

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 13 日現在

機関番号：17401

研究種目：基盤研究(C) (一般)

研究期間：2012～2015

課題番号：24500650

研究課題名(和文) 音声画像を用いたウェアラブル聴覚代行・補助システムの開発

研究課題名(英文) Development of a wearable hearing substitution system using speech visualization

研究代表者

上田 裕市 (UEDA, YUICHI)

熊本大学・自然科学研究科・教授

研究者番号：00141961

交付決定額(研究期間全体)：(直接経費) 4,100,000円

研究成果の概要(和文)：聴覚障害者の音声コミュニケーション支援を目的として、音声画像と補聴音声を同時に提示するマルチモーダルタイプのウェアラブルシステムの開発を行った。音声分析部と画像合成部は、汎用PC上でリアルタイム処理を確認しているC言語ベースの音声特徴推定と音声画像化の両エンジンをDSPプログラムに移植し、エミュレータ上での実行と処理精度を確認した。更に、ポータブルタイプのDSPボードを設計製作して、オフライン条件での動作を確認した。補聴音声の生成と出力は過去の成果から容易に実装可能であるが、市販のHMDとDSPシステムとのインタフェースの完成までには至っていないことから音声画像表示機能の課題を残している。

研究成果の概要(英文)：For the purpose of supporting a speech communication of hearing impaired persons or deaf persons, we have developed a multi-modal type of wearable system that presented a visualized speech and a hearing aided speech sound, simultaneously. Since both of a speech analysis unit and an image synthesis unit have been implemented in the software engine based on C-language, we have revised those programs and designed a DSP program for real-time processing on a manufactured DSP board. In addition, after designing and manufacturing a portable type of DSP board, a real-time processing and its accuracy were verified under the offline condition. Generation and output of the hearing aid sound is easy to be implemented from our past research result. In order to establish the proposed whole system, we have left some problems of devising a display of the visualized speech image, because it has not yet completed an interface between the HDMI for image signal and the DSP system.

研究分野：音声情報処理

キーワード：音声分析 聴覚障害 補聴器 DSP ウェアラブル 画像表示 ホルマント周波数

1. 研究開始当初の背景

先に開発した音声画像化システムでは、音声ホルマント周波数の色彩表示機能を始め、複数の音声特徴パラメータを統合・融合化して表示することで連続音声の中の音素列を範疇的かつ即時に読解できることから、聾者向けの聴覚代行や難聴者向けの発声・発語学習ツールとしての可能性が期待された。また、音声ホルマントを圧縮増幅する単共振分解型補聴処理による補聴効果が既に確認されている。

このような補聴処理や音声画像処理のベースはホルマント推定である。これには、当研究室で開発した独自のホルマント推定手法である逆フィルタ制御法 (IFC: Inverse Filter Method) を採用している。IFC 法では、年齢・性別の異なる話者の音声について高次ホルマントまで安定かつ高精度で推定することができる。本手法は、その性能評価から、カリフォルニア大学バークレー校 (UCB) の 2011 年度プロジェクト (Voices of Berkeley^(*)) を始め、UCB と同大サンフランシスコ校 (UCSF) の研究グループによる脳科学研究用音声分析ツールとして採用された。

(*) http://voicesof.berkeley.edu/vowel_space.php

このような独自に開発した音声処理技術を背景として、最近、汎用 PC 上での音声画像化リアルタイムツールの開発に至った。一方、最近のハードウェア技術の進展により、DSP や FPGA による超小型の実時間システム化の可能性が広がると同時に、HMD (ヘッドマウントディスプレイ) に代表される視覚呈示インタフェースにより、ウェアラブルコンピューティング環境を実現できる状況にある。

2. 研究の目的

本研究では、これまでのシミュレーションを主とした研究背景に基づき、その有効性が確認されている視覚 (音声画像) と聴覚 (補聴音声) の双方から同時に音声情報を受容する新たな聴覚補償方式 (聴覚代行・補助機能) を、汎用 PC でのリアルタイム処理の実績を踏まえ、ヘッドマウントディスプレイ (HMD) あるいは、スマートグラスに代表されるウェアラブルシステムとして、難聴者・聾者のための口話音声伝達・受容システムを構築することを目的とする。

