

体内伝導インタフェース装置の開発 - Development of Body-Conducted Speech Interface System -

1 0 8 1 0 0 1

研究代表者 広島市立大学大学院 准教授 石光 俊介

[研究の目的]

音声は人間にとって重要なコミュニケーション手段であり、それを失うことは多大な精神的苦痛を伴う。現在、咽頭癌患者の増加により、手術で発声機能を失う機能障害者も増えている。そこで人間と機械の調和により機能障害者のコミュニケーション障害を解決すべく、骨導などの体内伝導音を利用した発声支援システムを開発する。申請者はこれまで体内伝導音を利用した高騒音下での音声通信システム開発を行ってきた。このシステムに機能障害者と一体化した体内伝導音連続認識を組み込み、発声機能障害者自身の体内伝導信号を用いて音声を生成させ、コミュニケーション障害を緩和できるシステムに発展させる。

このように人間と機械が調和し、高騒音下などの環境に左右されない、障害の有無にもかかわらない、人と人をつなぐコミュニケーションを実現するシステムを構築することを目的としている。

[研究の内容、成果]

開発する発声支援システムを図1に示す。発声障害者の方が発声する際に新声門付近から体内伝導音を抽出し、連続サブワード認識を行い、発声音声のサブワード間のフレーム推定を行う。推定したフレーム情報を基に、あらかじめ求めておいた食道発声時の体内伝導音と健常時の音声との間のサブワード単位の伝達関数を用いて変換音声を作成する。本システムでは、話者自身の伝達関数及び声帯振動を用いることを想定している。機能障害前の話者自身の発声した情報を用いることにより、話者自身の声に近い音質変換をすることが可能となることが特徴であ

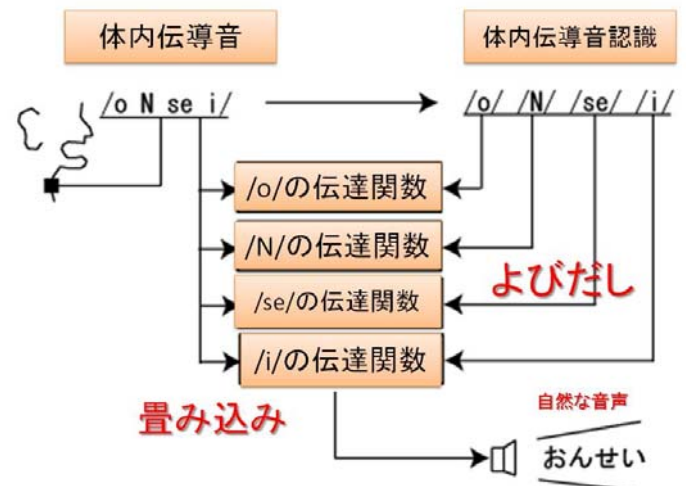


図1 発声支援システム

る。

また、音源として発声した体内伝導音を用いるため、サブワード単位の誤認識により異なる伝達関数を用いたとしても音質変換後の段階で補正されることも期待できる。

次に、単語より短いサブワード単位での音質変換の有効性を検討した。以下に電子協地名百選の"旭"をもちいて作成した変換音声の例を示す。音質変換時に用いられる伝達関数単位は、

- ・音素(/silB//a//s//a//h//i//silE/)
- ・半音節 (/silB//a-s//s-a//a-h//h-i//silE/)
- ・音節(/silB//a//sa//hi//silE/)

である。Julius3.4.2を用いて以上を単位とした連続サブワード認識を行った。最終的には機能障害者に適用するが、サブワード単位の体内伝導音認識と音質変換の有効性の基礎検討として、まず、健常者音声を用いた。体内伝導音もこれまでの検討で有効であった上唇左上部から採取した。推定された境界から音声と体内伝導音を単位ごと切り出し、クロススペクトル法で伝達関数を推定し、音質変換を行った。音声はマイ

クロフォンを用いて話者から 30cm, 体内伝導音は加速度ピックアップで上唇左上部より採取した。サンプリング周波数は 16kHz である。音響モデル及び言語モデルは Dictation Kit ver3.0 付属の不特定話者モデルと web ベースの 6 万語 bigram を用いた。認識対象は、電子協地名百選を用いた。音節単位は音素単位を基に変換して音節単位による境界と定義した。音素及び半音節は音節に比べて推定精度が低かったので、音節単位における認識を用いた場合の結果を示す。

