

機関番号：12401

研究種目：基盤研究 (C)

研究期間：2008 ～ 2010

課題番号：20500152

研究課題名(和文) 骨導音声を利用する高騒音環境下での音質確保とその音声認識、話者認識への応用

研究課題名(英文) Speech Quality Preservation with Bone-Conducted Speech in Highly Noisy Environments and Its Application to Speech and Speaker Recognition

研究代表者

島村 徹也 (SHIMAMURA TETSUYA)

埼玉大学・大学院理工学研究科・教授

研究者番号：40235635

研究成果の概要(和文)：骨導音声の低周波成分から高周波成分を推定する2つの方法を見出し、それらを話者認識に適用した。ハードウェアおよびソフトウェアの双方から認識実験を行い、提案する特徴量が有用であることを明らかにした。

研究成果の概要(英文)：Two methods to estimate the high frequency components of bone-conducted speech from its low frequency counterparts were derived, and applied to speaker recognition. Recognition experiments through hardware and software showed that the proposed methods could provide a very useful feature for speakers.

交付決定額

(金額単位：円)

	直接経費	間接経費	合 計
2008 年度	2,200,000	660,000	2,860,000
2009 年度	600,000	180,000	780,000
2010 年度	600,000	180,000	780,000
年度			
年度			
総 計	3,400,000	1,020,000	4,420,000

研究分野：音声工学

科研費の分科・細目：情報学・知覚情報処理・知能ロボティクス

キーワード：骨伝導

1. 研究開始当初の背景

近年では、情報の保持、安全な社会の構築のために、セキュリティの重要性が共通認識としてもたれてきている。中でも、人間の生理学的特徴(バイオメトリックス)を利用する個人認証の研究が多く注目を集めている。特に、マイクロホンとパーソナルコンピュータが準備できれば容易に実現が可能なことから、音声による個人認証の研究が盛んに進められている。しかしながら、我々の生活環境に見られる雑音環境下で高い認識率を保持する、確約される認証アルゴリズムは存在していない。

2. 研究の目的

本研究は、音声を利用する個人認証システムの質的向上および利用環境の拡大を目指し、雑音・騒音環境下においてさえも高認識率を確保する新しい認証システムを構築することを目的とする。

3. 研究の方法

従来着目されてこなかった骨導音声積極的に利用し、話者の個人情報として利用可能な特徴パラメータを見出す。また、認識を目的としたときの雑音低減法のパラメータ最適化を図り、人間の聴覚との関連性を明らかにする。最終的には、骨導音声およびそれに雑音低減を施した品質改善骨導音声から話者認識実験を行い、特徴量として利用する

ための骨導音声の有用性を明らかにする。

4. 研究成果

主な研究成果は、学会発表(1)で発表された。以下にその概要を述べる。

骨導音声の低周波成分から高周波成分を推定する方法を見出し、それを話者認識のための特徴量とした。ここでは、具体的に2つの推定法(提案法1, 提案法2)を導出し、利用した。提案法1と提案法2をまとめておくと、以下のようになる。

<提案法1>

- Step 1 : 雑音の付加された骨導音声を取得する。
- Step 2 : 骨導音声にスペクトル引き算を行うことにより、ある程度雑音を取り除く。
- Step 3 : 短時間フレームに切り取り、高速フーリエ変換により振幅スペクトルに変換する。
- Step 4 : Step 3 で得た周波数特性からケプストラム法によりスペクトル包絡を取得する。
- Step 5 : 取得した振幅スペクトルから低周波数帯の成分のみを取り出す。
- Step 6 : Step 5 で取得した特性を時間波形と見て再び高速フーリエ変換を行う。
- Step 7 : Step 6 で取得した特性にハイパスフィルタリングを行い、基本周期成分を取り除く。
- Step 8 : Step 7 で取得した特性に逆高速フーリエ変換を行い、元の周波数領域に戻す。
- Step 9 : Step 8 で得た周波数特性を重み係数を掛け合わせながら高周波数帯に複写する。
- Step 10 : Step 4 と Step 9 で得たそれぞれの周波数特性を掛け合わせる。
- Step 11 : 逆高速フーリエ変換により時間信号に変換する。
- Step 12 : 全フレームが終了するまで、Step 3 から Step 11 までの処理を繰り返す。

<提案法2>

- Step 1 : 雑音の付加された骨導音声を取得する。
- Step 2 : 骨導音声にスペクトル引き算を行うことにより、ある程度雑音を取り除く。
- Step 3 : 短時間フレームに切り取り、高速フーリエ変換により振幅スペクトルに変換する。
- Step 4 : Step 3 で得た周波数特性からケプストラム法によりスペクトル包絡を取得する。
- Step 5 : Step 3 で取得した振幅スペクトルから低周波数帯の成分のみを取り出す。
- Step 6 : Step 4 で取得したスペクトル包絡

