Akita University

口唇の特徴に着目したコマンド識別および発話認識に関する基礎研究

2013

髙　橋　　毅

Akita University

内 容 梗 概

近年，高機能化の著しい各種情報機器の利用を支援するため，ユーザビリティに主眼を置くヒューマンマシンインタフェースの研究・開発が行われている．その中で，日常一般的な動作である「発話」の音声情報に着目したインタフェースの研究・開発も行われており，カーナビゲーションなど多くのシステムに応用されている．この発話動作には，音声情報に加えて口唇の動きという視覚情報も含まれ，この口唇の動きをコマンド入力や発話認識システムに適用した研究事例はこれまでにも多数報告されている．しかしながら，多くの利用者が共用する状況下において良好な認識を可能にする要素技術，ならびに自然な発話条件を実現するための要素技術の開発は十分とは言い難い．このため，口唇の動き特徴を用いたヒューマンインタフェースの要素技術として，(1) 自然な発話状態での入力操作においても発話区間を良好に自動抽出する手法，(2)多数の利用者がシステムを共用する使用環境へ対応するための手法，(3)発声に起因する口唇の動き特徴変動を考慮した手法の開発が望まれている．本論文は上記項目に関わる課題について研究を行い，工学上の進歩に寄与することを目的とするものである．本論文は全5章より構成されている．

第1章を緒論とし，ここでは本研究の背景とその目的を述べ，本研究に対する筆者の立場を明らかにした．さらに，本論文の主題である口唇の動き特徴を用いたコマンド識別・発話認識システムならびに実用化のための要素技術について，現在までの研究状況を概観するとともに，本研究の内容について述べている．

第2章では，口唇の色彩特徴と形状の時系列変化量を用い，複数の単語を任意の間隔で発話する状況においても発話フレームを検出可能な手法を提案した．被験者5名に3つの単語を任意の間隔で発話させたデータを用いて実験した結果，提案手法は各単語の発話フレームを高精度で検出可能であり，複数の単語を含む場合における発話区間の推定に有用であることを明らかにした．

第3章では，口唇の局所部位から得られる形状特徴に着目した口唇形状のグループ化法について検討を加えた．上下唇の厚さ，口裂の凹凸方向，ならびに口唇のアスペクト比の3種類の形状特徴についての統計的な解析を行い，各形状それぞれ3クラスから構成される27の形状カテゴリを構築した．さらに，口唇の局所形状に基づいて被験者を27カテゴリに自動分類するアルゴリズムを提案し，被験者52名を対象にした評価実験において，80%以上の精度で登録データおよびその類似形状に分類可能であることを示した．また，分類結果は照合対象の絞り込みにも有用であることを示した．

第4章では，発話フレーム数および口唇の動き特徴（横幅，縦幅，面積，アスペクト比）に着目し，各特徴量と発声との関連について検討を加えるとともに，同一取得日における無声発話データと有声発話データの判別についても検討を加えた．その結果，無声発話時は有声発話時と比較し，発話区間が長くなる傾向

Akita University

を有すること，発話全体を通した口唇の動作量が大きくなる傾向を有することが明らかになった．また，発話フレーム数および口唇動作量を用いた線形判別手法は，約 92%の精度で同一取得日における無声発話データと有声発話データを判別可能であることを明らかにした．

第 5 章は結論で，本研究で得られた主な成果と本論文の工学的意義および今後に残された課題について述べている．

Akita University

目　次

第 1 章　緒論　1
1.1 本研究の目的　1
1.2 口唇の動き特徴に着目したコマンド識別・発話認識に関する研究（分野）の概観　2
1.2.1 発話区間の推定に関する従来研究　2
1.2.2 利用者の増加を想定したコマンド識別・発話認識精度向上に関する従来研究　3
1.2.3 発話状態とコマンド識別・発話認識精度の関連に関する従来研究　4
1.3 本論文の内容　5
1.4 本論文で用いる主な用語　6
1.5 本論文における個人情報を有するデータの取扱い　6
第1章　文献　7

第 2 章　口唇の色彩情報および形状情報に着目した発話フレームの検出　9
2.1 はじめに　9
2.2 使用データ　10
2.2.1 データ取得方法　10
2.2.2 口唇領域の特徴解析用データ(画像データセット A)　11
2.2.3 評価用データ（画像データセット B)　12
2.3　口唇領域の色彩情報解析　13
2.3.1　色彩情報を用いた従来研究　13
2.3.2　口唇の色彩情報とその特徴　14
2.4　発話フレーム検出法　16
2.4.1　口裂判定処理　17
2.4.2　口唇形状の時系列変化判定処理　18
2.5 検出条件の検討および評価方法　19
2.5.1　口裂判定位置に関する比較　19
2.5.2　口唇形状の時系列変化検出に関する比較　20
2.5.3　発話フレーム検出の評価方法　22
2.6　実験結果および考察　23
2.6.1　発話フレーム検出結果　23
2.6.2　検出失敗事例に関する考察　28
2.7　まとめ　29
第2章　文献　30

第 3 章　口唇局所領域の形状解析に基づいた顔画像のグループ化手法　31
3.1 はじめに　31
3.2 使用データ　33
3.2.1 データ取得環境　33

Akita University

3.2.2 解析用データ 33
3.2.3 分類実験用データ 33
3.3 口唇の形状特徴 34
3.3.1 形状特徴と局所領域分割 34
3.3.2 口唇領域の明度情報 35
3.3.3 特徴量の算出 37
3.4 特徴量による形状分布解析と分類カテゴリ 39
3.4.1 形状特徴の分布 39
3.4.2 形状カテゴリの定義 48
3.5 局所形状に着目した顔画像のグループ化法 51
3.5.1 口唇領域分割・特徴量算出処理 51
3.5.2 局所形状クラスの判定処理 52
3.5.3 口唇形状の分類処理 55
3.6 分類実験および考察 56
3.6.1 登録データ生成 56
3.6.2 評価基準 56
3.6.3 分類結果に関する検討 56
3.6.4 k-means 法による分類結果 58
3.6.5 絞り込みに関する検討 59
3.6.6 分類不良となった事例 59
3.7 まとめ 61
第3章 文献 62

第 4 章 発話に伴う口唇の動き特徴と発声の関連に関する検討 63
4.1 はじめに 63
4.2 使用データ 65
4.2.1 データ取得の流れ 65
4.2.2 使用コマンドの選定 67
4.3 特徴量抽出処理 68
4.3.1 前処理および口唇特徴計測 68
4.3.2 特徴量の算出 69
4.4 口唇の動作量変化に関する検討 70
4.4.1 発話区間の変動 70
4.4.2 口唇横幅および縦幅の変動 71
4.4.3 面積（raS_i）およびアスペクト比（A_i）の変動 73
4.5 口唇の動作量に着目した発声データの判別 76
4.5.1 発話データセット 76
4.5.2 無声，有声判別に関する検討 78
4.6 まとめ 83
第4章 文献 84

Akita University

第 5 章　結論　85
5.1　本論文により得られた主な知見　85
5.2　本論文の工学的意義　87
5.3　今後に残された諸問題　88

謝辞　89

本研究に関連する発表論文　91

Akita University

第 1 章 緒論

1.1 本研究の目的

近年，高機能化の著しい各種情報機器の利用を支援するために，ユーザビリティに主眼を置くヒューマンマシンインタフェースの研究・開発が行われている．その代表的なものに音声認識[(1)-(7)]による非接触式インタフェースがあり，カーナビゲーションや携帯端末などに広く応用されている．これは，手による操作を必要とせず，人間が日常的に行っている「発話」を介して操作が可能である．「発話」による操作は，操作盤を視認する必要がなく，操作に熟達する必要もないことから，非常に利便性が高い．しかしながら，音声認識は雑音環境下では認識精度が低下すること，発声が制限される状況下では使用に適さないこと，発話内容が第三者に知られてしまう可能性のあることなどの課題を有している[(8)]．

一方，発話時の発声動作は，声帯の音波発生や咽頭・口腔などでの共鳴動作と共に，口唇・舌の動作が必要不可欠である[(9)]．したがって，人間が発話した場合には，音声に加えて「口唇の動き」を発話情報として得ることができる[(10)-(16)]．発話に伴う口唇の主な動きは，上下方向の開閉と左右方向の伸縮に大別され，これらの動きから得られる時系列的な変化を口唇の動き特徴パターンとして捉えることができる[(10)(14)]．この口唇の動き特徴パターンは，発話内容に関する情報を包含しているため，発話に伴う口唇の動き特徴を視覚情報としてビデオカメラなどで捉え，ヒューマンマシンインタフェースに利用した場合，下記に示す多くの利点が考えられる．

