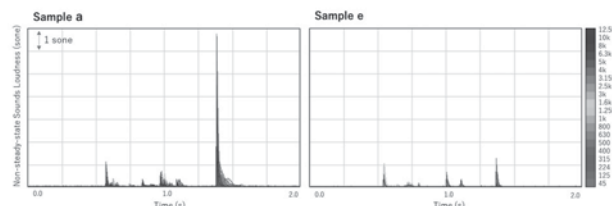


Fig.1: Non-steady-state sounds loudness of closing sounds of lipstick packages

2-4-6

2-4-6 クリック数に着目したロータリースイッチ
操作感覚のオノマトペによる評価

Onomatopoeic evaluation of rotary switch operating sensations
focusing on the number of clicks

☆宮入徹(中央大院/都産技研), △下村尚登, △白坂剛,

△大友貴史(アルプスアルパイン株式会社), 戸井武司(中央大)

- ◆ロータリースイッチのクリック数に着目し, その操作感覚から連想されるオノマトペ表現と物理量との対応について検討した。
- ◆実験では, スイッチ操作時の感触 (Fig.1 のシミュレータより提示), および操作音 (ヘッドホンより提示) から連想されるオノマトペを, 自由回答形式で被験者から抽出した。
- ◆結果から, 感触および操作音に共通して, クリック数が多いほど濁音を含む語, クリック数が少ないほど促音を含む語が多く用いられる傾向にあった。
- ◆一方, 長音表現においては, 感触から連想されるオノマトペはクリック数が少ない条件 (Fig.2(a)), 操作音ではクリック数の多い条件 (Fig.2(b)) で用いられる傾向にあり, 両感覚における差異も確認された。

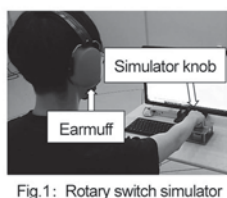


Fig.1: Rotary switch simulator

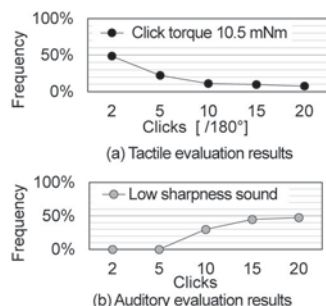


Fig.2: Frequency of word occurrence (Long vowel)

2-4-8

2-4-8 ニューロフィードバックを用いた
バイノーラルビート生成手法の検討

A Study of Binaural Beat Generation Method Using Neurofeedback.

○伊藤 弘章, 野口 賢一, 千葉 大将, 加古 達也, 日和崎 祐介 (NTT)

- ◆個人に合わせた音環境を作り出すことで, ユーザの集中力向上を支援する知覚制御システムを提案
- ◆脳波と音刺激に着目した知覚制御プロトタイプシステムとして, ニューロフィードバック型バイノーラルビート生成手法を提案
- ◆ユーザ本人の脳波信号から最適なビートを決定し, 脳波状態に応じたバイノーラルビートを生成
- ◆一般被験者による認知タスク実験により, 提案手法によるタスク成績向上効果を確認

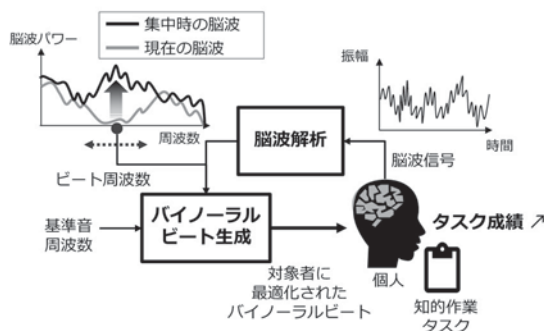


Fig.1: ニューロフィードバック型バイノーラルビート生成による知覚制御システム

2-4-7

2-4-7 音楽および映像の印象の強さと
主成分分析の累積寄与率との関係について

Relation between the cumulative contribution of PCA for music contents
and strongness of impact from video information

☆藤井香奈江, 脇田由実 (大阪工大)

- ◆音楽と映像を同時に視聴した場合の印象において, 音楽または映像どちらかの印象が支配的になる場合がある。各メディアの印象の支配度とメディアにおける主成分分析の累積寄与率との関係を主に音楽メディアを取り上げて分析した。
- ◆映像から受ける影響度合いの違いは, ピアノアルバムから抜粋した3曲 (図のP1~P3) と, メロディのみをDAWソフトにて作成した打ち込み音源3曲 (図のM1~M3) の間の差が顕著であった。
- ◆ピアノから抜粋した曲とメロディのみの打ち込み楽曲のΔMFCC に対し主成分分析を行った結果, 累積寄与率はFig.1 のようにピアノ曲に比べメロディのみの曲の立ち上がりが鋭くなった。
- ◆同ジャンル内で低次の累積寄与率に差がある楽曲の印象変化を調べた結果, 累積寄与率との相関は見られなかった。

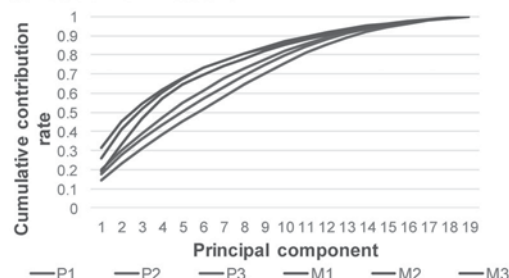


Fig.1: Cumulative contribution rates of variable music's ΔMFCC.

2-4-9

2-4-9 バイノーラルビートと性格特性に
おける集中力への影響

Effects of Binaural Beats and Personality Traits on Concentration

○野口 賢一, 伊藤 弘章, 千葉 大将, 加古 達也, 日和崎 祐介 (NTT)

- ◆個人毎に最適な周囲音環境を制御することで, 集中力向上を支援する知覚制御システムを提案し, 実現に向けた実験検証を行ってきた。
- ◆集中を支援する提示刺激音としてバイノーラルビート (Binaural Beats, 以降 BB) に着目し, 認知タスク実験を実施, 個人性の解明のために, Big Five 性格特性で分類比較した結果を報告する。
- ◆下図に結果例を示す。BBの提示により, タスクパフォーマンスに影響を受けやすい性格特性について, 分析した。

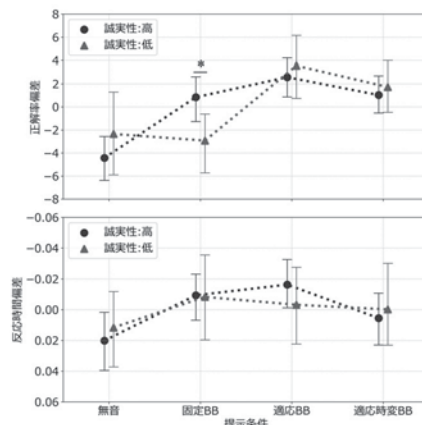


Fig. 1 Big Fiveの誠実性を考慮した音源種別ごとのN-Back タスク成績比較

2-4-10

2-4-10 Social force モデルを用いた背景音が歩行者モデルに及ぼす影響の分析

Experimental research on the effect of background music on the pedestrian model – Social force model

☆神戸美里, 塩見康博(立命館大)

- ◆本研究では複数の背景音下におけるボトルネック歩行実験から得られた移動軌跡データを用い、背景音の情緒的性質が Social force モデルのパラメータに及ぼす影響を分析した。
- ◆背景音は、印象評価アンケートの結果を基に「安心感」と「陰陽」という印象因子に特徴を持つ既成曲5曲を使用した。
- ◆Social force モデルのパラメータのうち、目標速度、加速時間、相互作用の大きさおよび影響範囲を歩行者ごとにチューニングしたパラメータを背景音ごとに比較した。
- ◆ヒストグラムの形状、基本統計量、有意差検定によって比較した結果、背景音がある場合、パラメータの分散が小さくなることが分かった。しかし、「安心感」と「陰陽」による明確な差異は確認されなかった。

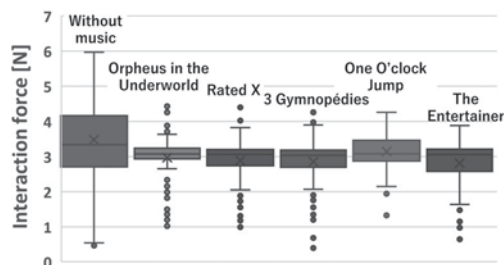


Fig.1: Distribution of parameters for each BGM

2-4-12

2-4-12 ウイルス感染症の飛沫感染リスクに対する音環境の評価指標に関する研究

Study on evaluation indicators of sound environment for risk of droplet transmission of viral infectious disease

☆窪前海斗, 木谷恒陽, 辻村壮平(茨城大院)

- ◆音環境に着目した飛沫感染リスク評価指標の構築を目指し、本研究では騒音レベルや発話レベルと飛沫量の関係性を捉えるために、心理音響実験を実施した。さらに発話時間を考慮した飛沫感染リスク評価を試みた。
- ◆心理音響実験の結果と既往研究の結果を比較し、騒音レベルが60 dB未満では60 dB以上に比べて飛沫量が減少する傾向がみられた。
- ◆意識的に大きく発声している場合の発話レベルは60 dB以上が多く、このような条件では飛沫量は増加する可能性がある。
- ◆会話時間を考慮した飛沫感染リスクの評価を行った結果、騒音レベルが60 dB以上で30分以上の会話では飛沫感染リスクは非常に高いが、騒音レベル50 ~ 60 dBの範囲では感染リスクは大きく低減する。

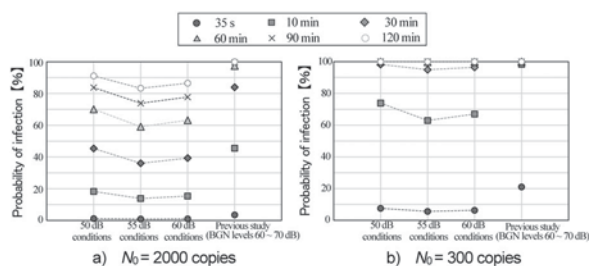


Fig.1 Risk of droplet infection considering BGN levels and conversation

2-4-11

2-4-11 音環境のアノニマス性その4: カフェにおける実験的検討

Anonymity of Sound Environment Part 4: An experiment in cafe environment

☆安田駿, 小橋侑生(神奈川工科大), 三浦貴大(産総研), 田中哲雄, 松本一教, 上田麻理(神奈川工科大)

前報では音環境のアノニマス性の音響的要素の解明を目的として、カフェの音環境をヘッドホンで呈示し、キーボード入力タスクを行い、その結果、自分の発生音への注意意識因子や発生音への鈍感化に関する因子がアノニマス性的一端である可能性を報告した。

本稿では、カフェの実環境下でキーボード入力タスクを実施し、音響的状況や他者の存在がどのようにアノニマス性へ影響するのか調べることにした。



Figure: Anonymity of the sound environment

2-6-1

2-6-1 生物模倣音を用いた水中音響測位用送信信号の基礎的検討 ~6種のイルカ鳴音の比較~

A basic study of transmitted signals for underwater acoustic localization using bio-mimic sound: comparison of bio-sonar signals generated by six species of dolphins and porpoises.

☆佐藤絵梨, 浦川騰太, 黒山喬允, 小笠原英子, 森和義(防衛大)

- ◆極浅海域の音波伝搬環境下での生物模倣音を用いた測位について検討する。送信信号として6種のイルカ鳴音を利用した検出性能の比較及び検証を行う。
- ◆ガウス分布雑音を用いた雑音評価から、ハンドウイルカとスナメリの鳴音を送信信号として選び、実海域における音波伝搬シミュレーションを行った。さらに実際に観測した沿岸域生物雑音を合成し、送信信号と受信信号との相互相関関数のピークと直達波の時間差を誤差として評価を行った。
- ◆この結果、ハンドウイルカの方がスナメリに比べて誤差の生じるS/N比が大きくなった。ハンドウイルカ波形はテッポウエビ雑音との相関が大きいたことが示唆された。

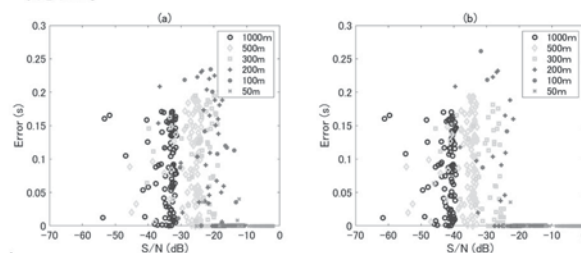


Fig. 1 Detection error vs. S/N ratio at Kagoshima in winter (a) using signal of bottlenose dolphin (b) using signal of finless porpoise.

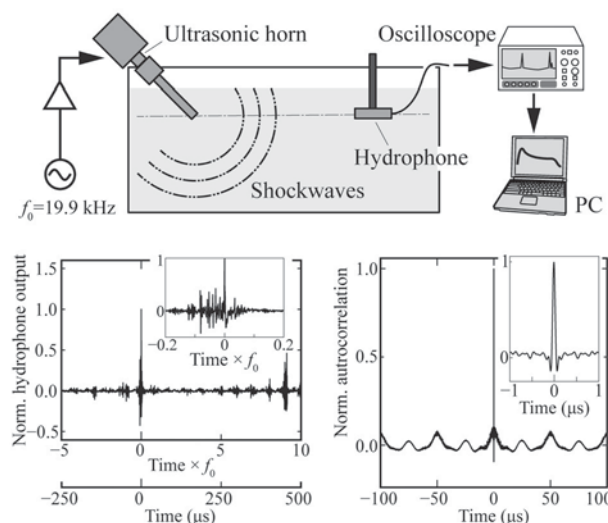
2-6-2

2-6-2 水中音響計測用音源としての
音響キャビテーションノイズの検討

Characterization of acoustic cavitation noise as a sound source
for underwater acoustic measurements

◎黒山 喬允, 小笠原 英子, 森和義 (防衛大)

- ・超音波ホーンにより水中に音響キャビテーション気泡を発生
- ・音響キャビテーションノイズをハイドロホンで計測
- ・受波信号の周波数特性と自己相関特性を解析
- ・受波信号は最低でも10 MHzの周波数成分を含む
- ・共振周波数10 MHzのハイドロホンでの受波信号の自己相関関数は68 nsの半値全幅を持つメインローブを形成



2-6-4

2-6-4 地震計波形記録からの遠地海底火山噴火
イベント抽出の検討

Study on extraction of remote submarine volcanic eruption events from
seismographs

◎岩瀬良一(JAMSTEC)

- ◆2021年8月13日の福徳岡ノ場及び2022年1月15日のフンガ・トンガーフンガ・ハアパイ火山(以下、「フンガ火山」)でそれぞれ発生した海底噴火について、筆者は先行研究において、日本海溝沿いに展開した海底ケーブル型観測網 S-net の海底地震計の観測波形を用い、スペクトログラム比較により、深海サウンドチャンネル(SOFARチャネル)を伝搬した噴火に伴う信号の存在を視覚的に識別した。
- ◆今回、この結果をもとにして、船舶や現在も比較的高頻度で発生する2011年東北地方太平洋沖地震の余震等の背景雑音を考慮し、海底地震計の波形データに5 Hzから10 Hzまでの周波数帯域を透過するバンドパス・フィルタを適用し、フィルタ適用後の波形に対して観測点毎に時間軸を伝搬時間分ずらし、更にそれを伝搬距離順に並べる解析を行った。その結果、噴火に伴う信号は横軸の同じ位置に出現し、噴火イベントとしての識別・抽出を可能とした。
- ◆福徳岡ノ場噴火については、2021年8月13日05:57 JST頃のパルス形状の波形、及びそれに先立つ05:55 JST頃からのイベント発生が確認でき、その後も複数のパルス形状の噴火イベントが確認できる。
- ◆フンガ火山噴火については、2022年1月15日13:20 JST頃から20分弱にわたる連続的なイベント及びそれに先立つ複数の比較的大振幅のパルス形状のイベントが確認できる。
- ◆今後手法の高度化及び噴火活動全体の把握を試みる。

2-6-3

2-6-3 複数の線形チャープ信号とM系列の組み合わせによる信号の相関特性

Correlation characteristics of signals generated by combining multiple linear
chirp signals and M-sequence

◎浦川 騰太, 佐藤 絵梨, 黒山 喬允, 小笠原 英子, 森和義 (防衛大)

- ◆多音源マルチスタティックソーナ用探信音として複数の線形チャープ信号(LFM信号)を用いて符号化された信号を用いて音源を識別可能とする手法であるVFMDM(Varying Frequency Modulation Division Multiplexing)を提案してきた。
- ◆VFMDMは送信ごとに傾きの異なるLFM-up及びLFM-downを割り当て、M系列で符号化した送信信号を生成する。
- ◆LFM信号の組み合わせによる識別性能の差異を調べるため、条件設定して選定した送信信号について、同一チャンネルの送信信号の曖昧度関数特性(Fig.1(a))及び異なるチャンネル間の相互相関関数特性(Fig.1(b))を評価した。
- ◆LFM信号の組み合わせは設計上考慮しなければならないほど識別性能に影響を及ぼさないことが示唆された。

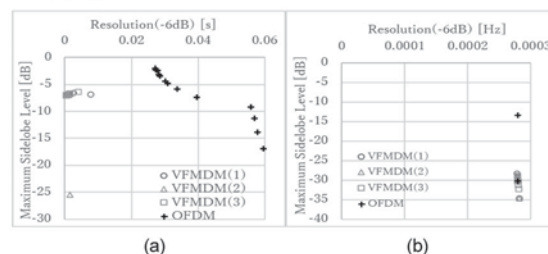


Fig.1: Characteristics of Ambiguity Function
(a) at Zero Doppler and (b) at Zero Delay.

2-6-5

2-6-5 PVDF フィルムを超音波エミッタに用いた
パラメトリックスピーカの高調波歪み低減の
評価

Evaluation of harmonic distortion reduction of parametric speakers using
PVDF film as ultrasonic emitter

☆広崎真早志, 野村英之(電通大)

- ◆本研究では、パラメトリックスピーカの持つ課題点である音質劣化の原因として高調波歪み(Total Harmonic Distortion, THD)に着目した。
- ◆課題解決のため、高調波歪みが小さい変調方式特徴を持つ平方根振幅変調(Square Root AM, SR-AM)を採用した。
- ◆SR-AM方式では広帯域音源が必要なため、セラミックタイプと比較し広帯域なPVDFを超音波エミッタに用いた。
- ◆音圧を確保するためにPVDFへ高電圧を印加することによって生じる歪みは、復調波に影響を及ぼさないことを確認した(Fig. 1 左)。
- ◆PVDFを用いた広帯域エミッタでは、遠距離場においてセラミックタイプと比較して高調波歪みが低減することを確認した(Fig. 1 右)。

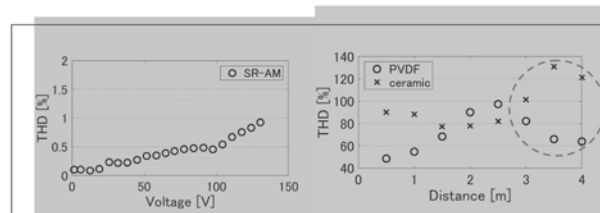


Fig1: Voltage dependence of PVDF vibration distortion (left), and Distance dependence of the distortion of the demodulated wave component (right)

2-6-6

2-6-6 振動面近傍に配置された小径穴を有する円板による噴出流の流速分布

Flow velocity distribution of the jet from a small hole of cylinder close to a vibrating surface

◎青野浩平, 孔徳卿, 青柳学 (室蘭工大・院)

- ◆振動面近傍に配置された小径穴付き円板からの噴出流の流速分布を Particle Image Velocimetry (PIV)により計測した。
- ◆Fig.1 のようにアクリル容器内に振動面と小径穴を持つ円板を1mm 間隔で配置し高速度カメラを用いて撮影した。
- ◆Fig.2 は噴出流時の画像であり, Fig.3 は Fig.2 の破線内(A)における PIV 結果である。Fig.3 よりギャップ内に空気は流入しないことがわかった。噴出流は穴の周囲の空気を吸い込み, 吸い込んだ空気を穴軸上に排出するように分布することがわかった。

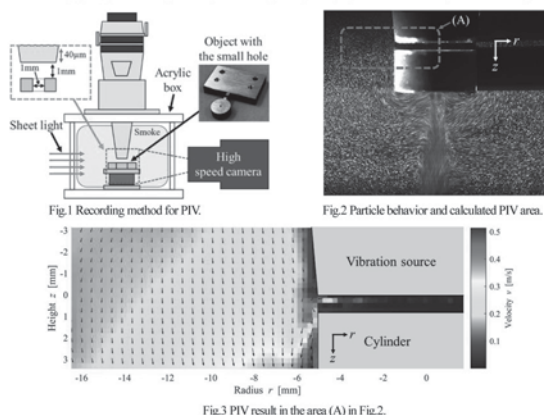


Fig.3 PIV result in the area (A) in Fig.2.

2-6-8

2-6-8 超音波霧化による撥水面から定在波への液滴の転送

Study on the transportation of droplets from the hydrophobic surface into the standing wave

☆頃安祐輔(筑波大), 星貴之(PxDT), △落合陽一(筑波大)

音響浮揚において, 超音波定在波に液滴を挿入する技術は実験の自動化のために必要である。本研究では, 撥水加工した反射板上に液滴を置き, そこに収束超音波を照射することで, 浮揚トラップの広さに応じた 2 通りの浮揚機序が観察され, それぞれについて報告する。1 つは液滴の霧化が発生し, 節に霧が引き寄せられる。もう 1 つは, 液滴が節に引き寄せられて撥水面上で弾み, 徐々に液滴が小さくなっていくことで安定した浮揚に至る。

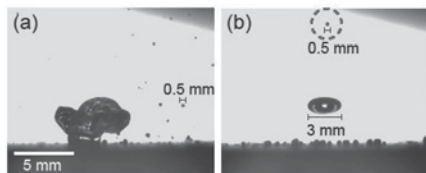


Fig. 1 焦点距離が 105 mm で転送される液滴

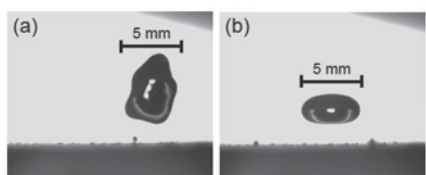


Fig. 1 焦点距離が 122 mm で転送される液滴

2-6-7

2-6-7 非平面振動面を有する振動子を用いた浮揚距離制御の試み

An attempt to control levitation distance using vibrator with a non-planar vibrating surface

☆唐木澤直樹, 青野浩平, 孔徳卿, 梶原秀一, 青柳学 (室蘭工大・院)

- ◆Fig.1 のような非平面振動面を有する振動子を用いて浮揚距離制御を行った。
- ◆Fig.2 のように振動子の印加する交流電圧の振幅を周期的に変化させ, そのときの浮揚物体の浮揚距離をレーザー変位計により測定した。
- ◆Fig.3 に電圧振幅の大きさと浮揚距離 H の時間変化を示す。印加電圧は 10.6 V を 100 ms 間, 18.0 V を 500 ms 間を周期的に変化させた。浮揚距離は最大で 1264 μm , 最小で 1106 μm の大きな値が得られた。

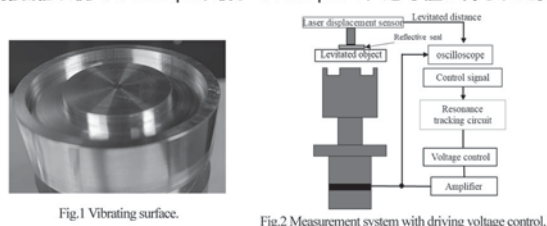


Fig.1 Vibrating surface.

Fig.2 Measurement system with driving voltage control.

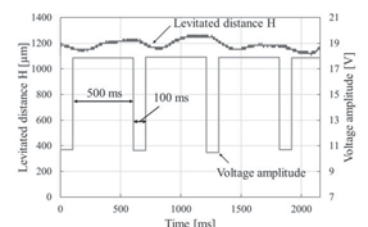


Fig.3 Time variation of levitation distance H and voltage amplitude.

2-6-9

2-6-9 超音波曝露の許容値に関する検討
第三報 超音波焦点における飽和

Study on Tolerance of Ultrasound Exposure (3rd Report) Saturation at Ultrasonic Focal Point

○星貴之、大岡佳生(ピクシーダステクノロジーズ株式会社)

- ◆近年、パラメトリックスピーカの用途展開や超音波フェーズドアレイの登場により強力な超音波がより身近なものになってきている。
 - 前々報において、140 dB SPL 未満の 40 kHz 超音波が危険であるというエビデンスは現在のところ存在しないことを指摘した。
 - 前報において、強力超音波が非線形吸収により減衰してやがて飽和し、平面波の場合は音源から 1.2 m、球面波の場合には 0.4 m 以上離れれば初期音圧によらず 140 dB SPL を下回することを示した。
- ◆本報では、収束する球面波の非線形吸収と飽和について検討し、周波数 40 kHz、焦点距離 10 cm、仰角 45 deg の場合の上限値 184.8 dB SPL を理論値として得た。
- ◆焦点音圧 172 dB SPL までの範囲で上記検討を実測により確認した。

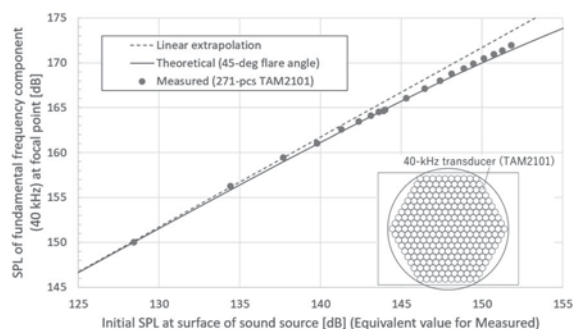


Fig.1: Theoretical vs measured SPL at focal point for 40 kHz ultrasound.

