

# ホルマントベースのスペクトルモーフィングによる劣化音声復元手法の開発

代表研究者 上 田 裕 市 熊本大学大学院自然科学研究科教授  
 共同研究者 坂 田 聡 熊本大学大学院自然科学研究科助教

## 1 はじめに

喉頭摘出・声帯麻痺などの声帯機能不全や発話器官の変形・運動性低下などの構音機能不全による発声障害を持つ人たちは発話音声の不明瞭さのために日常会話に多くの支障をきたしている。このような劣化音声であっても、多くの場合、その音声信号には低明瞭度ではあるが、発話者の意図する音声特徴はわずかながら残存していることから、これを抽出・補間して通常音声に近い音声に復元できれば、正常な音声コミュニケーション手段の回復にもつながる。本研究では、そのような症例ごとに異なる発声特性や声の質に合わせて柔軟な声質改善が期待できる劣化音声復元手法の確立を目的とする。具体的には、音声の韻律情報と音韻情報を担う音声特徴量としての声帯振動周波数であるピッチ（音声ピッチ）と声道の共振周波数に対応するホルマント周波数を高精度でリアルタイム推定し、そこから音声劣化特徴を顕在化させる。その特徴を正常化するための補正処理を施して、総合的な音声特徴量として再合成処理を行うものである。まず、その利用目的上、処理のリアルタイム性は必須であることから、上記の分析システムのリアルタイム処理を実現する。さらに、再合成方式として、ホルマントベースのスペクトルモーフィング<sup>[1]</sup>に基づく音声合成方式および標準的なソース・フィルタモデルであるホルマント分析合成方式<sup>[2]</sup>との併用を検討する。

以下、2章で新たに開発した汎用PC上でリアルタイム処理が可能な音声特徴ベクトル推定エンジンの構成と劣化音声対象とした処理例として、健常者の囁き母音と声帯疾患者の嗄声の分析を行い、それぞれの音声特徴を忠実に推定・抽出し得ることを示す。3章では、劣化音声の音声特徴量の補正例について示し、再合成方式としてホルマント合成方式を適用した例を示す。4章では、再合成音声の個人性の保持が期待できるスペクトルモーフィング方式の原理を説明し、話者変換原理に基づく劣化音声復元処理例を示す。最後に、5章において可搬型音声復元システム化に向けた課題を含め、今後の展開について述べる。

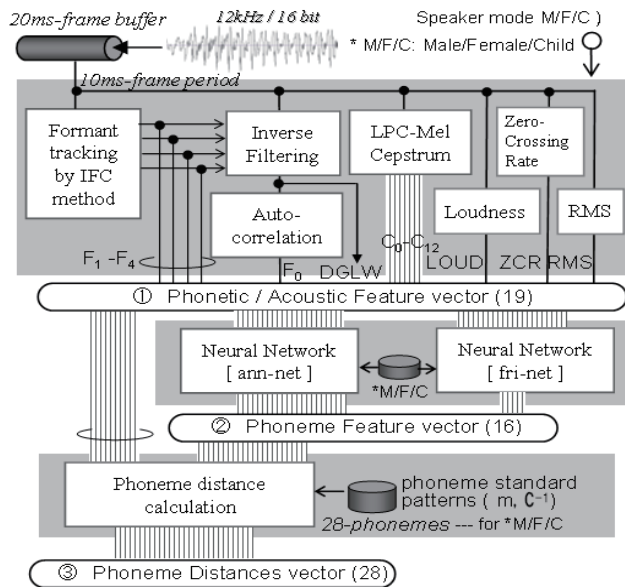
## 2 音声特徴推定エンジンの構築

### 2-1 音声特徴ベクトル推定エンジン

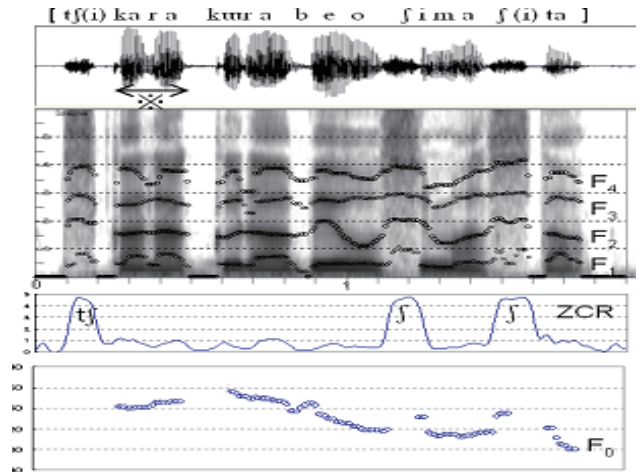
筆者らがこれまでに開発してきた音声分析処理ツール群を統合化して、音声・音響特徴量～音素特徴・音素距離成分からなる複合パラメータ（音声特徴ベクトル）を汎用PC上でリアルタイムで推定する分析処理ツールを開発した。