3. 研究の方法

(1) 音声処理装置の設計と製作

ホルマント周波数などの音声特徴抽出処理をリアルタイムで実現するための可搬タイプのハードウェアを設計・製作する。ソフトウェア開発については、従来、汎用 PC 上に実装して、シミュレーションを行ってきた C 言語ベースのプログラムを改良し、新たな DSP 機能として実装するものとする。

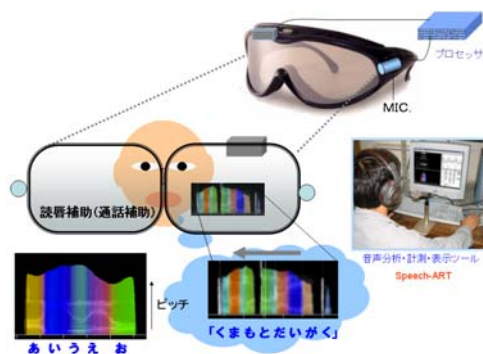


図1 ウエアラブル音声画像化システム

(2) 音声・画像処理の実験システム構築

図1のような実験システムを構築する。

[A] **通話補助機能**：相手話者の読唇と同様、視野の一部にHMDを通した音声画像を電光ニュース式に表示し、同時に本人の聴覚特性に適合するデジタル補聴音声を聴取しながら、実環境における視聴覚情報を融合した通話補助機能を実現する。ホルマント分解補聴処理や合成フィルタ圧縮補聴処理など、これまでの補聴処理に関する研究成果を背景として、音声画像による視覚フィードバックとの同時呈示可能な聴覚情報を与えるための補聴音声の分析合成装置を開発する。

[B] **発話訓練機能**：音声画像は自己発声の視覚フィードバックとしての利用も可能であることから、訓練プログラムの内蔵により、接話マイクを用いた発音・発話に関する自己学習に供する機能も付加する。

なお、ウェアラブル化においては、使用者の日常生活において、次のような機能が利用できるシステム構成とする。

(3) 視聴覚併用による音声受容能力の評価

リアルタイム表示される音声画像について、音声情報の持つ様々な側面 (了解性・直感性・自然性など) から、聴覚受容と視覚受容を主にした心理実験により、音声画像の特徴を生かした人への呈示形態を明らかにし、次年度のヘッドマウントディスプレイベースの視覚受容方式設計の基礎資料とする。

4. 研究成果

4.1 ハードウェアシステム構築

(1) システム系統

開発システムの主要部となる市販のデジタルシグナルプロセッサモジュール (DSP10001; TMS320C6713, 300MHz/SDRAM-16MB, /FLASH-1MB) の概観を図2に示す。TMS320C6713は、32bit浮動小数点演算を高速 (300MHz) で処理する能力を有する DSP である。全体のシステム設計においては、演算処理負担を考慮して、本ボード2枚を用いた dual-DSP 構成とし、外部インタフェースを付加して図3に示すハードウェア構成を最終版とした。音声信号処理部と補聴音声生成と音声画像生成を担うパタン生成部を2つの

DSPモジュールで分担することで処理の独立性を保持している。さらに、DSP間の高速データ授受を確保するために、DualPort-RAM(DPRAM)の採用により、双方のメモリ空間を共有する。音響信号入出力(音声入力と補聴音声出力)では、USBオーディオ機器の利用ができるように、USBポート経由での信号入出力構成(USB-port~SL811-HST~H8-processor)とした。

(2) 回路基板設計・製作

このような機能構成を実現するために、両面6層の多層構成によるシステム基板設計を行った。図4に、設計・製作した多層基板のパターン図例とDSPモジュール、DPRAMおよびインタフェース部品を実装した完成基板の写真を示す。基板サイズは、138mmx134mmとなった。

(3) ソフトウェア開発環境 (図2)

