

3-1-1

3-1-1 アトミックノルム最小化による スパース時間周波数表現

Sparse time-frequency representation using atomic norm

◎草野翼, 矢田部浩平, 及川靖広(早大理工)

- ◆音響信号処理において時間周波数領域での処理が広く行われる。
- ◆短時間 Fourier 変換によって得られた表現は窓関数をかける影響が広がりをもつ。
- ◆本研究では、アトミックノルムを用いた周波数方向にグリッドレスなスパース時間周波数表現の推定手法を提案する。
- ◆アトミックノルムを用いることで ℓ_1 ノルムを用いたスパース推定において生じるグリッドの不一致を回避し、より局在化したスパース時間周波数表現が得られることを確認した (Fig. 1)。

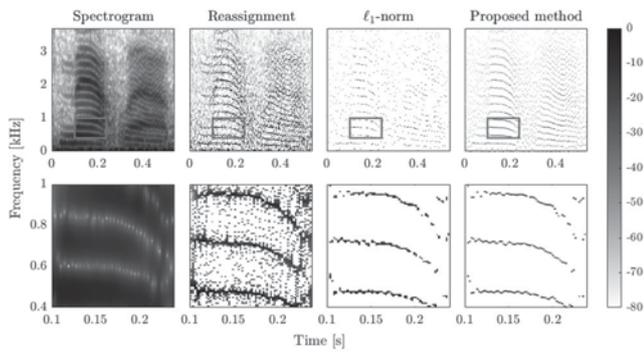


Fig. 1 Time-frequency representations of a speech signal. Each column shows (from left to right) the representations obtained by the discrete Gabor transform, the reassignment method, ℓ_1 -norm minimization, and the proposed method, respectively. The bottom row illustrates these enlargements in the gray box.

3-1-3

3-1-3 パラメトリックスピーカにおける矩形波キャリアに 基づく復調音の音圧改善の評価

Evaluation on Sound Pressure Enhancement of Demodulated Sound Using Carrier with Periodic Square-wave in Parametric Loudspeaker

☆旭浩平(立命館大院), 中山雅人(阪産大), 西浦敬信(立命館大)

- ◆パラメトリックスピーカは超音波の直進性を利用することで超指向性を実現し、特定の範囲にのみ音を届けることが可能である。しかしながら、パラメトリックスピーカは空気非線形性に基づいて可聴音を復調しているため、復調音の音圧が小さいという問題がある。
- ◆そこで、本稿ではキャリアが持つエネルギーおよびキャリアと復調音間における音圧の比例関係に着目し、矩形波キャリアに基づく復調音の音圧改善手法を提案する。Fig. 1に示すように矩形波キャリアは、キャリアの波形を矩形波とするため、従来のキャリアとして利用されている余弦波と比較して、信号が持つエネルギーが大きい。そのため、提案手法を用いることでキャリアの音圧が増幅し、それに伴い復調音の音圧改善が可能となる。
- ◆評価実験の結果、Fig. 2に示すように提案手法を用いることで復調音の音圧を改善可能であることを確認した。また高調波歪の評価実験、音質の評価実験より従来手法と同等の音質を保持したまま音圧改善が可能であることを確認した。

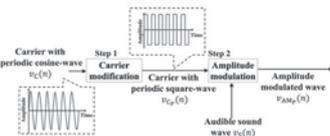


Fig.1: Overview of the proposed method.

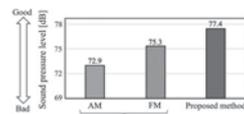


Fig. 2: Experimental results for sound pressure of the demodulated sound.

3-1-2

3-1-2 パラメトリックスピーカにおける調波構造の スペクトル傾斜制御に基づく音質改善

Sound Quality Improvement Based on Spectral Slope Control for Harmonic Structure in Parametric Array Loudspeaker

◎秋庭庭(立命館大院), 中山雅人(阪産大), 西浦敬信(立命館大)

パラメトリックスピーカ(PAL)は原理上、再生音圧が周波数に関係し、高調波歪みと混変調歪みが混入するので、復調音と音源の傾斜が異なり、音質が劣化する。調波信号の場合、成分の数が多いので、復調音の理論分析が極めて困難となる。そこで、本稿では、復調音と音源のスペクトル傾斜の差に着目し、音源にあえて傾斜をつけることで、復調音に本来の傾斜を再現し、音質の改善を試みる。予備実験で復調音と音源のスペクトル傾斜のモデルを構築し、評価実験で提案手法の有効性を確認した。

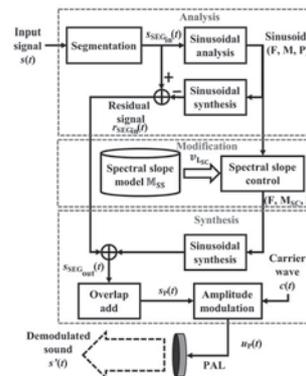


Fig. 1 Block diagram of proposed method

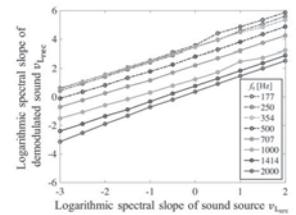


Fig. 2 Relationship of spectral slopes

Table 1 Results on spectral slope (average of difference to target value)

Non.	Conv.	Prop.
2.0	0.5	0.1

3-1-4

3-1-4 アルミ電解キャパシタの高調波歪み測定 方法に関する基礎的検討

Basic Measurement Method of Harmonic Distortion of Aluminum Electrolytic Capacitor

☆庄司晃太, 折野裕一郎, 宮岡洋平, 黒澤実(東工大・工)

- ◆今日において、音質の評価は定量的でなく主観に依存したものである。音楽保存のデジタル化により、再生のためのアナログ回路の重要性が高まった。本研究ではアナログ回路の素子の一つのアルミ電解キャパシタの高調波歪みの測定方法を検討した。
- ◆高調波歪みを測定するために、Fig. 1に示すブリッジと計装アンプで構成された回路を作成した。(i)フィルムキャパシタのみのブリッジと(ii)フィルムキャパシタとアルミ電解キャパシタのブリッジの2種類のブリッジを製作し測定した。
- ◆440 Hzの正弦波を入力し、出力信号を測定した結果をFig. 2に示す。(ii)フィルムキャパシタとアルミ電解キャパシタのみのブリッジは(i)フィルムキャパシタのみで構成されたブリッジより大きな高調波歪みを得られ、高調波の違いを観測することができた。

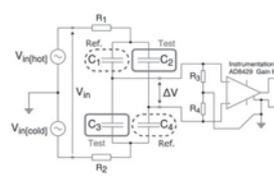


Fig. 1 Measurement circuit with a Wheatstone Bridge and Instrumentation Amp.

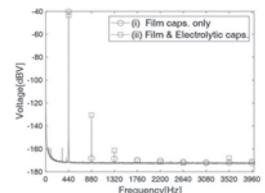


Fig. 2 Compare the bridge consisting of Film & Electrolytic caps. with that consisting of only Film caps.

3-1-5

3-1-5 音響レンズの焦点距離の縮小による 距離選択收音性能の向上

Improvement of distance-selective acoustical performance
by reducing the focal length of the acoustic lens

☆村上龍斗(産技高専), 高橋義典(工学院大), △高野邦彦(産技高専)

- ◆本研究は、距離選択收音を目的とした音響レンズの性能の向上を目指したものである。
- ◆周囲雑音に有効な收音方法としてマイクロホンアレーと信号処理を用いた音源分離があるが、実時間での処理は計算コストの問題から実現が困難とされている。また指向性マイクロホンは特定の方向からの收音に対しては高感度であるものの、距離についての選択性はない。
- ◆また、距離選択收音を目的とした音響レンズは焦点距離を縮小することで、距離選択性能が向上するとされているものの、実証例がない。本研究では実際に複数のホーンで製作された短焦点距離の音響レンズを試作し(Fig. 1) 特性の評価を試みた。(Fig. 2)

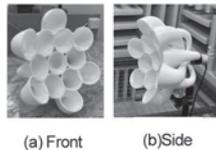


Fig. 1 Acoustic lens composed by hyperbolic

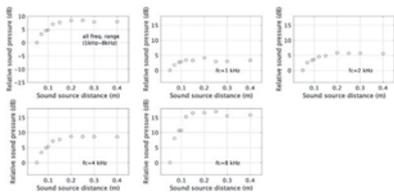


Fig. 2 Distance selection performance of normalized

3-1-7

3-1-7 ライン露光順次読み出し方式カメラ による抽出音声の調波構造復元

Harmonic Structure Reconstruction of Extracted Speech
with Rolling-shutter Camera

☆吉田敦史(立命館大院), 岩居健太, 西浦敬信(立命館大)

音波は空気中を疎密波として伝搬するため、音波が周囲の物体に衝突した際に、その物体を微小に振動させる。これを利用し、我々は、民生用カメラの1つであるライン露光順次読み出し方式カメラにより撮影した動画像から音声抽出する手法をこれまでに提案してきた。

しかし、抽出可能な音声の最大周波数はカメラの露光時間に依存するため、露光時間よりも短周期の振動は抽出できない。そのため、抽出音声の高域成分が欠落し、抽出音声の音質が低下する。そこで、Fig. 1 に示すように半波整流制限および GL アルゴリズムを用いた調波構造復元による音質改善法を提案する。Fig. 2 より、提案法により抽出音声の高域成分が復元されることが確認された。

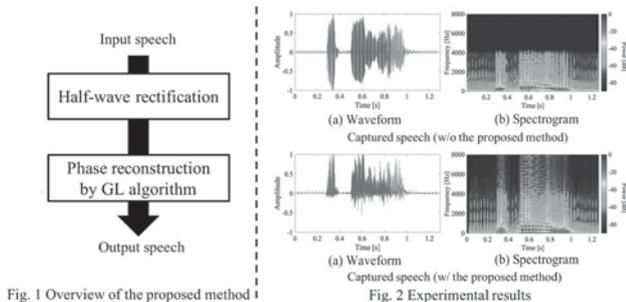


Fig. 1 Overview of the proposed method

Fig. 2 Experimental results

3-1-6

3-1-6 抵抗器における抵抗値変動が音響信号に 与える影響の検討

Effect of Resistance Fluctuation in Resistors on Acoustic Signals

◎宮岡洋平, 折野裕一郎, 黒澤実(東工大・工)

- ◆これまで議論されてきた抵抗器における高調波歪みに関する数理モデルでは、2次高調波歪みは交流信号に加えて直流信号を入力した際に生じるとされている。
- ◆しかし、これまでの測定結果から複数の抵抗器において直流信号を入力せずとも2次高調波歪みを確認している。
- ◆そこで抵抗器にて生じる $1/f$ ゆらぎを起こす抵抗値の変動 $\Delta r(t)$ に着目し、新たなモデルを提案した。そして Fig.1 のように抵抗値変動にランダムウォークの考えを用いることで、信号歪みに関するシミュレーションを行った。
- ◆Fig.2 にシミュレーション結果を示す。結果より入力信号 440 Hz の2次高調波にあたる 880 Hz 及び3次高調波 1320 Hz で信号歪みが確認された。

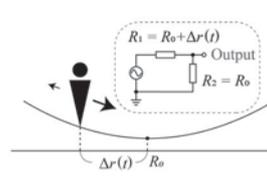


Fig.1: Model and simulation circuit for resistance fluctuation

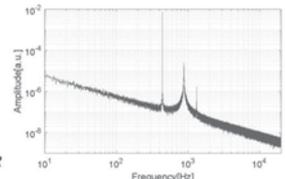


Fig.2: Results of resistance fluctuation simulation

3-1-8

3-1-8 ライン露光順次読み出し方式カメラを用いた 撮影動画像の RGB 輝度変位に基づく 音声検出

Speech Detection Based on RGB-Intensity Variation in Videos
Captured with Rolling-shutter Camera

☆芳澤翼(立命館大), 吉田敦史(立命館大院),

岩居健太, 西浦敬信(立命館大)

近年、音波によって振動する物体をカメラで撮影し、その撮影動画像から音声を検出する手法が提案されている。ライン露光順次読み出し方式カメラを用いた撮影動画像による音声検出手法では、その処理で位相限定画像を使用するため、グレースケール動画像から音声検出される。しかし、グレースケール動画像は階調のダイナミックレンジが小さく、音声の検出精度が低くなるという問題がある。

そこで、本稿では、音声の検出精度を向上させるために RGB 動画像を使用し、RGB 輝度変位から音声を検出する手法 (Fig. 1) を提案する。提案手法では、画像の行ごとの R, G, B それぞれの勾配を使用し、その類似度を算出することで音声を検出する。提案手法の有効性を確認するために、客観評価指標を用いて評価実験を実施した。客観評価実験の結果 (Fig. 2) より、従来手法に比べ音質が改善されることが示された。

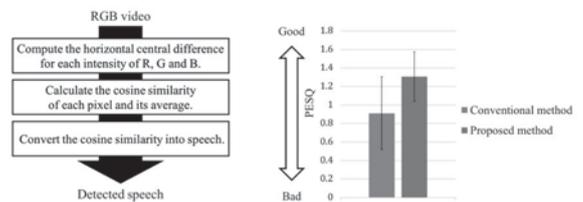


Fig. 1: Overview of the proposed method

Fig. 2: Evaluation results of PESQ

3-1-9

3-1-9 光差動検出による
ミッドフリンジロック干渉計の低雑音化
Reducing noise of mid-fringe locked interferometer
by optical differential detection

○石川憲治, △白木善史, 守谷健弘, △石澤淳, △日達研一, △小栗克弥 (NTT)



差動化による雑音低減量 単位周波数当たりの等価雑音量 計測用マイクに迫るSN比

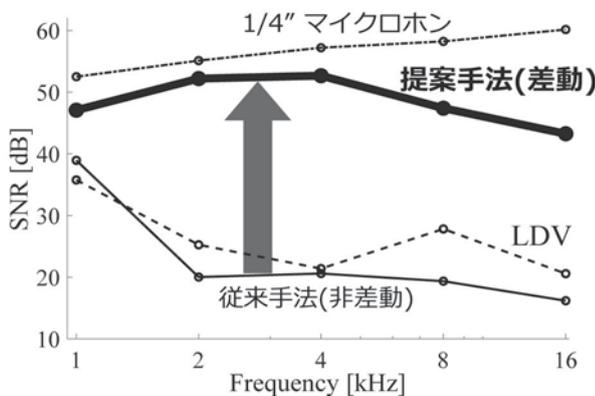


Fig.1 Results of 50 dB SPL sound-field measurement

3-1-11

3-1-11 ブリッジ回路を用いたケーブルでの相互変調歪み測定方法の検討

Measurement of Passive Intermodulation in Cables with Wheatstone Bridge

☆杉山公洋, 宮岡洋平, 折野裕一郎, 黒澤実(東工大・工)

- ◆ケーブルでの相互変調歪み(PIM, Passive Intermodulation)は通信分野にて問題となっており,オーディオの分野においてもケーブルの非線形性について評価する必要がある。
- ◆本研究ではブリッジ回路を用いた歪みの測定回路を提案した。2つの周波数の合成波をブリッジ回路に印加し,ブリッジの平衡点での電圧値の周波数スペクトルの評価を行った。
- ◆ブリッジ回路内およびブリッジ回路後段で発生する相互変調歪みを測定することができた。

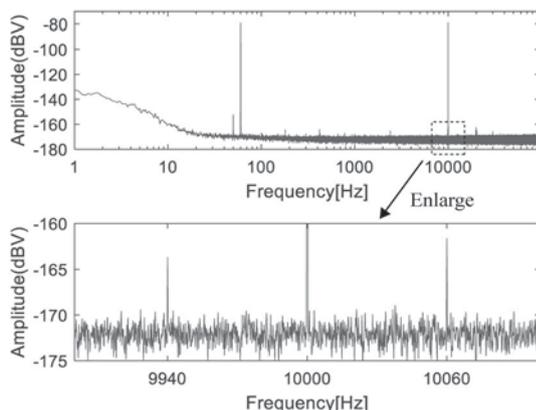


Fig. 1: The frequency Spectrum of a bridge circuit with tinned solid copper wire and tinned stranded copper wire

3-1-10

3-1-10 超音波素子の周波数特性に適した有限近似に基づく振幅変調を用いた高調波歪み低減

A Low Harmonic Distortion Amplitude Modulation for Bandlimited Parametric Loudspeaker using Taylor Expansion.

☆王浩南(立命館大院), 中山雅人(阪産大), 西浦敬信(立命館大)

パラメトリックスピーカの応用において, 超音波素子の周波数特性に適していない変調方式を用いると, 高調波歪みが多く発生する恐れがある。本稿では, 抑圧搬送波両側波帯(DSBSC: Double sideband with suppressed carrier)と呼ばれる振幅変調方式が素子の周波数特性に適していないとの問題点に着目し, テーラー展開を用いた有限近似に基づく歪み低減策を提案した。THD による評価実験の結果を下図に示す。500 Hz 以下の帯域で他の手法より歪みが少ないことを確認した。

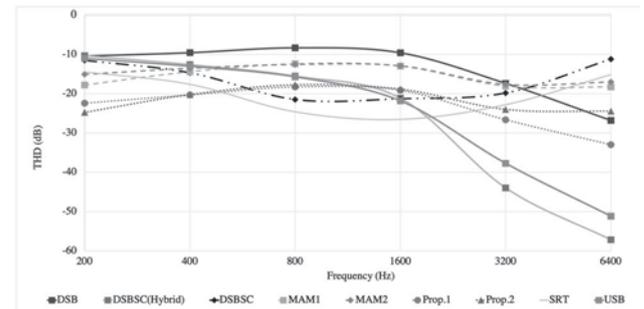


Fig. THD of several modulation techniques.