#### (1) 推定エンジンの構成

図1(a)に開発ツールのブロック図を示す。システムは、音声情報に関する3つの階層から成る。最下層は音響・音声特徴量、第2層は音素的特徴量、第3層は日本語各音素に対する音素距離量をそれぞれ抽出する機能を持つ。本ツールの主要部（本研究テーマでの主要な音声特徴）となるのは最下層の特徴群であり、音韻情報を担うホルマント周波数（ $F_1 \sim F_4$ ；逆フィルタ制御法<sup>[3]</sup>）、日本語アクセントやイントネーションを決定する有声音の基本周波数  $F_0$ （音声ピッチ）、声の大きさに対応する実効値（RMS）やラウドネスレベル（LOUD）である。各パラメータは、12kHz サンプリングの入力信号から10ms フレーム周期にて抽出される。基本周波数については、抽出ホルマント周波数を零周波数に持つ逆フィルタを用いたフィルタリング処理により声帯音源微分波信号を得て、その自己相関関数のピーク位置により抽出している。声道の共振特性を除去した信号の自己相関処理であることから高精度のピッチ推定が期待できる。また、 $F_4$ までの高次ホルマント周波数までの高速・高精度推定処理は、すでに筆者らの報告によりその実時間性と有効性は健常音声に対して確認されている。なお、本研究報告では、母音を対象としていることから最下層の特徴量を用いているが、今後の研究展開によっては、更に音素特徴量や音素距離の利用も可能である。



(a) 音声特徴ベクトル推定エンジン



(b) 音声特徴抽出例～男声「力比べをしました」

上段：ホルマント周波数軌跡 ( $F_1 \sim F_4$ )

中段：零交差周波数

下段：ピッチ軌跡 (基本周波数  $F_0$ )

図1 音声分析処理部 (特徴ベクトル推定エンジン) の構成と抽出処理例

(2) 音声特徴推定例 (健常者による通常音声の例)

推定エンジンで抽出した特徴量の例を図1 (b)に示す。音声は男性話者の文音声「力くらべをしました」である。

2-2 劣化音声の音声特徴推定

劣化音声例として2種類の音声について検討を行った。

①健常者の囁き母音

②声帯疾患患者音声 (成人男性の嘎声)

発声は、日本語5母音の単母音と連続母音 [aiueo] である。

1) 聴覚的印象

ともに、母音音韻性は残しているが、韻律的特徴として、

①囁き母音: 顕著な氣息性により無声化しパワーも小さい。

②嘎声: 症状に応じて、次のような特徴が知覚される。

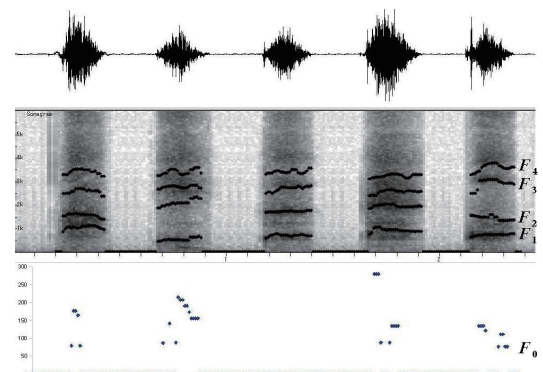
- A) 粗ざう性 (R): 不規則な声帯振動によるガラガラ声
- B) 努力性 (S): 発声器官の異常な緊張によるりきみ声
- C) 氣息性 (B): 声帯の不完全閉鎖による息漏れ声
- D) 無力性 (A): 呼気がうまく出せないため弱々しい声

2) 音響的特徴

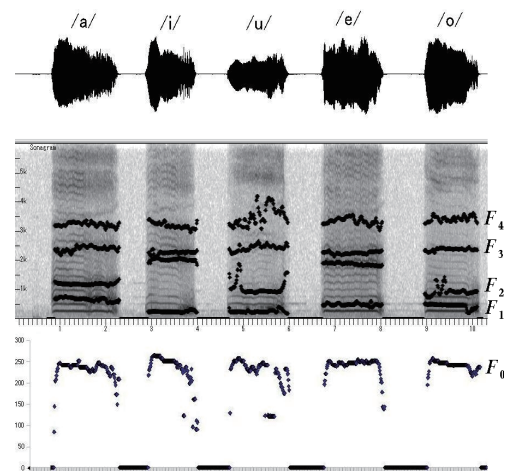
声道共鳴であるホルマント周波数は残存するが、有声音特徴であるスペクトル高調波成分が消失 (囁き母音)、もしくは弱くて、非常に不安定 (嘎声) である。

各劣化音声について、推定エンジンで抽出したホルマント軌跡とピッチ軌跡を図2に示す。

嘎声では、ピッチ、ホルマント共に変動が大きく、聴覚的印象を裏付けている。囁き母音では、ホルマント周波数は安定しているが、ピッチは抽出できず、ほぼ完全に無声化していることが確認できる。しかし、母音の音韻性を決定する第1、第2ホルマント周波数 ( $F_1, F_2$ ) については、その相対的な関係を維持しつつ、平均的には通常音声に比べて妥当な値となっている。これより音韻性の残存が確認できる。



(a) ①健常者の囁き母音



(b) ②声帯疾患患者の嘎声 (努力性)

