特集論文 ファジィ画像情報処理

口形の多角形近似と ファジィ類似度を用いた読唇†

添田 満*¹ 古谷 忠義*¹ 張 偉京*² 黒須 顕二*³

読唇は、聾啞者が他人の音声言語を受信する方法として、発音時の口辺、特に唇の形状、舌の位置および運動から発音を理解するものである。本研究は、この読唇の工学的実現を目標として行ったものである。顔画像をTVカメラで取り込み、画像処理を施し、口形の特徴を抽出し、ファジィ類似度を用いたパターン認識により、発声母音の識別を試みた。その際、TVカメラと被験者の距離や、個人差による口形の違いに影響を受けないように、口形の部分を六角形に近似して、その辺の比、内角度、縦横比を特徴量として用いた。

キーワード：読唇、口形、六角形近似、ファジィ類似度、パターン認識、母音識別、画像処理

1. はじめに

読唇とは、人間の発音内容を、発音時の口辺、特に唇の形状(以下、“口形”とよぶ)、舌・顎の運動、顔の表情などから理解する方法で、聴覚障害者のコミュニケーションの重要な手段として用いられている。また、騒音環境下での音声認識の代用的・補助的役割を果たすものとしても重要である。一般に、日本語には約 100 種類の異なった音節があるといわれているが、発音の際に識別できる口形変化は、僅か 15 種程度しかないため、同口形異音が多く、読唇において内容を把握するには高度の推論を必要とする^[1]。そのため、この技術の修得には、多年の訓練と努力を必要とし、聾教育における成果は満足すべきものとは言えないのが現状である。

そこでこれらの聾啞者の日常的コミュニケーション

の困難を排除するため、読唇の工学的な実現を目的として、聴覚障害者のために読唇・読唇の練習をするソフトの開発や画像処理などを用いて読唇の実現を目指す研究が行われている^{[2]–[5]}。また、音声認識の補助システムとしての読唇に関する研究も行われている^[6]。筆者らも口形の画像情報から母音の識別・単語の認識をする方法として、パターン認識による方法^[7]、ファジィ論理による方法^[8]、ニューラルネットワークを用いた識別法^[9]について報告している。また、X-Yトラックによる口唇の画像情報を用いた、音声情報との併用による単語認識法について報告している^[10]。

これらの研究では、主にデータとして、口形の正面像を取り入れ、その特徴量を求め、母音の識別や単語の認識を行っている。これらの方法では口形部分の画像が抽出されていることが前提であるものや、識別のための処理が複雑なものも多い。また、一般にこれらの方法による識別では、パターン作成に伴う誤差などのために完全なマッチングをとれないことが多い。これらの手法において実用的な実現を目指すには、計算機のメモリの制約、処理速度の問題が考えられ、口形の抽出、認識率の改善を含めた、より簡略な識別方法が望まれる。

† Lip Reading using Polygonal Approximation of Mouth-Shape and Degree of Fuzzy Similiarity
Mitsuru SOEDA, Tadayoshi FURUYA, Weijing ZHANG and Kenji KUROSU

*¹ 北九州工業高等専門学校
Kitakyushu College of Technology

*² 国際ファジィ工学研究所
Laboratory for International Fuzzy Engineering Research

*³ 近畿大学九州工学部
Kinki University

本論文は、自動的な読唇をめざしており、TV カメラで取り込んだ顔の正面画像をもとに母音の識別を自動的に行うシステムについて提案するものである。本手法では、口だけでなく目、耳、鼻などが写っている顔画像から半自動的に口形を抽出する。より発音状態に近い口形の特徴をとらえるため口形を六角形に近似し、この六角形がどの程度似ているかを、辺の長さ、内角、縦横比を特徴量としてそれぞれの類似度を計算することにより求め、母音の識別を行う。ここでは、マシン制約上の都合から入力画像を間引いて画像処理を行う。この口形抽出や六角形近似の段階で生じる誤差を考慮して、類似度を評価する際にファジィ論理を導入したファジィ類似度を用いる。

