

**科学研究費助成事業 研究成果報告書**

平成 29 年 6 月 16 日現在

機関番号：12612

研究種目：基盤研究(C) (一般)

研究期間：2014～2016

課題番号：26330188

研究課題名(和文) 時間周波数平面上の成分移動による高耐性音声の研究と研究者用音声データベースの作成

研究課題名(英文) Study on robust speech by component movement on time frequency plane and construction of speech database for researchers

研究代表者

高橋 弘太 (Takahashi, Kota)

電気通信大学・大学院情報理工学研究科・准教授

研究者番号：10188005

交付決定額(研究期間全体)：(直接経費) 3,600,000円

研究成果の概要(和文)：音声の妨害音に対する耐性を高めて音提示する方法については、線形予測分析と変調スペクトル解析法を組み合わせ、母音部分の聴き取り向上と子音部分の聴き取り向上を別個に解決する方法を提案した。さらに、実験により提案法の有効性を示した。また、音響信号を時間周波数平面上に展開しその成分移動を行うことで聴き取りを向上させることが有効であることも示した。また、本課題では、音声の聴き取り実験を行うために作成し公開している独自の音声データベースを拡充することも目指した。当該期間中に、「1文字違い」音声セットと、「擬似的カーナビ」音声セットを充実させ、研究者が利用しやすい形で公開した。

研究成果の概要(英文)：For the method of enhancing the robustness of speech against the disturbance sound, we proposed a method to solve separately improvement of hearing of vowel part and hearing improvement of consonant part by combining linear prediction analysis and modulation spectrum analysis method. Experiments showed the effectiveness of the proposed method. We also showed that it is effective to improve listening by expanding the acoustic signal on the time frequency plane and moving its components. In addition, in this study, we aimed to expand our original speech database which was created and made public for audition test of speech. During that period, we enhanced the "one character difference" speech set and the "pseudo car navigation" speech set, and made it available for researchers to use easily.

研究分野：デジタル信号処理

キーワード：話速変換 音声データベース 時間周波数平面

### 1. 研究開始当初の背景

テレビ放送の多チャンネル化と、ネットによる番組配信により、視聴覚コンテンツ量は爆発的に増大している。一方、コンテンツを収録するHDDレコーダ等の記憶容量も着実に増加し、結果として現代人がレコーダに録り溜めているコンテンツ数は著しく増大している。ある研究機関の調査によれば、HDDレコーダに録り溜めてある番組の平均数は37.3番組であり、1年で4割の伸びがあるという。このような背景のもと、究極の高速再生を目指してはじまった「効率的視聴法」の研究が本研究の第一の源流である。

本研究の第2の源流は、平成21年度より取り組んでいる「スマートミキサー」の研究である。これは、ミキシングエンジニアが経験と勘を駆使して行っていたミキシングを、信号処理技術によって誰にでも行えるものとして実現することを目指したものである。技術の根幹は、ソース信号を時間周波数平面に展開し、時間周波数平面同士の非線形操作での混合としてミキシングを行う枠組みである。この研究を通して、時間周波数平面を土台とした信号処理の有効性を認識できただけでなく、今回計画した研究を進める上で必須となるソフトウェアやハードウェア環境を整備することができた。例えば、ソフト面では、時間周波数平面上での効率的な処理に特化した開発環境とライブラリを独自に構築しており、ライブラリ上に準備された各処理要素を選択し結合するだけで、処理の実行と結果の評価の実験が手際よく行えるようになっている。

第一の源流である「フレキシブルな時間軸を持つ音声再生法」は、効率的視聴のために、時間軸方向に信号成分の移動を行う方法であった。これに、第二の源流であるスマートミキサーの時間周波数平面上で処理するという発想を取り入れ、時間軸だけでなく時間周波数平面上で、自在に信号の成分移動を行い、聴取者によって聞き取りやすい音声を作成する方法を確立しようと考え、本研究を開始した。

また、研究代表者は、本研究開始前から、研究者用音声データベースの構築を手掛けており、本研究のために録音する音素材を含めてこのデータベースをさらに充実させ、国内外の研究者に利用してもらうことも願って、本研究を計画した。

### 2. 研究の目的

以下を目的にして研究を進めた。

(1) 目的の第一は、騒音耐性の強い音声の作成方法の研究である。たとえば、騒音化でメガホンによって音声拡声を行うとき、あらかじめ音声に信号処理を施しておくことで、発話内容を聞き取りやすくする方法について研究する。

