

## 科学研究費助成事業 研究成果報告書

令和 3 年 6 月 7 日現在

機関番号：17401

研究種目：基盤研究(C) (一般)

研究期間：2017～2020

課題番号：17K01568

研究課題名(和文)話者性と音韻性の音声分離モデルに基づく発話機能回復支援システムの構築

研究課題名(英文) Development of speech articulation recovery support system based on a phonological / individual separation model

研究代表者

上田 裕市 (UEDA, YUICHI)

熊本大学・大学院先端科学研究部(工)・教授

研究者番号：00141961

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)：ホルマント復元関数に基づく音韻性・個人性分離モデルを提案した。復元関数に含まれる個人性特徴に関する検証の結果、関数形状と話者性の相関から話者識別への応用が可能である。音声合成では、縦続型ホルマント合成方式に代えて、並列型とすることで、復元音声におけるホルマントレベル制御が可能となり、障害音声の特徴の一つ(こもり声)の明瞭化に寄与できる。さらに、h<sub>v</sub>構音空間に基づく音韻変換として規格化空間を介する異話者音声変換アルゴリズムを提案し、PCベースのリアルタイムシステムを構築してその効果を確認した。実用化のためには、規格化空間における健常化アルゴリズムの構成手法について検証する必要がある。

研究成果の学術的意義や社会的意義

ホルマント周波数に含まれる音韻性と話者性(話者の個人性)を分離するための新たな手法であり、ホルマント分析・統計過程で生成されるホルマント復元関数に着目することで、ホ音声合成における話者変換、音声変換、話者識別・照合などの様々な音声応用システムの基本特徴量として利用できると考えられる。また、発話障害者の劣化音声の健常化へ利用することで、障害者・健常者間のオーラルコミュニケーションに寄与できるが、そのためには、障害毎に異なると予想される健常化アルゴリズムの導入が課題として残されている。

研究成果の概要(英文)：We proposed a phonological / individual separation model based on the formant restoration function. As a result of verification of the personality characteristics included in the restoration function, it is possible to apply it to speaker identification from the correlation between the function shape and the speaker. In speech synthesis, by using a parallel type instead of the cascade formant synthesizer, it is possible to control the formant levels in the restored speech, which can contribute to the clarification of a characteristics of the impaired speech (muffled voice). Furthermore, we proposed a different speaker's voice conversion algorithm via a standardized space as a phoneme conversion based on the hv-articulation space, constructed a PC-based real-time system, and confirmed its effect. For more practical use, it is necessary to verify the construction method of the normalization algorithm in the standardized space.

研究分野：音声情報工学

キーワード：ホルマント周波数 音声合成 障害者音声 個人性 音韻性 発話訓練

## 様式 C-19、F-19-1、Z-19 (共通)

### 1. 研究開始当初の背景

音声には、意味的特徴を担う音韻性と個人的特徴である話者性に関する情報が融合して含まれている。音響音声学の立場からは、声帯振動数である音声ピッチや声道共振を特徴づけるホルマント周波数が主要な音声特徴量とされ、後者のホルマント空間は、話者の生理学的特徴である声道長に関する情報(話者性)を含みつつも、声道形状に対応する情報(音韻性)の表現に有用なものとして、音声合成や認識、あるいは発声訓練の現場で用いられてきた。話者毎に異なる領域サイズで定義されるホルマント空間で音韻性特徴を見出すために話者の声道長正規化手法が種々提案されてきた。一方、声道長とホルマント周波数の関係や声道形状とホルマント比の関係から、個人性特徴を見出す試みがなされてきた。我々は、これまでの研究において、リアルタイムホルマント抽出方式を用いた聾児の母音構音学習向けの色彩画像表現や声道形状表現などのシステム開発において、母音ホルマントの正規化問題を多く扱ってきた。その中で、構音障害者向け発話補助マイクロホンの開発として行った「ホルマント変形規則に基づく音声復元技術に関する研究」の中で、本申請の主要テーマであるホルマントベースの構音空間における母音音韻性と話者性の分離モデルの着想を得て、これまで予備的検証を行ってきた。