3-1-12

講演取消

3-1-13

3-1-13 ファジィ理論とベイズ推定の融合による音声信号に対する騒音抑制法

Noise suppression method for speech signal by jointly using fuzzy theory and Bayesian estimation

○折本寿子, 生田顯 (県立広島大), △長谷川浩治 (広島県立総合技術研究所)

- ◆騒音の混入した気導音を観測値とし、気導音観測データが得られた後の事後分布から、ベイズ推定を活用することにより新たな騒音抑制法を提案する。
- ◆骨導音の測定値から推定されるパラメータを反映した確率分布を事前分布として採用した。
- ◆骨導音をあいまいさが混入したファジィ・データとして捉え、ファジィ事象の確率を適用することにより、気導音と骨導音の関係を反映した手法である。
- ◆音声信号に対するシステム方程式を必要とせず、任意の音声信号に適用可能であることも特徴の一つである。

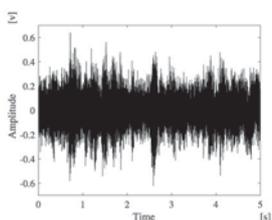


Fig.1: Observation of female noisy speech signal

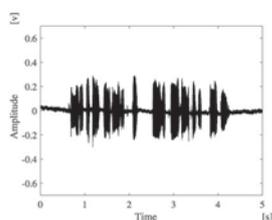


Fig.2: Estimation of female speech signal

3-1-15

3-1-15 音響イベント長とイベント非活性区間長の不均衡が検出性能に及ぼす影響

Impact of Imbalance Between Sound Event Duration and Inactive Frames on Performance of Sound Event Detection

○井本桂右(同志社大学), 美島 咲子, 荒井 友督, 近藤 玲史(NEC)

- ◆音響イベント検出におけるデータ不均衡問題について詳細に調査

- 音響イベントの継続長によって生じるデータ不均衡
- 音響イベントの発生/未発生区間に起因するデータ不均衡

- ◆4つの損失関数を用いて評価

- Simple reweighting loss
- Inverse frequency loss
- Focal batch Tversky loss
- Asymmetric focal loss (AFL)

$$E_{AFL}(\theta) = - \sum_{m,n=1}^{M,N} \left\{ \underbrace{(1-s(y_{n,m}))^{\gamma}}_{\text{Rewighting factor for active frames}} z_{n,m} \log(s(y_{n,m})) + s(y_{n,m})^{\zeta} \underbrace{(1-z_{n,m})}_{\text{Rewighting factor for inactive frames}} \log(1-s(y_{n,m})) \right\}$$

- ◆評価実験結果

- 音響イベントの発生/未発生区間に起因するデータ不均衡がより大きな影響を及ぼすことが明らかになった

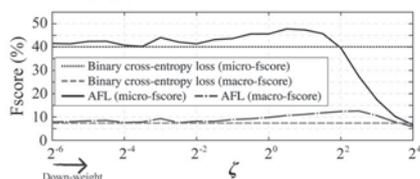


Fig.1: Detection results of sound events using BCE loss and AFL with CNN-BiGRU

3-1-14

3-1-14 音響イベントとシーンのマルチタスク学習における評価関数の重みの自動調整

Self-adjusting Loss Function for Multitask Learning of Sound Events and Scenes

○岩前 玲那, 白波瀬 壮, 高橋 皓大, 井本 桂右, 土屋 隆生(同志社大学)

- ◆マルチタスク学習に基づく音響イベントとシーンの同時分析
 - 音響シーンの情報が音響イベント検出性能を向上させる
 - 従来のマルチタスク学習では音響イベントと音響シーンの損失関数の重みの比率は学習を通して一定

- ◆提案手法

- 音響イベントと音響シーンの学習難易度が異なる点に着目
- モデル学習の進み具合に応じて自動的に評価関数の重みを調整する手法として、Multi-focal loss に基づくマルチタスク学習法を提案

- ◆評価実験結果

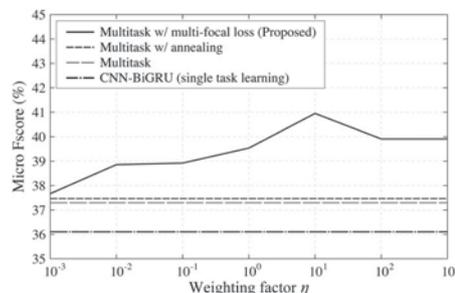


Fig.1: Detection results of sound events (Micro Fscore)

3-1-16

3-1-16 深層学習を用いた音をクエリとする類似効果音検索システム

A Similar Sound Effects Search System by Audio Query Using Convolutional Neural Network

☆小宮寛季, 小坂直敏(東京電機大)

- ◆電子音色辞書: 表記ゆれに強いテキストによる音検索を実現
 - 問題点: 求める音を正確にテキストで表現すること自体が不可能

- ◆目的: 音をクエリとして類似効果音を検索

- ◆手法: CNN ベースの分類モデルを用いて特徴量抽出 (128 次元)
 - 抽出した特徴量を主成分分析で圧縮 (20 次元)
 - ユークリッド距離を算出、距離が小さい順に上位 10 個を表示
 - 20 次元から 3 つを選び、音を空間に表示 (音色空間ブラウザ)

- ◆実験① 効果音分類モデルの構築 ⇒ 精度 83.4%

- ◆実験② 類似音検索: 15 種類のクエリで 2339 個の音から 10 件ずつ ⇒ 聴感上似ている音の件数: 119/150 = **79.3%**

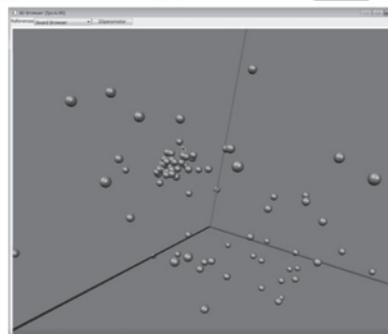


Fig.1: A prototype of 3D timbre browser

3-1-17

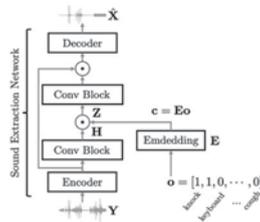
3-1-17 深層学習に基づく音響イベント抽出の検討

Investigation of deep learning-based acoustic event extraction

○落合 翼, デルクロア マーク, 小泉 悠馬
伊藤 弘章, 木下 慶介, 荒木 章子(NTT)

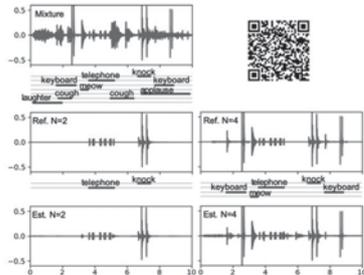
◆本稿の目的

日常生活において我々は、多様な音響イベント信号が混ざり合った音響環境の中で暮らしている。本稿では、状況に応じてユーザが聞きたい音のみを聞くことを可能とする技術(混合信号から特定種類の音響イベント信号のみを選択的に抽出/除去する技術)の検討について報告する。



◆評価実験 (Freesound Dataset Kaggle 2018 corpus を利用)

- 抽出事例 (7クラスの音響イベントを含む混合音声)
 - 1) 「knock, telephone」の2クラス抽出
 - 2) 「knock, telephone, keyboard, meow」の4クラス抽出



3-1-19

3-1-19 分散マイクロホンアレイにおけるDOA パーミュテーションの反復更新による複数音源定位

Multiple sound source localization with iterative updates of DOA permutations in distributed microphone arrays

☆田中一希(徳山高専), 若林佑幸, 小野順貴(都立大), 宮崎亮一(徳山高専)

- ◆空間的に分散配置されたマイクロホンアレイを用いて推定されたDOAから複数音源を定位する手法について検討を行う。複数音源の定位にはDOAと音源との対応関係が必要である。
- ◆本研究では、推定されたDOAと各音源との対応関係を表すパーミュテーションを導入し、これを反復更新する事で複数音源を定位するアルゴリズムを提案する。
- ◆本実験では残響の無い部屋をシミュレーションし、複数のマイクロホンアレイと音源を配置した。MUSIC法でDOAを推定し、提案アルゴリズムを用いて複数音源の定位を行なった。
- ◆実験により提案アルゴリズムは少ない計算時間で収束することを確認(Fig.1)し、各アルゴリズムの誤収束回数(Table 1)からアルゴリズムの性能を確認した。(M:マイクロホンアレイ数, N:音源数)

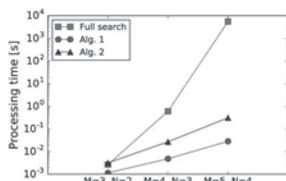


Table 1: Number of false convergence.

	M = 3, N = 2			M = 4, N = 3			M = 5, N = 4		
	A	B	C	D	E	F	G	H	I
Alg. 1	2	2	4	11	14	50	1	0	3
Alg. 2	0	0	0	0	0	14	0	0	0

Fig.1: Time to converge of each algorithm.

3-1-18

3-1-18 因果的DNNによる音響イベント検出・音源定位のオンライン化

Causal DNN for online sound event detection and source localization

☆長友健人(早大理工), 安田昌弘(NTT), 矢田部浩平(早大理工), 小泉悠馬(NTT), 及川靖広(早大理工)

◆音響イベント検出・音源定位(SELDO)

- 音源到来方向(DOA)と音源数の推定と音響イベント検出を同時に行う
- 運転アシストシステムなどへの応用が考えられている
- ⇒ 応用を考えるとオンライン動作が好ましい

◆従来手法は非因果的な構造を持つオフラインのDNNベースの手法

- オンライン化すると音響イベント発生付近の性能が低下するのが課題

◆物理-DNNベースのDOA推定手法のオンライン化を提案

- 物理ベースの手法は音響イベント発生区間に関わらず安定したDOA推定
- オンライン化によるDOA推定性能の低下を抑えることを期待

◆結果

- 提案手法はオンライン化によるDOA推定の性能低下の抑制を確認(Fig.1)

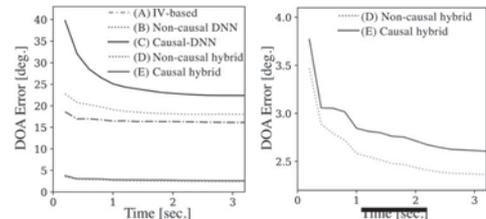


Fig.1 時間の経過によるDOA推定性能の変化。0 sec.は音響イベントの発生時を表す。右の図は左の拡大。(A)は物理ベースの手法。(B)はDNNベースの手法。(C)は(B)のオンライン化。(D)は物理-DNNベースの手法。(E)は提案手法。

3-1-20

3-1-20 複数マイクロホンアレイを用いた音源追跡に関する初期的検討

A study of sound source localization by multiple microphone arrays

☆岸本楓莉, △赤塚俊洋, 高橋徹, 中山雅人(阪産大)

- ◆3m間隔で配置した3つのマイクロホンアレイにより、1m/sで(0,0)から(9,0)へ直線移動する音源をGCC-PHATで定位する問題を考える。
- ◆3つのアレイで音源定位すると音源から遠いマイクロホンアレイの定位結果がばらつき結果を統合すると誤差が大きくなる問題がある。
- ◆音源の近傍のアレイで推定することを提案。
- ◆計算機シミュレーションにて3つのアレイより近傍2アレイの定位精度が高いことを確認。真値からの平均誤差が0.09mに改善。

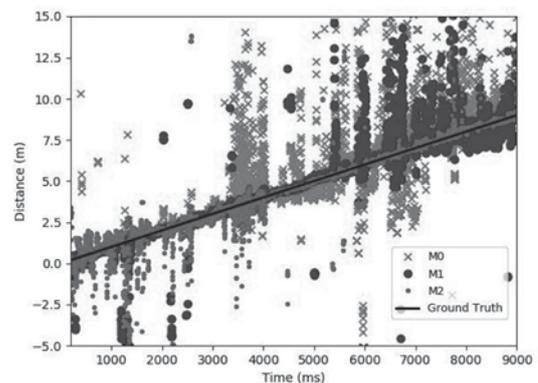


Fig.1: Distance between (0,0) and estimated position by three microphone arrays (×: MA0 (0,1.5), ∙: MA1 (3,1.5), ●: MA2 (6,1.5))

3-1-21

3-1-21 Method of estimating DOA based on monaural modulation spectrum

☆Rui Wang (JAIST), △Nguyen Khanh Bui (JAIST),
Daisuke Morikawa (TPU), Masashi Unoki (JAIST)

- ◆Aim: To propose a method of estimating DOA in a 3D space by using the monaural modulation spectrum (MMS).
- ◆Problems: Realization of monaural DOA estimation in 3D space and the improvement of the fitting model to reduce the root mean square error (RMSE) for monaural DOA.
- ◆Solution: Based on the modulation transfer function (MTF) and auditory perception of temporal modulation, extracting useful features in MMS and redesigning a polynomial fitting model for monaural DOA in 3D space using these features.
- ◆Evaluation: 51,840 simulations with several signal types and multiple subjects were carried out to simultaneously estimate the azimuth and the elevation of an incoming sound source.
- ◆Summary: The improved method could adequately estimate the DOA in 3D space with an overall RMSE of 5.59 degrees.

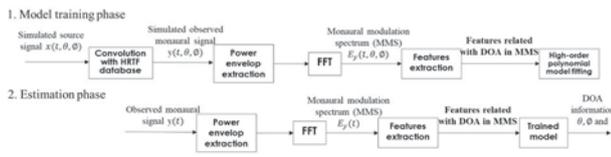


Fig.1 Scheme of the proposed method for monaural DOA estimation

3-2-1

3-2-1 ボコーダ波形生成における励振源の群遅延操作に向けた声帯音源特性の解析

Analysis of glottal source characteristics towards group delay manipulation of excitation source for waveform generation based on a vocoder

☆小口純矢, 森勢将雅(明治大)

- ◆目的: 声帯音源の群遅延特性を考慮した音源パルスの時間伸長
 - 最小位相フィルタ: 応答が時刻 0 に集中し音質劣化
 - 振幅特性はそのままに励振源の群遅延特性を操作し時間伸長
 - 声帯音源の群遅延特性を考慮する上で効率的な表現はないか?
- ◆従来法: 時間領域平滑化群遅延 [坂野 '01]
 - 群遅延を逆フーリエ変換→高次の成分を打ち切って平滑化
 - 群遅延の概形が音質に寄与
- ◆提案法: 時間領域平滑化/パワースペクトル重み付き群遅延
 - パワースペクトルを乗じた群遅延に対数を取り平滑化
 - ◇ 重み付け: 持続時間への影響が小さい成分の除去!
 - ◇ 対数: 概形成分を分離しやすくなる!

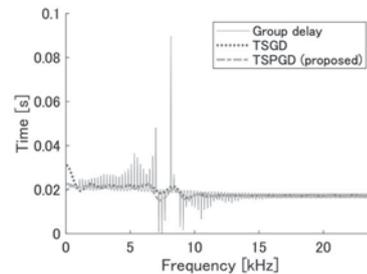


Fig.1: The glottal source's group delay and the smoothed group delays obtained by conventional and proposed method.

3-2-2

3-2-2 音声分析合成の波形に用いる励起信号の違いが知覚に与える影響

Influence of perception on the difference of excitation signal used for waveform generation in speech analysis synthesis system

森勢将雅, ☆田鎖佑弥(明治大)

- ◆目標: ボコーダの音声生成における無声音の励起信号の差を検証
 - 主観評価結果から差が生じた音声を調査
 - 主観評価結果と無声音の継続時間の相関を分析
 - 差が生じた音素について波形を観察し原因を解析
- ◆従来法: ホワイトノイズ
 - チャネルボコーダの無声音の生成で一般的に用いられる
 - 短時間パワーが不安定
 - 品質劣化につながる恐れがある
- ◆提案法: Modified velvet noise (MVN)
 - ベルベットノイズを音声合成向けに改良
 - DC オフセットを低減する
 - 短時間パワーの安定性に優れる
- ◆音源を個別に解析した結果
 - 摩擦音の合成では短時間パワーの安定性に優れる MVN が有利
 - 破裂音の合成ではホワイトノイズが MVN より有利

3-2-3

3-2-3 音声合成のためのサブバンド符号化音声に対する音声分析の検討

A study on speech analysis of subband-coded speech for speech synthesis

○西澤信行, △服部元 (KDDI総合研究所)

筆者らは音声合成のための波形生成を疑似 QMF バンクに基づくサブバンド符号上で行う方法を提案しているが、それに対応する音声分析方法として、正弦波の重量による音声波形モデルのパラメータをサブバンド符号上で推定する方法について検討した。予測残差最小化に基づく基本周波数推定において、波形上で行った同様の推定結果との間の差の RMS は約 0.1 セミトーンである。

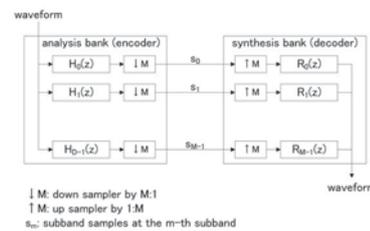


Fig. 1 Block diagram of the used subband coding system.

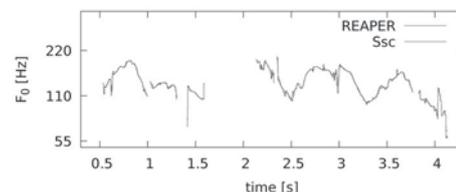


Fig. 2 Estimated fundamental frequency of A01 of ATR503 spoken by a male speaker (Ssc: the proposed method).

3-2-4

3-2-4 シャドーイング時の自己聴取音に起因する発話の乱れに対する分析とその対策

Analysis of inarticulate shadowing utterances caused by hearing shadowers' own voice

☆箱田峻, 朱伝博, 峯松信明, 齋藤大輔(東大)

- ◆シャドーイングとは、外国語発話能力訓練の一つで、モデルの外国語音声を聴取すると同時にその内容を復唱するという訓練法であるが、その円滑さが自己聴取音(発話時に聞こえる自分の声)を抑制することによって向上するかを分析的に実験した。
- ◆日本人及び中国人大学生に、英語音声及び日本語音声を、カナルイヤホン、オンイヤードホン、ノイズキャンセリングヘッドホン等、計6つのヘッドホンをそれぞれ着用してシャドーイングさせ、その音声と提示音声の書き起こしを見ながらシャドーしたスクリプト・シャドー音声との posteriorgram DTW 歪みをシャドーの歪み=即時聴解の崩れとして、ヘッドホン条件別に比較した。
- ◆主観評価により最も自己聴取音が抑制できていないと判断されたオンイヤードホンと他のヘッドホンのシャドー歪みを有意差検定すると、英語音声の場合は、予想と反して、シャドーの崩れは、オンイヤードホンの歪みが最も小さいという結果となった。これは、英語音声が発験者にとってシャドーが難しく、ヘッドホンごとの傾向が現れないほど崩れが不安定だったからだと考察される。
- ◆一方日本語では、ノイズキャンセリングヘッドホンの崩れがオンイヤードホンと比べ有意に小さくなった。また、オンイヤードホンより自己聴取音が抑制されたと評価されたヘッドホンも、有意差こそないものの、シャドー歪みが小さくなった。

3-2-6

3-2-6 マルチリンガルマシンスピーチチェーンを用いたゼロショットコードスイッチングの音声認識と音声合成

Multilingual machine speech chain for zero-shot code-switching ASR and TTS

☆中山佐保子, チャンドラ アンドロス, サクティ サクリアニ, 中村哲

- ◆会話の中で言語が切り替わるコードスイッチングは、音声とテキストがペアになったデータを集めるのが難しく、教師あり学習による音声認識と音声合成の開発が難しい。そのため、我々はマシンスピーチチェーンを利用した半教師あり学習の開発を進めている。
- ◆マシンスピーチチェーンは音声認識と音声合成をループ結合して互いに学習させる仕組みであり、入力に対する正解データがないラベルなしデータに対しても、互いのモデルを用いて推測させることで学習可能になる。
- ◆本研究では、Fig.1 のようにマシンスピーチチェーンの音声認識に言語識別(LID)、音声合成に言語IDの分散表現を組み込み、言語ID情報を用いることで、ゼロショットと呼ばれる学習に含まれない未知のコードスイッチングに対してもロバストな音声認識と音声合成を実現する。

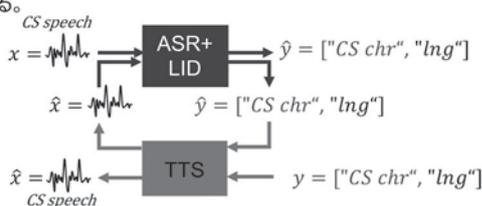


Fig.1: Multilingual machine speech chain.