(2) 目的の第二は、時間周波数平面上での信号処理技術に取り組み、音声などの音源の信号を時間周波数平面に予め展開しておき、

時間周波数平面上で成分移動を行うことでより耐性の高い音を作成する方法について研究し、その有効性を見極めることである。

(3) 目的の第三は、音声データベースの拡張である。研究開始時には、話速を可変とした音声については、十分拡充していたが、音声の細かな差異が識別できるか否かを実験するための素材は十分とは言えなかった。そこで、拡張にあたっては、一文字のみの差異のある原稿の発話や、カーナビ音声を模擬した発話を含めてさらに充実させるようにして、より多くの研究者に利用してもらうことを目指した。

### 3. 研究の方法

#### (1) 耐性の高い音声の作成

一般に利用されているメガホンを用意し、その伝達特性を計測した上でその特性を積み込む方法で、騒音化における拡声をシミュレートし、どのような処理が効果的であるかを検討した。その結果、全帯域を単一の処理方法で統一的に変更するよりも、帯域ごとに適切な方法で処理することが有効であることがわかった。具体的には、ランニングスペクトルを求めた結果を使った変調スペクトル分析により子音判定を行い、高域においては重要な子音部分を強調することが良い結果をもたらすことがわかり、また、低域においては、線形予測分析結果を使ったフォルマント分析により母音の必要部分を強調することが有効であることがわかった。具体的には、図1の仕組みでゲインマスクを作成して、

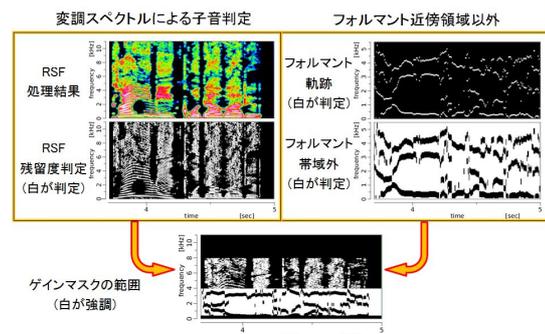


図1 耐性の高い音声の作成

このゲインマスクで時間周波数平面上での強調と抑制を行うことが有望と考え、本手法を提案法として実験を行った。

#### (2) 時間周波数平面上での成分移動

例えば、音楽を再生中にカーナビの音声を割り込ませるとき、0.2秒だけカーナビ音声を遅らせることで、重要な音韻を確実に聞き取らせることができるような状況は決して珍しくない。このようなとき、我々が提案する「時間移動」を行うことが有効である。また、ある音韻を聞かせるとき、音声の重要な周波数成分をわずかに1kHzだけシフトすれば、一定量の音色変化に目を瞑れば音声を確実に聞き取らせることができる一方で背景の

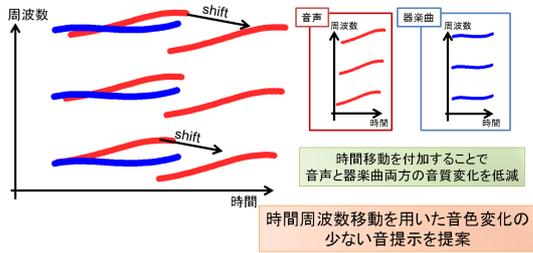


図2 時間周波数平面上での成分移動

音楽に対する悪影響も軽減できるという場面もありうる。このような場面で有効なのが、「周波数移動」である。本研究では、時間周波数平面において、時間移動と周波数移動を自在に駆使するための手法を確立した。具体的には、信号成分を時間周波数平面上で局在化した領域に分割し、その各領域を自在に移動することで、各音源にどのようなダメージが与えられ、また、聴き取りに関してどのようなメリットが与えられるかを評価できるようにした。複数の移動候補に対してこの評価を行うことで、最も効果の高い移動が達成できるようにした。

33. 彼は具にこだわる。
34. 彼は紀伊にこだわる。
35. 彼は杭にこだわる。
36. 彼は刑にこだわる。
37. 彼は恋にこだわる。
38. 彼は差巽にこだわる。
39. 彼は鯛にこだわる。
40. 彼は地位にこだわる。
41. 彼は二位にこだわる。
42. 彼は墨にこだわる。

43. あとご飯ください。
44. あと五品ください。
45. あと五分ください。
46. あと五篇ください。
47. あと五本ください。
48. あと五円ください。
49. あと五点ください。
50. あと五年ください。
51. あと五万ください。
52. あと五問ください。

