

令和 4 年 6 月 6 日現在

機関番号：10103

研究種目：若手研究

研究期間：2019～2021

課題番号：19K15146

研究課題名（和文）マイクロホンネットワークと高品質音声合成による「よく聴こえる」拡声システム

研究課題名（英文）A "easy hearing" public address system with microphone network and high quality speech synthesis

研究代表者

小林 洋介（Kobayashi, Yosuke）

室蘭工業大学・大学院工学研究科・助教

研究者番号：10735103

交付決定額（研究期間全体）：（直接経費） 3,200,000円

研究成果の概要（和文）：屋外拡声器の聴き取りを改善するために、複数マイクロホンを利用した音声品質予測システムとその品質をフィードバックした音声合成システムに関する基礎検討をおこなった。音の計測にはマイクロホンアレイを用いたプロトタイプを作成した。音声品質には深層学習を利用した品質予測モデルを開発した。音声合成に関しては通信時にフィルタ処理される高域を補完する信号処理を実装した。本研究期間ではこれらの単体テストが主だったが、複合的に連携した拡声システムが実現すればより良い聴こえが実現につながると考えられる。

研究成果の学術的意義や社会的意義

屋外拡声に聴き取りにくいことがあることは多くの人が理解していることである。なぜ聴こえにくいのかの要因は騒音や風の影響、残響（響き）の影響など多数ある。この問題を解決するには、どの程度聞き取りにくいのかを定量化する事と、この値をフィードバックしながら聴こえやすく音声処理し直すことの二つが必要である。本研究では、拡声音声の計測（録音）、計測信号の品質評価、品質値の音声合成へのフィードバックを一体的に行うシステムの基礎検討をおこなった。

研究成果の概要（英文）：In order to improve the listening of outdoor public address (PA) system, we conducted a basic study on a speech quality prediction using multiple microphones and a speech synthesis that feeds back the PA speech quality. First, we developed a prototype using a microphone array for PA speech measurement. Next, we developed a speech quality prediction model that uses deep learning. Finally, for speech synthesis, we implemented a statistical speech synthesis model that complements the high frequencies filtered by wireless communication. In this study, only these unit tests were performed. However, if a PA system that cooperates in a complex manner is realized, it is thought that better hearing will be realized.

研究分野：音響工学

キーワード：屋外拡声システム 音声了解度 音声品質 音声合成

### 1. 研究開始当初の背景

拡声システムは音声情報を多くの人に伝達するため、学校・駅・空港など公共施設のアナウンス、防災行政無線の放送塔、スーパーマーケット内の販促用音声などあらゆる場所に設置されている。しかし、直感的に利用できるが故に最適とは程遠い利用が多く、過度なゲインによる音割れやボソボソとした発話など放送音声からの情報収集が困難になることが多い。

一方、最近ではスマートフォンなどから確実に目的の情報収集が可能であるため、日常生活では拡声装置の意味は薄れてきている。しかし、都市部の駅や地下街のような多くの市民が集まる環境で災害が発生した場合の避難誘導に拡声装置は現代でも欠かせず、聞き取りやすい(=高明瞭化)ことは重要である。

### 2. 研究の目的

本研究では、図1における 1.主観品質予測値を拡声音声へフィードバックするための音声処理の研究開発を行う。 具体的には、元の音声の発話速度などの聞き取りに影響する音声パラメータを操作する音声合成を行う。加えて、提案システムの社会実装を目指すため、2.実際の防災用行政無線の屋外拡声音をできるだけ模擬した実環境でシステムの性能評価を行う。

