

発話障害者のコミュニケーション支援のための携帯電話用読唇システムの開発 (092108002)

Lip reading based communication support system of mobile phone for voice disordered people

研究代表者

齊藤剛史 九州工業大学

Takeshi Saitoh Kyushu Institute of Technology

研究分担者

なし

研究期間 平成 21 年度～平成 23 年度

概要

本研究開発課題では、発話障害者のコミュニケーション支援として、カメラで撮影した顔画像を入力として、読唇技術により発話内容をリアルタイムで認識するシステムを開発する。課題を実現するために、撮像機構・手段、認識対象とする発話内容や認識に有効な特徴量など本システムに有用な読唇手法を模索、提案した。これらの成果をもとにプロトタイプシステムを開発した。被験者実験を通じて開発システムの有効性を示した。

Abstract

In this research project, I have developed a real-time lip reading system for speech disorder people to support their communication. In order to achieve the objective, I investigated the photography environment, target words, extraction method, features, and recognition method. The prototype system was developed based on these results. The validity of development system was shown through the subject experiment

1. まえがき

先天的な聴覚障害により言語を習得するのが難しく二次的な聴覚性の発話障害を併発する人、脳の障害により言語機能が失われる人、あるいは気管切開術や喉頭切除術などにより一時的あるいは後天的に発話障害をもつ人がいる。厚生労働省「平成 18 年身体障害児・者実態調査結果」によると、聴覚・言語障害者数は 34 万人いる。また赤木らは口頭摘出による失声者数は 2～3 万人いると報告している（日本医療機械学会大会 2008 年）。

発話障害者は音声による意思伝達が困難なため、他人とのコミュニケーションにはコミュニケーションボードや筆談、手話を利用する。筆談は筆記用具などの道具や指文字などによりコミュニケーションを行えるが長い文章を伝えるには労力がかかる。また手話は、手指動作と非手指動作を同時に使う視覚言語で、音声言語と並ぶ言語である。聴覚障害者が中心となって使用しているが、健常者に使える人が少ない問題がある。また口が動くものの意思伝達が難しくストレスが溜まる問題がある。このような状況から、障害者の生活の質（Quality of Life, QOL）の向上は急務であり、読唇技術を利用したコミュニケーション支援システムに対する強いニーズがある。

我々はこれまでカメラ画像から得られる視覚情報を利用して発話内容を認識する読唇に関する研究に取り組んできた。本研究開発課題では、読唇技術を携帯電話の機能として搭載することにより、発話障害者のコミュニケーション支援を実現するためのリアルタイム読唇システムの開発を目的とする。これを実現するために、撮像機構・手段、認識対象とする発話内容や認識に有効な特徴量など本システムに有用な読唇手法を模索、提案する。本稿では特に開発システムについて説明する。

2. 研究内容及び成果

本研究開発課題で目的とするコミュニケーション支援システムを構築するために、(A)プレ評価実験の実施、(B)発話障害者およびそのコミュニケーション支援に関する

調査、(C)読唇システムにおける撮像機構・手段の確立、(D)認識対象の発話内容の決定および発話シーンの撮影、(E)撮影画像から自動的に口唇領域を抽出する手法の提案、(F)口唇の動きと認識精度に関する検討、(G)発話シーンのフレームレートと認識精度の検討、(H)読唇に有効な視点の検討、(I)単語ベース読唇と文章ベース読唇の検討、の 9 項目について取り組んだ。

プロトタイプシステムの開発にあたり、本研究開発課題で得られた以下に示す成果をシステム開発に反映した。

- (1)単語ベース読唇と文章ベース読唇の検討結果に基づき、認識対象とする発話内容は単語単位あるいは短文単位（以後、定型文と記す）とする。
- (2)読唇に有効な視点の検討結果に基づき、正面顔画像を用いる。
- (3)発話シーンのフレームレートと認識精度の検討結果に基づき、リアルタイム処理時間を 1 フレームあたり 333ms 以下、すなわち 10fps 以上を目標にする。さらにニーズ調査に基づき、以下の要件を満たすシステムを目標とした。
- (4)研究のためのシステムでなく、研究代表者以外も利用可能なシステムを開発する。そのため認識処理だけでなく、学習データの登録作業も手軽に操作可能にする。
- (5)本研究開発課題の目的は製品開発でなく、提案手法の実用性を見出すことである。このことを考慮し、携帯電話にアプリケーションを組み込むのではなく、ノート PC ベースでシステムを構築する。
- (6)発話終了後 2 秒以内に認識結果を出力することを目標にする。
- (7)認識結果をテキストとして画面に表示するだけでなく音声メッセージも出力する。
- (8)操作の利便性を考慮して把持しやすいシステムの操作にはキーボードでなくワイヤレスコントローラを利用する。

以上の項目に基づき、開発したプロトタイプシステムの構成図を図 1 に示す。本システムのハードウェア構成は、

カメラと PC、コントローラから構成される。本システムは事前に登録した定型文をリアルタイムで認識し、認識結果を音声メッセージとして出力する。本システムは登録モードと認識モードの二つの操作モードをもつ。登録モードはユーザが事前に定型文を登録する場合に利用する。認識モードはユーザが話者とコミュニケーションする際に利用する。

本システムはこれまで開発を進めてきた読唇アルゴリズムを単にリアルタイムで実現するだけでなく、コミュニケーションを支援することを目標としている。これを実現するために、発話区間の自動検出機能、目標文入力のための誤認識文の回避機能、メッセージを一文ずつ伝達するだけでなく複数の単語を組み合わせたメッセージを伝える 2 種の達機能、および光源環境の影響を軽減するためのカメラ制御機能を実装した。