図 2 及び 3 に音声及び体内伝導音を示し、図 4 及び 5 に対応した伝達関数を用いた場合の変換音声と認識誤りが生じたとして異なる伝達関数を用いた変換音声の音声波形及びスペクトログラムを示す。図 4 の変換音声において、損失していた高域成分の周波数特性を回復していることが確認できる。また、図 5 の変換音声でも同等の回復が見られた。

次に、目標となる音声を用いず体内伝導音のみによる明瞭度の高い音質変換手法を提案した。さらに、提案法によって得られた信号に対し従来法を併用することにより極めて明瞭度の高い信号を得ることができることを示す。その方法とは加速度差分を用いることである。加速度差分信号は一定の雑音に重畳された音声の形態となるので、雑音除去により有益な信号を得ることができる。本研究では、音声認識において一般的に用いられているスペクトルサブトラクション法と音声特徴を考慮したスペクトル推定を行うウィナー法を用いて雑音除去を行った。

スペクトルサブトラクション法による雑音除去では一定に含まれていた雑音を除去することができたが、高周波数特性の回復を得ることができなかったうえ、ミュージカルノイズを含む信号となった。そこで雑音が混入された信号から線形予測係数を用いて音声のスペクトル包絡

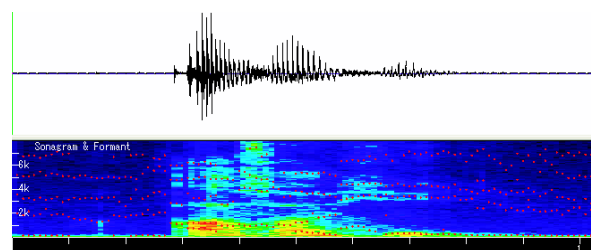


図 2 音声

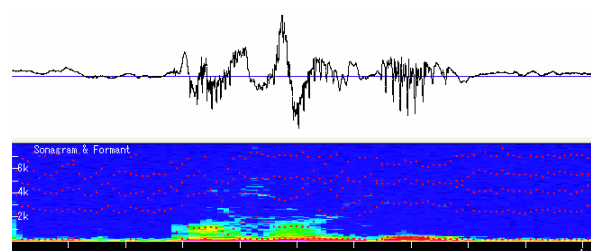


図 3 体内伝導音

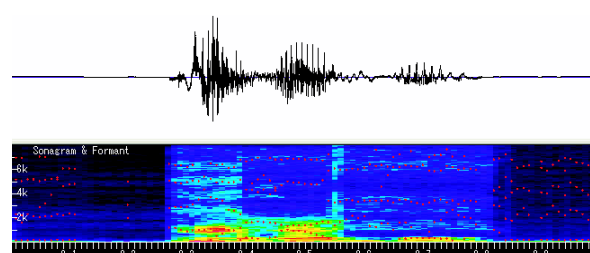


図 4 対応した伝達関数を用いた変換音声

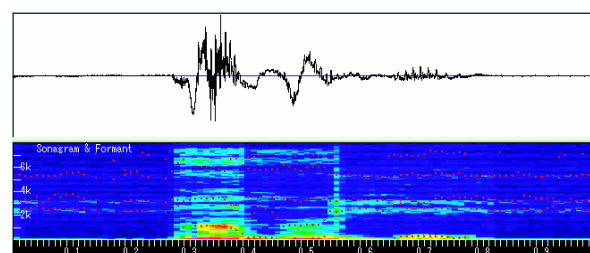


図 5 異なる伝達関数を用いた変換音声
を推定するウィナー法により、雑音除去を試みた。その結果、得られた信号では体内伝導加速度信号において損失していた周波数特性の回復がみられるとともに、スペクトルサブトラクション法において観測されたミュージカルノイズも含まれていなかった。

