東海大学紀要情報通信学部
vol.5,No2,2012,pp.36-44

論文

口唇特徴点動作のパワースペクトルの相関を用いた非発声の発話認識手法の可能性に関する検討

齋藤 翼[*1,2]，大城 政人[*1]，尾上 拓[*1,3]，笠原 篤之[*1,4]，新川 達矢[*1]，山田 光穂[*5]

A Study of an Utterance Recognition Method using Correlation of Power Spectrum of Characteristic Lip Movements

by

Tsubasa SAITO[*1,2], Masato OSHIRO[*1], Taku ONOUE[*1,3], Atsuyuki KASAHARA[*1,4], Tatsuya SHINKAWA[*1] and Mitsuho YAMADA[*5]

(received on September 28. 2012 & accepted on February 7, 2013)

Abstract

Currently, many voice recognition technologies are available; however, the use of these technologies in noisy environments is difficult. Therefore, a wide range of techniques for individual authentication and utterance recognition using lip movements without voice data have been proposed and are achieving excellent results. To date, verification of these methods has involved only specific words, and we believe more general verification is needed to aim for practical use. In our proposed method, the spectrum of lip movements during words utterance is calculated by Fourier transforms, and the correlation of the spectrums between uttered words is estimated. Using this method, we studied the possibility of recognizing words with the same number of characters that were considered likely to be difficult to recognize. Furthermore, we analyzed the correlation of the power spectrum when the same speaker said the same words but on different days and at different times, and our results showed strong correlation between these power spectrums. We also introduced difficult examples of recognition by correlation of the power spectrum using words with similar or matching vowel sounds. Finally, we discuss the validity of our proposed method to recognize words only by lips movements without saying the words aloud.

Keywords: *Lip-movement, Voice Recognition and Human interface*
キーワード: 口唇動作、発話認識、ヒューマンインタフェース

1. まえがき

意思伝達するためのコミュニケーションの手法として，文字や絵・発話などの様々な手法が用いられている．その中でも発話は、最も有用なコミュニケーションの手法として位置づけられている．そして，この発話における音声情報や口唇形状の変化は，個人毎での特徴を有しており，発話認識および個人識別を行う上で有効なパラメータとして重要視されている．それゆえ，昨今では発話認識を用いた工学機器（カーナビゲーション，バイオメトリクスなど）の普及が進み，音声情報が手入力に変わる新たなスタンダードとして注目されている．しかし，発話認識技術の現状は，静穏環境下という状況を前提として使用されなければ，高い認識率を得ることが困難である．これは，外部からのノイズなどが対象音声への干渉を行うことが主な要因であり，環境によって大きく異なって変化するノイズには，定型化されたノイズフィルター等の処理を行っても，対象音声の抽出・ノイズ除去することは困難である．また，一般的な発話認識のソフトウェアは，静的環境内のユーザーが発話認識を意識して発話を行い，雑音を含まない音声情報が得られた場合に良好な認識率を提示するが，前述したような雑音を含まない音声情報の取得が困難な環境下におかれると，対象の音声情報に干渉がかかり，認識率の低下を引き起こしてしまう．よって，発話認識を幅広く活用するためには，対象音声が外因ノイズの影響下にある状況であっても，発話内容を抽出することが重要である．そこで，外因ノイズの性質に干渉されない発話認識のために，口唇形状の発話に伴う視覚的時間変化（以下，口唇動作と表現する）に注目し，個人認証[1)~5)]，読唇による発話認識[6)~7)]，コミュニケーション支援[8)]などを目的とした研究が行われている．

口唇研究を進めるうえでは，口唇部の抽出，口唇動作の数値化，口唇動作の評価という３つのステップが必要である．

口唇部の抽出法としては，色差情報を用いて肌色部分から口唇部分を抽出する手法がよく用いられている[9),10)]．

口唇動作の数値化には，口唇左右上下の時系列的な動きを図る方法[1),6),8)]，口唇の縦幅、横幅、周囲長を求める方法[2),3)]．フーリエ記述子により唇形状を抽

*1 工学研究科情報理工学専攻 修士課程
Graduate School of EMG, Master's program
*2 現・東海旅客鉄道株式会社
Central Japan Railway Company
*3 現・株式会社 リアルビズ
Present RealViz,Inc.
*4 現・株式会社 アイ・ディ・ケイ
Present IDK,Inc.
*5 情報通信学部情報メディア学科 教授
School of Information and Telecommunication Engineering, Professor

出する方法[4]等がよく用いられているが，多くの例では照明条件等による変動を除くため，手動による不良データの除去も併用されている．輪郭や周囲長の抽出には，動的輪郭法（Snakeなど）を適用した例が多い[6),9)]．

口唇動作の評価法としては，口唇左右上下の時系列的な動きからマハラビノス距離を用いて個人識別する手法[1)]，口唇の縦幅，横幅，周囲長等のパラメータから部分空間法を用いて識別する手法[2)]，口唇動作の時間的変化からトランジェクトリ特徴量を求め、口唇の横幅と縦幅から要素ベクトルを求めDPマッチングで類似度を検出する手法[3)]，口唇特徴点の動作履歴のパワースペクトルの相関を比較する我々の手法[7)]，DPマッチングにより認識する手法[9)]など，多くの提案が行われ口唇動作を用いて個人認証や発話認識が可能なことが示されている．

実際に個人認証や発話認識に用いられる単語は，個人認識については，あらかじめ決められた単語を発話させ検討したもの[5)]が多く，発話認識では，病院で用いられる限定的な単語[8)]や，前進，後退，止まれなど車椅子の制御に用いられる単語[9)]など特定の単語認識に焦点を絞ったものが多い．