(1)発話に伴う口唇の動き特徴は，発声の有無に係らず特徴量を取得可能であるため，雑音環境下や静粛性を要求される環境でも使用可能である．
(2)無声発話の場合には，周囲の第三者に対する発話内容の秘匿が可能である．
(3)特殊な設備を必要とせず，市販のカメラなどで特徴量が取得できる．
(4)入力が容易であり，利用のために特殊な知識やスキルを習得する必要が無い．
(5)口唇の動きは行動的特徴であるため，パスワードなどのログイン情報として利用する場合に登録データの変更が可能である．
(6)口唇は発話動作において最も特徴的な部位であるため，口唇の特徴を用いたコマンド識別・発話認識技術は，顔全体の特徴に着目した非接触インタフェースの開発においても重要な技術となる．

口唇の動き特徴を入力情報として適用した研究事例はこれまでにも報告されており，個人識別[(15)-(17)]やコマンド識別・発話認識[(18)-(22)]へ応用可能であることが明らかにされている．また，口唇の動き特徴変動は，体調・心情変化を検出する指標になり得ることが明らかになっている[(23)]．しかしながら，①自然な発話状態での入力を想定した発話区間自動推定手法が確立されていないこと，②多数の利用者が共用する状況を想定した研究事例が見当たらないこと，

Akita University

③発声の有無や体調・心情変化など，口唇の動きに影響する各種要因について検討された事例は少ないなどの課題を有している．特に，発声の有無に起因する特徴量の変動は，より自然な利用環境を実現する上で重要な課題である．

以上の観点から，本論文では，口唇の動き特徴を入力情報とするヒューマンインタフェースの実用化に向けた要素技術について基礎的な検討を加え，工学上の進歩に貢献することを目的とした．具体的には，①自然な発話状態での入力操作およびコマンド認識・識別処理の自動化を図るため，「口裂」の色彩情報および形状情報に着目した発話区間の自動推定手法について検討を加えた．②多数の利用者・コマンドを用いた場合の認識・識別精度を向上させるため，非発話状態の口唇形状特徴に着目した利用者のグループ化法に関する検討を加えた．③発声の有無に起因する口唇の動き推移の変動について解析を行い，無声発話データと有声発話データの判別に関して工学的な観点から検討を加えた．

1.2 口唇の動き特徴に着目したコマンド識別・発話認識に関する研究（分野）の概観

1.2.1 発話区間の推定に関する従来研究

発話に伴う口唇の動き特徴に着目したコマンド識別・発話認識を行う場合，識別・認識処理の前段階として全撮影区間から発話区間のみを抽出することが求められる．発話区間の抽出処理は，コマンド識別・発話認識自体の精度に直結する非常に重要なプロセスである．このため，従来研究においても様々な発話区間抽出手法が試みられており，その例として，音声情報と画像情報を併用した手法[24][25]，画像情報のみを利用した方法[20][21][26]，ならびにオペレータの手動による方法[22]が挙げられる．

二宮氏ら[24]は自動車内におけるドライバユーザビリティの見地から，音声と画像情報の統合による発話区間の検出手法を報告している．また，山口氏ら[25]はハンズフリーインタフェース機器の操作を行うための手法として，音声と口唇縦線画像を融合した発話区間の検出手法について報告している．二宮氏ら，山口氏らの手法は，いずれも音声と画像情報のバイモーダルによる手法であり，基本的には音声認識を主体として構成している手法である．このため，オフィスなどの静粛性を求められる環境下や，喉の不調によって発声が困難となった状況下での使用は想定されていない．

一方，画像情報のみを対象として発話区間を検出する手法[20][21][26]に関する研究では， Huang 氏ら[26]は YIQ 表色系に着目した口唇領域抽出を行い，その 2 値化画像における口唇領域幅と領域サイズ変動を特徴量とした発話区間検出を行っている．Huang 氏らの報告では発話区間の検出が可能であることを明らかにしているものの，実験に用いた発話内容および被験者数が明示されていない．齋藤氏らのトラジェクトリ特徴量による単語読唇手法[20]では，発話区間を口唇領域の高さ変化から検出している．この手法では，初期状態（非発話状態）の平

Akita University

均値を基準とした判定を行っており，初期状態フレーム取得のために複数フレームを必要とする．このため，1 フレームの画像情報のみで口の開閉状態を判別するには至っていない．中西氏らの手法[21]では，発話開始時に一定時間開口状態にすることを被験者に指示しており，自然な形での発話特徴利用という観点において課題を有する．したがって，上記課題を解決する手法の開発は，画像情報による非接触型インタフェースに極めて有用である．

1.2.2 利用者の増加を想定したコマンド識別・発話認識精度向上に関する従来研究

口唇の動き特徴を利用したコマンド識別・発話認識は音声を伴わずに利用可能であるため，静寂を要する環境や雑音環境下など，様々な場面で利用可能である．しかしながら，口唇形状の時系列変化などは，発話慣れや体調・心情変化などの影響を受けるため[23]，得られるデータにはあいまいさが存在する．このため，システム利用者の増加に伴い，(1)口唇の動き特徴の類似する利用者も増加すること，(2) 登録情報の類似するケースが増加すること，(3)登録データとの照合処理の負荷が増大することが予想される．

発話に伴う口唇の動き特徴に着目した従来研究[18]-[22]において，最も多くの被験者を対象としているのが，渡邊氏らの障がい者・高齢者とのコミュニケーション支援への応用に関する研究であり，被験者 25 名を対象に実験を行っている．しかしながら，認識対象の口形パターン数は 4 パターンと非常に少ない[19]．一方，佐藤氏らは，個人認証（ログイン）からコマンド入力装置まで一体的に行うインタフェースを提案し，被験者数 9 名を対象に，比較的多数の発話内容である 20 コマンドを用いて検討を行っている．その結果，20 コマンドを認識階層に基づいてグループ化することは，コマンド識別におけるロバスト性向上に寄与することを明らかにしている[22]．すなわち，佐藤氏らの研究成果は，利用者についても特定の指標に基づいてグループ化し，照合対象を絞り込むことで識別精度の向上が期待できることを示唆するものである．

口唇は「形状」や「色」といった身体的特徴も有しており，独特の形状を有するいくつかの局所部位（図 1.1 参照）により各個人の口唇が形成されている[27]．この口唇の局所形状特徴は個人ごとに異なるが，器官としての基本的な形状は共通しているため，いくつかの類似形状群に大別できることが予想される．この点に着目し，口唇形状に関する統計的な解析を行うことは，利用者をグループ化する指標を得る上で重要な課題と考える．

身体的特徴としての口唇形状に着目した従来研究では Travieso 氏らによる顔と口唇のバイモーダル認証[28]，Sforza 氏らによる口唇形状の加齢変化や性別による差異に関する研究[29]などが行われている．Travieso 氏らの研究は顔と口唇のバイモーダルによる認証であり，口唇の局所形状に着目した特徴量抽出は行っていない．また，Sforza 氏らの研究は，口唇形状に着目した識別対象者のグループ

Akita University

化や絞り込み手法を目的としたものではなく，口唇形状の計測には特殊な機器である電磁式 3 次元ディジタイザを必要とするなどの課題を有している．

以上のように，一般的なビデオカメラで取得した口唇画像をコマンド識別へ応用するという観点から，口唇局所部位の形状解析，ならびに分類による対象者絞り込みを行う手法，特に口裂（Oral fissure）形状を主眼点として行われた研究は，筆者らの調査した範囲では見当たらない．

1.2.3 発話状態とコマンド識別・発話認識精度の関連に関する従来研究

発話に伴う口唇の動き特徴は「行動的特徴」[1] であるため，同一の利用者が同一の内容を発話した場合においても，得られる特徴量は発話ごとに僅かずつ異っている．また，利用者の体調や心情が定常時と異なる場合には，口唇の動きにおけるばらつき度合が増加することも明らかになっている[2 3]．したがって，口唇の動き特徴を応用するシステムの構築には，「行動的特徴」であることに起因する口唇の動き特徴の変動を考慮した認識技術の開発が必要である．特に，発声の有無にかかわらず特徴量が取得可能という利点をコマンド識別・発話認識システムで実現するためには，発声の有無に起因する口唇の動き特徴変動の影響を考慮した認識技術が必要となる．この技術は，利用環境に関する条件を緩和し，さらには音声認識との併用においても重要となる．しかしながら，筆者らの調査した範囲において，発声と口唇の動き特徴の関連を研究した事例は見当たらないのが現状である．