以上のように加速度差分信号を用いることにより体内伝導音のみで周波数特性の改善を得る

ことを明らかにした．さらなる改善のために単語単位適応フィルタ法を導入し，極めて明瞭度の高い信号を得ることに成功した．その結果を図 6 に示す．次に単語単位よりも短いサブワード単位に対してアルゴリズムを適用した．サブワード単位で音質変換が可能になると発声機能障害者支援システムに利用することも可能となる．まず，音声に対して連続サブワード認識を行い，発声された単語内のサブワード区間の推定を行う．認識デコーダは **Julius**，サブワード単位は音素単位で認識し，音節単位に変換して用いる．図 7 にその結果を示す．結果より，単語単位と同様にサブワード単位でも適応アルゴリズムの有効性を確認することができた．単語単位では若干エコーがかかっていたようであったが，サブワード単位ではエコーがなくなった．また，サブワード単位で推定した信号を連結するときに，不連続点が生じるためインパルスが混入されることが予測されたがその問題もなかった．

以上のように健常者ベースでシステムの妥当性が示せたので，発声機能障害者音声並びに体内伝導音の評価を行った．以下に示す被験者 3 名の音声と体内伝導音を収録し，音声認識を行った．

- ・ 22 才健常者，男性
- ・ 28 才声帯ポリープ患者，男性
- ・ 61 才声帯全摘出患者，音声再建法発声，男性

各収録点における文章認識の結果を図 8 に示す．全被験者において音声と比べて体内伝導音の認識率は非常に低いことが分かる．また，音声では健常者と機能障害者 1 に違いは見られなかったが，体内伝導音では 20%以上の認識率の低下を確認した．

