音声画像表現を用いる聴覚障害児用発話訓練システム*

宮崎 恵† 坂田 聡†† 岩田 一樹†† 渡邉 亮††† 上田 裕市†
(†熊本大学大学院自然科学研究科 ††熊本大学工学部 †††熊本県立技術短期大学校)

1. はじめに

我々は自己の発声をフィードバックさせることで発声の制御を行い，発話を習得していく．しかし聴覚に障害がある場合，自分の発声した音も聞こえないため，自己の発声をフィードバックさせることが困難となる．そこで発話訓練を行い発声方法を習得する必要がある．このような発話訓練教育は，できるだけ早期の段階から訓練を開始することが望ましいとされているが，発話訓練を指導する言語聴覚士の数が不足しており，十分な訓練ができていないのが現状である．この問題に対処するため，近年様々な発話訓練システムの開発が行われている．

本稿では音声を視覚化した音声画像を用いて，調音とピッチ制御を同時に行うことができる聴覚障害児用の発話訓練システムの提案を行う．訓練システムは，拡張性を考慮し，Java 言語を用いてシステム本体を構築している．提案システムとその性能は次節以降で詳しく説明する．

2. 発話訓練システム

2.1 音声画像表現

提案する訓練システムの概要を Fig1 に示す．(a) はシステムの構成，(b) は音声画像 (/kumamotodaigaku/) の表示例である．本システムで用いる音声画像 [1] は，音声を直観的に捉えることができるように音声の音響的特徴を画像イメージとして表現したものである．使用するパラメータは逆フィルタ制御法 (IFC 法)[2] により抽出されるホルマント周波数，基本周波数，ニューラルネットワークにより抽出される音素的特徴 (有声/無声，調音様式，調音位置など) である．

まず音声の音韻情報は，ホルマント周波数 (F1,F2,F3) の巡回比によって定まる色彩で表現し，色彩の縦方向の幅は基本周波数によって決定する．この基本周波数を用いた表現は，発話者の性別，年齢などの個人情報を画像に反映し，イントネーションの変化を確認することも可能とする．

次に音声の子音情報は，有声/無声，調音様式，調音位置の組み合わせによって表現される．各音素の有声/無声は画像の色彩の有無，調音様式は割り当てたテクスチャパタンの種類，調音位置はテクスチャパタンの表示位置で表現する．子音情報を表すテクスチャパタンは白色で表示し，音韻情報を表す色彩に重畳することで音声を一つの画像イメージとして表現する．ここで重要なのは，音素を推定して一つに限定するのではなくニューラルネットワークによって抽出された要素すべてをアナログ的に色彩に重畳し，音声の特徴がそのまま画像表現されるということである．こうすることで，音声画像の視覚的認識率の向上が期待できると同時に，発話の不完全さが画像イメージ上に直接表現される．Fig1(b) では母音部は色彩で表されている．子音部は (a) で示されるテクスチャパタンにより /k/,/m/,/t/,/d/が表され，/k/,/t/には冗長な要素も重畳されている．このように本方式は，認識手法を使用しない音声の画像表現として提案されている [1]．

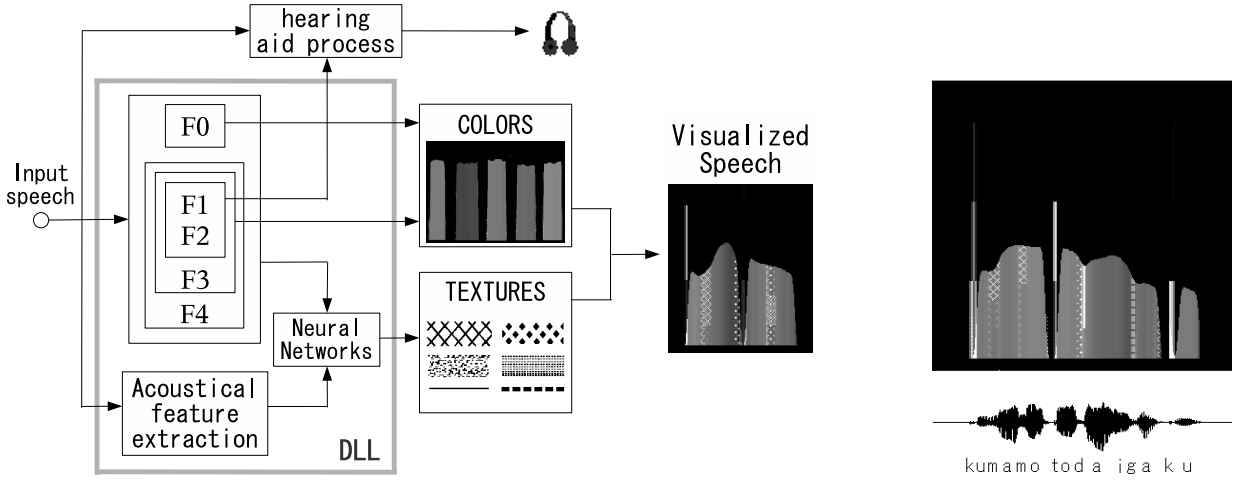
この音声画像を使った聴能訓練システムは池田らによって報告されており，聞き取り能力の改善が認められている [3]．本研究では，このシステムに訓練者が発声した音声を分析・画像化する機能を付加することで，難聴者が単独で聴能と発話 (特に調音とピッチ制御) を統合的に訓練できるシステムの構築を目的としている．

