

科学研究費助成事業 研究成果報告書

平成 26 年 6 月 24 日現在

機関番号：12401

研究種目：基盤研究(C)

研究期間：2011～2013

課題番号：23500203

研究課題名(和文) 複数のマイクロホンによる骨導振動のピックアップと骨導音声の品質改善

研究課題名(英文) Pick-up of Bone Conduction with Multiple Microphones and Quality Improvement of Bone Conducted Speech

研究代表者

島村 徹也 (SHIMAMURA, Tetsuya)

埼玉大学・理工学研究科・教授

研究者番号：40235635

交付決定額(研究期間全体)：(直接経費) 2,900,000円、(間接経費) 870,000円

研究成果の概要(和文)：単一の骨導マイクを用いた場合における骨導音声の分析において、まず骨導音声を利用した基本周波数抽出を検討し、高騒音環境下においても良好な推定結果を与える抽出方法の導出に成功した。また、骨導音声の時間領域での振幅特性に着目し、気導音声の振幅特性に近づける変換を施すことにより、骨導音声の品質改善をもたらす事実を明らかにした。複数の骨導マイクを用いた場合においては、適応フィルタを利用し、多入力1出力システムを構築し、骨導音声の最良の品質改善を与えるアルゴリズムを検討した。

研究成果の概要(英文)：For the case of single use of a bone conducted microphone, fundamental frequency detection with bone conducted speech was investigated, and a detection method was derived, which provides a good performance even in highly noisy environments. Amplitude of bone conducted speech was investigated in the time domain, and it was found that a transform of amplitude of bone conducted speech into that of air conducted speech leads to an improvement of speech quality. For the case of multiple use of bone conducted microphones, a multi-input single-output system was constructed involving adaptive filters. Algorithms to provide the best improvement of speech quality were investigated.

研究分野：総合領域

科研費の分科・細目：知識情報処理・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

音声認識、話者認識においては、無雑音環境下においてほぼ 100%に近い認識率を得るに至っている。しかしながら、実環境としての雑音環境下においては、認識率が 50%以下になるなど大幅に低下してしまうことがよく知られている。一方、電話においても、駅や交差点など、高騒音を伴う環境では、情報を正確に伝達できず、これらを例としても明らかかなように、音声処理システムの今後の進展においては、雑音対策が極めて重要である。申請者の研究室では、広い意味での音声処理システムへの応用を念頭にこれまで音声分析の研究を進めてきた。また、実システムとしての利用を目的としたときの雑音環境下での音声処理の重要性を考慮し、雑音環境下での音声特徴パラメータの抽出及び雑音付加音声の音声のみの強調の問題（音声強調）に着手してきた。最近では、特に後者とも関連し、特殊なマイクロホンを利用することにより、雑音環境下、特に高騒音環境下においてさえも高い信号対雑音比(SNR)が確保できることから、骨伝導マイクロホン（骨導マイクロホン）の利用に着目してきた。本研究では、この骨導マイクロホンから収録された音声、骨導音声、を利用し、その品質改善を図り、広く音声処理システムへの応用に努めるものである。

2. 研究の目的

これまでに、骨導マイクで取得した骨導音声のある程度品質改善することには成功してきた。しかしながら、骨導音声の 1kHz より高い周波数帯の振幅レベルを減衰させてしまうその本質的な性質から、気導音声に品質を近づけるためのフィルタリング処理は必然的に高周波数帯域を増幅する特性となり、わずかにでも雑音が骨導マイクから混入する場合には雑音強調が発生し、聴感的には付加雑音が増幅された印象を持つことがある。元々の音声信号の明瞭性は改善される