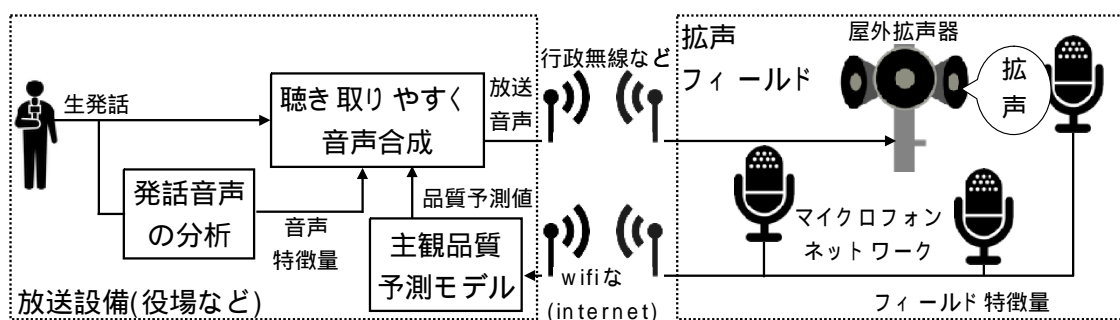


図1 本研究での提案

### 3. 研究の方法

#### 3.1. 屋外での録音

実際に屋外で拡声器メーカーと協力して拡声器を設置し、マイクロホンアレイ、騒音計、ダミーヘッドを立てて、インパルス応答や拡声音声を録音しさまざまな分析に用いることができる音声データセットを作成した。

#### 3.2. 音声品質予測

3.1 と同等の手法で 2017 年度に収集したインパルス応答を利用してモデル学習用音源を作成し、深層学習やスパースモデリングの手法を用いた音声品質予測モデルを学習し、その性能を評価した。

#### 3.3. 音声品質の改善

最終的な品質フィードバックの前に行政無線などの通信の仕様で劣化する音声の高域部分の補完に Pix2Pix モデルを利用することを検討し、その学習音源数や基本的な品質評価を行った。Pix2Pix モデルの概念を図2に示す。このモデルは入力音声の音響特徴量を目的音声の特徴量に U-net を用いて変換し、変換後のスペクトログラムを高精細化する 2 弾構成の変換モデルである。今回は元音声を帯域制限された音声をアップサンプリングした信号、声質変換後音声を高域が保持されている原音声として実験を行った

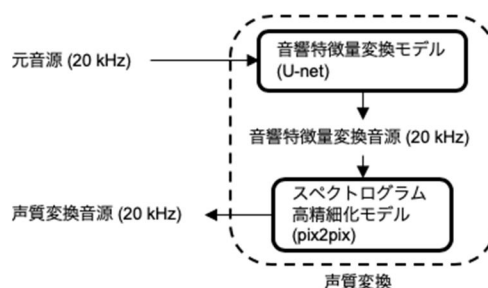


図2 Pix2Pix のフロー図

### 4. 研究成果

#### 4.1. 屋外での録音

2019年6月26日に北海道白糠町で3種類のスピーカ(従来型のレフレックス, 大型のホーン

アレイ、小型のスリムアレイ)の音源を録音した。録音条件を Table 1 に、各録音地点で録音したダミーヘッドのインパルス応答を評価音源にたたみ込んで作成した評価音源を室蘭工業大学の防音ブースで 24 名(男性 22 人, 女性 2 名, 平均年齢 22.2 歳)の聴取実験により評価した聞き取りにくさ尺度 (LDR: Listening Difficulty Rating) の結果を Table 2 に示す。LDR は聞き取りにくいときに 1, 聞き取りにくくはないときに 0 となり, 同一条件で LDR の低い方が聞き取りにくくはない拡声装置といえる。

聴取実験の結果より, 平均的に従来型のレフレックスホーン単体より, 新しいアレイ/スリムタイプの方が聞き取りにくくはないことがわかる。特に拡声地点から直線上にある条件 B はアレイタイプがスリムタイプより LDR が低く, 拡声地点から斜め方向にある条件 C ではスリムタイプの方が LDR より低いことから, スピーカの指向性も結果に影響していると考えられる。

表 1 録音条件

条件	距離 (m)	備考
A	650	スピーカの正面で遮蔽物等なし
A'	650	拡声方向を 50° 傾ける
B	1200	スピーカの正面で途中に遮蔽物あり
C	1000	丘を隔てて録音
D	400	拡声方向と逆側