2. 口形画像の抽出

ビデオカメラにより被験者の顔画像を取り込む。装置の概略を図1に示す。口全体を含む顔の画像が得られるように、被験者の前方約1 mのところからカラービデオカメラを用いて顔画像を写し、フルカラービデオデジタイザ(Hyper-Vision+ : 画素数640×400、色数R,G,B各8ビット)に取り込む。取り込まれたデータは、32ビット・マイクロコンピュータ(NEC: PC-9801 DA)を用いて処理する。

このとき、撮影は次の条件のもとに行う。

図1 入力装置の概要

a. 口形を鮮明に描き出させるために、口に口紅をつけた状態にしておく。

b. TV 画面のなかに口全体が入っているようにしておく。このとき、口の他に目、耳、鼻などが写っていてもかまわない。3.2で述べるように、得られた画像から抽出した口形を正規化するため倍率は固定ではないが、ここでは、処理の上で、640画素フレームに対して約150画素程度、口の横幅が占めるように入力倍率を設定する。

c. 照明条件としては、口辺や首などに影が写らないようにすることが必要で、ここでは補助ライトを使用する。

ここでは、取り込んだ R,G,B データのうち G データを使用する。これは、G データでは口の部分と肌の部分の濃度値の差が大きいため、2値化によって口の部分が現れやすくなるためである。

取り込んだ画像をそのまま使用すると、口形抽出などの画像処理にマシンの制約上非常に時間を要す。そこで本報告では、前処理としてこの入力データを160×100、256階調のデータに間引いて、メモリの節約、処理速度の改善をはかっている。ここでの間引き処理は単純に4×4ドットを平均して1ドットとする方法を用いている。

(1) 平滑化^[11]

入力画像中に混入した雑音を除去するため平滑化を行う。ここでは、ある点(i,j)を中心とした3×3の平方領域の各点の濃度の均等重み平均を求め、その値を平滑化画像の点(i,j)の新しい濃度とする。

(2) 2値化

口の部分の濃度は小さく、肌の部分は大きい。そこでしきい値 X_0 で2値化処理する。このしきい値 X_0 は以下の式で算出される。

$$X_0 = X_{\min} + (255 - X_{\min}) \times 0.6 \quad (1)$$

(3) 輪郭線の抽出^[13]

2値化された画像で、図2に示すような3×3の小画素領域からなるマスクの中心部分を、常に輪郭線上に置きながらマスクを移動していき、その中心の軌跡を輪郭線とする。はじめに輪郭線上の

一点を抽出するため、画面上を図3のように走査して図形の最上点・最左点の黒領域を見つけ、この点を P_1 とし、輪郭線追跡の始点とする。次に P_1 を中心に 3×3 のマスクを考え、 p_1 から図2に示す番号順に各点が黒領域かを調べ、最初に見つかる点を p_2 とする。ここで、①から④まで調べても黒領域とならないときは、 p_1 は孤立点である。 p_n 点が見いだされ次に p_{n+1} 点を見いだす場合も同様に、 p_n 点を中心に 3×3 のマスクを考え、 p_{n-1} 点から反時計方向に順に黒か白かを調べ、最初の黒領域の点を p_{n+1} とする。この操作を順次行い、 p_{n-1} 点と p_1 点、 p_n 点と p_2 点がそれぞれ一致したとき、輪郭線の追跡は終了したことになる、 p_1 点から p_{n-2} 点が輪郭線となる。

(4) 口形の抽出

(3) の処理で得られた輪郭線の中から、閉じた輪郭線(輪郭線追跡の際、始点に再び戻ってきた輪郭線)を残す。こうして得られた閉曲線の中には口形を表すものの他に、目、肌に現れる小さな影、顎の影なども含まれる。一般に、顔の中における口形の占める面積は大きいと考えられるため、それぞれ残った輪郭線に囲まれる部分の面積を求め、面積があるしきい値以下の小さいものを削除する。ここでは処理を簡単にするため、輪郭線に囲まれる部分の最大の縦幅と横幅を求め、両者の積を面積の近似値として使用する。また、しきい値は、入力画像において口形の占める画素数をもとにした口形の大きさに対してどの程度の大きさまで削除するかの数値を処理者が設定する。さらに、口形と同程度の面積をもつ目、鼻、耳などと区別するため、それぞれの位置関係を考慮して、最も下に位置する輪郭線のみを残す。以上の操作で最後に残った輪郭線を“口”の輪郭線とする。これらの処理手順の概要を図4に示す。