以上に述べた様な研究から，口唇動作を用いて，個人認証や発話認識を行うことは極めて有望な手段であることが示唆される．しかし，いずれも特定の単語に限って検証が行われており，実用化をめざすにはより汎用的な検証が必要である．そこで，我々は，あえて同じ語数のみを用いたとき認識可能か，同じ個人でも日にちや昼夜など発話時間を変えたときに変動が生じるか，また，本方式で認識が困難な単語対はあるのかすなわち本方式の限界について検討することを目的とした。

対象とする単語としては、同じ語数の単語が大量にあり，その中から無作為に選択することが可能な和歌に着目した．このような理由から，本稿では百人一首を用い，上述した項目を検討した．

今回の提案では，口唇動作による鉄道の駅の券売機の駅名入力など，騒音下で視覚障害者や高齢者が快適に使用できるヒューマンインタフェースの開発を念頭においている．鉄道の駅名では，同一語数の駅名も多い．また，不特定多数の人が利用する．語数が同じでも，認識できる可能性があること，同一発話者がいつ発話しても認識できる可能性があることを示すことは，本提案を実用に結びつけるために重要な要素となり，これまでの研究に比べ新規性があると考えている．他方，認識が困難な例を検証する事により，鉄道なら沿線の駅情報など，他の知識情報を組み合わせる手法の開発に役立てることができる．

2. 母音認識に用いる各口唇の特徴点

発話に伴う口唇の主な動きは，上下方向の開閉および左右方向の伸縮に二分されている[1)]．そこで本論文では，口唇特徴点としてFig.1に示す口唇部の上下左右端の４点と下顎端の１点を含めた計５点に注目した．

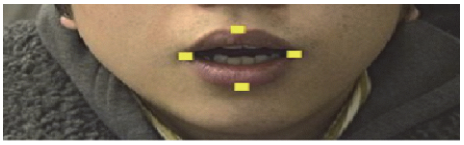

Fig.1　Characteristic points of lip.

これらの口唇特徴点は，各母音発話時に口唇が最も動作し，以下のような動作をする．

「あ」: 左右特徴点間はほぼ動作せず，上下特徴点間・下顎は非常に大きく動作する．

「い」: 左右特徴点間は大きく開くが，上下特徴点間・下顎はあまり大きく動作しない．

「う」: 左右特徴点間はすぼむため縮まり，上下特徴点間・下顎はあまり大きく動作しない．

「え」: 左右特徴点間は大きく開き，上下特徴点間はあまり大きく動かず・下顎は動作する．

「お」: 左右特徴点間はすぼむため縮まり，上下特徴点間・下顎は大きく動作する．

これら特徴点の動きの違いから母音の分別を行い，発話内容を読み取る．

3. 提案手法

提案手法のシステム構成を Fig.2 に示し，手順の概要を以下に述べる．

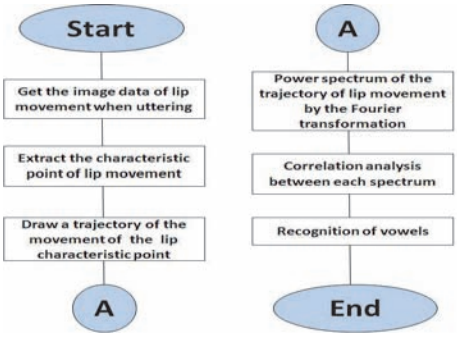

Fig.2　System configuration.

(1)　被験者に指標とする母音を発音させ，その発音時の口唇動作を Web カメラ（Logicool 製：QuickCam for Notebooks Pro）を用いて撮影し，発音時から発音終了までの発話全体の動画像を得る．

(2)　得られた動画像を一秒当たり 30 フレームの BMP 形式の画像に分割し，Fig.1 に示す各口唇特徴点の座標点を以下に示す基準を用いて手動で取得する．

①　口唇特徴点は外側口唇輪郭に注目し，口唇部の赤色と肌色の境目の画素を特徴点とする．

②　特徴点の取得において，取得者は一人のみとし，①の基準を統一して習得を行う．

(3)　発話開始のフレームで得られた口唇特徴点の座標を基準とし，各フレームの座標値から発話開始のフレームの座標値を引くことにより，口唇特徴点座標の移動量を求める．これは解析をする口唇特徴点ごとで行う．

(4)　(1)～(3)で得られた口唇移動量の時系列データを動作履歴とし，その動作履歴全体を関数と見て，

フーリエ変換を施し，その関数におけるパワースペクトルを計測する．フーリエスペクトルは，時間関数をフーリエ変換したもので，周波数の波の振幅幅を表す．これを二乗したエネルギーがパワースペクトルである．

(5)　一回目および二回目の発話データから得られるパワースペクトルの出力量から相関係数を計測する．解析する各口唇特徴点における相関が強相関である場合，各周波数で得られるスペクトルの出力量が似通うため，同一の発話をしたものと仮定し，逆に一点でも強相関が得られない場合は，別の母音列を発話したものと仮定した．

4. 使用データと実験環境

本研究の目的は語数が同じ母音列で認識できるかどうかを検証することである．そこで，実験に用いた発話単語は百人一首の上の句５文字である．これらの単語は，母音列数（５文字）が統一されており，語数が統一されていることで，認識区別の要因が各母音発話時の口唇動作の違いによるものと断定できる．また，実験者の作為なく多数の母音列を容易に取得できたため採用した．

解析する特徴点は口唇左端特徴点・下顎端特徴点の２点に注目した．これまでの我々の研究から(7)，口唇の左右端特徴点は，「あ」，「い・え」，「う・お」の特徴点動作に分類でき，他の特徴点より細かく分類できる．同様に下顎部特徴点は，他の特徴点と比較して大きく動作し，「あ・え・お」と「い・う」の特徴点動作に分類できる．左特徴点部の動作からでは区別することが困難な「い・え」と「う・お」の分類が下顎端特徴点では可能であり，この２点の特徴点における解析を組み合わせることで，各母音「あ・い・う・え・お」全ての区別がつきやすいことが示されている．