ものの、この雑音強調で、処理された骨導音声の品質が必ずしも保証されず、この問題を解決する必要がある。そこで本研究では、骨導マイクの人体へのフィッティング問題に着目する。骨導マイクは、人体に伝わる骨導振動をピックアップして電気信号に変換することを基本原理とする。よって、人体へのフィッティングの仕方によって音質が変わってくる。しかしながら、強すぎるフィッティングは振動ピックアップに圧迫を与え、弱すぎるフィッティングは振動ピックアップをそもそも困難にする。よって、容易に解決できないのが現状と考えられる。そこで本研究では、これまで未解決問題とされてきた振動ピックアップの具体的な方法を検討する。まず、一カ所のほぼ類似する位置からの複数の振動ピックアップされた骨導音声から、骨導マイクのフィッティング依存性を調べ、最適フィッティングの方法を見いだす。続いて、頭の頂点、側頭部のような離れた位置からの振動ピックアップされた骨導音声の類似性、関連性を明らかにし、複数マイクから得られる複数の骨導音声入力から、含有される付加雑音を最大限に抑圧し、最良品質な骨導音声を出力する多入力1出力型の音声強調システムを構築する。そして、その有効性を室内実験により明らかにする。また、本音声強調システムにさらに、骨導音声入力のみを利用し、気導音声の音質を近づける復元フィルタリング手法を新たに導出しつつ、統計信号処理技術をも加味し組み合わせることにより、多入力1出力型復元フィルタリングシステムを構築する。そして、その有効性を実際の様々な環境における実験を介して明らかにする。

3. 研究の方法

申請者の研究グループでこれまでに提案してきた骨導音声の品質改善法は、主として 2004 年での国際会議論文での方法に基づく、骨導音声と気導音声の長時間スペクトルを

利用する方法である。ここでは、その方法にホルマントの帯域幅を調整する機構を導入し、高周波数成分を擬似的に復元する方法へと発展する。そして、これまでの知見と結びつけ、データ長、フィルタ長、雑音量等を変化させて、復元音声への影響を実験的および解析的に調べることにより、骨導マイクのフィッティング問題との関連性を明らかにする。

マイクの間隔を十分に空けて並べるヘッドセット型マイクロホンシステムを、複数の骨導マイクを利用し試作する。TEMCO EM7 はイヤホン型骨導マイクであり、これもヘッドセット型マイクに取り付ける形を考える。ここでは、各マイクから取得される振幅レベルはもちろんであるが、位相成分すなわち時間遅延量を厳密に測定し、複数のマイクから得られる骨導音声の類似性と関連性を明らかにする。そして、骨導音声同士を強調し合う複数マイクの設置方法を明らかにし、それを実装する。

ここまでで得られる知見を基に、各マイク入力での骨導音声の振幅レベルを調整し、平均化して混入される雑音を低減し、音声強調を図る。少なくとも、この方法により、取得される骨導音声の強相関性と混入する付加雑音の弱相関性の性質から、単一の骨導マイク利用の場合に比べ、雑音抑圧が図れると期待できる。また、各マイクからの骨導音声の音質を客観的に評価する方法を導入し、その評価値を基に、各マイク入力からの骨導音声を重みづけ処理する方法へと発展する。

ここまでで得られた知見から、上記の多入力1出力型音声強調システムの出力に、先に得られた品質改善アルゴリズムを付け加えることで、多入力1出力モデル型の復元フィルタリングシステムを構築する。ここでは、特に品質改善アルゴリズムのパラメータの最適化に取り組む。さらには、復元されるべき気導音声を所望特性とみなし、最適フィ

ルタ設計の立場から、各骨マイク入力に処理を施す方法へと発展させる。

4. 研究成果

(1)雑音環境下における基本周波数の高精度抽出のために気導音声と骨導音声を組み合わせる抽出法を検討した。提案手法は、気導音声と骨導音声の振幅スペクトルをあらかじめべき乗することによって拡張、圧縮しておくことにより音声成分の強調と雑音成分の圧縮を行う。さらにその2つのスペクトルをかけ合わせることで、相関の強い音声成分を強調、相関の弱い雑音成分を抑圧することで自己相関関数のピークを強調し、ピーク検出を容易にすることができた。気導音声と骨導音声のそれぞれのべき乗数、 α と β の最適値が、雑音SNR、抽出精度の閾値を変化させることでどの程度変化するかを確認した。結果として、雑音量を変化させると高雑音環境にするほど α を小さく、 β を大きくしていく傾向が見られた。また、男女における傾向の違いが見られた。男性の場合は気導音声で圧縮処理、骨導音声で拡張処理の場合が多かったが、女性の場合は、どちらも拡張処理となることがほとんどであった。気導音声と骨導音声の相関の強さも見てみたが、相関の強い男性の方が一致率が高くなるわけではなかった。