図2 劣化音声の音声特徴抽出例

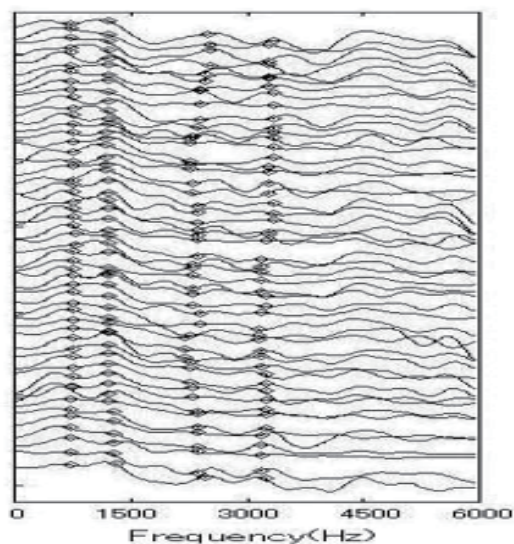
これらのパラメータ変動が、聴覚的あるいは音響的な劣化として、本質的なものであるか、あるいは不安定な音声信号に起因する分析誤りによるものなのか、すなわち本研究で用いている処理系（特徴推定エンジン）の入力音声に対するロバスト性を確認することは研究の成否を定める上で重要な要素である。このことから、実験に用いたいくつかの音声試料について、以下の検討を行った。

i) ホルマント周波数変動

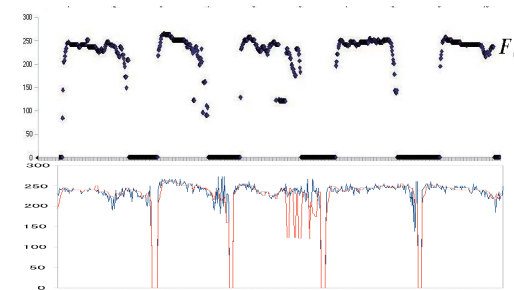
フレームごとの短時間スペクトルについて、Cepstrum 平滑化スペクトル（0次～20次）系列と抽出ホルマント軌跡を図3（a）に示す。ホルマントに対応するスペクトルピークは確認できるが、健常音声に比べるとフレーム毎の変動が大きく、抽出ホルマント軌跡の変動とよく対応している。このことから、ここで用いている逆フィルタ制御法（IFC法）<sup>[3]</sup>によるホルマント推定手法は劣化した音声にも適用できることが確認できる。

ii) ピッチ変動

各母音について、波形視察によりピークピッキングを行い、ピッチ周期をさだめ、その逆数としての瞬時基本周波数（ピッチ）を算出した。図3（b）に結果を示す。抽出ピッチと視察ピッチの変動がよく対応しており、劣化音声に対するピッチ推定も信頼できるものとなっていることが確認される。



(a) 母音[a]の抽出ホルマント軌跡(O印)と平滑化スペクトル系列



(b) 抽出ピッチ軌跡(橙)と波形視察系列(青)

図3 嘔声(努力性)の特徴抽出値と実測値との比較(図2(b))

### 3 音声特徴量の平滑化処理とホルマント分析合成方式による復元処理

#### 3-1 パラメータ平滑化による補正処理

##### (1) MA フィルタリング

ここでは、抽出された劣化音声のパラメータ軌跡で確認された変動成分を平滑化するために3点ハミング窓を用いたMA（移動平均）フィルタリングを適用した。図4（a）、（b）に、嘔声（努力性：図2の例）の各パラメータ軌跡の平滑化結果を示す。図よりわかるように、大局的な変動を吸収することはできないが、知覚上耳障りな細かな微細変動による調音の変動は改善できると考えられる。

##### (2) ホルマント分析合成方式による再合成

音声特徴量（ホルマント軌跡、ピッチ軌跡など）から直接音声を合成できるホルマント分析合成方式により、パラメータ平滑後の音声再合成を試みた。図5に合成器のブロック図を示す。声帯音源波信号はLFモデル波形を使用し、カスケード接続されたホルマント共振回路を駆動するターミナル・アナログ方式音声合成方式として知られている。