3. 口形の特徴量

読唇を行うには、口の形を表す特徴量が必要となる。本研究では、口形の輪郭線を六角形に近似することにより特徴量を求める。

図2 輪郭線追跡マスク

図3 始点の走査

図4 口形抽出手順

図5 口形の六角形近似と特徴量

3.1 口形の六角形近似

口形の輪郭線から六つの特徴点を選び、この六点を結ぶことにより口形を六角形に近似する。特徴点の抽出は、図5に示すとおり、以下のように行う。

まず、画面の左上から縦方向に探索していき、最初に輪郭線とぶつかった点(輪郭線の最左端)の X 座標を x_1 とする。 X 座標が (x_1+1) となる点を上と下から探索し、それぞれぶつかった二点の midpoint の Y 座標を y_1 とする。このとき、点 (x_1, y_1) を特徴点 A とする。同様に、画面右上から縦方向に探索して求めた輪郭線の最右端の X 座標を x_2 として、 X 座標が (x_2-1) となる点を上下から探索し、現れた二点の midpoint の Y 座標を y_2 とする。このとき、点 (x_2, y_2) を特徴点 B とする。

次に、画面の左上から横方向に探索していき、最初に輪郭線とぶつかった点(輪郭線の最上端)の Y 座標を y_3 とする。 Y 座標が (y_3+1) となる点を左と右から探索し、それぞれぶつかった二点の X 座標を x_3, x_4 とする。このとき、点 (x_3, y_3+1) を特徴点 C 、点 (x_4, y_3+1) を特徴点 D とする。同様に、画面左上から横方向に探索して求めた輪郭線の最下端の Y 座標を y_4 とし、 Y 座標が (y_4-1) となる点を左右から探索し、現れた二点の X 座標を x_5, x_6 とする。このとき、点 (x_5, y_4-1) を特徴点 E 、点 (x_6, y_4-1) を特徴点 F とする。

これらの六つの特徴点 $A \sim F$ を結ぶことにより、口形を六角形に近似することが出来る。

3.2 特徴リスト

六角形は、辺の長さと内角が与えられれば、その形を決定することができる。そこで近似された六角形の辺の長さ h_i 、内角 k_i を図5のように与える。辺の長さ h_i 、内角 k_i 、六角形の縦横比 ty (=横幅/縦幅)を六角形の特徴量とし、これらの特徴量から、口形の特徴リストとして、次の三つの特徴リストを定義する。

1. 六角形の辺の長さをもとにした特徴リスト

$$S_h : (h_1, h_2, h_3, h_4, h_5, h_6)$$

ただし、辺の長さは、カメラと被写体との距離な

ど、画像を撮る条件により変化するため正規化する必要がある。ここでは、 $h_1 \sim h_6$ は処理者の指定した特定の位置の辺を1としたときの各辺の長さとした。

2. 六角形の内角度をもとにした特徴リスト

$$S_k : (k_1, k_2, k_3, k_4, k_5, k_6)$$

ただし、 S_h, S_k は要素として、六角形の辺、角をある辺、角から右回りに順番に並べたもので、リストの第一要素はどの辺、角からスタートしてもよい。

3. 六角形の縦横比をもとにした特徴リスト

$$S_{ty} : (ty)$$

これらの特徴リストをもとに、口形の識別を行う。

4. 口形の識別

本研究では、口形のパターン認識を行うことにより、母音の識別を行う。パターン認識の基本的な過程は図6のように考えられる。特徴抽出法としては、3節の口形を六角形近似する方法を用い、得られた口形の特徴リストと基本パターンの特徴リストの類似度を評価し、母音の識別を行う。