3-2-5

3-2-5 音響特徴を用いたパーソナリティおよび年齢との関連に関する研究

Studies on the relationship between personality and age using acoustic features

○呂 昂, 党 建武(北陸先端大)

- ◆近年、パーソナリティの自動分類は注目されている。先行研究においては、機械学習、深層学習などの手法を用いて、音響特徴からパーソナリティを分類するモデルが提案されてきた。しかし、人の音響特徴は、特に青少年期に対して年齢に連れてかなり変化しており、パーソナリティは年齢によって変化していることも考えられる。本研究では、機械学習を用いて、異なる音響特徴を利用し、小・中・高と大学の学生のパーソナリティを分類し、その結果を分析した。結果として、音響特徴量によってパーソナリティ分類への貢献は異なり、その貢献度も年齢にも関連があることがわかった。

3-2-7

3-2-7 自己類似特徴量を用いたゼロ資源言語の音声検索の検討

Spoken Term Detection Using Self-Similarity Features for Zero-Resource Language

☆水落 智, 伊藤 彰則, 能勢 隆(東北大)

- ◆本研究では、ゼロ資源言語のデータベースに対する単語検索として、音声の自己類似性を用いる手法を検討した。検索対象言語はカクチケル語とする。
- ◆自己類似性を用いた単語検索には Self-Similarity Matrix (SSM) を使用する手法があるが、この手法では単語検出後の後処理として SSM を使うため、処理が複雑である。
- ◆SSM を用いる手法より簡単な処理で自己類似性を扱うために、自己類似性を特徴量として表した自己類似特徴量を用いて単語検出実験を行った。
- ◆自己類似特徴量を用いることで、単語検出精度の向上が確認された。
- ◆今後は SSM を用いる手法との比較を検討する予定である。

Table.1: Result of STD using Self-Similarity Feature

Feature	Target Language		
	Japanese	English	Kaqchikel
Baseline	76.28	59.90	40.36
Self-Similarity Matrix	77.56	60.22	42.32

※Evaluation index: Mean Average Precision (%)

3-2-8

3-2-8 MFCC-DPGMM Features for Enhancing Low-Resource ASR

○ Bin Wu (NAIST)
 ○ Sakriani Sakti (NAIST; RIKEN AIP)
 ○ Satoshi Nakamura (NAIST; RIKEN AIP)

Recently DPGMM features achieved top results in discriminating between linguistic units in Zerospeech challenges. This paper proposes to concatenate DPGMM posteriorgrams with speech features (MFCCs) to increase the phonemical discriminabilities, which enhances the ASR performances. By experiments on TIMIT and WSJ corpora, our proposal consistently improved the performances of the large vocabulary continuous speech recognition (LVCSR) and more absolute improvements at ASR with smaller datasets with fewer resources.

3-2-9

3-2-9 Building low-resource speech recognizer: Transfer learning and data augmentation

Narangerel Purevdorj, Ryota Nishimura (Tokushima University)
 Altangerel Ayush (Mongolian University of Science and Technology)
 Kengo Ohta, Norihide Kitaoka (Toyoashi University of Technology)

In this work, we attempt to use well-annotated data to build a low-resource speech recognizer using a hybrid attention/CTC network (Fig. 1). Especially the Mongolian language, which is much less data than that is normally needed to train conventional ASR systems. We used the following two methods:

- Data augmentation (DA) is a common, low-cost strategy that is often adopted to increase training data quantity.
- We adapted a high-resource language-independent end-to-end ASR model to a target language.

This paper presents a comparative study of training methods for ASR of low-resource languages, using DA and multilingual transfer learning methods with an S2S hybrid attention/CTC model and only a limited amount of target language training data. In Table 1 shows some results of DA based monolingual and multilingual transfer learning ASR.

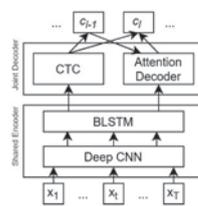


Fig. 1 Hybrid attention/CTC network.

Table 1: Results for DA based monolingual and multilingual transfer learning ASR

Results of DA based monolingual ASR (CER / WER)			
MN baseline	MN + DA (speed)	MN + DA (pitch)	MN + DA (All)
19.3 / 37.6	13.3 / 28.6	14.3 / 28.6	12.1 / 26.8
Results of multilingual transfer learning ASR (CER / WER)			
Language	with out script transliteration	with script transliteration	
EN	11.0 / 34.1	14.8 / 34.6	
JP	10.7 / 34.4	13.5 / 33.0	
KR	20.8 / 42.7	9.7 / 24.7	
TR	10.0 / 25.3	11.3 / 24.0	

3-2-10

3-2-10 Text to speech system for low resource languages by cross-lingual transfer learning and data augmentation

○ Zolzaya Byambadorj (Tokushima Univ.), Ryota Nishimura (Tokushima Univ.), Altangerel Ayush (Mongolian Univ. of Science and Tech.), Kengo Ohta (Toyoashi Univ. of Tech.), Norihide Kitaoka (Toyoashi Univ. of Tech.)

Deep learning techniques are now being applied in text-to-speech (TTS) systems, resulting in significant improvements in performance. These methods require large amounts of text-speech pair data however, and the task of collecting this training data is costly. In this paper, we propose a single-speaker, TTS system for languages with very limited amounts of text-speech pair data.

- ◆ We used the following methods to train the spectrogram prediction models of our TTS systems:
 - > cross-lingual transfer learning
 - > data augmentation
 - > a combination of the previous two methods

We also compared single-speaker and multiple-speaker training methods, using sequential and simultaneous training, respectively. We evaluated the performance of our models using a mean opinion score (MOS) test and found that using both cross-lingual transfer learning and augmented data during training resulted in the most natural synthesized Mongolian speech output.

- ◆ Next, we trained two Parallel WaveGAN neural vocoders, one using our augmented data and one using the entire 12 hours of target language data. Our subjective AB preference test indicated that the neural vocoder trained with augmented data had the same perceived speech quality as the vocoder trained with the entire target language dataset.

Only about 30 minutes of target text-speech pair data was used to train the spectrogram prediction models and generate augmented data using the Sox tool, demonstrating that our proposed TTS system could achieve reasonable performance using a very limited amount of target language training data.

3-2-11

3-2-11 Training data proliferation for environmental sound classification with neural network

☆Zhang Yutao(Ritsumeikan Univ.),
 Wakabayashi Yukoh(Tokyo Metropolitan Univ.),
 Iwai Kenta and Nishiura Takanobu(Ritsumeikan Univ.)

Insufficient of categories included in neural network acoustic model is one of key factors that influence the performance in practice application. However, preparing labeled training data for adding new categories always come at a high cost. In this paper, we proposed a training data proliferation method with variational autoencoder (VAE) to generate training data for new class from very little labeled sample, shown in Fig. 1. In the proposed method, a mean vector and a standard vector corresponding to the sample of the new class are calculated with VAE which is trained by unlabeled training dataset. Proliferation training data for new class can be generated with these two vectors. The experiment results have shown the effectiveness of our proposed method.

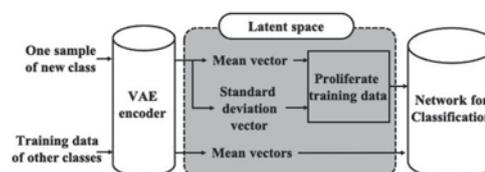


Fig. 1 Flowchart of the proposed method

3-2-12

3-2-12 余韻を含む打撃音や水滴落下音などの音響特徴量

Acoustic features of resonant transient sound such as waterdrop

○中島風太, △佐々木健(東京大学)

- ◆環境音認識は動画のシーン解析や機械の異常検出等に用いられる。その認識精度の向上には、各種の音を的確に識別できる特徴量の抽出が重要である。本研究では水滴音に着目した。水滴音は特徴的な音であり、水滴音の認識は家庭環境や産業現場における水漏れ等の異常検出にも利用できる。
- ◆まず、ヒトの聴覚により水滴音のおおまかな分類を行った。スペクトログラムの分析により、その分類は突発部の周波数成分の分布や時間的に持続する周波数成分の違いに対応していることがわかった。
- ◆水滴音の特徴と似た特徴を持つ水滴音以外の突発音、例えば木製品や金属製品を叩く音、についても同様の特徴量を計算した。さらに突発部の周波数分布の積分値をシグモイド曲線等の曲線に回帰したパラメータを用いて周波数分布の特徴を抽出した。
- ◆これらの特徴量を用いてk-最近傍法で突発音を分類し、識別精度を評価した。また音声認識における標準的なMFCCを特徴量として用い、識別精度を比較した。その結果、分類全体の正確さは汎用性の高いMFCCの方が優れていたが、水滴音を水滴音として特定するという観点では提案した特徴量の有効性が確認された。

3-2-13

3-2-13 聴取者ごとのグループソフトマックスによる音声感情認識

Speech emotion recognition by using listener-wise group softmax

○太刀岡 勇気(デンソーアイティラボラトリー)

- ◆聴取者ごとのばらつきに対応するため、聴取者ごとに分類器を作る方法が提案されている
- ◆一方、分類器の不均衡に対応するため、複数の分類ラベルに対してそれぞれの分類器のsoftmax結果を対応させるgroup softmaxの方法が提案されている
- ◆ここでは従来法と同様に聴取者ごとに分類器を作りつつ、group softmaxにより1つのモデルで同時にそれらを学習することで認識精度を向上させる方法を提案する

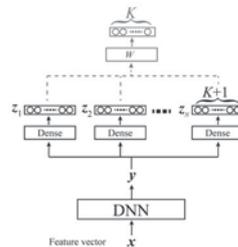
Fig.: Emotion recognition model for N listeners and K emotions

Table: recognition accuracy.

	speech	text	speech & text
conventional	56.3	54.1	64.9
proposed	60.7	54.8	66.3

3-2-14

3-2-14 Effects of Gender and Language Features on Text-independent Singer Verification

★Fanbin Mao, Daisuke Saito, Nobuaki Minematsu (UTokyo),

Chang Zeng (NII)

- ◆For singer verification, it's crucial to extract a robust embedding, which can fully represent one singer's distinctive feature. However, most current models utilize only singer labels, which may limit the discrimination performance of the networks.
- ◆Multitask learning (MTL) is an approach to inductive transfer that improves generalization by using the domain information contained in the training set of related tasks as an inductive bias. This paper introduces MTL method to combine language and gender features for enhancing the effect of embeddings.
- ◆Three effective MTL-based singer verification systems are proposed, including MT-SL, MT-SG and MT-SLG.
- ◆The proposed systems achieved 13.40%, 4.49% and 18.37% relative improvements compared with the i-vector baseline model in the EER respectively.

3-2-15

3-2-15 クロスチャネル言語識別におけるwav2vecを用いた教師なし特徴量学習

Unsupervised Feature Learning based on wav2vec for Cross-channel Spoken Language Identification

☆吉本拓真(神戸大/NICT), 沈 鵬, Xugang Lu (NICT), 高島遼一, 滝口哲也(神戸大), 河井 恒(NICT)

- ◆言語識別における特徴抽出手法としてi-vectorやx-vectorなどがあり、いずれも優れた性能を示すが、学習データと異なるチャンネルのデータを入力した際に性能が劣化してしまう問題がある。
- ◆先行研究では、音声認識モデルから得られる音素ベクトルをx-vectorの補助知識として用いる手法が提案されているが、大量のラベル付きデータが必要であり、対応できる言語に限られてしまう。
- ◆本研究では、教師なし学習に着目し、x-vector抽出モデルの入力として用いる、言語情報が多く含まれた特徴量を作成することを考える。
- ◆wav2vecの構造を用いて自己教師あり学習を行い、その出力をx-vector抽出モデルの入力とする。これをx-vector抽出モデルの入力にMFCCを用いるベースラインと性能を比較する。
- ◆実験より、ラベル付きデータが少ない場合においても自己教師あり学習を用いることで言語識別の性能を改善できることが分かった。

Table 1 Experimental results of baseline and the proposed method.

Method	Feature	C_{avg}	EER%
baseline	30-dimensional MFCC	0.32	32.7
	C (512-dimensional \mathcal{Z})	0.20	20.9
	C (256-dimensional \mathcal{Z})	0.19	21.1
proposed	C (128-dimensional \mathcal{Z})	0.16	18.5
	C (64-dimensional \mathcal{Z})	0.13	13.4
	C (32-dimensional \mathcal{Z})	0.21	23.3

3-2-16

3-2-16 日本語アイヌ語混合音声における言語識別

Language Identification on recordings of Ainu and Japanese Mixed Speech

◎松浦孝平, 三村正人, 坂井信輔, 河原達也(京大)

◆ アイヌ文化の一次資料は日本語とアイヌ語の混合音声(例: 図1)

- ◇ 一話者が両言語を話すため一般の言語識別より高難度
- ◇ 後続の処理のため、あらかじめ言語の識別が必要

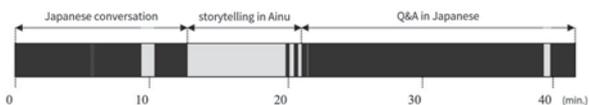


Fig. 1: A visualized example of a recording. White parts are spoken in Ainu.

◆ 提案法の概略図は右図2

◆ 言語識別の結果は下表1と図3

- ◇ 比較手法は言語識別の従来手法 (=Cai モデル)
- ◇ 同一話者の混合音声 (Mixed Speech) を利用すると高精度だが現実的ではない
- ◇ 提案手法では混合音声を用いなくても一定の認識性能を確認

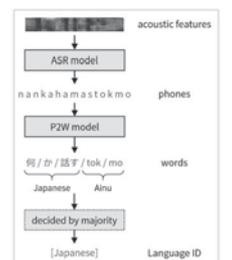


Fig. 2: An overview of our proposed method

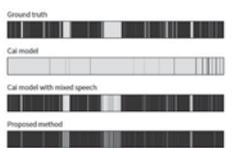


Fig. 3: The visualized ground truth and hypotheses. White parts are spoken in Ainu.

Table. 1: Framework F-measure values and accuracies (%) on each method

	Cai Model	Cai Model w/ Mixed Speech	Proposed
precision	31.0	84.2	88.3
recall	98.8	49.5	25.5
F-measure	47.2	62.4	39.6
Macro-F1	25.8	75.1	62.2

3-2-18

3-2-18 ドメイン適応と相互情報量最小化による disentangled な話者・言語表現に基づいたクロスリンガル音声合成

Disentangled speaker and language representations using domain adaptation and mutual information minimization for cross-lingual TTS

☆辛 徳泰(LINE, 東大院・情報理工), 小松 達也(LINE), 高道 慎之介(東大院・情報理工), 猿渡 洋(東大院・情報理工)

- ◆ ドメイン適応と相互情報量最小化による disentangled な話者・言語表現に基づいて、日本語(英語)話者の英語(日本語) 音声合成できるクロスリンガル音声合成の方法を提案する。
- ◆ 従来法は entangled な話者・言語埋め込みを使っている為、言語埋め込みのみを変えても話者性が変わる(Fig.1 上)。
- ◆ 提案法はまずドメイン適応を使い、違う言語の話者埋め込みの分布を統一し、更に話者・言語埋め込みの相互情報量を最小化することにより、disentangled な埋め込みを生成する。これにより、話者性を保つことが可能になる。(Fig.1 下)
- ◆ 実験結果は提案法がより高い自然度と話者類似度を持つクロスリンガル音声合成できることを示す。

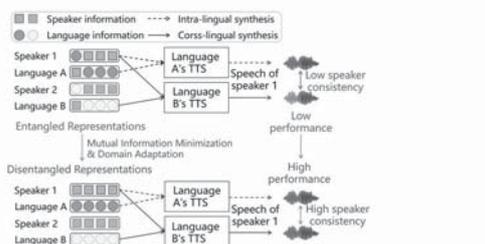


Fig.1: Comparison between entangled representations (top half) and disentangled representations (bottom half) for cross-lingual TTS synthesis.

3-2-17

3-2-17 Multi-granularity analysis of online intelligibility of L2 speech based on reverse shadowing

☆Chuanbo ZHU, Ryo HAKODA, Daisuke SAITO,

Nobuaki MINEMATSU(The University of Tokyo),

Noriko NAKANISHI(Kobe Gakuin University)

- ◆ In this study, DTW-based annotation of intelligibility using reverse shadowing is calculated in different units: frame, phoneme, and word.
- ◆ These annotations in different units are compared to objectively calculated intelligibility scores, which are based on manual transcripts.
- ◆ DTW scores were observed highly correlated to objective intelligibility scores based on transcripts in phoneme and word unit. This proved the validity and effectiveness to annotate online intelligibility in variable granularity.

Correlation between		
Shadowing-based score	Transcription-based score	
Phoneme unit	Phoneme unit	0.937
word unit	word unit	0.950

3-2-19

3-2-19 A Supervised Model for Emotion Embeddings and Its Application to Controllable and Emotional Speech Synthesis

☆Jiaqi ZHANG, Daisuke SAITO,

Nobuaki MINEMATSU(The University of Tokyo)

- ◆ In this paper, we propose a supervised CVAE-Tacotron 2 model for building an emotional Text-To-Speech (TTS) system where synthesis can be controlled by emotion categories. Previous work focus on using unsupervised variational autoencoder (VAE) to discover latent representations of prosody and then interpret them after training, which requires a lengthy post-processing stage and not practical for emotional TTS task.
- ◆ Also, as it is hard to acquire a public and large emotional dataset with high quality labels, we investigate the possibility to use transfer learning to adapt a neutral system to an emotional TTS system.
- ◆ Experiment results show conditional variational autoencoder (CVAE) learns emotion embeddings reliably and helps Tacotron 2 to generate emotional speech controllably, where emotions can be easily perceived by listeners.
- ◆ Future work will focus on extension methods for controlling speaker identities and possible improvements on audio quality by extra noise reduction.

3-2-20

3-2-20 CTC-Transformer 音声認識における自己知識蒸留の検討

A Study on Self-Distillation for Improving CTC-Transformer-based ASR Systems

◎森谷崇史, 落合翼, 菊田成樹, 佐藤宏, 田中智大, 芦原孝典, 増村亮, 篠原雄介, デルクロアマーク(日本電信電話株式会社)

- ◆先行研究では Encoder-Decoder (S2S) の特徴である注意重みを活用した Connectionist Temporal Classification (CTC) への知識蒸留を提案
- ◆本研究では, 高性能な Encoder-Decoder である CTC-Transformer の改善に提案手法を“自己蒸留 (SD)”として応用
- ◆ランダムに初期化したモデルを学習する上で自己蒸留の損失は性能低下につながるため, 学習時の Decoder のトークン正解率によって動的に変化する係数についても提案
- ◆5つのコーパスを用いた音声認識の実験において, 提案手法の自己蒸留 (+SD) により学習したモデルは, 全てのタスクにおいて S2S+CTC 損失のみで学習したモデルよりも認識性能の改善を確認

Table 1: Results of each CTC-Transformer.

