様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成　３０　年　　６　月　２２　日現在

機関番号：　３４４０６
研究種目：　基盤研究(C)（一般）
研究期間：　2015～2017
課題番号：　１５Ｋ０１４８７

研究課題名（和文）多様なユーザーに適応可能な複合機能を有する発声支援装置に関する研究

研究課題名（英文）Development of Adaptive and Multi-functional Speech Enhancement System for Both EL-users and Esophageal Speech Users

研究代表者

松井　謙二（MATSUI, Kenji）

大阪工業大学・ロボティクス＆デザイン工学部・教授

研究者番号：３０６１３６８２

交付決定額（研究期間全体）：（直接経費）　　3,600,000円

研究成果の概要（和文）：喉頭がんなどにより発声困難になった方々が用いる発声支援に関して、まず、超小型のワイヤレス拡声装置を開発し、ユーザー団体での量産試作が行われるようになった。次に、完全なハンズフリーで発声するためのフォトリフレクタを用いた口唇の動き検出器を開発し、口を動かすだけで人工喉頭の制御が可能であることを確認した。また、スマートフォンによる口唇認識を目指し、基本的な実験を行い、良好な認識結果を得た。

研究成果の概要（英文）：People who have had laryngectomy require speech enhancement system.This study was undertaken to explore the feasibility of using compact loudspeaker with wireless function and EQ. Also, lip motion detection device with photo sensor was developed and tested. Then, lip reading function was developed and obtained close to 90% vowel recognition accuracy.We confirmed the total system using those functions will be improve the user's communication performance.

研究分野：　音声信号処理

キーワード：　人工喉頭　ハンズフリー　食道発声

１.研究開始当初の背景

喉頭摘出などの発声障害に対して、　過去５０年間、人工喉頭や拡声器などの発声支援器具が用いられているが、その技術進化は極めて乏しい。しかし、現代は長寿化、就労の高齢化など、健常者と同等に元気に働くユーザーが増加し、これらのユーザーに対して発声支援が一層重要になりつつある。一方、昨今の IT、特に AI、IoT など急速に進展する技術により革新的機器開発が可能になりつつある。

２.研究の目的

これからの発声支援には障害のタイプ、程度や発話スキルに個人適応することが重要となる。本研究の目的は、（１）周囲の視線が気にならずストレスなく使える利点、（２）カメラ、マイク、ディスプレイ連携によるマルチモダル機能を有する、（３）広く普及しており新たな機器購入の必要がない、などの特性を活かした、見た目に自然、低コスト、軽量、かつ高度な信号処理技術を用いる革新的発声支援技術の実現を目指す。

３.研究の方法

真に役に立つ発声支援装置を開発するためには、ユーザー視点の開発プロセスは極めて重要である。本研究では、デザイン思考的アプローチ、およびイノベーション手法の一つである Foresight & Innovation を参考にしつつ、常にユーザーである銀鈴会の方々のご意見を伺いながら研究開発を行った。

４.研究成果

（１）ハンズフリーユーザーインタフェースの開発

発声支援装置の全体的なシステムのイメージの概略図を図１に示す。このシステムの特徴として、①より小型で省電力な振動子を利用できる、②ハンズフリーなユーザインタフェースを提案できる、③ EL 発声機能と拡声機能を兼用したり、片方の機能だけを使ったり切り替えることができるようにすることで、周囲のユーザ自身の状況に臨機応変に対応できる。といった３点が挙げられる。

