様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成 ２８ 年　 ４ 月 １４ 日現在

機関番号：３２６４４
研究種目：基盤研究(C)（一般）
研究期間：2013～2015
課題番号：２５３３０４１８

研究課題名（和文）口唇動作による発話支援装置と読唇補助装置の開発

研究課題名（英文）Development of utterance training support-equipment and lipreading assistance equipment by lip movement

研究代表者
山田　光穂（Yamada, Mitsuho）

東海大学・情報通信学部・教授

研究者番号：６０３６６０８６

交付決定額（研究期間全体）：（直接経費）　　2,900,000円

研究成果の概要（和文）：日本語と英語を対象として口唇動作による発話支援装置の開発を行った。まず、口唇動作を取得しデータベース化できる発話トレーニングシステムを開発した。次にこの装置を用いて、日本語については放送局アナウンサー、英語についてはネイティブの英語講師により、模範となる発話データベースを作成した。さらに、この装置を用いて、日本語と英語の発話トレーニングを行い、発話能力が向上することを示した。本研究の成果により、日本語や英語の発話トレーニングだけでなく、母音発話能力の向上による読唇や発話認識の精度向上に役立てることができる。

研究成果の概要（英文）：Utterance training support-equipment by lip movement has been developed targeted for Japanese and English. First, We developed automatic extraction software for the characteristic points of the lips and we created a database of utterances by Japanese television announcers and English teachers. Utterance training of Japanese and English was performed using this equipment, and showed that the utterance ability improved. It will be possible to use this system for lipreading by improvement of the vowel utterance ability and improvement of recognition rate of utterance recognition as well as Japanese and English utterance training by an outcome of this research.

研究分野：ヒューマンインターフェース

キーワード：学習支援システム　発話支援　口唇動作　発話認識

１．研究開始当初の背景

口唇動作による発話内容解析の取り組みは、我が国では西田〔1〕、菅原〔2〕の研究が代表的である。西田らの方法は、まず口唇の左右と上下の動きから、口形の動きを表す動きベクトルを求める。単語を発話した際に連続的に生じる動きベクトルの変化を元に、ニューラルネットやファジーにより、あらかじめ登録した単語から最適なものを求め発話内容を認識する。菅原らは動的輪郭モデルを提案し、唇形状を抽出して母音認識を行っている。彼らの研究により、口唇動作による発話認識の有用性が示唆された。我々は口唇動作を用いた非発声のヒューマンインタフェースが様々な機器に適用できる基幹的なインタフェースになると考え、平成２２～２４年度の科研費の支援を受け、口唇動作のフーリエ変換を行う手法を提案し、小型で処理が軽く安価な認識装置の開発を行った。

その研究成果を説明する。発話者の口唇動作から、最も動作が顕著な点である口唇の上下左右端、またそれに加えて下顎端の５点を検出し特徴点とする。発話時の特徴点の動きをフーリエ変換し、そのスペクトルとあらかじめ記録した発話との相関を求め、発話時の母音や単語を認識することができる。口唇動作に支障が無い健康な発話者では、上、左、下顎端の３特徴点のみから十分に認識可能であることを示した。

以上の成果を得て、本研究を社会的な要望の強い、言語学習者の発話訓練、聴覚障害者の読唇補助に展開することを目的として、新たにアルゴリズム研究の専門家を共同提案者に加え、本研究課題に応募した。

参考文献：(1)石井、佐藤、西田、景山：時系列口唇画像を用いた読唇のための特徴抽出と唇の動き解析、電学論 D,119,4,465-472(1999)　(2)佐々木、川村、菅原：動的輪郭モデルのハードウェア化とその読唇母音認識への応用､信学技報､VLD、VLSI 設計技術 104(509)、13-17(2005)

２．研究の目的

口唇動作の抽出により、言語学習者の発話訓練、聴覚障害者の読唇補助に役立てることができる。そこで、本研究では、この成果を発展させ、外国語学習者、朗話者の口唇動作を可視化、数値化し、より正しい発話に導く発話学習支援装置、騒音下や声を発することができない環境下だけでなく、大学等への入学者が増え対策が急がれている、教師と聴覚障害者のコミュニケーションを補助する読唇補助装置を実現し、この研究成果をより広く社会に役立てることをめざす。