類似度の評価において、入力パターンと基本パターンの特徴リストを比較する場合、二つの特徴リストで表される図形は“同じ”ではないが“似ている”といえることが多く、このような場合には、特徴リストの各要素は、“一致”はしていないが、“似た値”になっている。このように、二つの特徴リストがどの程度一致しているかということはいまいなことが多い。そこで、このような類似性、すなわち“似ている”程度のあいまいさを表すために、ファジィ論理を導入したファジィ類似度を用いて、特徴リストの評価を行う。

図6 パターン認識の過程

4.1 ファジィ類似度

(1) 数の類似度

類似度の評価を行う際に、特徴リストの各要素は、図7に示すように、三角形メンバーシップ関数 $\mu(x)$ をもっているファジィ数とする。また、その三角形の底辺の幅はその数の絶対値に一致しているものとする。全ての数をこのように定義すると、二つの数 A と B の類似度 $S(A, B)$ は、図7に示されるように、

$$S(A, B) = \text{Max}\{\text{Min}\{\mu_a(x), \mu_b(x)\}\} \quad (2)$$

と定義される。

この数と数との類似度を用いることにより、リスト同士の類似度を評価する。リスト $C: (c_1, c_2, \dots, c_n)$ とリスト $D: (d_1, d_2, \dots, d_n)$ の類似度 S_L は

$$S_L = \text{Min}_{1 \leq i \leq n} \{S_i\} \quad (3)$$

と定義される。ただし、 S_i はリストの各要素 c_i と d_i の類似度で、(2)式により求められる。

(2) 特徴リストの類似度

二つのパターンの特徴リストを、それぞれ

$$L_1: (x_1, x_2, \dots, x_n)$$

$$L_2: (y_1, y_2, \dots, y_n)$$

とする。一般に、二つの図形パターンの類似度を評価する場合、片方の特徴リスト L_1 については、その要素を順次ずらしてゆき、

$$\begin{aligned} L_1: L_{11} &= (x_1, x_2, \dots, x_n) \\ L_{12} &= (x_2, x_3, \dots, x_n, x_1) \\ &\dots \\ L_{1n} &= (x_n, x_1, \dots, x_{n-1}) \end{aligned}$$

の n 個のリストを作成し、それぞれのリストと基本パターンのリストを比較して類似度を求める。これは一方の形状を回転しながら、他方の形状と比較しそれぞれの類似度を求めていることに相当する。ここで、二つのパターン図形の位置関係がうまく対応した時がもっとも図形の類似度が高いと考えられる。そこで、リスト L_{1k} とリスト L_2 の類似度を S_{Lk} とすると、 S_{Lk} は(2)式により求めることができ、二つの特徴リスト L_1, L_2 の類似度 S_F は

図7 二つのファジィ数の類似度

$$S_F = \text{Max}_{1 \leq k \leq n} \{S_{Lk}\} \quad (4)$$

と定義される。

ここで使用する特徴リスト S_n, S_k は要素として、六角形の辺、角をそれぞれある位置から右回りに順番に並べたものである。辺のリストについては、特定の位置の辺の長さを1として正規化しているため、たとえばその辺から右回りのリストを作成するなどして、標準リストと入力リストの要素の位置対応が一致していれば、要素を順次ずらして n 個のリストを作成する必要はないが、ここでは、基本パターンと入力パターンのリストの第一要素が同じ位置対応の辺からスタートしていない場合を想定して、 n 個のリストを求め比較している。

今回は、処理者が指定する特定の辺を1として正規化しているため、半自動化になっているが、各図形とも最小(最大)辺で正規化すればより一般的になり、自動化に近づく。この場合は、合同図形の類似性を求めることと等価になり、上記のように順次ずらしたリストにより比較する方法をそのまま適用することができる。