System	WER [%]						CER [%]					
	WSJ		Switchboard		Librispeech		CSJ			NITVS		
	Dev93	eval92	SWB	CH	clean	other	E1	E2	E3	E1	E2	E3
Baseline	7.9	4.9	9.1	18.0	4.4	10.8	5.1	3.8	4.2	2.7	3.6	5.0
+SD	6.9	4.2	8.9	17.3	4.4	10.0	5.1	3.6	3.9	2.3	3.3	4.2
++L+LM	-	-	7.9	15.7	2.4	5.6	4.6	3.4	3.7	2.0	3.1	3.7

3-2-22

3-2-22 雑音のある通信路モデルに基づく E2E ASR システムにおける言語モデル増強

Language Model Augmentation in End-to-End ASR Systems

Based on Noisy Channel Model

☆Zhuo Gong, Minematsu Nobuaki, Saito Daisuke(The University of Tokyo)

Language models (LM) have played crucial roles in automatic speech recognition (ASR) to enhance performance of end-to-end ASR systems. Basically, there are two categories of approaches, one is finding better ways to integrate LMs into ASR systems, the other one is to adapt on LMs to the task domain. This article will start with reflection of interpolation based integration methods of end-to-end ASR's scores and LM's scores. Then we will focus on LM augmentation approaches based on noisy channel model, which is intrigued by insights obtained from the above reflection. The experiments show that we can enhance an ASR E2E model which is based on encoder-decoder architecture by pre-training the decoder with text data. This implies that the decoder of an E2E model can be treated as a language model and reveals the possibility of enhancing the E2E model without an external LM.

3-2-21

3-2-21 挿入操作に基づく End-to-End モデルによる音声認識と音声区間検出

END-TO-END ASR AND AUDIO SEGMENTATION WITH NON-AUTOREGRESSIVE INSERTION-BASED MODEL

◎藤田悠哉(ヤフー), 渡部晋治 (Johns Hopkins Univ.), 大町基(ヤフー)

- ◆音声区間検出と非自己回帰的な音声認識処理を単一のモデルで実現する手法を提案する。
- ◆挿入操作に基づく非自己回帰モデルと CTC を同時学習するモデルに対し, 因果的な自己注意を導入した。
- ◆因果的な自己注意の導入により, CTC 部分を用いた音声区間検出を実現した。
- ◆日本語話し言葉コーパスを用いた評価により, 提案したモデルは CTC よりも高い認識精度を達成した。

Tab. 1: CER of CSJ of unsegmented audio. B is the block length which is the theoretical latency of the model.

BlockSA	Eval1		Eval2		Eval3	
	CTC	Prop.	CTC	Prop.	CTC	Prop.
$B = 4$	16.1	11.1	12.8	8.4	23.3	16.6
$B = 8$	14.1	10.8	10.8	8.0	20.8	16.7
$B = 16$	12.8	10.5	9.4	7.6	18.7	16.4
$B = 32$	11.6	10.2	8.3	7.2	16.8	16.0

3-2-23

3-2-23 隠れセミマルコフモデルに基づく構造化アテンションを用いた

Sequence-to-Sequence 音声合成

Sequence-to-sequence speech synthesis using a hidden semi-Markov model based structured attention mechanism

☆角谷健太, 吉村建慶, 高木信二, 橋本佳, 大浦圭一郎, 南角吉彦, 徳田恵一(名工大)

本研究では, HSMM の構造を組み込んだ Seq2Seq モデリングに基づく音声合成手法を提案する。提案手法は Variational Auto-Encoder(VAE) に基づく生成モデルとして構成されており, HSMM と同様にアライメントを隠れ変数とする生成モデルとして定義される。提案法の核となるアイデアは, アライメントを推定するアテンション機構(VAE におけるエンコーダ)に隠れセミマルコフモデルの構造を組み込むことであり, これは構造化アテンションの一種と見なすことができる。提案法によって得られるアライメントはモノトニックであり, 系列として一貫性のあるアライメントを正確に取り扱うことが可能なため, 従来の Seq2Seq モデルに比べて少量の学習データで高品質なシステムの構築が期待できる。また, 継続長を陽に取り扱うために, 音声合成として継続長のコントロールが可能となる。

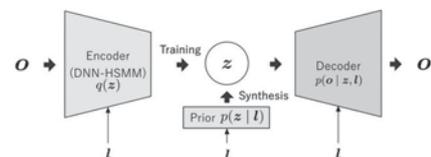


Fig.1: 提案モデルのイメージ図

3-2-24

3-2-24 学習時と合成時の一貫性を考慮した VAE に基づく自己回帰型 sequence-to-sequence 音声合成

Variational autoencoder-based autoregressive sequence-to-sequence speech synthesis considering consistency between training and synthesis
☆藤本崇人, 橋本佳, 南角吉彦, 徳田恵一(名工大)

本稿では、自己回帰モデルにおける学習時と合成時の不一致を軽減するための手法として、学習時と合成時の一貫性を考慮した VAE に基づく自己回帰型 sequence-to-sequence 音声合成を提案する。提案モデルは最終的な観測ではなく、潜在変数において自己回帰を行うことで、学習時と合成時に共通した自己回帰生成が可能となる。

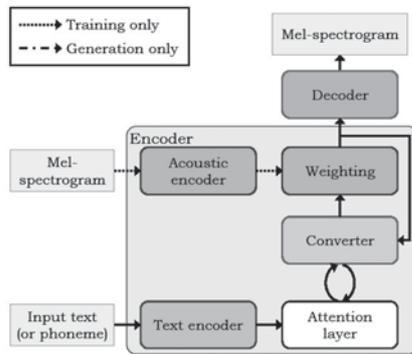


Fig. 1: 提案モデルの概要図

3-2-26

3-2-26 言語モデルによる文横断情報を用いた オーディオブック音声合成の検討

Audiobook Speech Synthesis using Cross Sentence Information by Language Models
☆中田亘(東大工学部), 郡山知樹, 高道慎之介(東大院・情報理工), 井島勇祐, 増村亮 (NTT), 猿渡洋(東大院・情報理工)

- 背景: オーディオブック音声合成
 - 長い文脈に対応した韻律を実現する必要がある
- 手法: 言語モデルに複数文を入力することによって得られる文横断情報を用いた、音声合成
 - BERT に当該発話文+前の 2 文を入力することで、文脈を考慮した単語分散表現、文書ベクトルを用いて Tacotron2 エンコーダ出力を拡張
- 実験: 言語モデルに単文を入力するモデルと複数文入力するモデルを比較
 - 平均メルケプストラム距離 (MCD) および Gross Pitch Error が言語モデルに複数文入力することにより改善
 - 当該発話文が短い場合において、言語モデルに複数文入力することにより MCD の値が改善

3-2-25

3-2-25 複数話者を対象とした非自己回帰型ニューラル音声合成

Non-autoregressive neural speech synthesis for multiple speakers
☆上乃聖, 三村正人, 河原達也(京大)

- ◆ 多数話者音声合成は音声認識のデータ拡張に有用であり、近年、非自己回帰型の音声合成により、より高速に生成ができるようになった。
- ◆ 本研究では、FastSpeech 2 を用いて、話者 ID を挿入するモデルと挿入しないモデルを構成する。
- ◆ 主観評価・音声認識の拡張実験の両実験において、話者 ID を挿入した方が主観評価が高く単語誤り率も改善した。

Table 1: MOSs for naturalness and generation speed (original: 1.32 hours) using synthesized speech of CSJ

Model	Naturalness	Speed (s)
Original	4.93 ± 0.29	—
Tacotron 2	3.20 ± 1.08	117.3
FastSpeech 2		
with speaker ID	3.08 ± 1.21	22.5
without speaker ID	2.97 ± 1.14	24.1

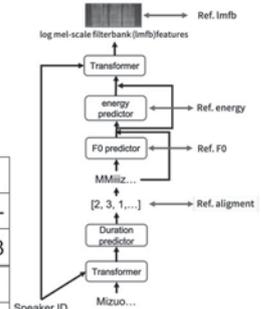


Fig. 1: FastSpeech 2 with speaker ID

Table 2: ASR performance (WER (%)) for Librispeech. We used paired-data of train-clean-100 in Librispeech (100h). We generated speech from train-clean-360 in Librispeech (360h).

Model	dev-clean	test-clean
Oracle (460h)	5.58	5.80
Baseline (100h)	11.39	11.70
FastSpeech 2 with speaker ID (100+360)	8.67	8.88
FastSpeech 2 without speaker ID (100+360)	9.23	9.93

3-4-1

3-4-1 感音性難聴者の語音明瞭度と初期音響エネルギーとの関係

Relationship between speech intelligibility of sensorineural hearing loss and initial energy of Japanese monosyllables
○下倉良太, 赤松義浩(大阪大)

- ◆ 本研究は、医療機関で得られた感音性難聴者の語音明瞭度のデータをもとに、単音節別正答率と初期音響エネルギーとの関係を検証した。
- ◆ 感音性難聴者 75 名を対象に、57-S 語表 (女性発話の単音節 50 語のデータセット) を用いた語音弁別検査を行い得られた明瞭度別に明瞭度 70~100% の群 (軽度), 明瞭度 50~68% の群 (中等度), 明瞭度 0~48% の群 (重度) の 3 群に分類し、それらの各群での各単音節別正答率を引用した。
- ◆ 初期音響エネルギーは、単音節の Sound Exposure Level (SEL) を計算した。SEL は A 特性音圧レベルの時間累積和を示す。
- ◆ Voice Onset Tim が非負の単音節 28 語の結果を重度別に Figure 1 に示す。横軸は SEL がある値に到達するまでの時間を対数化したもので、短いほど初期音響エネルギーの立ち上がり急峻であることを表す。重度難聴ほど両者の相関が高く、初期音響エネルギーの立ち上がり急峻な単音節ほど聞き取りやすい (Figure 1c)。

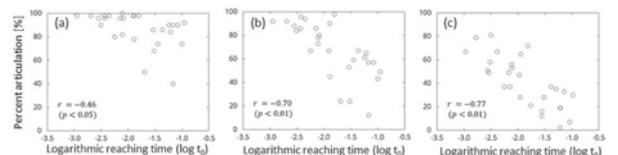


Figure 1. Relationships between percent articulation and the logarithmic times until the $L_{50}(f)$ for (a) mild, (b) moderate and (c) severe hearing loss.

3-4-2

3-4-2 異常検知技術を用いた嚥下機能評価手法に関する研究 (第2報:音響特徴量を用いた嚥下機能評価手法に関する研究)

Evaluation of Swallowing Function Using Anomaly Detection Techniques
(2nd Report : Evaluation of Swallowing Function Using Acoustic Features)

◎岡田真(ITIC), △本村美和(IPU)

- ◆ 昨今 IoT や AI, 5G といった最先端のデジタル技術を活用した遠隔医療や予防医療サービスの創出が始まっており, 生体信号のデジタル化とデータ活用技術は今後一層飛躍が期待される分野である。
- ◆ 在宅などで簡単に嚥下機能評価ができるようになれば, 症状が悪化する前に専門的な診療に患者を導くことができる。
- ◆ 前回の我々の報告では, 嚥下音から音響特徴量を抽出し, 教師あり学習を用いて嚥下機能の評価を行うアルゴリズムを提案したが, 真陽性率の改善が課題であった。
- ◆ 今回, 嚥下機能評価に有用な特徴抽出手法を検討し, 異常検知アルゴリズムを用いて新たな嚥下機能評価アルゴリズムの構築を行った。
- ◆ 構築したアルゴリズムを一個抜き検証法により評価した結果, Table1 に示す結果となり, 真陽性率 92%, 偽陽性率 17% を達成した。

Table 1 Result of leave-one-out cross-validation

		真値	
		陰性	陽性
予測値	陰性	53	1
	陽性	11	12

3-4-4

3-4-4 骨伝導イヤホンを使った 保育園環境下での聞こえ A study of listening capability of bone conductive earphone under nursery school environment

☆堀内契介, △赤塚俊洋, 高橋徹, 中山雅人(阪産大)

- ◆ 保育園環境下, すなわち園児達による LAeq 81.4dB 雑音環境下での骨伝導イヤホンの聞こえについて調査。骨伝導イヤホンで保育士に音声提示による情報支援の可能性について検討。
- ◆ 4 モーラ無意味単語を 4 つの異なる SNR (= -11dB, -14dB, -17dB, -20dB) で提示し, 骨伝導イヤホン, 密閉型ヘッドホン, オープンエア型ヘッドホンで聞こえを MOS で比較。3 名による聴取実験を実施。
- ◆ 骨伝導イヤホンによる -11dB 音声提示時の聞こえが, ヘッドホンによる -20dB 音声提示程度またはそれ以下の聞こえであると確認された。

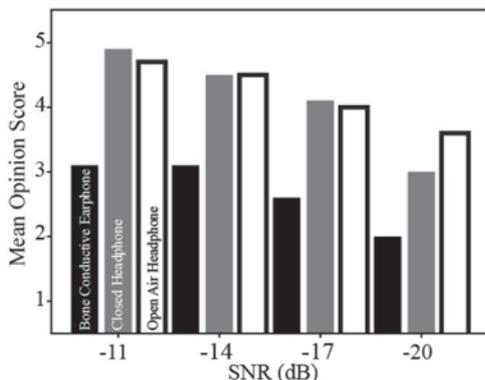


Fig. 1: Mean Opinion Score by subject A (Black: Bone Conductive Earphone, Gray: Closed Type Headphone, White: Open Air Headphone)

3-4-3

3-4-3 宣伝音声の聴取によって生じられる感情と 購買意欲の関係性の検討

Initial analysis of the relationship between consumer's willingness to purchase and emotions perceived from advertising speech

◎長野瑞生, 井島勇祐, 廣谷定男(NTT)

- ◆ 感情を媒介とする消費者行動モデル [A. Mehrabian+, 1974]
 - 消費者の購買行動を外部刺激→感情→購買意欲の 3 ステップで説明するモデル
 - 先行研究では混雑状況や音楽のような外部刺激を用いた検討が多く, 音声による感情や購買意欲への影響は不明
- ◆ 本研究では F0, 話速, スペクトル傾斜が感情及び購買意欲へ及ぼす影響について検証し, 従来の消費者行動モデルへの適用を検討
- ◆ 宣伝音声を用いた大規模な主観評価実験を実施し, パス解析, 媒介分析により音声特徴と感情, 購買意欲への影響を分析
 - F0 と話速は快感情と覚醒に対して, 快感情と覚醒は購買意欲に対して有意な影響
 - パス解析の結果, 音声特徴から購買意欲を求める 2 層モデルよりも, 感情を媒介とした 3 層モデルの適合度が高い
 - 媒介分析の結果, 快感情と覚醒の間接効果が有意
 - 従来と同様の消費者行動モデルが適用可能であると考えられる

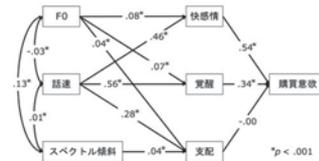


Fig. 1: Path analysis of three layers model.

3-4-5

3-4-5 聴覚における外有毛細胞の運動と 内毛細胞励起のモデル

A model of inner hair cell excitation based on interlocking of outer hair cell motility in the hearing

◎石渡信吾(横浜国大), △東原昂輝(横浜国大理工)

ヒトの内耳にある蝸牛は音波の検出と周波数分解・強度分解を行っている。蝸牛を長さ方向に仕切る基底膜の上にコルチ器と呼ばれる器官がある。コルチ器の上部には内外 2 種類の有毛細胞があり, 外有毛細胞の毛束はコルチ器を上から覆う蓋膜に刺さっている。音波の圧力によって基底膜が持ち上がると, コルチ器と外膜の間隔が縮み, 外有毛細胞の毛束は外側に傾く。これがスイッチとなって, 外有毛細胞は収縮する。従来の解釈は, この収縮が基底膜振動を増幅させて, 内毛細胞がこの振動を検出し, 蝸牛神経を発火させる, というものであった。しかし内毛細胞の励起機構には不明な点が多い。我々は, 蓋膜の下側を内毛細胞より少し内側に走るヘンゼン線条に着目した。最近, 外有毛細胞の毛束が能動的に動くことが報告されたが, この毛束の運動が蓋膜を外側に引っ張り, ヘンゼン線条が内毛細胞の毛束を直接外側に押して励起する機構を提案した。

我々のモデルをマクロな模型で検討した結果, ヘンゼン線条が内毛細胞を励起するには, 外有毛細胞の毛束の変位量では不十分であると結論した。そこで新たに外有毛細胞の湾曲運動を取り入れてモデルの修正を図った。今回はこの湾曲運動を実現する方法について議論する。

3-4-6

3-4-6 数値解析を用いた耳小骨再建モデルの伝音特性に関する検討

Numerical analysis on the sound transmission via the reconstructed ossicular model
○朝倉 巧, 伊藤 瑠哉 (東京理科大).

倉科 佑太 (東工大), 平林 源希, 栗原 渉 (東京慈恵医大)

- ◆ 耳小骨再建手術は、真珠腫性中耳炎、耳小骨離断などの気骨導差を生じる疾患に対し、伝音系を再建することで患者の聴力の回復を図る術式である。しかしながら、鼓室形成術の中には術後聴力成績が芳しくない症例が存在する。また既往の研究では健康なヒト中耳の伝音特性に関する報告が多いが、耳小骨再建手術によって再建された中耳の伝音特性に関する研究は小池らによる報告はあるものの、多くは見られず、手術成功のための知見を蓄積する必要がある。本検討では、耳小骨形成術において用いられる各種術型の伝音特性を調べるための基礎的検討として、有限要素法解析を用いた健全なヒト中耳の伝音特性に関する予測検討を行った。
- ◆ 有限要素法を用いた振動音響連成解析を行うことにより、健全な中耳の伝音特性を求め、既往の研究結果との比較により、解析モデルの妥当性を検証した。

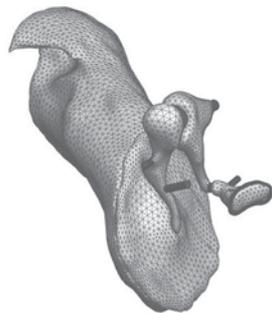


Fig. 1: 3D FEM models of a human intact middle ear.

3-10-1

3-10-1 巨大メガホンを用いた音声の長距離伝搬と集音に関する実験

Experiments of long distance sound propagation and sound collection using huge megaphone.

○高橋 義典 (工学院大), 小谷 信公 (日テレアクセス),
伊東 歌寿彦 (伊東NC工業), 土屋 敬 (木村製造所)

- * 博物館のポピュラーな屋外展示としてパラボラ集音器が挙げられる。パラボラはアンテナとしては一般的であるものの、音響学分野では、ホーンの方が日常生活でよく目にするデバイスである。
- * しかし、ホーンの原理はパラボラよりも複雑であり、原理を体験的に学べる博物館展示の登場が望まれる。巨大なホーンであれば、パラボラ集音器と同様に博物館の人気アトラクションとなるのが期待できる。
- * 本報告では、2020年12月24日に放送された、読売テレビ「日テレアクセス制作の番組「クイズ距離感カジノ」の「最強のメガホンを使って人が会話できる限界の距離」を実証する企画として行われた実験を紹介する。



Fig. 1 huge megaphone

- * 遮断周波数100Hzのハイパボリックホーン (Fig. 1) を使って使用したところ、電気を使わずに1.1kmでも会話できることが確認された。
- * 計算上、3.5km程度まで拡声可能であり (Fig. 2)、同様の実験には静かで互いに見通せる実験場所の確保が必要である。

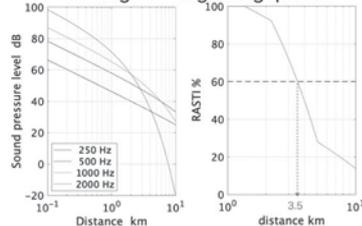


Fig. 2 Theoretical value of the distance which speech can be transmitted using the huge megaphone. (90dB at the mouth)

3-4-7

3-4-7 Cartilage conduction applied on a smartphone – Measurements of ear-canal sound pressure, HATS, and acceleration

© Xiuyuan Qin (Chiba Univ.), Sho Otsuka (Chiba Univ.), Junghun Noh (Samsung Display), and Seiji Nakagawa (Chiba Univ.)