2-7-1

2-7-1 構音障害支援のための ウェアラブルデバイスにおける 簡易声質維持機能とインタフェース

A simple voice reproducing function and an interface on wearable device for assisting articulation disorder.

○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ている。
- ◆さらに、発声機能が残存している構音障害を対象として、話者が発した声に対してリアルタイムに音韻を補正する方式や、簡易的な声質維持機能を提案し、試作器でその可能性を示してきた。
- ◆それらの入力インタフェースとして、ジョイスティックの2次元座標入力から、ホルマントの可動範囲を考慮した座標変換を行い、構音をリアルタイムに制御するシステムの試作機を報告する。

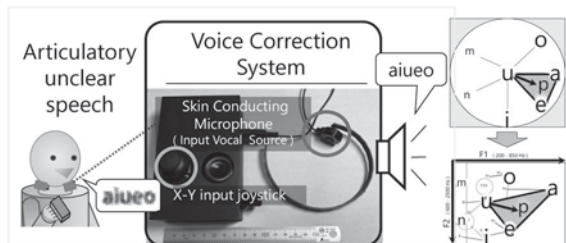


Fig.1: A concept and a prototype of voice correction system

2-7-3

2-7-3 ラウドネスレベル一定の条件における ダイアログエンハンスメントの効果法

Effect of dialogue enhancement under constant loudness level condition

○杉本岳大 (NHK)

- ◆ダイアログエンハンスメント(DE)は、番組音声におけるダイアログの聞き取りやすさ向上を目的とした、オブジェクトベース音響の機能であり、ダイアログと背景音との音量比(D/B 比)をカスタマイズできる仕組みである。本研究では、ラウドネスレベル一定の条件における DE の効果を主観評価実験で調査した。
- ◆図に、スピーカ再生による全 10 番組分の聞き取りと総合印象に関する評価結果の平均と 95%信頼区間を、DE 前の D/B 比(初期 D/B 比)からの DE 量を変数として示した。
- ◆実験結果から、「DEによって聞き取りは改善する。ただし、DE 量 ≥ 12 dB では改善が飽和する。」「DE 量が+3 dB の場合、どの初期 D/B 比でも、聞き取りに関する評価カテゴリが替わるほどの効果はない。」「DE によって、総合印象が低下する場合がある。」の知見を得た。

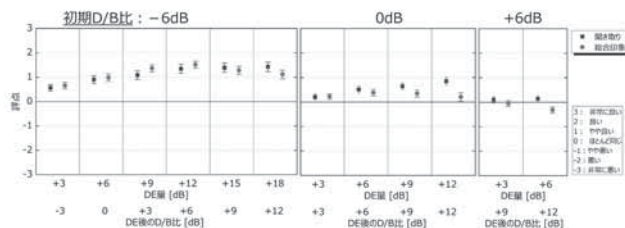


図 スピーカでの実験結果

2-7-2

2-7-2 五感に配慮した高齢者の 食事の QOL に関する基礎的検討

Basic study of dietary QOL for the elderly with consideration for the five senses

☆杉山洋介(神奈工大院/横浜創英大), 塩川茂樹, 饗場直美,
△小川喜道, △高橋勝美△松本一教, 上田麻理(神奈工大)

「住み慣れた地域、住まいでの尊厳のある自立した生活」にはフレイル予防や高齢者の社会参加が重要である。難聴を起因とした引きこもり高齢者も多いがフレイル予防のための聴覚データは乏しい。神奈川工科大学における食事支援と e スポーツを柱にした、高齢者社会参加・聴覚ビッグデータ取得のためのプロジェクトの紹介と聴覚分野からの食事支援について報告する。



Fig.1 Data acquisition cycle image that encourages social participation from the enjoyment of meals

2-8-1

2-8-1 デコーディング時の探索を考慮した 系列識別学習による End-to-End 音声認識

End-to-end speech recognition with sequence discriminative training under constraint of decoding algorithm

☆白木佑弥, 橋本佳, 南角吉彦, 徳田恵一(名工大)

本稿では、学習時と推論時の不一致による exposure bias を緩和するために、連続近似したビームサーチに基づくニューラルネットワーク学習を音声認識に適用する。また、損失関数と評価指標の不一致も緩和するために、系列識別学習も導入する。さらに、ビームサーチをより効果的に機能させるために従来のモデル構造の改良も行う。これらにより、学習時と推論時の不一致が緩和され、性能の改善が期待される。

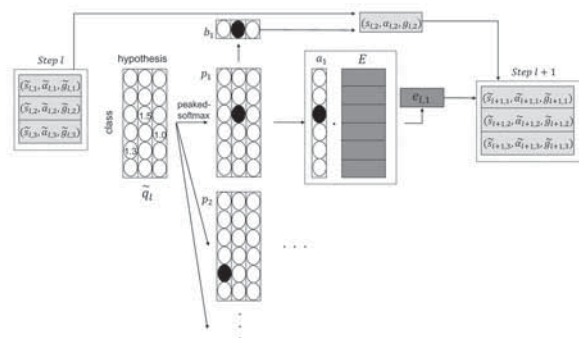


Fig.2: ビームサーチを考慮した学習の概要図

2-8-2

2-8-2 話者整合性損失と段階的最適化に基づく
音声認識と音声合成の半教師あり統合学習

Semi-supervised joint training of automatic speech recognition and text to speech based on speaker consistency loss and step-wise optimization

©牧島直輝, 鈴木聡志, 安藤厚志, 増村亮(NTT)

- ◆本発表では、小規模のペアデータ及び大規模のテキストデータを用いて音声認識と音声合成の両者を学習する半教師あり統合学習を扱う
- ◆従来法の半教師あり統合学習では、音声合成モデルが音声認識モデルに過適合し、音声認識モデルが認識しやすい合成音声を生産するように学習されるという課題が生じる
- ◆提案法では、従来の半教師あり統合学習に対して、話者整合性損失と段階的最適化を導入することでこの課題を改善し、音声認識性能及び音声合成性能が従来法よりも改善することを示す

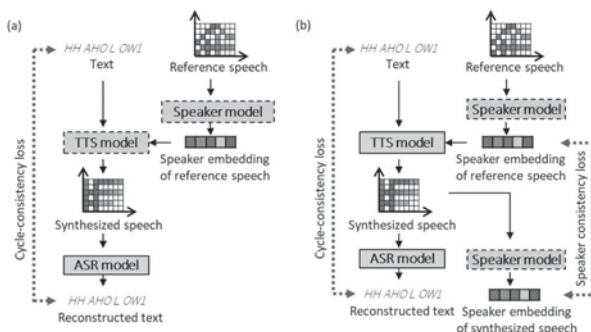


Fig.1: Overview of proposed method in (a) first-step training and (b) second-step training. Dashed and solid squares represent frozen and updated models, respectively.

2-8-4

2-8-4 自己教師あり学習したモデル表現の
知識蒸留: 生徒モデルの構造が
タスク精度に与える影響

Deep versus Wide: An Analysis of Student Architectures for Task-Agnostic Knowledge Distillation of Self-Supervised Speech Models.

©声原孝典, 森谷崇史, 松浦孝平, 田中智大(日本電信電話株式会社)

- ◆自己教師あり学習は、音声分野においても大いに注目を浴びているが、一般的にそのモデルサイズは膨大なため、DistilHuBERT に代表されるような、下流タスクに非依存な知識蒸留 (knowledge distillation: KD) を介した省モデルの構築は、重要な研究課題である。
- ◆Deep neural network は、モデル構造の深さ(層数)と幅(次元数)に関してスケールする根本的な性質を備えているため、KD する際の生徒モデルの構造は、その2要因に対し浅くするか狭くするかを選択することになる。しかしながら、そのような基本的な構造変化に起因した音声表現への影響については、ほとんど分かっていない。
- ◆そこで本稿は、SUPERB における各音声タスクの精度を通じて、モデル構造の変化に対する汎用性・表現力の変化を分析した。結果から、深い構造では内容系タスク、広い構造では話者系タスクに適していることが示された。

Model	KD Loss	PR	ASR (w/ LM)	KS	QbE	SID	ASV	SD	IC	SF	ER
		PERL	WERL	Acc	MTWV	Acc	EERL	DERL	Acc	F1 / CERL	Acc
HuBERT BASE	-	5.41	6.42 (4.79)	96.30	0.0736	81.42	5.11	5.88	98.34	88.53 / 25.20	64.92
DistilHuBERT	PLD	16.27	13.34 (9.21)	95.98	0.0511	73.54	8.55	6.19	94.99	82.57 / 35.59	63.02
12-L HALF	(a) PLD	13.09	11.87 (8.07)	96.97	0.0501	69.11	6.32	6.67	94.91	84.49 / 32.54	62.76
	(b) L2LD	10.67	10.96 (7.68)	97.24	0.0604	69.52	6.13	6.81	96.97	86.11 / 30.93	63.24
12-L FOURTHS	(c) PLD	18.92	14.02 (9.35)	96.44	0.0495	49.51	6.74	7.12	97.03	81.21 / 37.27	62.83
	(d) L2LD	16.96	13.84 (9.30)	96.40	0.0562	47.67	6.41	7.12	91.62	84.81 / 32.77	61.84
3-L ONE	(e) PLD	13.34	12.23 (8.64)	96.69	0.0489	75.71	6.48	6.56	94.15	82.89 / 34.65	63.95
	(f) L2LD	13.96	12.94 (9.11)	96.52	0.0568	47.76	6.18	7.17	96.02	85.99 / 32.38	62.57
3-L HALF	(g) PLD	18.62	13.91 (9.27)	96.22	0.0462	62.59	6.56	6.69	91.68	82.78 / 35.75	61.83
	(h) L2LD	18.11	14.48 (9.86)	96.48	0.0502	60.40	6.82	7.31	94.91	81.82 / 37.36	62.78

Table 1: Evaluation result for each model distilled from HuBERT BASE and each task on SUPERB. PLD means the predication-layer distillation and L2LD indicates the layer-to-layer distillation in the second column.

2-8-3

2-8-3 音声認識のデータ拡張のための
話者情報およびマスクを用いた合成音声の
周波数スペクトrogram強調

Enhancing synthesized speech using speaker information with phone masking for data augmentation of speech recognition

©上乃聖, 李晃伸(名工大), 河原達也(京大)

- ◆音声合成を用いた音声認識のデータ拡張では、通常音声合成側ではボコーダを用いて音声波形を生成しより拡張の効果を高める。
- ◆先行研究で生成に時間がかかるボコーダに代わる直接対数メルフィルタバンク特徴量を強調するネットワークを提案。
- ◆本研究ではそのモデルをもとに話者情報と音素単位のマスクとにより効率的な拡張手法を提案。
- ◆以前提案したモデルと比較し、提案手法は音声認識のデータ拡張の効果が高いことを示した。

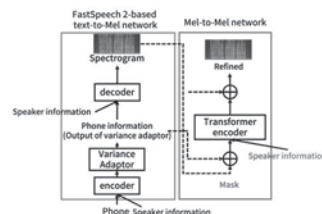


Fig. 1: The overall flow of proposed

Table 1: ASR performance (WER (%)) for TED-LIUM 2. We used 100h paired data (speech and transcription) from Librispeech100. We generated speech from TED-LIUM 2 (211h) transcription.

Model	test	Data generation time
Real (100h) [baseline]	27.60	—
Real + TTS w/o vocoder and post-processing	16.40	× 1
Real + TTS w/ vocoder	16.09	× 2.15
Real + TTS w/ Mel-to-Mel [previous work]	16.19	× 1.19
Real + TTS (100 + 211) [proposed]	15.66	× 1.19
Real + Real (100 + 211) (lmf) [oracle]	8.56	—

2-8-5

2-8-5 End-to-End 非流暢整形音声認識システム
の対話音声による評価

Evaluation of End-to-End Disfluency Shaping Speech Recognition System Using Spoken Dialogue

©堀井こはる(豊橋技科大), 福田芽衣子(徳島大), 太田健吾(阿南高専),

西村良太(徳島大), 小川厚徳(NTT), 北岡教英(豊橋技科大)

- ◆フィルラーを「#」、言い淀みを「@」といったラベルとして認識することで、これらを整形した認識結果を得る「End-to-End 非流暢整形音声認識システム」を対話音声によって評価実した。
- ◆実際の認識結果を以下に示す。

実際の発音: 「うーん 嬉しいって言うより し 信じられない」
 正解文: # 嬉しいって言うより @ 信じられない
 認識結果: # 嬉しいって言うより @ 信じられない
 目的文: 嬉しいって言うより信じられない

- ◆正しく認識できており、ラベル部分を削除することで、非流暢発話から言い淀みが整形された目的文を得られた。
- ◆文字誤り率 (Character Error Rate: CER)、文誤り率 (Sentence Error Rate: SER) とともにラベル付けにより大幅に削減できた。
- ◆検出率を見ると、過半数のラベルを正しく認識できており、特にフィルラーを正しく認識できていることがわかる。
- ◆対話音声認識においても、このシステムが精度改善に有効であることが示された。

Model	CER [%]	SER [%]
Baseline	16.0	62.8
Proposed	10.3	32.8

Label	Precision [%]	Recall [%]	F-measure [%]
#	81.7	70.2	75.5
@	69.6	59.8	64.3

2-8-6

2-8-6 MoCoVC: モーメンタム対照表現学習
によるノンパラレル声質変換MoCoVC: Non-parallel Voice Conversion
with Momentum Contrast Representation Learning

☆大西弘太郎、中鹿亘(電通大)

- ◆ 声質変換では話者性と発話内容をよく解きほぐすのが重要
- ◆ 対照表現学習を用いることで、中間表現に発話内容が残るような制約を明示的に加えることができる
- ◆ 対照表現学習の枠組みとして MoCo を取り入れた声質変換手法である MoCoVC を提案
- ◆ 自然性と話者性の主観評価実験において、提案手法の有効性を確認

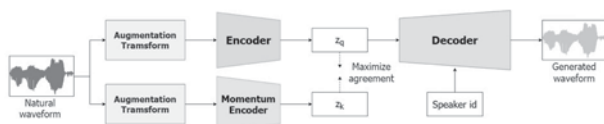
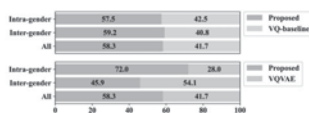


Figure 1: The overview of MoCoVC.

Table 1: MOS scores on naturalness
with a five-point scale

Method	Type	MOS
Ground truth	-	4.40 ± 0.06
MoCoVC (ours)	Intra-gender	3.83 ± 0.12
	Inter-gender	2.86 ± 0.14
	All	3.37 ± 0.10
VQ-baseline	Intra-gender	3.20 ± 0.15
	Inter-gender	2.57 ± 0.15
	All	2.87 ± 0.11
VQVAE	Intra-gender	3.50 ± 0.19
	Inter-gender	3.19 ± 0.14
	All	3.26 ± 0.10

Figure 2: Results with XAB test for
speaker individuality evaluation.

2-8-8

2-8-8 Sequence-to-sequence voice conversion
training using synthetic parallel data
for electrolaryngeal speech enhancement○ Ding Ma, Lester Phillip Violeta, Kazuhiro Kobayashi, Tomoki Toda
(Nagoya University)

- ◆ **Aim:** To propose an enhancement method to optimize the electrolaryngeal (EL) speech to normal speech conversion (EL2SP) based on sequence-to-sequence modeling.
- ◆ **Problems:** Seq2seq models need a large amount of training data. Otherwise, significant performance degradation would occur. And in general, the dataset for EL2SP is low-resource.
- ◆ **Solution:** A seq2seq-based two-stage training method by leveraging the largely imperfect synthetic parallel data is proposed to improve the EL2SP under the small original dataset.
- ◆ **Evaluation:** Objective evaluation and subjective evaluation.
- ◆ **Summary:** The proposed method can significantly optimize the performance of EL2SP under only 10 mins of original dataset.

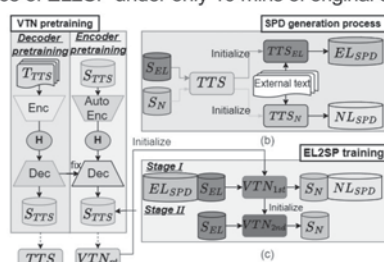


Fig. 1: The overall process of proposed method.

2-8-7

2-8-7 Three-stage voice conversion framework
for noisy and reverberant speech

☆ Yeonjong Choi, Chao Xie, Tomoki Toda (Nagoya Univ.)

This study presents a new voice conversion (VC) framework that can deal with both additive noise and reverberation. To leverage large amount of noisy-reverberant target speech data, we propose a three-stage VC framework based on denoising process using a pretrained denoising model, dereverberation process using a dereverberation model, and VC process using a nonparallel VC model based on a variational autoencoder. The experimental results show that the proposed method alleviates the adverse effects caused by both noise and reverberation, and significantly outperforms the baseline directly trained on the noisy-reverberant speech data.

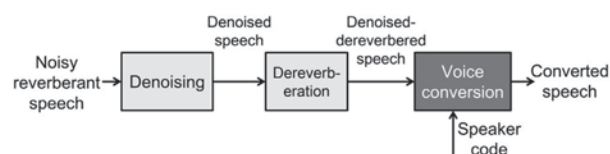


Fig. 1: Overview of the proposed three-stage VC framework for noisy and reverberant speech.

2-8-9

2-8-9 実用的なささき声の生成法:
Phantom Silhouette 方式の多言語対応と
ユーザによる声色調整A Practical Whisper Generation Method: Phantom Silhouette Method with
Multilingual Support and User Adjustment of Voice Tone

○内田照久(大学入試センター 研), 森勢将雅(明治大 総合数理)

- 目的** 通常音声から疑似ささき声を生成するパラメトリックな変換方式における多言語対応と機能の追加
- 方法** パラメタをユーザが調整して音色を作る仕様に変更
- (1) F_1 - F_2 シフト量・高域補填量・低域抑圧位置を操作可能に
 - (2) 発話速度・声色の変換機能を新たに追加
 - (3) ユーザが音作りを愉しみつつ 声の表現の自由度を広げる
- 評価** 日本語・英語・韓国語・中国語の母語話者が疑似ささき声を試聴、聴感的にささき声と認識、言語内容も了解

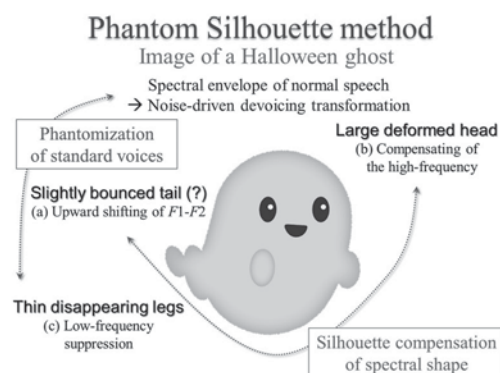


Fig. 1 Image of phantom silhouette method

2-8-10

2-8-10 自己教師あり学習と分類器チェーンを用いた非言語的発声に対する感情識別

Emotion Recognition of Nonverbal Vocalizations using Self-supervised Learning and Classifier Chains

☆辛 徳泰, 高道 慎之介, 猿渡 洋(東大院・情報理工)

- ◆非言語的発声とは、笑い声や泣き声のように、言語的な意味を持たず、かつ色んな感情を表せる発声である。
- ◆自己教師あり学習と分類器チェーンを用いた、非言語的発声に対する感情識別の方法を提案する。
- ◆提案法は、自己教師あり学習モデルで非言語的発声から特徴量を抽出して、分類器チェーンを用いて感情スコアを逐次的に予測する(Fig. 1)。
- ◆提案法は、通常のクラス分類損失ではなく、相関を測る一致相関係数を学習時の損失に使用する。
- ◆実験結果より、提案法が四つの baseline より一致相関係数を大幅に改善できることを示す。

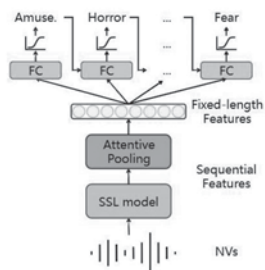


Fig.1: Architecture of the proposed method.

2-8-12

2-8-12 音声認識とのマルチタスク学習を用いた CTC モデルに基づく短区間音声感情認識

Short-term speech emotion recognition based on CTC model with multitask learning of automatic speech recognition

☆永瀬 亮太郎, 福森 隆寛, 山下 洋一(立命館大)

【概要】

- 文字付き感情ラベル列を予測する音声感情認識の提案
- 文字付き感情ラベル列の予測と類似する音声認識を補助タスクとしたマルチタスク学習の提案

【目的】

数フレーム単位で評価が可能な短区間音声感情認識の実現

【提案手法】

入力音声から文字に感情の情報を付与したラベル列を推定

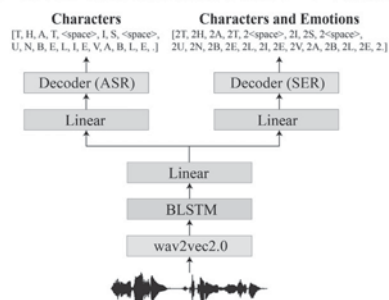


Fig1. The network structure of multitask learning with SER and ASR

【結果】

- 数フレーム単位の音声から約 47%~52% の感情認識を実現
- 発話単位の認識率において、従来の短区間感情認識より性能が向上

2-8-11

2-8-11 Transformer-based Speech Emotion Recognition with Multiple Acoustic Features

複数の音響特徴量を用いたトランスフォーマーに基づく音声感情認識

☆ Yurun He, Nobuaki Minematsu, Daisuke Saito (The University of Tokyo)

- ◆Speech Emotion Recognition (SER) is the process of detecting the emotional state of a speaker.
- ◆The existing SER systems always have only one kind of input feature.
- ◆Our previous work has verified the effectiveness of cross-attention transformer (CAT) for joint-encoding embeddings from bi-sources.
- ◆In this study, we propose a novel SER system named SMW_CAT utilizing CAT to fuse three acoustic features: log mel spectrogram (S), MFCC (M), and raw waveform data (W), as depicted in Fig. 1.
- ◆Experiments on the IEMOCAP dataset shows that our system achieves 73.80% WA and 74.25% UA, which surpasses the existing state-of-the-art approaches.

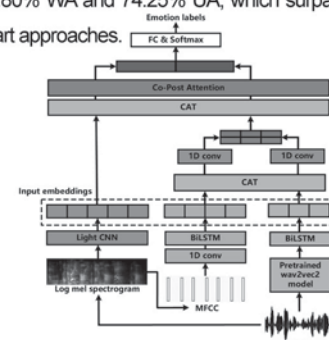


Fig. 1: Block diagram of our proposed SMW_CAT.

2-8-13

2-8-13 感情知覚特性に基づいた 自発的な叫び声の分類と音響的特徴量の関係

Spontaneous scream categorization based on perceptual emotional information and its acoustic tendency.

○大久保 港, △井岸 渉, 有本 泰子(千葉工大)

- ◆背景：叫び声を検出することで危機的状況を検知するシステムが開発されているが、叫び声には喜びや怒りの表現もある
- ◆目的：人がどのような音響的特徴から叫び声の表現を区別しているかを明らかにする
- ◆分析方法：叫び声に感情情報を付与する評価実験
付与された感情評価値を基に階層的クラスタリングによる分類
音響的特徴量を目的変数、叫び声の表現を説明変数とする回帰分析
- ◆結果：叫び声の表現は3種類に分類された
叫び声の表現の区別には基本周波数の平均値、ジッター、シマ、継続時間、音圧レベルの最大値が重要であることが示唆された

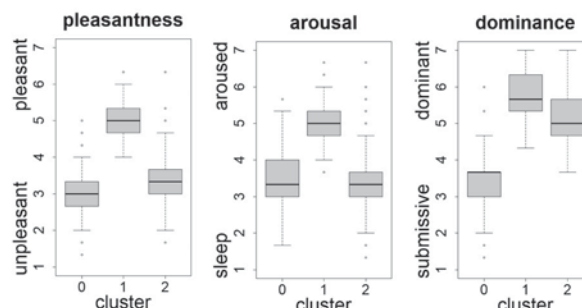


Fig.1: Boxplot of emotional scores for each cluster

2-9-1

2-9-1 話者交替に注目した実商談対話場面におけるハイパフォーマー分析

High Performer Analysis in Real Business Dialogue
Focusing on Switching Pause

○浅井拓也(早大・人), △神長伸幸(ミイダス), 菊池英明(早大・人)

- ◆実際の対話場面では**参加者全員が対話を継続する動機を有しているとは限らない**。
- ◆対話の目的の達成には、対話相手に対し**対話そのものの継続を促すことが重要**である。
- ◆本研究では**セールス側が比較的制御しやすいと考えられる話者交替に注目**し、解析を行なった。
- ◆ハイパフォーマーは自身の話者交替のみならず**対話相手の話者交替の時間も制御している**ことが示唆された。

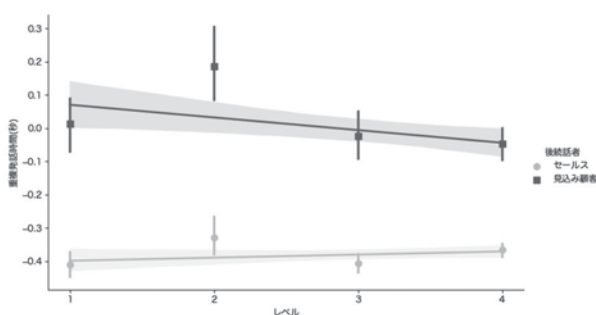


Fig.1: The duration of switching pause in some dialogs to make an appointment. (Participants ●: Sales, ■: Customer)

2-9-3

2-9-3 「STの音響学活用相談会」
求められているものWhat is required for "a meeting for speech therapists
to discuss how to use acoustics."