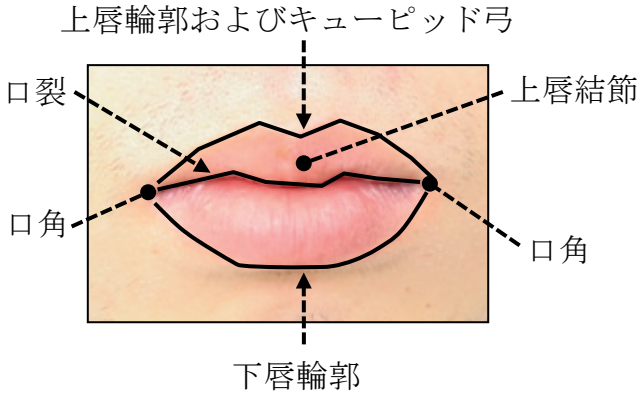


図 1.1　口唇各部位の名称

Akita University

1.3 本論文の内容

本論文は全 5 章より構成され，第 1 章を緒論とした．

第 2 章では，発話区間の自動推定処理ための要素技術について検討を加え，発話時の口唇画像における L*a*b*表色系[30]の色彩情報および口唇形状の時系列変化を特徴量とする発話フレーム自動検出法を提案した．具体的には，口唇画像における垂直方向の L*および a*の推移特徴から口裂（閉口時の上唇と下唇の境界）を検出し，各フレームにおける口の開閉状態を判定する．次に，3 フレーム間における口唇形状の時系列変化から発話の過程で閉口状態となったフレームを検出する．5 つの母音全てを含む人名を発話内容として，被験者 5 名による実験を行った結果，約 99%と高精度で発話フレーム検出が可能であることを明らかにした．

第 3 章では，発話内容の識別精度向上を目的とし，非発話状態の口唇形状特徴を解析し，解析結果に基づいた口唇形状のグループ化法を提案した．具体的には，口唇を 3 つの矩形領域に分割して，①上唇および下唇の厚さ比率特徴量，②口裂の凹凸形状特徴量，③アスペクト比を算出し，その解析結果に基づいて口唇形状を 27 のカテゴリに分類した．なお，撮影環境の微小な変化によるあいまいさを考慮し，ファジィ推論を用いて形状分類処理を行った．被験者 52 名を対象に分類実験を行ったところ，80%以上の精度で登録データおよびその類似形状と同一の形状に分類され，4 位カテゴリまでに分類可能であった被験者において，照合対象を約 1/8.5 に絞り込み可能であることを示した．

第 4 章では，発話内容の識別精度向上を目的とし，発声の有無に起因する口唇の動き特徴変動に関して検討を加えた．同一の被験者が同一の内容を発話した場合においても，無声発話と有声発話では発話の長さや口唇の動き特徴に変動が生じる．そこで，発話フレーム数および口唇の動き特徴（横幅，縦幅，面積，アスペクト比）に着目し，発声に起因する特徴量の変動について検討を加えた．その結果，無声発話時は有声発話時と比較し，発話フレーム数および口唇の動き特徴量の累積差分値が大きくなること，特徴量の変動が大きな被験者と微小な被験者に大別されることを明らかにした．さらに，同一取得日の無声発話データと有声発話データの判別手法についても検討を加え，発話フレーム数および口唇動作量に着目することで，約 92%の割合で無声発話データと有声発話データを判別可能であることを明らかにした．

第 5 章では，本研究で得られた主な成果と本論文の工学的意義および今後に残された課題について述べている．

Akita University

1.4 本論文で用いる主な用語

本論文で使用する用語について以下に解説を加える．なお，用語については文献(1)，(27)，(30)，(31)を参考にしてまとめた．

- **L*a*b*表色系**：1964 年に CIE（国際照明委員会）が均等知覚色空間として勧告した表色系の一つであり，色の分布を人間の感覚に近い形で捉えることが可能である．また，色相を等間隔の直線，彩度を等間隔の同心円として近似することが可能な表色系であり，明るさ変化に対するロバスト性が高い．なお，L*は明度指数，a*および b*は色相と彩度を表現する量であり，a*は正方向に大きいほど赤，負方向に大きいほど緑であることを示す．また，b*は正方向に大きいほど黄，負方向に大きいほど青であることを示している．
- **身体的特徴**:生体計測的な生体情報のことを指し，指紋，静脈パターン，虹彩などが「身体的特徴」に相当する．
- **行動的特徴**:行動計測的な生体情報のことを指し，声紋，署名，歩き方などが「行動的特徴」に相当する．随意的な要素があり，一般的には「癖」と認識される．
- **口唇領域**:人体顔面の口部において，表皮の角化度が低く，血管が透けて赤く見える領域．
- **口裂**:上唇と下唇が接する境界領域．

1.5 本論文における個人情報を有するデータの取扱い

本論文に関する各データ（顔画像，音声，被験者のアンケートなど）は「秋田大学手形地区における人を対象とした研究に関する倫理規定第 6 条第 2 項」に基づいて倫理審査の申請を行い，承認された研究計画の下に，被験者本人の了承を得て取得し，これを解析および実験に使用している．

なお，データの取り扱いについては，JIS TR X0086:2003 [32] に記載されている「生体情報データベースの取扱い」に準じている．

Akita University

第 1 章 文献

(1)瀬戸，織茂，寺田，佐藤：「情報セキュリティ概論」，日本工業出版(2008)
(2)中川：「音声認識研究の動向」，信学論, Vol.J83-D-II, No.2, pp.433-457(2000)
(3)荻原，新谷，土居，福永：「視聴覚融合を用いた HMM 音声認識」，電学論 C, Vol.115-C, No.11, pp.1317-1324(1995)
(4)伊田，森，中村，鹿野:「据置き型情報提供端末向き雑音処理を用いた音声入力インタフェース」，信学論(D-Ⅱ), Vol.J84-D-Ⅱ, No.6, pp.868-876(2001)
(5)中川，鳥居，甲斐，中西:「任意語彙の追加登録可能な単語音声認識システム」，電学論 C, Vol.118-C, No.6, pp.865-872(1998)
(6)D. Kahn, L. R. Rabiner, and E. Rosenberg, “On Duration and Smoothing Rules in A Demisyllable-Based Isolated-Word Recognition System”, J. Acoust. Soc. Am., Vol.75, No.2, pp.590-598(1984)
(7)谷萩:「音声と画像のディジタル信号処理」，コロナ社(1996)
(8)真鍋，平岩，杉村:「無発声音声認識 : 筋電信号を用いた声を伴わない日本語 5 母音の認識」，信学論(D-Ⅱ), Vol.J88-D-Ⅱ, No.9, pp.1909-1917(2005)
(9)坂井，河原 編：「カラー図解　人体の正常構造と機能（改訂第 2 版）」，日本医事新報社(2012)
(10)根田，西田，石井，佐藤:「口唇の動き特徴の個人識別法への適用」，電学論 C, Vol.120-C, No.5, pp.765-766(2000)
(11)市野，坂野，小松:「核非線形相互部分空間法を用いた話者認識」，信学論(D-Ⅱ), Vol.J88-D-Ⅱ, No.8, pp.1331-1338(2005)
(12)E. Gómez, C. M. Travieso, J. C. Briceño, M. A. Ferrer, “Biometric Identification System by Lip Shape”, IEEE ICCST'02, pp.39-42(2002)
(13)U. Dieckmann, P. Plankensteiner, and T. Wagner, “SESAM: A Biometric Person Identification System Using Sensor Fusion”, Pattern Recognition Letters, Vol.18, No.9, pp.827-833(1997)
(14)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, Vol.119-D, No.4, pp.465-472(1999)
(15)寺田，吉田，大恵，大橋:「口の形状をパスワードに用いた本人認証」，画像電子学会誌，Vol.30, No.3, pp.267-275 (2001)
(16)白澤，三浦，西田，景山，栗栖:「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, Vol.60, No.12, pp.1964-1970 (2006)
(17)佐藤・西田：「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 C, 125-C, 8, pp.1282-1289(2005)
(18)内村，道田，桃田，佐藤，相田:「人間の読唇傾向に基づく口唇画像情報を用いた単語自動認識の試み」，システム制御情報学会論文誌，Vol.4，No.4, pp.163-172(1991)