2.2 システム構成

訓練システムは，汎用 PC，音声入出力インターフェース及びシステムのプログラムにより構成されている．システムのプログラムは大部分を Java 言語 (Java アプリケーション) により

∗ A speech training system for hearing impaired children using the speech visualization

By Megumi Miyazaki† Tadashi Sakata†† Kazuki Iwata†† Akira Watanabe††† Yuichi Ueda†
(†Graduate School of Science and Technology, Kumamoto University ††Faculty of Engineering, Kumamoto University †††Kumamoto Prefectural College of Technology)



(a) System structure

(b) An example of visualized speech

Fig1. The proposed speech training system with a visualized speech

構築している．システムの音声分析処理部にはCやFortran言語によって記述された既存のプログラムをJNI(Java Native Interface)を用いて実装している．JNIは，Java言語で開発されたプログラムから，他の言語で開発されたプログラムを利用するための標準プログラミングインターフェース(API)である．JNIを使用するのは，Javaのクラスライブラリでカバーしていないプラットフォームに依存するような操作を行いたい場合，すでに別の言語で作成したライブラリがあり，Javaアプリケーションで使用したい場合などが挙げられ，提案システムでは後者の目的により使用する．JavaプログラムとCプログラムの連携は，CプログラムからDLLファイルを作成し，そのDLLの動作をJavaから制御することで実現している．

本システムでは訓練者に聴覚情報として訓練者自身の発声音声，視覚情報として訓練者自身の発声を画像化した音声画像を呈示し，教師用の音声と音声画像と比較しながら訓練を行う．

視覚情報は，まず音声からIFC法によりホルマント周波数(F1〜F4)，基本周波数(F0)と，ニューラルネットワークの入力となる音響特徴を抽出する．次にF0，F1〜F4，音響特徴をニューラルネットワークの入力とし，音素を特徴付ける有声/無声，調音様式，調音位置を抽出する．抽出されたF0とF1〜F3を用いて音声画像の色彩(RGB値)と画像の縦方向の高さを決定し，ニューラルネットワークによって抽出された音素の特徴で子音情報を決定する．その後それらの情報を統合して音声画像を生成する．音声からホルマント周波数(F1〜F4)，基本周波数(F0)，音響特徴を抽出し，ニューラルネットワークを用いて子音情報を抽出する部分はDLLによって実現し，その他の部分はJavaで構築している．

聴覚情報である補聴音声は，訓練者に応じて音声に補聴処理を加え，聴き取りやすくしたものである．補聴音声は，DLLによって抽出されたF1,F2を使用して生成する．現状のシステムでは音声をそのまま呈示しているが，将来的には訓練者が聴き易いような補聴処理を行う予定である．

2.3 システム機能

Fig2に訓練システムのブロック図を示す．システムは学習モードと訓練モードからなり，訓練者は学習モードで発声方法等の学習をした後，訓練モードに移って自ら発声訓練を行う．学習モードは教師用音声の選択，訓練モードは発声音声の録音再生を行う．

Fig3にシステムのプログラム構造を示す．システムを起動すると，まずメインのウィンドウが開く．訓練者はボタンを選択し，各モードの処理を行う．学習モードと訓練モードについて，以下で詳しく説明する．

2.3.1 学習モード

発話訓練の目的として，(1) ピッチの安定化，(2) 音素の習得，(3) 抑揚(イントネーション)の習得があげられる．そこで本システムでは以下の3つの方法により(1)〜(3)の訓練を目指す．