(2)CV音を用いて気導音声と骨導音声の相対振幅の違いを調べた。気導音声ではほとんどの音声において、横に長い紡錘形をしており、子音部分の振幅は母音部分の振幅よりも小さい。しかし、骨導音声の場合では調音によって、子音部分の振幅が母音部分の振幅よりも大きくなったり、ほとんど観測できないほど小さくなったりする。その傾向としては、

・発声による傾向では、有声音は振幅が増大し、無声音は振幅が減少する

・調音による違いでは、鼻音が一番振幅が大きく、次いで破裂音が大きい。摩擦音の場合は振幅が減少するという傾向が有ることが分かった。これらは、声帯振動の有無や、高周波数域の減衰などの骨導音声の特徴や従来の見地と一致する。また、頭頂型骨導マイクロフォンとイヤークラス型骨導マイクロフォンと比較すると、上述した傾向は変わらないが、イヤークラス型骨導マイクロフォンによる骨導音声ではクラス2に分類される音声が増えており、気導音声に近い傾向を持っている事がわかる。気導音声と骨導音声の波形例を図1に示しておく。

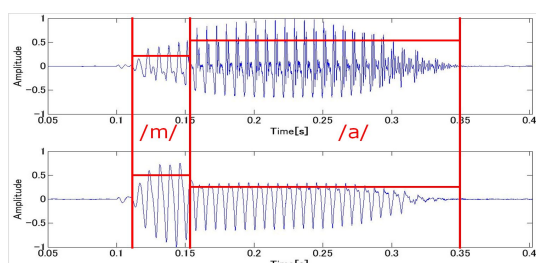


図1 波形比較(上：気導音声、下：骨導音声)

(3)骨導音声の相対振幅を気導音声の相対振幅へと調整・変換する品質改善法を検討した。具体的には、振幅調整の方法をとった。実際の環境での使用を想定し、子音区間と母音区間が既知である場合と未知である場合のそれぞれの処理を検討した。前者ではブレボイスパーの扱いによって3種類の処理を、後者ではフレームベースでの処理の合計4通りの処理を検討した。振幅調整法により導出された音声と、プリアンファシス処理と振幅調整法を組み合わせた場合に導出された音声を用いて明瞭度試験を行った。実験の結果、プリアンファシス処理とフレームベースでの振幅調整法を組み合わせた導出音声は最も高い明瞭度を示した。これは他の処理と比べ、急激な振幅の変化が少なく、明瞭度に影響を与えるようなノイズが発生

しづらいためであると考えられる。この結果から、周波数領域にのみ着目し、骨導音声の品質改善を図る手法よりも、時間領域である相対振幅の違いも考慮した骨導音声の品質改善手法の方がより効果的であると言える。また、振幅調整法は、簡便なアルゴリズムなため、他の音声強調手法との併用が容易に行える事も有益である。一方で、振幅調整法はその処理に平均振幅値を用いているため、雑音の混入に対して脆弱であるという問題点がある。そのため、実際に使用する際には、雑音低減手法との併用が必要になる。

(4)複数マイクロフォンを用いた音声強調手法を検討した。図2にあるようなマイクセットを利用した。

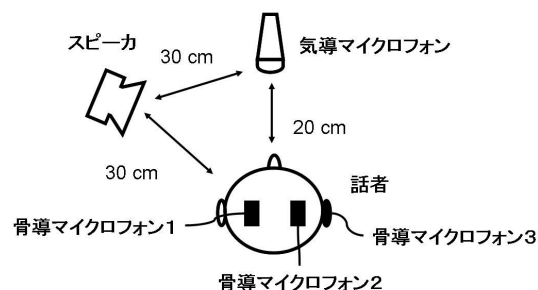


図2 骨導マイクセット

まず、適応処理による骨導音声の品質改善について述べる。従来の適応処理による骨導音声の品質改善法では、骨導音声から気導音声へと近づける適応処理が提案されている。しかしここでの提案法では、従来手法と異なり複数骨導マイクロフォンにより得られた複数の骨導音声同士を、入力信号と所望信号として適応処理に用いることで品質改善を図る。聴取による評価においては、どの適応アルゴリズムにおいても、フィルタ次数が小さいときは、雑音の音量は大きくないが音声部分に集中して雑音がのっているように聞こえるため、耳障りな音声である。これは適応に用いる信号数が少