4.2 母音の識別法

抽出された口形の六角形データの特徴リスト(辺 S_h 、内角度 S_k 、縦横比 S_{ty})と発声母音の基本パターンの特徴リストから、それぞれ、辺、内角、縦横比について、特徴リストの類似度を求める。この場合、 S_h と S_k の類似度については、(4)式により求め、 S_{ty} の類似度については(2)式により求める。

各母音(あ・い・う・え・お)ごとに、入力データと基本パターンの類似度を、辺、角、縦横比について求めていく。辺、角、縦横比についてそれぞれ、一番類似度の大きい母音に2ポイント、二番目に大きい母音に1ポイントを与える。ポイントの合計が一番大きい母音を発音母音と認識する。ポイントの合計が同じになった場合は、その母音の三つの類似度の平均をとり、これが大きいものを発音母音と判定する。以上から、入力パターンがどの母音に一番近いかを判定し、発音母音の識別を行う。

5. 母音認識実験

本論文で示した手法の評価を行うため、発音母音“あ”・“い”・“う”・“え”・“お”を発声時の口形から識別する実験を行った。

データとして、同一人物から、“あ”～“お”の母音を発声したときの顔画像を4組取り込む(これらをデータ①～④とする)。また、他の人物の同様の顔画像を取り込む(これをデータ⑤とする)。次に、これらのデータからそれぞれ、前述の方法により、口形を抽出し、特徴リスト(辺 S_b 、内角 S_k 、縦横比 S_{lv})を求める。入力画像から口形部分を抽出し、六角形に近似した結果の例を図8に示す。

図8 口形抽出・六角形近似例

ここで、同一人物から取り込んだ3組のデータ①～③を平均したものを基本パターンとし、入力パターンの特徴リストと基本パターンの特徴リストのファジィ類似度を(2)～(4)式により求め評価することにより、入力データの発音母音の識別を行った。

実験1：基本パターンを構成するデータ①～③の中の一つを入力データとして、基本パターンと比較し、発音母音の識別を行った。識別結果を表1に示す。表中、辺・角・縦横比の数字はファジィ類似度を表し、一番類似度の大きいものに○印を、二番目に大きいものに△印を付している。また P はポイントの合計を示している。

実験2：実験1と同一人物で、基本パターンを構成するデータには含まれていないデータ(④)を入力データとして、基本パターンと比較、識別を行った。このときの結果を表2に示す。

実験3：基本パターンを与える人物とは別の人物のデータ(⑤)を、入力データとして発音母音の識別を行った。このときの結果を表3に示す。

実験1と実験2から、入力データが、基本パタ

表1 判別結果(実験1)

入力	対象	辺の比	角 度	縦横比	P	判別結果
あ	あ	○ 0.8930	○ 0.9714	○ 0.9118	6	あ
	い	0.6438	0.7719	0.1110	0	
	う	0.6316	0.8103	0.6057	0	
	え	△ 0.7830	0.8196	0.4312	1	
	お	0.6038	△ 0.8727	△ 0.7544	2	
い	あ	0.6581	0.7014	0.1871	0	い
	い	○ 0.7973	○ 0.8811	○ 0.9953	6	
	う	0.5187	△ 0.8800	0.4622	1	
	え	△ 0.7831	0.7938	△ 0.6380	2	
	お	0.5817	0.7400	0.3236	0	
う	あ	0.5276	0.7817	0.7027	0	う
	い	0.4151	△ 0.8807	0.4469	1	
	う	△ 0.6065	○ 0.8872	○ 0.9882	5	
	え	0.4968	0.8447	0.8035	0	
	お	○ 0.6755	0.8311	△ 0.8593	3	
え	あ	△ 0.7299	0.8387	0.5106	1	え
	い	0.6994	△ 0.9057	0.6362	1	
	う	0.6500	0.8586	△ 0.8123	1	
	え	○ 0.8204	○ 0.9736	○ 0.9971	6	
	お	0.5349	0.8353	0.6623	0	
お	あ	0.6670	△ 0.8521	△ 0.8653	2	お
	い	0.4233	0.7471	0.2985	0	
	う	△ 0.7980	0.7882	0.8240	1	
	え	0.6379	0.7600	0.6421	0	
	お	○ 0.8808	○ 0.9075	○ 0.9763	6	