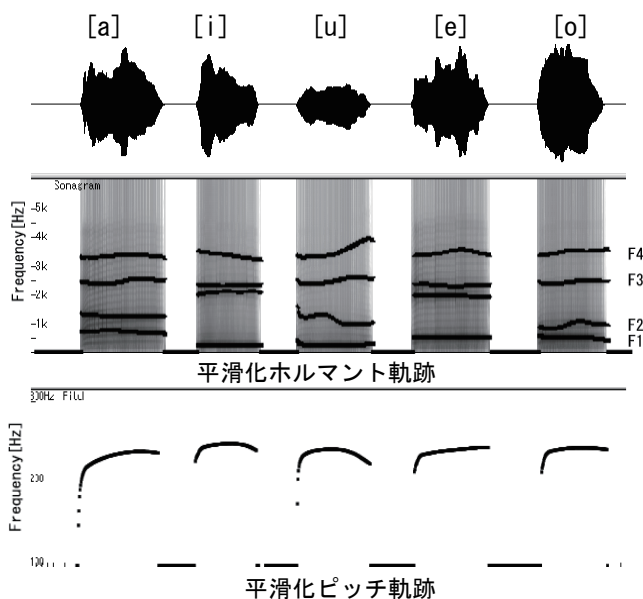


図4 劣化音声(図2の嘔声)特徴量の平滑化補正

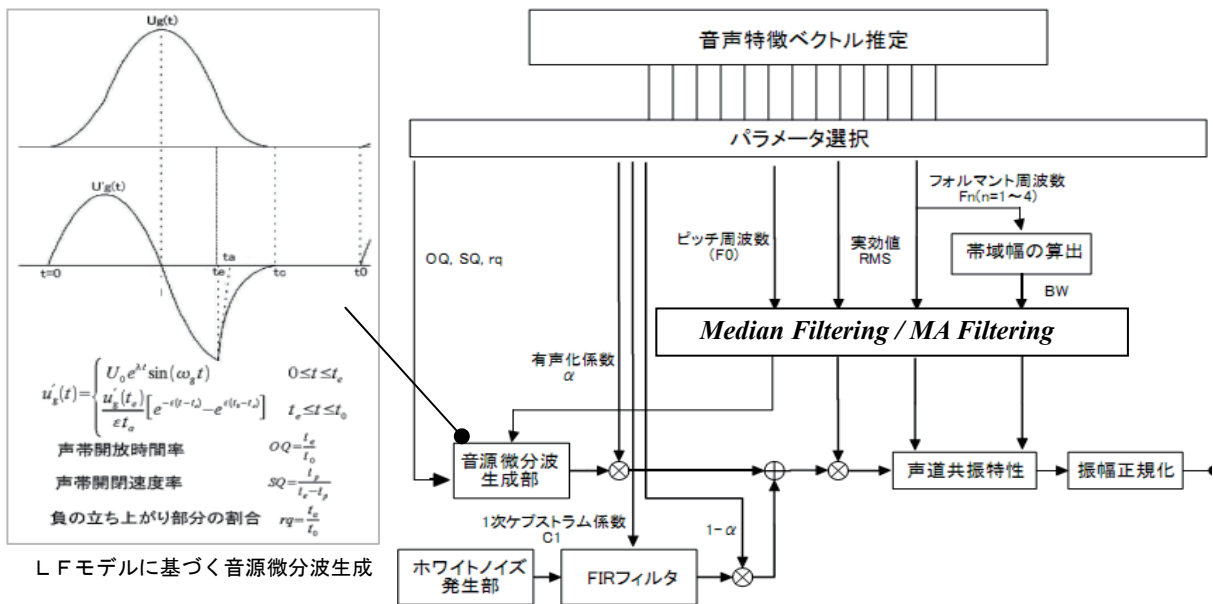


図5 ホルマント分析・合成型音声再合成方式

(3) 音声の安定性に関する聴覚評価実験

パラメータ補正後の再合成音声の評価するために、以下の聴取実験を行った。

[実験条件]

- ・音声試料：4症状の嘆声[a]（粗ぞう性/気息性/無力性/努力性の各症状）の原音声と再合成音声
- ・実験環境：静かな部屋でヘッドホンの両耳提示
- ・呈示方法：各症状の音声
- ・評価方法：音声の安定性に関する5段階評価（1:原音声と同程度～5:通常音声と同程度）
- ・被験者：男子大学生4名

[実験結果]

図6に、各症状について被験者4名の平均評価値を示す。粗ぞう性や努力性に比べて、気息性や無力性の評価値が小さくなっている。これは、図7の例（努力性と気息性の各母音のホルマント分布）に示すように、日本語5母音のホルマント分布の特徴であるF1-F2平面上での相対配置（母音五角形）が残存している音声（努力性）では、その変動性の平滑化により劣化度の改善は期待できるが、相対配置が歪んでいる場合（気息性）には、平滑化補正では比較的改善度が低いことが推測される。