表 2 主観評価による LDR の値

条件	レフレックス	ホーンアレイ	スリムアレイ
A	0.34	0.04	0.08
A'	0.96	0.44	0.33
B	0.90	0.44	0.72
C	0.91	0.73	0.67
D	0.98	0.99	1.00
平均	0.82	0.53	0.56

#### 4.2. 音声品質予測

4.1 の実験以前の 2017 年度に収集した拡声声源のインパルス応答を用いて原音と雑音抑圧音声との時間-周波数分析に基づく知覚的歪みをモデル化した Short-Time Objective Intelligibility (STOI), 原音と雑音抑圧音声との相互情報量を用いた Speech intelligibility prediction based on mutual information (SIMI), STOI の相関係数計算部を正規化指標とした Extended STOI (ESTOI) による LDR 推定モデルの評価を行った。結果を表 3 に示す。この表では SNR ごとに相関係数と主観評価値との平均二乗誤差 RMSE で比較している。3 指標とも, 相関係数が 0.97 以上であり相関が非常に高く, 平均 RMSE も 0.1 以下と十分な性能を持つ。

表 3 従来型了解度指標による推定精度

SNR (dB)	overall		0		10	
	$\rho$	RMSE	$\rho$	RMSE	$\rho$	RMSE
Index						
STOI	0.973	0.090	0.136	0.011	0.496	0.036
SIMI	<b>0.981</b>	<b>0.076</b>	0.227	<b>0.002</b>	0.549	0.042
ESTOI	0.978	0.081	<b>0.327</b>	0.006	<b>0.633</b>	<b>0.033</b>
SNR (dB)	20		30		40	
Index	$\rho$	RMSE	$\rho$	RMSE	$\rho$	RMSE
STOI	0.873	0.059	0.601	0.166	0.619	0.091
SIMI	<b>0.899</b>	<b>0.052</b>	<b>0.781</b>	<b>0.135</b>	<b>0.815</b>	0.080
ESTOI	0.893	0.074	0.687	0.142	0.777	<b>0.078</b>

しかし, 従来型の指標は原音との比較が必要であるため, 自動フィードバックシステムに用いるには原音が入手できない場合に利用できない問題がある。この問題を解決するために, 指標値自体を予測する機械学習モデルを学習した。音響特徴量に Mel-Frequency Cepstral Coefficients (MFCCs) を利用し, 指標値を予測し, 表 3 の計算で用いた各指標の LDR の推定を行う。決定木を用いたランダムフォレスト(RF), スパースモデリングの一手法である Elastic net(EN)を用いた場合と深層学習によるニューラルネットワーク(DNN)を用いた場合の推定二乗誤差 RMSR の一覧を表 4 に示す。結果より, 原音がわかる表 3 の RMSE より全体的に数値が大きく悪化している。平均的には DNN による STOI の予測モデルが最も予測誤差が小さくなることが示された。しかし, SNR によって旗の組み合わせがよく, ターゲットとする拡声フィールドに合わせて最適に選択する必要がある。

表 4 機械学習による指標予測を入れた了解度推定精度

SNR (dB)	overall			0			10		
	RF	EN	DNN	RF	EN	DNN	RF	EN	DNN
STOI	0.218	0.379	<b>0.159</b>	0.234	0.167	0.171	0.185	0.167	0.125
SIMI	0.328	0.479	0.189	0.112	<b>0.033</b>	0.100	0.060	<b>0.049</b>	0.051
ESTOI	0.219	0.406	0.170	0.262	0.166	0.136	0.219	0.174	0.081
SNR (dB)	20			30			40		
Algorithm	RF	EN	DNN	RF	EN	DNN	RF	EN	DNN
STOI	0.216	0.161	0.205	0.158	0.400	0.151	0.276	0.690	<b>0.131</b>
SIMI	<b>0.062</b>	0.117	0.095	0.413	0.573	0.232	0.588	0.896	0.320
ESTOI	0.253	0.171	0.090	<b>0.139</b>	0.433	0.201	0.200	0.743	0.265