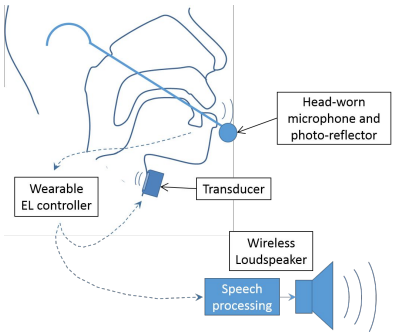


図１　システムのイメージ

口の開閉を検出するために、本研究ではフォトリフレクタからの出力電圧の変化量を用いた EL コントローラを試作した。試作装置は図２のように構成した。

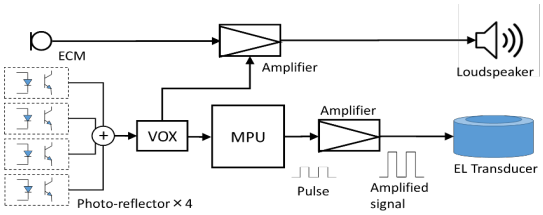


図２　試作装置の構成

図２の構成をもとに作成したプロトタイプを図３に、装着の例を図４に示す。

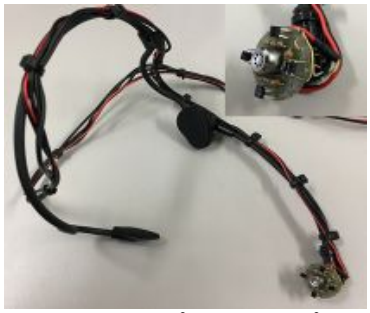


図３　EL コントローラのプロトタイプ

図４　EL コントローラの装着の様子

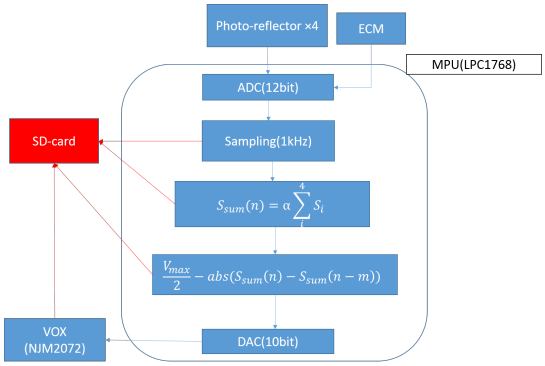


図５　評価実験のためのハードウェア構成

以上の EL コントローラの性能の評価実験を行った。測定は図５のようなハードウェア構成で行った。マイク（ECM）は発話時にのみ EL を起動するスイッチが入ることを確認するために用いた。

マイクロプロセッサーユニット(MPU)に用いた LPC1768 の内部では以下の、

$$S_{sum}(n) = \alpha \sum_{i=1}^{4} S_i(n) \qquad \cdots\cdots(1)$$

$$V(n) = \frac{V_{max}}{2} - |S_{sum}(n) - S_{sum}(n-m)| \qquad \cdots\cdots(2)$$

式(1)、(2)を実行させている。式(1)の $S_i(n)$ は各フォトリフレクタの値で、4 つのセンサ値を加算させている。式(2)で m サンプル前の値との差分を求めて、その差分信号 $V(n)$ を VOX に入力させている。また、V_{max} は LPC1768 の電圧レベルの 3.3[V]である。本システムをテストした結果を図 6 に示す。

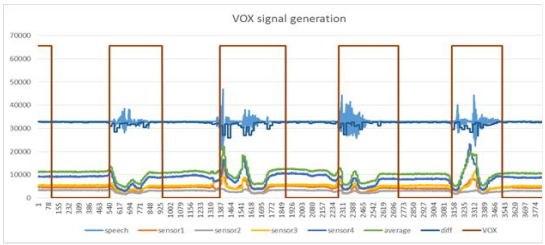


図 6　音声、光センサ出力とそのときの VOX 出力波形（sample rate = 1kHz）

Speech は音声、S_1~S_4 までがフォトリフレクタ出力を示している。S_sum が式(1)、V が式(2)でそれぞれ求めた値、VOX が EL へのスイッチング操作である。発話時に VOX 出力が立ち上がっていることがわかる。

実際にユーザに使用してもらうことを想定したとき、唇の下（右下のあたり）であれば、唇の上や前にセンサを持っていかなければならない場合よりは比較的ユーザの邪魔にならないと考える。今回はプロトタイプの作成ということで、目立つ形状の EL コントローラになったが次世代型を作成するときは、目立たない形状で唇の下にフォトリフレクタを近づける形状の設計ができる。

また現時点の問題点として、フォトリフレクタの位置をうまく固定しないと測定結果で示すような綺麗な波形が出ないことがある。所定の位置に固定する方法の検討が今後の課題である。

（2）食道発声のための拡声器の開発

拡声器の内部の構成を図 7 に示す。拡声器の箱の構成は、マイクアンプ、D 級アンプ、VOX が回路に含まれており、ニッケル水素の単 4 電池が 4 つ、2 つのスピーカからなる。D 級アンプと VOX 機能を用いることで電力消費を抑えることができる。また実際に作成したプロトタイプを図 8 に、回路図を図 9 に示す。