３．研究の方法

研究の方法と研究成果について、研究代表者が主に行った言語学習者の発話訓練に関する研究成果と、研究分担者が主に行った聴覚障害者の読唇補助に関する研究に分けて述べる。

（１）言語学習者の発話訓練に関する研究

①　発話トレーニング装置の開発

まず口唇特徴点自動抽出および認識・トレーニング装置の開発を行った。開発には C++ 言語の Windows フォームアプリケーションとし Visual Studio 2010 を用いて開発した。顔認識には seeingMachines 社の顔認識ソフト faceAPI を用いた。faceAPI は Viola らの研究(3)を基に高速かつ高精度な顔認識を行うソフトウェアである。取得できるデータは顔の向きや位置だけではなく目、鼻、口、眉などの情報も取得でき、口唇の点だけでも上唇と下唇がくっつく内側の 8 点、唇と皮膚の境界で 8 点も取得でき計 16 点も取得できる。多くの点が取得でき、かつ、30fps でデータを取得できるソフトウェアであることから本研究では faceAPI を用いることとする。faceAPI の精度は外眼角間の距離が最小の 40 画素である場合 1cm 以下となる。本研究の撮影条件では外眼角間距離は最小値の 4 倍以上であるため口唇動作を十分な精度で取得可能であることが考えられる。また、faceAPI の処理速度は、CPU が Core 2 Duo で 2.4GHz 帯を使用する場合、頭部運動が生じない条件で 0.3 秒であることが示されている。本研究で用いる機器は、Core i7 の 1.9GHz 帯を使用しているため唇の動作を検出するための十分な処理速度を有している。

②　教師用日本語・英語発話の取得

日本語のデータは発話訓練用の本「声がよくなる簡単トレーニング」(4)の中から「あ」から「わ」の発話トレーニングの文のうち 69 文を抜粋し取得した。また、より長い文と短い文を取得するためアナウンサーの発話練習にも使われている文章である小学校 3 年生の国語教科書に含まれている「たんぽぽ」(5)の文章を文節で区切ったデータ及び句読点で句切ったデータを取得した。アナウンサーが文章を読みやすいように文章は縦書きにした。例として図 1 の文章を使用した。

英語のデータはテレビ講座のテキスト「3ヶ月トピック英会話」(6)に取り上げられた日本人に発話の難しい母音や子音を含む単語と文章を用いた。例を表 1 に示す。全部で文章は 12 個、単語は 65 個である。日本語の発話には NHK アナウンサー、英語には英語を母国語とする本学の英語教師に依頼した。

参考文献
(4)福島: 声がよくなる簡単トレーニング; 成美堂出版, (2006)
(5)　改訂新しい国語二上 たんぽぽ改; 東京書籍, (1985)
(6)　3ヶ月トピック英会話 話して聞き取るネイティブ発音塾; (2009,1～2009,3), 日本放送出版協会, (2009)

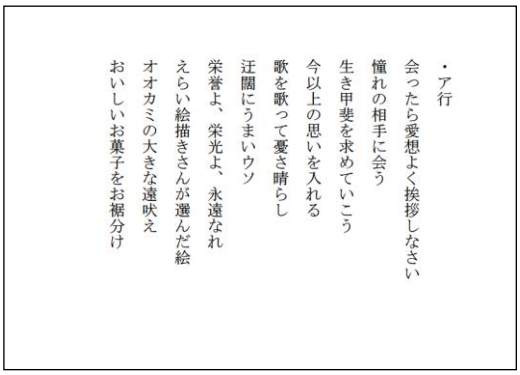


図 1. 日本語発話トレーニング用の文章例[4]

表 1. 英語トレーニング用文章と単語例[6]

単語			
long	road	cup	cat
right	left	look	book
文章			
The Long Road to Little Rock.			
She saw a Sinking Ship.			
The Cop Found the Cup.			