ーンを構成するデータに含まれている、いないにかかわらず、同一人物の場合は、かなり正確に判別されていることが分かる。従来の正面画像により母音識別を行う方法では、特徴量として口形の縦横幅などを使用しているが、これらは特徴量としての分離性はあまり良くなく、本当の発音状態の口形の特徴をとらえているとは言い難い面があった。本論文では、口形を六角形に近似し、この六角形を特徴づける辺、内角、縦横比を特徴量としたため、より発音状態に近い口形の特徴をとらえることができる。従って、従来の特徴量では分離し難かった母音“あ”、“え”、“お”の認識も正確に行えるようになっていく。また、これらの場合、完全なマッチングがとれず、特徴リストの要素が一致していなくても、ファジィ類似度を用いたことにより、“似た状態”であることをとらえて的確に判別を行っている。一方、実験3のように、基本パターンと入力パターンが別人物の場合には、“あ”と“お”の識別に誤った結果がでていく。これは、発音時の口の開け方の個人差によるものと考えられる。また、基本パターンの特徴量として、三つの画像データの平均をとったものを使用

しているため、誤差が大きく、本来の基本パターンとなるべき特徴量からずれていることも考えられる。この問題は、基本パターンを決定する際、いろいろな人のデータを数多く取り込むことにより改善されると考えられる。

ここでは、口形の辺、内角、縦横比の3つの特徴リストから、各々の類似度を求めているが、これにより、それぞれの特徴(項目)で比べるとどの母音に近いのかを知ることにもできる。また、ここでは、ポイントの総合点を比較することにより、最終的にどの母音かを判定している。したがって、たとえば発音時における口の突き出し量など、発音状態に大きく関わっていると思われる特徴量をリストとして加えれば、より正確な識別を行うことも期待できる。また、比較するときに重視したい項目があれば、各項目ごとに重みをつけてポイントを加算することにより、希望する条件での識別を行うことができる。

6. おわりに

本論文では、顔画像の中から口形を抽出し、得られた口形のもつ特徴量を用いて読唇を行うシス

表2 判別結果(実験2)

入力	対象	辺の比	角 度	縦横比	P	判別結果
あ	あ	○ 0.9068	○ 0.9538	○ 0.9723	6	あ
	い	0.6848	0.8036	0.1603	0	
	う	0.5205	0.7913	0.6643	0	
	え	△ 0.8247	0.8065	0.4874	1	
	お	0.6265	△ 0.9074	△ 0.8143	2	
い	あ	△ 0.2386	○ 0.8565	0.1754	3	い
	い	○ 0.2587	0.7323	○ 0.9906	4	
	う	0.0251	0.7236	0.4491	0	
	え	0.1493	0.7689	△ 0.6243	1	
	お	0.1488	△ 0.8033	0.3111	1	
う	あ	0.3527	0.6080	0.6665	0	う
	い	0.3587	○ 0.7600	0.4813	2	
	う	○ 0.5261	△ 0.7204	○ 0.9747	5	
	え	△ 0.4933	0.6739	△ 0.8403	2	
	お	0.4456	0.6211	0.8224	0	
え	あ	0.5000	0.8605	0.4374	0	え
	い	○ 0.7392	0.7982	0.7128	2	
	う	0.3280	0.7667	△ 0.7347	1	
	え	△ 0.6571	○ 0.8855	○ 0.9185	5	
	お	0.2700	△ 0.8645	0.5864	0	
お	あ	0.7688	△ 0.8521	△ 0.8653	2	お
	い	0.4233	0.7471	0.2985	0	
	う	△ 0.7980	0.7882	0.8240	1	
	え	0.6379	0.7600	0.6421	0	
	お	○ 0.8808	○ 0.9075	○ 0.9763	6	