- ◆ Cartilage conduction (CC) has been proposed and applied to various applications. Recently applications of CC on mobile phone have been developed.
- ◆ In this study, ear-canal sound pressures (ECSPs) and acceleration produced at the pinna in CC transmission were measured with different conditions by using a smartphone panel.
- ◆ The measurements were done when the panel contacted or did not contact the pinna as well as the panel covered the pinna fully and partly to estimate the frequency characteristics of CC and AC.
- ◆ A comparison between ECSP and HATS showed the difference between CC and AC, and the influence of vibration at the pinna on CC.

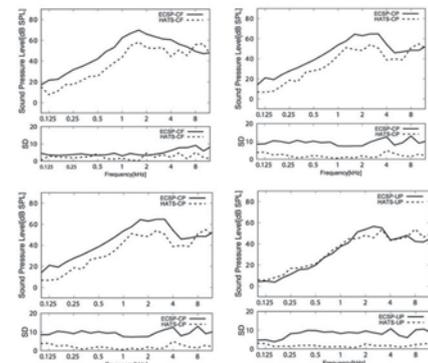


Fig. 1. The results of ECSP and HATS measurements in four conditions. CF: Contacted, fully-covered condition; CP: Contacted, partly-covered condition; UF: Uncontacted, fully-covered condition; UP: Uncontacted, partly-covered condition.

3-10-2

3-10-2 言語聴覚士と音響学の関係を考える

The relation between acoustics and Speech-Language-Hearing-Therapist

○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)

- ◆ ことばのリハビリを行う言語聴覚士の養成校では、音響学 (聴覚心理学も含む) の授業が必修である。決められた授業時間内で臨床に役立つような音響学を教えるのには言語聴覚士がなぜ音響学が必要であるかについて考えなくては行けない。
- ◆ 本発表では、現役養成校の言語聴覚士と音響学教員が、音響学・聴覚心理学で、現状はどのような知識を学んでいるのか、今後どのようにすべきかについて意見を出し合い、今後予定している全国の言語聴覚士養成校の教員対象のアンケート作成のための予備調査を行った。その内容について報告する。
- ◆ 言語聴覚士のための音響学・聴覚心理学を担当するにあたって、教師は、以下のことを考えることが重要である。
 - (1) 一般教養ではない、臨床に関する専門科目を学ぶために特化した基礎的知識であることを意識すること、
 - (2) 現状では、音声障害など、一部の分野、機関において、高価な機材を使った「音響分析」や、研究としての実験しか行われていないが、音響分析の結果を応用した訓練や、AACの作成など、どんな機関に就職しても気軽に実現可能な例を紹介し、敷居を下げていくこと。
- ◆ 本発表は、科研費番号 20K03074 と 18K02988 の成果である。

3-10-3

3-10-3

聴覚検査を学ぶための音響学

Acoustics for learning the Hearing tests

○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)

- ◆言語聴覚士養成校の基礎科目である音響学や聴覚心理学は、その後に学ぶ様々な専門科目と深く結びついている。
- ◆それゆえ、単独で学ぶよりも、両者の関係を考えながら学習することによって、学習の目的がはっきりし、より理解が深まると考えられる。
- ◆関係ある項目の一つに聴覚検査がある。本発表では、養成校の現役言語聴覚士教員と音響学教員がこれらの関係についての予備調査として意見を出し合った結果をまとめて報告する。
- ◆音響学教員に、聴覚検査の内容については、あまり知られていない。また、独学で聴覚検査を学習することはとても大変である。
- ◆それゆえ、音響学、聴覚心理学との「つながり」を示す指針やそれを示す教材が必要である。
- ◆学生にとって苦手な音響学・聴覚心理学であるが、この「つながり」は、学習の目的となる。
- ◆本発表では、言語聴覚士養成校で、音響学、聴覚心理学を学ぶにあたり、専門科目である聴覚検査実習とのつながりを考えた。
- ◆今後は、この結果をもとに、全国の言語聴覚士専門学校の音響学・聴覚心理学の教員、言語聴覚士教員に対するアンケートを行い、聴覚検査を学ぶための音響学・聴覚心理学の授業を支える教材作りを目指したい。

◆本発表は、科研費番号 20K03074・18K02988 の成果である。

3-10-5

3-10-5 音に関する科学リテラシー調査票の設計
—大学生を対象とした予備的調査—

Developing a questionnaire to investigate scientific literacy about the sound – preliminary survey of university students–

○網野加苗(科警研), 程島奈緒(東海大)

- ◆科学リテラシー (SL) とは？
 - 生きるために必要な科学的な知識
 - 社会生活においてそれを活用する能力
- ◆音に関するSLを調べるための質問を含む調査票を作成した
 - <4段階評価の質問>
 - オーディオ機器の音質の違いには敏感な方だ
 - 生活騒音を軽減するために床にマットやカーペットを敷いている
 - イヤホンで音や音楽を聴くときは、大音量にならないように注意している
 - <正誤を答える質問>
 - 話し声は、話し手の顔が向いている方向にのみ伝わる
 - 打ち上げ花火は音が聞こえた後に花火が見える
 - 救急車が通った後に、サイレンの音が低くなったように聞こえるのは、音の振幅が減少するからである
 - クラシック音楽を聞かせて熟成させたワインは、味がまるやかになる (ほか3問)
- ◆大学生(3年生以上)、大学院生16名から回答を得た
 - 日常生活において音に配慮している人が半数以上いた
 - 音に関する知識は分野によって正解率が大きく異なっていた
 - 音に関する疑似科学を「正しい」とする回答も見られた

3-10-4

3-10-4

BGMが学校教育環境に及ぼす効果
—中学校の現場調査を例にして—

Effects of background music on the school educational environment
-Taking the example of a field study at a junior high school-

○小松正史(京都精華大学)

- ◆教育機関(中学校)の環境改善を目的とした背景音楽(BGM)の活用のある方を模索するための、現場調査を報告する。
- ◆福岡県にあるT中学校を対象に、「音楽なし」と「音楽あり」の2条件で、校内の音環境の印象評点を実施した。
- ◆授業時間以外の12の時間帯でUSEN放送による特別プログラムを設定し、40~45dB程度の音量で再生した。
- ◆BGMを導入することで、動的な心理状態を促進させ、教育活動の活性化が生まれることが分かった。

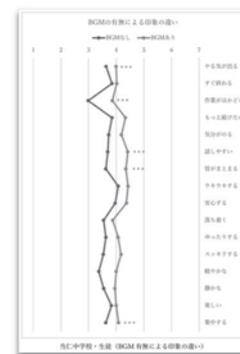


Fig. Profile diagram with 2 conditions

3-10-6

3-10-6 音によるEMS(Examination Monitoring System)
開発の試み

An attempt to develop EMS (Examination Monitoring System) by sound.

○上田麻理, 田中哲雄(神奈川工科大), 長谷川英之(富山大)

昨今のコロナ禍のオンライン授業で最も問題となる課題の一つが試験である。対面式の試験以上に、カンニング、複数名での受験や替え玉受験等が世界中で問題となっており、その不正防止策として、様々な試験監視システムの開発・導入が試みられている。これに伴い米国では、webカメラ等の画像による監視ツールの使用は学生のプライバシーの侵害であり、しばしば公民権の侵害でもあるとの議論が勃発している。プライバシー及び、ネットワークの速度に配慮すべく、画像(録画)を使用せずに試験監視ができないか?に対応すべく、音による試験監視ができないか?をリサーチクエストとして、音によるEMS(Examination Monitoring System: Medusa)の開発とその実践を試みた。

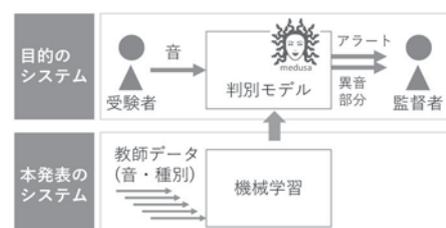


Figure: Application overview

3-10-7

3-10-7 この音何ヘルツ？
 -コロナ禍におけるオンライン授業用教材作成の試み-
 What is this sound in Hz?
 - An attempt to create teaching materials for online class in CoronaV-

○上田麻理, 田中哲雄(神奈川工科大)

コロナ禍において今後対面式授業が推奨される場合であってもハイブリッド形式の授業が求められる。そこで、引き続きオンライン授業教材の充実化を目的として、音の高さに関するwebアプリケーションを開発したので紹介する。



Fig.1 What is this sound in Hz? Fig.2 Which sound is this frequency?



Fig.3 What hertz sound do you hate? Fig.4 What is this range in Hz?

3-10-9

3-10-9 音声に関わる音響学のオンライン授業のために
 For online courses on speech-related acoustics
 ○荒井隆行(上智大・理工)

- ◆近年、電子教材などオンライン上で教材が公開されたり、動画が配信されるなど、教育におけるICT活用は高まってきている。
- ◆2020年においては新型コロナウイルス(COVID-19)感染拡大に伴い、多くの講義がオンラインで実施されることとなり、教育におけるオンライン化は結果的に加速した。
- ◆音響学に関わる分野も例外ではなく、音に関わる実験などもオンラインでいかに見せるかなど、様々な工夫が行われた。



- ◆オンライン教材であるAcoustic-Phonetics Demonstrations (APD)が見直され、同名のYouTubeチャンネルの動画も活用された。
- ◆そこに、短い動画を含むコンテンツが次々と追加された。
- ◆オンラインで工作を可能にするため、参加者が身近な素材で作ることができる声道模型が見直され、実際にオンライン授業で活用された。

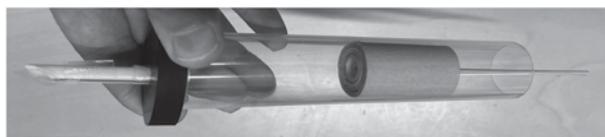


Fig. 1: Simple vocal-tract model with sound source.

3-10-8

3-10-8 オンライン教育でのシミュレータ教材の活用事例

A Case Study of Simulator Teaching Materials for Online Education
 ○須田 宇宙(千葉工大・情報科学部)

- ◆音響工学の教育分野では、不可視現象を扱うことが多い。
- ◆筆者らは、パラメータの変更に即座に対応して可視化・可聴化するシミュレータ教材を中心としたマルチメディア教材を開発し、教育に利用することによって教育効果を高めてきた。
- ◆コロナ禍での対応として、大学ではオンライン授業を実施せざるを得ない状況に陥り、多くの授業がオンデマンドで実施された。
- ◆しかし、映像メディアによる学習では、対面式の授業と比較して学習者の反応を得づらいため、様々な場面を想定した長時間の講義になりがちなことや、学習者の理解度を確認するために課題の量を増やさざるを得ず、学生が時間に追われることとなり不満の声が上がった。
- ◆この問題を解決するために、筆者はオンライン授業にシミュレータ教材を用いた簡易的な実験を取り込む形式の授業を実施した。
- ◆本稿では、この取組に使用した教材の概要と学習者の反応を報告することを目的としている。

3-10-10

3-10-10 (招待講演) 真耳 Online を活用したリモート聴能形成の取り組み

The trial for remote technical listening training

○森尾謙一, △藤原賢(日本音響エンジニアリング)

- ◆聴能形成の特色として「集団訓練」がある。しかし、コロナ禍において、人が集まって音を聴く「クラス訓練」(集団訓練)は「密を避ける」という観点から聴能形成の実施を困難にしている。
- ◆筆者らは聴能形成に特化したオンラインサービス真耳 Online の開発に携わっているが、コロナ禍における訓練実施の困難さに直面した。
- ◆コロナ禍における音響教育実施の取り組みとして、真耳 Online を活用した、リモート聴能形成について複数の方法で検討を行った。
- ◆それら検討結果について、弊社新人研修事例におけるリアルタイム音声伝送による方法、および神奈川工科大学でのオンデマンド方式での訓練事例を紹介する。
- ◆併せてリモート環境特有の課題についても併せて報告する。



Fig.1: Shiji Online System Configuration

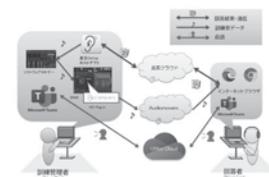


Fig.2: Remote Training System Configuration

3-10-11

3-10-11 (招待講演)遠隔環境における MATLAB を活用した音響教育

Education in acoustics using MATLAB for distance learning
 ○村松未輝雄(MathWorks Japan)

- ◆従来より、MATLAB®は音響教育において、理論計算やアルゴリズムの可視化、数値解析、シミュレーションなどで活用されてきたが、特に近年では、ハードウェアとの連携や、クラウドソリューションも充実させている。
- ◆本講演では、コロナ禍で明らかになった遠隔授業における課題に関し、特に遠隔環境で学生に興味を持ってもらう音響教育を行うために、オンライン教育をサポートするクラウド環境、自宅での音響実験、インタラクティブに使える教材など、MATLAB を活用することで対応可能な取り組みを、授業事例を交えて紹介する。



Fig: Example of interactive textbook using Live Editor

3-11-1

3-11-1 野球における歓声等の環境音が選手に与える影響

Effects of environmental sounds such as cheers in baseball on players
 ○春日秀雄, 上田麻理, 三春知史, 依田裕雅, 二村亮平, 田中哲雄,
 △磯川昇汰, △高橋勝美, △新田晃司(神奈川工科大)

- ◆新型コロナウイルス流行の影響で、様々なスポーツで無観客試合が実施された。しかし、歓声のない試合ではモチベーションを上げられないという選手も存在する。そこで本研究では、**歓声等の環境音の有無が選手の動きに与える影響について調査している。**
- ◆神奈川工科大学のKAIT スタジアムに歓声等の環境音を発生させる音響AR環境を構築し、硬式野球部の練習試合で音響の影響を比較する実験を行った。この音響ARを用いた時に選手にどのような影響が出るか、アンケートを用いた心理評価により測定した。
- ◆本研究で用いた音響ARでは、スタジアムに設置された11台のスピーカーを使って「ざわめき」、「拍手」、「歓声(得点)」、「歓声(得点が入らなかった)」の4種類の環境音を再生することができる。
- ◆音響ARを用いた試合に参加した選手へのアンケート結果から、「**歓声があったほうがモチベーションを高めることができ、試合に集中しやすい**」という結果が得られた。
- ◆一方で、今回構築した音響ARの臨場感に関しては、「**臨場感がない**」、「**タイミングに違和感がある**」ということが明らかになり、**音響ARの臨場感の改善が課題**であることが判明した。



3-10-12

3-10-12 オンライン会議システムの音質とその客観評価

Objective measurement on the audio quality of the video conference system
 ○西村 明(東京情報大)

2020年4月より、Covid-19感染拡大防止のため、多く大学においてビデオ会議システム(VCS)を用いたリアルタイム・オンライン講義が開講された。音響関連授業でも、音のデモをVCS経由で行わざるをえない場合がある。本稿は、代表的なVCS(Zoom, Webex Meetings, Microsoft Teams)における、音声および音楽の伝送時の客観品質と、送信設定との関係を、PESQおよびPEAQを用いて測定した結果を報告する。

図は、PESQ(ITU-T P. 862. 2)により測定した、音声信号の伝送路のMOS-LQO(客観聴取品質スコア)値である。PESQはAACなどの新しいコーデックを対象としていないため、VCS間の品質比較はできないが、送信設定により音質が変化する場合があることが分かる。

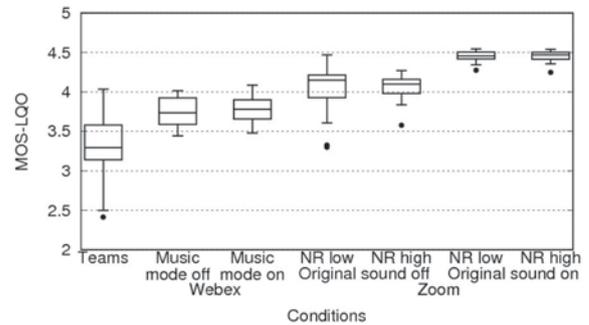


図 ITU-T P. 50 付録英語および日本語 32 発話ファイルの MOS-LQO

3-11-2

3-11-2 球技の聴覚情報利用に関する基礎的検討その2

-試合時の観客有無による競技者・観戦者の音による臨場感
 Basic consideration on auditory information in ball games
 A feeling of presence by a player or spectator depending on whether they are spectators

☆二村亮平, 依田裕雅, 磯川昇汰(神奈川工科大), 三浦貴大(産総研), 中村健次郎(東工大),
 春日秀雄, 新田晃司, 高橋勝美, 田中哲雄, 上田麻理(神奈川工科大)

2020年現在、COVID-19の流行に伴い、中継でしかスポーツ公式試合を観戦することができない状態がほとんどである。例えば、プロスポーツの競技では無観客試合で開催することが増えている。そんな中、6月のプロ野球開幕を盛り上げるため、福岡ソフトバンクホークス応援中するために「Pepper」と「Spot」が応援に駆けつけた。



そこで、ARでの声援と実際の声援の臨場感印象の違いやタイミングのズレによる違和感と、相対的に観客の有無による競技者視点と観客者視点での音による臨場感の違いを明らかにするため、野球を対象とした無観客試合に関するWebアンケート調査を行った。

3-11-3

3-11-3 スマートグラスを用いた 競技場の音の認識と可視化 Stadium Sound Recognition and Visualization Using Smart Glass

☆小林雅彦, △村手涼雅, 西田昌史, △綱川隆司, 西村雅史 (静岡大学)

◆聴覚障がい者のスポーツ観戦を支援するため、実時間で競技場の音を認識しスマートグラス上で可視化するシステムを提案した。また、CatBoostを用いた競技場の音の認識手法を提案した。

◆提案システムはスマートグラスで収集された音をサーバーに送信し、環境音の認識を行う。その後認識結果に応じてスマートグラス上で可視化を行う。

◆環境音の認識精度及び認識速度を改善するため、SVM、XGBoost、LightGBM、CatBoostの4つの機械学習手法に対して比較実験を行った。

◆競技はバスケットボールを対象とし、ブザー、応援歌や歓声、ホイッスルの音、その他の4イベントを認識対象として実験を行った。実験結果から、F値が0.81、認識時間が4.0msでCatBoostが最も優れているという結果が得られた。

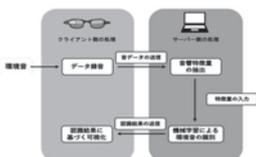


Fig.1: Flow of the sound visualization system



Fig.2: Example of the sound visualization system

Table.1: Result of the experiment

	SVM	XGBoost	LightGBM	CatBoost
F値	0.75	0.76	0.76	0.81
認識時間[ms]	203	22.6	24.5	4.0

3-11-5

スポーツ音響への応用に適したアンサンブル平均を用いた吸音特性の現場測定法 - 体育館における現場測定 -

An in-situ measurement method using ensemble averaging technique for sound absorption of materials installed in athletic fields - Applications of the EA method to materials in realistic gymnasiums -
○大鶴 徹、富来礼次、岡本則子、ラワンワデークル シワット(大分大)

◆アンサンブル平均を利用する吸音特性測定法を2箇所の屋内体育館で試行的に適用した概要を報告し、スポーツ施設の音響に資する現場測定手法の可能性と課題を議論した。

◆測定風景を Fig.1 に示す。

◆Gym-OU における表面インピーダンスと吸音率の現場測定結果は残響室における値と概ね近似した。

◆2箇所の施設の木質床面の吸音率に多少の相違が認められた: 床面の硬さが影響したものと推察している。



Fig.1: An EA method measurement conducted in-situ at the main arena of Gym-OU.