### 2. 研究の目的

本研究では、ホルマント空間をベースとした新たな正規化構音空間を構築し、そこから(音響的・統計的に困難視されている)「音韻性と話者性の分離抽出」手法を確立する。さらに、これらを選択的・融合的に組み合わせ、視覚パターンあるいは聴覚パターンとして発話者へフィードバックすることで、聴覚障害者や構音障害者の発話訓練や音声リハビリに有用な発話機能回復システムの開発と訓練プログラムの構築を目的とする。特に、「ホルマント縮退」により音韻性を表象する話者毎の正規化構音空間を推定し、この正規化空間から(逆に)実ホルマント空間復元の過程で得られるホルマント復元関数に含まれる話者性特徴を分離・関数化するものである。

### 3. 研究の方法

#### (1)ホルマント空間から正規化構音空間の構築:

多数話者群の音声試料を収集してデータベース(音声DB)を構築する。高精度ホルマント抽出手法(IFC手法)を用いて、試料中の母音区間から抽出されたホルマント周波数を用いて正規化構音空間を話者毎に推定する。

#### (2)「音韻性」モデルの構築と検証:

話者正規化の観点から、日本語5母音分布の分離度・同母音群内分布の局所化の傾向について検討し、話者性分離(話者正規化)後の音韻性空間の数理モデルを構築する。多数話者での音韻性モデルの汎化性と普遍性の検証が本課題での主テーマとなる。

#### (3)「話者性」モデルの構築と検証:

ホルマント空間に含まれる話者性(声道長や声道形状の個人差)に起因する特徴要素の分離手法を提案し、話者毎の特徴空間を構築する。多数話者について、それぞれの特徴空間の詳細な分析を行い、男女性・年齢層および声質による特徴空間類別の可能性を調べる。このことから、「話者性」における知覚的類別、解剖学的・生理学的類別および分離モデル(話者性特徴空間)の定量的類別の相関性を明確にする。

#### (4)話者性/個人性分離モデルに基づく発話支援システムの構築:PCベースのソフトウェア

### 4. 研究成果

#### (1)ホルマント空間から正規化構音空間の構築

音声試料として、男女12名の音声コーパスの母音フレームを使用する。母音フレームは、音声特徴推定エンジンを用いて音声コーパスから母音性(vow)を抽出し(vow $\geq$ 0.95)となるフレームを母音区間として使用した。図1に男性1名(黒点)・女性1名(赤点)の実ホルマント空間でのF1-F2分布とhv正規化構音空間での分布例を示す。F1-F2分布では、明らかに話者特徴である性差としての声道長の違いがホルマント周波数の絶対的な違い(相似形)として現れるのに対して、hv正規化構音空間では両者の重複が見られ、性差がかなり解消されている、即ち話者正規化の効果が顕著に現れていることが分かる。

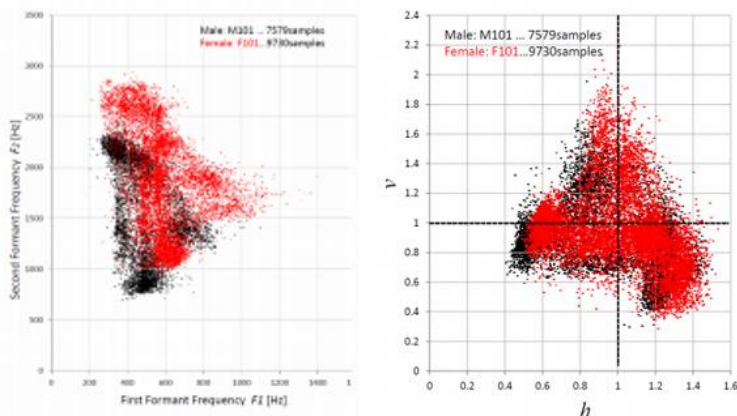


図1. 実ホルマント空間(F1-F2分布)とhv正規化構音空間(例:男性M1と女性F1)