Akita University

(19)渡邊，西:「口部パターン認識を用いた日常会話伝達システムの研究」，電学論 C, Vol.124-C, No.3, pp.680-688(2004)
(20)齊藤，小西:「トラジェクトリ特徴量に基づく単語読唇」，信学論 D, Vol.J90-D, No.4, pp.1105-1114(2007)
(21)中西，寺林，梅田：「インテリジェントルームのための DP マッチングを用いた口唇動作認識」，電学論 C, Vol.129, No.5, pp.940-946(2009)
(22)佐藤，景山，西田：「口唇の動き特徴を用いた非接触コマンド入力インタフェースの提案」，電学論 C, Vol.129, No.10, pp.1865-1873(2009)
(23)景山，安東，西田：「発話に伴う口唇の動き特徴を用いた心情変化の検出」，電学論 C, Vol.131, No.1, pp.201-209(2011)
(24)二宮，坂，前野，根木，宮島，森，北坂，末永：「音声と画像の統合によるドライバの発話区間検出」，映情学誌, Vol.62, No.3, pp.435-441(2008)
(25)山口，浜田：「音声と口唇縦線画像を融合した発話区間検出法」，信学技報, HIP2007-161, pp.13-18(2007)
(26)Y. Huang, H. Dohi, M. Ishizuka: “Man-Machine Interaction Using a Vision System with Dual Viewing Angles”, IEICE Trans. INF. & SYST., Vol.E80-D, No.11, pp.1074-1083(1997)
(27)小野，飯塚，吉竹：「口腔外科学（第 6 版）」金芳堂(2002)
(28) C. M. Travieso, J. Zhang, P. Miller, J. B. Alonso, and M. A. Ferrer: “Bimodal biometric verification based on face and lips”, Neurocomputing, Vol.74, pp.2407-2410(2011)
(29) C. Sforza, G. Grandi, M. Binelli, C. Dolci, M. D. Mendes, and V. F. Ferrario: “Age- and sex-related changes in three-dimensional lip morphology”, Forensic Science International, Vol.200, pp.182.e1-182.e7(2010)
(30)日本色彩学会編：「新編　色彩科学ハンドブック（第 3 版）」,東京大学出版会(2011)
(31)瀬戸：「サイバーセキュリティにおける生体認証技術」，共立出版(2002)
(32)日本規格協会：「顔認証システムの精度評価方法」，TR X 0086：2003(2003)

Akita University

第 2 章　口唇の色彩情報および形状情報に着目した発話フレームの検出

2.1 はじめに

発話に伴う口唇の動き特徴を応用したコマンド識別・発話認識手法において，発話区間の抽出処理は，コマンド識別・発話識別の精度に直結する重要なプロセスである．佐藤氏らの手法では，オペレータの手動による発話区間の抽出を適用しており[(1)]，Huang 氏ら，齋藤氏ら，中西氏らは発話区間の自動抽出を適用したコマンド識別・発話識別手法を報告している[(2)－(4)]．なお，二宮氏らや山口氏らも発話区間の自動抽出法[(5)(6)]を報告しているが，これらは音声情報を主体とするものである．

画像情報のみを利用して発話区間を決める手法として，Huang 氏ら[(2)]は YIQ 表色系に着目した口唇領域抽出を行い，その 2 値化画像における口唇領域幅と領域サイズ変動を特徴量とした発話区間検出を行っている．Huang 氏らの報告では発話区間の検出が可能であることを明らかにしているものの，実験に用いた発話内容および被験者数が示されていない．一方，齋藤氏らの手法[(3)]では，初期状態（非発話状態）の口唇領域の高さの平均値を基準として判定を行っており，平均値取得のために初期状態フレームを複数フレーム必要とする．さらに，中西氏らの手法[(4)]では，発話開始の合図として一定時間開口状態を維持する必要があるなどの課題を有している．

そこで本章では，口唇の形状特徴の 1 つである口裂に着目し，口裂における色彩情報を特徴量として，発話動画像の画像情報のみを用いた発話区間に属する各フレーム（以後，「発話フレーム」と呼ぶ）の自動検出手法を提案した．提案手法は，L*a*b*表色系[(7)]で表現した口唇の色彩情報を特徴量として口の開閉状態を判別する処理，ならびに口唇画像の時系列変化を指標として発話状態を判別する処理の 2 段階で構成される．まず前段の処理では，色彩情報に基づき，フレーム単位で口の開閉状態を判別する．次に，前段の処理において閉口状態と判別された画像を対象に口唇形状の時系列変化を検出し，発話中に生じる閉口状態か否かを判定する．これら前段・後段の処理によって，フレーム単位で高精度に発話フレームの検出を行う手法である．

なお，色彩情報解析は被験者 24 名（Ex1-id001～Ex1-id024）を対象として実施した．また，提案手法の有用性を検証するため，日本語の母音すべてを含む 3 つの発話内容（5 つの母音全てを含む人名）を設定し，被験者 5 名（Ex1-id025～Ex1-id029）を対象として発話フレーム検出の評価実験を行った．

Akita University

2.2 使用データ

2.2.1 データ取得方法

通常の発話において，口は発話に伴い開閉動作を繰り返す．このため，口の開閉状態を判別することで, 発話区間の大部分を決定することが可能と考える. そこで, 発話時における口唇の開閉状態を解析するための画像データセットである「データセット A」，ならびに提案手法を評価するための画像データセットである「データセット B」の 2 種類を使用データとして取得した．データ取得環境を図 2.1 に，取得データ例を図 2.2 にそれぞれ示す．また，データセット A，B に共通するデータ取得条件を以下にまとめる.

- 日常一般的と考えられる室内環境(蛍光灯下，照度 600～1000lx).
- 口紅の塗布や補助的な照明は使用しない.
- 発話の前後には，軽く口を閉じるよう指示.
- 口唇の動きを捉えやすくするため，被験者を正面から撮影.
- 被験者とカメラの位置は約 60cm.
- 被験者は全員モンゴロイド（日本人）20 代男女.
- 3CCD ビデオカメラ（SONY 製 DCR-VX2100）を使用.

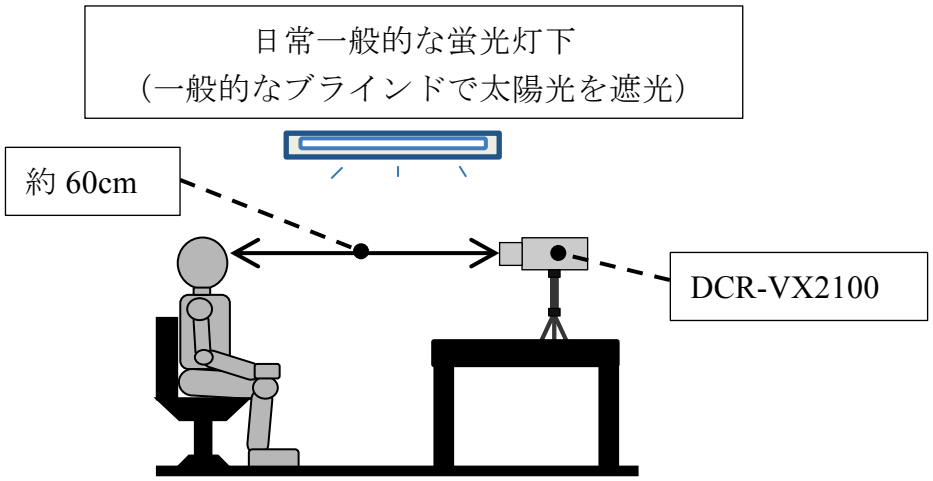


図 2.1　データ取得環境

図 2.2　取得データ例

Akita University

2.2.2 口唇領域の特徴解析用データ（画像データセット A）

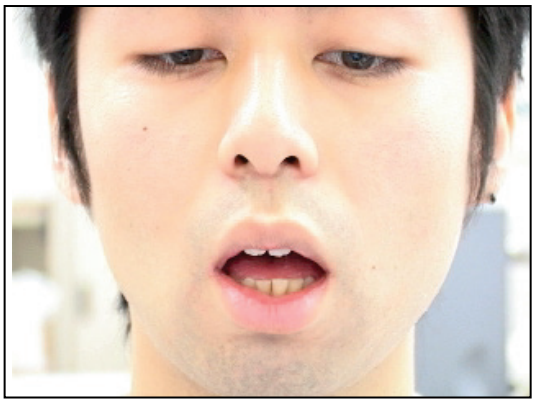
口の開閉状態における特徴解析を行うため，被験者 24 名（被験者 Ex1-id001～Ex1-id024）を対象とし，開口状態と閉口状態の顔画像をそれぞれ撮影した．