#### 4.3. 音声品質の改善

元音声の高域が通信で帯域制限されるのは狭帯域通信である電話系と共通しており, 本研究では帯域制限後の電話系での高域復元を実験対象とした。図 4(a)に通信した音源を, (b) に通信による帯域制限後の音源のスペクトログラムを示す。(a)と比べ(b)は 4000 Hz 以上の成分が失われているのがわかる。

Pix2Pix による音響特徴量変換モデルの学習には、電話音源と元音源の音響特徴量を用いた。学習に用いた音響特徴量は、声高を表す基本周波数と声質を表すスペクトル包絡の 2 つである。スペクトル包絡はメルケプストラムに変換することで、音響特徴量の次元数を減らしている。スペクトログラム高精細化モデルの学習は、元音源のスペクトログラムを学習データに用いる。ペア学習データとして、元音源を低次元音響特徴量で再構成したスペクトログラムが用いられる。音源の発話内容は ATR 音素バランス 503 文の J セットを除く 450 文を学習データ、J セット 53 文をテストデータとして声質変換の各モデルに用いた。モデルの反復学習回数であるイテレーションは、音響特徴量変換モデルを 100,000、スペクトログラム高精細化モデルを 300,000 とした。Pix2Pix による復元後の音声を(c)に示す。(b)と比べ(a)に近いスペクトログラムが復元されている。音声品質の客観品質を 4.2 でも用いた STOI で原音と帯域制限音声および帯域復元音声を比較し、表 5 に示す。結果より、話者によって STOI 値で改善しないことがあった。これは STOI が原音声のパワースペクトル包絡の復元度で評価するため、包絡構造が高域補完によりより大きく変化したことが原因と考えられ、最適な評価指標とはならなかった。今後は性能評価指標も含めた検討が必要である。

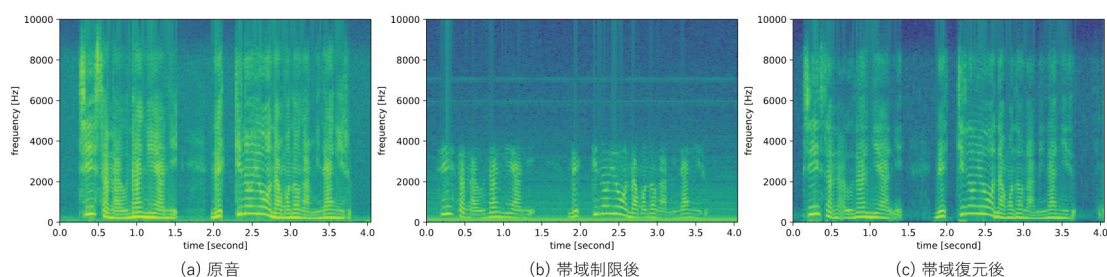


図 4 音声のスペクトログラム

表 5 話者別の帯域制限音声と帯域復元音声の STOI 値

Speaker	FKN	FYM	MMY	MSH
Tel. band	<b>0.131</b>	0.233	<b>0.884</b>	0.213
Wide band	0.080	<b>0.243</b>	0.820	<b>0.229</b>

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 2件/うち国際共著 0件/うちオープンアクセス 1件）

1. 著者名 Noguchi Keita, Kobayashi Yosuke, Kishigami Jay, Kurisu Kiyohiro	4. 巻 41
2. 論文標題 Listening difficulty estimation model using short-time objective intelligibility measure for outdoor public address systems	5. 発行年 2020年
3. 雑誌名 Acoustical Science and Technology	6. 最初と最後の頁 420 ~ 422
掲載論文のDOI (デジタルオブジェクト識別子) 10.1250/ast.41.420	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Kobayashi Yosuke, Kondo Kazuhiro	4. 巻 156
2. 論文標題 Japanese speech intelligibility estimation and prediction using objective intelligibility indices under noisy and reverberant conditions	5. 発行年 2019年
3. 雑誌名 Applied Acoustics	6. 最初と最後の頁 327 ~ 335
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.apacoust.2019.07.034	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計13件（うち招待講演 0件/うち国際学会 2件）