更に、復元関数  $g(h, v)$  を用いてホルマント復元処理を行った例を図2に示す。図2 (a) は MIAES (男性話者 M1 の朗読音声 AES) の復元関数曲面である。図2 (b) は (a) の MIAES の復元関数を用いて復元されたホルマント周波数の例である。横軸が復元対象音声のホルマント周波数、縦軸が復元されたホルマント周波数を示しており、ホルマント復元で対応する箇所に F1~F3 の点をプロットする。復元誤差率が  $\pm 10\%$  の領域に概ねプロットされている。図2 (c) は、各条件下 (話者毎 M1~M6、F1~F6・音声試料毎 ATR/KYO/AES) における平均復元誤差率である。同一話者における文脈 closed/open test では、ホルマント復元関数については、関数作成元の話者と復元対象音声の話者が同一であれば、話者や文脈によらず 5% 程度の平均復元誤差率でホルマント復元が可能であることがわかった。また、性別 closed/open test では関数作成元の話者と復元対象音声の話者が同性であれば 10% 以下、異性では 20% 前後の平均復元誤差率となることがわかった。

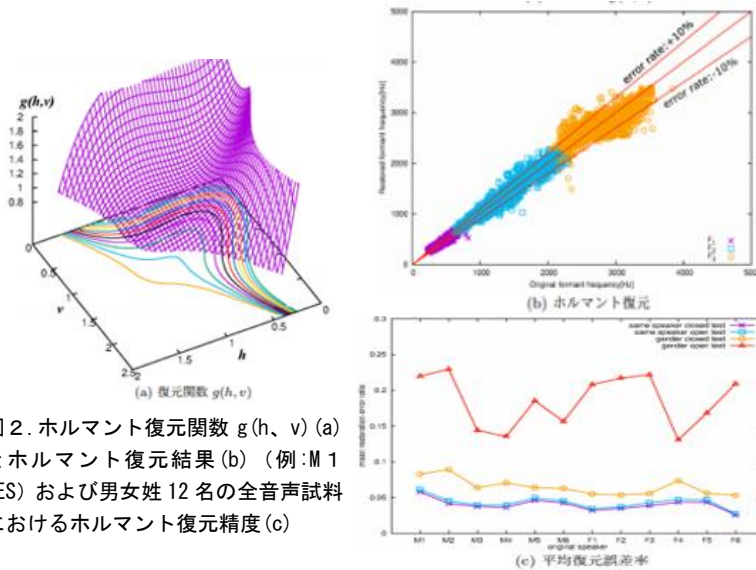


図2. ホルマント復元関数  $g(h, v)$  (a) とホルマント復元結果 (b) (例: M1 AES) および男女姓 12 名の全音声試料におけるホルマント復元精度 (c)

### (2) 「音韻性」モデルの構築と検証

図3 (a) に、日本語 5 母音を男女各 75 名が 2 回発声した音声試料 (VOW) の  $h, v$  分布を示す。図中の青点は男性の  $h, v$  分布、赤点は女性の  $h, v$  分布を示している。図では、各母音が明瞭に発話されており、各母音に対応する領域に分布していることがわかる。男女間においても同様の母音領域に分布しており、性別による違いも見られないことから話者性特徴の正規化効果を確認することができる。また、本研究での音声試料群ごとに全話者分を  $h, v$  平面に写像した結果を図3 (b)~(d) に示す。+ 点 ((b)~(d)) は各話者の  $h, v$  分布の重心を示している。各音声群・各話者の重心 (+ 点) は  $h, v$  平面上の中性母音点 (1, 1) の近傍に分布しており、音声試料に含まれている各音素のバランスがとれていることがわかる。また、ホルマント空間において話者ごとに異なる声道長の違いを直接反映する中性母音構音について、 $h, v$  平面上では中性母音点 (1, 1) に収束し話者性を正規化できていることから、ホルマント空間における音韻性特徴を  $h, v$  分布として分離できることを示唆している。単語音声・文音声・朗読音声の場合は単母音とは異なり、発話において、調音結合により各母音間を遷移するため、母音三角形の内側にまんべんなく分布している。これは、男女間においても性差による分布の違いはなく、ほぼ同様の領域に分布している。これらの傾向は図3 (c)、(d) においてもみられることから、 $h, v$  平面が音韻性特徴の表現に適していることがわかる。このことから、 $h, v$  正規化構音空間 ( $h, v$  平面) はホルマント空間から話者の声道長という話者性特徴を正規化した空間であり、実ホルマント空間から音韻性特徴を分離・表現した母音「音韻性」モデル空間と見なすことができる。