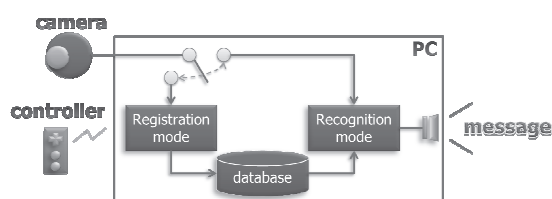


図 1 開発システムの構成

プロトタイプシステムではノート PC (CPU: Intel Core i5-520M、2.40GHz)、Point Grey Research 社製 USB カメラ Chameleon、コントローラとして無線で把持しやすい形状をもつ任天堂 Wii リモコンを利用した。カメラより取得される画像サイズは 640×480 画素であるが、抽出処理の高速化を図るため 320×240 画素に縮小した。また音声メッセージの出力には、アクエス社 AquesTalk2 を利用した。また前述のハードウェア構成における処理速度は 22.3fps であり目標をクリアし、リアルタイム性を実現した。

開発したプロトタイプシステムを用いて被験者 4 名に対して評価実験を実施した。認識対象とする定型文は日本語会話文 50 文とした。被験者は椅子に着席し、カメラは三脚で固定した。登録モードを利用して各発話内容においてそれぞれ 10 サンプル登録した。1 サンプル 50 文の登録に要する時間は 5～10 分であった。一人 10 サンプルずつの登録作業を終えた後に認識実験を実施した。50 文の認識実験を 1 セットと定義し、認識実験は全ての被験者について 9～12 セット実施した。実験実施期間は被験者により異なるが 37 日～53 日であった。また登録作業および認識実験は全て各被験者が操作して行なった。その結果、各被験者の平均認識率は 87.6%、97.5%、94.6%、97.6% であり 4 人の平均認識率は 94.4% となり高い認識精度を得られた。また認識時間は約 0.2 秒でありリアルタイムで認識が行えていることを示した。さらに発話区間終了後から音声メッセージが出力されるまでは 2 秒以内であり目標をクリアした。図 2 に実験の様子を示す。



図 2 プロトタイプシステムによる実験の様子

次に被験者がカメラを手を持った状態における認識実験を実施した。被験者 1 名に対して前述の実験と同様の実験を実施した結果試行回数 5 回の平均認識率は 71.5% であった。これは 1 回の試行で 50 文を実施したため、カメラを持つ手に疲れが生じて発話時にカメラが動いてしまったためと考える。また医師からの要望にあった入院患者の利用を想定し、被験者 1 名に対してベッドに寝た状態で認識実験を実施した。試行回数 5 回の平均認識率は椅子に着席した姿勢とほぼ同じ精度が得られた。

3. むすび

本研究開発課題では発話障害者のコミュニケーション支援を目的として、リアルタイム読唇システムを開発し、評価実験により有効性を示した。

読唇技術は発声による音声信号を用いた音声認識と同じアプリケーションが考えられる。近年、急速に普及しているスマートフォン向けのアプリとして iPhone は 2011 年に「siri」、NTT ドコモは 2012 年に「しゃべってコンシェル」の提供を開始した。これらは声を発する必要があり、騒音時に認識精度が低下する、発話障害者は利用できない、公共の場所などで声を発することが望まれない場所で利用できないなどの問題が残っている。本研究により得られた成果は、これらの代わりとなるインタフェースとしての利用が期待されている。さらに読唇技術を利用して家庭内で電化製品を制御する研究、個人認証に関する研究などが報告されている。本研究開発課題ではコミュニケーション支援に焦点を絞っていたが、本研究開発課題の成果はこれらの関連研究へ応用することは十分可能である。今後はこれらの応用研究を進めていきたい。

【誌上発表リスト】

- [1] Takeshi Saitoh, Ryosuke Konishi, “Profile Lip Reading for Vowel and Word Recognition”, 20th International Conference on Pattern Recognition, pp.1356-1359 (2010 年 8 月 23 日～26 日)
- [2] 齊藤剛史、森下和敏、小西亮介、“発話シーンからのキーフレーム検出とキーフレームに基づく単語読唇”、電気学会論文誌、Vol.131、No.2、pp.414-424 (2011 年 2 月)
- [3] Takeshi Saitoh, Ryosuke Konishi, “Real-time Word Lip Reading System based on Trajectory Feature”, IEEJ Trans. On Electrical and Electronic Engineering, Vol.6、No.3、pp.289-291 (2011 年 5 月)

【申請特許リスト】

- [1] 齊藤剛史、ワードスポッティング読唇装置及び方法、日本、出願番号：特願 2010-201629、出願日：2010 年 9 月 9 日
- [2] 齊藤剛史、コミュニケーション支援システム、日本、出願番号：特願 2011-182594、出願日：2011 年 8 月 24 日

【報道発表リスト】

- [1] “携帯電話カメラで読唇”、山陰中央新報、2009 年 4 月 10 日
- [2] “テクノロジー Powered by 九工大 シリーズ 13 視覚情報に基づく発話内容の認識に関する研究”、西日本新聞、2010 年 10 月 20 日
- [3] “医療と大学連携模索”、西日本新聞、2011 年 7 月 23 日

【本研究開発課題を掲載したホームページ】

<http://www.slabb.ces.kyutech.ac.jp>