本実験で用いた発話時の動画像データの取得環境を以下に示す．

(1)　被験者は健康で発話に支障のない２０代の本学男子学生３名で行う

(2)　一般的な白色蛍光灯による照明環境の下，口紅を塗布せず，髭はない

(3)　Web カメラの解像度は，携帯電話などで普及している QVGA サイズ(320*240)とする

(4)　発話の際，動作する下顎端特徴点を画面内に捉えつつ，顔を大きく捉えて解像度をあげるために，話者とカメラまでの距離を約 50cm に固定し，話者にはカメラの正面を向いてもらい，発話中は頭を動かさない

(5)　一母音を発話するごとに口を閉じてもらい，はっきりと口を開けて発話をする

(6)　一回目の上の句を発話し終えたら，数秒のインターバルを設けた，引き続き同一の上の句を発話してもらう．

尚，この実験結果で報告する発話単語の上の句は，実際には百人一首のほとんどの上の句を用いたが，分析結果としてはその一例を挙げ，各々の検討項目ごとに解析結果を記載する．

5. 実験結果および検討

5.1　同一被験者の同一上の句発話時の検討

ここでは，話者が同一の上の句を発話した際の動作履歴グラフ・口唇動作スペクトルおよび相関関係を示す．ここで示す上の句の母音列は，左端特徴点における動作履歴に差異があると推測される上の句，「わびぬれば」，「あまのはら」，「もろともに」の３つである．検討したデータの総数は被験者３×上の句の数３×２回の 18 個である．

Fig.3-a および Fig.3-b は左端特徴点および下顎端特徴点における，上の句「あまのはら」を発話した際の口唇動作履歴グラフを示したものであり，Table 1 は他の被験者の発話も同様であったので，うちの１人の例について動作履歴について，同一の上の句発話時の一回目と二回目の相関係数を表したものである．上の句発話時は Fig.3-a,b から分かるよう，各特徴点で固有の動作をもつことが見受けられた．しかしながら，同一人物による実験でありながら，発話のタイミングおよび発話区間は異なり，特徴点の移動量も変化するため，動作履歴そのものでは Table 1 から分かるように安定して高い相関係数を得ることは難しい．Fig.4-a 及び Fig.4-b では，Fig.3-a および Fig.3-b に示した各特徴点での動作履歴から得られるパワースペクトルを示したものであり，各周期で得られるスペクトルの出力量は異なるものの，ピークを持つ周期および似通うスペクトル出力の比をもつことが分かる．Table 3,Table 4 は左端特徴点、下顎端特徴点のパワースペクトルの相関係数を求めたものである．この表からパワースペクトルの相関係数は同じ上の句間では高い値を示すことが分かる．

ここで示したように、パワースペクトルを求めることにより，各母音の発話タイミング，発話時間の長さの影響を除去し，口唇動作の変化量だけを抽出することができる．パワースペクトルの値は周波数の増加とともに，高周波の口唇動作を示し，これまでの研究結果から，ほぼ 7Hz で収束することが分かっている．

なお，ここで述べる相関係数は母音単位ではなく，上の句を構成している母音を総合したパワースペクトルの相関係数である．また，動作履歴には子音の発音も含まれるが，子音は主に声帯の気流制御もしくは気流振動と，歯，舌，唇によって呼気を制御する調音によって実現される．[p]の様な破裂音を除き，口唇の動きには反映されない．それゆえ，子音の影響は少ないと考え，ここで求めたパワースペクトルは上の句発話時の母音列のパワースペクトルと見なして，相関係数の計算を行った．

しかし，下顎端特徴点の動作履歴のパワースペクトルを示す Table 3 の「わびぬれば」と「あまのはら」間に示すよう，特徴点の動作が似通うことで，異なる母音列であっても高い相関係数を得ることがある．この要因は，下顎端特徴点は左端特徴点とは異なり，

動作軌跡が一方向に，どの程度動作するのかという特徴量しか持たないため，口唇の動作履歴が似通いやすいからである．したがって下顎端特徴点では高い相関係数を示しているが，左端特徴点では低い値を示している．

また，百人一首上の句には，Table 4 の「あまつかぜ」と「やまかわに」に示すよう，左端特徴点では高い相関係数を示すが，下顎端特徴点では低い相関係数を示す上の句同士もある．これらの結果から，一点のみの特徴点解析からでは誤認を引き起こす可能性が生じてしまうが，二点の特徴点解析を併せて用いることにより，誤認を回避した発話認識が可能であると考えられる．

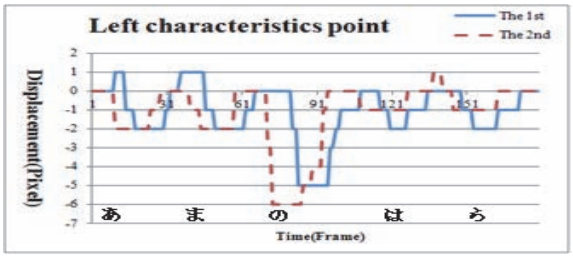

Fig. 3-a　Trajectory of lip movement when uttering Amanohara by subject1. (Left characteristics point)

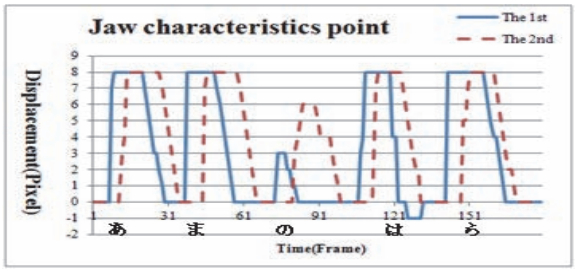

Fig. 3-b　Trajectory of lip movement when uttering Amanohara by subject1.(Jaw characteristics point)