音声認識率低下の原因を探るため，音声パラメータでの比較を行った，これは認識率低下の原因の究明のみではなく，発声障害者音声の特

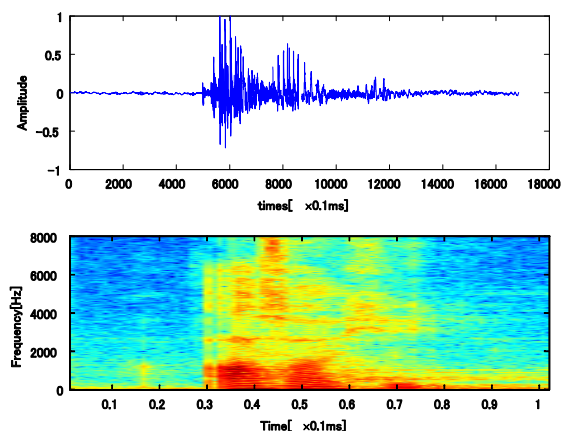


図 6 単語単位適応フィルタ法

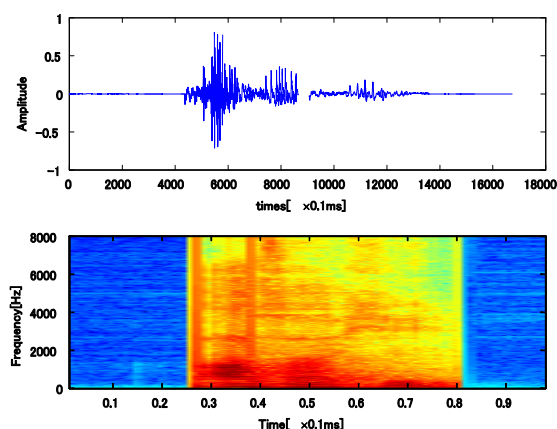


図 7 可変長サブワード単位適応フィルタ

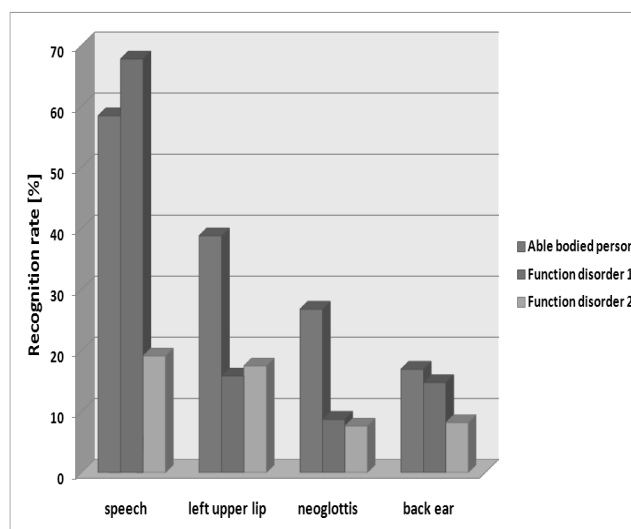


図 8 音声認識率

まず、知覚線形予測分析(PLP 分析)を用いて調音器官の伝達特性(スペクトル包絡)を比較した。PLP 分析によって得られた「はちのへ」のスペクトル包絡を図 9 に示す。PLP 分析は声帯の音源信号がインパルスであると仮定してスペクトル包絡を推定しているため、声帯に異常がある機能障害者の場合、正常なインパルスを得ることができないことで音源信号の特徴を含んだスペクトル包絡を推定していると考えられる。健常者と比較して、機能障害者 1, 2 の周波数のピーク（フォルマント）の位置が異なり、機能障害者 2 は全体的なパワーも弱いことが確認できた。これが、機能障害者の音声聞き取りづらい原因の一つであると考えられる。

一方、多くの音声認識では MFCC（Mel-Frequency Cepstrum Coefficient）を用いて特徴量を推定している。そこで、MFCC を用いて特徴量比較を行った。調音器官の伝達特性だけを比較するために、低次の 8 次成分までを用いた。この結果より、健常者に比べ機能障害の症状が重くなるにつれて特徴量の違いが大きくなることが確認できた。特に声帯全摘出者は全体的なパワーも弱く、健常者のような強弱が見られない。このように、一つの音素内で機能障害者の特徴量推定できないことが音声認識率の低下へと繋がったと考えられる。

〔今後の研究の方向、課題〕

今後は今回被験者になっていただいた音声再建法による発声を行われている方に特化したシステムを完成させながら、汎用化を図っていく方針である。そのためには発声機能障害者データベースのさらなる拡充とシステムの小型軽量化の検討が必要である。このようにデータベースの完成度を上げると同時に音声認識率向上のための検討も同時に行っていく予定である。それによりシステムの完成度も上げていきたいと考えている。

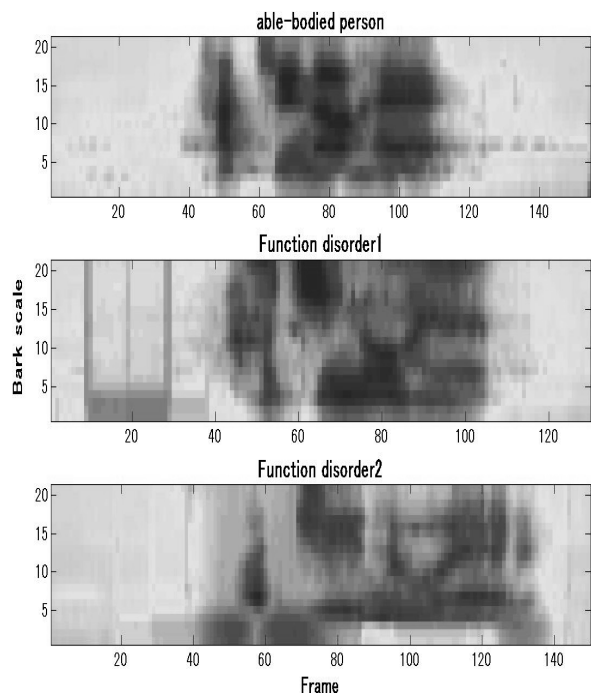


図 9 PLP 分析によるスペクトル包絡

〔成果の発表、論文など〕

- [1] M. Nakayama, S. Ishimitsu and S. Nakagawa, "Speech Estimation from Body-conducted Speech with Differential Acceleration," 2009 IEEE International Conference on Networking, Sensing and Control, pp.858-863
- [2] M. Nakayama, S. Ishimitsu and S. Nakagawa, "Speech recognition with body-conducted speech using differential acceleration," Acoustics'08 Paris, pp.1407-1412
- [3] 小田康平, 中山仁史, 石光俊介, 発声機能障害者の音声特徴分析, 日本機械学会中国四国支部 第 47 期総会・講演会 講演論文集, pp.83-84
- [4] 中山仁史, 石光俊介, 中川誠司, 雑音に頑健な音声-骨伝導マイクの基礎検討, 第 14 回高専シンポジウム in 高知 講演要旨集, pp.238