○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)

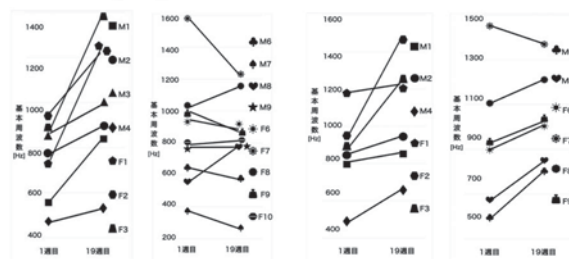
- ◆言語聴覚士は、ことばのリハビリを行う職業である。その養成校では音響学(聴覚心理学を含む)が必修であるが、学生の苦手で嫌いな科目の筆頭である。
- ◆さらに、現役の言語聴覚士も苦手であることに変わりなく、卒業後も臨床において音響学の知識を学び、活用する必要性を感じつつも、実現できないようである。
- ◆現役のSTのために開催している音響学の講習会(第8回STのための音響学)で、臨床における音響学の活用法を話し合った。本発表では、その経過と同時にに行ったアンケート調査の結果を報告する。
- ◆講習会の前半は、臨床での音響学の使い方の例の紹介を行い、後半は少人数グループに分かれて、各自テーマ(他のSTに聞いてみたいこと、自分がやってみたいと思っていること、実践していること)を決め、話し合いを行った。
- ◆参加者のディスカッションによって、臨床で具体的に実践している者は少ないが、現在模索している者が多く、今回のように他人に相談できる会は、効果的であり、必要であることが分かった。
- ◆今後は、現役STからの実際の臨床での様子の発信をもとに、どのようなことが可能であるかを考え、講習会で紹介できる内容を増やしたい。さらに、その内容を養成校の授業内容に還元していきたい。

2-9-2

2-9-2 チューブ発声法による基本周波数の
上昇の検討Examination of fundamental frequency increase by tube phonation
exercise

☆井上智瑛, △小野里真優, 竹本浩典(千葉工大)

- ◆チューブ発声法とは、ストローを用いる音声訓練技法の1つで、嚔声の治療や声学の訓練などに広く用いられている。
- ◆本研究では20代日本人成人(男性9名、女性10名)の実験参加者を用いて9ヶ月間にわたりチューブ発声法を行い、基本周波数の最高値と音圧の変化について検討した。
- ◆実験1では19週にわたり、グループ1はチューブ発声法の訓練を行い、グループ2は訓練を行わなかった。
- ◆実験2では10週にわたり、グループ1は実験1で使用したストローよりも細い直径のストローでチューブ発声法の訓練を行い、グループ2は実験1のグループ1と同じ訓練を行った。
- ◆結果、チューブ発声法により基本周波数の最高値が有意に上昇し、音圧も上昇する傾向が見られた(Fig.1, Fig.2)。



グループ 1

グループ 2

グループ 1

グループ 2

Fig.1: 実験1の結果

Fig.2: 実験2の結果

2-9-4

2-9-4 話速変化における音素長の変化の個人差

Individual differences in phoneme length changes due to speech rate

☆関純香, △鍋島由輝, △武井俊輔, △下条未来, 竹本浩典(千葉工大)

- ◆本研究では、15名の音声データから話速の変化に伴う音素長の変化などについて、基礎的な検討を行った。
- ◆実験参加者は、自身にとって普通(normal)、速い(fast)、遅い(slow)の3段階で朗読文「北風と太陽」(7文、195モーラ)を読み上げ、その音声から音素ラベリングに基づいて各音素の継続時間長を抽出して分析を行なった。
- ◆本研究では、発話の休止区間を含む時間を総発話時間、休止区間を含まない時間を実発話時間とした。また、この休止区間を「間」と呼ぶ。
- ◆普通に話すときに比べて速く話すときの時間変化(fast-normal)をf-n、遅く話すときの時間変化(slow-normal)をs-nで表す。
- ◆音素長の分析の結果、話速を変化させるときは、母音、閉鎖、間が積極的に利用されていることが明らかになった。
- ◆相関分析の結果(Fig1, Fig2)、速く話すときは母音と間を利用して発話時間を短縮していたが、遅く話すときは速く話すときより間を積極的に使い発話時間を伸長していることが明らかになった。

f-n	総発話時間	実発話時間	母音	閉鎖	その他	間
総発話時間	1.00					
実発話時間	0.73	1.00				
母音	0.66	0.73	1.00			
閉鎖	-0.23	-0.12	0.01	1.00		
その他	0.48	0.67	0.06	-0.14	1.00	
間	0.74	0.08	0.24	-0.21	0.04	1.00

Fig.1: f-nの各変量の相関(s)

s-n	総発話時間	実発話時間	母音	閉鎖	その他	間
総発話時間	1.00					
実発話時間	0.65	1.00				
母音	0.57	0.92	1.00			
閉鎖	0.27	0.52	0.26	1.00		
その他	0.63	0.81	0.66	0.21	1.00	
間	0.90	0.25	0.20	0.04	0.34	1.00

Fig.2: s-nの各変量の相関(s)

2-9-5

2-9-5 オペラ歌唱の演奏技法の違いによる
肺圧の制御に関する研究

Research on the control of lung pressure
due to different performance techniques in opera singing

☆戸田菜月, 竹本浩典(千葉工大), 高橋純(大阪芸大), 足立整治(帝塚山大)

- ◆前報では、オペラ歌手が高音から低音へ移行する際に用いる「支え直す」技術を検討した。このとき肺の容積を一時的に増大させて低音に必要な低い肺圧まで急激に肺圧を下げた後、再び上昇させていた。しかし、これは跳躍時に演奏技法を指定しない場合の技術であった。
- ◆本研究では演奏技法の指示なし、跳躍時に強く演奏するアクセント、滑らかに演奏するポルタメントの3つの演奏技法による肺圧の制御をバリトンとテノールの歌手各1名で比較検討した。
- ◆指示なしの場合、肺の容積の増加は見られなかったが、高音発声の途中から肺圧を下降させ、低音への移行後に肺圧を上昇させた (Fig. 1)。
- ◆アクセントの場合、各高音の発声開始時に急激に肺圧を上昇させ、発声途中から急激に肺圧は減少した。また指示なしと同様、低音への移行後に肺圧を上昇させた (Fig. 1)。
- ◆ポルタメントの場合、肺圧を連続的に上昇させながら高音へ移行し、下降させながら低音へ移行した後、急激に肺圧を低下させた (Fig. 1)。

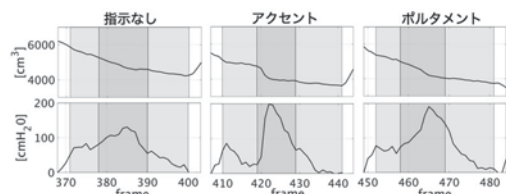


Fig.1:演奏技法ごとのバリトン歌手における肺の容積(上)と肺圧(下)

2-9-7

2-9-7 人声の距離減衰特性の
聴取印象評価

Sound impression evaluation of characterization of distance attenuation
for human voice

◎木下光太郎, 杉本岳大(NHK)

- ◆AR/VRなどの視聴時に音源の放射特性を再現することで、音源の向きが自由に変化した際に自然な可聴化を行うステアリングを検討している。
- ◆前報では、距離方向の放射特性を外挿する人声の距離減衰特性の付与法を提案し、シミュレーション上で従来法よりも有効であることを確認した。
- ◆本稿では提案法が聴感上でも有効か、主観評価実験により検証を行った (Fig.1)。

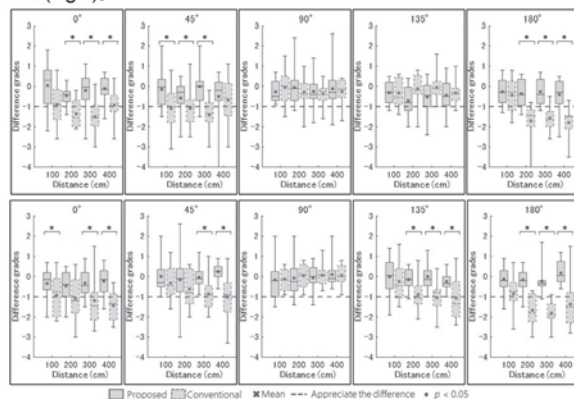


Fig.1 Subjective values of characterization method
(Upper: Male, Lower: Female)

2-9-6

2-9-6 オペラ歌唱における声の明暗の表現と
声道形状の制御の検討

A study on expression of voice brightness and darkness and
control of vocal tract shape in operatic singing

◎高橋純(大阪芸大), 戸田菜月, 竹本浩典(千葉工大)

- ◆オペラ歌手が「普通」の声、「明るい」声、「暗い」声を歌い分けた際の音響特性的変化と声道形状の変化を分析し、歌い分けのメカニズムを検討した。
- ◆3つの表現でプロの歌手は、「明るい」では第2, 3フォルマント(F2, F3)が上昇し、「暗い」では第1, 3フォルマント(F1, F3)が下降した。声楽の経験がない大学生(ama)では変化は見られなかった。
- ◆fMRI画像からamaの声道形状は変化の度合いは小さかったが、プロは「明るい」では口唇が開き、「暗い」では喉頭が下降して咽腔が伸長・拡大した。
- ◆声道断面積関数を用いたシミュレーションの結果、喉頭を伸長・拡大するとF1が下降し、口唇を開くとF2が上昇したが、F3は変動しなかった。これらはMRI画像と音声の分析結果にF3を除き一致した。
- ◆オペラ歌手は、喉頭を伸長・拡大してF1を下降させて「暗い」声を、口唇を開いてF2を上昇させて「明るい」声を表現していることは明らかになったが、F3の制御については明らかにできなかった。

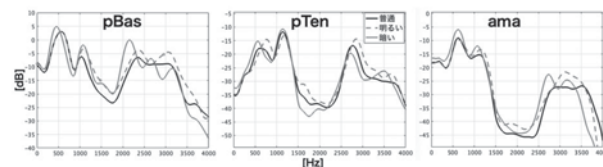


Fig.1 3つの表現における歌声のスペクトルの比較
(プロのバスバリトン:pBas, テノール:pTen, 経験のない大学生:ama)

2-9-8

2-9-8 人口声帯と声道模型を用いた会話時のエアロゾル生成量の定量評価

Quantitative evaluation of aerosols production during conversation used
artificial vocal cords and the vocal tract model

☆山谷悠介, 飯田明由, 吉永司(豊橋技科大)

- ◆COVID-19の感染拡大により、会話などから生成されるエアロゾル量の増加が注目されており、呼吸器系のエアロゾル発生源として声帯振動がある。
- ◆本研究では人工声帯と声道模型を用いて声帯振動によるエアロゾル生成量を定量的に評価する。
- ◆リード式人工声帯を口腔モデルに取り付けた発話模擬装置の声帯部に人口唾液を注入し、空気を一定量流入させた時に声帯から生成されるエアロゾルの量をパーティクルカウンターによって計測する。
- ◆同じ空気の流量で声帯振動時(発声)と静止時(呼吸)における各粒径のエアロゾル数を比較すると発声時のエアロゾル生成量は呼吸時の3.1倍に増加することを確認した。

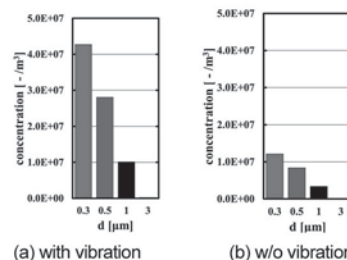


Fig. 1 Dependence of aerosol production on the vocal cord vibration
(27 L/min)

2-9-9

2-9-9 リード式人工声帯が発生する 喉頭粘液の液膜と飛沫の可視化

Visualizing liquid film and droplets of laryngeal mucus
produced by a reed-type artificial vocal fold

○荒井隆行(上智大), 吉永司, 飯田明由(豊橋技科大)

- ◆目的: ヒトが発話時に飛沫を生成する機序を明らかにする
- ◆実験:
 - 人工声帯に喉頭粘液を模擬した液体を塗布
 - 液膜と飛沫が発生する様子を可視化
- ◆結果:
 - リードと支持部の断面との間に液膜が出来た (Fig. 1)
 - その後、消滅
 - 別の瞬間では、上方には飛沫が飛び出ていく様子が観測

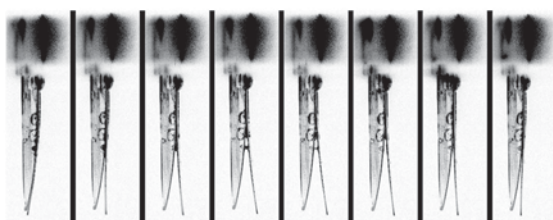


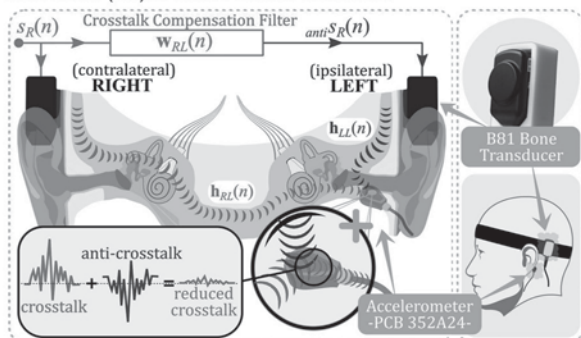
Fig. 1: Experimental results.

2-10-2

2-10-2 Unilateral Crosstalk Cancellation in Bone Conduction Using an Accelerometer Placed at the Mastoid

© IRWANSYAH, Sho OTSUKA, Seiji NAKAGAWA (Chiba University)

- ◆Bone conduction (BC) sound presented on either side of the head reaches the cochlea in both ears.
- ◆In this study, to attenuate BC crosstalk, an accelerometer placed at the mastoid was used to measure the vibrations caused by the crosstalk and then adjust the BC sound from the ipsilateral bone transducer (BT) to cancel out the vibrations.



- ◆We confirmed that even though the cancellation was designed to occur at the mastoid, participants could perceive the cancellation. They benefited from the crosstalk cancellation with an average improvement in the hearing threshold of up to 9.5 dB, depending on the frequency between 250 Hz and 1 kHz.

2-10-1

2-10-1 回転速度に寄与するパラメータが 頭部回転中の音像位置知覚の変位に 及ぼす影響

Effects of various parameters related to rotational speed
on the displacement of perceived sound position during head rotation.

☆守部颯一郎, 坂本修一(東北大)

- ◆本研究では、頭部の回転速度に寄与するパラメータの影響をそれぞれ相殺できる実験条件を設定し、頭部回転中に生じる音像位置の変位に各パラメータが及ぼす影響について検討した。
- ◆聴取者が様々なパラメータで左方向に回転する 5 実験条件 (頭部回転セクション)、聴取者が静止する 1 実験条件 (静止セクション) を設定し、聴取者正面に対して左 10° と右 10° での音像定位実験を行った。
- ◆知覚された音像位置の変位に着目すると、どの実験条件間でも有意な差は見られなかった。このことから各種パラメータが音像位置の変位に与える影響は小さいと考えられる。
- ◆一方で、音像判断を行った角度間で差が見られ、自身から離れていく角度となる 10° で、より頭部の回転方向へと変位する結果となった (Fig. 1)。頭部が回転することによって生じる音の相対的な接近・遠離によって差が生じたと考えられる。

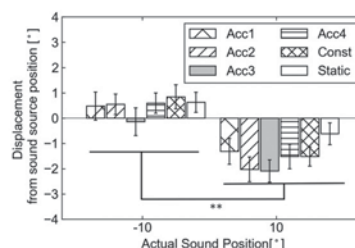


Fig.1.判断角度における知覚された音像位置の変位 (** $p < .01$) (正面を 0° とし右方向がプラス, 左方向がマイナスとなっている)

2-10-3

2-10-3 3D スキャンしたスピーカを HMD で提示する視覚刺激の背景として 提示したときの腹話術効果

The ventriloquism effect presented 3D-scanned loudspeaker image
as background image

©河井薫, 武藤憲司(芝浦工大)

- ◆本研究は、HMD 利用時における被験者にスピーカ位置が明らかである時の腹話術効果を調査した。腹話術効果とは、音像と映像の位置が異なる時に映像の位置から音が聞こえる効果のことである。
- ◆従来、腹話術効果は被験者に音像位置が視認できない条件で調査されてきた。本研究では被験者に音像位置を視認させ、音像と映像の位置を水平にずらした時のそれらのずれを知覚する角度を調査した。我々はそれを角度弁別域の評価値で調査している。
- ◆その結果スピーカ位置視認条件の方が、スピーカ位置非視認条件より角度弁別域が大きかった。これはアバタ映像がより音像位置を動かして腹話術効果が強く働いたということである。

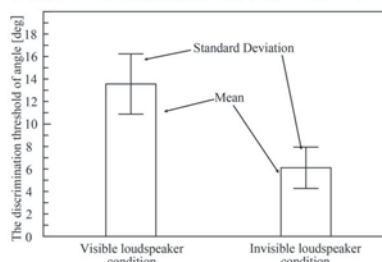


Fig. 1 Both results of the discrimination threshold of angle under visible loudspeaker and invisible loudspeaker conditions



Fig. 2 Visual stimuli under visible loudspeaker condition

2-11-1

2-11-1 アクティブ制御による防音壁の遮音効果を向上させるための二次音源の配置最適化

Improvement of noise barrier performance by optimizing secondary source arrangement of active control system

☆原健登, 川瀬遼我, 高野靖(京大院・工学研)

- ◆防音壁上部に二次音源を設置し、その振幅を制御するアクティブ防音壁において、音の減衰が顕著な領域を一般に“Dark zone”と呼ぶ。
- ◆本稿では、有限サイズのアクティブ防音壁において、三つの二次音源配置を用意し、遮音量を目的関数として音源の振幅最適化を行った。
- ◆結果、コの字状の二次音源配置は、直線状の配置よりも挿入損失が大きく、かつ小さな出力での遮音が可能なが分かった。

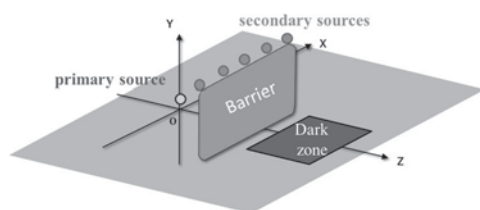


Fig.1 Constituents of active noise barrier

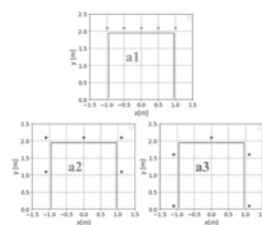


Fig.2 Three types of secondary source arrangement

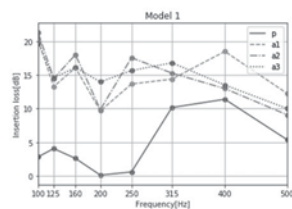


Fig.3 Insertion loss at evaluated frequency

2-11-3

2-11-3 2種の支持部材を用いた狭帯域低周波騒音制御向け圧電スピーカの開発

Development of piezoelectric speaker for low-frequency narrowband noise control using two types of support

©高安 俊貴, 後藤 達彦(株)東芝 研究開発センター)

- ◆ファンやモータなどの回転機による騒音、特に低周波成分の対策において狭帯域の能動騒音制御(Active Noise Control: ANC)システムが有効とされており、制御スピーカの小型・軽量化が強く望まれている。
- ◆本研究では、2種の支持部材からなる弾性支持機構を利用して圧電振動子の共振周波数を下げることにより、高出力の低音再生を可能とする小型で軽量の圧電スピーカを提案した。
- ◆バイモルフ型の圧電振動子と二分割型の円環ホルダ(2種)を利用して、提案スピーカの原理検証機を構築した(Fig.1(a))。
- ◆圧電振動子への質量付加による低周波化、保持部の側面支持による小型化、共鳴室設置による高出力化を実施した(Fig.1(b), (c), (d))。
- ◆正弦波掃引(500 Hz ≤ f ≤ 2 kHz)駆動時、最大92 dB(672 Hz)の低音を再生できることを確認した(Fig.2)。

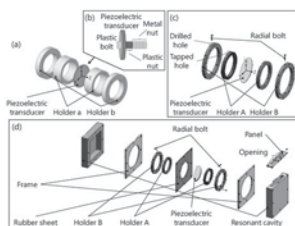


Fig.1: Proposed piezoelectric speaker.

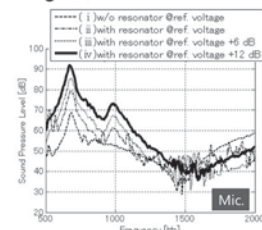


Fig.2: Frequency characteristics of radiated sound pressure from proposed piezoelectric speaker.

2-11-2

2-11-2 超低周波音による建具のがたつきを対象としたアクティブ振動制御 —試験家屋を対象にしたフィールド実験—

Active vibration control for rattling fittings caused by infrasound: Field experiment for the test house

©岩永景一郎, 土肥哲也(小林理研), 尾本章(九大・芸工)

- ◆超低周波音による窓のがたつき影響を対象にして、加振器と変位計を用いたアクティブ振動制御(AVC)を開発(Fig.1)
- ◆①窓の振動変位、②室内への伝搬音を、約10 dB低減(Fig.2)

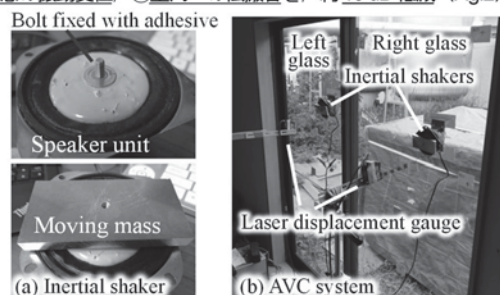


Fig.1 AVC system for rattling fittings caused by infrasound

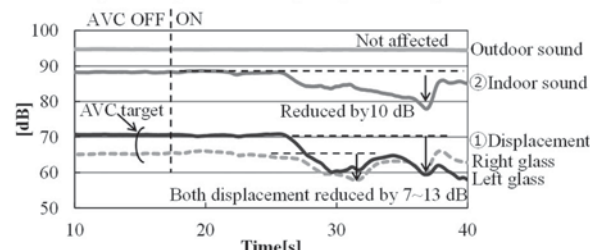


Fig.2 Vibration displacement of both glasses were controlled

2-11-4

2-11-4 機械振動のエネルギー伝達特性に基づく解析

Analysis based on energy transmissibility of mechanical vibration

○山崎徹(神奈川大工), △山崎弘隆(三菱自動車),

△武田正利(電通国際情報サービス), △岡建樹(電通国際情報サービス)

- ◆自動車業界は100年に一度の大変革期の真っ只中にあり、ものづくりはモノ・コトづくりへ大きな変化が必須である。低コスト・高価値を実現するために、「形で考えないNVHモデル研究会」を発足し、産学連携にて、多性能適正化のための全体俯瞰・共創活動を始めた。
- ◆「形で考えないモデル」は、設計初期に、設計空間の拡大、新発想の創出も意図し、機能を数式などで記述されるものである。
- ◆本報では、取り組み例として、振動性能の形で考えないモデルとして、高周波および低周波の二つの振動エネルギー伝達モデルを概説し、それらの有効性を示す。
- ◆振動エネルギー伝達モデルの特長は、いずれも対象系の要素間のエネルギー伝達に着目する(図1)。エネルギーを蓄える要素(ブロック)と要素間の経路(線)を表すエネルギーブロック図である。この図により、対象全体を俯瞰でき、また例えば、図中に#newで示す「エネルギーを蓄える新たな要素(機能)」の追加(Additional function)や、要素#4と#7に点線で示す「新たな伝達経路」の追加(Additional path)などといった新発想がしやすい。



Fig.1 An example of energy-block diagram

2-11-5

2-11-5 マッハカットオフ騒音のハイブリッド解析の試み

Hybrid analysis of nonlinear acoustics and FDTD method for Mach cutoff noise
○土屋 隆生 (同志社大・理工), 金森 正史 (JAXA)

本稿では, Mach cutoff 騒音の解析手法として, プーム波伝搬と擾乱波伝搬を分離して計算するハイブリッド解析法を提案している。プーム波は拡張 Burgers 法で解析し, caustic 近傍の回折波は非線形 Tricomi 法で解析している。得られた高忠実度波形からプーム全体の音圧分布を計算し, FDTD 法へ入力することで Mach cutoff 騒音のみの伝搬計算を実施している。Mach cutoff 騒音が入射波, 集束波, 回折波のいずれの影響を大きく受けているのかを検証したところ, Fig. のようにほとんどが入射波に影響を受けていることが示された。

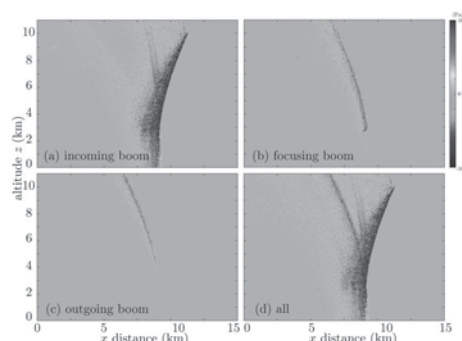


Fig. 1: Sound pressure distributions of Mach cutoff noise caused by the incoming, focused, and outgoing waves.

2-11-6

2-11-6 低周波音モニタリングシステムの検討

-風雑音判別とアレイを用いた音源方向推定-
Examination of low frequency sound monitoring system

土肥哲也, 高山徹也, 岩永景一郎, 小林知尋(小林理研), 中島康貴 (リオン)

◆風雑音判別手法とマイクロホンアレイを組み合わせた低周波音モニタリングシステムを検討した (Fig. 1)。

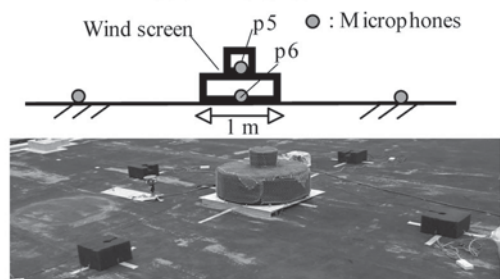


Fig. 1: Overview of low frequency sound monitoring system.

◆雷鳴を観測した際の風雑音の判別結果と, 音の到来方向の推定結果 (Fig. 2) を示した。

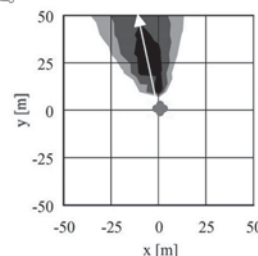


Fig. 2: The result of estimating the direction of arrival of thunder sound.