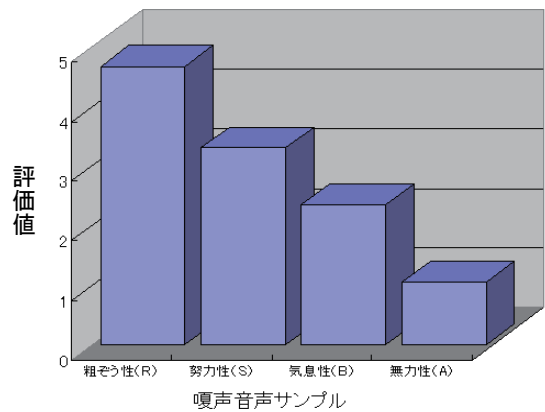


図6 復元音声の聴取評価結果

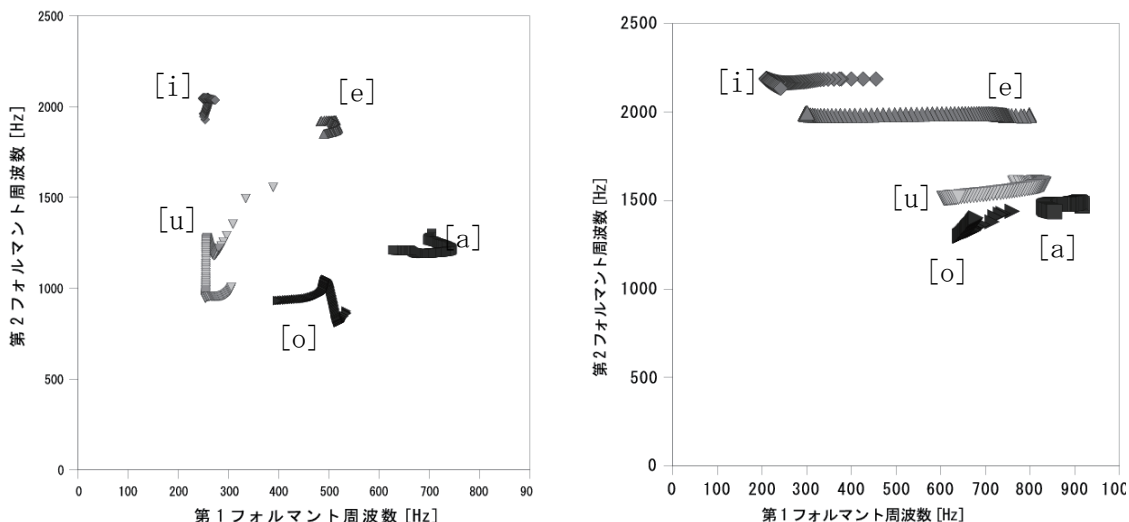


図7 平滑処理後のホルマント軌跡例 (左：嘆声(努力性) / 右：嘆声(気息性))

## 4 ホルマントベースのスペクトルモーフィングによる復元処理

### 4-1 スペクトルモーフィング方式<sup>[1][4]</sup>

ホルマント空間での5母音の相対配置の改善を行うには、その目標配置としての標準的な5母音配置を想定する必要がある。本研究では、そのような指定ホルマント位置で連続音声（母音）を再合成するために、スペクトルモーフィング合成方式[1]を適用する。

#### (1) 合成方式

本合成方式は、基本的には、Source-Filter モデルによる合成であるが、有声音源には周期的インパルス列、Filter には、声帯音源特性と声道特性を含む最小位相インパルス係数<sup>[5]</sup>を用い、FIR フィルタリング処理によりフィルタ応答としての合成音声を得るものである。特に、FIR フィルタ係数をフレーム周期で更新することで連続母音を合成できる。さらに、ホルマント軌跡の連続性を保持しながら連続的なスペクトル遷移を実現するために、正規化母音スペクトルという概念を導入している。

#### (2) スペクトルモーフィング原理

図8に、本モーフィング処理の概念を示す。図8(a)左が合成すべき目標母音のスペクトル(Cepstrum 平滑化スペクトル)とする。周波数軸の非線形伸縮により、ホルマント位置(F1~F3)を周波数軸上で等間隔に配置した正規化母音スペクトル(同図(a):右)を生成後、データベースとして格納しておく。合成においては、隣接する母音間の補間率(r)に従って、正規化空間上にて2母音の正規化スペクトルを補間する。同図(b):右は、正規化スペクトル空間での母音[i]から[e]への連続遷移を表している。さらに、得られた補間スペクトル系列を原スペクトル空間に戻すことで、同図(b):左のようにホルマント軌跡の連続性を保持するスペクトル系列が得られる。これらを各フレーム処理において最小位相インパルス応答(FIR フィルタ係数)に変換して、周期インパルス列と畳み込むことで音声信号を得ている(同図(c))。

#### (3) スペクトル・モーフィングの実際例(スペクトル補間処理)

あるホルマント(F<sub>1</sub>-F<sub>3</sub>)を持つ母音スペクトルを既知の5母音正規化スペクトル(データベース)から補間推定することを考える。推定の流れを図9に従って、以下に示す。

##### (i) 母音特徴空間(ホルマント空間)での任意点xの指定(図9(a))

図8(a)の点xが目標とするホルマント周波数を持つ補間スペクトル位置とする。各母音位置および重心点(中性母音)の各点にそれぞれの正規化母音が配置されているものとする。

##### (ii) 補間点ベクトルxの算出(図9(b))

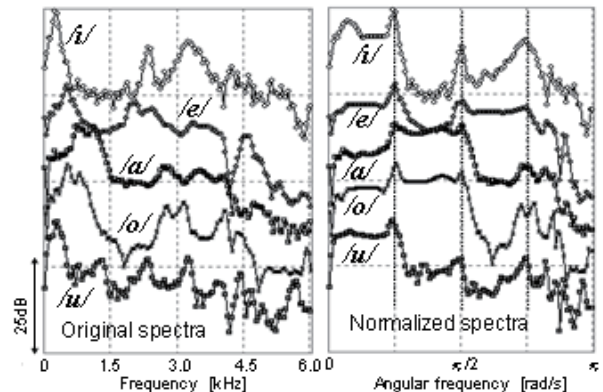
母音隣接空間(図では、/e/と/a/ではさまれた領域)での補間ベクトルの関係から重み(a<sub>1</sub>, a<sub>2</sub>)を求める。

##### (iii) 補間比(R<sub>a</sub>, R<sub>b</sub>)から正規化補間スペクトルS<sub>x</sub>を求める(図9(c))

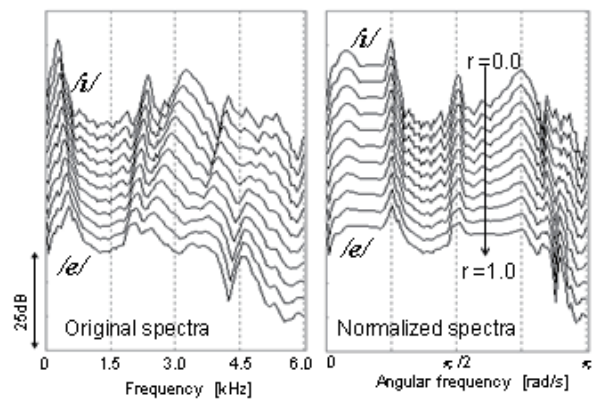
図のベクトル演算に従って、隣接母音補間比R<sub>a</sub>とさらに、原点の中性母音[N]との補間比R<sub>b</sub>を算出する。この補間比を用いて、隣接母音と重心母音の各正規化スペクトルの補間により、目標とする点での正規化スペクトルS<sub>x</sub>を求める。