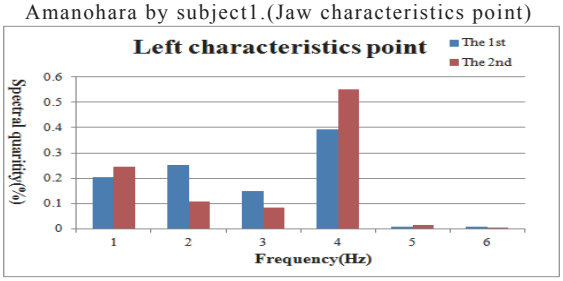

Fig. 4-a　Power spectrum of the trajectory of lip movement when uttering Amanohara by subject1.(Left characteristics point)

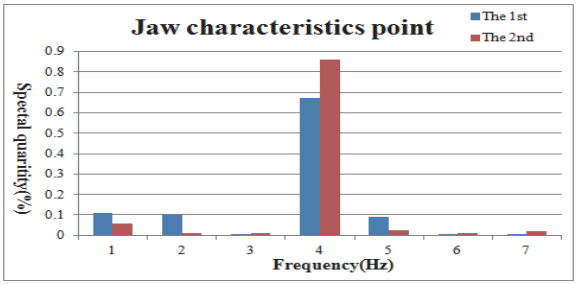

Fig. 4-b　Power spectrum of the trajectory of lip movement when uttering Amanohara by subject1.(Jaw characteristics point)

Table 1　Correlations of trajectory of the left characteristic point when uttering same phrase twice

Correlations of trajectory of the left characteristic point when uttering same phrase twice			
	Wabinureba	Amanohara	Morotomoni
Subject1	0.404	0.404	0.683

Table 2 Comparison of the correlations of power spectrum when uttering three phrases for subject1. (Left characteristics point)

Correlations when uttering three phrases (Left characteristics point)			
	Wabinureba	Amanohara	Morotomoni
Wabinureba	0.983	0.098	-0.093
Amanohara		0.923	-0.201
Morotomoni			0.994

Table 3 Comparison of the correlations of power spectrum when uttering three phrases for subject1. (Jaw characteristics point)

Correlations when uttering three phrases (Jaw characteristics point)			
	Wabinureba	Amanohara	Morotomoni
Wabinureba	0.953	0.842	-0.096
Amanohara		0.972	-0.147
Morotomoni			0.996

Table 4　Correlations when uttering [Amatsukaze] and [Amanohara] of each characteristics point.

Correlations when uttering [Amatsukaze] and [Amanohara] of each characteristic point.		
	Yamakawani (Left point)	Yamakawani (Jaw point)
Amatsukaze(Left point)	0.886	
Amatsukaze(Jaw point)		-0.204

5. 2　同一上の句発話時の各話者間の検討

ここでも 5.1 で述べた上の句「わびぬれば」,「あまのはら」,「もろともに」を用い，左端特徴点の結果について述べる．Fig.5 は,「わびぬれば」を発話した時の各話者の動作履歴グラフを示したものである．Fig.6 に，3 名の被験者がそれぞれの句を発話したときの口唇動作スペクトルを示す．表では，Fig.6 に示したようなパワースペクトルの結果から求めた同一の上の句発話時の各話者間の相関関係を上の句ごとに示し，Table 5 は「わびぬれば」，Table 6 は「あまのはら」，Table 7 は「もろともに」を示す．

発話のタイミングおよび長さは話者によって異なるが，特徴点の動作軌跡は固有の動作をもつことが Fig.5 より見受けられる．Fig.6 から分かるように，同一の上の句を対象とした各話者間の動作履歴は同様の固有動作をもち，これにより得られるパワースペクトルも同様の傾向を有するため，その相関係数は安定して高い値を示すことが Table 5,6,7 よりわかる．下顎端特徴点でも同様の結果が得られ，これらの結果から，個人間で使用する発話認識だけではなく，

複数の話者間における発話認識が可能であると考えられる．

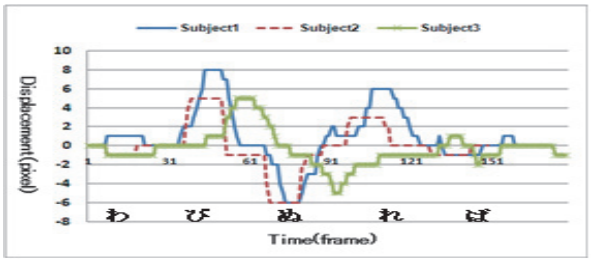

Fig. 5　Trajectory of lip movement by each subject.(Wabinureba, Left characteristics point)

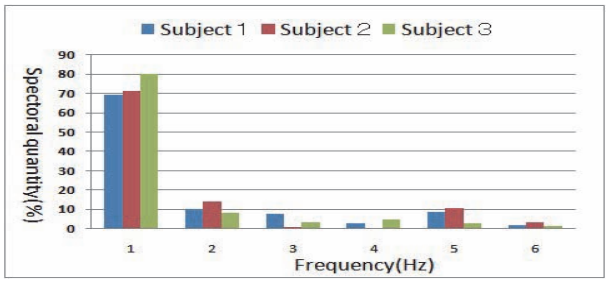

Fig. 6　Power spectrum of the trajectory of lip movement when uttering Wabinureba by each subject.(Left characteristics point)

Table 5　Correlations when uttering Wabinureba between each subject. (The left characteristic point)