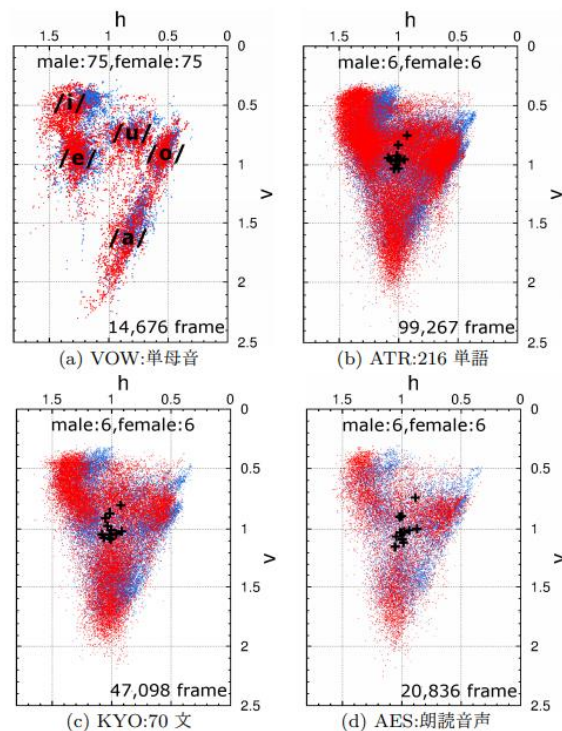


図3.  $h, v$  正規化構音空間における音韻性モデル

### (3) 「話者性」モデルの構築と検証

表1に音声試料 AES (朗読音声) のホルマント周波数群を用いて、話者別に推定された復元関数の係数  $a_0 \sim a_9$  の数値を示す。この関数の違い、即ち図2 (a) の  $h, v$  平面上で定義される 3 次元的な復元曲面形状の違いが「話者性」を特徴づけるパラメータであると考えられる。



(a) 【話者性】としての声道長比の推定

いま、話者毎に表1の復元関数(係数)を用いて計算したホルマントスケーリングファクタ(FSFR)を求めた結果を図4に示す。図より、男性群と女性群で、FSFR値に大きな違いが見られる。すなわち、FSFR値の示す(標準的な成人男性に対する)ホルマント周波数の比に逆比例する声道長の違いの傾向を見ることができる。

更に、FSFR値の妥当性を確認するために、話者群を児童にまで拡張してFSFR値推定実験を行った。用いた音声試料は、子供の声データベース(CIAIR-VCV)である。これは、6~12歳の男児135名、女児132名の計268名を対象に童話から引用した文章を室内(通常的生活環境下)で収集したものである。

図6に年齢ごとに平均した児童のFSFR値及び図5を男性群と女性群で平均したFSFR値を示す。図より、児童の年齢が上がるにつれてFSFR値の下降、原理的には声道長の増加が確認でき、成人男女(M1~M6、F1~F6)の平均FSFR値に漸近するという生理学的な成長過程(個人性)をこの傾向から類推することができる。

表1. 各話者の復元関数係数値(音声:朗読音声AES)