3-11-4

3-11-4 画像解析からみる運動時の音楽の効果 -OpenPose を使用した屋外でのウォーキングの動作解析の試み-

Effect of music during exercise from image analysis
-An attempt to analyze the behavior of outdoor walking using OpenPose -,
☆三春知史, 春日秀雄, △高嶋渉, △高橋勝美, 上田麻理(神奈川工科大)

◆前報では、OpenPose という姿勢推定手法を用いて屋内でのトレッドミルを使用した歩行動作の解析をすることで、運動時における音楽聴取が及ぼす効果を調べた。その結果、通常時と音楽聴取時の歩行動作の違い及び OpenPose の有用性を確認することができた。

◆前方の研究ではトレッドミルの歩行動作の解析であったため動作解析の対象者の移動量が少なかったが、フリーフィールドによる歩行など対象者の移動量が多い運動において OpenPose による姿勢推定の精度が落ちることが予測される。

◆本研究では、対象者の移動量が多い運動時においても音楽聴取の効果を OpenPose により解析することができるかを明らかにするために、屋外のウォーキングを動作解析の対象として調べた。

◆2種類の音楽聴取歩行、2種類のメトロノーム音聴取歩行、通常歩行の合計5条件の歩行動作を OpenPose により動作解析を行った。

◆音楽やメトロノーム音の聴取により1歩あたりの歩行速度が速くなることが確認できた。

3-11-6

3-11-6 OpenPose による野球の動作解析と音響解析

OpenPose-based motion analysis and acoustical evaluation of baseball batters

☆依田裕雅, △磯川昇汰, △新田晃司(神奈川工科大), 三浦貴大(産総研)
中村健太郎(東工大), 春日秀雄, 田中哲雄, 上田麻理(神奈川工科大)

◆野球において選手の掛け声や観客の歓声以外に、打球音や捕球音が発生する。その様な音は試合中の野球選手達にとって大事な判断材料とされている。本研究では、打撃力による違いに打撃音の音響的特徴について解析を行う、その結果を元に音と動作の関係性が、**野球の競技力向上や指導における示唆**に音を利用することが可能か検討することを目的とした。

◆音響解析は1/1 オクターブバンド解析を行った後、8つの中心周波数の周辺帯域のレベル差において、打球方向、打球の種類、打ち方の他、指導者視点での評価での相違点について、分散分析で求めた。

◆動作解析は OpenPose を利用し、25 点の Key Points の中から首と足の付け根の2点を繋いだ直線の傾きを求め、バッターの体の軸の傾きとした。構えからインパクトまでの軸の傾きの時間変化をグラフ化し、バッティングの特徴について調査した。

◆OpenPose を利用したバッティング動作時の簡易評価を試み、指導者視点で評価が高い・低い打者で、**体の軸の傾き変化に差がある可能性と 250 Hz 以下の周波数帯域**で打球方向・打球の種類・打ち方・指導者視点での評価に関する違いが生じる可能性があることが確認できた。

◆十分なデータが採れたとは言い難いため、今後もデータの収集をする必要がある。

3-11-7

3-11-7 画像解析からみる運動時の音楽の効果 —テンポの違いが歩行リズムに与える影響— Effect of music during exercise from image analysis —The effect of different tempos on walking rhythm—

☆三春知史, 春日秀雄, △高嶋渉, △高橋勝美, 上田麻理(神奈川工科大)

- ◆ウォーキングやランニング, サイクリングなどの律動的運動時における, 音楽リズムと運動リズム関連性を調査する先行研究から, 運動リズムと音楽リズムの高い運動同調発生率が確認されている。
- ◆先行研究によると自分でペースの調節がしやすいサイクリング運動において運動同調発生率が確認されていた。しかし, 歩行速度が規定されているトレッドミルの歩行において音楽リズムの違いが歩行動作に及ぼす影響や, どのようなテンポの音楽が安定した動作をもたらすのかはまだ明らかになっていない。
- ◆本研究では速度が規定されているトレッドミルを使用した歩行動作において音楽リズムが歩行動作に与える影響を明らかにするために, OpenPoseによる姿勢推定手法を用いて調べてみた。
- ◆聴取する音楽のテンポの違いによって歩行動作に違いが出ていることを確認できた。歩行動作に影響をもたらす条件は人によって差があることが確認できた。

3-11-9

3-11-9 (招待講演)バレーボール競技ににおける セッターが収集する必要な情報としての音 Sound as indispensable information for volleyball setters

○ヨヨーコ セッターランド(日本女子体育大学)

6人制バレーボールを本格的に始めたのは中学校1年生(12歳)、バレーボールの強豪校で、すでに全国大会出場を果たしていた学校の部活動で競技生活がスタートした。

バレーボール未経験者であったが、セッターのポジションを希望してかかったのであるが、技術用語はおろか、ルールさえよく理解していない状況であった。当然のことながら技術が未熟だったこともあり、いつ、どこで、どのように集中してプレーをしてよいのかもわからず、試合中には周囲のすべての音が聞こえてきてしまい、緊張が頂点に達して混乱することも多々あった。経験を積み重ねるうちに技術も向上し、試合も場数を踏んだことにより、過度に緊張することも少なくなった。また、様々な試合会場で試合を体験したことにより、特に大きな会場での音の反響にも混乱することは無くなった。

一定のキャリアに達した後の試合では、自陣のコート上以外の「音」はインプレー中には小さくなる感覚に気付くようになった。ホーム、アウェーにかかわらず、周囲が大音量であればあるほど、自陣のコート上の「音」、すなわちチームメイトの指示の音だけをはっきりと拾うことができるようになっていくことに気づいた。

バレーボールは3次元の空間でボールを扱うため、特にセッターのポジションは味方が360度あらゆる方向から返球してくるボールに対応し、配球しなくてはならない。視野に入りきらないエリアは必ず存在するため、反復練習で習得した空間認知の感覚と、チームメイトが発する有音である「声」と無音の「呼吸」から配球したい空間を特定していく。

キャリアを積み上げていくほどに「感覚」は無駄なものをそぎ落として鋭敏になっていくように思う。プレーに必要な「音」を確実に収集していく感覚もそのうちのひとつである。

ボールを使用した反復練習や、フィジカルトレーニングからのアプローチによる競技力向上だけでなく「音のとらえ方」といったアプローチも今後、考えられるのではないだろうか。

3-11-8

3-11-8 (招待講演) 水泳競技と音

(その2 東京オリンピック・パラリンピックの開催が
議論される中で)

Water sports and sound -Par 2 In suspended situation of Tokyo
Olympic/Paralympic games -

○源 純夏(徳島ライフセービングクラブ)

(2020年春季研究発表会講演論文を改訂)

- ◆スポーツと音響、この2つを並べ考えたときに最初はなんの言葉も思いつかず、むしろ文系と理系のようなまったく違った本質で交わることのないイメージを思い浮かべてしまったことが自分の中の始まりであった。
- ◆競泳選手時代、泳いでいるときに聞こえていたモノが「水の音」だと考えることもなく過ごした時間が長かった影響も考えられる。
- ◆そんな小学生の頃からの競泳選手のキャリアと現在はライフセービングやSUP(Stand Up Paddleboard)などのマリンスポーツを実施する中で、「音」に関わる事柄を探し出し、考えていくことは私にとっても、スポーツ界にとっても新たな視点になることは間違いがないと感じている。
- ◆今年、東京オリンピックパラリンピックが開催され多くの感動や涙、そしてたくさんの物語が誕生するだろう。しかしこの2020年をピークにしてはならない。この年を契機にスポーツの価値そのものが見直され、多くの人々がスポーツに参加し、親しむためにもこの分野の発展は大いに役立つだろう。(2020年春季原稿をそのまま記載)

3-1P-1

3-1P-1 船舶航行音を対象とした窓スピーカ ANCシステムによる室内環境の評価*

* Evaluation of room noise by window speaker ANC system for boat noise.

☆深津遠貴, 武藤憲司(芝浦工大)

- ある運河では小型船舶が朝から夕方にかけて多く航行し、運河沿いの住宅に騒音影響をもたらしている。
- 本研究の目的は、窓振動を用いた室内静音化のためのマルチ適応騒音制御システムの構築である。本報告では、窓スピーカを用いた単一周波数音の制御時の室内空間の音圧測定位置の違いによる騒音低減を報告する。
- 音圧測定距離の違いによる音圧低減の分布をみると、音圧低減の大きい箇所はほぼ一致することがみられる。この結果から提案した窓振動マルチANCシステムで室内空間の音圧を低減させることが可能であることがわかった。

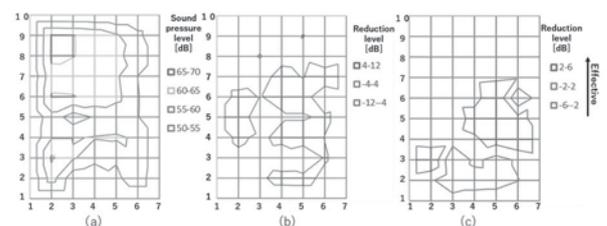


Fig. 1. Map of sound pressure level and reduction level by 120 Hz (Vibration pickup's position:(X),(X'), Actuator's position:(Y),(Z),

(a): Sound pressure level map, (b): Reduction level map (50 mm), (c): Reduction level map (300 mm)

3-1P-2

3-1P-2 低域通過フィルタによる能動正弦波騒音制御システムの有効性検証

Confirmation on effectiveness of active sinusoidal noise control system using low pass filters

○藤井健作(コダウエイ研), △棟安実治(関西大), 菅木禎史(千葉工大)

本稿では、送出正弦波の振幅調整を低域通過フィルタで行う騒音制御法の有効性検証結果を報告する。この方法では到来正弦波と参照正弦波、参照余弦波との積を印加した低域通過フィルタの出力を使って振幅と位相の調整が行われる。その調整にFiltered-x LMS法が使用されないことから、ステップサイズの選定に2次系の特性や到来正弦波と周囲騒音のパワー比を考慮する必要はない。Fig.1は本法によって低減された正弦波の波形を示している。但し、この波形では重畳している周囲騒音は省略され、初めの1秒間に本法は適用されていない。この波形において到来正弦波が50 dB低減されていることから、試聴しても正弦波の残留は知覚されず、算定誤差による白色雑音的な音が聞こえるだけである。

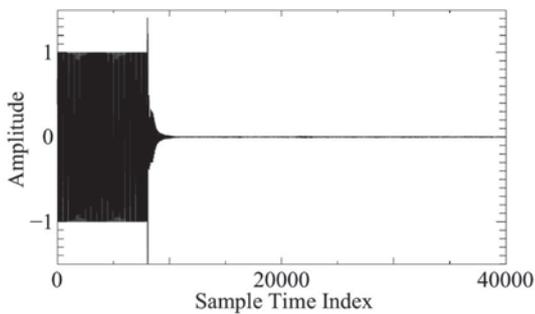


Fig.1: Sinusoidal wave reduced by proposed method (SN20dB)

3-1P-4

3-1P-4 カートシス変化抑制を考慮したスパース性に基づく音楽音響信号の雑音除去

Noise reduction of music signal based on sparseness remaining kurtosis unchanged

☆中野佑里, △伊藤和寿(芝浦工大)

- ◆音声強調の研究で非線形信号処理前後のミュージカルノイズの発生量を非音声区間におけるカートシスの変化量により定量化する方法が提案されている。
- ◆本研究では、ミュージカルノイズの発生を低減するために楽音信号のスパース性に基づくノイズ除去信号処理において、カートシスの変化を抑制する正則化項を追加することを提案し、その効果検証を行った。
- ◆提案項を追加することで、スペクトログラムにおいてミュージカルノイズの成分となる要素が目立たなくなることが確認された。

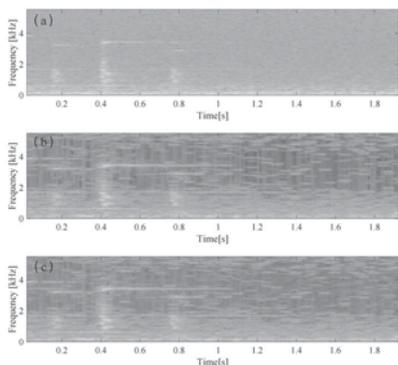


Fig.1:(a)Spectrogram of input signal, (b)Spectrogram of processed signal w/o kurtosis term, (c)Spectrogram of processed signal w/ kurtosis term

3-1P-3

3-1P-3 パワースペクトルを用いたGANによる光レーザマイクロホンのための音声強調

Speech enhancement based on power spectrum with generative adversarial networks for optical laser microphone

☆CAI Chengkai(立命館大院), 岩居健太, 西浦敬信, 山下洋一(立命館大)

光レーザマイクロホンは、レーザ光を音源近傍の物体に照射し、音により発生した振動を計測することで音を受音するシステムである。レーザ光は強い直進性を持つため遠隔収録が可能であり、マイクロホン周囲の雑音に頑健である。しかし、被照射物体の形状及び振動特性により、雑音の混入や音声成分の欠落が発生し、音質が劣化する。従来、深層学習に基づいた様々な音声強調手法が提案されている。これらの手法は多種類の歪みを含む光レーザマイクロホンの観測音声に直接適用した際に十分な復元精度が得られない。本稿では、光レーザマイクロホンのための音声強調手法を提案する(Fig.1)。まず、観測音声のパワースペクトルをスペクトル減算法及び第1, 2フォルマント強調により処理し、その後GANを用いて音声波形を生成する。音声強調実験の結果(Fig. 2)より、提案法の有効性が示された。

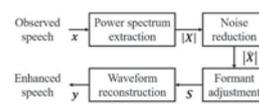


Fig. 1 Block diagram of the proposed method

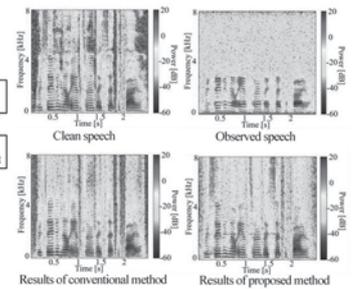


Fig. 2 Enhanced speech with each method.

3-1P-5

3-1P-5 ヴァーチャル観測信号を用いたニューラルネットワークの指向特性先鋭化

Sharpening directivity of neural network beamformers using virtual observation signals

☆金清祐紀(徳山高専), 水町光徳(九工大), 宮崎亮一(徳山高専)

- ◆ニューラルネットワークビームフォーマーでは、音声などの低周波数帯域の信号において先鋭な指向特性を得ることが難しい。
- ◆また、マイクロホンの数がそのまま入力ノード数にあたるため、ネットワークの柔軟性を増すためには多くのマイクロホンが必要となる。
- ◆本稿では、マイクロホンアレイによる観測信号を疑似的に多素子化したヴァーチャル観測信号を導入したニューラルネットワークビームフォーマーを提案する。
- ◆実験では、ヴァーチャル観測信号の導入個数や配置の異なる4通りの学習データを用意し(Fig.1, Table 1)、低周波数帯域において指向特性の大幅な先鋭化を確認した(Fig.2)。

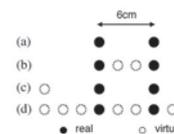


Fig.1:Arrangement of virtual microphone.

Table 1:Arrangement of virtual signals.

	(a)	(b)	(c)	(d)
Interpolation	None	2	None	2
Extrapolation	None	None	2	6
Number of mic.	2	4	4	10
Mic. interval [cm]	6	2	6	2
Array length [cm]	6	6	18	18

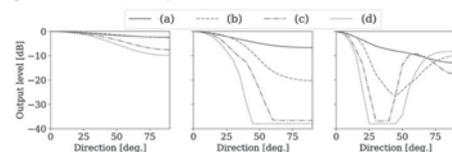


Fig.2:Directivity of each virtual microphone array. (left) 700 Hz, (center) 1.5 kHz, (right) 2.7 kHz

3-1P-6

3-1P-6 空間フィルタ出力を補助情報として用いた音源の移動に頑健なニューラル音声強調

Robust Neural Speech Enhancement against Sound Source Movement using Spatial Filtering Output as Auxiliary Information

☆西城耕平(早大), 藤枝大, 片桐一浩(沖電気), 小林哲則, 小川哲司(早大)

◆提案手法

➢ マイクロホン正面の扇形領域を強調するマスクを DNN (Fig.1) で推定.

◆学習・評価方法

➢ 学習: 目的音源をマイクロホンアレイの真正面に固定したデータ.
 ➢ 評価: 目的音源を真正面 (Matched), 真正面からずらした位置 (Mismatched) に配置したときの強調音声の評価し, 提案法の音源移動に対する頑健性を調査.

◆音声強調実験・結果

➢ 既存のスペクトルサブトラクションによる扇形空間フィルタを上回る精度を達成.
 ➢ 提案法の音源位置移動に対する頑健性には改善の余地がある.

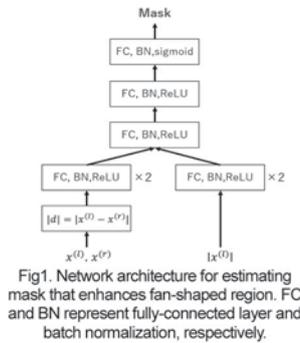


Table 1. SNR for microphone observation, SS-based fan-like BF, and proposed DNN-based BF under matched and mismatched conditions.

	SNR [dB]	
	Matched	Mismatched
Observation	0.0751	0.168
Fan-like BF	10.2	7.59
Prop.	15.1	8.96

3-1P-8

振動スピーカを用いた運動錯覚を誘発する振動刺激の検討

Study on the vibrational stimulation lead to an illusion of motion by using a vibration speaker

○平澤昭樹, 朝倉巧

◆強い刺激は知覚しやすく, 弱い刺激は知覚しにくい人間の知覚反応特性に基づいて, 運動錯覚を生起できることが検討されているが, その運動錯覚を誘発する振動刺激の物理量と運動感覚量の関係は未だ確立されていないため, 運動感覚量が物理量によってどのように変化するのか調査し, 運動錯覚を誘発する最適な振動刺激を検討した.

◆Fig.1 のようにのこぎり波 (周波数 75Hz) をフーリエ級数展開し, 次数によって立下り時間 (点線部) と立ち上がり時間 (実線部) を非対称的にすることで単位時間当たりの振幅変化量が大きい部分 (点線部) 小さい部分 (実線部) の存在する周期運動を振動刺激とした.