③　トレーニング装置の概要

顔画像を取得するカメラを指定しカメラの解像度やフレームレートなどを設定する。次に顔認識を行い口唇の特徴点データを取得する。取得した口唇特徴点を時系列に並べ、上下端の差を用いてある程度口が開いたら発話区間の始めと設定し口が閉じ終わり停止するまでを発話区間とする。その後比較をするために表示の方法を選択する。表示方法として模範の口の動きと発話訓練者の口の動きを重ねて表示する方法、上下に配置し口の左右の動きの違いが分かりやすくする方法、左右に配置し口の上下の動きの違いを分かりやすくする方法が選択でき、さらにベクトルを用いてより詳しく口の動きの差がどれくらいあるのかを表記できる。

④　取得した発話の例

図 2 はアナウンサー、図 3 は発話訓練者が「会ったら愛想よく挨拶しなさい」と発話した時の動作履歴である。図中の縦の破線は文章を文節で区切っている。図 4 は英語教師、図 5 は発話訓練者が右という意味の「right」を発話した際の動作履歴である。図中の縦の破線は発話区間を示している。

それぞれ縦軸が移動量で横軸が時間となっている。移動量が増えた時は口が開き、減った時は口を閉じた状態となる。

アナウンサーと発話訓練者の動作履歴を比較した結果、「あったら」と発話する際、

アナウンサーは 1.2 秒かけて発話しているのに対し、発話訓練者は 0.8 秒で発話を行っておりアナウンサーの発話の方がゆっくりである。また、「あ」について比較するとアナウンサーでは「あったら」「あいそよく」「あいさつ・・」と順に下唇特徴点の動作が大きくなっているのに対して、一般大学生にはそのような特徴は見られない。

英語教師と発話訓練者の発話時間は"right"では 0.5 秒前後であり発話時間の差はない。しかし英語教師には、発音が始まる前に一度口を軽く開き予備動作と考えられる口唇の動きが取得できた。