話者	$a_0$	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$	$a_7$	$a_8$	$a_9$
M1	1.909	-1.926	-0.923	2.751	1.486	-1.310	-0.931	-0.666	-0.105	0.674
M2	3.818	-8.142	-1.696	8.988	1.340	0.588	-3.360	-0.101	0.374	-0.856
M3	1.840	-3.748	0.586	4.759	0.022	-0.841	-1.608	-0.039	-0.100	0.142
M4	1.463	-2.428	0.681	3.670	0.172	-1.573	-1.447	-0.101	0.611	-0.126
M5	3.031	-5.055	-1.561	4.176	0.174	2.792	-0.943	-0.085	-1.722	0.152
M6	1.730	-2.640	0.091	4.791	1.271	-3.107	-2.197	-0.302	1.474	-0.098
F1	1.943	-1.881	-0.473	1.538	0.393	0.574	-0.287	-0.201	-0.526	0.144
F2	1.804	-1.242	-0.641	1.566	0.843	-0.410	-0.600	-0.131	0.353	-0.321
F3	2.400	-2.999	-1.019	2.577	0.852	1.137	-0.467	-0.496	-1.244	0.495
F4	3.420	-7.403	-0.196	7.750	0.313	0.005	-2.483	-0.053	-0.147	-0.088
F5	3.167	-5.849	0.667	5.535	0.193	1.135	-1.502	-0.089	-0.089	0.136
F6	2.232	-2.703	-0.713	2.225	0.125	1.512	-0.469	-0.025	-0.934	-0.031

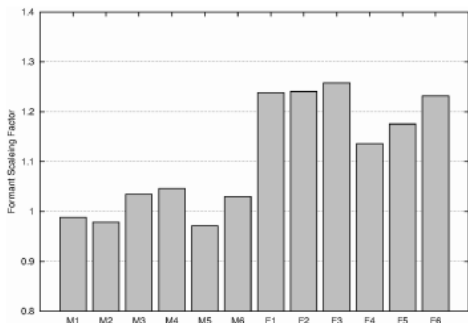


図4. 男女各6名の復元関数より計算されるFSFR値

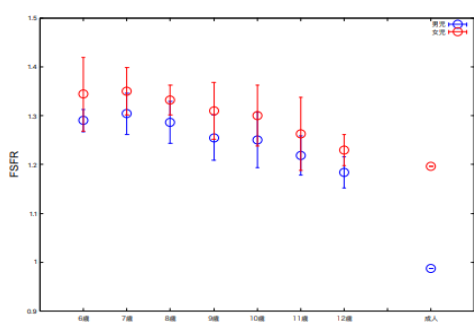


図5. 男女児において推定されるFSFR値及び成人男女群の復元関数より計算される平均的なFSFR値

(b) 【話者性】としての復元関数曲面の特徴

図6(話者固定:M4)、図7(音声試料固定AES:話者3名)は、各音声試料群から算出される復元関数 $g(h, v)$ を $hv$ 平面上での等高線分布( $g(h, v)$ 値)として描いたものである。図7より、話者が同じであれば、異なる音声試料群でもそこから推定されるホルマント復元関数は類似している。それに対して、図8より、同じ音声試料群でも、話者が変われば、復元関数形状が異なる(個人性特徴が現れる)傾向として観測することができる。すなわち、これらの等高線分布形状そのものに、話者の個人性特徴が含まれていると解釈できる。

復元関数の持つ話者性特徴を確認するために話者識別実験を行った。結果として、男性6名・女性6名の計12名の話者について、各音声試料の抽出ホルマント周波数と各復元関数で計算される復元ホルマント周波数の復元誤差の評価から、第2候補話者まで認めれば、平均80%の話者識別が可能であった。

これらの結果から、表1の係数群で定義されるホルマント復元関数(～図7の復元関数曲面の等高線分布)は、各話者の個人性(話者性)を反映する特徴を併せ持つことが示された。