Correlations when uttering same phrase between each subject. (The left characteristics point)		
	Wabinureba（２）	Wabinureba（３）
Wabinureba（１）	0.988	0.973
Wabinureba（２）		0.982

Table 6　Correlations when uttering Amanohara between each subject. (The left characteristic point)

Correlations when uttering same phrase between each subject. (The left characteristic point)		
	Amanohara(2)	Amanohara(3)
Amanohara(1)	0.947	0.967
Amanohara(2)		0.839

Table 7　Correlations when uttering Morotomoni between each subject. (The left characteristic point)

Correlations when uttering same phrase between each subject. (The left characteristic point)		
	Morotomoni(2)	Morotomoni(3)
Morotomoni(1)	0.991	0.983
Morotomoni(2)		0.983

5.3　異なる日時における検証

ここでは，異なる日時の発話における実験結果を示す．具体的には朝と夜に一回ずつ発話したものを４日分，計８回であり，例示する上の句として「わびぬれば」を無作為に選択した．Fig.7 では，８日分の朝における口唇の動作履歴を示し，Fig.9 では，夜における口唇の動作履歴を示す。Fig.8 と Fig.10 は，Fig.7,9 に対応する口唇動作スペクトルを示す．Table 8 は各日・時間における相関関係を示したものである．尚，5.1 および 5.2 では被験者３名の結果について記述したが，被験者３名ともに同様の傾向を示したので，ここでは被験者１の実験結果を代表例として報告した．

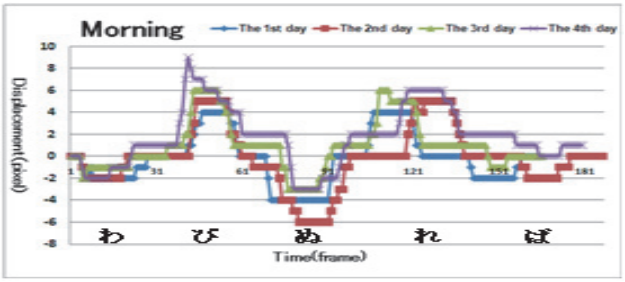

Fig. 7　Trajectory of lip movement when uttering Wabinureba. (Morning, Left characteristics point)

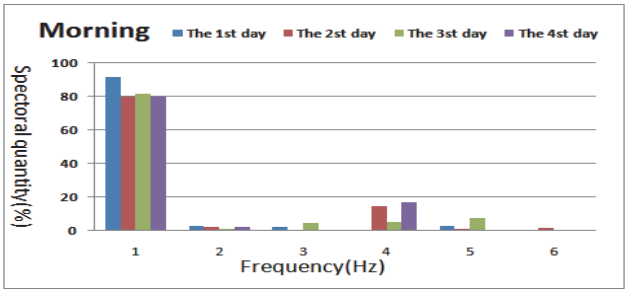

Fig. 8　Power spectrum of lip movement when uttering Wabinureba. (Morning, Left characteristics point)

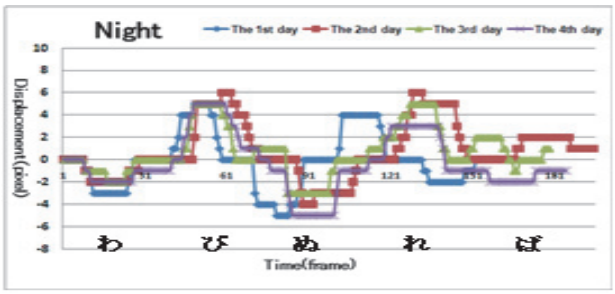

Fig. 9　Trajectory of lip movement when uttering Wabinureba. (Night, Left characteristics point)

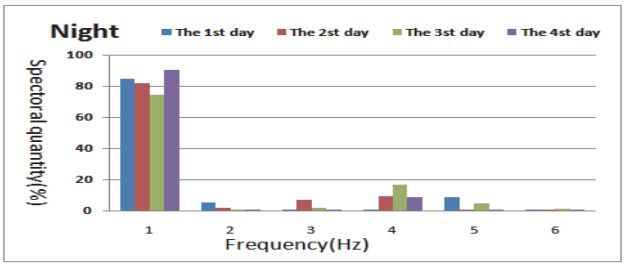

Fig. 10　Power spectrum of lip movement when uttering Wabinureba. (Night, Left characteristics point)

Fig.7 および Fig.9 に示すよう，同一話者の同一の上の句における発話であっても，その発話のタイミングおよび区間は，日時によって異なる．しかし，特徴点の動作履歴が固有の動作を持つことは変わらない．その結果，計８個の動作履歴から得られるパワースペクトルの相関係数は Table 8 に示すよう，全体を通して 0.9 以上の値を計測し，日時が異なる発話であっても安定して強相関が得られることを示した．

Table 8　Correlations of when uttering [Wabinureba] on different days and at different times (Left characteristics point). Mor.:Morning,Eve.:Eveneing

Correlations when uttering [Wabinureba] on different days and at different times (Left characteristics point).							
	1 Eve	2 Mor	2 Eve	3 Mor	3 Eve	4 Mor	4 Eve
1 Mor	0.997	0.981	0.991	0.996	0.973	0.973	0.993
1 Eve		0.975	0.982	0.995	0.969	0.967	0.988
2 Mor			0.993	0.985	0.998	0.999	0.997
2 Eve				0.993	0.989	0.990	0.998
3 Mor					0.983	0.979	0.994
3 Eve						0.998	0.992
4 Mor							0.993