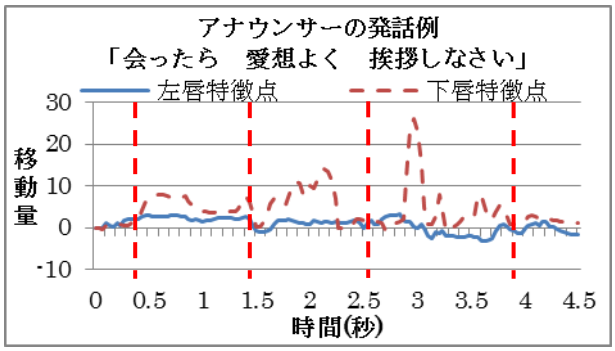


図 2. アナウンサーによる発話例

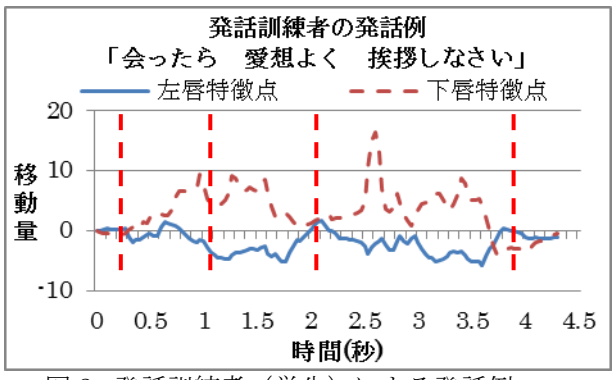


図 3. 発話訓練者（学生）による発話例

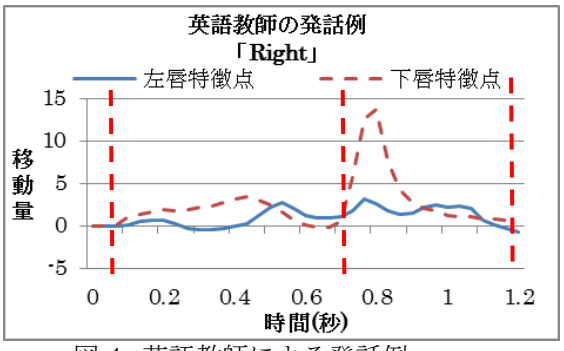


図 4. 英語教師による発話例

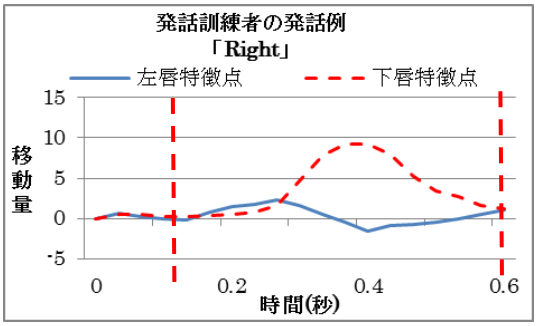


図 5. 発話訓練者（学生）の発話例

⑤　トレーニング手法

図 6 はデータを取得する画面である。まず、

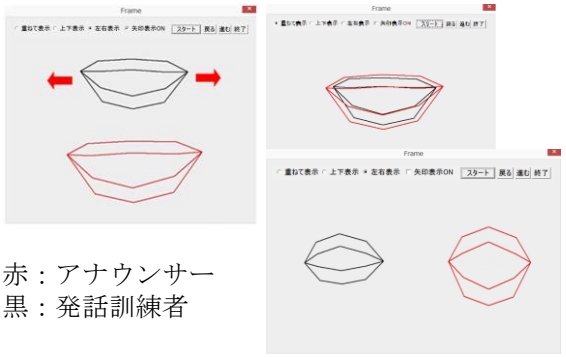
トレーニングの文章を選ぶために辞書選択を行う。次に、データを保存するために被験者番号とデータの番号を設定する。データの取得は認識開始ボタンを押し発話を行い、発話が終わったら認識停止ボタンを押してデータの取得を終わらせる。トレーニングボタンを押すと取得したデータのトレーニングを行うことが出来る。

図 7 はトレーニング表示画面である。上部に「重ねて表示する」「上下に表示する」「左右に表示する」を選択できるラジオボタンがあり、その横に矢印を表示するためのチェックボタンを配置している。スタートボタンを押せば口唇動作をアニメーションとして表示でき、また、戻ると進むボタンで 1 コマずつ発話を確認することが出来る。黒線が訓練者の取得データ、赤線が模範データである。

図 6. データ取得画面

⑥　日本語の検証

発話トレーニング装置を用いてアナウンサーの口唇動作データと比較して被験者に 10 回練習させた．被験者は本学の 20 代の学生男性 3 名、女性 1 名である。さらに発話トレーニングを行った被験者 4 名とは異なる新たな被験者から男性 8 名、女性 2 名の計 10 名により一被験者あたり 4 回、合計で 16 回の音声比較を主観的評価によって行った。比較を行うにあたって、アーティキュレーション(歯切れの良い発音)、声の速さ、声の大きさの 3 つを比較項目として評価を行った。



赤：アナウンサー
黒：発話訓練者

図 7. トレーニングの表示画面

⑦　日本語の検証結果

図 8 は「歌を歌って憂さ晴らし」と発話し

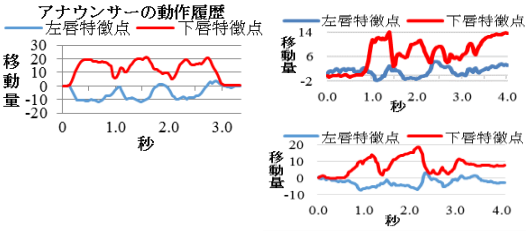


図 8　アナウンサーと被験者 A の動作履歴

左 : アナウンサー　右 : 被験者 A　上 : 1 回目　下 10 回目

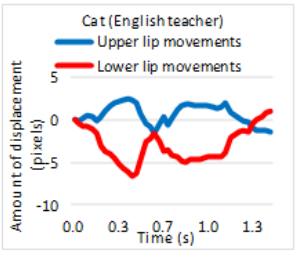
た際の動作履歴の例である。はアナウンサーの動作履歴と比較すると、被験者 A の 1 回目の発話では不安定な動作であるのに対して、トレーニングを経た 10 回目の発話では振幅の大きさなどがアナウンサーの発話に動作履歴にかなり近くなっていることが分かり、このトレーニング法の成果が示された。