2-11-7

2-11-7 トンガ噴火由来のシグナルの分離と特徴の整理

Extraction of 3 types of signals from observed atmospheric pressure wave induced by the Tonga eruption

○新井伸夫, △乙津孝之, △本橋昌志, △岩國真紀子(日本気象協会)

- ◆現地時間 2022 年 1 月 15 日に発生した南太平洋トンガのフンガ・トンガ・フンガ・ハアパイ火山の噴火に由来するシグナルについて, 国際的な観測網, および国内に展開されている観測点で得られた波形を整理・分析。
- ◆参照したのは, CTBT のもとで展開されている観測網, 米国大学間地震学研究連合(IRIS)の観測網, および国内の防災科学技術研究所の火山観測ネットワーク(JVDN), 大阪市大, 筆者らが展開している観測点で得られた連続波形。
- ◆大規模な噴火を想起させる特徴的なシグナルが, 姿を変えずに長距離伝搬していることを確認。その周期の長さとともに形状を保持して伝搬しているその特徴から, このシグナルは大気境界波(ラム波)として伝搬したものと同定。遠方でのシグナルの到達時刻から推算されるその平均的な伝搬速度は, 300 - 320 m/sec。
- ◆音波の帯域に着目して長距離伝搬の様子をみてみると形状が変化し継続時間が延びているようにも。この帯域には短周期の大気境界波と音波が混在している(遠方になるほど境界波が支配的か)と考えられるが, その特徴分析は課題。
- ◆国内の観測波形からは, 重力波の存在が示唆され, その伝搬速度は 240m/sec.程度と推定。気圧変動が潮位変動を励起したとされるが, この重力波がそれを担った可能性。

2-11-8

2-11-8 低周波音による圧迫感・振動感の主観評価

—純音を用いた実験的検討—

Subjective evaluations of oppressive or vibratory feeling induced by low-frequency sound: experimental investigation using pure tone sound

○横島潤紀(神奈川県環境科学セ/神奈川県), 森長 誠(神奈川県), 牧野康一, 土肥哲也, 横山 栄, 小林知尋(小林理研), 山崎 徹(神奈川県)

- ◆低周波音により生じる圧迫感・振動感について, この感覚が生じる人体の部位にも着目した主観評価実験を実施した。
- ◆実験に用いた刺激は, 周波数の範囲が 10~160 Hz, 音圧レベルの範囲が 40~100 dB にある 40 種類の純音(継続時間 20 秒)とした。
- ◆下表は, 本実験結果を整理して, 音圧レベルと圧迫感・振動感の知覚割合との関係を周波数別に整理したものである。表中の [] に示す記号は, 当該の提示条件において知覚割合が最も高い部位(「耳の周り」: a, 「耳の奥」: b, 「頭部」: c, 「胸部」: d, および「腹部」: e) を示している。

Table: Relationship between sound pressure level and perceptual proportion of oppressive or vibratory feeling by frequency		SPL (dB)						
FRQ (Hz)		40	50	60	70	80	90	100
10		—	—	—	—	3.3% [abcde]	20.0% [b]	46.7% [b]
20		—	—	—	10.0% [b]	46.7% [b]	73.3% [b]	70.0% [bc]
40		10.0% [b]	13.3% [b]	36.7% [b]	56.7% [c]	66.7% [b]	73.3% [b]	80.0% [b]
50		10.0% [b]	30.0% [b]	43.3% [b]	60.0% [b]	66.7% [b]	66.7% [bc]	76.7% [b]
63		20.0% [ab]	30.0% [a]	40.0% [b]	53.3% [b]	50.0% [c]	76.7% [d]	—
80		20.0% [b]	26.7% [ab]	33.3% [b]	43.3% [c]	46.7% [b]	60.0% [b]	70.0% [b]
160		—	33.3% [b]	20.0% [b]	43.3% [c]	53.3% [c]	56.7% [bc]	73.3% [b]

2-11-9

2-11-9 低周波数の純音による圧迫感・振動感の
閾値実験—調整法による検討—

Laboratory study on the threshold of oppressive or vibratory feeling induced by low-frequency pure-tone: investigation by the method of adjustment.

○森長誠(神奈川大), 横島潤紀(神奈川県環境科学センター/神奈川大), 小林知尋, 横山栄, 牧野康一, 土肥哲也(小林理研), 山崎徹(神奈川大)

- ◆低周波音による圧迫感・振動感の閾値を周波数別に検討することを目的として, 低周波音実験室での音響心理実験を行った。
- ◆手法: 上昇系列2回+下降系列2回の調整法を用いた。実験参加者がボリュームコントロールで音量を調整した。
- ◆刺激: 20 Hz, 40 Hz, 50 Hz, 63 Hz, 80 Hzの純音。
- ◆実験参加者: 20歳から59歳までの30名の成人(分析は29名)。
- ◆各参加者の4回の系列における調整値の平均値を算出し, 29名のデータで周波数別の調整値を分析した結果をFig.1に示す。
- ◆重回帰分析の結果, 高感受性群, 女性, 40歳未満は調整値を有意に小さく設定していた($p < 0.05$)。

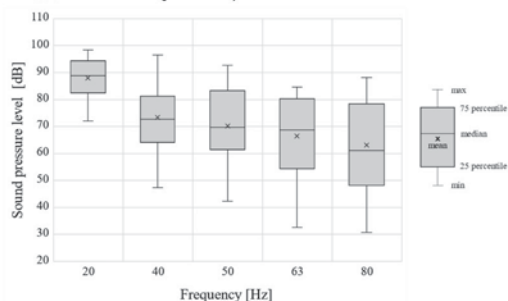


Fig.1: Ranges of participants' thresholds for oppressive or vibratory feeling.

2-11-11

2-11-11 厚木市における防災無線放送の長期観測
その5:

～季節の違いに関する検討～

Long-term observation of disaster prevention radio broadcasting in Atsugi City Part 5:

Examination of seasonal difference

☆仲村雄大, 橋本卓己(神奈川工科大), 三浦貴大(産総研), 上田麻理(神奈川工科大)

前報では防災無線放送の悪天候時における聞き取り易さに及ぼす影響の解明を目的として, 雨天時の防災無線放送の音源をヘッドホンで呈示し, 聴感評価を行い, その結果, 雨天時において, 降雨量によって周囲の音へ影響を及ぼし, 防災無線放送の聞き取り易さにも影響している可能性を報告した。

本稿では, 防災無線放送の聞き取り易さが季節によってどのように変化するか, 1/3 オクターブバンド分析により調べることにした。



Figure: Image of listening to disaster prevention radio broadcasting

2-11-10

2-11-10 分散型音響計測を用いた
音源探索システムに関する基礎的研究

A basic study on sound source identification system using distributed acoustic measurement

☆池見樹(近畿大院), 原田和典(岡山県立大), 菅原彬子, 平栗靖治(近畿大)

- ◆本研究では, 分散配置させたマイクロホンから信号に相互相関関数を用いることで, 一般的なマイクロホンアレイを使用する方法では困難な, 広域での音源探索を可能なシステムの構築に向けた基礎的検討を行った。
- ◆複数の信号とその座標から音源位置を推定するプログラムの構築, および音声信号を収録するためのデバイスの開発を行った。
- ◆単一の固定音源を対象とした精度検証では誤差が20mm以下と精度よく推定できた。また移動音源に対してもおおよその位置を推計できた。
- ◆収録用のデバイス開発ではGPS信号を用いた時刻同期等を搭載したが, A/Dコンバータを中心に機器選定に課題が残った。

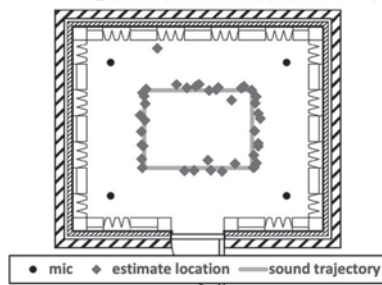


Figure. Results of location estimation of moving sound sources

2-11-12

2-11-12 国内外における環境騒音測定における
暗騒音の定義について

Definition of background noise for measurement and assessment of environmental noise

小林知尋, 横山 栄 (小林理研)

発生源が異なる複数の騒音が同時に存在する一般環境騒音を扱う際, 騒音をどう分類して分析するかが重要であるが, 規制や規格等で騒音の分類方法や定義が異なっているのが現状である。本研究では, 環境騒音の測定方法に関する国内外の規格を中心に調査を行い, 環境騒音の構成要素の分類方法および定義について, とりまとめた。日本国内のJIS 1996-1だけを見ても, 2019年に初めて本編で定義された「残留騒音」や一般的に馴染みのある「背景騒音(暗騒音)」等の定義は正しく理解されていないように思われる。さらに, 国際的にも, 環境騒音の構成要素の分類としての“ambient noise”や“residual noise”, “background noise”等の定義は様々であった。今回は一般環境騒音を中心に調査したが, 風車騒音等の他の騒音に関する規格についても, 定義が統一されていない状況が見られ, 調査を継続する予定である。

Table Classification of environmental noise in JIS Z 8731.

分類	英訳	1966年版	1983年版	1999年版	2019年版
総合騒音	total noise	—	環境騒音 (ambient noise)	○	○
特定騒音	specific noise	—	○	○	○
残留騒音	residual noise	—	△解説に記載	△備考に記載	○
暗騒音	background noise	○	○	△備考に記載	○背景騒音 (暗騒音)
		英訳なし	ISO 1996/1-1982 を参照	ISO 1996/1-1982 を翻訳	ISO 1996-1:2016 を基に作成

2-11-13

2-11-13 サウンドスケープ評価における Perceived Affective Quality をめぐる一考察

On Perceived Affective Quality Scales for Soundscape Evaluation

○永幡 幸司 (福島大学)

- ◆ISO/TS 12913-2 の Annex C に掲載されているサウンドスケープの“perceived affective quality”を評価するとされている尺度(PAQS)は、“Perceived Affective Quality”を名乗っているものの、正しい意味での“Affective Quality”の尺度となっていない。
- ◆PAQS の基となっている Axelsson らの提案は、Russell らの環境心理学と感情心理学における Circumplex モデルについて1980年頃の古い議論のみを参照しており、近年の知見が反映されていない。
- ◆Circumplex モデルについての近年の知見と、永幡らの複数のサウンドスケープの評価実験の結果を照らし合わせると、PAQS の評価モデルは、Fig. 1 のように表されると考える。

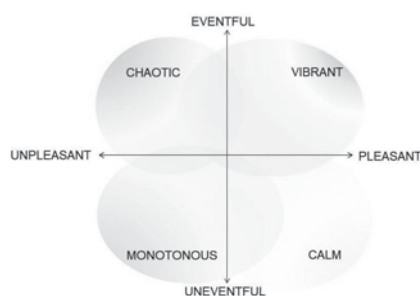


Fig.1: A revised conceptual model of evaluation of “perceived affective quality” of soundscapes using PAQS proposed in ISO/TS 12913-2

2-12-2

2-12-2 卓上衝立設置時の会話印象に関する心理評価実験

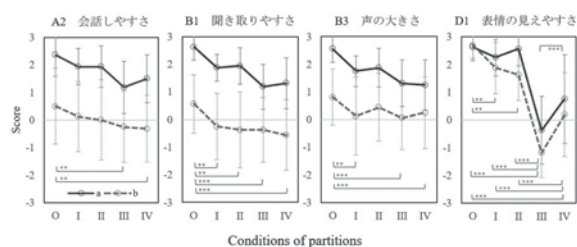
—光反射性・遮音性の影響について—

Psychological experiment on impression of conversation across desktop partitions – Effect of light reflection and sound insulation

☆中島美百合, 佐久間哲哉(東大・工),

△小崎美希(東大・新領域), △松山麻珠, △酒井美希(大日本印刷)

- ◆近年の感染症流行に伴い普及する卓上透明衝立については、衝立面への反射光の映り込みと衝立による減音が会話のしやすさに影響を及ぼす可能性が指摘されている。
- ◆本稿では、光反射性・遮音性の異なる衝立5種類と背景音2条件を設定し、2人1組での会話を模擬した心理評価実験を行い、視・聴覚的側面から会話しやすさへの影響を検討した。
- ◆特に暗騒音の小さい条件では、会話・聞き取りのしやすさだけでなく、声の大きさに関する項目でも、遮音性に関わらず光反射率の低い面材で評価が高くなる傾向がみられた。

Fig. 1 Average scores and two-sided 95% confidence intervals (**: $p < 0.01$, ***: $p < 0.001$).

2-12-1

2-12-1 リモート会議の音声伝送と室の響きの関係についての実験的検討

Experimental examination about the speech transmission of remote meeting and the reverberations in the rooms

○土屋裕造, 小泉穂高(戸田建設)

- ◆近年リモートによる会議方法が浸透し、社会に欠かせない手段となっている一方で、同じ室で、対面の会議は聞き取れるが、リモートの会議は聞き取りにくいといった事象が起きている。音声空間を伝搬する場合は、室内音響の影響も大きい。
- ◆そこで、リモート会議を行う上で、室の響きはどの程度が適当か、実験的に検討した。
- ◆その結果、リモート会議を行う室は、最適残響時間と室容積との関係の、講演・話をする部屋の線を目標に残響時間を設定すればよい可能性が示唆された。

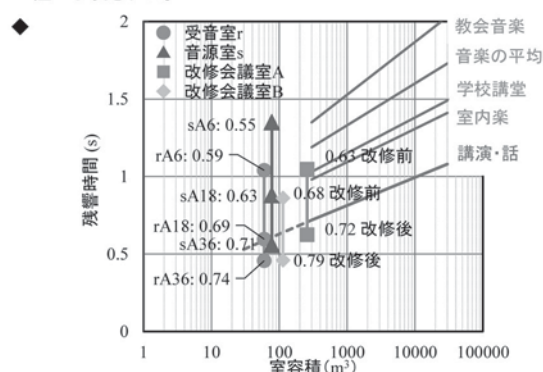


Fig.:用途に応じた最適残響時間(500 Hz)と室容積との関係 STI併記

2-12-3

2-12-3 障壁がある空間におけるスピーチプライバシー簡易予測

Prediction of speech privacy for a room with a barrier using statistical energy theory

○増田 潔(大成建設技術センター)

- ◆設計・計画段階において、間仕切り壁から透過する会話音のスピーチプライバシー評価量を簡易的に予測する手法として、筆者は「拡散音場を仮定した幾何音響理論」のみで STI の予測計算が可能な SNR_{stat} を用いる方法を提案している。
- ◆本報では上記手法を応用して、室内に障壁がある場合のスピーチプライバシー評価指標の簡易予測式を導出した。
- ◆開口から透過する初期反射音や回折音を明瞭度に寄与する成分として、後部残響音および暗騒音を明瞭度の阻害成分として計算できる予測式を「拡散音場を仮定した幾何音響理論」のみで導出した。
- ◆上記予測式による予測値と障壁で区切られた2室間のスピーチプライバシー評価量の実測値は良好な対応を示した (Fig.1)。

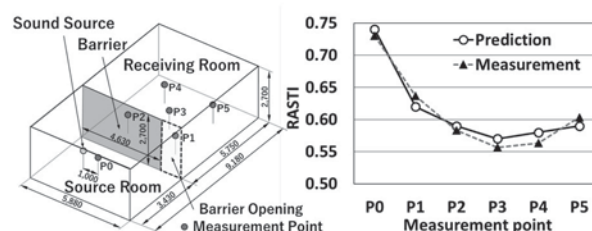


Fig.1: Comparison between prediction and measurement.

2-12-4

2-12-4 オープンプランオフィスにおけるパーソナルブースの音漏れ低減方法の検討

Study of methods to reduce sound leakage from personal booths in an open plane office.

◎清家 裕喜子, 石塚 崇(清水建設 技研)

- ◆Web 会議での利用を想定した、反射性のパーティションで構成されたパーソナルブースの音漏れ低減方法を検討するため、実際のオープンプランオフィスに配置されたパーソナルブースに対策を講じて、その効果の測定を行った。
- ◆ブース内の壁面・天井下面の吸音処理は、音漏れ対策として有効であることが確認された。
- ◆一方、ブース上部開口を一部塞いだ場合、ブース内の音圧レベルが上昇してパーティションからの透過音が大きくなるため、隣接するブースでは効果が小さいことが明らかになった。

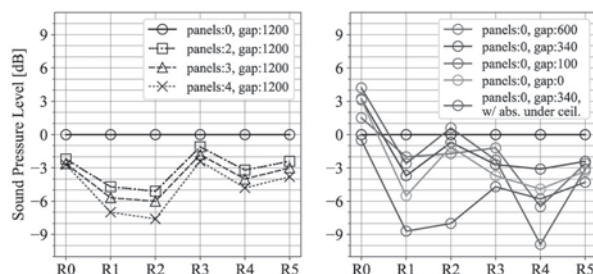


Fig.1: Sound pressure levels in the 500 Hz band relative to those of the case applying no sound leakage reduction methods.

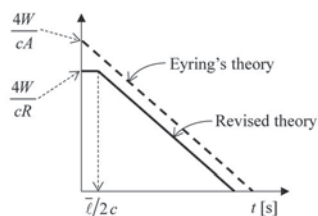
2-12-6

2-12-6 室内インパルス応答の確率統計モデルと室内音圧レベルの修正理論

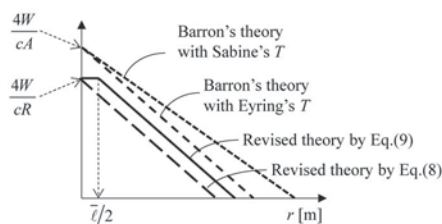
Stochastic model of a room impulse response and revised theory of sound pressure level in a room

◎羽入敏樹(日大・短大)

- ◆直接音の平均自由行程が反射音の1/2になることを前提に、Eyringの残響理論の修正理論を提案した。さらにこれを基に、残響音レベルの距離減衰に関するBarronの理論についての修正理論を提案した。



Conceptual diagram of differences between Eyring's theory and the revised theory with respect to reverberation energy decay



Conceptual diagram of differences between Barron's theory and the revised theories with respect to sound pressure level

2-12-5

逆解析を用いたインパルス応答の連結成分可視化

Visualization of connected components of impulse responses by using inverse analysis.

○太刀岡 勇気(デンソーアイティラボラトリ)

- ◆インパルス応答の細かな構造の差に着目するため、パーシステントホモロジーを用いた解析に注目している
- ◆既報ではパーシステンス図を使ってエコーの有無がそれらに与える影響を調査したが本報ではその際の連結成分の可視化を行った

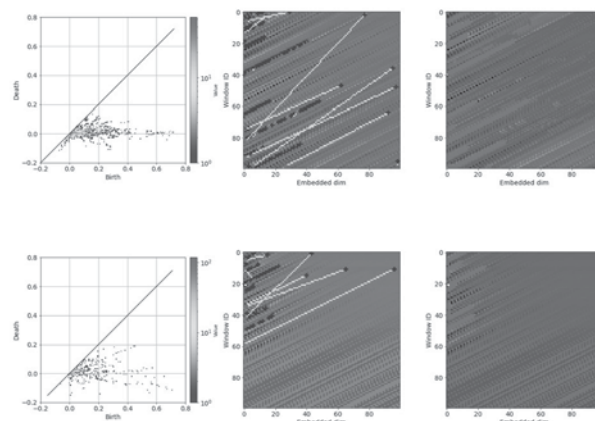


Fig.: Persistence diagrams of impulse responses and their inverse analysis results observed in a room with many echoes (upper) and less echoes (lower).

2-12-7

2-12-7 クラリティ C_{50} の理論式の適用性に関する研究—幾何音響解析による検討—

Study on applicability of the theoretical equation for Clarity C_{50} using geometric acoustic analysis

☆陳科吉(東大・工), 廣瀬量子(日本環境アメニティ), 佐久間哲哉(東大・工)

- ◆拡散場の仮定の下で音声明瞭度 C_{50} の理論式が示されているが、各種の室条件においてその適用性まだ不明確である。
- ◆本研究では、幾何音響解析に基づき、室容積、音源・受音点配置および吸音分布により、 C_{50} の理論値と解析値の対応を分析し、適用性を検討する。
- ◆ C_{50} の解析値は10 m点において理論値と概ね対応しているが、2.5, 5 m点では全体的に5 dB程度高く、吸音率が低いほど乖離が大きい。吸音偏在の影響としては、天井が高吸音性となる平均吸音率0.3の場合、小規模空間では C_{50} の解析値は低下するのに対して、大規模空間では逆に上昇する傾向が見られる。
- ◆ C_{50} の理論式の適用性に関しては、一部の条件で過大評価が生じたものの、最大2 dB程度であり、全体的には安全側の評価となることが確認された。

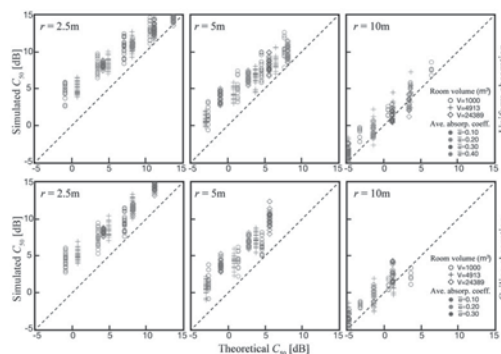


Fig. 1: Theoretical and simulated C_{50} under two conditions of absorption.

2-12-8

2-12-8 矩形室における残響時間測定事例から考察する
室内音響予測の注意点についてCaution for room acoustics prediction method studied by a measurement
example of reverberation time in the rectangular room

○小泉 穂高, 土屋 裕造 (戸田建設)

- ◆天井のみが吸音された矩形室において残響時間を測定した。可動間仕切の有無によって室容積を変更可能であり、机椅子の設置数との組み合わせを測定条件とした。その結果、残響時間は完全拡散音場を前提とした場合の予測値とは大きく乖離した。また、設置した机椅子は単純な付加吸音としては機能せず、拡散体としての役割が大きいことが考察された。

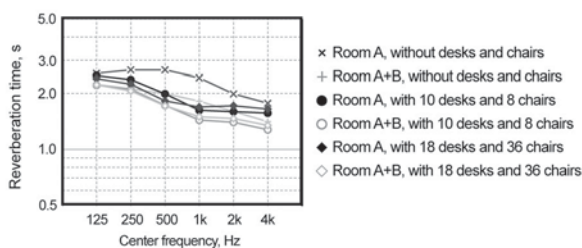


Figure: Reverberation time in the rectangular room

2-12-9

2-12-9 3次元ボロノイ分割を利用した
音響拡散体の有限要素法解析による検討Examination by finite element method analysis of acoustic diffuser
using 3D Voronoi decomposition

☆濱谷晃希, 菅原彬子, 平栗靖浩(近畿大)

- ◆周期性を持たない非晶質構造を音響拡散体を利用することで、広帯域で、無指向性に近い拡散特性を得られる可能性がある。
- ◆非晶質構造をボロノイ分割という手法を用いて作成した。また、試料に対して幾何パラメータより、形状を評価する。(Fig.1)
- ◆有限要素法を用いた数値解析より、拡散係数を算出する。試料を90度ずつ回転させることで4方向からの解析を行う。
- ◆高周波数域ほど拡散係数は高くなるが、低域においてもある程度の性能を有することが確認できた。また、同じ厚みで多様な拡散特性を実現できる可能性が示唆された。(Fig.2)

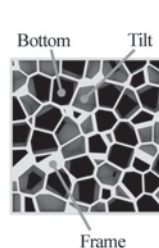


Fig.1: Each element that constitutes an amorphous structure

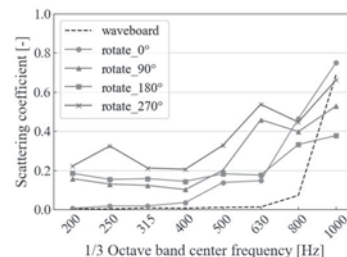


Fig.2: Scattering coefficients calculated From numerical analysis

2-12-10

2-12-11

2-12-11 クラウドコンピューティングを用いた
並列陽的 FEM による室内音響解析A parallel explicit time-domain FEM
with cloud computing for room acoustics simulation

◎吉田卓彌(安藤ハザマ技研, 神戸大院), 奥園健, 阪上公博(神戸大院)

- ◆Amazon Web Service (AWS)のクラウド HPC 環境における並列陽的 TD-FEM の性能を検証する。
- ◆6144 コアまで使用した大規模並列に対するスケーラビリティを2種の実用的な建築音響問題の解析を通して調査した。
- ◆並列陽的 TD-FEM は6144 コアを用いたプロセス並列により自由度数が3千万、ならびに、1億の大規模室内音場の6 kHz までの周波数成分を含む1 s 間のインパルス応答をそれぞれ217 s, 613 s と超高速に計算可能であった。
- ◆主要演算部である時間進行スキーム計算部については並列ノード数と共にリニアに計算時間が短縮されており、優れたスケーラビリティが確認できた。

講演取消

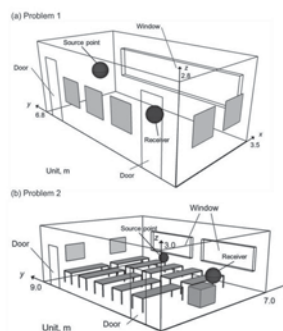


Fig.1 Two practical architectural acoustics problems: (a) Problem 1 and (b) Problem 2.

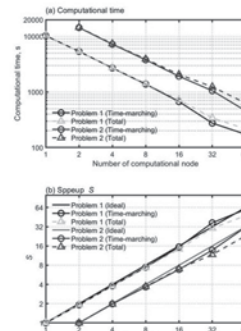


Fig.2 The results of scalability test (a) computational time and (b) speedup S.