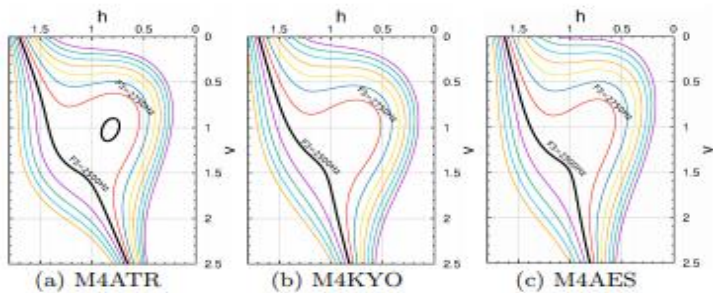


図6. 異なる音声群での同一話者の復元関数の等高線分布(男性M4の例)

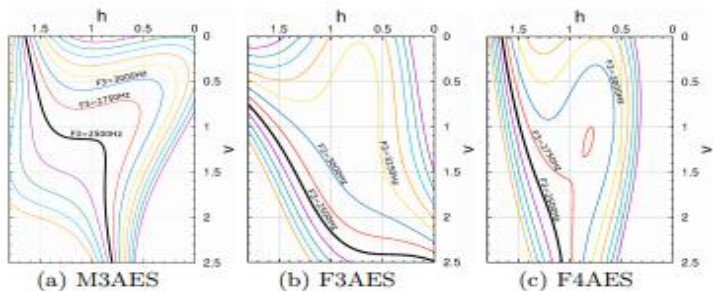


図7. 朗読音声AESの復元関数の等高線分布(男性M4・女性F3, F4の例)

#### (4) 話者性/個人性分離モデルに基づく発話支援システムの構築

障害音声において、話者性を損なうことなく母音の音韻性改善を行うことを目的とした発話支援システムは図8のように分析部・健常化部・合成部の3つの基本ブロックで構成される。

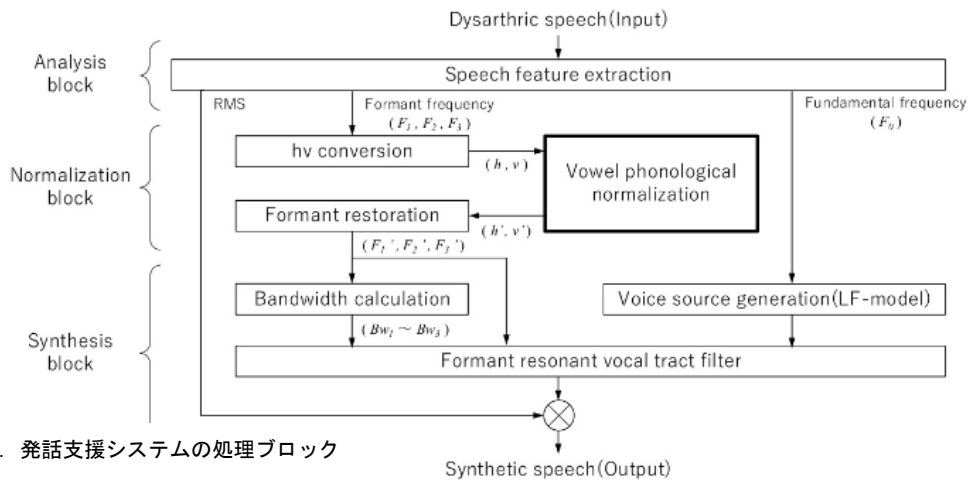


図8. 発話支援システムの処理ブロック

図9は、健常部での音韻性制御手法を合成母音の音韻性制御において確認するために作成したWindowsベースのリアルタイム音声変換ツールである。図左と図右のh v平面が、それぞれ原話者構音空間（障害音声）と目標話者構音空間（音韻性健常化音声）、中央が日本語5母音をレーダーチャート状に等角度で配置した音韻性規格化空間である。図左のマーク（□）が当該障害話者の通常の5母音位置、図右のマークが目標話者の5母音位置を示す。図右の目標話者は（テスト用として）研究に用いた男女性12名の復元関数を任意に選択できるように設計している。