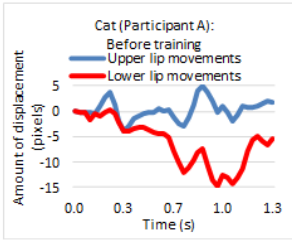
⑧　トレーニングの効果・英語

トレーニングに使用した英単語はデータベースの中から、短母音を含む英単語 5 つ（cat、 pet、 little、top、cup）を抜粋した。被験者は本研究室から 20 代の学生 10 名（男女各 5 名）に協力して頂いた。トレーニングは計 10 回行った。

⑨　英語の主観評価結果

ここでは“cat”の結果を代表例として述べる。図 9 は英語教員が”cat”と発話した際の口唇動作履歴を示したグラフとトレーニング前の被験者 A が”cat”と発話した際の口唇動作履歴、トレーニング 10 回目の被験者 A が”cat”と発話した際の口唇動作履歴を示したグラフである。”cat”は発音記号では/kæt/と表される。英語教員は/t/と発話する前に一瞬口を閉じていることが動作履歴からわかる。また、/kæ/の部分と/t/の部分でほぼ同様の口の開き具合であることが分かる。これは，/t/が破裂音の子音であるためである。/t/の発音は上あごの裏に舌を当てて発話する子音であり、口を閉じたまま発話することはできない。一方、被験者 A はトレーニング前では口の開きは英語教員と同じくらいであるものの、口唇動作が安定しておらず，英語教員のような一瞬口を閉じる動作も見られない。しかし、トレーニング 10 回目では一瞬口を閉じる動きが見られたため，発音が改善したと考えられる．また，被験者 A のトレーニング 10 回目での/t/の部分の発話時間が英語教員に比べて短く、上唇下唇ともに大きく開いていなかった。その要因として、子音の/t/の発音が日本語にはないため，トレーニング前ではそのような特徴を考慮することが出来なかったのがトレーニングによって改善されたと考えられる。





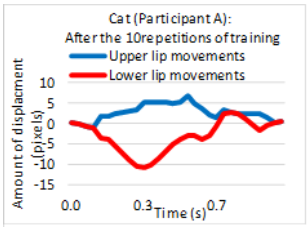


図 9　"cat"と発話した時の動作履歴
左：英語教員
右：被験者 A
上 1 回目下 10 回目

⑩. 英語の客観評価結果

大塚は, 音声のスペクトログラムから第 1、第 2 フォルマントを検出し、周波数値をもとに具体的で細かな母音発音指導ができると報告している(7). 音声の第 1 フォルマントと第 2 フォルマントをプロットすることで、発話時の口の開け方と舌の位置を表すことができる IPA（International Phonetic Association）チャートというものが世界の英語教育で用いられている。第 1 フォルマントは舌の高さを表しており、第 2 フォルマントは舌の位置を表している(8)。これらを参考に我々は考察を行った。ここでは短母音/æ/の結果を例として示す。図 10 にトレーニング前と 10 回目の英単語"cat"に含まれる短母音/æ/の被験者別フォルマント散布図を示す。赤い三角が日本語の「あ」、青い三角が英語の/æ/、そしてピンクの三角が 10 人の被験者の平均の位置である。また、A~J の被験者をそれぞれ色を変えて○でプロットした。