2-12-12

2-12-12 FMBEM 入射指向性解析における
近傍要素の寄与計算Calculation of the contribution from near elements
in FMBEM incidence directivity analysis

☆萩原諒(東大・工), 榎本貴之(サイバネット),

佐久間哲哉(東大・工), 安田洋介(神奈川大・建築)

- ◆筆者らが提案してきた入射指向性の計算手法では、FMBEMにおける受音セルの内向係数が用いられている。
- ◆一方、受音セルの内向係数には近傍要素の寄与を含められないため、受音セルが境界面に近接する際、近傍要素についての計算が別途必要となる。本報告では、その近傍要素の寄与計算手法を示すとともに、室内音場への適用を試みた。
- ◆近傍要素からの寄与を、境界上のモノポール・ダイポール音源からのセル中心点における局所展開係数として計算した。
- ◆室内音場へ適用すると、受音セルが大きくなり周波数が高くなるほど、指向性が鋭く多方向に現れるようになった。
- ◆近傍要素を含む大きなセルでは、小さなセルの際に比べて指向性のピークの方向が異なるなどの特徴が見られた。

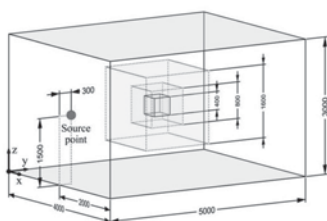


Fig. 1: Analysis model.

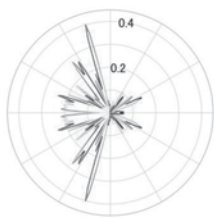


Fig. 2: Incidence directivity.

2-12-13

2-12-13 FDTD法の音源について

On sound source of FDTD method

○豊田政弘(関西大), 矢田部浩平(農工大)

- FDTD法の音源について、音源の条件を解析に導入する方法には、空間的な音圧分布、ならびに、粒子速度分布を与える方法と、音源となる点に時間波形を与える方法の2種類がある。
- 後者の方法には、ハード音源、ソフト音源、トランスペアレント音源の3種があり、それぞれに利点・欠点がある。
- インパルス応答を算出したい場合に最も有効なのがトランスペアレント音源であるが、既存の手法は煩雑であり、計算負荷も大きい。
- 本報では、FDTD法における簡便なトランスペアレント音源の導入方法を提案する。
- Fig. 1に本報で提案する音源セルのモデルを示す。このモデルに従って、定式化を行い、音圧更新式に音源項を導入する。
- 自由空間を対象とした検討により、提案手法の有効性を示す。

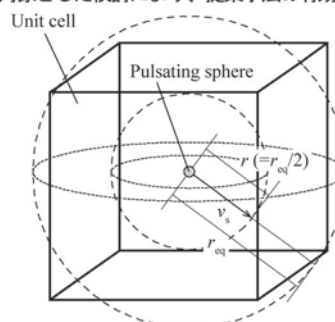


Fig. 1 Suggested source model.

2-P-1

2-P-1 音響イベント検出を用いた超音波通信
システムの雑音下における性能評価Performance evaluation under noise of ultrasonic communication system
using sound event detection

○尾岡剛成, 青木直史, 土橋宜典(北大)

- ◆音波通信は、電波通信に比べると雑音やドップラー効果、エコーなどの影響によって、信号検出精度が劣化する特性がある。
- ◆本研究では、非可聴域の超音波通信システムにおいて、機械学習に基づく音響イベント検出を用い、雑音環境下における信号検出精度の評価を行う。

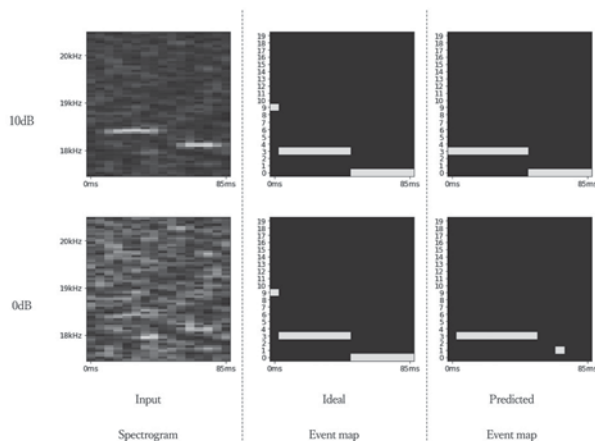


Fig. 1: Results of sound event detection of ultrasonic signals

2-P-2

2-P-2 Rayleigh 散乱と Brillouin 散乱による
培養液の評価

Evaluation of culture medium by Rayleigh and Brillouin scatterings

☆稲本脩人, 斎藤駿, 橋川亮平, 小山大介, 松川真美(同志社大)

- ◆ Brillouin 散乱法では、媒質中の熱揺らぎである熱フォノンとフォノンの非弾性衝突による光散乱を計測する。
- ◆ Brillouin 散乱と Rayleigh 散乱の積分強度から、媒質の比熱比 γ にかかわる量である Landau-Placzek (L-P) 比を求めることができる。
- ◆ 既存の方法では計測時のセルや光学系内のレンズなどからの反射や散乱の影響で、散乱光の正確な測定が難しい。そこでセルなしの簡易な構造で培養液や細胞懸濁液の測定を試みた。(Fig. 1(a))
- ◆ セルなし構造を用いて培養液と HeLa 細胞の L-P 比を観測した(Fig. 1(b))。培養液は細胞密度がゼロの場合に相当する。本結果より、細胞密度の増加に伴い L-P 比が増大したことから HeLa 細胞の L-P 比は培養液より大きい可能性が示唆された。

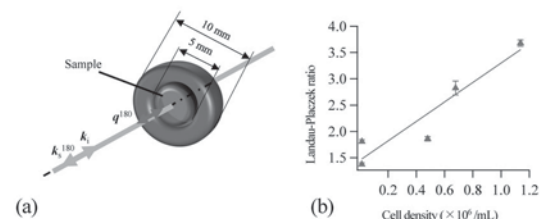


Fig. 1(a) Back (180°) scattering geometry with a simple cell and (b) relationship between L-P ratio and cell density of HeLa cell solutions.

2-P-3

2-P-3 光学式機械物性計測のための光干渉を用いた光音響波検出システムの構築

Development of photoacoustic signal detection system by an optical interferometer

○田村和輝(浜松医大), △橋本研也(電子科大), △大川晋平(浜松医大)

- ◆細胞の種類・炎症の有無や細胞周期の段階に応じて、体積弾性率に関連する機械物性が変化する。
- ◆我々のグループは測定速度と空間分解能を向上させることを目標に、完全光学式の機械物性計測法を提案している。
- ◆ポリスチレン水槽の内側底面に対してナノ秒パルスレーザーを照射し厚み方向に光音響波を伝搬させた。(Fig. 1 (a))
- ◆底面外側に到達した光音響波を表面変位として偏光マイケルソン干渉計を使って計測した。ポリスチレン水槽の厚み(2.5 mm)と縦波伝搬速度 2330 m/s に対応した時間に表面変位を計測した。(Fig. 1 (b))
- ◆パルスレーザーを用いて発生させた光音響波がポリスチレンの厚み方向に伝搬し、表面変位として検出可能なことが確認された。

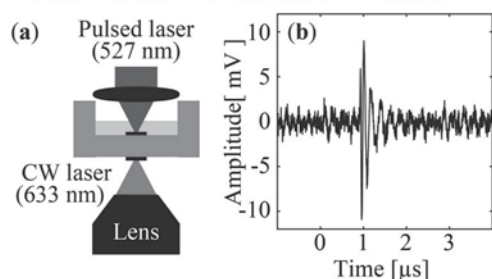


Fig. 1: (a) Schematic illustration of the experiment. The tank filled degas water. (b) Measured interferometry signals

2-P-5

Si, Ge 添加による AlN 薄膜の極性制御と極性反転薄膜音響共振子への応用

Polarity control of AlN films by Si, Ge doping and applications for polarity inverted film bulk acoustic wave resonators

○関本淳, 鈴木雅視, 垣尾省司(山梨大)

薄膜音響共振子(Bulk acoustic wave 共振子: BAW 共振子)は今後さらなる高周波化が予想される。しかし、高周波動作する BAW 共振子は薄い圧電膜厚による共振特性の劣化が懸念される。一方で極性反転構造 BAW 共振子は高次モードで共振するため、高周波動作における問題を解決する可能性がある。本研究では AlN 薄膜への Ge, Si 添加による極性制御、結晶配向性、電気機械結合係数(k_t^2)に及ぼす影響を調査した。また逆方向の極性を持つ c 軸配向 AlN 薄膜を交互に成膜することで極性反転構造多層膜を作製した。Fig.1 に示すように作製した 8 層極性反転 SiAlN (GeAlN) / AlN 薄膜共振子は 8 次モードで共振することを確認した。

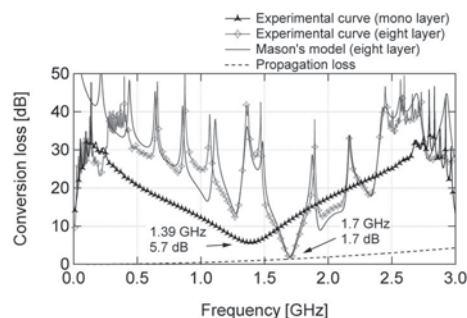


Fig. 1: Frequency characteristics of longitudinal conversion loss of the eight-layer polarity inverted SiAlN/AlN HBAR.

2-P-4

2-P-4 液体充填弾性細管の光音響信号の放射指向性

Directivity of photoacoustic signals in a liquid-filled elastic tube

○Wang Kun, 和田有司, 中村健太郎(東工大)

- 外径 1.0 mm の細管に黒インクを充填し、半導体レーザー(波長 637 nm, 出力 1300 mW) によるパルス光 (パルス幅 230 ns) を照射し、中心周波数 5 MHz の超音波トランスデューサで光音響信号を受信した。
- 細管には内径 0.75 mm のガラスキャピラリーと内径 0.50 mm のシリコンチューブを用い、ソフトファントム内に埋め込んだ。
- ガラスキャピラリーでは管内に共振モードが励振され、光音響信号はモード形状に依存した顕著な指向性を示すことがわかった。
- シリコンチューブでは強い共振は起きず、ブロードな指向性となった。
- これらの実験結果をほぼ再現するレーリー積分による解析や有限要素法解析を行った。

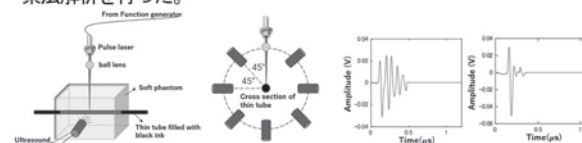


Fig. 1 Experimental setup.

Fig. 2 Photoacoustic signals for glass capillary (left) and silicone tube (right).

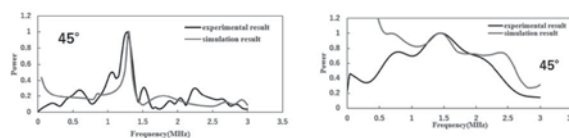


Fig. 3 Frequency characteristics of photoacoustic signals for glass capillary (left) and silicone tube (right) in the direction of 45 degrees.

2-P-6

2-P-6 The Characteristics of Surface Acoustic Wave Atomization with Unidirectional Interdigital Transducer

☆Yimeng Wang¹, Taihei Tsubata¹, Minoru Kurosawa², Manabu Aoyagi¹, Deqing Kong¹ (¹Muroran Institute of Technology; ²Tokyo Institute of Technology)

- ◆The surface acoustic wave (SAW) atomization with a unidirectional interdigital transducer (U-IDT) by 9.59 MHz driving frequency is proposed, which aims at improving the atomization rate and decreasing the size of atomized drops.
- ◆Since R-IDT can reflect the wave towards it in the opposite direction, U-IDT has higher atomization efficiency than those transducers with only D-IDT.
- ◆By using U-IDT, the atomization rate is increased greatly, as shown in Fig.1. Furthermore, the diameters of most of the atomized drops can reach the level of less than 1 micron, as shown in Fig.2.
- ◆SAW atomization with U-IDT is expected to be applied in medical applications such as pulmonary targeted drug delivery.

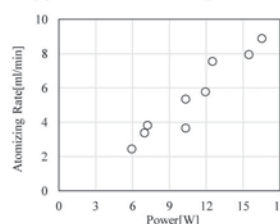


Fig. 1 Atomizing rate vs. driving power

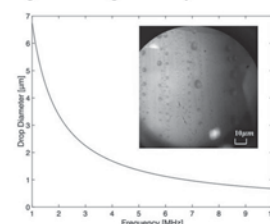


Fig. 2 Numerical analysis and observation of the atomized drop diameter

2-P-7

2-P-7 サファイアスライダを用いた弾性表面波モータの出力特性の検討

Characteristics of surface acoustic wave motor using sapphire slider

○孔徳卿, 雲浩(室蘭工大院), 黒澤実(東工大), 青柳学(室蘭工大院)

◆本研究では、弾性表面波モータの耐久性を改善するため、サファイアスライダを用いた弾性表面波モータの出力特性を検討する。予圧特性により、モータの推力、速度、ステッピング駆動特性を議論し、サファイアスライダの性能を評価する。耐久性を改善した上、高精度ステータのアクチュエータとしての実用化が期待される。

◆Fig. 1 に示すように、サファイアスライダ（5mm×10mm×1.5mm）を作製した。Fig. 2 に示すように、サファイアスライダを用いた弾性表面波アクチュエータを作製した。ステータは一方方向性 IDT 電極を用いた 128 Y-X LiNbO₃ 基板で作製した。

◆Fig. 3 を示すように、25 N の予圧の場合、100 の Cycle のパルス駆動を設定し、80 V_{pp} の駆動電圧で平均変位 48 nm の高精度ステッピング制御を確認した。

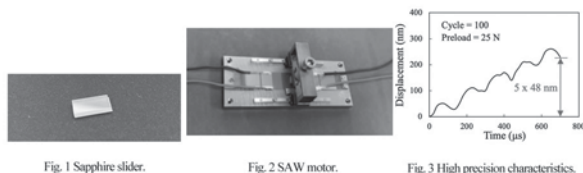


Fig. 1 Sapphire slider.

Fig. 2 SAW motor.

Fig. 3 High precision characteristics.

2-P-9

2-P-9 集束超音波音場における円錐型堅牢ハイドロホンの評価

Evaluation of Conical Tough Hydrophone in Focused Ultrasound Field

○見瀬不二丸(桐蔭横浜大), 椎葉倫久(日本医科学大), 岡田長也(本多電子), 森下武志(桐蔭横浜大), 佐藤敏夫(桐蔭横浜大), 竹内真一(東京都市大/桐蔭横浜大)

◆これまでに私たちの研究グループは前面板の厚さが 3 mm の堅牢ハイドロホンの評価してきた。このハイドロホンは超音波洗浄器に 100 時間暴露しても受波感度が低下しないことがわかっている。

◆今回の実験では凹面振動子を用いた集束超音波照射システムによる音場に堅牢ハイドロホンを暴露し、堅牢性の評価及び音響キャビテーションが及ぼす影響について検討した。

◆堅牢ハイドロホンを集束超音波の音場に 30 時間暴露しても受波感度が低下することはなかった (Fig.1)。また、暴露実験中ではハイドロホンの先端近傍にキャビテーションバブルが生じており、その挙動を観察した。

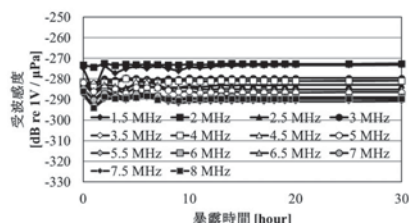


Fig. 1 集束超音波音場における円錐型堅牢ハイドロホンの堅牢性評価

2-P-8

2-P-8 堅牢ハイドロホンの感度校正を目的とした電極分割振動子型

高強度超音波音源システムの検討

Investigation of an Electrode Divided Transducer Type High Intensity Ultrasound Source System for Calibration of Sensitivity of Tough Hydrophones

○小林篤史(桐蔭大院), 五十嵐茂(職能大), 佐藤敏夫(桐蔭大), 竹内真一(都市大)

◆ハイドロホンの感度校正を実際の治療で用いられる音響強度で行うためには高強度な平面波が必要とされる。

◆先行研究では、球面型振動子の焦点に円筒状音響導波路を設置し、その球面型振動子の面積を大きくすることで、高強度な平面波を照射する音源システムを提案した。しかし、大面積化に伴う入力インピーダンス低下の影響が音響強度が確保できなかった。

◆本研究では、実効直径 100 mm の球面型振動子の電極を分割し、1 つ 1 つの電極片を独立した送信回路で駆動することで、大面積化と入力インピーダンスの確保を実現する音源を提案し、その照射音圧や入力インピーダンスについて、Mason の等価回路に基づく数値計算を行った。

◆本研究で提案した電極分割振動子型音源の入力インピーダンスを送信回路の出力インピーダンスに整合させて投入電力を最大にするよりも、球面型振動子の電極を細かく分割して入力インピーダンスを大きくすることにより印加電圧を増加させる方が高強度な超音波を照射できることが示唆された。

2-P-10

2-P-10 咬合力計への応用を目的とする安全性を考慮した音響式口腔内用圧力センサの検討

Examination of a safety-oriented acoustic intra-oral pressure sensor for application in bite force gauges

☆星野拓真, 田原麻梨江(東工大)

◆感圧部に金属を含まず、口腔内で使用しても通電リスクがない咬合力計として、チューブ内の音波伝搬特性を利用した圧力センサを提案した (Fig. 1)。

◆5種類のチューブの中で最も加圧材料として適していたウレタンチューブ (Fig. 2) を用いて音響式圧力センサを構築した。100 ~ 1000 N までの加圧実験を行い、800 N までは出力が増加することがわかった (Fig. 3)。

◆音響式圧力センサと市販の圧抵抗効果式圧力センサで、噛み締め試験を行なった。最大咬合力で 3 秒間噛み締めて得た出力から、圧力を推定すると、音響式と圧抵抗効果式とで結果が一致し、咬合力計としての応用可能性を示した。

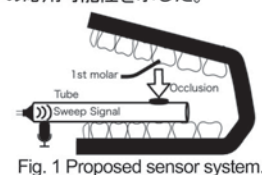


Fig. 1 Proposed sensor system.

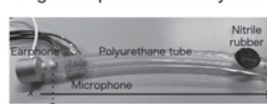


Fig. 2 Configuration of experiments.

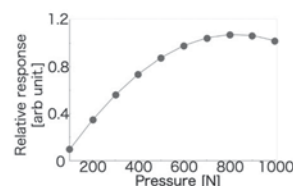


Fig. 3 Relationship between output and pressure.

2-P-11

2-P-11 空中超音波を用いた青果物の非接触弾性計測における経時変化の検討

Study of change over time in non-contact elasticity measurement of fruits and vegetables using airborne ultrasound waves

☆池田将基, 田原麻梨江(東工大)

- ◆広帯域超音波振動子を用いた空中超音波システムにより 9 種類の青果物(Table 1)の複数試料(10 個)を用いた非接触弾性計測を行った。
- ◆測定で得た振動速度波形から、時間経過とともに振動振幅が大きくなり、また、周期が長くなることが分かった (Fig. 1)。

Table 1 Measurement samples.

種類	測定頻度	測定回数
ブドウ	1 日毎	5 回
さくらんぼ	1 日毎	4 回
桃	1 日毎	4 回
イチゴ	1 日毎	4 回
バナナ	1 日毎	5 回
リンゴ	7 日毎	5 回
レモン	7 日毎	5 回
マンゴー	4 日毎	4 回
キウイ	4 日毎	4 回

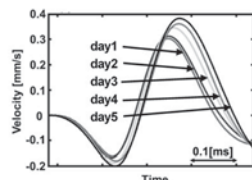
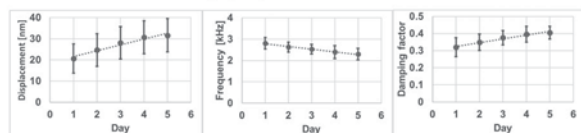


Fig. 1 Results of waveform of vibration velocity (Grape).

- ◆青果物の弾性パラメータの解析の結果、ほとんどの試料で時間経過とともに最大変位と減衰係数は増加傾向であり、中心周波数は減少傾向であることが分かった (Fig. 2)。



(a) Maximum displacement (b) Center frequency (c) Damping factor

Fig. 2 Results of elastic properties of fruits.

2-P-13

2-P-13 超音波を用いた下肢浮腫の定量診断に向けた研究
—散乱体の音響インピーダンスの変化が与える影響についての基礎検討—Study for Quantitative Diagnosis of Lower Limb Edema Using Ultrasound
-Fundamental Study on the Influence of the Change of Acoustic Impedance of Scatterers-

☆神戸大希, 坂本真一, 越山雅文, 生田実里, 渡邊友美子(滋賀県立大)

- ◆重病の症状の一つである浮腫の定量診断の確立が急務である。
- ◆浮腫は進行に伴って水分量が増加し組織構造が変化するため、超音波伝搬波形に変化が生じる。
- ◆浮腫の超音波伝搬波形の周波数特性から定量診断を試みる。
- ◆基礎検討として、ガラス(GP)およびアルミナ(AP)の球状散乱体を用いてファントムを製作し、超音波伝搬波形の周波数特性を算出した。
- ◆数値計算では FDTD 法を用いて実験同様のファントムモデルを作成し、超音波伝搬波形の周波数特性の解析を行い、散乱体の音響インピーダンスの違いが周波数特性に与える影響を検討した。
- ◆音響インピーダンスが大きい散乱体を含む媒質の超音波伝搬波形の周波数特性は実験および解析とともに振幅が増幅した (Fig. 1)。

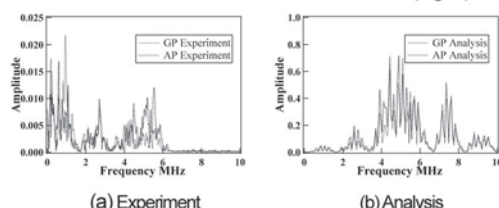


Fig. 1 Frequency characteristics of ultrasound propagation waveforms.

2-P-12

2-P-12 ソノルミネッセンス強度と音響スペクトルの初期成長過程におけるアルコール添加の影響

Influence of alcohol addition in sonoluminescence intensity and acoustic spectrum in the initial process of bubble's growth

☆中村 悠人, 山本健(関西大)

- ◆マルチバブルソノルミネッセンス(MBSL)の現象について検討する際、気泡のサイズ及びその成長速度といった気泡ダイナミクスは重要な要素である。
- ◆本研究ではアルコールの添加における気泡成長速度の影響を光と音の観点から検討した。
- ◆アルコール(エタノール、1-ヘキサノール)を添加し、超音波照射開始直後の MBSL 強度の時間変化を測定し、発光強度が定常状態になる時間を比較した。
- ◆気泡成長の詳細なメカニズムを検討するために、アルコールを添加し、音響スペクトルを測定した。
- ◆光と音の観点からアルコールを添加することで気泡合体が抑制され気泡成長速度が低下することが示唆された。

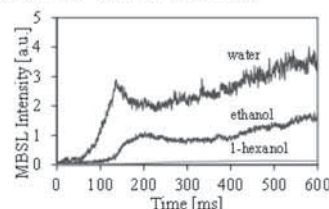


Fig. 2 Result for time change in the MBSL intensity of water, ethanol aqueous solution, and 1-hexanol aqueous solution

2-P-14

2-P-14 血管変形モデルシミュレーションによる狭窄部壁せん断応力の超音波計測の精度評価

Evaluation of accuracy of ultrasonic measurement of wall shear stress at stenosis based on simulation of vascular deformation model.

◎茂澄倫也, 大村真朗, 長岡 亮, 長谷川英之(富山大)

- ◆狭窄部近傍での流速分布について、非対称的な分布にフィッティングする信号処理手法を提案した。その検証のために、狭窄部を有するシミュレーションファントムを有限要素法に基づき作成し、WSS 超音波計測時の精度評価を行った。
- ◆円筒管内を流れる血液流れを模擬し流路経路途中に本研究報告ではブラーク径を 1 mm、または 2 mm に変化させ、WSS 推定時の狭窄厚みの影響を検証した。
- ◆図 1、表 1 に提案法を適用することで得られる速度断面および WSS の偏り誤差を示す。図 1 の通り、提案法による速度断面の壁近傍での傾きは真値の傾きに近づいている。また表 1 の通り、WSS の偏り誤差を 30 %未満に抑えることができた。

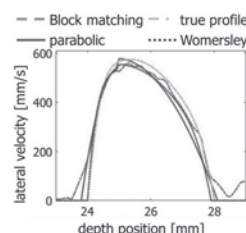


Fig. 1 Velocity profile obtained in simulation at a plaque thickness of 2 mm.

(a)	parabolic	Womersley
WSS (near wall)	-40.0 %	-4.4 %
WSS (far wall)	-26.1 %	-16.1 %

(b)	parabolic	Womersley
WSS (near wall)	-39.7 %	-16.3 %
WSS (far wall)	-53.0 %	-23.3 %

Table. 1 Bias errors of WSS at plaque thicknesses of (a) 1 mm and (b) 2 mm.

2-P-15

2-P-15 骨増生のあるウマ第3中足骨モデル中の
音波伝搬

Wave propagation in third metatarsal bone with bucked shin

☆津幡大聖¹, 鈴木英寿¹,
△千葉恒², △三田宇宙³, △田村周久³, 松川真美¹
(1. 同志社大, 2. 長崎大, 3. JRA 総研)

- ◆弾性 FDTD 法による 2 次元シミュレーションを用いて、馬のトレーニングなどで生じる骨増生の評価への Axial Transmission 法の適用可能性を検討した。
- ◆競走馬の第三中足骨の CT 画像から、小さな骨増生部の突起高さが 0 % から 100 % まで異なるモデルを作成した。
- ◆骨表面の長軸方向にアレイ状受波器を設置し、隣接する各受波器の音波の到達時間差 Δt から長軸方向に伝搬する波のみかけの音速を算出し (Fig. 1(a)) みかけの音速の標準偏差を算出した (Fig. 1(b))。
- ◆骨増生部が高いと偏差も大きくなり 10% でも変化を確認した。音速のばらつき評価で、簡易に骨増生部を検出できる可能性がある。

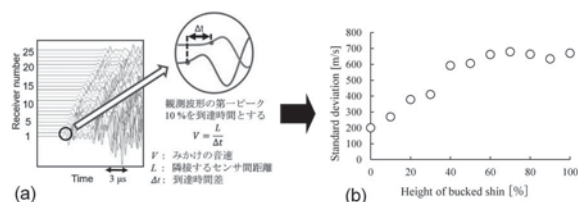


Fig.1(a) Observed waveforms in simulation and (b) velocity dispersions of wave fronts as a function of the height of bucked shin.