5.4　似通うおよび同一の上の句間の検証

百人一首すべての上の句を解析していく上で，母音列の構成が同一である上の句の数は少ないが，似通う母音列で構成されている上の句は多い．このような似通う母音列を動作履歴としてもつ上の句間では，その内の一つの母音において動作が大きく異なっていたとしても，動作履歴全体は類似した動作履歴を辿ることに変わりはなく，異なる上の句間であっても似通うパワースペクトルを計測してしまう．ここでは，共通の母音が多い似通う母音列構成の上の句間として「あまつかぜ」と「あまのはら」，全く同一の母音列構成をもつ上の句間として「あうことの」と「はるのよの」を例として結果を示し，これらの上の句間における発話認識の検証を行う．尚，ここでも５.３と同様に，他の被験者も同様の結果であったので、代表として被験者１の実験結果を報告する．似通う母音列指標である，「あまつかぜ」と「あまのはら」における，左端端特徴点での動作履歴を Fig.11 に示し，下顎端特徴点における動作履歴を Fig.12 に示す．また，それぞれの口唇動作スペクトルを Fig.13，Fig.14 に示す．そして，左端特徴点と下顎端特徴点における相関関係を Table 9 に示す．また，同一の母音列指標である，「あうことの」と「はるのよの」における，左端特徴点における動作履歴を Fig.15 に示し，下顎端特徴点における動作履歴を Fig.16 に示す．同様に，それぞれの口唇動作スペクトルを Fig.17，Fig.18 に示す．そして，左端特徴点と下顎端特徴点における相関関係を Table 10 に示す．

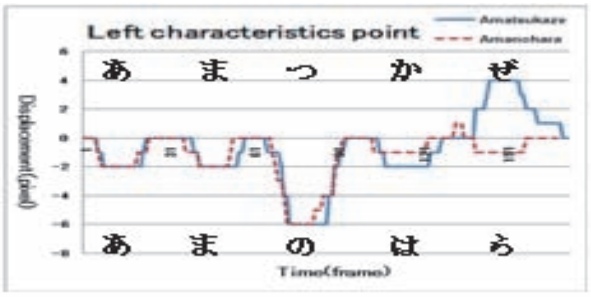

Fig. 11　Trajectory of the left characteristics point when uttering resemble vowels (Amatsukaze and Amanohara)

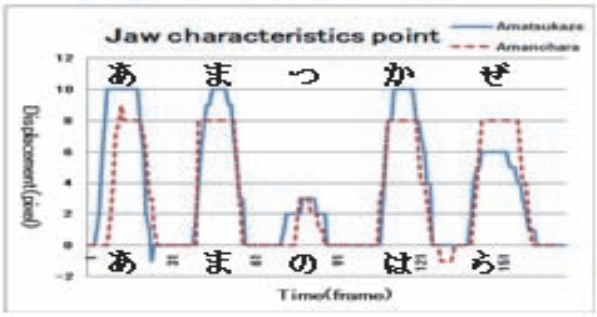

Fig. 12　Trajectory of the jaw characteristics point when uttering resemble vowels (Amatsukaze and Amanohara)

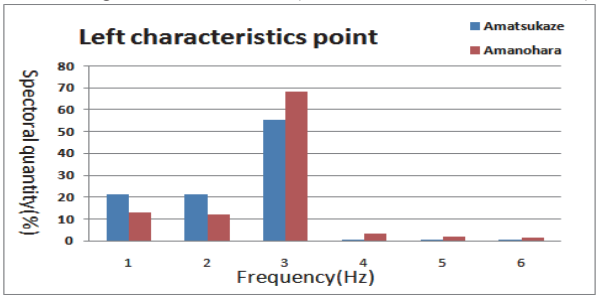

Fig. 13　Power spectrum of the trajectory of the left characteristics point when uttering resemble vowels (Amatsukaze and Amanohara)

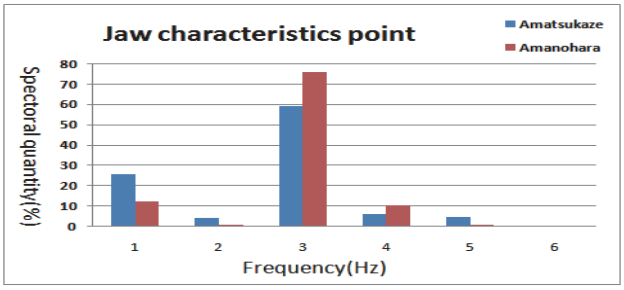

Fig. 14　Power spectrum of the trajectory of the jaw characteristics point when uttering resemble vowels (Amatsukaze and Amanohara)

Table 9　Correlations when uttering each phrase of resembling vowel of left and jaw characteristic point.

Correlations when uttering each phrase of resembling vowel of left and jaw characteristic point.		
	Amanohara(Left)	Amanohara(Jaw)
Amatsukaze(Left)	0.957	
Amatsukaze(Jaw)		0.957

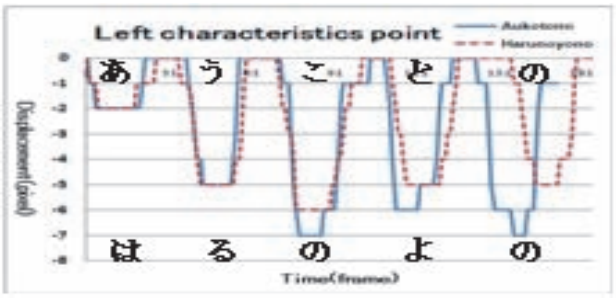

Fig. 15　Trajectory of the left characteristics point when uttering same vowels (Aukotono and Harunoyono)

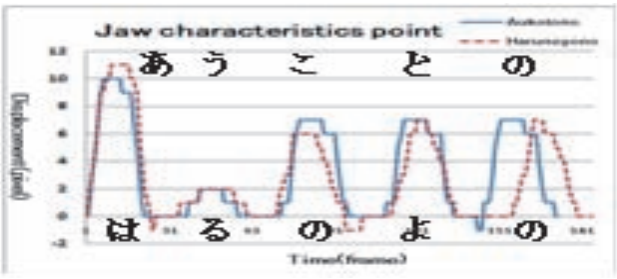

Fig. 16　Trajectory of the jaw characteristics point when uttering same vowels (Aukotono and Harunoyono)