10 回のトレーニング後、平均位置が英語の平均値に近づいたことが分かる。しかし、10 回のトレーニングを行っても、被験者 D や E のように発音の改善が見られない者もいた。また、英語のフォルマント値とトレーニング前，5 回目，10 回目のフォルマント値の平均値の 2 点間距離を求め図 11 に示す。トレーニング前からトレーニング 10 回目にかけて値が小さくなった。t 検定を行った結果、トレーニング前と 10 回目の平均値の 2 点間距離について有意差が認められた(p=0.02< 0.05)。トレーニング前から 10 回目にかけて英語らしい発音に近づいたと考えられる。

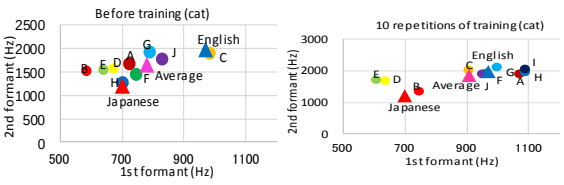


図 10　短母音/æ/のフォルマント散布図（トレーニング前　左、10 回目　右）

参考文献

[7]大塚、"フォルマント周波数値を利用した母音発音指導の可能性についての一考察"、東京女子大学紀要論、第 64 巻 2 号,pp.311-333,2014.

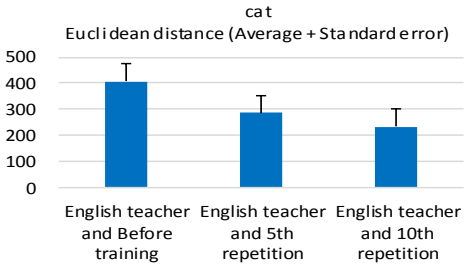


図 11　各指標のフォルマントの 2 点間距離

[8]今泉他、"音声学・言語学、"今泉敏(編),(社)医学書院,2009.

（２）　読唇補助装置開発に関する研究

研究代表者が主に行った発話トレーニング装置の研究成果を読唇補助装置に発展させるのが共同研究者の研究目的である。読唇補助装置は、対象とする単語を特定の分野の言語に限ることにより容易に実現することができる。しかし、朗話者の読唇支援、音声認識の補助手段となる読唇支援を考えた場合、分野を限定するとその適用範囲が限られてしまう。より広い分野の単語を対象とした場合、これまでの手法では単語数を拡張していくことは難しいことがわかった。そこで、より多くの単語を扱え、随時拡張できる口唇動作履歴データベースの開発に注力した。

①　SQLite の組み込み

我々の開発した口唇動作を用いて声による発話認識を補助するシステムでは辞書および学習者の口唇動作に関するデータは全てCSVファイルで管理されている。CSVはテキスト形式であるため読み書きが簡単ではあるが、辞書やユーザデータの更新時に間違ったファイルに上書をしてしまう可能性がある。また、単語や文の数が増えれば、それだけCSVファイル数も増加し、辞書数そのものが増えるとディレクトリ数が増加する。また、プログラムのコーディング量が多く、読み書きに時間もかかってしまい、使い勝手が良いとは言えない。そこで、このシステムにDBMSの1つであるSQLiteを組み込むことにした。

②　データベースの導入

先述の発話練習に使用する辞書データはフォルダで管理され、辞書に含まれる 1 単語ごとに 1 つのサブフォルダを作成し、複数の CSV 形式ファイルで保存している。
辞書と単語の数が非常に多くなった場合、データ管理やアクセスに問題が発生しやすくなる。CSV ファイル形式のままでは、誰でも簡単に書き換えることができてしまい、ファイルの数が増加し過ぎると、入出力処理に遅延を引き起こす可能性もある。

また、我々の既存の発話トレーニングアプリケーションは、最新の口唇動作履歴を格納し、アプリケーションの画面に線画によって、母音発音のための唇の動きを再現することができた。しかし、この助言は最新の発話に対しての

みの提供である。学習者が継続的に発話練習を繰り返すのであれば、具体的なトレーニング結果のフィードバックができるようにするべきである。このような場合、CSV ファイルを使用するとデータ処理が非常に煩雑になり、時間がかかってしまう。そこで、この問題を解決するために SQLite を導入した。