1. 発表者名 小林洋介, 野口啓太, 鳴野寛, 栗栖清浩,
2. 発表標題 バイノーラル録音による屋外拡声スピーカ音の主観評価
3. 学会等名 日本音響学会2020年秋季研究発表会
4. 発表年 2020年

1. 発表者名 小林洋介, 野口啓太, 栗栖清浩
2. 発表標題 屋外での長距離伝搬条件で計測した HRIR を畳み込んだ拡声声の了解度評価
3. 学会等名 日本音響学会2021年春季研究発表会
4. 発表年 2021年

1. 発表者名 Keita Noguch, Yosuke Kobayashi, Jay Kishigami and Kiyohiro Kurisu
2. 発表標題 Comparison of STOI-type Intermediate Feature for Listening Difficulty Rating Prediction Model
3. 学会等名 The 8th IEEE Global Conference on Consumer Electronics (国際学会)
4. 発表年 2019年

1. 発表者名 Hirokazu Akadomari, Yuhi Sato, and Yosuke Kobayashi
2. 発表標題 The 8th IEEE Global Conference on Consumer Electronics
3. 学会等名 The 8th IEEE Global Conference on Consumer Electronics (国際学会)
4. 発表年 2019年

1. 発表者名 野口啓太, 小林洋介, 岸上順一, 栗栖清浩
2. 発表標題 屋外拡声品質予測モデルの中間特徴量の検討
3. 学会等名 情報処理学会SIGMUS
4. 発表年 2019年

1. 発表者名 野口啓太, 小林洋介, 岸上順一, 栗栖清浩
2. 発表標題 STOI型客観了解度指標を用いた屋外拡声音の聴き取りにくさ予測モデルの評価
3. 学会等名 電子情報通信学会応用音響研究会
4. 発表年 2019年

1. 発表者名 小林洋介, 近藤和弘
2. 発表標題 原音予測モデルを利用した客観音声了解度指標の提案
3. 学会等名 日本音響学会2020年春季研究発表会
4. 発表年 2020年

1. 発表者名 小林洋介, 山部匠, 野口啓太, 赤泊寛和
2. 発表標題 pix2pix話者変換を用いた狭帯域音声の高域補間
3. 学会等名 日本音響学会2020年春季研究発表会
4. 発表年 2020年

1. 発表者名 野口啓太, 小林洋介, 岸上順一
2. 発表標題 中間特徴量を用いた屋外拡声音評価システムの実装
3. 学会等名 日本音響学会2020年春季研究発表会
4. 発表年 2020年

1. 発表者名 野口啓太, 小林洋介, 岸上順一, 栗栖清浩
2. 発表標題 STOI型客観了解度指標を用いた主観評価予測モデルの詳細評価
3. 学会等名 日本音響学会2019年秋季研究発表会
4. 発表年 2019年

1. 発表者名 野口啓太, 小林洋介, 岸上順一, 栗栖清浩
2. 発表標題 Elastic Netを用いた屋外拡声音の聴き取りにくさ予測
3. 学会等名 令和元年度電気・情報関係学会北海道支部連合大会
4. 発表年 2019年

1. 発表者名 赤泊寛和, 佐藤裕飛, 小林洋介
2. 発表標題 2段階声質変換システムにおける学習文章数の比較
3. 学会等名 令和元年度電気・情報関係学会北海道支部連合大会
4. 発表年 2019年

1. 発表者名 山部匠, 野口啓太, 赤泊寛和, 小林洋介
2. 発表標題 pix2pix話者変換を用いた狭帯域音声の高域補間
3. 学会等名 令和元年度電気・情報関係学会北海道支部連合大会
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究協力者	近藤 和弘  (Kondo Kazuhiro)  (10312753)	山形大学・大学院理工学研究科・教授    (11501)	



6. 研究組織（つづき）

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究協力者	岸上 順一  (Kishigami Jay)	室蘭工業大学・大学院工学研究科・教授    (10103)	
研究協力者	栗栖 清浩  (Kurusu Kiyohiro)		

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関