表3 判別結果(実験3)

入力	対象	辺の比	角 度	縦横比	P	判別結果
あ	あ	○ 0.8126	0.8475	0.8169	2	お
	い	0.4848	0.8318	0.3417	0	
	う	0.6149	0.7692	△ 0.8725	1	
	え	△ 0.6985	○ 0.9193	0.6896	3	
	お	0.6617	△ 0.8632	○ 0.9749	3	
い	あ	○ 0.4708	0.7817	0.2074	2	い
	い	△ 0.4466	△ 0.8618	○ 0.9711	4	
	う	0.1440	0.8545	0.4847	0	
	え	0.3597	○ 0.8916	△ 0.6614	3	
	お	0.1862	0.7376	0.3451	0	
う	あ	0.6134	△ 0.7951	0.7666	1	う
	い	0.2799	○ 0.8396	0.3874	2	
	う	○ 0.8042	△ 0.7951	△ 0.9232	4	
	え	0.4140	0.7462	0.7394	0	
	お	△ 0.7310	0.7732	○ 0.9241	3	
え	あ	△ 0.8427	0.7757	0.6525	1	え
	い	0.7629	○ 0.9175	0.4948	2	
	う	0.5668	△ 0.8621	○ 0.9602	3	
	え	○ 0.8607	0.8235	△ 0.8547	3	
	お	0.6284	0.7617	0.8081	0	
お	あ	0.3811	0.8182	△ 0.8653	1	う
	い	0.4092	○ 0.8958	0.2985	2	
	う	○ 0.6161	△ 0.8477	0.8240	3	
	え	0.3967	0.8020	0.6421	0	
	お	△ 0.5042	0.8000	○ 0.9763	3	

テムを構成し、基礎的な発声母音の認識実験を行い、その結果について報告した。TV カメラにより取り込まれた顔画像に、画像処理を施すことにより、口形の輪郭線の抽出を行い、口形を六角形近似することにより得られた特徴量から、特徴リストを作成する。基本パターンと入力データの特徴リストのファジィ類似度から発声母音の識別を行った。ここでは特徴量として発音時の口形の近似六角形の辺長、内角、縦横比のリストを作成し、リスト同士の類似度を求め、その一番大きいもの同士がもっとも近いと判断するため、従来の最大尤度関数やニューラルネットワークなどを用いた識別方法よりも単純で、物理的にも把握しやすい。

本手法では、基本パターンとの比較において、ファジィ論理を導入したファジィ類似度を定義し、このファジィ類似度を用いて識別を行っている。したがって、従来のパターン認識のように完全なパターンマッチングが取れていない場合にも、“一致した状態”でなく“似た状態”を評価に取り入れることができる。また、処理が簡単で物理的に把握しやすいため、環境・条件を考慮した特徴量の項目を増やすことで、簡単に識別精度をあげることが可能である。さらに、各特徴量ごとに、どの程度似ているかといった、項目ごとの類似性を評価することもできる。したがって、本手法は、システムを構成する上で、拡張性をもった読唇手法であるということが出来る。さらにこの研究を基礎として、頬や顎との区別、光の反射による問題、輪郭線抽出の精度の問題などを考慮し、口形の抽出精度の向上をはかることが必要である。