##### (iv) 原スペクトル空間への逆写像(図9(d)):赤~原ホルマント軌跡、緑~目標軌跡

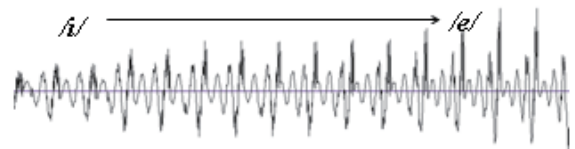
図中、緑の各点に対応するホルマント値を与えて、正規化スペクトルの非線形変換により原スペクトル



(a) 母音スペクトル(左:原 / 右:正規化)



(b) スペクトル補間例(左:原 / 右:正規化)



(c) 補間スペクトル列による合成音声例

図8 正規化スペクトルモーフィング合成処理

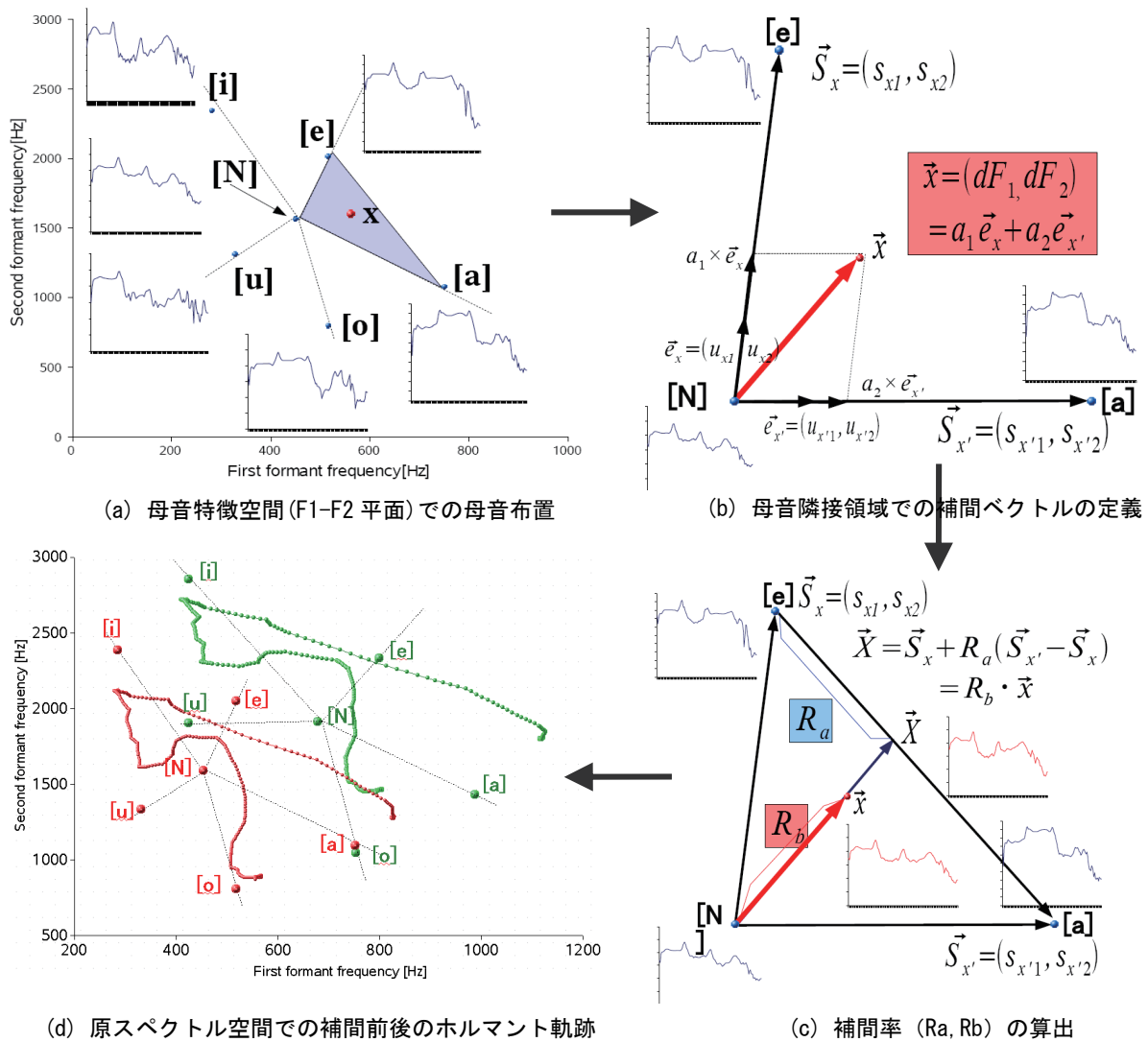


図9 スペクトルモーフィング手法による補間率算出とスペクトル補間の流れ

空間に逆写像して原スペクトル空間での補間スペクトルを得る。図(d)は原話者の連続母音[aiueo]のホルマント軌跡の連続性を保持したまま、異なる母音ホルマント周波数を持つ別話者の目標ホルマント軌跡に従う連続母音へモーフィングした例を示している。