状態①: 口を軽く閉じた状態（図 2.3(a)参照）

状態②: 口を開けた状態（図 2.3(b)参照）

撮影した静止画像（データセット A）に対して色彩情報を用いた口唇形状自動抽出法[8]を施し，得られた口唇領域内すべての色彩情報を抽出した．なお，口唇領域内において，口唇クラス以外の画素についても色彩情報を取得し，これを解析に用いた．

(a)口を軽く閉じた状態（状態①）



(b)口を軽く開けた状態（状態②）

図 2.3　口の開閉状態の例（Ex1-id013）

Akita University

2.2.3 評価用データ（画像データセット B）

被験者 5 名（被験者 Ex1-id025～Ex1-id029）を対象に，3 つの単語を任意の間隔で発話させた場面をそれぞれ 4 回分取得した．発話内容は，全ての母音（ア，イ，ウ，エ，オ）を含む人名「オオダテケンシロウ」(Word1)，「カゲヤマヨウイチ」(Word2)，「アキタウメコ」(Word3）である．なお，発話に伴い，口唇の形状が変化した場合においても画像中に口唇が含まれる必要があるため，データ取得の際に用いる確認用モニタに横幅 75 画素，縦幅 40 画素の矩形を表示し，被験者が発話を行うときの指標とした（図 2.4 参照）．さらに図 2.5 に示すように，撮影した被験者 Ex1-id025～Ex1-id029 の発話動画像を 30fps の時系列静止画像（320×240 画素，32 ビットのカラー画像）に変換して得た合計 4447 枚の静止画像を発話フレーム検出用データ（データセット B）とした．なお，被験者により差異はあるものの，各発話内容（人名）のフレーム数は，「オオダテケンシロウ」の場合では 38～51 フレーム程度（約 1.26～1.70 秒)，「カゲヤマヨウイチ」の場合では 34～38 フレーム程度（約 1.13～1.26 秒)，「アキタウメコ」の場合では 23～44 フレーム程度（約 0.77～1.47 秒）である．

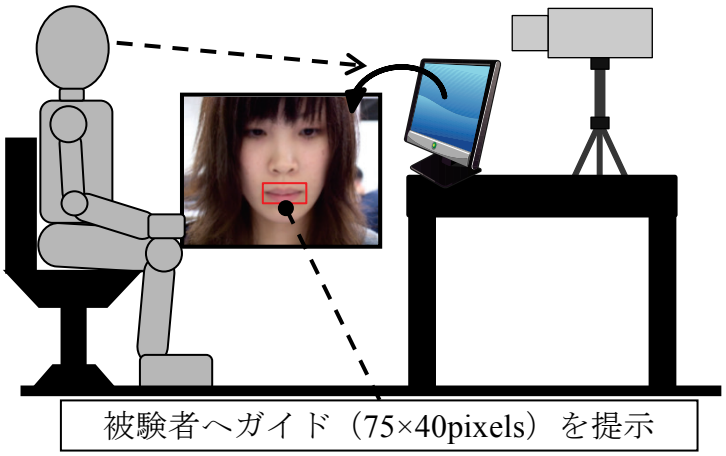


図 2.4　データ取得時の被験者へのガイド

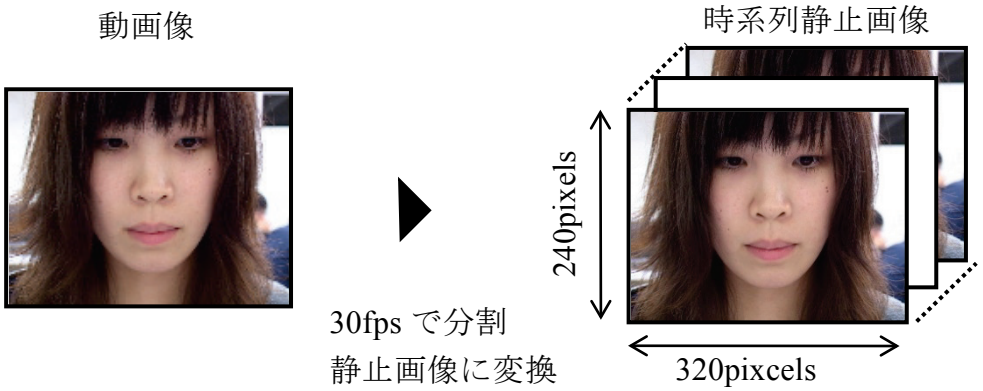


図 2.5　時系列静止画像への変換

Akita University

2.3 口唇領域の色彩情報解析

口を閉じた場合，口唇領域内に上唇と下唇の境界領域である「口裂」が生じる（図 1.1 参照）．この口裂の色彩情報は，上唇および下唇と異なる傾向を有する．そこで本研究では，データセット A に対して「口唇の赤み」および「口唇領域の明度」に着目し，L*a*b*表色系を用いた口唇の色彩情報の解析を行った．

2.3.1 色彩情報を用いた従来研究

ディジタル画像で一般的に使用される RGB 表色系は，明度が色彩情報から分離されていないため，照明の変動が RGB 値に大きく影響を与える[7]．このため，明度と色彩情報とが独立した表色系を用いた研究が行われている．口唇領域を良好に抽出している従来手法としては，YIQ 表色系を用いた手法[2]，HLS 表色系を用いた手法[3]，HSV 表現を用いた手法[9]，L*a*b*表色系を用いた手法[1][8]が挙げられる．Huang 氏らの YIQ 表色系を用いた手法[2]は，口唇領域の抽出に色彩情報を用いており，抽出領域の幅と領域サイズの変動量による発話状態の検出を試みている．齋藤氏らの HLS 表色系を用いた手法[3]は，色彩情報を口唇位置の推定に利用しており，得られた口唇輪郭長または口唇輪郭の高さの変動量を用いて発話区間の検出を試みている．上記に示した 2 種類の手法は良好に口唇形状を抽出し，得られた特徴量から発話区間を自動推定している．しかしながら，複数の単語を任意の間隔で連続して発話する場合，ならびに母音などの発話内容に関する検討は十分に行われていない．また，黒田氏らの HSV 表現を用いた手法[9]，白澤氏らの手法[8]ならびに佐藤氏らの手法[1]では，発話区間または発話状態の検出に至っていない．この様に，色彩情報を利用した発話状態検出に関して，自然な発話状態までを考慮した検討は十分に行われていないのが実情である．

白澤氏らが提案した L*a*b*表色系に着目した口唇形状自動抽出法は，口唇を良好に抽出可能であり，個人識別にも応用可能であることが明らかとなっている[8]．また，佐藤氏らは白澤氏らの手法を用いて口唇を抽出し，コマンド入力インタフェースに応用可能であることを明らかにしている[1]．このことは，L*a*b*表色系を用いることで，個人差を考慮した発話区間の推定が可能であり，さらに個人識別および発話認識へと連携した処理が可能であることを示唆している．そこで本研究では，色彩情報に L*a*b*表色系を採用し，白澤氏らの提案した口唇形状自動抽出法[8]を用いて口唇抽出を行い，その色彩情報の特徴解析を行った．

Akita University

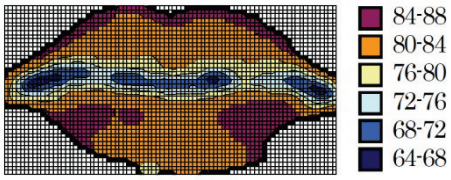
2.3.2 口唇の色彩情報とその特徴

口裂は，上唇と下唇の境界領域であり，口唇の中間部分に谷間のような形状を形成している．このため，口裂は他の口唇領域と比較して明度が低く，赤みが深く見える領域である．L*a*b*表色系は明度値 L*，赤みを表す知覚色度指数 a*を有していることから，口唇の陰影特徴や赤み分布を良好に解析可能と考える．そこで，閉口状態ならびに開口状態における口唇明度値 L*と赤み a*の色彩情報解析を行った．

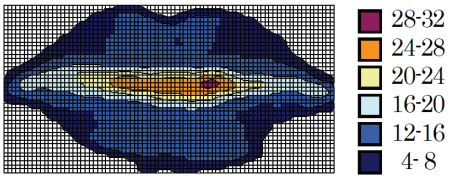
(1)閉口状態の特徴

口を閉じた場合の L*および a*の分布例を図 2.6 に示す．L*に着目すると，上唇や下唇と比較し口裂付近において低い値を有していることがわかる．一方，a*の場合，上唇や下唇と比較し口裂付近において高い値を有している．

次に，口唇中央の垂線上における L*および a*の推移を調査した．推移例を図 2.7 に示す．口裂近傍の画素において，L*の最小値（極小値）および a*の最大値（極大値）を有していることがわかる．なお，これらの口裂における色彩情報は，被験者 24 名全員のデータにおいて，類似した傾向を有することを確認している．