デバッグを実装した汎用PCとDSPモジュールをJTAG経由で接続してインタフェースして、DSP処理のためのソフトウェア開発環境を構築した。

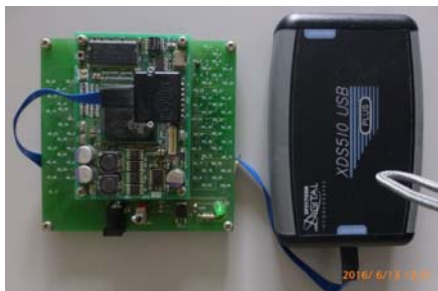
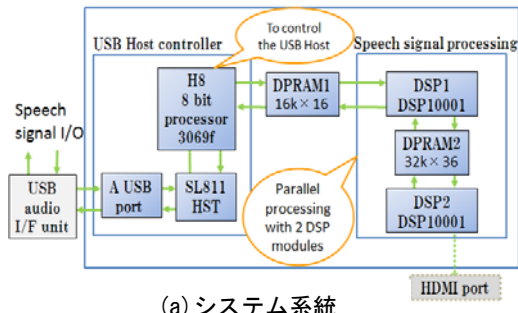
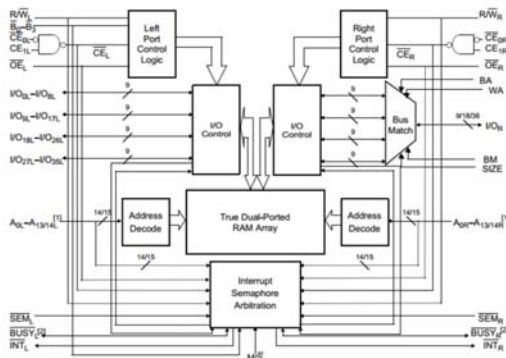


図2 DSP基板 (DSP10001:mtt 製) と JTAG



(a) システム系統



(b) ハードウェア構成

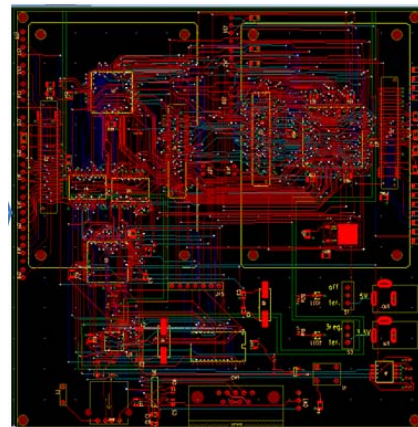
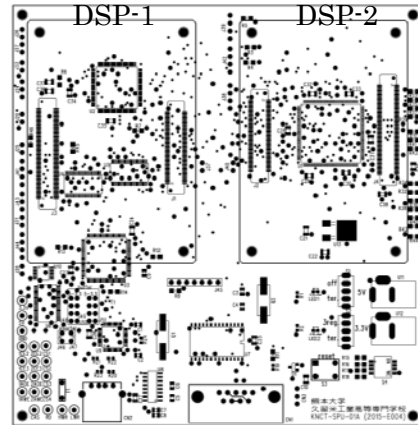
図3 ハードウェアシステム構成

4.2 音声分析・合成処理機能の実装

本システムにおける音響・音声特徴推定を含めたすべての処理はフレーム長 20ms、フレーム周期 10ms で実行されることから、フレーム処理を 10ms 以下で終了する必要がある。これには、音声分析処理と音声画像生成処理および補聴音声生成処理が含まれる。

(1) 音声分析機能

音声分析部(音声特徴ベクトル推定エンジン)のブロック系統図を図5に示す。本機能は、既に汎用PCに実装して動作しているC言語で構築されていることから、DSP処理



(a) 回路パターン(上)と配線済み基板(下)



(b) システム基板 (部品実装済)

図4 完成した可搬型DSPシステム (138mm X 134mm: 6層基板)