ないために、音声信号の適応が不十分であり、誤差信号が大きくなってしまふ事が原因だと考えられる。フィルタ次数を大きくしていくと、徐々に音声品質が向上していく。これは、適応に最適なフィルタ次数に近づくとつれて誤差信号が減少していくため、出力も雑音成分が小さくなり、音声品質が改善するためである。また、最適なフィルタ次数を超えて、さらにフィルタ次数が大きくなると徐々に出力音声はこもったような音声になっていく。これは、フィルタ次数の増加によって適応フィルタの性能としては向上していくが、適応線スペクトル強調器の特徴として、入力レベルの高い成分から適応していくことが知られている。そのため、骨導音声が高いパワーを持っている低周波数域から線強調が始まるために、相対的にこもったような音声になってしまうと考えられる。距離尺度である LAR 距離による評価では、フィルタ次数 20 以下の場合 LAR 距離も安定しない。フィルタ次数の増加に合わせて LAR 距離は増加していくが、あるフィルタ次数を超えると減少傾向を示す。これらの理由としては、上述したようにフィルタ次数が小さいときは、誤差信号が大きいため出力信号も安定せず、LAR 距離も上下する。フィルタ次数の増加に合わせて音声品質が良くなっていくが、LAR 距離の基準となる骨導音声から遠くなってしまふためである。また、最適なフィルタ次数を超えると音声がかもっていき、LAR 距離は小さくなっていく。これは基準となる骨導音声により近づいているためである。もう一つの良く用いられる距離尺度である WSS 距離による評価では、フィルタ次数が小さいときはあまり安定せずに上下しているが、多くのアルゴリズムで低次数の方が WSS 距離が小さく、フィルタ次数の増加に合わせて WSS 距離も増加している。この理由として、低次数の場合では音声部分に集中して雑音が

のっているが、パワーが小さい WSS 距離は大きくならない。しかし、フィルタ次数が増加するにつれて、音声品質が良くなる一方で雑音信号はややパワーが大きくなっていくため、WSS 距離が増加していると考えられる。

最適なフィルタ次数による導出音声を聴き比べると、AP アルゴリズム、NLMS アルゴリズムによる導出音声は同程度の品質改善した音声であり、LMS アルゴリズムによる導出音声は一段劣る。RLS アルゴリズムは、誤差信号による雑音の付加が著しく適応処理には向いていない。これらの適応アルゴリズムの収束速度は一般に RLS アルゴリズムは最も早く、続いて AP アルゴリズム、NLMS アルゴリズム、LMS アルゴリズムとなると知られている。これらより、収束速度が速すぎても遅すぎても良質は導出音声を得ることは難しくなることがわかる。

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 5 件)

(1) T. Shimamura, Noise Robust Speech Analysis, WSEAS International Conference on Applied Computer Science, 2014.1.29-1.31, Cambridge USA.

(2) M.S. Rahman and T. Shimamura, A Study on Amplitude Variation of Bone Conducted Speech Compared to Air Conducted Speech, Proc. Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, 2013.10.29-11.01, Kaohsiung Taiwan.

(3) 箴恒介, 島村徹也, 気導音声と骨導音声を組み合わせた基本周波数抽出, 音声研究会, 2012.12.21, 東京

(4) T. Kato and T. Shimamura, Consonant Behavior of Bone Conducted Speech in Japanese, Proc. IEEE International Symposium on Intelligent Signal Processing and Communications, 2012.11.04-11.07, New Taipei Taiwan

(5)T.Shimamura, Speech Enhancement Using Bone Conducted Speech, WSEAS International Conference on Signal Processing, Computational Geometry and Artificial Vision, 2012.8.21-8.23, Istanbul Turkey.

〔図書〕(計 0 件)

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