図左のh v平面上のカーソル（+）で示される構音点が規格化空間を介して、図左のカーソルに写像されて、このh v座標を選択された目標話者のホルマント復元関数を用いて実ホルマント周波数に復元して、ホルマント合成方式による合成音声を作成、リアルタイム音声出力するものである。図では、障害音声[a]が目標音声[a]の位置に変換されている。実際の音韻性改善では、規格化空間処理において健常化アルゴリズムを適用して音声変換を行う。

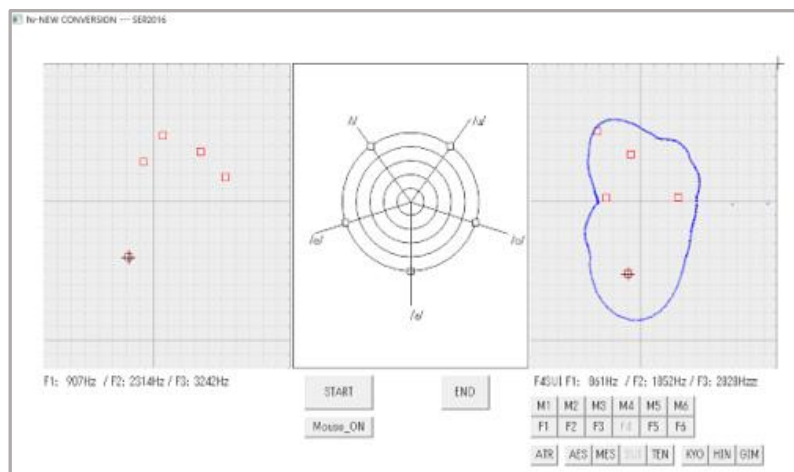


図9. 規格化構音空間を介した障害音声から健常音声へのホルマント変換ツール

図10に、発話支援（母音発声）のために開発したプロトタイプシステムの実行面を示す。ただし、障害の種類や程度毎に、個別に設定する必要のある規格化空間での健常化アルゴリズム（規則）が未完成のため、現状では、左下のカーソルを使ってマニュアルで、実験的にh v座標に変位を与えて、ホルマント合成音声（赤マーカー）音韻性改善の効果を試聴できる状況である。

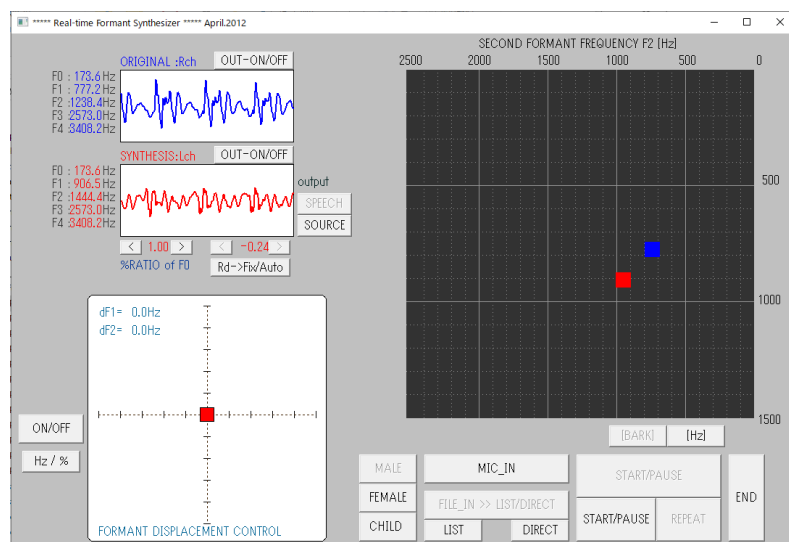


図10. 母音音声のリアルタイムホルマント健常化ツール

5. 主な発表論文等

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

1. 著者名 Tadashi Sakata, Naomitsu Ikeda, Yuichi Ueda, Akira Watanabe	4. 巻 29
2. 論文標題 Vocal Tract Length Estimation Using Accumulated Means of Formants and Its Effects on Speaker-Normalization	5. 発行年 2021年
3. 雑誌名 IEEE/ACM Transactions on Audio, Speech, and Language Processing	6. 最初と最後の頁 1049-1064
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/TASLP.2021.3060172	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12件(うち招待講演 0件/うち国際学会 2件)