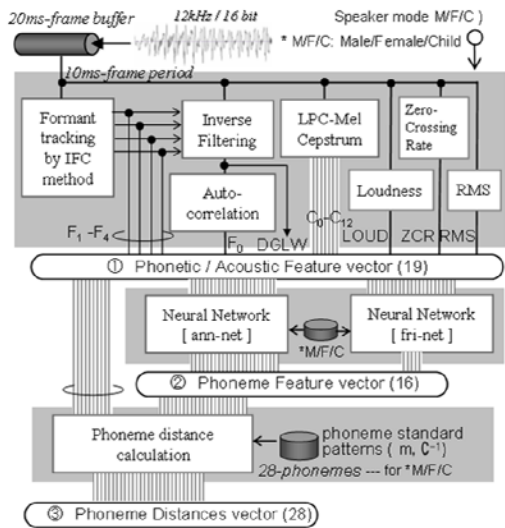


図5 音声特徴推定処理ブロック系統

として実現するためには、高速化処理を前提としたプログラム改良が必要である。例えば、対数計算では、Cコンパイラ内蔵の算術関数 \log では相応の処理時間を要するためある程度の誤差を前提とした近似計算で実現する必要がある。あるいは、DSP特有の並列演算を実行するためのパイプライン処理、double型演算からfloat型演算への変更などが必要である。このような高速化の効果は、表1（音響特徴量と音の特徴量の例）に示すような結果として現れる。

表1: 高速化処理の適用前後の処理時間

	高速化前	高速化後	推定される要素
音響特徴量	6.98ms	1.09ms	ラウドネス
音素特徴量	2.32ms	0.98ms	母音性、鼻音性
合計	9.30ms	2.07ms	

図5のブロック系統図中、ピッチとホルマント周波数推定処理部の詳細を図6に示す。本処理が音声分析のメインであり、相応の処理時間を要することから、本処理の成否がシステム構築の鍵となる。DSP特有の高速処理を考慮したコード修正の結果、図7に示すように、音響特徴量と音声特徴量の必要なパラメータすべてを8.7ms程度(after)で抽出することが可能となった。