(a) L*の分布例



(b) a*の分布例

図 2.6　閉口状態（状態①）における L*および a*の分布例（Ex1-id013）

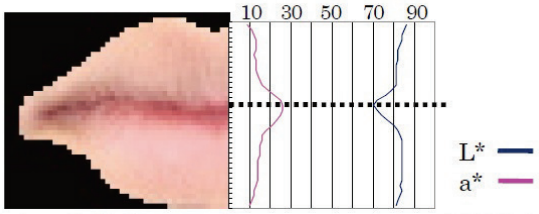


図 2.7　閉口状態（状態①）における L*および a*の推移例（Ex1-id013）

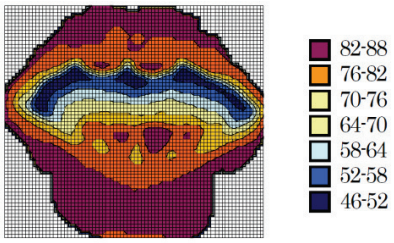
Akita University

(2)開口状態の特徴

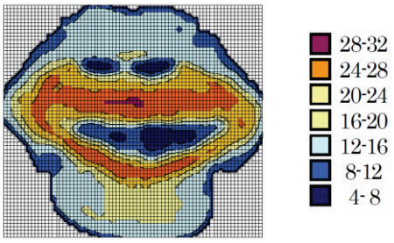
口を開けた場合の L*および a*の分布例を図 2.8 に示す．開口状態では上唇と下唇の境界に口内領域の現れていることがわかる．また，口内領域は各部（口腔，舌，歯）において，それぞれ特有の色彩情報を有している．L*に着目すると，上唇や下唇周辺と比較し，口内領域（特に口腔の奥）において低い値を有している．一方，a*では上唇や下唇および舌の領域において高い値を有する傾向，歯の領域において低い値を有する傾向をそれぞれ認めた．

次に，口唇中央の垂線上における L*および a*の推移を調査した．推移例を図 2.9 に示す．L*では口腔の奥行きの深い部分に極値（極小）を持つ傾向が見られたものの，a*では開口幅や歯並びなどにおいて個人差が大きく現れ，被験者に共通する推移の形状を得ることができなかった．

以上の結果は，L*および a*を指標とし，閉口時に存在する口裂を検出することで，口の開閉が判別できる可能性のあることを示唆している．この特徴解析結果を踏まえ，本研究では口裂の有無に着目した発話フレーム検出を行った．



(a) L*の分布例



(b) a*の分布例

図 2.8　開口状態（状態②）における L*および a*の分布例（Ex1-id013）

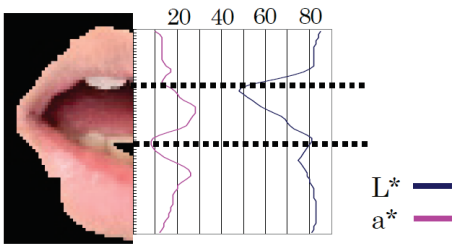


図 2.9　開口状態（状態②）における L*および a*の推移例（Ex1-id013）

Akita University

2.4 発話フレーム検出法

開口状態，あるいは口唇形状が時系列的に変化している状態を発話状態であると仮定し，その状態に適合する発話画像フレームを発話フレームとして検出する．図 2.10 に発話フレーム検出処理の流れを示す．また，処理の流れを以下にまとめる．

① 口唇形状自動抽出法[8]を施し，原画像から口唇形状を抽出する．
② 処理①で得られた口唇画像に対して，口裂の有無を判定する処理を行う（2.4.1 項に後述）．処理②において，口裂が認められない場合，対象フレームを発話フレームである（開口状態）と判定し，処理を終了する．
③ 処理②において，口裂の認められる場合，口唇形状の変化を判定する処理を行う（2.4.2 項に後述）．
④ 処理③において，口唇形状の変化が認められない場合，対象フレームは発話フレームではないと判定し，処理を終了する．
⑤ 処理③において，口唇形状の変化が認められた場合，対象フレームが発話フレームである（口唇形状が変化）と判定し，処理を終了する．

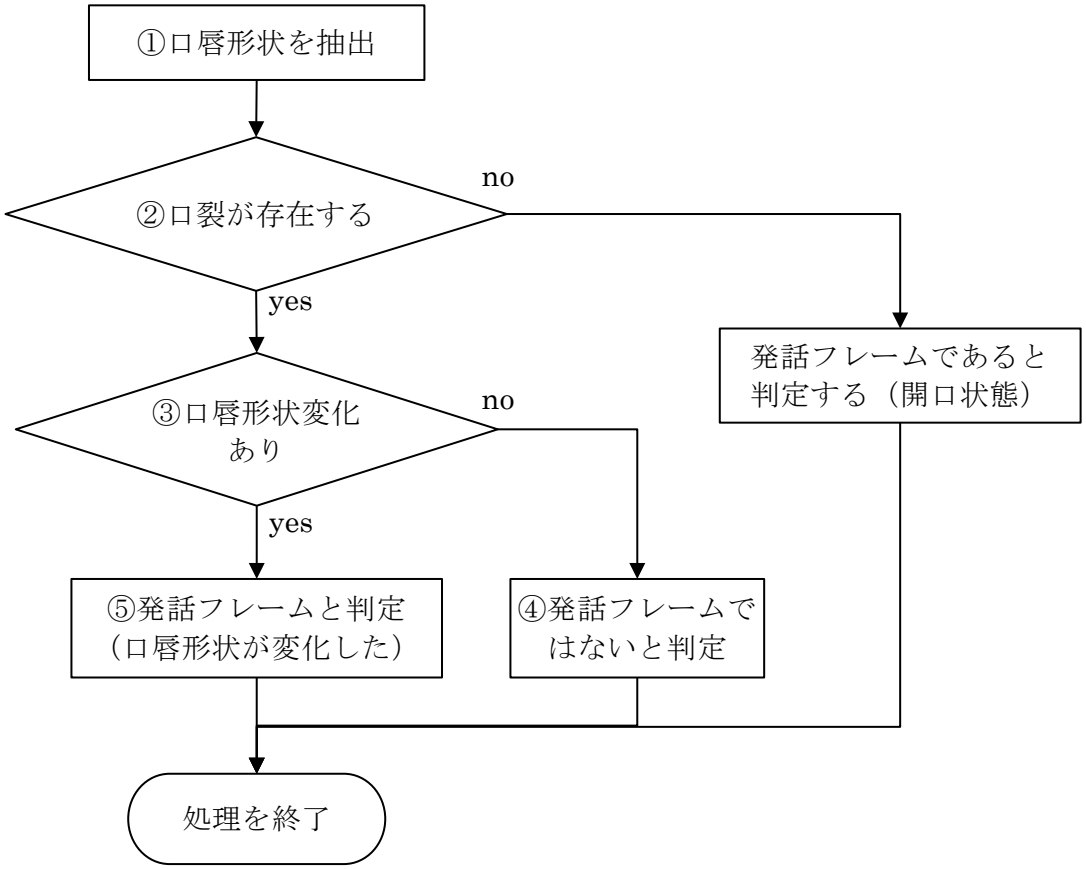


図 2.10 発話フレーム検出処理の流れ

Akita University

2.4.1 口裂判定処理

(1)色彩情報の取得

口唇の垂直方向の色彩情報を取得するため，図 2.11 に示す口唇の垂直方向の色彩情報を走査する垂線 P_k を設定した．本論文では，これを「口裂判定垂線」と定義する．上唇結節周辺は，口の開閉動作が最も顕著に表出すると考えられるため，口裂判定垂線 P_k は，口唇の横幅を 1:1 に内分する上唇結節周辺の位置を基準として設置した．

(2)口裂判定処理

口を開き始める時は口角（口唇の左右端）周辺と比較して，上唇結節（上唇の下中央部のふくらみ）周辺の動きが顕著である様子を認めた．そこで提案手法では，口裂判定垂線 P_k の L*および a*の推移を求め，全ての P_k において，以下に示す判定基準Ⅰ，Ⅱを満たす場合に「口裂あり」と判定した．

- **判定基準Ⅰ**：垂線上の L*および a*の推移において，L*推移における最も小さい極小値を持つ画素の近傍 6 画素に a*の極大値をもつ画素が存在する．

- **判定基準Ⅱ**：垂線上の画素を口唇クラスとその他のクラスに分類（色彩情報を用いたファジィ推論による自動分類[8]）した結果において，口唇クラスのみ存在する．