(1) 母音を持続的に発声する訓練

(2) 各音素を50音順，または発声が易しい順で発声する訓練

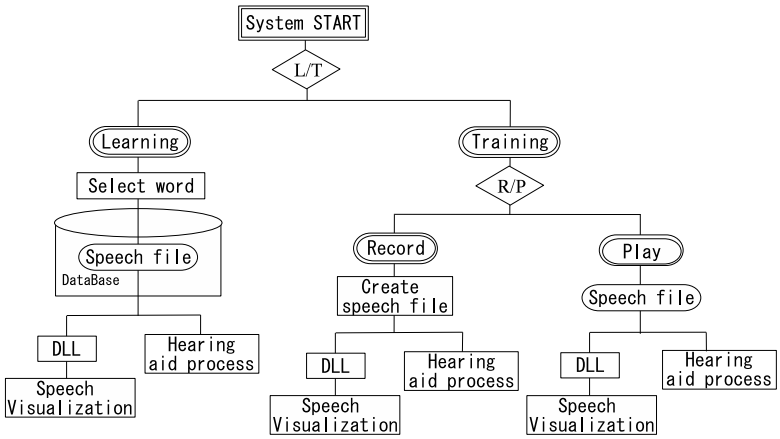


Fig2. Functional diagram of the learning and training modes in the proposed system

(3) 各音素を含んだ単語を発声する訓練

学習モードでは，訓練方法に応じて学習できるように教師用データを上記 3 項目に分類して呈示する．呈示方法は訓練者がまず (1)〜(3) を選択し，さらにその中から学習することばを選択する．すると選択されたことばにリンクする音声によって音声画像と音声の呈示が行われる．

訓練の方法としては，習得が易しい (1) から順に訓練を行っていく．また，声道断面図等を用いて発声方法の呈示も補助的に行う予定である．

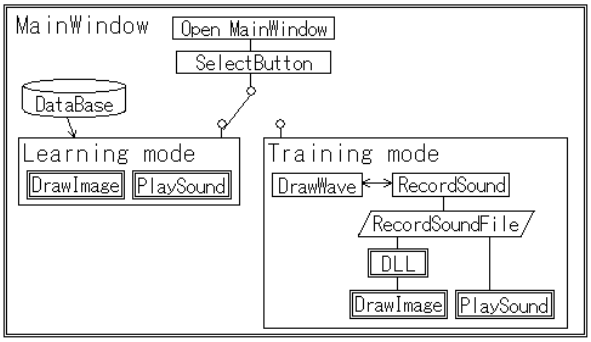


Fig3. Software structure of speech training system

2.3.2 訓練モード

訓練モードでは，学習モードの教師画像を手本として発声の練習を行う．訓練者の発声は録音後，音響特徴を抽出され，学習モードと同様に音声画像と音声の呈示が行われる．訓練者が発声した音声波形はシステム起動時から状態 (各モード) に関わらず表示されるが，音声録音は訓練者の操作によって行われる．

訓練モードでは，発声の向上を確認するために履歴がわかるように発声を保存していく．発声履歴を訓練者が参照できるようにすることで，以前の発声との比較や発声の上達具合の確認などに役立つと考えられる．

2.4 システム実行例

Fig4 にシステムの表示画面，Fig5 にシステムの実行例を示す．システムの実行画面はメインウィンドウとメインウィンドウ内の 5 つのサブウィンドウで構成されている．サブウィンドウ {1}〜{5} の上部が学習モード，下部が訓練モードである．サブウィンドウは各々役割を持っており，サブウィンドウ {1} は教師用音声画像の表示，{2} は学習モードの操作ボタンとリストの配置，{3} は訓練者が発声した音声の音声画像の表示，{4} はマイクからの入力波形のリアルタイム表示，{5} は訓練モードの操作ボタンの配置を行っている．

学習モードでは，まずウィンドウ {2} で学習することばを選択する．ウィンドウ {2} には

(1) 訓練項目選択ラジオボタン
(2) ことば選択リスト
(3) (2) に対応する音声画像，音声の呈示ボタン
(4) 声道断面図表示ボタン

を設定している．訓練者はまず訓練項目 (2.3.1 参照) を選択する．訓練項目によって (2) のリスト表示が切り替わるため，さらにそのリストから学習したいことばを選択する．その後ボタン (3) を押すと，選択されたことばに対応した音声ファイルを元に音声画像の生成が行われ，ウィンドウ {1} で音声画像，ヘッドホンもしくはスピーカで音声が訓練者に呈示される．ボタン (4) は学習の補助として声道断面図を別ウィンドウで表示させる．

訓練モードでは，発声練習のためにマイク等の外部入力により訓練者の発声を取り込み，音声画像化したいときには発声を録音する．ウィ

ンドウ {4} に外部入力の入力波形を出力し，レベルオーバーを防止するため声の大きさを確認しながら発声を行なう．発声を音声画像で確認するときはウィンドウ {5} で，録音〜分析〜音声画像表示の操作を行う．ウィンドウ {5} には