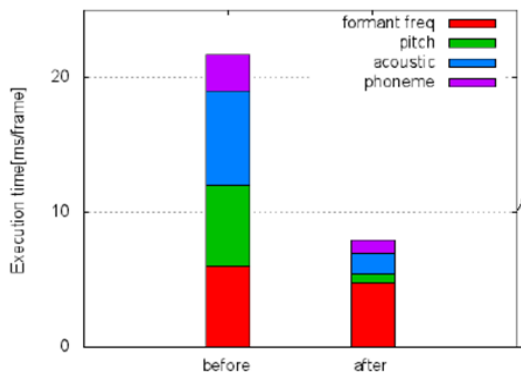


図7 音声分析総処理時間（高速化前後）

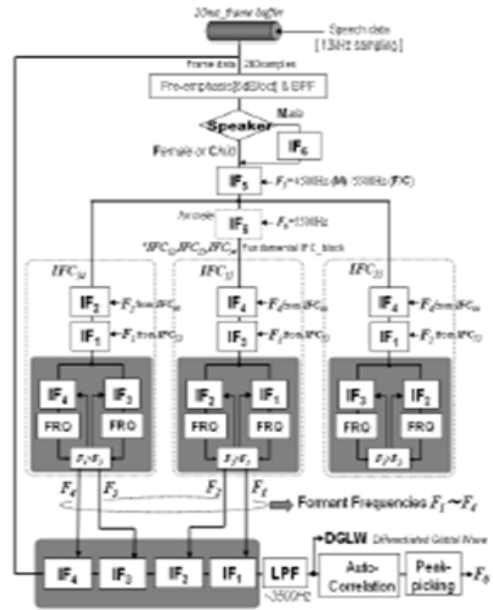


図6 ピッチ・ホルマント推定処理

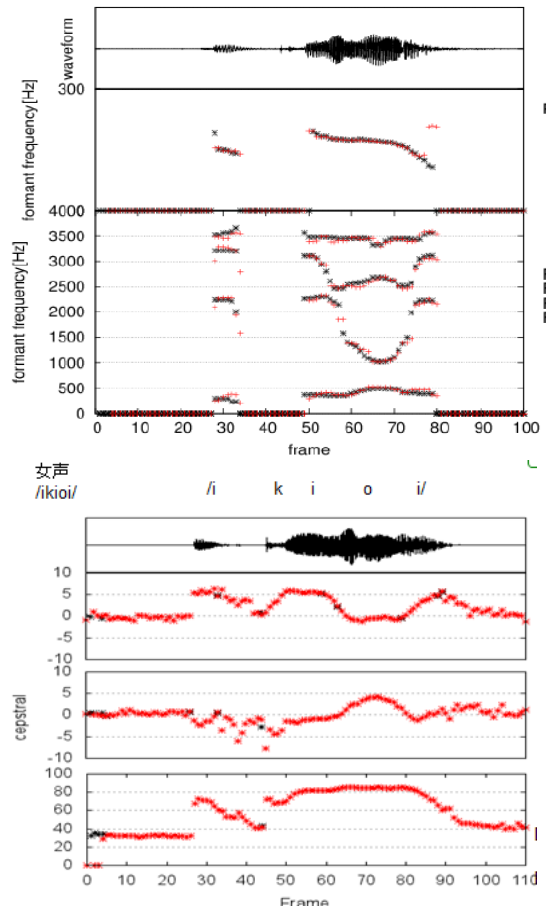


図8 音声特徴抽出例

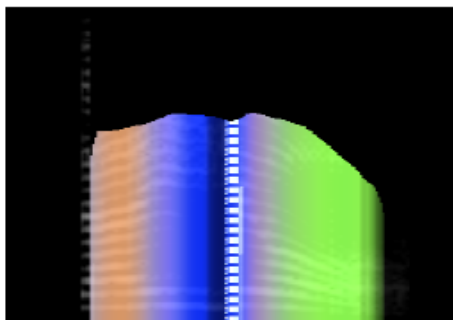
上: ピッチ・ホルマント周波数

下: ケプストラム係数 (1-3次)

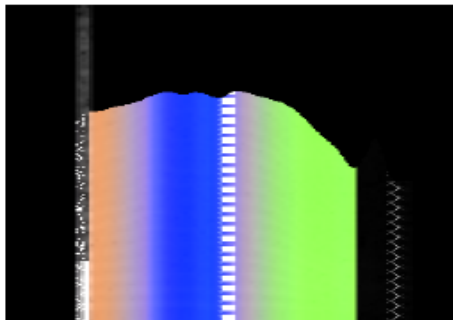
図8に、単語音声 (/ikioi/: 女声) の特徴抽出例を示す。黒点が高速化前、赤点が高速化後の抽出結果であり、両者の誤差はほとんど無視できる程度に小さいことが分かる。

(2) 音声画像生成機能

本システムでは、音声画像も1フレーム周期(10ms)毎に生成される。用いる主なパラメータは、図5のホルマント周波数の巡回比を三原色信号に対応させた母音色彩パターンとニューラルネット出力群(母音性、子音性、摩擦性・・・等)の値(0.0~1.0)で、それぞれ特徴毎に用意されているテクスチャ画像(画素)を重みづけした子音パターンを重ね合わせることで実現される。図9(a)に、従来の汎用PCで生成・表示された音声画像の例(女声単語/zairyo:/)を示す。同図(b)は、DSP処理で抽出されたパラメータ群で生成した画像ファイルをPC画面上に非リアルタイムで表示したものである。両者を比較して、ほぼ同様の視覚イメージが得られていることが分かる。



(a) PCベース画像化群



(b) DSPベース画像化群

図9 音声画像表示例(女声単語:/zairyo:/)

上: 汎用PC処理での生成画像

下: DSP処理での生成画像

(3) Dual-DSPでの分散処理

上述した(1)音声特徴推定部と(2)音声画像生成部の処理について、(1)の方は、図7の処理時間評価より、フレーム周期10msの境界付近にあり、リアルタイム処理の安定性と他のオーディオ信号や画像信号の外部インタフェース処理を考慮するとsingle-DSPで、(1)(2)の処理を同時に実行することは困難である。このため、最終システム設計においては、上述したように2個のDSPモジュールを用いたDual-DSP方式とした。