#### 4-2 モーフィング合成手法の評価実験

スペクトルモーフィング手法による再合成音声について、3章で述べた（ホルマント軌跡とピッチの平滑化補正処理に基づく）ホルマント分析合成手法による再合成音声との比較において、その効果を確認するために以下の聴取実験を行った。  
[実験条件]

- 音声試料：3症状の嗄声[a]（粗ざう性/氣息性/努力性の各症状）のホルマント分析合成音声とスペクトルモーフィング再合成音声。図10に嗄声（氣息性）におけるモーフィング合成音声例を示す。なお、ここでは、モーフィング目標話者は、

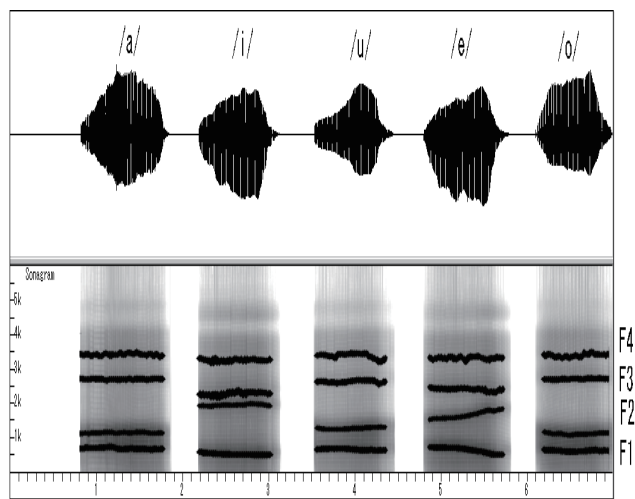


図10 スペクトルモーフィング手法による嗄声（氣息性）5母音の復元音声

健常男性話者のホルマント布置と正規化母音スペクトルを用い、原話者である声帯疾患者音声パラメータは、各発声母音から抽出したピッチ軌跡とホルマント軌跡（いずれも、MA フィルタリング平滑化処理後にモーフィング適用）である。

- ・実験環境：静かな部屋でヘッドホンの両耳提示
  - ・呈示方法：上記2方式の再合成音声を対で呈示
  - ・評価方法：どちらが健常音声に近いかを判定（二者択一）
  - ・被験者：男子大学生3名、女子大学生1名
- [実験結果]

図11に各症状について被験者4名と4名の平均評価値として、各症状音声での結果を示す。図より、男性1名（M2）を除けば、いずれの被験者でも、モーフィング処理による再合成音声での改善度を評価しており、劣化音声の復元において、パラメータ軌跡の平滑化だけでなく適切なホルマント位置へのモーフィング処理の必要性を示している。

### 4-3 モーフィング処理の改善

前節までの処理では、ホルマント空間のうち F1-F2 平面でのホルマント軌跡の連続性のみを考慮したモーフィングに基づくものであった。再合成される音声特徴量の連続性からは、高次ホルマント軌跡を考慮することが望ましい。そこで、図9の補間率算出過程において、第3ホルマント軸を加えた3次元空間において考察した。図12は、この3次元モーフィング処理の効果を示している。ここでは、例として健常者の連続母音のホルマント軌跡から、目標として同一話者のホルマント空間でのモーフィングを行ったもので、図12(a)と(c)の両軌跡の比較、あるいは同図(b)の3次元空間での軌跡（原軌跡とモーフィング軌跡の一致度）よりわかるようにモーフィングのための補間率推定がうまく機能している。

## 5 おわりに

本研究では、音声障害や構音障害に起因する劣化音声の復元を目的として、ホルマントベースのスペクトルモーフィング方式再合成処理を提案し、主に声帯疾患者音声の母音発声音声に関してパラメータ復元と再合成を行い、モーフィングの効果を確認した。日常音声のリアルタイム復元処理へ適用するには、子音を含む連続音声に関する検討が今後の課題として残される。子音特徴の劣化特性に関しては、本文中で述べた音声特徴ベクトル推定エンジンの階層構造を持つ特徴量（音素特徴量）を積極的に利用した異常性の検出を行う必要がある。また、処理のリアルタイム性に関しては、もっとも処理の負荷が重いと予想される音声分析部について、同じく特徴推定エンジンが汎用PC上で、リアルタイムで動作する（～マイク入力音声に関して