から低周波数帯の成分のみを取り出す。

Step 7 : Step 6 で取得したスペクトル包絡の逆数をとる。

Step 8 : Step 5 と Step 7 で取得した特性を掛け合わせる。

Step 9 : Step 8 で得た周波数特性を重み係数を掛け合わせながら高周波数帯に複写する。

Step 10 : Step 4 と Step 9 で得たそれぞれの特性を掛け合わせる。

Step 11 : 逆高速フーリエ変換により時間信号に変換する。

Step 12 : 全フレームが終了するまで、Step 3 から Step 11 までの処理を繰り返す。

図1に、それぞれの方法により、低周波数成分から高周波数成分を推定した例を示す。提案法2では、提案法1に比べ高周波数帯において低周波数成分の基本周期が保たれていることがわかる。

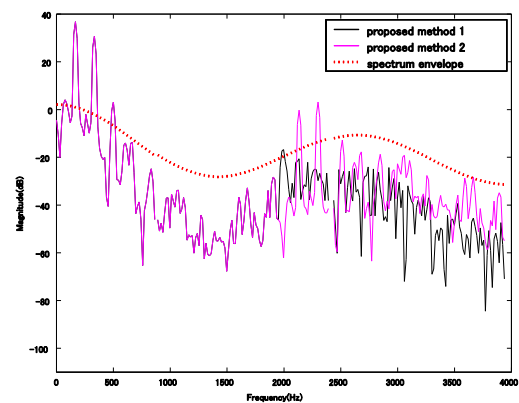


図1 2つの提案法の特性の違い

図1で示した処理波形を具体的に示しておく。図2が入力の雑音が付加された骨導音声波形である。図3は、スペクトル引き算を適用し、雑音低減した処理波形を示している。図4、図5はそれぞれ提案法1, 提案法2で処理して得られた処理波形を示している。

図3からは、スペクトル引き算により、雑音低減がなされていることが確認できる。図4と図5では、処理波形がほとんど類似しており、微細特性は異なるものの、提案法1と提案法2は同程度の波形処理特性を有するとと言える。

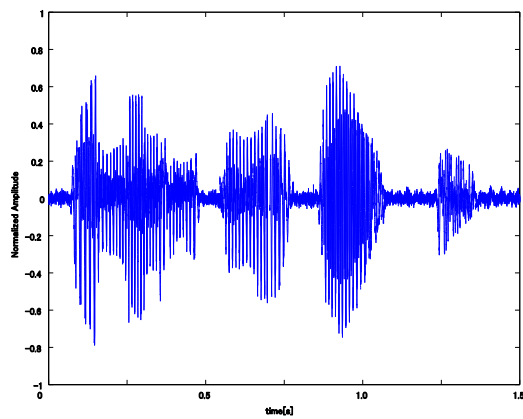


図2 雑音付加骨導音声

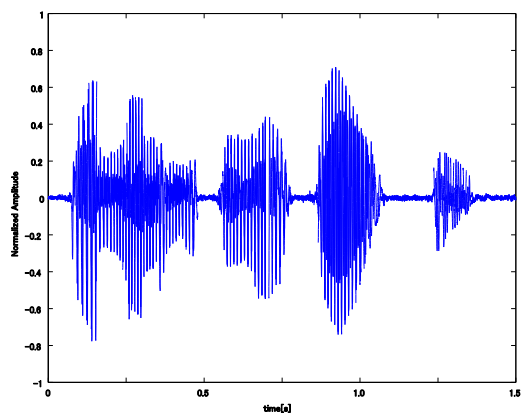


図3 雑音付加音声にスペクトル引き算を施した処理波形

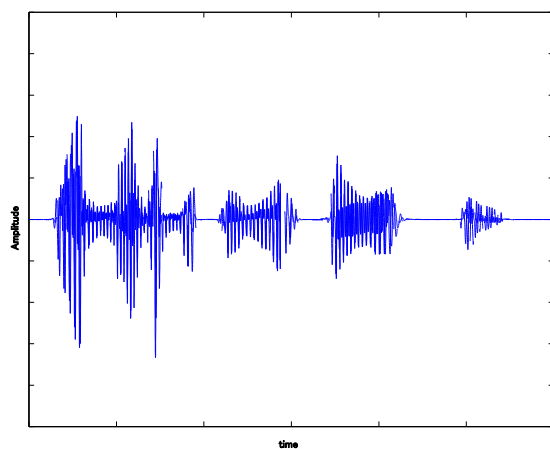


図4 提案法1で処理した波形

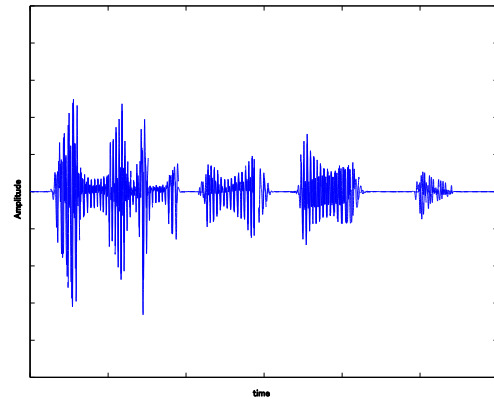


図5 提案法2で処理した波形

上記の2つの提案法を利用し、得られたスペクトル特性を話者認識の特徴量とし、話者認証実験を行った。

実験に用いた音声データを以下にまとめておく。

話者：男性18名、女性17名

気導音声収録マイク：Shure SM58

骨導音声収録マイク：Temco HG17

発生語：島村研究室

発生回数：12回(学習用6回、認証用6回)