◆ME 法による主観評価実験を通して, 単位時間当たりの振幅変化量の違いによって運動感覚量がどのように変化するのか調査した.

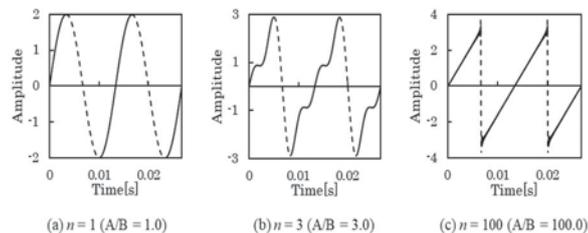


Fig.1 The vibration waveform changes caused by the degree n

3-1P-7

3-1P-7 アクティブ音響センシングにより得られる生体内音響伝搬特性と手指形状の関係性に関する実験的検討

Experimental study on the relationship between in-vivo acoustic propagation characteristics obtained by active acoustic sensing and finger shape

○清水涼平, 朝倉巧(東京理科大・理工)

◆ 手指姿勢の推定手法の一つに, アクティブ音響センシングを用いた周波数応答解析がある. この手法では, 手指形状毎の周波数応答よりエネルギースペクトル密度(ESD)を算出し, それを機械学習の特徴量として用いることで, 姿勢推定を行う. 本研究では, 手指姿勢の際の特徴量を増やすことを目的とし, 単純な手指形状と ESD の関係性について実験的検討を行った.

◆ 手首の機骨及び尺骨上に送波用のトランスデューサーと受信用のピエゾ素子を, 医療用両面テープを用いて取り付け, 3 kHz~6 kHz の線形チャープ信号を送波した. グー, チョキ, パーの三種類の手指形状について, 収録した応答を収録して, Swept-Sine 法を用いて実験系のインパルス応答を求め, Fig.1 に示すような ESD を得た.

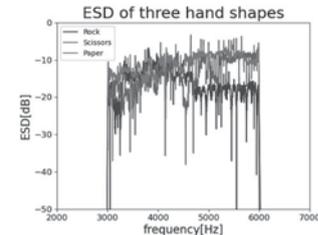


Fig. 1 ESD of Three hand shapes

3-1P-9

3-1P-9 耐騒音性デュアルマイクロホンの混合比による明瞭度と印象の検討

Study of articulation and impression according to mixing ratio of anti-noise dual microphone

☆松本飛竜, △郡雅治, △社本進也, 青木茂明, 島内末廣 (金沢工大)

◆気導マイクロホンと骨導マイクロホンを組み合わせた耐騒音性デュアルマイクロホンを用いて録音した音声の合成比率を変更したときの明瞭度および聴いた時の印象について検討した.

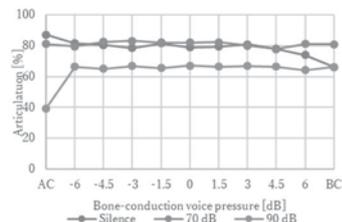
◆合成時に音圧バランスを変更することで気導音を強めにする, 骨導音を強めにするなどの 11 種類の混合比での単音節明瞭度試験を行った.

◆無騒音下では骨導音の比率が大きいと正答率が下がるような影響を受けたが, 騒音下では混合比による影響はほぼなかった.

◆単音節明瞭度試験の聴取者から「音によって聴き心地が違う気がした」との感想があったため, 混合比を変更した音声についてオピニオン評価法を用いて主観的な音質や印象について検討した.

◆無騒音下では気導音の割合が多い方がわずかに優位だった. 一方 90 dB の騒音レベル(A 特性)では骨導音の割合が多い方が優位だったが, 70 dB の騒音レベルでは混合比による印象への影響はほぼなかった.

◆混合比ごとの正答率と印象の傾向に差があることがわかった.



3-1P-10

3-1P-10 作動音識別における短パルス異音認識性能改善
 Performance improvement of short pulse abnormal sound recognition in operating sound identification

○伊藤秀浩, 高部晃好(アイシン精機),
 西浦敬信(立命大/名産研), 王浩南(立命大院)

- ◆自動車部品の製造現場では数値化し難い音色等の合否判定をベテランの官能検査員が検査している。そこで今回、自動車部品異音検査自動化の一助となる短パルス異音の識別技術を提案する。
- ◆作動時間にバラつきがあるため時間伸縮を行った自動車部品の作動音に対し、時間伸縮により鈍ったパルス成分を群遅延スペクトログラムを用いて強調する。Fig. 1
- ◆工場製品の検査では異常製品を流出させないことが目的となるため、異常品の検出率を100%とし、そのとき正常品を異常と判断してしまう過検出率を下げる必要がある。
- ◆提案法により過検出率を従来法の72.8%から2.8%と大幅に抑圧できることを確認できた。

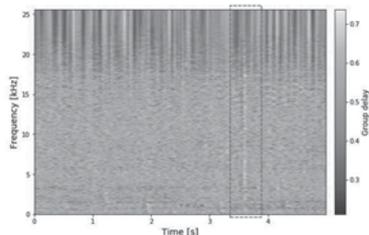


Fig. 1 Group delay with gaussian filter

3-1P-12

3-1P-12 音響説明文生成におけるキーワード推定サブタスクの効果
 Effect of keyword estimation sub-task for audio captioning

◎竹内大起, 大石康智, 小泉悠馬, △仁泉大輔, 原田登, 柏野邦夫 (NTT)

- ◆音響説明文生成は、与えられた音響信号からその説明文を生成する音と言語のクロスモーダルなタスクである。
 - これまでにマルチタスク学習とデータ拡張と後処理に注目したシステムを提案 (DCASE 2020 Challenge において1位を獲得)
 - 詳細な比較実験によって、データ拡張と後処理は効果的であるが、マルチタスク学習には効果がないことを発見
- ◆本研究では、キーワード推定サブタスクの学習によって音響説明文生成の精度が向上しなかった原因について明らかにする。

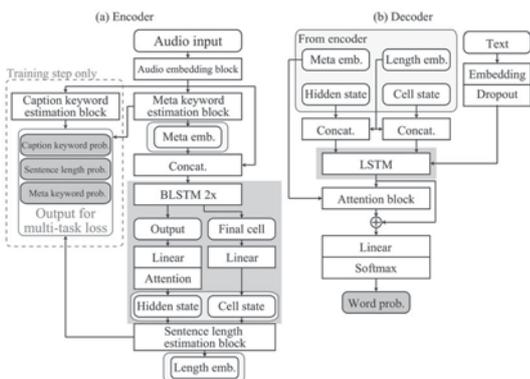


Fig. 1: DNN architecture of our model submitted to DCASE 2020 Challenge Task6

3-1P-11

3-1P-11 類似環境音識別における背景雑音が与える影響
 Effect of background noise on similar environmental sound identification

☆山之上大輝(静大・院・総合科技研), 竹内太法(静大創造科学技術大学院),
 立蔵洋介(静大・院・総合科技研)

- ◆目的：雑音環境下における類似環境音の識別精度向上
 - 類似環境音：音響的特徴が類似し、人間が聞き分けづらい音
- ◆手法：機械学習に基づく識別手法の前処理部に雑音抑圧手法を追加
 - 雑音抑圧手法：目的音を強調するためにWFにおける目的音推定と雑音の減算処理を反復するIWFを使用
 - 識別手法：類似環境音の細かな特徴を直接モデル化することが可能なGMMを識別に使用
- ◆実験：類似環境音(4種類)と環境音(4種類)の計8種類識別実験
- ◆結果：雑音抑圧前後で誤識別したクラスに違いがあることを確認。類似環境音と環境音で音の歪みや雑音の消し残りが識別率へ与える影響を確認

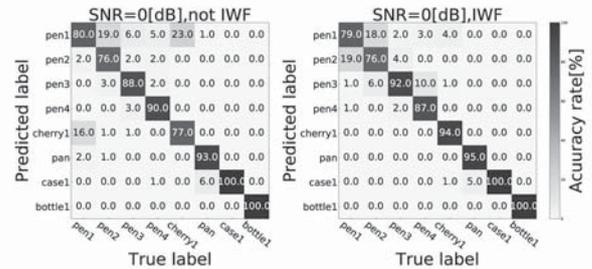


Fig. 1: Discrimination rate by class before noise suppression (left) and after noise suppression (right)

3-1P-13

3-1P-13 体導音と気導音を用いた作業行動認識に関する基礎的検討
 A study on human action recognition based on air and body-conducted sounds

○佐宗 晃, 蛇島 伸吾, 陳 暘, 児島 宏明, 長久保 晶彦(産総研),
 小木曾 里樹(岐阜高専)

- ◆人口急減と超高齢化による労働力人口の加速度的な減少により、製造現場などでは協働ロボットの導入による省人化、生産の効率化、熟練技術の継承などが喫緊の課題となっている。
- ◆作業者が視覚的に遮蔽された状況でも作業行動を認識可能とするために、作業行動に伴って発生し、空气中を回折しながら伝搬する気導音に基づいた音響イベントの認識方法を構築する。
- ◆指と物体との接触による振動が手の皮膚、組織、骨などを介して伝搬する体導音を、手首装着型の体導音センサーで計測することで、指先での接触状況や詳細な手作業の内容を反映する振動イベントの認識方法を構築する。
- ◆音響・振動イベントの種類と、それらの発生するタイミング・順序などのコンテキスト情報、更に音響イベントの発生方向などの空間情報を統合的に判断して最終的に作業行動を推論する。
- ◆本稿では、エッジデバイスが、少量データでも汎用性のあるモデルを学習できるような時系列処理モデルを提案し、それを用いた音響・振動イベント認識に関する予備実験結果を紹介する。

3-1P-14

音色属性に係る評価指標を用いた異常音検知の検討

A study on anomalous sound detection using timbral attribute

☆倉誠吾, 鶴木祐史(北陸先端大)

【背景】異常音検知の技術開発では、外れ値検知など識別器の検討に重点が置かれている。

【問題意識】高精度な異常音検知法を確立するには、識別器の検討だけでなく、特徴量について深く検討する必要がある。

【目的】音色に着目した異常音検知法を提案すること

【方法】

- 特徴量: 音色属性 (Timbral attribute) に関する指標
- 識別器: サポートベクターマシン
- 比較評価: DCASE2020 task2 の Baseline で用いられた対数メルスペクトログラムを特徴量としたもの

【結果】特定の音色属性の組み合わせを特徴量とした結果、対数メルスペクトログラムを特徴量としたものと比較し 16 機種中 6 機種 (37.5%の機種)で精度が向上した。

【結論】提案法の有効性が確認できた。

Table 1. Result of classification (id: 0)

Machine	Accuracy		FNR		FPR		F-measure	
	TA	LME	TA	LME	TA	LME	TA	LME
Fan	0.83	0.91	0.51	0.31	0.03	0.00	0.63	0.81
Pump	0.97	0.96	0.23	0.30	0.00	0.00	0.86	0.82
Slide rail	0.98	1.00	0.07	0.02	0.01	0.00	0.95	0.99
Valve	1.00	0.89	0.00	1.00	0.00	0.00	1.00	-

3-1P-16

音響イベント検知のための複数フレーム位相投票に基づく突発音検出法の検討

A study of impulsive timing detection method based on multi-frame phase voting for acoustic event detection

◎美島 咲子, 近藤 玲史 (NEC)

◆存在時間長の短い突発音イベントに着目し、複数フレーム位相投票に基づく突発音検出法について検討

◆従来法

- 突発音とそれ以外の非目的音との音圧比(Peak Signal to Noise Ratio, PSNR)が小さい場合に検出性能が低下

◆提案法

- 突発音の位相直線性に基づき分析フレーム毎に位相傾きを算出
- 各フレームの位相傾きを時刻に変換し、同一時刻に対応するスコアを重みづけ投票で統合

◆評価実験

- 実環境音と衝突音を加算して PSNR が 30, 20, 10 dB の音源を作成し、検知性能を比較
- 提案法は従来法と比較してフレーム F 値が平均で 0.14、最大で 0.20 向上
- 従来法で性能劣化が著しい環境においても、提案法は頑健な検知が可能

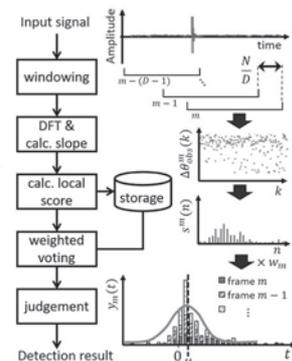


Fig.1: Overview of proposed method.

3-1P-15

3-1P-15 コンクリート橋の遠隔点検支援を目的とした打撃音による浮き検出に関する研究

Void defect detection by impact sound to support remote inspection of concrete bridges

☆藤井一貴, 宮崎亮一(徳山高専)

◆高度経済成長に伴い大量に建設が行われた橋梁は耐用年数を超え、鉄筋からコンクリートが乖離する「浮き」が発生

→重大事故の原因となり危険 定期打音検査の実施が義務付け

◆検査を実施する自治体の負担を減らすため、無人飛行機(UAV)を用いた遠隔打音検査支援システムが開発されている。

◆しかし、このシステムは人間が音を聴いて異常の判定を行うため、UAVのプロペラ音に打撃音が掻き消され打撃音が聞きづらく、技術者の熟練度により検出精度の幅があるという問題点がある。

◆そこで本研究は、密度比推定を用いた異常検知により、異常判定を自動化し技術者の熟練度を問わない異常検知手法を提案する。

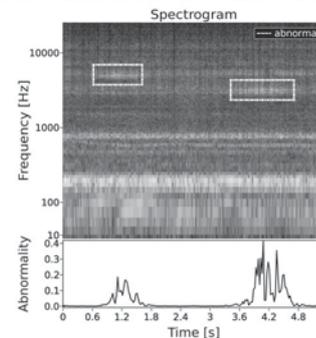


Fig.1: Spectrogram of recorded sound and graph of anomaly level

3-1P-17

3-1P-17 2台のパラメトリックスピーカを使用した音像定位の検討

—聴取者の頭部伝達関数を用いて—

Study on sound localization with two parametric loudspeakers -Using listener's HRTFs-

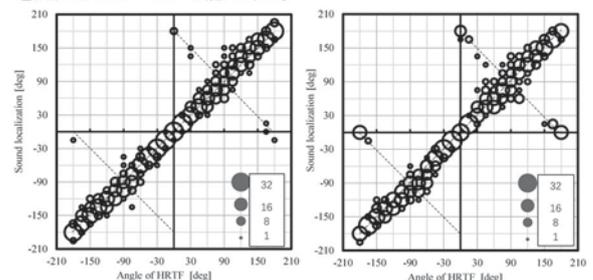
☆福田隆人, △秋山航佑, △野賀琢斗, △吉田壮来,

青木茂明, 島内末廣(金沢工大)

◆これまでの検討で使用されていた MIT で公開されている HRTF と聴取者本人の HRTF を用いてパラメトリックスピーカとヘッドホンによる前後左右方向定位試験を行った。

◆試験には HRTF を畳み込んだピンクノイズを使用し、スピーカは聴取者から見て方位角±60° の方向に 120 cm 離れた。

◆聴取者本人と MIT で公開されている HRTF を使用したパラメトリックスピーカによる定位方向の分布図を示す。音像は全体的に HRTF の角度の方向に頭外定位した。ヘッドホンと同様にパラメトリックスピーカでも本人の HRTF を用いることで水平方向の任意の角度に音像を定位させることが可能である。



Sound localization

3-1P-18

3-1P-18 シヤント音における複数センサ間の差分平均を用いた血管狭窄検出

Detection of Angiostenosis using the average of differences between multiple sensors in shunt murmurs

○西島恵介, 古家賢一(大分大)

- ◆血液透析患者はバスキュラーアクセス(患者と透析装置との間で血液循環するための仕組み)を管理し、狭窄の早期発見が必要である。
- ◆狭窄位置は様々であり狭窄位置に影響しない識別手法が必要である。
- ◆複数位置の聴診から得られたシヤント音を用いて、時間周波数領域において、ある位置のパワースペクトルと残りの位置のパワースペクトルの差分の平均を特徴量として狭窄の識別を試みた。
- ◆データ数の少なさに起因する過学習への対応や精度向上を図るために、時間軸方向の移動・伸縮によるデータ拡張を試みた。
- ◆時間軸方向の前後0.01秒の移動、0.01秒の伸縮によるデータ拡張をしたデータによる学習、データ拡張なしの学習と単独位置による学習、複数位置による学習の組合せで比較した。
- ◆識別にはサポートベクターマシンを用い、最大6箇所まで計測された20症例(正常7、狭窄13)のデータを用いた。
- ◆単独位置の正解率50%に対し、複数位置の正解率65%と向上した。

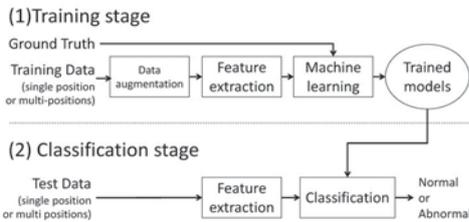


Fig. 1 Classification method

3-4P-2

3-4P-2 振動覚刺激による有声および無声両唇閉鎖音の知覚への影響：周波数に関する探索

Frequency and effect of vibro-tactile stimuli on the perception of voiced and voiceless bilabial stop consonants

◎小野明日香, △中谷桃子, △中根愛(NTT サービスエボリューション研究所), △渡邊淳司, 廣谷定男(NTTコミュニケーション科学基礎研究所)

- ◆近年、音声知覚における聴覚と体性感覚の統合について、声帯振動を模した振動覚刺激により、有声性の対立のある子音間の雑音下での識別が向上する現象が報告されている。
- ◆本研究では、有声および無声両唇閉鎖音を含む1モーラの音声刺激とともに、音声刺激の基本周波数と合致する/合致しない周波数の振動覚刺激を手に呈示し、本現象における周波数の影響を調べた。
- ◆実験の結果、音声刺激の話者の基本周波数に振動覚刺激をチューニングしても効果は強化されなことが示唆された。
- ◆また、聴取者の発話基本周波数と聴取実験の振動覚刺激の周波数の近接性も効果に関連があったとは言えなかった。
- ◆個人の発話特性への刺激のチューニングの影響を確認することが今後の課題である。

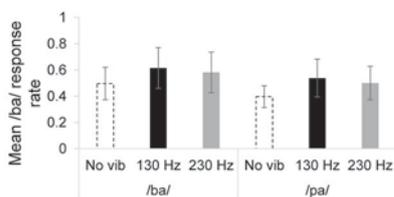


Fig. 1. Mean /ba/ response rate for the different vibration frequency conditions. Error bar indicates 2SE.