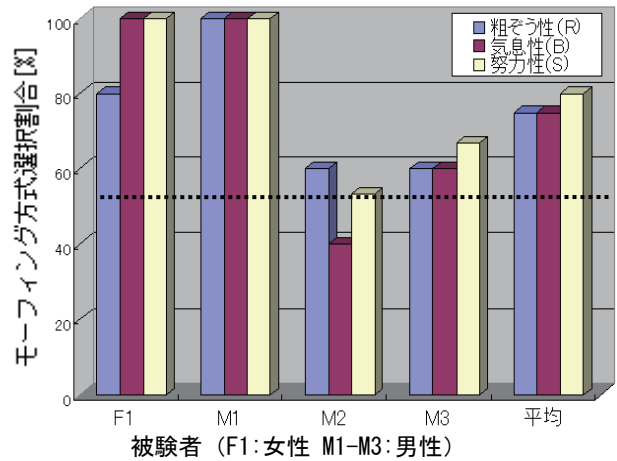
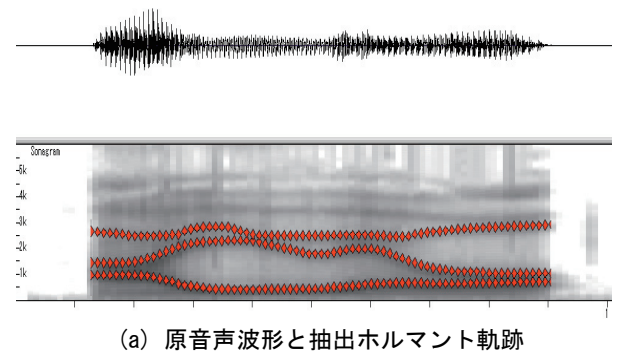
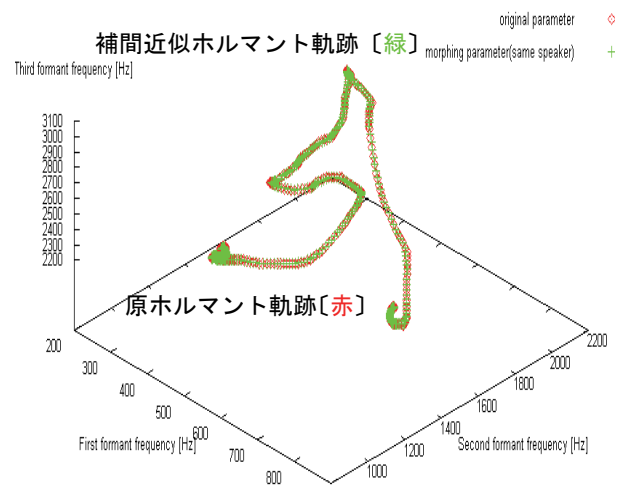


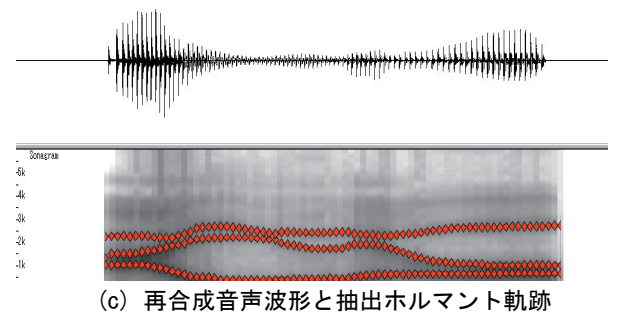
図11 モーフィング復元音声の聴取評価実験結果



(a) 原音声波形と抽出ホルマント軌跡



(b) F1-F2-F3 空間での補間率によるホルマント軌跡



(c) 再合成音声波形と抽出ホルマント軌跡

図12 スペクトルモーフィング音声合成例

リアルタイムで音声特徴ベクトルが画面上に表示できる機能を実現)ことを確認しており、処理のリアルタイム性<sup>[6]</sup>については本研究のシステム化においてほぼ解決したと考えられる。このような背景から、図13に示すような、DSP内蔵の信号処理装置(S-BOX)を中心としたハードウェア構成により、ベースのスタンドアロン型音声復元機能のシステム化を今後の課題としている。



図13 スタンドアロン型音声復元装置の開発システム概観(今後の課題)

### 【参考文献】

- [1] 上田裕市, 廣田美菜子: "スペクトルモーフィング手法に基づく母音合成と話者変換への応用," 電子情報通信学会誌, Vol. J90-D, No. 5, pp. 1318-1322, 2007
- [2] 原田剛輔, 上田裕市, 坂田聡: "ホルマント分析合成方式におけるホルマント伸縮法に関する研究," 電気関係学会九州支部連合大会講演論文集, 06-1P-02, 2006
- [3] A. Watanabe, "Formant estimation method using inverse-filter control," IEEE Trans. Speech Audio Process., vol. 9, pp. 317-326, 2001.
- [4] 上田裕市, 前田大輔, 渡邊亮, "ピーク位置正規化空間での補間処理に基づくスペクトルモーフィング手法", 電子情報通信学会論文誌 D-□ Vol. J83-D-□ No. 7 pp. 1704-1708, 2000
- [5] 今井聖(著), 音声信号処理, 森北出版, 1996
- [6] Y. Ueda, T. Hamakawa, T. Sakata, A. Watanabe: "A real-time formant tracker based on the inverse filter control method," Acoustical Science and Technology, Vol. 28, No. 4, pp. 271-274, 2007

### 〈発表資料〉

題名	掲載誌・学会名等	発表年月
発話インタフェース応用のための階層化音声特徴ベクトルのリアルタイム推定エンジン	電気関係学会九州支部連合大会講演論文集 (CD-ROM)	2008年9月
発声障害音声復元のためのリアルタイム音声特徴量の検討	電気関係学会九州支部連合大会講演論文集 (CD-ROM)	2008年9月
音声応用のためのリアルタイム音声特徴推定エンジンの構築	電子情報通信学会技術研究報告, SP2008-67	2008年10月