2-P-17

酵母菌増殖促進を目的とした
レーザ超音波の連続発生に関する研究
Study on continuous generation of laser ultrasound
in order to promote yeast proliferation

○宮崎晃汰(金沢工大院), 會澤康治(金沢工大院)

- ◆パルス超音波照射による酵母菌増殖効果を確認する。
- ◆酵母にレーザ超音波のパルス連続照射を可能にするターゲット材料を検討する。具体的にはレーザアブレーションによるターゲット材料の不可逆的な破壊に伴う音圧の変動を解決する。
- ◆結論として、黒色インクをプラスチックのチューブで循環させる液体ターゲットを使った方法は、2 MPa 以上の音圧を 100 発キューベツ内に導入できるので、酵母菌増殖促進実験において効果的であると考えられる。

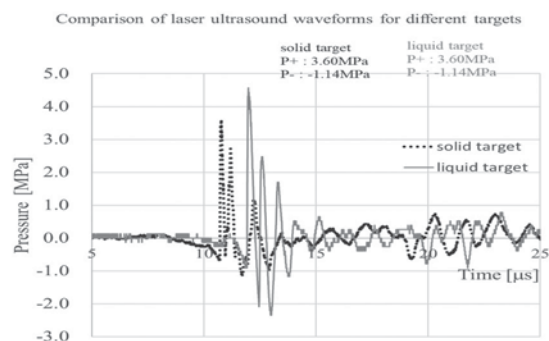


Fig.1: Comparison of laser ultrasound waveforms for different targets (dot line: solid target, solid line: liquid target)

2-P-16

2-P-16 橈骨の誘導電位に及ぼす超音波入射角度
の影響

Effect of incident angle of ultrasound on induced potentials in the radius

☆鈴木英寿, 津幡大聖, △前原溪吾(同志社大), 細川篤(明石高専),
土屋隆生, 松川真美(同志社大)

- ◆超音波骨折治療法の効率化を目指し、骨中に発生する電界と超音波入射角度の関係をシミュレーションにより検討した。
- ◆66 歳女性の橈骨の CT データから 2 次元モデルを作成した。入射角度を骨表面に対して 90° あるいは 45° としてシミュレーションを行った (Fig. 1(a))。
- ◆骨中の電界を算出して厚さ(Y)方向の電界強度を求めたところ、入射角度 45° の方が最大電界が大きいことがわかった (Fig. 1(b))。
- ◆骨の圧電定数を考慮すると、せん断波の伝搬が骨中の電磁気現象の鍵となる。

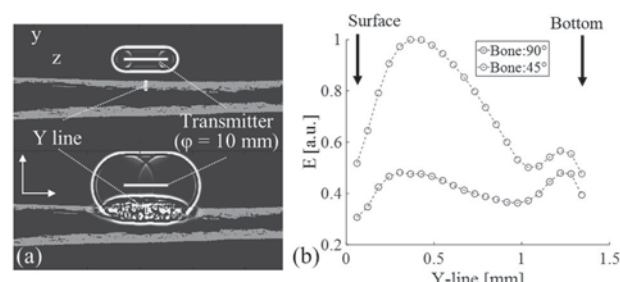


Fig. 1(a) Distribution of the waves radiated from the transmitter to radius. (b) maximum electric field strength values in the thickness (Y) direction.

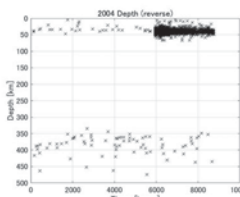
2-P-18

2-P-18 紀伊半島南東沖地震のエネルギー推移

Energy transition of the Kii Peninsula southeast offshore earthquake

○菊池年晃(防大), 水谷孝一(筑波大)

先に、紀伊半島南東沖地震の震源振動構造を解析した。その際、震源深度の構造は 2 分生であることが判明した。2004 年に、この地域で発生した M 1.5 以上の地震の震源深度を下図に示す。縦軸は発生した地震の震源深度を、横軸は元日からの経過時間である。図から明らかのように、震源深度は 50km 以浅と 380km 以深に 2 分化されている。この分布図は観測されている 2002 年より今日に至るまで変わらない。したがって、この 2 分性は紀伊半島の生成に起因した地勢学によるもので、これらの中間に付加体が存在する。この付加体が浅震源域と深震源域とで共有しているので二つの震源域で発生する地震に関連性が生じる。この関連性を調べるために浅震源域と深震源域で発生する地震のエネルギー推移を調べた。その結果、深震源域で付加体に加えられたエネルギーと浅震源域から放射されるエネルギーに関連性があることが明らかになった。上図が示すように、浅震源域では 2004 年 9 月に、M 7.4 の地震が発生している。この時点では浅震源域のエネルギーが大きい、そのエネルギーは余震として放出されていき、2016 年に深震源域からのエネルギーレベルと一致する。その後は深震源域から、即ちプレートからのエネルギーが蓄積され、巨大地震の源が形成されていく。これはプレートの挙動と直結した蓄積エネルギーの監視や地震の予知に活用出来る。



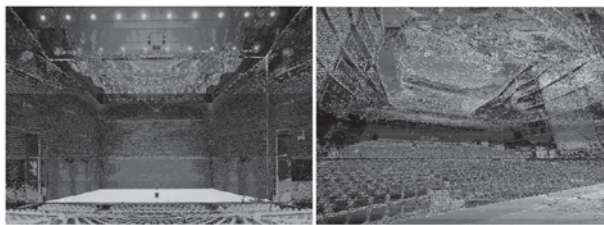
2-Q-1

2-Q-1 あきた芸術劇場「ミルハス」の音響設計
—2000 席規模の多機能ホールにおける音響性能バランスの最適化—

Acoustical Design of the Akita Arts theatre "Mille Has" —Optimization of the acoustic performance balance in the 2000-seat multipurpose hall—

○中川貴美子, 宮崎秀生(ヤマハ), 岸永伸二

- ◆あきた芸術劇場ミルハスは高い音響性能と舞台機能を併せ持つ大ホール(2007 席)と臨場感を重視した中ホール(800 席)をはじめ、二つの小ホール、練習室等を備えた複合施設である。本報では 2000 席規模(客席 2 層)の大ホールにおいて、音響上の課題とそれを解決するために行った音響設計を中心に報告する。
- ◆「残響感と音量感・明瞭性のバランス」「2 階席後方やバルコニー下の音量感・明瞭性の低下」「1 階席中央部の音量感・拡がり感不足」という課題に対し、「残響感の最適化」「天井からの初期反射音の最適化」「音量感と拡がり感の最適化」の 3 つの観点から室形状・内装仕様を検討し、どの席でも豊かな響きを感じられるとともに細かい演奏音まで聴き取れる明瞭性のある空間を実現した。



2-Q-3

講演取消

2-Q-2

桜美林芸術文化ホールの音響設計

Acoustical Design of J. F. Oberlin Hall of Arts and Culture

○竹林 健一(鹿島技研), 箱崎 文子, 池田 寛(永田音響設計)

- ◆桜美林芸術文化ホールではパイプオルガン(29 ストップ, パイプ本数 1,788)が設置された音楽ホール「プロビデンスホール」、演劇・ダンス対応の「ストーンズホール」が整備され、学生の実技や成果発表の場として活用されている。
- ◆両ホール間の遮音性能は 86 dB/500 Hz 以上であった。
- ◆プロビデンスホールは側壁や天井を不均一な折板形状とし、音場の拡散性を高めた。またオルガン演奏や室内楽に対応するため、満席時 1.8 / 2.1 秒(カーテン展開/収納時)と 0.3 秒の可変幅を設けた。
- ◆ストーンズホールは演劇・ダンス対応であり、明瞭性を確保するため低域まで 1 秒程度の残響時間となっている。
- ◆パイプオルガン設置前後の残響時間測定より得られた透過吸音面積(Fig.2 赤線)は設定値や参考値(Fig.2 青線及び黒線)より小さい値であった。

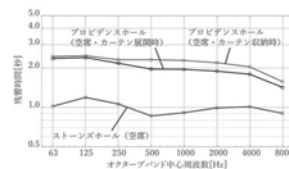


Fig.1 Reverberation time (black line: Providence hall w/ curtain, red line: Providence hall w/o curtain and blue line: Stones hall)

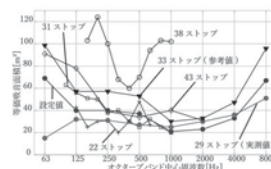


Fig.2 Absorption power of pipe organ (red line: measured, blue line: assumed and black line: referenced)

2-Q-4

さいき城山桜ホールの音響設計

Acoustical design and characteristics of Saiki Shiroyama Sakura Hall

○箱崎文子, 小野朗, 稲生真(永田音響設計)

大分県佐伯市に 2020 年 10 月に開館した「さいき城山桜ホール」は、約 900 席を収容する大ホール、平土間形式の小ホール、スタジオ 5 室をはじめ、多くの機能が組み込まれた施設である。

- 遮音計画: 本施設では、開放的な吹抜空間に面して様々な機能を持つ室が配置されている。各室間の同時使用を出来るだけ可能にするため、大・小ホールや各スタジオの離隔距離を確保した上で、小ホール・各スタジオに防振遮音構造を採用した。

- 大ホールの室内音響計画: 大ホールは、舞台音響反射板や移動観覧席・客席ワゴンなどの転換要素を持つ多目的ホールである。

〔客席天井〕舞台天井反射板から連続した面として見せたいという意匠的な意図と、音響的に望ましい反射面の設置位置への解決策として、客席上部のキャットウォーク下天井面・床面を音響的に透過とみなせる仕様で構成し、それらの約 2.5 m 上部に音響的な反射面を設けた。

〔側壁〕意匠性と壁面の仕上げを散乱形状にしたいという音響的な意図から、下部はルーバー仕上げ、

上部は傾斜壁が採用された。ルーバーはフラッターエコーの防止・規則的なリブ配列による特異音の防止のため、ランダムなリブ配列+リブ間に部分的に埋木をはめ込む仕様とした。

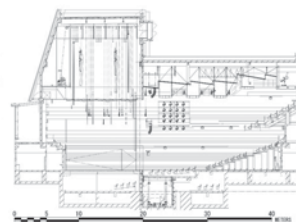


Fig.1 Section of Large Hall

2-Q-5

2-Q-5 みやま市総合市民センター「MIYAMAX」の音響設計 —芸術文化活動から運動利用までを網羅する 多目的ホールの音場最適化について—*

Acoustic Design of the Miyama City Civic Center "MIYAMAX" - Sound field optimization of multipurpose hall covering from arts and culture activities to exercise use -

○日根野翔太, 宮崎秀生(ヤマハ), 岸永伸二

- ◆本施設は、文化・芸術、健康増進、子育て支援の3つの部門を柱として、多目的ホール、リハーサル室、トレーニングルーム、キッズルーム、調理室などから構成される複合施設である。
- ◆多目的ホールは体育館機能を有しており、「ホール機能」と「体育館機能」の性質の異なる空間において、それぞれに最適な音場を実現することを目的として各種検討をおこなった。

Table.1 Performance required for each use

Form	Architectural Performance	Acoustic Performance	Examination Item
Gym (Parquet)	- Ceiling height that does not hit even if the ball flies - Simple wall shape for safety when touching - Wide room width that can be set courts - ease of conversion (rollback and stacking seats) - Protection of speakers in the event of a ball collision	- Ease of communication between players - Line of hearing announcements - Avoiding acoustic obstacle - Presence when watching a game	- Balance between reflective and absorbing surface - Ensuring reverberation adjustment elements (curtain) - Optimization of loudspeakers position, configuration, adjustment - Diffusion by wall and ceiling
Hall	- With orchestra shell - Easy-to-see audience space, appropriate stage floor - Diverting the architectural performance required in the gym	- Appropriate reverberation - Ensuring a sense of loudness - Sense of unity between stage and audience - Ensuring a sense of wideness - Avoiding acoustic obstacle - Easy to play stage space	- Adjust the absorption surface to prevent excessive reverberation - Continuity between stage and audience - Ensuring direct sound by arranging audience seats - Correspondence to reduction of reflected sound by high ceiling and wide side wall (setting of reflective surface) - Ensuring side reflected sound effect of acoustic waves - Diffusion by wall and ceiling - Optimization of stage shape
	Without orchestra shell	- Uniform and clear loudspeaker - Suppression of reverberation that does not impair clarity - Ease of talking on stage	- Ensuring reverberation adjustment elements (stage, curtain) - Ensuring early reflected sound - Optimization of loudspeakers position, configuration, adjustment

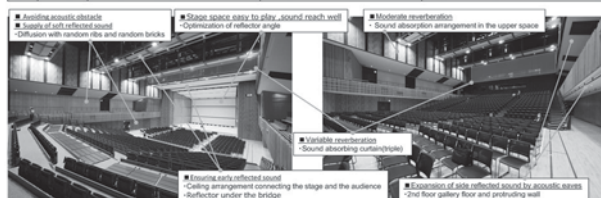


Fig.1 Acoustical Concept

2-Q-7

2-Q-7 高崎芸術劇場の音響設計

Acoustical Design of Takasaki City Theatre

○服部暢彦, 稲生真, 小口恵司(永田音響設計)

- ◆高崎芸術劇場は、大劇場（2027席）、スタジオシアター、音楽ホール（412席）、スタジオ（9室）からなる大規模な複合施設である。本稿では、施設内の遮音計画と大劇場、スタジオシアターの室内音響計画について報告する。

◆遮音計画

- スタジオシアターは、スタンディングのロックコンサートを主用途としており、120 dB 近い発生音が想定された。
- スタジオシアターに近接する大劇場や音楽ホールとの間に高い遮音性能を確保するため、スタジオシアターを別棟として、Exp.J によって大劇場との遮音性能を高めた。
- スタジオシアター、音楽ホールは防振遮音構造とし、躯体側からなるべく防振支持点をとらないような遮音構造を計画した。

◆室内音響計画

- 大劇場は、舞台音響反射板設置時に舞台と客席が一体の空間となるように、舞台音響反射板と客席の天井・側壁が、一つの大きな反射面として、段差なく滑らかに繋がるような室形状とした。
- 音楽ホールでは、音響的に透過するルーバー仕上の背面に音響的な反射面を設けた。ルーバーの裏側には、残響可変のために、昇降式の吸音カーテンを計画した。

2-Q-6

2-Q-6 House of Music Hungary の音響設計

Acoustical Design of House of Music Hungary

○鈴木航輔, 服部暢彦, 小口恵司(永田音響設計)

- ◆House of Music Hungary はブダペストに建設された音楽博物館で、2022年1月に開館した。展示機能の他に、コンサートホール(320席)、レクチャーホール(120席)、教室、図書館などで構成され、音楽に関する多種多様な利用が想定される。本稿では、主にコンサートホールとレクチャーホールの室内音響設計について報告する。

◆コンサートホール

- 生のクラシックコンサートや、拡声設備を用いた様々な音楽イベントに対応するホールとして計画された。
- 壁面がすべてガラスである。聴感的に鋭い反射音を和らげ、水平方向のモードが残ることを避けるため平面的に折板形状とし、部分的に上向きのガラスフィンを取り付けることとした。
- 内装天井を音響的に透明な構造とし、音響的な天井高さを上階スラブの高さまで確保した。透明度合いは、モックアップによる試験で確認を行った。
- 残響抑制のため、舞台後方の床面や客席後方の天井面に電動の吸音バナーを設置した。

◆レクチャーホール:

- 講演会などスピーチが主目的の、音楽利用もできるホールとして計画された。
- コン서트ホールと同様の空間構成だが、壁面の一部は微細穿孔板による吸音仕上げとした。

2-Q-8

2-Q-8 市民会館おおみや レイボックスホールの音響設計

Acoustical Design and Characteristics of Omiya Civic Hall

◎荒田耕平, 高橋藍子, 板垣直実, 北村浩一 (NHK-TECH)

- 【大ホール】
 - * 反射板の有無による約0.5秒の残響可変
 - * 直下にあるスタジオ群との遮音・防振対策
 - * スチールメッシュ下地の大開口スピーカサランネット
- 【小ホール】
 - * 左官仕上げによる高音域拡散
 - * 舞台機構を含めた完全浮き床構造

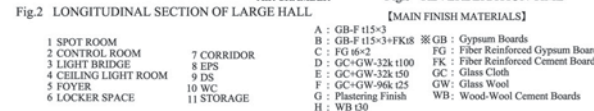
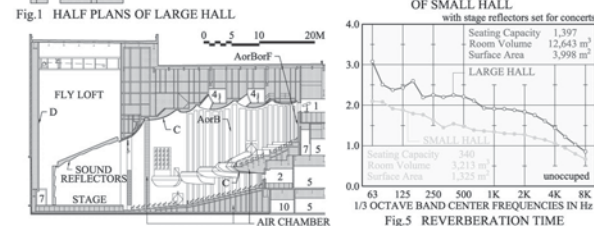
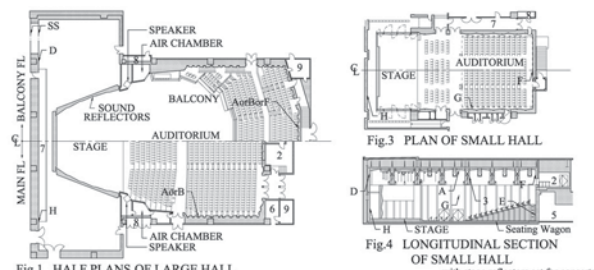


Fig.5 REVERBERATION TIME

【MAIN FINISH MATERIALS】
 A : GB-F 15-3
 B : GB-F 15-3+FK8
 C : FG 46-2
 D : GC+GW-32k 1100
 E : GC+GW-32k 150
 F : GC+GW-96k 125
 G : Plastering Finish
 H : WB 130
 ※ GB : Gypsum Boards
 FG : Fiber Reinforced Gypsum Boards
 FK : Fiber Reinforced Cement Boards
 GC : Glass Cloth
 GW : Glass Wool
 WB : Wood-Wool Cement Boards

2-Q-9

2-Q-9 枚方市総合文化芸術センターの音響設計

Acoustical design of Hirakata Performing & Visual Arts Center

○中川浩一(日建設計)

- ◆枚方市総合文化芸術センターは大・小・イベントの3つのホールと、リハーサル室、美術ギャラリー等から成る複合文化施設である。本稿ではホールを中心とした遮音計画、室内音響計画について報告する。
- ◆施設間の遮音性能確保のため、小ホールの周囲にエキスパンション・ジョイントを設け、小ホールとリハーサル室に防振遮音構造を採用した。小ホールの防振遮音構造は仕上のレンガ壁が浮置音層を兼ねる。
- ◆各ホールの壁面は、表面の平らなレンガと複数種の曲面レンガを組み合わせた仕上がりで、ランダムな凹凸面による音の散乱を意図した。
- ◆大ホールには国内のホール施設としては初の放射空調を採用した。天井面のアルミ製放射パネルによる吸音特性の影響確認のため、残響室法吸音率測定を実施しており、その結果についても報告する。



Fig.1: Ceiling radiation panel

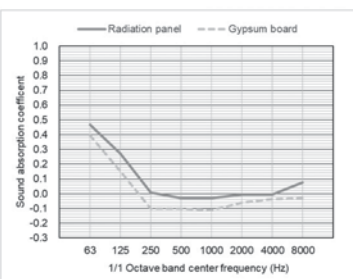


Fig.2: Comparison of sound absorption coefficient

2-Q-11

2-Q-11 減衰除去インパルス応答の周波数変動係数を用いた固有振動分析の数値シミュレーションによる検討

Normal mode analysis in a room using coefficient of variations calculated from the decay-cancelled impulse responses in frequency domain.

◎鈴木諒一, 星和磨, 羽入敏樹(日大短大)

- ◆周波数変動係数によって固有振動を評価できるか否か、吸音率を変化させたシミュレーション音場を用いて検討した。
- ◆音場Ⅰが 4.00 m×4.00 m×4.00 m の立方体室で、音場Ⅱが 5.10 m×4.05 m×3.05 m の最適な室寸法比として提案される音場の一つである。
- ◆吸音率が高い条件において、インパルス応答の周波数変動係数では音場Ⅰと音場Ⅱの周波数変動係数の差が小さく、室形の違いを捉えられないのに対し、減衰を除去したインパルス応答の周波数変動係数は両者の差が大きく、室形の違いを捉えられることがわかった。

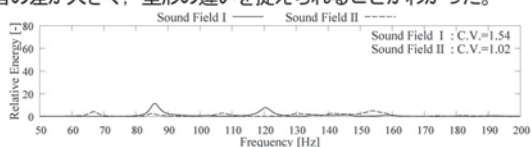


図1 インパルス応答のパワースペクトルと周波数変動係数(音場Ⅰ:1.54, 音場Ⅱ:1.02)

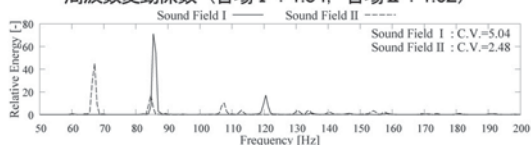


図2 減衰除去インパルス応答のパワースペクトルと周波数変動係数(音場Ⅰ:5.04, 音場Ⅱ:2.48)

2-Q-10

2-Q-10 VRを利用した室内音響指標の3次元空間分布の可視化

Visualization of 3D spatial distribution of room acoustical indices using VR

◎上原正志(都立産技高専), 大川茂樹(千葉工大)

- ◆VRを用いた音場の可視化
 - 室内音響指標(音声明瞭度)の3次元空間分布をVR空間上で表現・HMDで体験
- ◆音声明瞭度は多点測定された室内インパルス応答から算出
 - 移動ロボットによって室内インパルス応答を測定
 - SLAMを用いて環境地図および測定点の記録
- ◆三次元点群データとしての表現
 - UnityなどのVRが扱えるプラットフォーム上で読込・HMDへの出力

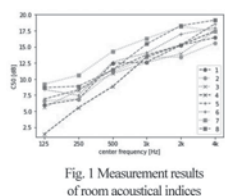


Fig.1 Measurement results of room acoustical indices



Fig.2 Example of visualization in VR space

2-Q-12

2-Q-12 原始根に基づく円形のリフレクター

Circular acoustic reflector based on primitive roots

☆金澤七海, 高橋義典, △牧野潔夫(工学院大)

- ◆数論の音響分野への応用として、平方剰余や原始根に基づく窪みを持ったリフレクターが知られている。
- ◆これまで提案され、実用化されているものは、様々な長さの角材を長方形に並べた形状のものが一般的であった。
- ◆本稿では、原始根に基づくランダムな凹凸を円形に並べた位相回折格子を提案する。
- ◆従来の四角いリフレクターと提案する円形のリフレクターについて、FDTDシミュレーションによって比較する。

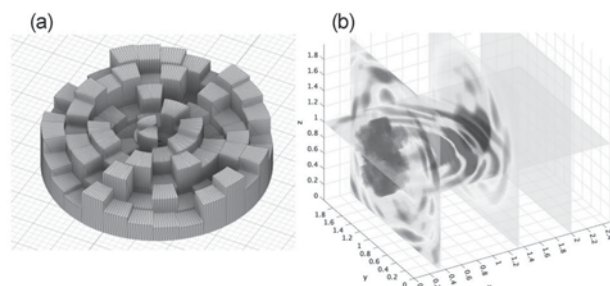


Fig.1 Circular acoustic reflector based on primitive roots (a) and diffused reflected sound shown by FDTD simulation (b).

2-Q-13

2-Q-13 側方エネルギー率における反射音到来方向による重み付けの再検証

Reverification of directional weighting for arriving reflected sound in the lateral energy fraction

○河合茜音(日大・院), 羽入敏樹, 星和磨, 鈴木諒一(日大・短)

- ◆ ASW と LEV の評価において、反射音到来方向と両耳軸のなす角度 θ による重みづけとして $\cos \theta$ と $\cos^2 \theta$ のどちらが適切か確認するために聴感実験を実施した。
- ◆ ASW、LEV の評価指標である側方エネルギー率として、 $\cos^2 \theta$ より $\cos \theta$ の到来方向の重み付けを用いる方が適切であることを確認した。
- ◆ この結果を踏まえ、以下のことが明らかになった。
 - ・ ASW の指標として LF より LFC が良い。
 - ・ LEV の指標として $\cos^2 \theta$ で重み付けしてさらに対数化した LG よりも、単に $\cos \theta$ の重み付けで評価の方が良い可能性がある。

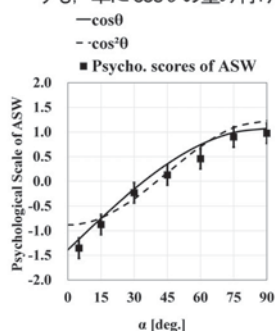


Fig. 1: Psychological scores of ASW

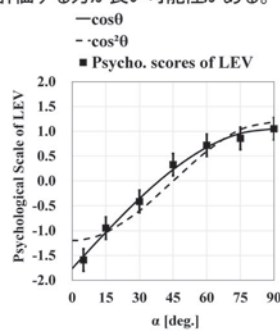


Fig. 2: Psychological scores of LEV

2-Q-15

2-Q-15 音圧-粒子速度センサを用いる吸音特性の in-situ 測定法に関する研究-温度・湿度毎の校正値の適用-

An in-situ sound absorption measurement method by ensemble averaging technique using a pressure-velocity sensor-Application of calibration values for each temperature and humidity-

☆池田夏菜(大分大院), 富来礼次, 岡本則子, 大鶴徹(大分大), 田原有彩(大分大院)

- ◆ 恒温恒湿器内で音響管を用いて音圧-粒子速度センサの校正値を測定する装置(Fig. 1)を製作し、温度・湿度毎の校正値を測定した。
- ◆ 温度・湿度条件として、約 10 °C から 26 °C の 8 °C 刻み、54 % から 75 % の 10 % 刻みの 9 条件を設定し、測定を行った
- ◆ EA 法による吸音率測定時と温度・湿度が最も近似した条件の校正値を利用した場合(#9)は、現場で測定した校正値を利用した場合と吸音率の最大差が 0.01 となった(Fig. 2)



Fig. 1 Measurement device installed in a thermo-hygrostat.