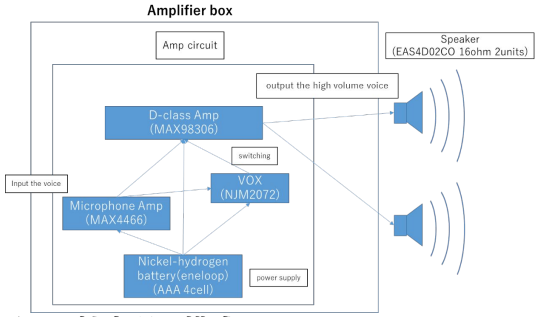


図 7　拡声器の構成

図 8 拡声器のプロトタイプ

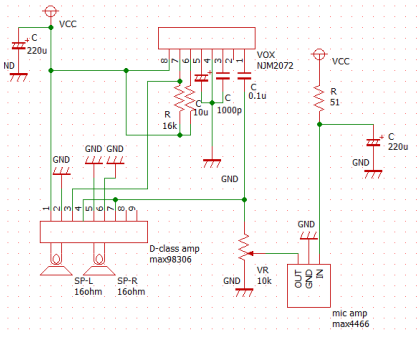


図 9 拡声器の回路

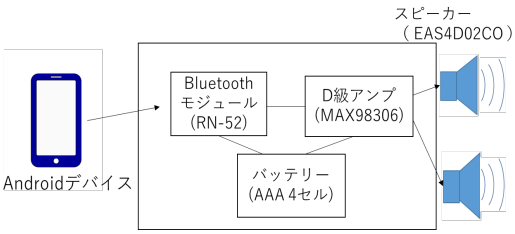


図 10 拡声器システムの概要

また本研究で製作した拡声器を実際に銀鈴会の方に使用してもらいフィードバックをいただいた。拡声器の筐体の大きさと音量については高い評価を貰い、実際にこの拡声器を今後も使いたいという意見もいただいた。またディスカッションの中では、スマートフォンアプリを用いた拡声器の開発への要望もあった。

Bluetooth モジュールを用いてスマートフォンと連携させることでワイヤレス拡声器の開発を可能にした。またマイク側にスマートフォンを用いることで食道発声に信号処理による強調ができる。

本研究では食道発声の音声全体にリバーブをかけることで母音の強調を行い、音声についての評価実験を行った。リバーブのイメージについて図 11 に示す。

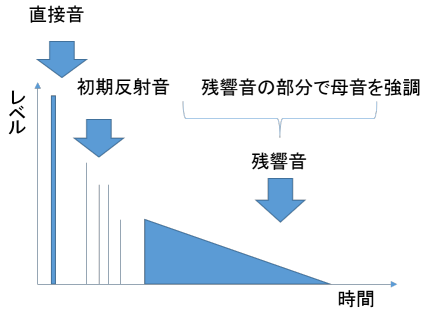


図 11 リバーブ生成のイメージ

リバーブのエフェクトをスマートフォンアプリに実装した。原理としては、入力データと出力データを格納するリングバッファを 1 つずつ用意する。ここでは、それぞれ S[ ],Y[ ]とする。録音した音声データをそのまま両方のバッファに格納する。その後、ディレイの時間の分、遅れた時刻の音声データを入力のバッファ S[]から取り出し、出力バッファ Y[ ]の現在の音声データと融合する処理を繰り返している。イメージ図を図 12 に示す。

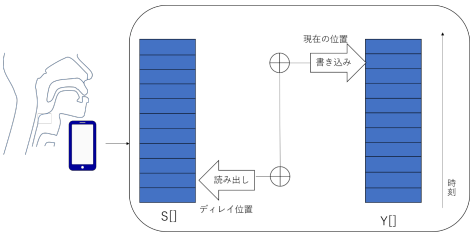


図 12 リバーブの実装のイメージ

リバーブの有無による効果を評価するために 10 人の食道発声のかたの音声から、母音が途切れ気味の音声を 2 名、母音が途切れず比較的はっきりと発声できている音声を 1 名選別した。関係性を表 1 に示す。

表 1

音声データベース	概要
食道発声 A	かなり母音が途切れている
食道発声 B	少し母音が途切れている
食道発声 C	はっきりと発声できている

シェッフェの一対比較法での解析データを図 13、図 14、図 15 に示す。
αをリバーブなし、βをリバーブ小、γをリバーブ大とする。

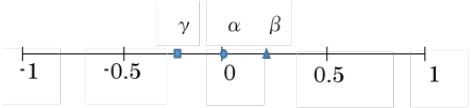


図 13 食道発声 A の結果

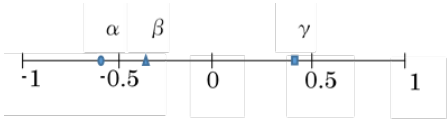


図 14 食道発声 B の結果

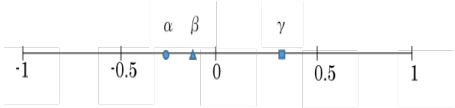


図 15 食道発声 C の結果

食道発声 A、B、C ともに有意な差は見られなかった。

リバーブの処理によって若干の聞きやすさ改善がみられたが、明瞭性に重要な広域の劣化が一部見られた。よって図 16 に示すように高域はそのままで、低域のみリバーブ処理をして合成する処理を行った。カットオフ周波数は 1kHz とした。