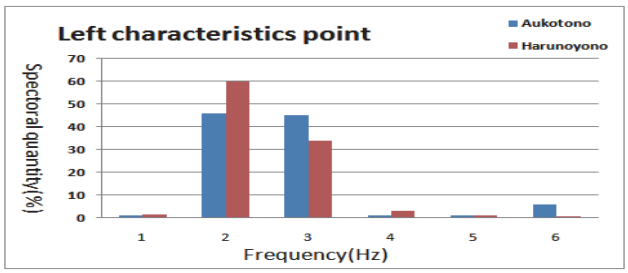

Fig. 17　Power spectrum of the trajectory of the left characteristics point when uttering same vowels (Aukotono and Harunoyono)

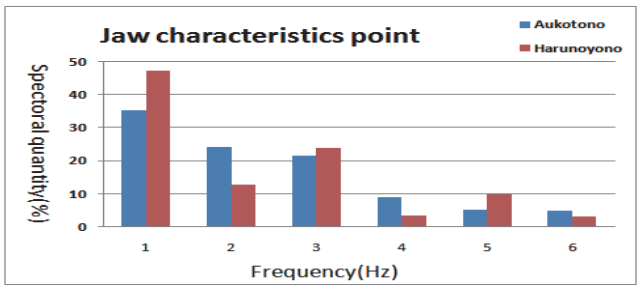

Fig. 18　Power spectrum of the trajectory of the jaw characteristics point when uttering same vowels (Aukotono and Harunoyono)

Table 10 Correlations when uttering each phrase of same vowels of left and jaw characteristic point.

Correlations when uttering each phrase of same vowel of left and jaw characteristic point.		
	Harunoyono(Left)	Harunoyono(Jaw)
Aukotono(Left)	0.942	
Aukotono(Jaw)		0.885

Fig.11 および Fig.12 に示す似通う母音列の上の句間における口唇動作履歴グラフより，各母音発話時の動作履歴が全体的に似通っていることがわかる．どちらの特徴点においても似通う口唇の動作履歴をもつことで，パワースペクトル（動作特徴）も似通ってしまい，その相関係数は，左端特徴点および下顎端特徴点ともに非常に高い値を計測することが Table 9 に示されている．また，Fig.15 と Fig.16 に示す同一母音列の上の句間における口唇動作履歴グラフより，５つの発話母音すべてが同じであることで，口唇の動作履歴がほぼ一致してしまうことがわかる．これらのパワースペクトルの相関係数は，似通う上の句間と同様に非常に高い値を計測することが Table 10 に示されている．このように，口唇の動作履歴が同一および似通う上の句間では，各特徴点におけるパワースペクトルの相関係数が高い値を示すため，相関関係からでは発話した内容を識別することは困難である．よって発話認識を行う際に，百人一首を例とすると，次の句を解析に用いて候補を絞ったり，競技会等では，既に読まれている句を参考にして候補を絞るといった，他の要素と組み合わせて解析する必要があると考えられる．

ここで示したように，日本語には，全く同じ母音列の異なる単語が存在する．例えば，我々が行っている駅名認識で用いられる小田急線の「はだの」「やまと」はどちらも母音は「ああお」である．しかし，それぞれ，小田原線と江ノ島線という別の路線上の駅である．したがって，母音認識だけで日本語の全ての単語を識別することは困難であるが，路線などの情報，使用環境の情報，前後の発話の文脈を考えることにより絞り込みは可能である．我々はここであげた例のように口唇動作による母音認識の課題を認識して，アルゴリズムを組んでいくことにより，様々なアプリケーションに適用が可能になると考えている．

5.5　検定結果

5.5.1　F 検定

各話者が同一の上の句を発話した際に，安定して高い相関係数を得ることができるのかを調べるため検定を行い，パワースペクトルの相関係数について被験者３者間のデータの分散が等しいか否かを確認するため F 検定を使用した．

帰無仮説「各群の母平均値は等しい」であり，対立仮説「各群の母平均値は等しくない」である．

被験者 1 人あたり，同一の上の句の発話における相関データとして 13 個使用した。従って，したがってデータ数の総数は３名で計 39 個であった．その結果，計測された F 値は 1.33 であり，このデータを対象とした 5%有意水準の棄却域境の値は 3.23 であった．「3 人の被験者間の同一の上の句の発話時のパワースペクトルの相関係数値の平均に差はない」という結論が得られ，計測値が棄却域を超えないので帰無仮説は棄却されず，各話者が同一の上の句を発話する際，その相関は安定して高い相関係数を得る傾向があることを示した．なお，統計解析にはマイク

ロソフト社 Excel 2007 を用いた.

5.5.2　t 検定

一被験者が同一および異なる上の句を発話した際,その相関係数が有意的な差であるかを調べるためｔ検定を行った.

帰無仮説「2 群の母平均値に差はない」であり,対立仮説「2 群の母平均値に差がある」である.

一被験者の同一上の句間における相関データを 13 個と,異なる上の句間における相関データを 13 個用い,データ数の総数は計 26 個である.他の被験者も同様の結果であったので,ここでは,被験者 1 のパワースペクトルの相関係数を用いた.その結果,計測されたｔ値は 10.68 であり,このデータを対象とした 5%有意水準の棄却域境の値は 2.06 である.計測値が棄却域を超えたので帰無仮説は棄却され,同一話者が同一の上の句を発話した際の相関係数と,異なる上の句を発話した際の相関係数には有意的な差があるという傾向を示し,これら相関係数の高低差を用いて発話認識を行うことができることを示すことができた.