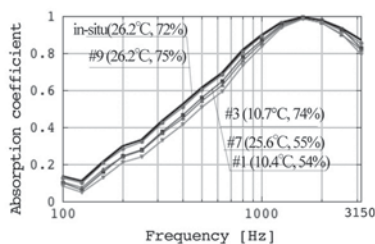


Fig. 2 Comparisons of absorption coefficient with calibration values applied for different temperature and humidity conditions.

2-Q-14

2-Q-14 残響室法吸音率の測定音場に関する数値解析的研究
-室形状と測定値の関係-

Numerical analysis study of sound field on measurement of absorption coefficient in a reverberation room.

-Relationship between room shape and measured values-

☆鈴木彰馬(大分大院・工学研), 富来礼次, 岡本則子, 大鶴徹(大分大・理工)

- ◆ 拡散板を設置していない場合を想定した 9 種の不整形残響室と 2 種の矩形残響室を解析対象とし、時間領域有限要素法による残響室法吸音率の測定に関する比較試験を行った。
- ◆ 対象とした全ての不整形七面体残響室で、推定した残響室法吸音率 α_{rFEM} は、数値解析で与えた試料の統計入射吸音率 α_{sp} に近い値を示した。
- ◆ 不整形七面体残響室の α_{rFEM} と α_{sp} の差 $\Delta \alpha_{rFEM}$ を求めたところ、250 - 630 Hz では 180 m³ 以上の室容積で残響室法吸音率が高くなり、拡散板がない場合でも、比較的良好な拡散音場で吸音率測定が可能であることを確認した。(Fig.1 (a), (b))

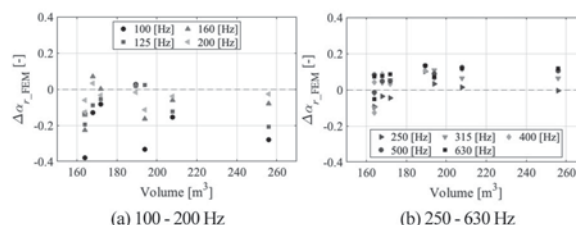


Fig.1 Relationship between $\Delta \alpha_{rFEM}$ and room volume, $\alpha_{sp} = 0.8$.

2-Q-16

カーボンナノチューブを用いた高性能な CNT/CMC 超軽量吸音材料の開発

○織田涼也, △柳怜旺, 瀬木貴裕, 上野智永(名大)

音は人々の心身に負荷を与え、健康被害を引き起こすため、対策が求められる。今後、大型ドローンや空飛ぶクルマの実用化が目指される中で、騒音問題は課題であり、社会受容性確保の観点から対策が必要である。騒音対策には、吸音材や遮音材などが用いられるが、輸送機器では、エネルギー効率やペイロードの観点から軽量性が強く求められる。不快な高周波音に加えて、1000 Hz 以下の低周波帯で問題になることが多いため、1000 Hz 以下の騒音が吸音可能な軽量材料が必要とされている。本研究では単層カーボンナノチューブ(SWCNT)とカルボキシメチルセルロース(CMC)を複合し、高性能な超軽量吸音材料を開発した。

面密度 40 mg cm² の試料は 500 Hz において 0.85 という高性能な吸音性能を示し、これまで報告された低面密度の吸音材の中でも最高値を示している。開発された SWCNT/CMC エアロゲルは軽量でかつ低周波吸音に適しており、ドローンを始めとする分野の騒音問題の解決法として革新的なものとなる。

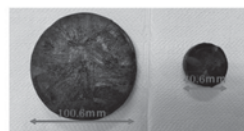


Fig.1 SWCNT/CMC aerogel

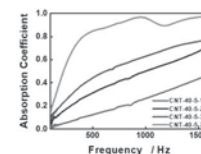


Fig.2 Difference in sound absorption coefficient due to thickness difference

2-Q-17

2-Q-17 無線ネットワーク化された小型端末を用いた大規模同期集音システムの高精度化

High Precision of Large-Scale Synchronized Recording System Using Microcomputer on Wireless Network

☆菅野正博, 大久保寛 (都立大)

- ◆本研究ではマイクロホンに接続した複数の小型端末をクライアントとして無線接続し、各クライアントの録音開始の時間的なずれをソフトウェアのみで補正することで、低コストで時間同期精度を向上させる手法を提案・評価する。
- ◆各クライアントの録音開始の時間的なずれが発生する大きな原因である以下の3つの時間を用いて、録音した音声の時間を補正することで、各クライアントが録音した音声の時間同期精度を高める。
 - ・サーバからクライアントへ録音開始命令を送信する時間
 - ・サーバ・クライアント間の通信遅延時間
 - ・クライアントが音声入力ライブラリを起動する時間
- ◆これにより、Fig.1に示すように、各クライアントの録音開始の時間的なずれを±2[ms]程度の偏差にまで縮めることができた。

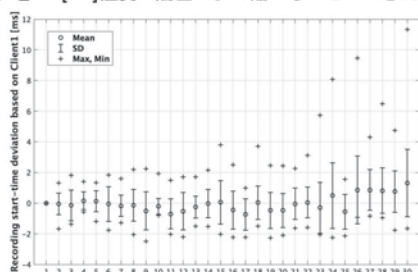


Fig.1: Deviation of recording start-time after correction (N=30).

2-Q-19

2-Q-19 環境音に対する印象評価と心拍変動の関係

Relationship between impression evaluation of ambient sounds and heart rate variability

☆赤塚亮太, 徳永泰伸(舞鶴高専)

- ◆本研究では、環境音を対象として、心拍計測を行い、被験者自身が動作を選択した場合における心拍変動と印象評価の関係について検討を行った。
- ◆実験では、着席状態にある被験者が3種類の環境音(セミの鳴き声、街路騒音、工事騒音)を受聴した場合の心拍変動を測定する。加えて、各音刺激について、印象評価への回答を求めた。
- ◆Fig.1は、副交感神経指標であるLF/HFが、刺激呈示前に比べて、刺激呈示の最中に増加するケースが多かったことを示している。またFig.2より、工事騒音呈示時の印象評価の回答についても、セミの鳴き声や街路騒音と比較して、評価値が低い傾向にあることが分かる。以上のことから、Fig.2の増加傾向と被験者のストレス反応には関連性が見られた。

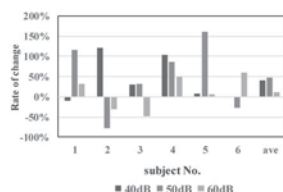


Fig.1 LF/HF rate of change before and during stimulation

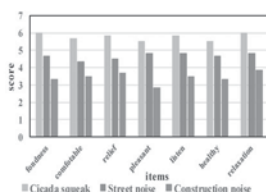


Fig.2 Answer results of impression evaluation in construction noise experiment

2-Q-18

2-Q-18 純音性成分を含む広帯域騒音のわずらわしさ評価におけるペナルティ関数—聴感評価実験による基礎的検討—

Penalty models of tonal components contained in broadband noises: fundamental studies by subjective evaluation tests

○米村美紀, 坂本慎一(東大生研)

- ◆広帯域騒音に卓越した狭帯域成分が含まれる「純音性騒音」は、設備機器などから発生することが多く、純音性成分によって不快感が増すと指摘されている。筆者らは、純音性成分の周波数や背景騒音の周波数特性を考慮した評価モデルの構築が必要であると考え、モデル構築のための基礎データとして、4種類の広帯域ノイズに50~2000Hzの純音を付加した試験音を用いてわずらわしさ評価実験を行った。
- ◆純音の強度が増すほどわずらわしさは高く評価される結果が得られた。この結果から、わずらわしさの増分を広帯域ノイズのA特性音圧レベルに換算したペナルティを試算したところ、0~8dB相当だった。
- ◆背景騒音の周波数特性によっては、純音の強さに対しペナルティが単調増加とならない場合があることがわかった。実在の騒音においても、背景騒音の周波数特性と純音周波数の相対関係によって、ペナルティ値を与える関数が複雑な形になる可能性が示唆された。

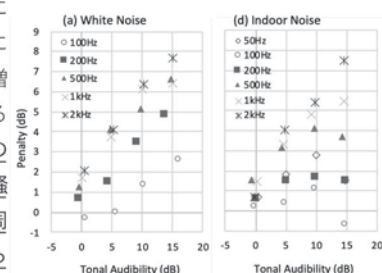


Fig.1: Penalties of tonal components obtained in noisiness evaluation tests.

2-Q-20

2-Q-20 音源移動が放射音の指向性の変化に及ぼす影響に関する実験的検討

Experimental study on the effect of sound source movement on the directivity of radiated sound

☆牧野 裕介, 高野 靖(京大院・工学研)、阿久津 真理子(早稲田大学/鉄道総研)、△宇田 東樹(鉄道総研)

- ◆模型発射実験 (Fig.1)
 - 模型にとりつけた音源からの放射音の指向特性の変化を計測

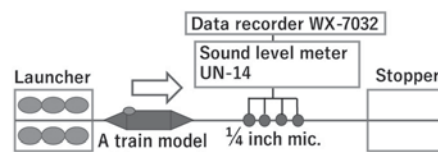


Fig.1: Setting of the train model launching experiment.

- ◆理論的に計算される放射音の指向性と比較 (Fig.2)
 - 理論値と実測値が概ね一致

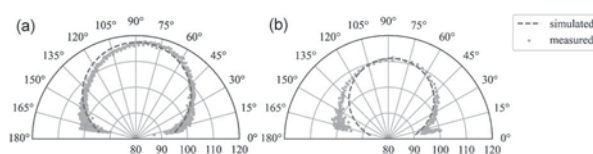


Fig.2: Directivity of radiated sound from a moving source [dB]. (Source velocity: (a)47.9 km/h, (b)166.2 km/h)

2-Q-21

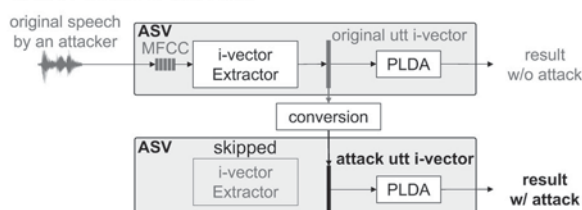
2-Q-21 i-vector の差ベクトルに基づく 話者認証システムへのなりすまし攻撃手法の検討 An Experimental Study of Spoofing Attacks on Speaker Verification System using Difference of i-vectors

☆菊地晏南, 齋藤大輔, 峯松信明 (東京大)

話者認証システムに対する攻撃手法を提案し
提案手法により誤り率の上昇が見られた



- ◆本研究では、話者埋め込みの一つである i-vector を用いた話者認証システムに対して、i-vector の差分を用いた攻撃手法を提案する。
- ◆i-vector を多次元空間内でさまざまな方向に並行移動させることで、攻撃話者の音声から抽出した i-vector を攻撃用の i-vector に変換した。
- ◆攻撃用 i-vector を用いた認証では誤り率が上昇したことから、i-vector が分布する多次元空間内の方向と、話者認証システムの誤り率に関係がある可能性が示唆された。



2-Q-23

2-Q-23 Hybrid RNN-T/Attention 音声認識モデルにおける Triggered Chunkwise Attention と 内部言語モデル統合の検討

A Study on Hybrid RNN-T/Attention-based Streaming ASR with Triggered
Chunkwise Attention and Dual Internal Language Model Integration

森谷崇史, 芦原孝典, 安藤厚志, 佐藤宏, 田中智大, 松浦孝平,

増村亮, デルクロアマーク(NTT), 篠崎隆宏(東工大)

- ◆先行研究では RNN-T と Attention-based decoder (AD) を組み合わせた Hybrid RNN-T/Attention モデルを提案
 - AD に Triggered AD (TAD) を応用することでストリーミング動作可能な Hybrid RNN-T/Attention 構造を提案
- ◆従来の TAD ではストリーミング処理を可能としたが、音声長によっては計算量やメモリ消費量の観点でオンデバイスでの動作が困難
- ◆本研究では認識精度を保ちながら計算コストが削減可能な Triggered Chunkwise AD (TCAD) を提案
- ◆認識精度のさらなる改善のため、様々な言語モデル統合の検討も実施
- ◆実験により、TCAD は認識精度を維持しながら認識速度を著しく改善し、様々な言語モデルや内部言語モデルと統合することでさらに認識精度が改善することを確認

Table 1: Results of each streaming ASR system.

System	TED-LIUM2				NTTWS			
	segment (8s)	WER	RTF	WER	command (2s)	CER	RTF	WER
RNN-T	11.4	0.53	14.7	0.53	5.2	0.41		
+TAD	11.0	0.62	14.4	6.88	4.8	0.53		
+TCAD	11.0	0.59	14.2	0.67	4.8	0.49		

2-Q-22

2-Q-22 中間層予測に対するノイズ付与による CTC 音声認識の頑健性向上

Augmenting Noisy Intermediate Predictions
for Improving Robustness of CTC-based ASR

◎中込 優, 小松 達也, 藤田雄介, 市村収太, 木田祐介 (LINE)

- ◆非自己回帰型音声認識において優れた性能を示している Self-conditioned CTC をベースとして、中間層予測に対して意図的にノイズを付与することで、モデルの頑健性を高める新しい学習法を提案する。
- ◆提案法では、推論時に典型的に生じるトークンの削除、挿入、置換誤りを模擬したエラーが付与された「ノイズの多い中間層予測」が後続の層の入力に条件付けされ、後続の層ではそれらのエラーを修正するように学習される。
- ◆エンコーダ層の複数箇所において、エラー付与とエラー補正が繰り返されることで、模擬して加えたエラーに対するモデルの頑健性の向上が期待できる。
- ◆LibriSpeech データセットを用いた評価実験において、提案法が全ての評価セットで Self-conditioned CTC の性能を上回ることを示した。

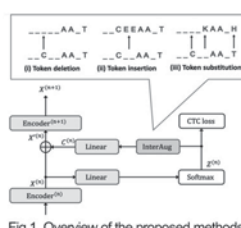


Fig.1. Overview of the proposed methods

Table 1. Word error rates (WERs) on LibriSpeech (100h). The results were obtained without language models.

Model	dev		test	
	clean	other	clean	other
Conventional models				
CTC	8.69	23.00	9.08	23.68
SelfCond	7.11	20.82	7.48	21.31
Prop. (Token space)				
Token del.	6.97	20.30	7.38	20.72
Token ins.	6.95	20.32	7.44	20.71
Token sub.	6.90	19.99	7.23	20.34

2-Q-24

2-Q-24 中間層予測にビームサーチを用いた 新しい CTC 推論

A New CTC Inference Method
with Beam Search for Intermediate Prediction

◎小松達也, 藤田雄介 (LINE), Jaesong Lee, Lukas Lee (NAVER),
渡部晋治 (CMU), 木田祐介 (LINE)

- ◆非自己回帰型音声認識において強力な性能を示している Self-Conditioned CTC をベースとして、中間層予測にビームサーチを用いた新しい CTC 推論を提案する。
- ◆提案手法は、中間層において推論を行う際に外部言語モデルを用いたビームサーチを用い、ビームサーチを施された「優れた中間層予測」を次層の音響エンコーダの入力へ加算する。
- ◆音響エンコーダに外部言語モデルの情報を取り入れることが可能となり、最終層における推論の改善が期待できる。
- ◆LibriSpeech データセットを用いた評価実験を行った。すべてのデータセットにおいて提案手法が最もよい性能を示した。
- ◆「優れた中間層予測」による音響エンコーダの条件付けは、最終的な推論制度の改善につながることがわかった。

Table 1 Experimental results using greedy search and LM+beamsearch for the final outputs.

model	output decoding	LibriSpeech (100h)				LibriSpeech (960h)			
		dev-clean	dev-other	test-clean	test-other	dev-clean	dev-other	test-clean	test-other
decoding with CTC greedy search									
CTC	CTC greedy	10.21	24.19	10.65	25.00	3.79	9.76	3.93	9.88
Intermediate CTC	CTC greedy	7.76	20.59	8.16	21.21	3.09	8.49	3.30	8.36
Self-Conditioned CTC	CTC greedy	7.51	20.04	7.86	20.65	2.86	7.87	3.06	7.91
Proposed method [†]	CTC greedy	6.45	18.45	7.11	18.57	2.68	6.95	2.98	7.07
decoding with Transformer LM + search									
Self-Conditioned CTC	beam search	5.05	15.44	5.40	15.61	2.19	5.96	2.45	6.08
Proposed method [†]	beam search	4.60	14.00	5.07	14.47	2.17	5.66	2.48	5.97

[†] For searched conditioning approach, we applied Transformer LM + beam search on the intermediate predictions.

2-Q-25

2-Q-25 A Comparison of Pretraining Frameworks for Improving Pathological Speech Recognition

☆ Lester Phillip Violeta, Wen-Chin Huang, Tomoki Toda (Nagoya University)

We evaluate the performance of different pretraining frameworks on pathological speech datasets used for automatic speech recognition (ASR). With only a small number of pathological speech datasets publicly available, modern end-to-end models do not generalize to these datasets very well due to their massive data requirements. A common solution to this problem is by first pretraining the model on a huge number of healthy speech datasets and then fine-tuning it on the pathological speech datasets. We investigate two types of pretraining frameworks such as supervised and self-supervised learning (SSL), both of which have been very successful with using minimal amounts of data. Our results show that the best supervised pretraining setup outperforms the best SSL setup, showing that more work needs to be done to improve self-supervised learning techniques with pathological speech.

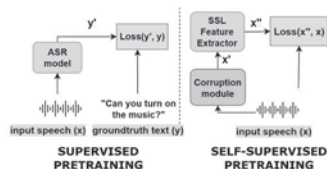


Fig. 1. A comparison of two popular ASR pretraining methods: supervised and self-supervised.

2-Q-27

2-Q-27 中間層予測に音節と表記を用いる日本語音声認識

Syllable-level and character-level intermediate conditioning for Japanese ASR

○藤田雄介, 小松達也, 木田祐介(LINE)

- ◆End-to-End モデルにより音声認識精度は飛躍的に向上したが、日本語の同音異字や各漢字が持つ複数の発音は依然として課題である。
- ◆そこで、日本語における音節と表記の間の相互作用を考慮した End-to-End モデルのマルチタスク学習法を提案する。
- ◆提案手法は、非自己回帰型モデルの一つである Self-conditioned CTC により、文字表記ラベルの中間層予測を次の層に供給する構成を用いるのに加えて、音節と文字の予測の相互作用をモデル化するために、別の中間層において音節ラベルをさらに予測する。提案手法では、予測を行う中間層を交互に配置して、音節→文字、文字→表記の相互作用を捉える。
- ◆日本語話し言葉コーパス (CSJ-APS)を用い、提案手法が従来のマルチタスク学習や Self-conditioned CTC を上回ることを示した。

Table 1 Character error rates on CSJ. The results were obtained without language models.

Method	中間層予測を行う位置		中間層予測の フィードバック	評価セット		
	表記ラベル	音節ラベル		eval1	eval2	eval3
従来モデル						
Baseline	18		N	5.4	3.9	9.9
Multitask	18	15	N	5.4	3.7	9.5
InterCTC	3,6,9,12,15,18		N	5.4	3.8	9.4
SelfCond	3,6,9,12,15,18		Y	5.3	3.7	9.2
提案モデル						
Parallel	6,12,18	6,12,18	Y	5.2	3.6	8.8
Hierarchical	12,15,18	3,6,9	Y	5.1	3.6	8.8
Alternate	6,12,18	3,9,15	Y	5.1	3.5	8.8

2-Q-26

2-Q-26 ドメイン外音響情報で補強した Encoder-Decoder 音声認識モデルの設計
Design of Encoder-Decoder speech recognition models augmented with out-of-domain acoustic information

☆森大輝(豊橋技科大), 太田健吾(阿南高専),
西村良太(徳島大), 北岡教英(豊橋技科大)

- ◆Density Ratio Approach の考え方を応用し、ドメイン外音声認識モデルから音響情報のみを抽出し、ターゲットドメイン音声認識モデルと併用するモデルを提案している。
- ◆CSJ データセットに含まれる学術講演コーパスと模擬講演コーパスの2つのドメインを用いて実験を行った。
- ◆提案法は Test セットのドメインと外部的な音響情報のドメインが異なる場合、認識精度を改善することができなかった。
- ◆クロスドメインタスクでの認識結果を以下の表に示す。
 - ・Baseline は純粋な RNN Based 音声認識モデルで、SF は Shallow Fusion である。
 - ・提案法は SF および Baseline と比較して、クロスドメインタスクでの認識精度を改善することができた。
 - ・しかし、この結果は音声認識モデル内部の音響情報と言語情報が完全に分離できていないことの証明であると考えられる。

統合手法	ASR モデル	Test セット	ドメイン外音響情報	$\lambda_{acoustic}$	λ_{sub}	λ_{add}	CER
Baseline	—	—	—	—	—	—	16.1
SF	学術講演	模擬講演	—	—	—	0.3	15.4
Proposed	—	—	模擬講演	0.1	0.5	0.3	14.8
Baseline	—	—	—	—	—	—	19.1
SF	模擬講演	学術講演	—	—	—	0.1	17.7
Proposed	—	—	学術講演	0.1	0.5	0.5	17.4

- ◆今後の課題として、音声認識モデルの音響情報と言語情報を分離させ、あらゆるドメインに適用できるようなスキームを考案したい。

2-Q-28

2-Q-28 文ベクトル系列の予測モデルを用いた雑談対話における適切な応答文選択法

A method for selecting appropriate response sentences in chat dialogues using a predictive model of sentence vector sequences

○袖谷紳太郎, 鈴木基之(大阪工業大学)

- ◆本研究では、ニューラル対話モデルが生成する応答候補文の中から、文脈を考慮した最も自然な応答文を選択する手法の提案と、実験・評価を行った。
- ◆先行研究では対話履歴に対して応答候補文のトピックが類似していれば、内容が矛盾していても自然であるとみなされていたため、SentenceBERT を用いて対話履歴の埋め込み表現を獲得し、そのベクトル系列から応答文を予測する事で自然な応答文選択の精度向上を試みた。
- ◆応答文選択の手法として、対話履歴ベクトルから応答文ベクトルを推定し、応答候補文ベクトルとの類似度から応答を選択する手法、対話履歴と応答候補文のベクトルから応答文が文脈として自然かどうかをモデルで判定し、応答の自然さで選択する手法、SentenceBERT を Fine-tuning し、2 つめ同様応答の自然さで選択する手法を提案した。
- ◆提案した各モデルに対し、対話履歴の直前 3 発話に、続く応答文である正例と、文脈として不自然な負例を 3 文与え、正例を適切な応答文として選択できるか評価した。結果は Table.1 の通りである。

	文ベクトル推定モデル	文ベクトル分類モデル	Fine-tuning モデル
正解率(%)	56.3	72.3	93.3

Table.1: Performance of the proposed model

2-Q-29

2-Q-29 未知のドメインに対して頑健な
音声表現学習の検討

Speech Representation Learning with Domain-Adversarial Neural Network

○田中 智大, 増村 亮, 佐藤 宏, 庵 愛,

松浦 孝平, 芦原 孝典, 森谷 崇史 (NTT)

- ◆ 音声の教師なし事前学習によって、音声認識の精度が改善
- ◆ 高精度な教師なし事前学習のモデルを構築するためには様々なドメインのデータを有効に利用することが必要
- ◆ これまで提案されてきた手法では、学習データのドメインの依存性が大きく、未知のドメインに対する頑健性は小さい
- ◆ ドメイン依存性を低減させる音声の教師なし事前学習手法を提案
- ◆ Domain-Adversarial Neural Network (DANN) を導入し、獲得された音声の表現がドメイン情報を識別できないように学習
- ◆ 実験では、事前学習の学習データに含まれる既知ドメイン、含まれない未知ドメインを含む複数ドメインの日本語音声認識
- ◆ タスクに対して再学習を行い評価
- ◆ 従来の wav2vec 2.0 と比較して、後段の音声認識タスクが未知のドメインであっても高精度に音声認識ができることを確認

Model	Known domain Lecture	Unknown domain		
		Call	Meeting	Dialog
wav2vec 2.0	5.51	5.64	23.51	16.27
+DANN	5.39	5.35	22.28	15.43

Fig. 1: Character error rates (%)

2-Q-31

2-Q-31 シングルチャネルマルチトーカー音声認識
と話者属性推定の結合モデリングSingle-channel multi-talker overlapped automatic speech
recognition with speaker attribute estimation○増村 亮, 牧島 直輝, 庵 愛, 田中 智大, 高島 瑛彦, 折橋 翔太
(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)

- ◆ 本稿では、シングルチャネルマルチトーカー音声認識と話者属性推定の結合モデリングを提案。
- 提案法では、性別や年齢など話者属性とし、それらを出力対象のトークン系列内で扱うことで、単一の自己回帰モデルで結合モデリングを実現

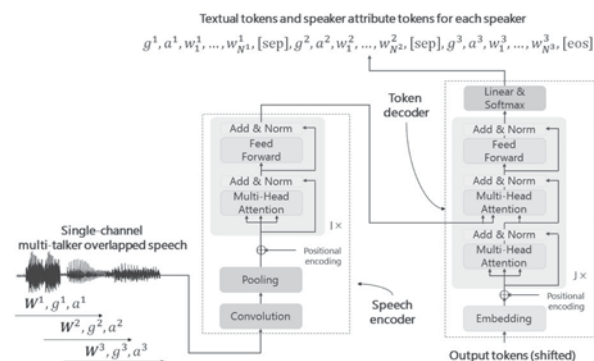


図 1: 提案手法.