(4) 音声画像表示・HMD/Smart-glass

DSP処理において画像イメージでメモリに一時格納される音声画像イメージをHDMIインタフェースにより外部機器への出力・表示を予定している。図10は、先に購入し、画像インタフェースの検討を行っているEPSONのHMD(スマートグラス)の外観である。本装置はHDMIインタフェースを備えていることから、製作した可搬型処理装置においてDSP-2で生成される音声画像イメージを出力するためのHDMIインタフェースを実装する必要がある(未対応)。



図10 HMDと可搬型処理装置の基板の概観

4.3 現状と課題

研究期間内で、可搬型DSPボードを設計・試作し、音声分析、補聴音声合成及び音声画像生成における卒とウェア開発を修了した。現段階では、USBオーディオインタフェースとHDMI画像インタフェースの実装が未了である。このため、現状では、実働確認と処理精度の評価では、音声ファイルをPCより、JTAG経由でダウンロードして実施して、実時間処理により抽出されるパラメータ値及び補聴音声及び音声画像イメージの検証を行っている。今後の課題として、早急に上記インタフェース実装に着手して、可搬型リアルタイムシステムとして完成させ、実証実験を実施しなければならない。

5. 主な発表論文等

[雑誌論文] (計 1 件)

- ① 池田隆, 上田裕市, 佐々木俊介, 戸川誠司, 井上敦, “音声応用システムのための可搬型実時間音声信号処理装置の開発,” 久留米高専紀要, 第 29 巻第 2 号, 査読有, pp. 23-27, 2015

[学会発表] (計 7 件)

- ① 七種貴紀, 上田裕市, 坂田聡, 池田隆, “高品質な音声信号処理を実現する汎用型装置に関する研究,” 第 21 回高専シンポジウム in 香川, 2016. 1. 23, 丸亀市民会館 (香川・丸亀市)
- ② T. Ikeda, T. Saigusa, T. Sakata, Y. Ueda, “A development of Wearable DSP Units to Extract Speech Parameters for Hearing Aids and Speech visualizer,” Proceedings of 12th Western Pacific Acoustics Conference 2015. 12. 8, Singapore
- ③ 井上敦, 坂田聡, 上田裕市, 池田隆, “汎用型音声信号処理装置の開発とその応用に関する研究,” 電子情報通信学会九州支部学生会講演会, 2015. 9. 19, 鹿児島大学 (鹿児島・鹿児島市)
- ④ 七種貴紀, 井上敦, 上田裕市, 坂田聡, 池田隆, “装用型音声信号処理装置の開発に関する研究,” 第 20 回高専シンポジウム in 函館, 2015. 1. 10, 函館高専 (北海道・函館市)
- ⑤ Y. Ueda, K. Tominaga, T. Sakata, “Formant-based Articulatory Normalization and its application to Vowel Restoration,” Proc. of 21st International Congress of Acoustics, 2013. 6. 6, Montreal (Canada)
- ⑥ 富永幸佑, 坂田聡, 上田裕市, “障がい音声の発話補償システムにおける母音音声の復元とその音質向上に関する検討,” 第 38 回感覚代行シンポジウム, 2012. 12. 3, 産総研臨海副都心センター (東京・港区)
- ⑦ 戸川誠司, 坂田聡, 池田隆, 上田裕市, “可搬型音声画像表示ツールにおける DSP ベースの音声複合パラメータ推定アルゴリズムの実装,” 電気・情報関係学会九州支部連合大会, 2013. 9. 24, 熊本大学 (熊本・熊本市)

6. 研究組織

(1) 研究代表者

上田 裕市 (UEDA, Yuichi)
熊本大学・大学院自然科学研究科・教授
研究者番号: 00141961

(2) 研究分担者

坂田 聡 (SAKATA, Tadashi)
熊本大学・大学院自然科学研究科・助教
研究者番号: 80336205

池田 隆 (IKEDA, Takashi)
久留米工業高等専門学校・教授
研究者番号: 80222884