今回は使用可能な辞書名を登録した辞書マスターDB と各辞書の DB を分けて作成することとした。従来の方法よりもファイル数が減り、およそ 4 分の 1 となる。また、DB にしたことにより、ユーザが誤って内容を変更してしまう可能性も少なくなる。辞書 DB のテーブル構成例を表 1 に示す。ユーザのトレーニング情報も同様に DB 化することで、ファイル数はおよそ 4 分の 1 となり、さらに学習履歴の抽出等も簡単にできるようになる。また、これによって、プログラムのコーディング量もオープンするファイル数減少により、データ入出力部分は 4 分の 1 程度になる。

表 2　辞書テーブルの例

表 2.1　マスター辞書

DicID	DicName
001	Odakyu
002	Tampopo
:	:

表 2.2　辞書「小田急」のマスター辞書

WordID	Word
001	Hadano
002	Kakio
:	:

表 2. 3　辞書「小田急」の中の「柿生」のパワースペクトラム

left	right	top	bottom
4.1948	0.0036	10.9897	436.3297
11.2565	0.0057	4.7877	2.9588
6.2338	0.0004	4.6497	55.9463
0.4302	0.0021	0.4758	1.1813
0.1987	0.0031	0.4919	0.2668
0.0722	0.0006	0.7744	1.2403
0.0375	0.001	0.2132	0.0455
0.0054	0.0001	0.0105	0.2448
:	:	:	:

4　研究成果

日本語と英語を対象として口唇動作による発話支援装置の開発を行った。本装置の実現にあたり、まず模範となる教師データを効率的に取得する必要があると考え、口唇動作を取得しデータベース化できる発話トレーニングシステムを開発した。次にこの装置を用いて、日本語については放送局アナウンサー、英語についてはネイティブの英語講師により、模範となる発話データベースを作成した。さらに、日本語と英語の発話トレーニングを行った。日本語では聞き取りやすく話すアナウンサーの発話時の強弱変化が数値化され、トレーニング者がこれを学ぶことにより、聞き取りやすさや歯切れの良さ（アーティキュレーション）が改善されることを示した。英語についても、トレーニングを重ねることにより、口唇動作履歴が英語講師に近づくことを示した。さらに、音声のスペクトログラムからを用いた客観的評価法を提案し、発話能力が向上し、ネイティブの母音発音に近づくことを示した。読唇補助装置の研究では、これまで特定のジャンルに限定して研究を行ってきたが、ニーズの増加に伴い、データベースの修正、拡充に柔軟に対応するシステムの開発が必要となった。そこで、SQlite をベースとする新たなデータベース管理手法を提案した。これにより、より多くの単語に対応した汎用的な読唇補助装置を実現できる。以上に述べたように、本研究の成果により、日本語や英語の発話トレーニングだけでなく、母音発話能力の向上による読唇や発話認識の精度向上に役立てることができる。

5．主な発表論文等

〔雑誌論文〕（計２件）投稿中

〔学会発表〕（計 15 件）

Yuko Hoshino, Tomoki Yamamura, Mitsuho Yamada, The learning data management of the utterance learning system using lip movements recognition,2015ICCAT,118-125,Matsue

Tomoki Yamamura, Miyuki Suganuma, Eiki Wakamatsu, Yuko Hoshino and Mitsuho Yamada Development of a speech training system by lip movements,2015ICCAT,62-65,Matsue

Miyuki SUGANUMA Tomoki YAMAMURA Yuko HOSHINO and Mitsuho YAMADA, How to evaluate English pronunciation learning by lip movements,IMQA2016,32-37,Nagoya

（他に 12 件）

〔その他〕ホームページ

http://www.yamadalab.net/UtteranceTrainingByLipMovement_JPN.htm　日本語

http://www.yamadalab.net/UtteranceTrainingByLipMovement_ENG.html　英語

6．研究組織

(1)研究代表者

山田光穂（Yamada Mitsuho）
東海大学情報通信学部情報メディア学科・教授　研究者番号：60366086

(2)研究分担者

星野祐子（Yuko Hoshino）
東海大学情報通信学部情報メディア学科・講師　研究者番号：80435271