53. そのお面は傾いている。
54. その仮面は傾いている。
55. その湖面は傾いている。
56. その譜面は傾いている。
57. その木綿は傾いている。
58. その路面は傾いている。
59. その画面は傾いている。
60. その座面は傾いている。
61. その図面は傾いている。
62. その場面は傾いている。

63. これが秋である。
64. これが雨期である。
65. これが駅である。
66. これが牡蠣である。
67. これが危機である。
68. これが四季である。
69. これが席である。
70. これが劇である。
71. これが時期である。
72. これが簿記である。

### (3) 音声データベースの拡張

一文字違いのデータベースについては、一文字ぶんの聞き違いがあっても日本語文章として矛盾のないものであることが好ましい。このような原稿を用意することは容易ではなかったが、左のような原稿を作成することができた。この原稿をプロのアナウンサーに朗読してもらい、防音室で収録した。研究用であるので、コンプレッサやイコライザなど、音声に魅力を高めるような人工的なエフェクタは一切使用しないことが原則である。ただし、アナウンサーが唇を開いたり閉じたりした瞬間のノイズ（一般にリップノイズと呼ばれるもの）は注意して発音してもらっても一定量は収録されてしまう。そこ

で、リップノイズを除去する GUI ツールを開発し、そのツールを使用して手作業で時間をかけてリップノイズを除去した。

## 4. 研究成果

### (1) 耐性の高い音声の作成

拡声器を発声デバイスとして用いた場合の実験結果を図3に示す。

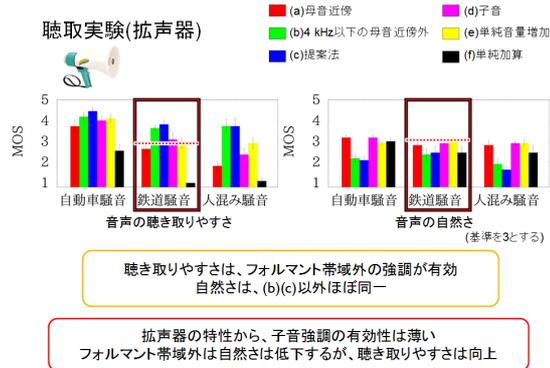


図3 耐性の高い音声に関する実験結果

本実験においては、騒音源として、自動車騒音、鉄道騒音、人混み騒音の3種類を用いている。グラフの縦軸は主観的評価である MOS 値 (mean opinion score) である。この結果よりは提案法は、音声の自然さについては若干のマイナスがあるものの、聴き取り易さにおいて良い評価を得られていることがわかる。

### (2) 時間周波数平面上での成分移動

3種類の音源セットを用意して単純加算、従来法、周波数移動のみを行うもの、時間周波数移動を行うもの、を MOS 値で比較した結果

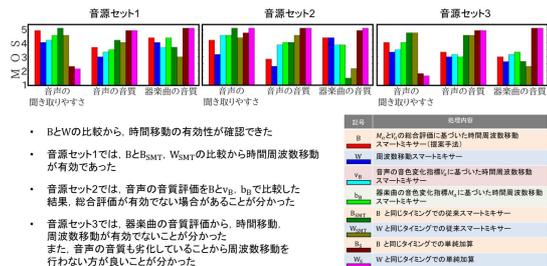


図4 耐性の高い音声に関する実験結果

を図4に示す。この結果より、音源セット1では時間周波数移動が有効であることがわかった。一方、音源セット2,3については、時間移動は有効であるが周波数移動については効果が薄いこともわかった。周波数移動は、音源のスペクトルを破壊することを認めるといった積極的音加工であるので、その利用に関しては、さらに研究を進めて詳しい知見を得る必要があることも認識された。