謝 辞

本研究は一部国際ファジィ研究所の受託研究として実施したものであり、また、実施に当たり九州工業大学情報工学部学生辻龍一君ほかに協力を得たことを感謝する。

参 考 文 献

- [1] Fukuda, Y., S. Hiki : Characteristics of the mouth shape in the production of Japanese stroboscopic observation, J. Acoust. Soc. Japan, (E), 3, 2, 75/91 (1982)
- [2] Mase, K., A. Pentland : Automatic lipreading by optical flow analysis, Trans. Inst. Electron. Inf. Commun. Eng., D-II, 6, 796/803 (1990)
- [3] Okada, K., C. Ohira, H. Nakamura : A method for lipreading, Trans. Inst. Electron. Inf. Commun. Eng., D-II, 9, 1532/1583 (1989)
- [4] Gulliams, I., A. Segui : Interactive videodisc for teaching and evaluating lipreading, Proceedings of An. Intern. Conf. of IEEE Eng. in Medicine and Biology Soc., Cat. 88Ch2566-8, 1551/1552, 4 (1988)
- [5] 渡部 : 読唇による二連母音識別、日本機械学会論文集 (C)、55-509, 134/138 (1989)
- [6] E. D. Petajan : Automatic Lipreading to Enhance Speech Recognition, IEEE Global Telecommunication Conference, GLOBECOM'84 Conference Record, 265/272 (1984)
- [7] 松岡、古谷、黒須 : 画像処理による読唇の試みー母音口形の識別およびそれに基づく単語認識、計測自動制御学会論文集、22-2、191/198 (1986)
- [8] Kurosu, K., T. Furuya, S. Takeuchi : Lipreading by fuzzy logic, Proc. of 2nd Congress of IFSA, 23-4C (1987)
- [9] Kurosu, K., T. Furuya, et al. : Lipreading by a neural net, Proc. of 8th European Annual Conf. on Human Decision Making and Manual Control, 329/339 (1989)
- [10] 古谷、添田、黒須、田村 : X-Y トラッカを用いた口形情報併用による単語音声認識、計測自動制御学会論文集、27-8、958/965 (1991)
- [11] 安居院、中嶋 : コンピュータ画像処理、秋葉出版 (1989)
- [12] 張、古谷、黒須 : 記号処理的手法による図形の分解と合成、計測自動制御学会論文集、26-1、81/86 (1990)
- [13] 田村 : コンピュータ画像処理入門、総研出版 (1985)

(1993年9月3日 受 付)
 (1994年4月9日 再 受 付)
 (1994年7月7日 再々受付)

[問い合わせ先]

〒820 飯塚市柏の森11の6
 近畿大学九州工学部経営工学科
 黒須 顕二
 ☎：0948-22-5655
 ☎：0948-23-0536

著 者 紹 介

添田 満 (そえだ みつる)

北九州工業高等専門学校電子制御工学科

1977年 九州工業大学制御工学科卒業。1979年 九州工業大学大学院修士課程制御工学専攻修了。同年舞鶴工業高等専門学校助手。同講師。北九州工業高等専門学校講師を経て、1992年 同助教授。現在にいたる。ロボットの制御、システム同定の研究に従事。日本ファジィ学会、日本機械学会、計測自動制御学会等の会員。

張 偉京 (ちやう いきやう)

国際ファジィ工学研究所

1985年夏 中国科学技術大学システム科学専攻卒業。1986年秋 政府派遣留学生として来日。1989年 九州工業大学大学院制御工学研究科修士課程修了。1992年 東京工業大学大学院総合理工学研究科システム科学専攻博士課程退学。同年渡米。Apronix社(San Jose, California)に入社。1993年秋 国際ファジィ工学研究所(横浜)に入所。現在にいたる。ファジィ理論とその応用および画像認識の研究に従事。日本ファジィ学会会員。

古谷 忠義 (ふるや ただよし)

北九州工業高等専門学校電子制御工学科

1965年 九州工業大学制御工学科卒業。1970年 大阪大学大学院博士課程電気工学専攻修了。工学博士。同年九州工業大学助教授。1990年 北九州工業高等専門学校教授。現在にいたる。プロセス制御、信号処理、情報処理応用の研究に従事。計測自動制御学会、電子情報通信学会等の会員。

黒須 顕二 (くろす けんじ)

近畿大学九州工学部経営工学科

昭和31年 東京教育大学理学部物理学科卒業。同年 運輸技術研究所(現船舶技術研究所)入所。昭和53年九州工業大学工学部教授。平成3年九州工業大学情報工学部教授。平成6年 近畿大学九州工学部教授。大規模システムの高信頼運転、ヒューマン・インタフェース、知的技術の応用に関する研究に従事。博士(工学)。