2-Q-30

2-Q-30 Attribute Prototype Network を用いた音響
イベントのゼロショット学習

Zero-shot sound event classification using Attribute Prototype Network

☆LIN YIHAN, 高島遼一, 滝口哲也 (神戸大)

- ◆ 音響イベント分類において、イベントによっては学習データが入手しづらい課題がある。
- ◆ 我々の先行研究では、音源の材質や音の高低などの属性情報をイベントごとに定義し、クラスラベルの代わりに属性情報を出力することで、音響イベントのゼロショット学習を行う手法を提案した。
- ◆ 本研究では Attribute Prototype Network を用いることで、属性とスペクトログラムの局所情報との関連を学習し、さらにゼロショット学習性能を向上する手法を提案する。

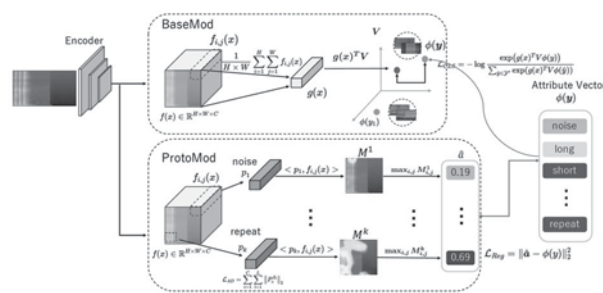


Fig.1: Overview of the proposed method based on an Attributes prototype network.

2-Q-32

2-Q-32 音声合成器の中間表現を利用した発話内
容に依存しない感情識別用特徴量の抽出Extraction of Feature for Emotion Identification Independent of Utterance
Content using Intermediate Expression of TTS

○八木崇(阪工大・情), △鈴木基之(鈴木研)

- ◆ 研究背景
 - 従来の音声感情識別における使用特徴量、MFCC やメルスペクトログラム等は発話内容によって大きく変動する問題が存在。
- ◆ 提案手法
 - End-to-End 音声合成の中間表現加工処理に着目
 - 感情音声から感情・言語情報を含む中間表現を生成
 - 発話文章から言語情報を含む中間表現を生成
 - 両者の差分を発話内容に依存しない感情識別用特徴量と提案



Fig. 1: Proposed Method

- ◆ 実験結果
 - 提案した特徴量が軽量ネットワークにおいて有効なことを確認

Table 1: Results of SER Accuracy

Feature	Small	Middle	Large
Middle Data(E)	28.8%	27.6%	24.3%
Middle Data(E&L)	26.8%	27.7%	29.1%

2-Q-33

2-Q-33 スイッチングトークンを活用した zero-shot モデリングに基づく話し言葉テキストのスタイル変換

Zero-Shot Modeling of Spoken Text Style Conversion using Switching Tokens

◎庵愛, 牧島直輝, 田中智大, 高島瑛彦, 折橋翔太, 増村亮

(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)

◆本稿では、音声認識の出力結果である話し言葉テキストに対し、その可読性を向上させるための複数のタスクを、そのデータセットなしに同時に実行する zero-shot モデリングの方法を提案する。

- 従来、複数のタスクを同時に実行するためにはカスケードリングが用いられてきたが、誤りの連鎖や接続順番を考慮する必要があること、計算量が多いことなどの課題がある。
- 提案手法では、系列変換モデルにスイッチングトークンを導入することにより複数のデータセットを同時に学習することが可能となり、実行するタスクを制御できるようになると考えられる。

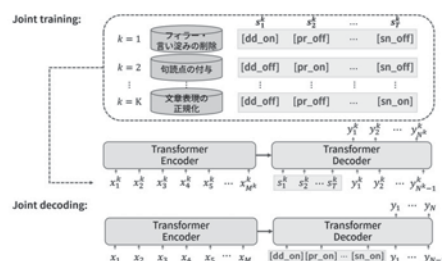


図1: スイッチングトークンを用いた複数データセットの同時学習

2-Q-35

分散マイクを用いた多チャンネルニューラルダイアライゼーション

Multi-channel neural diarization using distributed microphones

◎堀口翔太¹, 高島悠樹¹, △Paola Garcia², 渡部晋治³, 川口洋平¹
¹日立 ²JHU ³CMU

- ◆話者ダイアライゼーションを1つのネットワークで行う End-to-End Neural Diarization (EEND) に関する既存研究の多くは1チャンネル入力を対象としていた。
- ◆本研究では、分散マイクからの多チャンネル入力を取り扱えるように EEND を拡張した。設定した要件は以下の3点で、特に3つ目はハイブリッド会議が増加する昨今、リモート参加者の音声が一のスピーカーから再生され得る状況下においては重要である。
 1. マイクの数や配置に依存しない(マイク1つでも動く)。
 2. 位相差情報を用いない(マイク間のクロックドリフトを考慮)。
 3. 複数話者の位置が同じ場合でも話者を正しく区別できる。
- ◆多チャンネル EEND として、1チャンネル EEND で用いられる Transformer エンコーダに代えて、spatio-temporal エンコーダおよび co-attention エンコーダをそれぞれ用いるモデルを提案した。
- ◆2話者シミュレーションデータで検証した結果、特に co-attention エンコーダを用いた場合に、話者の位置が異なる場合にはチャンネル数増加の恩恵を受けることができ、また話者の位置が同じ場合には1チャンネルモデルと同等の精度を達成した。

Table 1: Diarization error rates on the simulated two-speaker conversations

	2 speakers at different positions					2 speakers at the same position				
Method	1ch	2ch	4ch	6ch	10ch	1ch	2ch	4ch	6ch	10ch
1ch + posterior averaging	5.13	4.60	4.31	4.19	4.10	6.07	5.68	5.42	5.38	5.33
Spatio-temporal	6.34	3.02	1.56	1.28	1.07	8.11	8.23	6.98	6.72	6.40
Co-attention	4.68	2.52	1.71	1.40	1.23	5.73	5.34	5.05	5.18	5.35

2-Q-34

2-Q-34 環境音合成の入力情報に応じた主観評価手法の検討

Investigation of subjective evaluation method for environmental sound synthesis according to input information

◎岡本悠希(立命館大学), 井本桂右(同志社大学),

高道慎之介(東大院・情報理工), 福森隆寛, 山下洋一(立命館大学)

◆環境音合成の入力情報に応じた主観評価手法を提案

◆提案①: 音響イベントラベルから生成された環境音の評価手法

- ✓ 従来: 音と音響イベントラベルを提示した「全体の印象」、「自然性」の絶対/相対評価
- ✓ 提案: 音と音響イベントラベルを提示した妥当性の絶対/相対評価

◆提案②: 擬音語から生成された環境音の評価手法

- ✓ 従来: 音と擬音語を提示した「許容度」、「表現性」の絶対評価
- ✓ 提案: 音と擬音語を提示した「妥当性」の絶対/相対評価

◆絶対評価と相対評価では評価結果が異なることを確認 (Fig.1, 2)

→ 環境音合成においては絶対/相対評価どちらも実施する必要がある

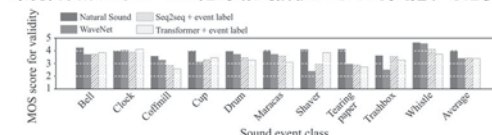


Fig. 1. MOS score for Validity by absolute evaluation method (Presented the sound and sound event label)

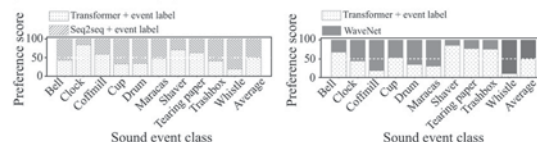


Fig. 2. Preference score by relative evaluation method (Presented pair sound and sound event label)

2-Q-36

2-Q-36 人間からの誤り訂正フィードバックに基づく適応的な日本語音声合成に関する検討

Adaptive End-to-End Text-to-Speech Synthesis Based on Error Correction Feedback from Humans

☆藤井一貴, 齋藤佑樹, 猿渡洋(東大院・情報理工)

◆テキスト音声合成 (Text-to-Speech: TTS) は、End-to-End(E2E)方式の登場で肉声に匹敵する品質の音声合成可能

◆しかし、DNN で記述される E2E TTS は人間による解釈性に欠ける。=合成音声に誤りが発生した際の訂正が困難!

◆特に日本語はアクセントの違いで言葉の意味を区別

→アクセントに誤りが発生した場合、異なる意味で伝わってしまう!

◆本研究では、合成音声の誤りを

- ・訂正可能な操作性
- ・非専門家でも容易に訂正可能とする適応性

以上の2つを兼ね備えた誤り訂正フィードバックに基づく E2E TTS を提案

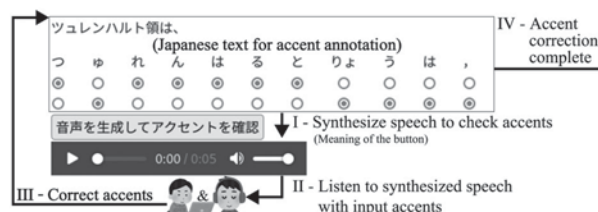


Fig.1: Interface and overview of accent error correction feedback.

2-Q-37

2-Q-37 拡散確率モデルとアライメントモデルを用いた潜在特徴系列変換に基づくテキスト音声合成

Latent feature mapping based text-to-speech synthesis using diffusion and alignment probabilistic model

©安田裕介, 戸田智基 (名古屋大学)

本稿では、拡散確率モデルを用いてテキストから波形モデルの潜在表現を予測することで、テキストから波形へ変換する音声合成手法を提案する。拡散確率モデルのモデル関数や分散の設計にはいくつかの選択肢があり、音声合成に適した設計を検証して明らかにする。また変数変換に基づくアライメント法を提案し、その性能を検証する。実験では聴取試験を行い、提案法を近年提案されたフローや拡散確率モデルに基づくテキスト-波形音声合成手法と比較する。実験結果では、拡散確率モデルにおいてモデル関数の出力の選択が合成音声の品質に大きく影響を与えることを示す。また提案法は文字を入力に用いた際に、音素を入力に用いた場合に比べて品質の劣化を示さず、文字入力の場合において近年の手法に匹敵する高い合成音声の品質を達成できることを示す。

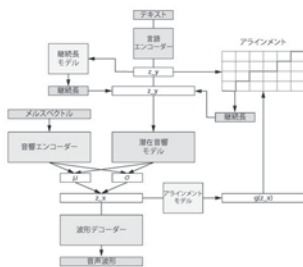


図1: 拡散確率モデルを用いた音声合成システム

2-Q-39

2-Q-39 構音障害者のための高明瞭度音声合成におけるHiFi-GANを用いた品質改善

Quality improvement using HiFi-GAN on high-intelligibility speech synthesis for dysarthric people

△松原圭亮^{1,2}, ©高島遼一¹, 岡本拓磨², 滝口哲也¹, 戸田智基^{3,2}, 河井恒²

(¹神戸大学, ²情報通信研究機構, ³名古屋大学)

- ◆明瞭な発音が困難な構音障害者のため、本人らしさを維持しつつ、かつ明瞭な音声を合成する手法を検討する。
- ◆我々の先行研究では、まず高明瞭に発話可能な健康者の音声を合成したのち、声質変換によって明瞭性を維持しつつ話者性のみを障害者本人の音声に変換する手法を提案した (Fig. 1)。
- ◆先行研究ではLPCNetをボコーダとして用いていたが、ボコーダの学習データと声質変換によって出力される音響特徴量とのミスマッチによる品質劣化が課題であった。
- ◆本研究では、未知話者でも高品質に合成可能なHiFi-GANと、その改良手法であるHarmonic-Netをボコーダとして使用することで、前述のミスマッチの影響を軽減し、品質の改善を試みる。



Fig.1: Overview of our high-intelligibility speech synthesis system based on TTS, VC, and vocoder.

2-Q-38

2-Q-38 応答テキストを介さないテキスト対話系列からの音声応答生成の検討

Speech-reply generation from text dialogue history not via text-reply

©藤田健一, 井島勇祐, 杉山弘晃 (NTT)

- ◆音声認識、雑談対話システム、音声合成 (TTS) を併せて用いることで、音声の雑談対話システムが実現されつつある
 - 途中にテキストを介するために、言い淀みなどテキスト化が難しい表現や、音声情報まで反映した音声応答の生成は困難
- ◆テキストを途中に介さないEnd-to-Endの音声対話システムの実現
 - 初期検討として、テキスト対話系列から音声応答を直接に生成する手法を提案・検討
- ◆TTSによる大規模な音声対話データの生成と音声量子化を用いることで音声応答の直接生成を実現。客観評価による評価の実施
 - 生成される応答の多様性などではテキスト対話モデルにやや劣るが、自然な音声応答の生成を確認

話者	発話
テキスト対話系列	
1	おいしいんですよ日本人に合うと思います。ラーメンとか基本中華ですし
2	マレーシアって安いんですか??
1	マレーシア安いんですよ、100円もあれば1食十分に食べれます
テキスト応答	
2	そうなんですね! 結構安いですね!
音声応答	
2	めっちゃ安いじゃないですか

Table.1: Example of generated dialogue

2-Q-40

2-Q-40 WORLDによる合成音声の位相特性の観察

Observation of phase characteristics of speech synthesized by WORLD

☆高津航輝, 矢田部浩平 (農工大)

- ◆背景 :
 - WORLDの再合成音声はバジーな音になってしまう
- ◆目的 :
 - WORLDの再合成音声の、調波間の位相関係がバジー感に与える影響について観察する
- ◆取り組み :
 - 元音声と再合成音声の調波間の位相関係の差を補償する
 - 時変オールパスフィルタを設計して、再合成音声に適用した

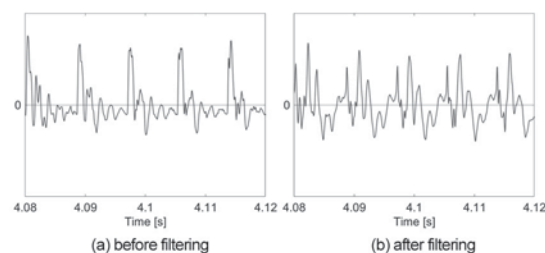


Fig 1: Waveform change by applying time-varying all-pass filter

- ◆結果 :
 - WORLD再合成音声の調波間位相関係を元音声に近づけた結果、合成音声のバジー感は軽減した

2-Q-41

2-Q-41 避難呼びかけ音声の持つ緊急性の分析と
音声合成への適用の検討

Analysis of urgency of evacuation announcement speech and its application to text-to-speech

☆原田 そら(木更津高専), 中田 亘(東大・工), 高道 慎之介,
齋藤 佑樹(東大院・情報理工), 齋藤 康之(木更津高専),
猿渡 洋(東大院・情報理工)

- ◆避難行動は災害時において重要である。避難を促すために、避難呼びかけは有効な手法となる。
- ◆呼びかけの効果を高めるための手法として、「避難呼びかけ音声」の適切な設計がある。先行研究では、緊迫感の高い音声で避難行動を促すことを示唆している[小林ら, 2018]。
- ◆本研究では、音声にある緊急性と影響する要因の実験的評価を行い、学習済みニューラル TTS への適用によるシステム化を試みた。

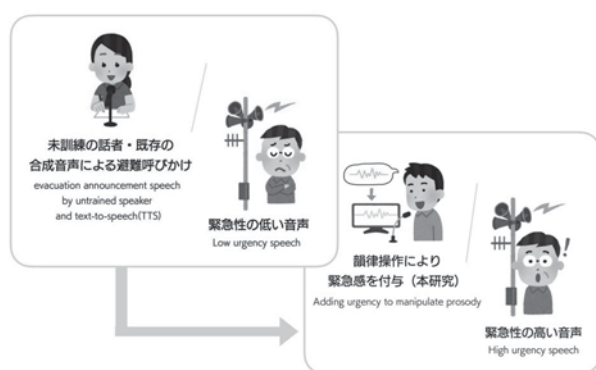


Fig1: Overview for Research

2-Q-43

2-Q-43 教師なし音声分割と単語埋め込み空間を
用いた L2 音声の競合単語分析

Competitive word analysis of L2 speech using unsupervised speech segmentation and word embedding space

☆勝瑞雄介, 峯松信明, 齋藤大輔(東大工), 中西のりこ(神戸学院大)

- ◆L2 音声の聞き取りやすさに関する調査を行った。
- ◆ある発声において音響的に似ている単語が多いほど聞き取りが難しくなると考えられる。
- ◆単語単位で、誤認識されやすい単語の数(競合単語数)についての調査がある[Shozui+, 2022]が、実際の聞き取りでは単語境界を正しく認識されない場合がある。
- ◆L2 音声を教師無しで音節単位に分割し、疑似 1 音節単語と疑似 2 音節単語について、L1 単語音声が音響的に埋め込まれた単語埋め込み空間上に埋め込み、分析を行った。(図 1)

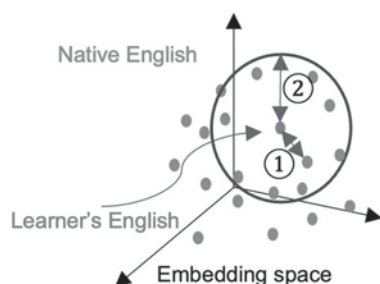


Fig. 1: How to analyze embedding vector of pseudo-words

2-Q-42

2-Q-42 時変複素分析を用いた音声の F_0 推定
における前処理の検討

A study of pre-processing on F_0 estimation of speech based on TV-CAR analysis

舟木慶一(琉球大)

ACELP や RCELP などの CELP 音声符号化で使用されている LPC 分析は最も広く使われている音声分析である。LPC 分析は最小二乗法に基づく方式のため、推定精度が低いという欠点がある。そこで、時間軸ならびに周波数軸 f_0 ノルム正規化時変複素 AR(TV-CAR) 音声分析法を提案した。これは正規化 LP(RLP) 法と時間 RLP(TRLRP) 法のハイブリッド方式であり、複素残差を用いた IRAPT[1] に基づく F_0 推定を用いて評価した [2][3]。さらに、骨伝導 (BC) フィルタ [4] ならびに適応プリエンファシス (APE)[5]を導入し、 F_0 推定の性能向上を図った。本稿では、BC フィルタと APE の両方を導入した改良型 F_0 推定を提案し、その効果を検証する。

2-Q-44

2-Q-44 歌唱音声におけるスペクトルの
微細変動増減処理による
ビブラート音声の作成と自然性の調査

A study on effect of naturalness of vibrato by increase and decrease processing for fine fluctuation of spectrum in singing voice.

◎鈴木千文, 坂野秀樹, 旭 健作(名城大)

- ◆本研究ではスペクトルの各ビンの時系列に含まれる微細変動の減衰処理と増幅処理によるビブラート合成音の作成を行った
- ◆メルケプストラム歪みの算出
 - 基準フレームをビブラート終了区間のフレームとした場合の歪みを算出
 - ビブラートの終了区間では、スペクトルの形状が大きく変化していることが分かった
- ◆スペクトルの各ビンの時系列に含まれる微細変動に対し減衰処理や増幅処理によってビブラート合成音を作成
- ◆微細変動に対して減衰処理を行った結果
 - 減衰処理での重み: 1.0~0 まで 0.1 刻みで設定
 - ◇ 重みを小さくしてもビブラート合成音はほとんど変わらない印象
 - 微細変動が聴感上与える影響はほとんどないと考えられる
- ◆微細変動に対して増幅処理を行った結果
 - 増幅処理での重み: 1.1~3.0 まで 0.1 刻みで設定
 - ◇ 2.1~3.0 まで重みを変化させるとビブラート合成音の自然性が低下する印象
 - 微細変動が 2 倍以上になると聴感上影響を与える可能性あり

2-Q-45

2-Q-45 自動推定された音素境界とフォルマントによる音声モーフィングの品質評価

Quality evaluation of speech morphing with automatically estimated phoneme boundaries and formants

☆堀部貴紀, 森勢将雅(明治大), 河原英紀(和歌山大)

- ◆ WORLD をベースとした音声モーフィングにおいて、時間軸に音素境界、周波数軸にフォルマント周波数を対応点の設定に用いた音声モーフィング手法について検討した。
- ◆ Julius による自動音素アライメントで得られる音素境界を中心に切り出したフレームに対して、線形予測符号によるフォルマント推定手法を用いる。
- ◆ 得られたフォルマント周波数のうち、第1フォルマント、第2フォルマントを音声モーフィングの対応点とする。
- ◆ 提案手法の妥当性を評価するため、対応点を手動設定して作成したモーフィング音声との自然性に関する主観評価実験を実施した。
- ◆ 手動設定されたモーフィング音声と比較したところ、自動設定されたモーフィング音声でも 40%程度は同等の品質であることが示唆された。

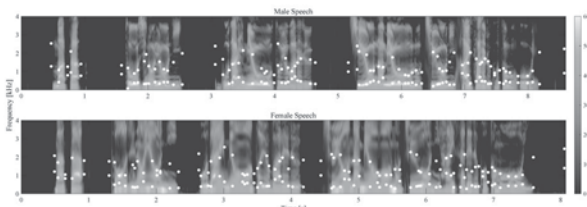


Fig.1: Spectrograms with anchor points by proposed method. The circles indicate F1 and F2 set as anchor points.

2-Q-47

2-Q-47 話者因子係数の量子化に基づく声色制御可能な話者変換

Voice-controllable speaker transformation based on quantization of speaker factor coefficients.

☆井路巧, 大西弘太郎, 岸田拓也, 中鹿亘(電通大)

- ◆ これまでの声質変換モデルの多くは、ある人物からある人物への声質変換にフォーカスしており、自分の声の擦れ具合や太さ等を補正して、理想の声を得るといったような事が出来なかった。
- ◆ そのため、本研究では、声の擦れ具合などの声色を決定する要素とその強弱を学習することで、声色を制御可能な声質変換モデルである VQ-SplitterNetVC を提案する。

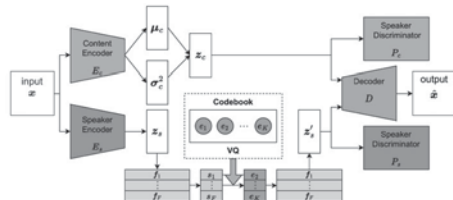


Fig.1: Structure of the proposed method

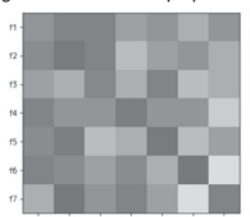


Fig.2: Evaluation of speaker coefficient control performance by filter

2-Q-46

2-Q-46 Towards Natural Emotional Voice Conversion with Novel Attention Module

☆Xunquan Chen¹, Jinhui Chen², Ryoichi Takashima¹, Tetsuya Takiguchi¹(¹Kobe University, ²Prefectural University of Hiroshima)

- ◆ Emotional Voice Conversion (EVC) technology aims to convert a neutral voice into an emotional one while keeping the linguistic information and speaker identity unchanged.
- ◆ In this study, we propose a novel EVC model, which can sufficiently learn the emotion information in both global-level and local-level. The global-level emotion information can be extracted by encoding the whole utterance into a fixed-length vector. For local-level emotion information, we adopt the attention mechanism for implementing time varying emotion representation.
- ◆ As show in Fig. 1, the generator is an encoder-decoder module in our work. Our attention module can appropriately embed an emotion feature which depends on the content information for another phoneme.

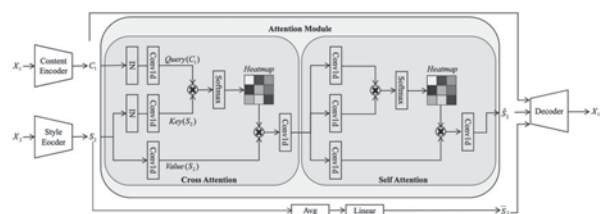


Fig. 1: The generator architecture of the proposed model.

2-Q-48

2-Q-48 条件付き制限ボルツマンマシンの平衡化傾向を利用したノンパラレル声質変換

Non-parallel voice conversion based on free-energy equilibration of conditional restricted Boltzmann machine

◎岸田拓也, 中鹿亘(電通大)

- ◆ 変換目標話者で条件づけて音響特徴量を学習したRBMを用いて、その自由エネルギーが平衡状態になるまで入力音響特徴量を反復的に更新し、変換後の音響特徴量とするノンパラレル声質変換手法を提案する。
- ◆ 自由エネルギーの特徴量空間に関するヘッセ行列は対角行列として近似することが妥当であると分かり、更新をニュートン法に近いやり方で効率的にできることが分かった。
- ◆ メルケプストラム歪みを指標とする客観評価実験において、提案法がARBMMベースの手法、StarGANベースの手法を上回る結果が得られた(Table 1)。
- ◆ 主観評価実験において、自然性評価ではMOSはStarGANを下回ったが、話者類似性評価ではStarGANと同等のMOSが得られた(Fig. 1)。

Table 1 MCD comparisons.

Pair	StarGAN	ARBMM	RBM
clb-to-bdl	7.17	7.58	7.24
clb-to-slt	6.84	6.38	6.27
clb-to-rms	6.93	7.83	6.64
bdl-to-clb	7.13	7.54	6.97
bdl-to-slt	7.44	7.75	6.94
bdl-to-rms	7.34	7.03	6.39
slt-to-clb	6.73	6.29	6.17
slt-to-bdl	7.45	7.90	7.48
slt-to-rms	7.35	7.92	6.92
rms-to-clb	7.01	8.00	7.40
rms-to-bdl	7.50	7.28	7.00
rms-to-slt	7.32	7.92	7.26

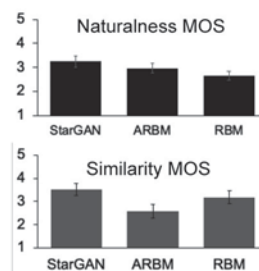


Fig. 1 MOS for naturalness and similarity. Error bars show 95%CI.