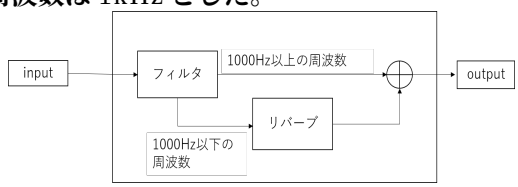


図 16 信号処理の構成

評価については同様のデータベースで行った。食道発声 A の結果を図 17、図 18 に、食道発声 B の結果を図 19、図 20 に、食道発声 C の結果を図 21、図 22 に示す。αはエフェクトなし、βはエフェクトありとする。

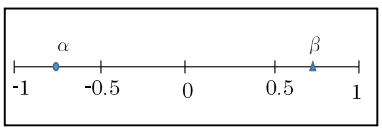


図 17 食道発声 A の結果（聞き取りやすさ）

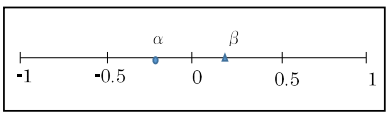


図 18 食道発声 A の結果（自然さ）
聞き取りやすさには危険度 1%で有意差がみられ、自然さには有意差は見られなかった。

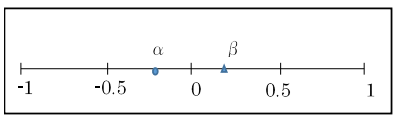


図 19 食道発声 B の結果（聞き取りすさ）

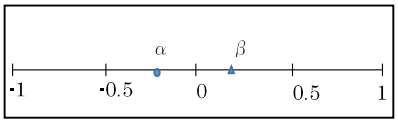


図 20 食道発声 B の結果（自然さ）
この結果には有意差は見られなかった。

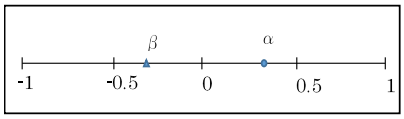


図 21 食道発声 C の結果（聞き取りやすさ）

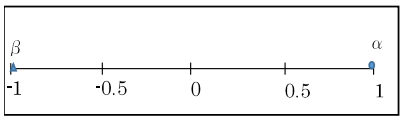
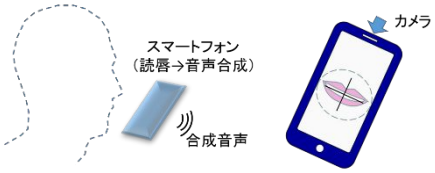


図 22 食道発声 C の結果（自然さ）

聞き取りやすさには有意差は見られなかったが、自然さには危険度１％で有意差がみられた。信号処理は食道発声の熟練度に応じて効果がみられた。今後はユーザの食道発声の熟練度に合わせ，リバーブのレベルを調整する。

（３）携帯機器と口唇情報利用による発声支援方式の基礎検討

声支援装置の全体的なシステムのイメージの概略図を図 23 に示す。このシステムの特徴として、① スマートフォンを用いることで既存のデバイスの利用が可能、② システムを使っている姿が不自然ではない、③ ディスプレイによるフィードバックで口唇認識の精度向上が期待できる。といった 3 点が挙げられる。



スマートフォンによる読唇→音声合成機能

図 23　発声支援システムのイメージ

口唇認識方式開発のための PC カメラから取入れた顔画像から口形素の認識実験を行った。現在、図 24 に示す母音推定のアルゴリズムを用いている。

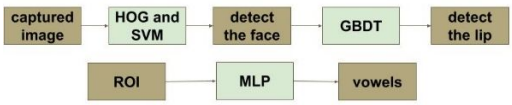


図 24　母音推定アルゴリズム

まず人の画像から、HOG 特徴量を抽出して SVM を用いて顔を検出する.次に勾配ブースティング決定木（GBDT）を用いて顔から、各部位を検出し、そこから口唇領域画像（50×100 pixel）を抽出する。最後に、抽出した部位画像（ROI）を 5000 次元のベクトルに変換して、ニューラルネットワークに入力して、“あ,い,う,え,お,無音”の 6 クラスへの分類を行い、母音を推定する。図 24 をもとに実装した母音推定システムのスクリーンショットを図 25 に示す。頭上の緑字が推定した母音であり、右上にシステム全体の処理速度をフレームレートで示している。