サンプリング周波数：48.0kHz (8kHz にダウンサンプリングして利用)

音声収録は図6の接続図のようにした。

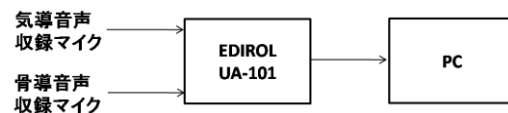


図6 音声収録の接続図

研究室にある遮音室にて、骨導音声収録マイクを装着した話者に、気導音声収録マイク前で発声してもらい、2チャンネル同時録音して利用した。

雑音データは、話者に骨導音声収録マイクを装着させたまま、周りで雑音を流すことで取得した。話者は発声せずに、気導音声収録マイク、骨導音声収録マイクそれぞれから雑音のみを収録することで、それらを雑音データとして扱った。この雑音収録の様子を図7に示しておく。

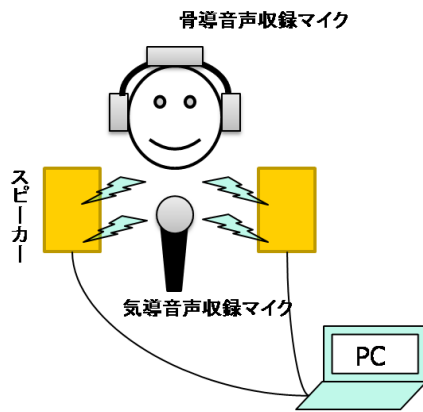


図7 雑音データの収録

ここでの雑音データは、計算機から生成される白色雑音とした。信号対雑音比 (SNR) を 0, 5, 10, 15, 20dB と変化させて利用した。

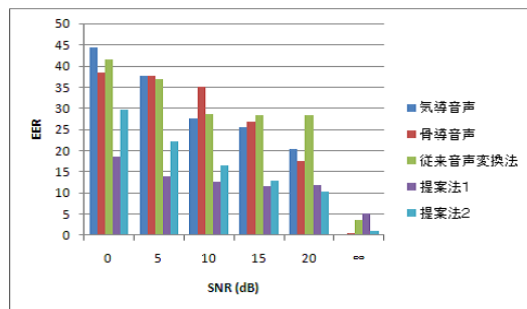


図8 話者認証実験結果(EER の場合)

図8は、本人拒否率と詐称者受理率が等しい Equal Error Rate (EER) の場合の認識率を算出した例を示している。ここでの算出値は、各音声データからの平均の結果である。図8からは、すべての SNR において、提案法1、提案法2が、気導音声、骨導音声および従来の音声変換法に比べ、認識率が改善されていることがわかる。また、提案法1と提案法2は、SNR が低くなるにつれて、大幅な特性改善を与えることが見て取れる。

図8は、SNR=20dB の場合を除けば、提案法1の方が提案法2より特性が良好であることを示している。特徴量の利用の仕方での認識率は変化してしまうため、今後より厳密な骨導音声の特性把握を必要とすることを裏付ける結果と考えられる。

話者認識専用ハードウェアで実験した結果においても、図8と類似する傾向が観測されたことをここに付記しておく。

個人認証の重要性は周知である一方、その技術的複雑さから、無雑音の環境を想定した研究が多くなされている。しかし、現実問題

としての雑音・騒音環境は、明らかに認証率を低下させる。この具体的対応策を見出す過程で、本研究での骨導音声の利用は、1つの有用な方向性であると言えよう。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計0件)

〔学会発表〕(計2件)

- (1) T.Shimamura, Accuracy Improvement of Speaker Authentication in Noisy Environments Using Bone-Conducted Speech, 2010.8.2, USA
- (2) T.Shimamura, Toward Speech Communication in Highly Noisy Environments Using Bone Conduction, WSEAS International Conference on Communications (Plenary Lecture), 2010.7.23, Greece

〔図書〕(計0件)

〔その他〕

本研究成果の一部は学会発表(2)の Plenary Lecture で発表した。また、2011年10月に依頼されている WSEAS International Conference on Applied Computer Science の Plenary Lecture でも発表予定である。

6. 研究組織

(1) 研究代表者

島村 徹也 (SHIMAMURA TETSUYA)
埼玉大学・大学院理工学研究科・教授
研究者番号：40235635

(2) 研究分担者

なし