また,同様に被験者 1 について朝と夕の発話認識において,同一の上の句を発話した際に得られるパワースペクトルの相関データが,安定して強相関といえる係数が得られるのかを調べるため,ｔ検定を行った.　5.3 で述べた「わびぬれば」の朝４回分・夕４回分の発話データを用いる.この朝４回分の発話データ間の相関データ６個と,夜４回分の発話データ間の相関データ６個を用い,データ数の総数は計 12 個である.その結果,計測されたｔ値は－0.12 であり,このデータを対象とした 5%有意水準の棄却域境の値は 2.226 である.計測値が棄却域を超えないので帰無仮説は棄却されず,朝４回分の発話データと,夜４回分の発話データにおける相関データの平均値には差がないということになり,発話する時間や日にちによらず,安定して強相関が得られる傾向があることを示した.なお,ｔ検定にはマイクロソフト社 Excel 2007 を用いた.

6. 実験結果および検討

今回の実験では,百人一首から一部の上の句を用い,あえて同じ語数のみを用いたとき認識可能か,同じ個人でも日にちや昼夜など発話時間を変えたときに変動が生じるか,また,本方式で認識が困難な単語対はあるのかすなわち本方式の限界について検討した.その結果,本報告で得られた成果を以下にまとめる.

(1)　単一の特徴点のみで発話認識を行う場合,特に今回のように語数が同じ場合に,誤認が発生する可能性が示唆された.しかしながら,左端特徴点と下顎端特徴点の二点を併せて相関解析を行うことにより,誤認の回避を行い,複数の単語に対する発話認識が行える可能性を示した.

(2)　同一母音列発話時は,発話者によらず,また同時に発話された子音の構成が異なるにもかかわらず,全体のパワースペクトルの相関係数が安定して高い値を計測した.これにより,発話者を限定しない認識システムを構築できる可能性を示した.

(3)　一個人の口唇動作においても,発話する日時が異なることで動作履歴も異なるが,同一母音列発話時は固有の動作を持つことに変わらず,そのパワースペクトルの相関係数が安定して高い値を計測した.これにより,発話者が発話時間を意識して発話する必要がなく,頑強な発話認識システムの構築が行える可能性を示した.

(4)　口唇特徴点の動作履歴をフーリエ変換したスペクトル相関を用いて,話者を問わず,良好な識別のできる単語がある一方で,パワースペクトルが似通うことにより,発話された単語の識別が困難な場合も存在することが示された.しかし,この様な単語に対しても,鉄道・道路など発話されている分野や騒音下の工場や工事現場で使用されている環境に関わる知識情報,使われる可能性のある単語など解析に用い条件を絞ることで,発話認識が可能であると考えられる.

(5)　各検定の結果より,3 人の発話者の比較から,同一の単語間におけるパワースペクトルの相関係数が安定して高い値を示し,異なる単語を発話した際の相関係数が発話認識を行ううえで有意的な差を持っていることを示した.これらのことから,話者が変わっても,何れかの母音列を発話したかをパワースペクトルの相関関係から識別できることを示し,本手法を用いて発話認識装置で実現できる可能性を示した.

7. むすび

本論文では,音声を伴わない発話認識システムの構築を実用化する際に検討が必要な,発話者間の識別,同一発話者内での安定性,本方式の限界について,識別が難しい同一文字数の母音列を対象として検討した.その結果,単語発話動作のパワースペクトルの相関関係を比較することで発話者によらず,また同一被験者が日にちを変えても安定して,発話認識が行えるという可能性および有効性を示すことができた。しかし,これらの実験結果は３名の被験者からなるデータを基としたものである.よって今後は,より多くの数日間分における発話データ,並びに被験者数を増加した実験を行い,本手法における信頼性のさらなる向上を目的として検証を続けていく.

謝辞

本研究は科研費(22500112)の助成を受けたものである.ここに深く謝意を表する.

参考文献

1) 根田雅稔,西田眞,石井雅樹,佐藤和人:”口唇の動き特徴の個人識別法への適用”,電学論 C,120,5,pp.765-766(2000)
2) 寺田賢治,吉田大輔,大恵俊一郎,大橋剛介:”口の形状をパスワードに用いた本人認証”,電子学会誌,

30，3,pp.267-275(2001)
3) 佐藤慶幸，西田眞：”音声と発話に伴う口唇の動き特徴を用いた個人識別に関する検討”，電学論，125,8,pp.1282-1289(2005)
4) 市野将嗣，坂野鋭，小松尚久：”核非線形相互部分空間法による話者認識”，信学論，J88-D-II，8,pp.1331-1338(2005)
5) 白澤洋一，三浦信，西田眞，景山陽一，栗栖怜史：”口唇の動き特徴を用いた個人識別に関する検討”，映情学誌，.60, No.12,pp.94-100(2006)
6) 石井雅樹，佐藤和人，西田眞，景山陽一：”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，119,4,pp.465-472(1999)
7) 柳朋宏，坂本篤史，山田光穂：”口唇動作を用いた発話認識法の構築”，HIS2007(2007)
8) 張斌，船久保昭夫，福井康裕：”口唇変位計測による読唇コミュニケーション支援システムの開発”，ライフサポート，10 3, pp.13-17(1998)
9) 齊藤剛史，小西亮介：“唇および口内領域形状に基づくトラジェクトリ特徴量に基づく読唇”，信学論，J90-D, 4, pp.1105-1114(2007)
10) 白澤洋一，西田眞，西健治：”ズームと顔の向き変化にロバストな口唇位置の推定”，信学技報,TL2004-83,PRMU2004-251,pp.115-120(2005)