1. 発表者名 Yukinori Hetsugi, Tadashi Sakata, Yuichi Ueda
2. 発表標題 A Phonological Control Method on A Speech Compensation System for Dysarthria Using A Standardized Space
3. 学会等名 ICIIBMS 2020 (国際学会)
4. 発表年 2020年

1. 発表者名 戸次 幸徳, 坂田 聡, 上田 裕市
2. 発表標題 障害音声健常化のための母音音韻性制御手法に関する検討
3. 学会等名 日本音響学会・春季研究発表会
4. 発表年 2020年

1. 発表者名 岸川義和, 上栗裕平, 坂田聡, 手塚征宏, 中村典史, 上田裕市
2. 発表標題 ディープニューラルネットワークによる口蓋裂患児音声の構音特徴分析
3. 学会等名 日本音響学会九州支部・第14回学生のための研究発表会
4. 発表年 2020年

1. 発表者名 戸次 幸徳・坂田 聡・上田 裕市
2. 発表標題 障害音声健常化のための母音音韻性制御手法に関する検討
3. 学会等名 日本音響学会春季研究発表会講演論文集
4. 発表年 2020年

1. 発表者名 荒川大貴, 岸川義和・坂田聡・手塚征宏・中村典史・上田裕市
2. 発表標題 口蓋裂患児音声の構音特徴空間とその客観的・主観的評価に関する検討
3. 学会等名 日本音響学会九州支部・学生のための研究発表会予稿集
4. 発表年 2019年

1. 発表者名 木下天介・坂田聡・渡邊亮・上田裕市
2. 発表標題 スペクトル圧縮関数の最適化に基づく補聴処理に関する検討
3. 学会等名 日本音響学会九州支部・学生のための研究発表会予稿集
4. 発表年 2019年

1. 発表者名 Tadashi Sakata, Yuichi Ueda, Akira Watanabe
2. 発表標題 Vocal tract length estimation from the long-term average formant frequencies
3. 学会等名 13th Western Pacific Acoustics Conference (国際学会)
4. 発表年 2018年

1. 発表者名 宮崎天斗, 戸次幸徳, 坂田聡, 上田裕市
2. 発表標題 並列型声道共振フィルタにおけるホルマントレベル制御に関する研究
3. 学会等名 電気・情報関係学会・平成30年度九州支部連合大会
4. 発表年 2018年

1. 発表者名 戸次幸徳, 坂田聡, 上田裕市
2. 発表標題 正規化構音空間からのホルマント復元における話者性特徴抽出に関する検討
3. 学会等名 電子情報通信学会・音声研究会
4. 発表年 2018年

1. 発表者名 戸次幸徳, 坂田聡, 上田裕市
2. 発表標題 母音ホルマント空間における話者性と音韻性の特徴分離
3. 学会等名 日本音響学会・2018年度秋季研究発表会
4. 発表年 2018年

1. 発表者名 日高将臣・宮崎天斗・戸次幸野徳・坂田聡・上田裕市
2. 発表標題 構音マップ変換による母音ホルマントの健常化
3. 学会等名 電子情報通信学会九州支部・学生会講演会, A-24
4. 発表年 2017年



1. 発表者名 中武佑太・上田雄大・坂田聡・上田裕市
2. 発表標題 正規化構音空間からの声道形状マッピング手法
3. 学会等名 電子情報通信学会九州支部・学生会講演会, D-25
4. 発表年 2017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
研究分担者	坂田 聡  (SAKATA Tadashi)  (80336205)	熊本大学・大学院先端科学研究部(工)・助教    (17401)	

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
---------	---------