[1] 訓練者選択ラジオボタン
[2] 名前入力ボタン
[3] 録音開始ボタン
[4] 録音停止ボタン
[5] 再生ボタン

を設定している．訓練者はまずボタン [1] で訓練者が該当する項目を選択する．これはホルマント周波数抽出のためのパラメータとなる．次にボタン [2] で訓練者の名前を入力すると，発声保存用のフォルダが生成される．その後ボタン [3] で録音開始の合図を行い，発声をする．録音中はウィンドウ {4} の波形がピンク色に変化し，そうでないときは黒色で表示される．発声終了後ボタン [4] を押すと，録音停止となり，生成された音声ファイルによって音響特徴が抽出され音声画像がウィンドウ {3} に表示される．またボタン [5] を押すと，最新の発声の音声画像と音声が呈示される．

このようにして学習モードと訓練モードを訓練者自身で繰り返し，反復することで，発話習得で重要なフィードバック (自分の発声を耳で聞きながら次の発声に反映させること) を視覚で補いながら行う．現在完成しているのは，学習モードではリストに応じたデータベースの音声，音声画像の呈示，訓練モードでは，外部入力の録音・再生と録音した音声の分析，その結果に基づく音声画像生成・表示である．今後はメニューバーやヘルプ機能等を追加して，幼児にも使いやすいシステムにする予定である．

3. まとめ

発話訓練は専門知識を持つ指導者により，訓練者とマンツーマンで行われるため，指導者不足により学習機会が不十分なのが現状である．本報告では，調音とピッチ制御の訓練を同時に行うことができる発話訓練を，独習で行えるシステムを提案した．今後はシステムの動作を安定させ，より使いやすいような機能を追加する予定である．また本システムの有効性を示すために評価実験を行い，さらに改良を加えていく予定である．

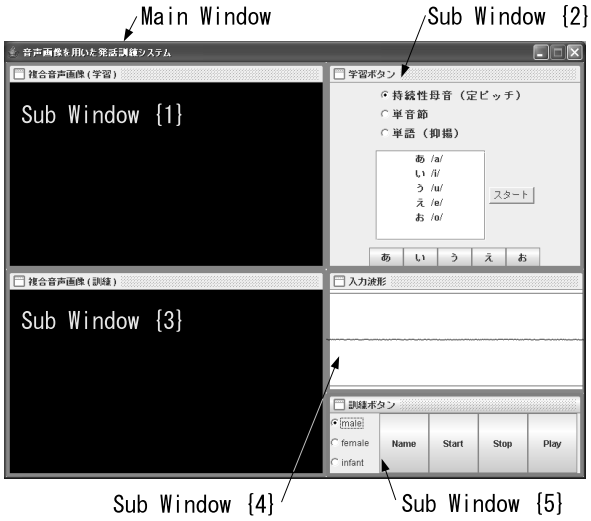


Fig4. Window representation of the proposed system

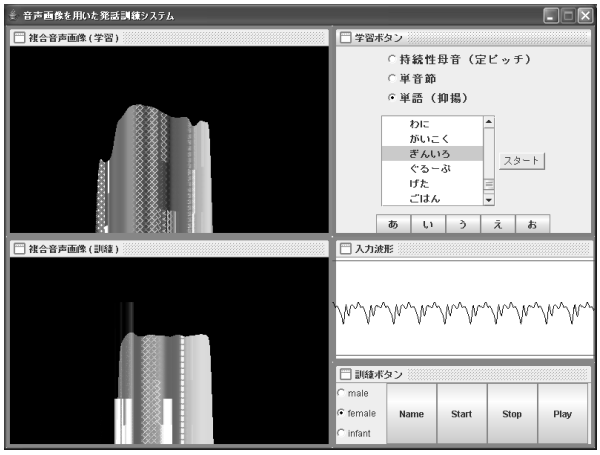


Fig5. An example of speech visualization in the training mode(/giNiro/)

参考文献

[1] A.Watanabe,S.Tomishige,M.Nakatake, “ Speech Visualization by Integrating Features for the Hearing Impaired ”, IEEE Trans. Speech and Audio Processing, Vol.8, No.4, pp454-466, 2000

[2] A.Watanabe,“ Formant estimation method using inverse-filter control ”, IEEE Trans. Speech and Audio Processing, Vol.9, No.4, pp317-326, 2001

[3] 池田隆 他, ”音声画像を用いた難聴者のための聴き取り訓練システム”, 映像情報メディア学会冬季大会講演論文集, 6-1, 1999

[4] 福迫陽子 他, ”言語治療マニュアル”, 医歯薬出版株式会社, 2002