### (3) 音声データベースの拡張

一文字違いのデータベースと、カーナビ音声の模擬発話のデータベースを充実させ、さら

## 録音条件と機材

使用マイク	ロホン
(1) SONY C-388 (指向特性: 単一指向性, ローカット特性: M, ハイカットスイッチ: OFF, PADスイッチ: 0 dB)	
(2) NEUMANN U87Ai (指向特性: 単一指向性, VF00, VF01, PF02に使用)	
プリアンプ	GRACE m101 × 2
AD変換器	Lynx AURORA-16
収録部屋	床面積: 24平方メートル (4 [m] × 6 [m]) 天井までの高さ: 2.65 [m] 絨毯敷, カーテン有.
サブソニックフィルタ	80[Hz]において-5dB落ちとなるHPFを, 直線位相FIRフィルタで設計し, 収録後にフィルタリング.
話速制御の単位	データセット3~6では, 自由発話における個々の文章間の相対的語速を一定に保つたまま, 記事全体(1ストーリー全部)での平均語速を設定値になるように制御. 全ての台詞が連結されたファイルにおいては, ファイルの先頭0.5秒, 末尾0.5秒に無音区間が挿入されているので, 「総モーラ数」を「無音時間を削除した時区間の長さ」で除算すると, 正確に表示の語速値になる. 台詞ごとに分割したファイルにおいては, 表示語速に $\frac{1}{1.2}$ における相対語速を乗算したものが, 各台詞での語速値になっている.
データセット7と8では, 個々の文章ごとの語速値が表示語速値になるように制御した.	
音質調整	データセット1では, 個々の音ファイルについて, AR19 TR-832「デジタルテレビ放送番組におけるラウドネス運用規定」に従い平均ラウドネス値を算出したとき, -20 LKFS相当となるよう調整した.

図5 公開中のデータベースの録音条件と機材部分

に過去に収録したデータについても整理して、Web ページでデータベースの公開を行っている。ファイル形式は、サンプリング周波数が 44.1 kHz の 16 bit 分解能である。フォーマットは wav 形式を利用している。利用者が利用しやすいように、ダウンロード前に、ひとつひとつの短い音声を視聴することができるように配慮してある。またダウンロードする際には、効率よく作業できるように「この行をダウンロード」「この列をダウンロード」というボタンを設け、さらに「一括ダウンロード」のボタンも設けている。利用状況についても調査した。2016 年度の後半の時期において、2 か月で約 600 のアクセスがあり、音声ファイルは 70 本程度がダウンロードされていた。また、本データベースの素材を研究に利用しているというメール連絡や、教育的コンテンツの作成に利用して下さったという連絡もいただいており、本データベースが研究者の方々から有効に活用されていることを確認している

図5に公開中のデータベースのページの中で、録音条件と機材について説明した部分を掲載する。テキストについては、図6に示すリストより入手することができるようになっている。

## テキストのダウンロード

データセット1から2の音声データのテキストファイルはこちらでご覧になります。

原稿名	一文字遣い	カーナビ
テキストデータ	<a href="#">EUC-JP \ Shift-JIS \ UTF-8</a>	<a href="#">EUC-JP \ Shift-JIS \ UTF-8</a>
PDFデータ	<a href="#">一文字遣い文</a>	<a href="#">カーナビ文</a>

声優パート(データセット4~6)の音声データのテキストファイルはこちらでご覧になります。

原稿名	「雪」
テキストデータ	役名と台詞だけが並んだテキストデータ
ReCok5用データ	ReCok5で原稿表示する場合に使用するデータ(原稿+漢字や特殊文字のモーラ数+相対語速値)
印刷用カラー原稿	原稿の1ページ目のPDFファイル 原稿の2ページ目のPDFファイル

データセット3, 7, 8の音声データのテキストファイルはこちらでご覧になります。

原稿名	ATR25	HENSHU00	HENSHU01	HENSHU02
テキストデータ	<a href="#">ATR25文</a>	<a href="#">編集手帳_原稿00</a>	<a href="#">編集手帳_原稿01</a>	<a href="#">編集手帳_原稿02</a>
振り仮名付きPDFデータ	<a href="#">ATR25文</a>	<a href="#">編集手帳_原稿00</a>	<a href="#">編集手帳_原稿01</a>	<a href="#">編集手帳_原稿02</a>

図6 公開中のデータベースのテキストへのリンク部

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[学会発表](計2件)

近藤聡太, 高橋弘太: 時間周波数平面上の成分移動に基づいた新しい音信号混合法, 電子情報通信学会 音声研究会, 2016年10月27日, 静岡大学(静岡県・浜松市)

長谷川政良, 大脇渉, 高橋弘太: 騒音下における音声再生の識別性を重視したスマートミキサー, 日本音響学会 騒音・振動研究会, 2015年8月28日, 横浜国立大学サテライトキャンパス(神奈川県・横浜市)

[その他]

この研究費を使い、収録、編集して公開している音声データベース(略称: SRV-DB)は、以下のアドレスからアクセスすることで、研究者から学生まで全データを無料でダウンロードして自由に研究に役立てることが可能である。

<http://www.it.ice.uec.ac.jp/SRV-DB/>

## 6. 研究組織

(1) 研究代表者

高橋 弘太 (TAKAHASHI, Kota)

電気通信大学・情報理工学研究所・准教授  
研究者番号: 10188005