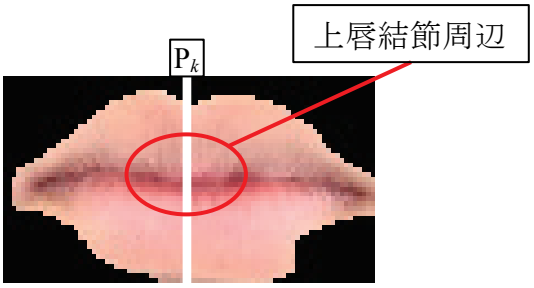


図 2.11　口裂判定垂線 P_k の設定イメージ

Akita University

2.4.2 口唇形状の時系列変化判定処理

口唇の形状は発話内容にしたがって連続的に変化する．このため，発話内容次第では，図 2.12 におけるフレーム No.88，No.89 のように，発話中に口を閉じる（口裂の生じる）動きが発生する場合が認められる．そこで，口裂が存在すると判定されたフレームに対し，フレーム間の口唇形状差分に着目した形状変化判定処理を施した．また，発話に伴う口唇の主な動きは上下方向の開閉と左右方向の伸縮に大別されることから，口唇の横幅 *diX*，縦幅 *diY* （図 2.13 参照）に変化が生じると考えられる．

そこで提案手法では，発話時の口唇の開閉動作に着目し，口唇形状のフレーム間差分から口唇の時系列変化を示す特徴量 *A* を算出し，*A* が形状変化のしきい値 Th_A 以上となった場合に「口唇形状変化あり」と判定した．

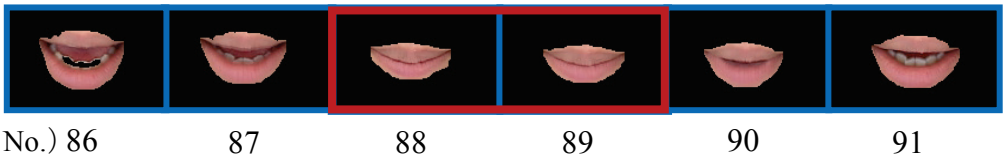


図 2.12　発話に伴う口唇形状の時系列変化例

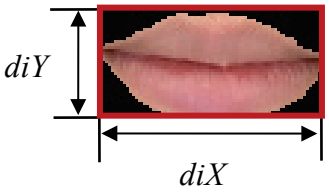


図 2.13　口唇縦幅 *diY* および口唇横幅 *diX*

Akita University

2.5 検出条件の検討および評価方法

2.5.1 口裂判定位置に関する比較

口裂判定垂線 P_k の有する情報は，口唇領域における設定位置によって異なると考えられる．また，その設定本数も口裂の検出精度に係る要因となることが予想される．そこで，上唇結節周辺の情報に加えて左右の口角に近い位置の情報も取得するため，図 2.14 に示すように，口唇の横幅を 1:1 に内分する垂線 P_1，2:3 に内分する垂線 P_2，3:2 に内分する垂線 P_3，1:4 に内分する垂線 P_4，4:1 に内分する垂線 P_5 の 5 本の口裂判定垂線を設定し，次の 3 通りについて発話フレーム検出精度の比較を行った．

(a)上唇結節の中央部の垂線 P_1 上の画素で判定．

(b)垂線 P_1～P_3 上の画素で判定．

(c)垂線 P_1～P_5 上の画素で判定．

用いた垂線全てにおいて，判定基準 I および判定基準 II を満たした場合に「口裂あり」と判定した．

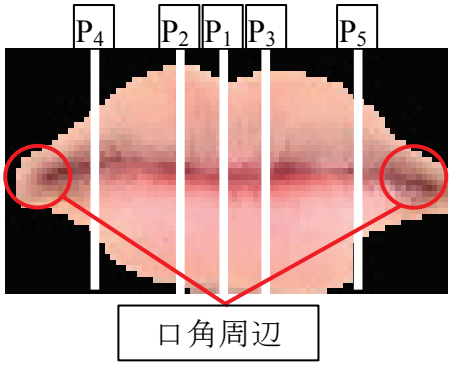


図 2.14　口裂判定垂線 P_1～P_5 の設定

Akita University

2.5.2 口唇形状の時系列変化検出に関する比較

(1)対象フレーム範囲の比較

図 2.12 の No.88，No.89 のように，発話時に口を閉じる動きが発生した場合，形状変化の判定対象となるフレームと直前フレームとの差分値が非常に小さくなる事例を多数認めた．そこで，形状変化判定における着目フレーム数について，2 フレームの場合と 3 フレームの場合の 2 通りを比較した．2 フレームの場合の特徴量は，判定対象フレーム F_i と直前フレーム F_{i-1} から算出される A_1 のみであり（図 2.15 参照），A_1 がしきい値 Th_A 以上となった場合に「口唇形状変化あり」と判定した．これに対して， 3 フレームの場合の特徴量は，判定対象フレーム F_i と直前の 2 フレーム F_{i-1}，F_{i-2} から算出される $A_{1\sim3}$ となる．得られた特徴量 $A_{1\sim3}$ のいずれかの値がしきい値 Th_A 以上となった場合に，「口唇形状変化あり」と判定した．なお，形状変化判定のしきい値については，Th_A を 1 から 6 まで 1 刻みで検討を行った結果，良好な判定結果を得た Th_A=3 を採用した．

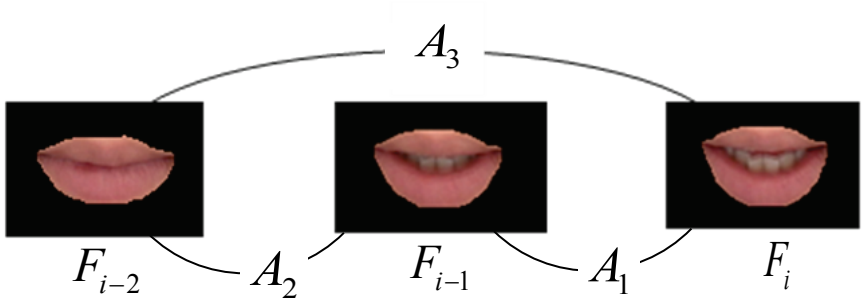


図 2.15　フレーム間における口唇形状変化判定の特徴量算出

Akita University

(2)口唇形状の時系列変化のための特徴量算出方式の比較

発話における口唇の主要な動作は左右方向の動きと，上下方向の動きである．また，発話時の口唇の動きにおいて上下方向の動きは開閉動作を司り，横方向の伸縮動作よりも顕著となることが推測される．そこで，縦方向の動きのみを考慮した特徴量 A_j（j=1 から 3）と縦方向に加えて横方向の動きも考慮した特徴量 A'_j を用いた 2 通りについて，発話フレームの検出率を比較した．

A_j は縦幅 diY のみを用いたフレーム間差分値として(2.1)～(2.3)式で算出し（以後，特徴量算出処理(ⅰ)と呼ぶ），A'_j は口唇の横幅 diX および縦幅 diY のフレーム間差分値から得られるベクトルとして(2.4)～(2.6)式で算出した（以後，特徴量算出処理(ⅱ)とよぶ）．

$$A_1 = | diY_i - diY_{i-1} | \quad \cdots\cdots\cdots(2.1)$$

$$A_2 = | diY_{i-1} - diY_{i-2} | \quad \cdots\cdots\cdots(2.2)$$

$$A_3 = | diY_i - diY_{i-2} | \quad \cdots\cdots\cdots(2.3)$$

$$A'_1 = \sqrt{(diX_i - diX_{i-1})^2 + (diY_i - diY_{i-1})^2} \quad \cdots\cdots\cdots(2.4)$$

$$A'_2 = \sqrt{(diX_{i-1} - diX_{i-2})^2 + (diY_{i-1} - diY_{i-2})^2} \quad \cdots\cdots\cdots(2.5)$$

$$A'_3 = \sqrt{(diX_i - diX_{i-2})^2 + (diY_i - diY_{i-2})^2} \quad \cdots\cdots\cdots(2.6)$$

Akita University

2.5.3 発話フレーム検出の評価方法

オペレータ（1 名）の目視により得られた正解発話フレームおよび正解非発話フレームとの比較を行い，検出されたフレームの評価を行った．評価の指標として，正解発話フレーム数 N_{sf} に対する検出フレーム数 N_{dt} の一致数から検出率 R_{hit} および未検出率 R_{miss} を，正解非発話フレーム数 N_{nsf} に対する誤検出フレーム数 N_{fdt} から誤検出率 R_{false} を求めた. R_{hit} は(2.7)式, 未検出率 R_{miss} は(2.8)式, R_{false} は(2.9)式によりそれぞれ算出した.