3-4P-1

3-4P-1 頭皮上で検出される骨伝導音声の明瞭性に関する検討

Articulation of bone-conducted speech detected on the scalp

☆南里聡志, 信夫大志, 大塚 翔, 中川誠司(千葉大)

- ◆機械工場等の強大騒音下では、通常のマイクロホンの使用は困難である。一方、骨伝導マイクロホンには周囲騒音を検出しにくいという特長があるが、装用性に問題があった。
- ◆我々は、頭皮上から骨伝導音声を検出するヘルメット型骨伝導デバイスを提案したが、頭皮上から検出される骨伝導音声の諸特性には不明な点が多い。
- ◆気導音および頭部の各部位で検出された骨伝導音声を用いた単音節明瞭度試験を行った。また、気導音声を基準として、頭皮上骨伝導音声の Speech Transmission Index (STI) の推定を行った。
- ◆気導音声には劣るものの、頭皮上部位 (Fpz, Cz) から検出された骨伝導音声の明瞭度は 60% 程度に達した。
- ◆一方、既存デバイスの検出部位である喉頭 (Lar) の明瞭性は、頭皮上部位よりも低下した。有声子音の正答率が低下していることから、高周波成分が減衰している可能性が示唆される。

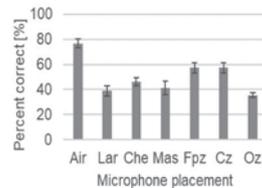


Fig. 1. Percent corrects of the mono-syllable articulation test

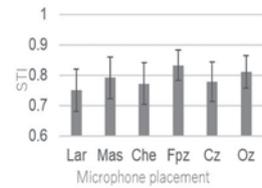


Fig. 2 Values of STI

3-4P-3

3-4P-3 自己聴取音の印象に関する個人性の調査

Investigation of speaker individuality for sense of self-perceived own voices

◎森田翔太(福山大), 鳥谷輝輝, 鶴木祐史(北陸先端大)

- ◆目的: 録音音声と自己聴取音(自身が発話しながら聴いている音声)の音高や音色といった印象の差異を明らかにすること。
- ◆実験: 実験協力者が発話した母音 /a/, /i/, /u/, /e/, /o/ と子音 /sh/, /ts/, /p/, 鼻音 /n/ を録音し、音の違いの有無、音が高い方の音声、音色の違いの有無、音色の印象(音色の印象に関する10種類の因子)について録音音声と自己聴取音を聴き比べさせて回答することを求めた。
- ◆結果:
 - 音高について、音が高いと回答した割合の結果を図1に示す。
 - ◇ /a/や/i/, /e/, /n/において録音音声の音高が高いと判断される割合が高かった。
 - 音色の印象に関する因子の一つ「重い - 軽い」について、自己聴取音に対して録音音声に有する印象の評価結果を図2に示す。
 - ◇ /a/や/i/において、録音音声の方が軽いと感じる割合が高かった。

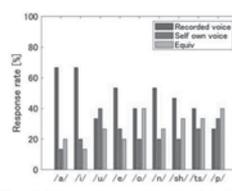


Fig. 1: Response rates for higher-pitch voice in each phoneme.

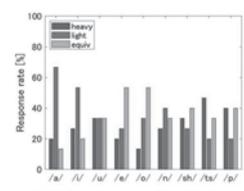


Fig. 2: Response rates for heavy-light in each phoneme.

3-4P-4

3-4P-4 振幅包絡線情報の局部時間反転による音声プライバシー保護の検討

Study on speech privacy protection with locally time-reversing temporal amplitude envelope

☆坂本貴望, 小林まおり, 鶴木祐史 (JAIST)

- ◆目的: 目的音声の振幅包絡線のみを局部時間反転した音声を用いて、目的音声を不明瞭化できるかどうかを明らかにすること
- ◆方法: 2つの親密度(最も高い親密度・最も低い親密度)と9つの局部時間反転の反転区間長(20, 40, 80, 160, 240, 320, 480, 640 ms, ALL(全区間))を条件として、親密度別単語理解度試験を行った。
- ◆結果: 次の3点が明らかになった。
 - 反転区間長を操作することで単語正答率が低下すること
 - 単語正答率の低下の度合いは、単語親密度で異なること
 - 最も効果的な反転区間長は、160 ms 前後であること

◆結論: 振幅包絡線が局部時間反転した音声により、目的音声を不明瞭化できることが明らかになった。

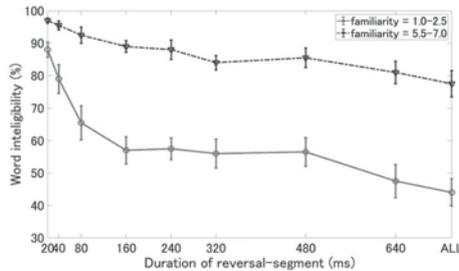


Fig.1: Result of word intelligibility test with familiarity-controlled word lists

3-8P-1

3-8P-1 日本語軟口蓋破裂音調音結合時の舌の前後位置,舌の横幅の変化とその機序の検討

Change of the position of the tongue tip and the width of the tongue during the velar plosive co-articulation and its mechanism.

☆古田尚久, 党建武, 鶴木祐史 (JAIST)

【目的】舌尖の前後位置(以下「舌尖位置」), 舌の横幅(以下「舌幅」)は、舌の引き込みや「棒状の舌」等の病理的運動の指標や、調音運動の機序を検討する手がかりになると考えられる。本研究では、軟口蓋破裂音調音結合時の、舌尖位置と舌幅の変化を計測し機序を検討した。

【方法】被験者10名, EMAにて調音運動を計測し, 下顎歯茎-舌尖のセンサ間距離を舌尖位置の指標, 左右センサ間距離を舌幅の指標とした。音声資料は、(1)単独母音条件(/a~/ /o~/), (2)先行母音変化条件(/aka~/ /oka~/), (3)後続母音変化条件(/akao~/ /okao~/)を用いた。

【結果】軟口蓋破裂音閉鎖時の舌尖位置は、先行、後続母音が /i/ と /e/ の際に前進, /a/ と /o/ の際に後退した。舌幅は、後続母音が /e/ の際に拡大, /o/ の際に縮小した。軟口蓋破裂音全条件全体では、舌尖位置は単母音より前進し下顎歯茎-舌尖間距離が安静時の0.93倍, 舌幅は、母音に比し縮小し左右センサ間距離が安静時の0.97倍となった。

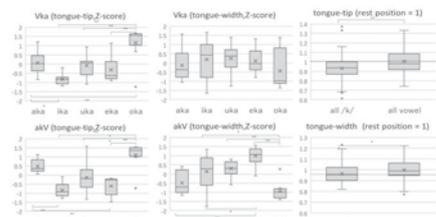


Fig.1: Z-score of tongue-tip position. Large numbers indicate forward movement. Fig.2: Z-score of tongue-width. Large numbers indicate width extension. Fig.3: Tongue-tip position & tongue width of all /V/ & all vowel (rest position = 1)

3-4P-5

3-4P-5 劣化音声知覚におけるトップダウン・ボトムアップ処理に関わる神経基盤: 脳波計測による検討

Neural basis for top-down processing and bottom-up processing in degraded speech perception: A study by EEG measurement

☆齊官重樹, 大塚翔, 中川誠司(千葉大)

- ◆音声知覚の神経基盤を探る研究では、音響的情報の一部を欠落させた劣化音声を用いられてきた。元音声に関する情報を聴取者に与える、あるいは、欠落させる音響的情報を操作することで、劣化音声に対する了解度を調節し、それに伴う脳活動の変化を調べるのが、よく使われる実験パラダイムである。
- ◆前者のアプローチでは、予測などのトップダウン処理が関与する一方で、後者のアプローチでは、音響的情報に基づく音声処理(ボトムアップ処理)が関与すると推察される。両者を比較することで、音声処理のトップダウン処理とボトムアップ処理の神経基盤の解明のために有用な知見が得られると期待できる。
- ◆本研究では、フォルマントの帯域幅を変えることで、劣化音声の了解度を調節し、それに伴う脳活動の変化と元音声提示前後での脳活動の変化を比較した。
- ◆元音声提示前後でハイガンマ帯域の脳波の振幅が増加したのに対して、フォルマントの帯域幅に応じた変化は見られなかった (Fig. 1)。音響的情報に基づくボトムアップ処理と元音声を聴くことで得られた情報に基づくトップダウン処理では、神経基盤が異なることが示唆された。

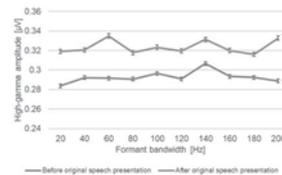


Fig. 1 High-gamma amplitude before and after original speech presentation. ***p<0.01

3-8P-2

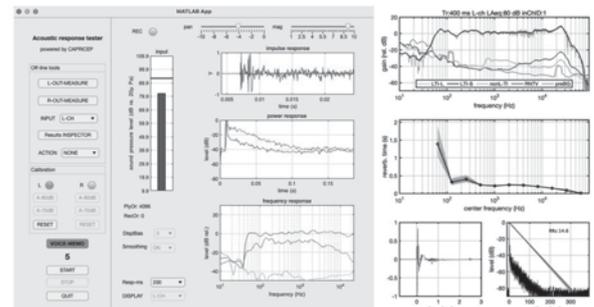
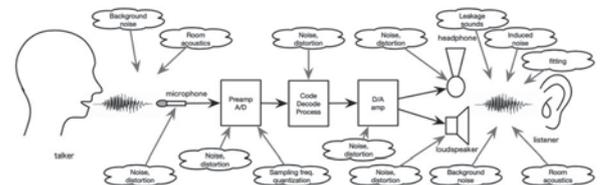
3-8P-2 音声資料の収録・再生環境の簡易な把握に向けて: オールパスフィルタの従属接続に基づく拡張された時間伸長パルスの応用

Toward handy understanding of environment for speech material acquisition and presentation: Application of a cascaded all-pass filters with randomized center frequencies and phase polarities

○河原英紀(和歌山大), 矢田部浩平(早稲田大), 榊原健一(北海道医療大), 水町光徳(九工大), 森勢将雅(明治大), 坂野秀樹(名城大), 入野俊夫(和歌山大)

- ◆音声資料の収録と再生に関わる要因と把握用の測定ツール。

<https://github.com/HidekiKawahara>



Factors contributing speech material acquisition and presentation. The bottom images show GUI of an interactive and real-time acoustic measurement and an example of measurement result.

3-8P-3

3-8P-3 発話のしにくさの自覚と音読潜時の関連の調査

Relationship between awareness of clumsiness while speaking and reaction time for speech production

☆中野智貴 北村達也(甲南大) 能田由紀子(国語研) 正木信夫(ATR)

- ◆背景: 医学的には健康に分類される人の中にも発話のしにくさを自覚する人が存在する。これらの人々の発話を改善するためには、この感覚が発話過程のどの段階に起因するのか明らかにする必要がある。
- ◆目的: 発話のしにくさの自覚の程度と音読潜時との関係を調査する。
- ◆方法: PCの画面に「か」と「た」から成る3モーラもしくは4モーラの無意味語(かかか、たかたか等)を2秒間提示し、それが消えてから0秒から3.5秒の後に合図を表示した。実験参加者はその合図の後できるだけ早くその無意味語を発話することを求められた
- ◆結果: 分散分析の結果、発話のしにくさの自覚の程度は音読潜時に対して有意に影響を及ぼすことが明らかになった。

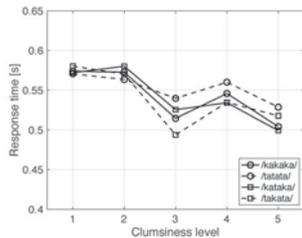


Fig. 1: Effects of awareness of clumsiness for reaction time of the utterance of 3-mora words.

3-8P-5

3-8P-5 中国人学習者, 小学3年生児童, 成人母語話者が産出する日本語真偽/疑問詞疑問文のイントネーション

Intonation of Japanese Yes-No Questions and WH Questions Read Aloud by Chinese Learners, Third Grade Elementary School Children and Adult Native Speakers

○伊達宏子, 花蘭悟(東外大), △郭田夫(元東外大院士)

- ◆日本語上級中国人学習者 (Fig.1 上段), 小学3年生児童 (Fig.1 中段), 成人母語話者 (Fig.1 下段) の真偽/疑問詞疑問文「箱の中に何か/何が ありますか」等のイントネーションを比較した。

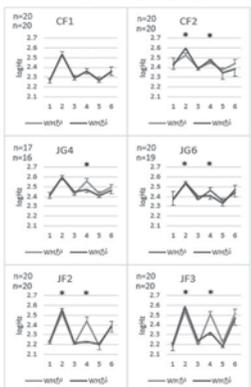


Fig.1 High and low patterns of discontinuous pitch in the phrases "PLACE ni(1) WH(2) ka/ga(3) V ma(4)su(5)ka(6)" uttered by CH learners in the upper row, JP children in the middle row and JP adult speakers in the lower row.

- ◆成人母語話者(下段)は真偽疑問文の述部隆起と疑問詞疑問文の述部抑制に明確な区別が見られ、疑問詞部にも有意な違いが見られた。
- ◆小3児童(中段)は両疑問文の述部に明確な区別が見られ、疑問詞部は有意な差が見られた話者と見られなかった話者がいた。しかし、成人と同様に両文を明確に発話し分けていることが分かった。
- ◆中国人学習者(上段)は両疑問文で同様なパターン、疑問詞疑問文で疑問詞部も述部も高くなるパターン等があった。話者内・話者間共にばらつきが大きく、上級話者でも母語話者パターンとは違いがあった。

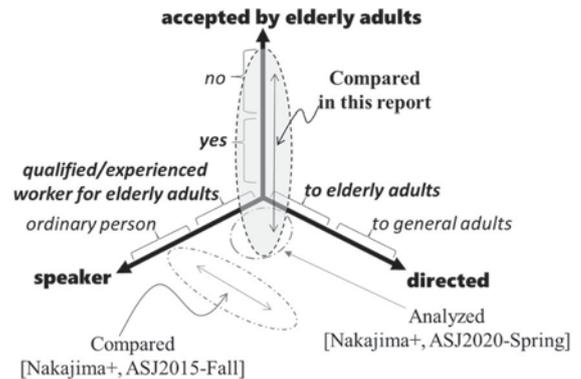
3-8P-4

3-8P-4 高齢者にとって聞いて分かりやすい話し方を解明するための高齢者向け模範音声と反模範音声との比較

Comparison between elderly adults directed exemplary and non-exemplary speech toward clarifying easy-to-understand speaking style for elderly adults

○中嶋秀治(日本電信電話(株) NTTコミュニケーション科学基礎研究所)

- ◆目的: 高齢者にとって聞いて分かりやすい音声や話し方の特徴の探索を行なっている。
- ◆進め方: 音声や話し方に焦点をあてるため、難聴や認知症を患っていない高齢者を対象として、文間ポーズ長の分析を現在は進めている
- ◆既報と本報との関係:



- ◆結果: 文間ポーズ長の使い分けの特徴として、5つの特徴(本文の※1から5)が、「高齢者対応の有資格者/経験者によって、高齢者を聞き手として想定して話され、かつ、高齢者が聞いて分かりやすいと高く評価した音声(上の図の灰色部分下側)」において確認された

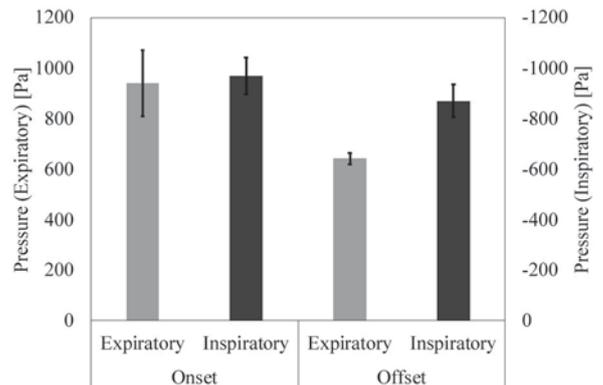
3-8P-6

3-8P-6 声帯物理モデルを用いた吸気発声実験

Experimental study of inspiratory phonation using physical model of the vocal folds

☆松本拓磨(立命館大), 金谷麻由佳(立命館大), △仲川拓臣(立命館大) △植村泰佑(立命館大), △松田竜樹(立命館大), 徳田功(立命館大)

- ◆声帯モデルにMRIモデルを用いて吸気発声実験を行った。
- ◆同一の声帯モデルを用いて、吸気発声時および呼気発声時における声門下圧、音声、高速度動画を計測した。
- ◆呼気発声と比較して、吸気発声では、オンセット圧、オフセット圧とも上昇傾向にあり、人の呼気発声に関する研究結果と矛盾しないことが確認できた。



◆Fig.1 Comparison of onset and offset pressures between expiratory and inspiratory phonations.

3-8P-7

3-8P-7 ブタ摘出喉頭における 内転レベルの発声効率への影響

Effects of adduction level on vocal efficiency in excised pig larynxes
 ☆宮崎琳太郎(立命館大), △吉井玲音(立命館大), △安川文雄(立命館大),
 石村憲意(立命館大), △西村剛(京都大), 徳田功(立命館大)

- ◆ 4個のブタ摘出喉頭を用いて吹鳴実験を行なった。
- ◆ 披裂軟骨間にアクリルプレートを挿入することで内転レベルを変化させ、発声効率を求めた。
- ◆ 仮声帯がある状態と除去した後の発声効率の比較を行なった。
- ◆ 4個の喉頭から発声効率の平均値を求めた。声門間隙を広げるに伴い、発声効率は減少した。仮声帯を切除した場合も同様であった。

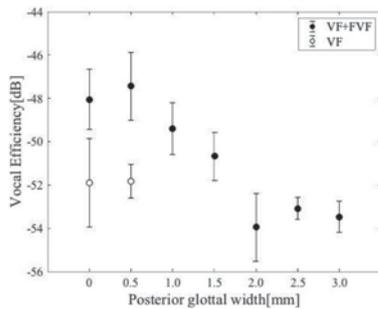


Fig.1 Vocal efficiency as a function of posterior glottal width with (black) and without (white) false vocal folds.

3-8P-8

3-8P-8 声帯膜を模擬した物理モデルの構築

Construction of physical model for vocal membranes

☆金谷麻由佳(立命館大), 松本拓磨(立命館大), 宮崎琳太郎(立命館大学)
 石村憲意(立命館大), △西村剛(京都大学), 徳田功(立命館大)

- ◆ 動物声帯に見られる声帯膜を模擬する物理モデルを構築した。
- ◆ 声帯膜の形状は、ブタやマカク属を参考にした。
- ◆ 声帯膜のある物理モデルと声帯膜のない物理モデルを作成し、吹鳴実験を通して、各モデルの声門下圧、音声、高速度動画を計測した。
- ◆ ブタ (*Sus scrofa domestica*) の摘出喉頭でも同様の実験を行い、振動特性を比較した。
- ◆ 声帯膜があることでオンセット圧が減少することが分かった。

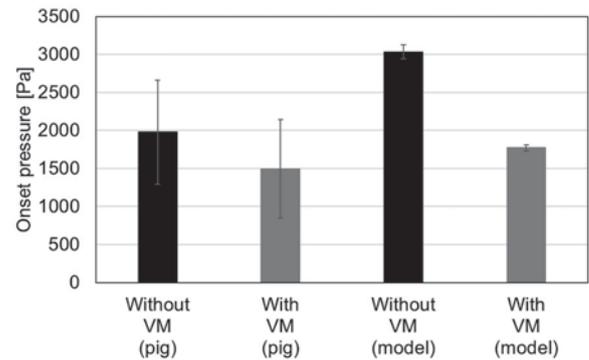


Fig.1: Onset pressure