図 25　母音推定システムのプロトタイプ

ニューラルネットワークには「多層パーセプトロン（MLP)」「畳み込みニューラルネットワーク（CNN)」「MobileNets」３つの手法を用いた。各ニューラルネットワークの構造を図 26～図 28 に示す。

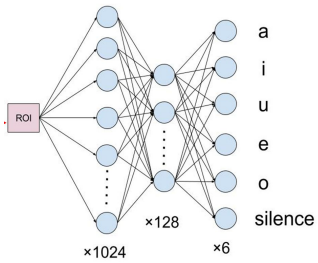


図 26　MLP の構造

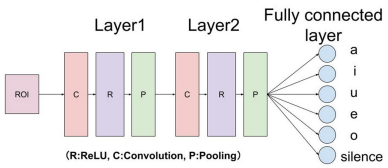
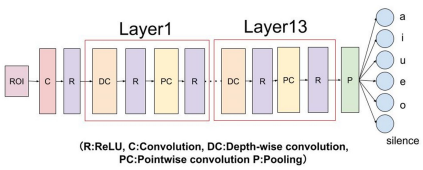


図 27　CNN の構造



$\alpha = 0.5$

図 28　MobileNets の構造

Google が提案した MobileNets の畳み込みフィルタは「depth-wise convolutions」と「pointwise convolutions」という２つのレイヤーで構成されている。通常の畳み込みを上記に置き換えることで、計算量が削減される。

各ニューラルネットワークの評価実験を行うために、5 人の実験協力者より得た 3000 枚の口唇画像（各母音 500 枚）のうち 2100 枚を学習に 900 枚をテストに用いて認識精度を求めた。また同時に 900 枚のテスト画像の予測にかかった時間を計測し、表 2 にまとめた。

表 2　認識精度と処理速度

Algorithms	Accuracy (%)	Processing time (s)
MLP	87.1	0.014
CNN	96.9	8.321
MobileNets	92.4	2.333

CNN が最も高い精度となったが処理速度が圧倒的に遅い結果となった。また、MLP は圧倒的に処理速度が速いが精度が低い。この中で MobileNets が高い精度と速い速度を保つ結果となった。

発声支援装置として用いる場合、遅延の少ない認識方式、安定した母音認識精度が必要である。今回の簡易評価実験より予想通り単純な口形素による方式では良好な母音認識は困難である。今後は安定した母音認識方式の開発、および、携帯機器によるディスプレイのフィードバックを用いた方式の開発でセグメンテーション精度の向上を検討する。また口形素以外の情報を学習させるために、連続的な画像から特徴ベクトルを抽出する方法を検討していきたい。

５．主な発表論文等

〔雑誌論文〕（計　1 件）

①Yuta Matsunaga, Kenji Matsui, Yoshihisa Nakatoh, Yumiko o. kato, “Development of Hands-free Speech Enhancement System for Both EL-users and Esophageal Speech Users”, Distributed Computing and Artificial Intelligence, 14th International Conference, pp334-341, 2017（査読有）

〔学会発表〕（計　2 件）

①松永勇太，松井謙二，中藤良久，加藤弓子，携帯機器と口唇情報利用による発声支援方式の基礎検討, 2018 年電子情報通信学会総合大会

②Yuta Matsunaga, Kenji Matsui, Yoshihisa Nakatoh, Yumiko o. kato, “Development of Hands-free Speech Enhancement System for Laryngectomies”, 2017 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing（査読有）

６．研究組織

(1)研究代表者
松井 謙二（MATSUI,Kenji）
大阪工業大学・工学部・教授
研究者番号：30613682

(2)研究分担者
中藤 良久（NAKATOH,Yoshihisa）
九州工業大学・大学院工学研究院・教授
研究者番号：10599955

(3)水町 光徳（MIZUMACHI,Mitsunori）
九州工業大学・大学院工学研究院准教授
研究者番号：90380740

(4)加藤 弓子（KATO,Yumiko）
聖マリアンナ医科大学・医学部・研究員
研究者番号：10600463