$$R_{hit} = \frac{N_{dt}}{N_{sf}} \times 100 \qquad \cdots\cdots\cdots(2.7)$$

$$R_{miss} = \frac{N_{sf} - N_{dt}}{N_{sf}} \times 100 \qquad \cdots\cdots\cdots(2.8)$$

$$R_{false} = \frac{N_{fdt}}{N_{nsf}} \times 100 \qquad \cdots\cdots\cdots(2.9)$$

Akita University

2.6 実験結果および考察

判定手法の組合せ条件の数は口裂判定に関して 3 通り，口唇形状の時系列変化判定に関して 4 通りの合計 12 通りとなる．これら 12 通りの条件全てについて発話フレームの検出実験を行い，その検出精度を比較した．

対象とした画像データは，データセット B の全データ（4447 枚）である．比較評価のために，オペレータ（1 名）の目視判別に基づいて正解発話データを決定した．なお，得られた正解発話フレーム数は 2470 枚，正解非発話フレーム数は 1977 枚である．

2.6.1 発話フレーム検出結果

(1)検出率および未検出率に関する考察

発話フレーム検出結果を図 2.16 および図 2.17 に示す．検討に用いた口裂・口唇形状判定条件 12 通り全てにおいて 90%以上と高い検出結果が得られた．一つ目の比較項目である口裂判定の垂線数に着目すると，垂線数 1 本（P_1）の場合の検出率平均は約 93.9%，垂線数 3 本（P_1～P_3）の場合の検出率平均は約 97.7%，垂線数 5 本（P_1～P_5）の場合の検出率平均は約 97.7%となり，3 本および 5 本とした場合に発話フレームを良好に検出可能であることがわかる．

次に，形状変化判定のフレーム数に着目すると，2 フレームを用いた場合の検出率平均は約 94.6%，3 フレームを用いた場合の検出率平均は約 98.2%となり，3 フレームを用いた手法が良好な結果となった．さらに，特徴量の算出処理における比較では，算出処理(ⅰ)を用いた場合の検出率平均は約 96.1%，算出処理(ⅱ)を用いた場合の検出率平均は約 96.7%となり，ほぼ同等の結果を得た．

以上の結果から，口裂判定の垂線数を 3 本または 5 本とし，形状変化判定には 3 フレームを用いることで良好な検出が可能になることが明らかとなった．また，算出処理(ⅰ)と算出処理(ⅱ)の正解フレーム検出率に大きな差は認められないことから，口唇の縦幅 *diY* は横幅 *diX* と比較し，形状変化の判定において重要度の高いことが示唆された．

Akita University

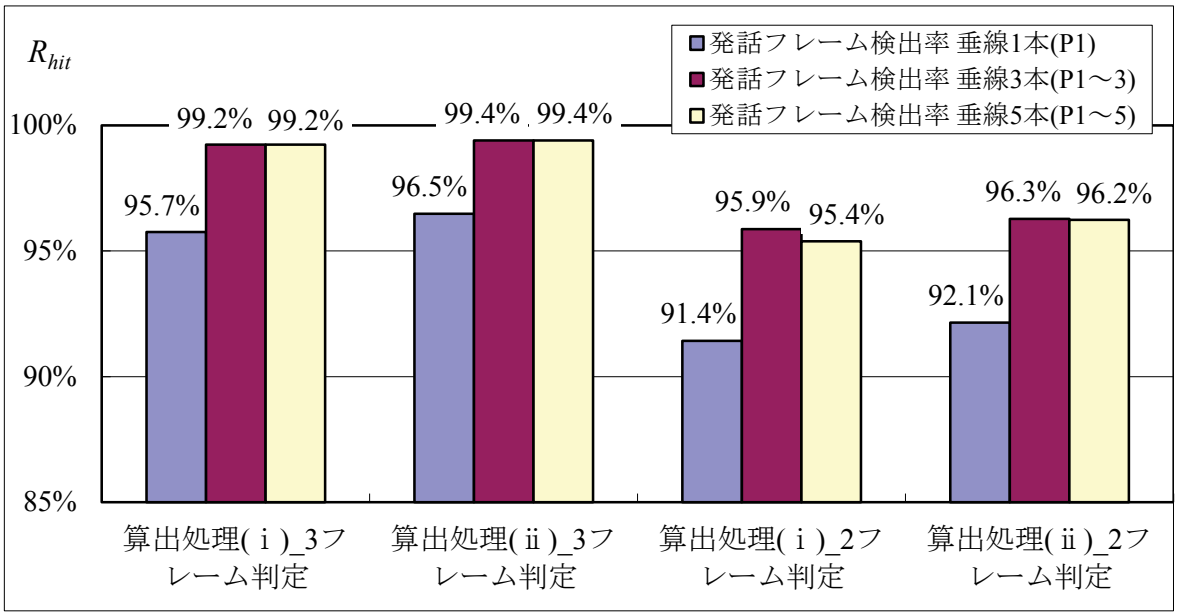


図 2.16　発話フレーム検出率 R_{hit}

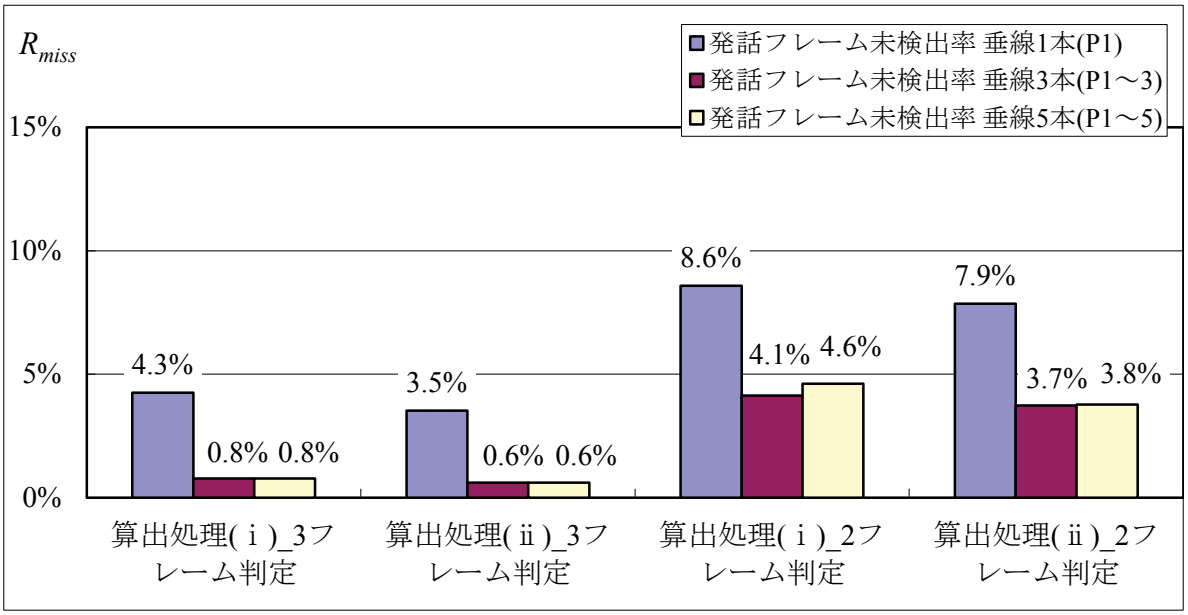


図 2.17　発話フレーム未検出率 R_{miss}

Akita University

(2)誤検出に関する考察

発話フレームに対する誤検出率を図 2.18 に示す．口唇の横幅 *diX* を考慮した形状変化特徴量算出処理（算出処理(ⅱ)：特徴量 $A'_{1\sim3}$）を用いた場合に誤検出は増加することがわかる．すなわち，縦幅 *diY* を特徴量とした算出処理（算出処理(ⅰ)：特徴量 $A_{1\sim3}$）と比較し，約 2〜3 倍程度誤検出フレームが増加している．このことは，横幅 *diX* を加味して算出された特徴量にはノイズを含んでいる可能性があり，誤検出の一因となり得ることを示唆している．

以上の結果から，口裂判定垂線数の条件では，垂線数 5 本の手法と垂線数 3 本の手法が検出率，未検出率，誤検出率のすべての評価指標において良好であり，その精度は同等であることが示された．しかしながら，垂線数が少ないほど，判定処理における計算量は低減するため，判定処理の計算量軽減の見地から垂線数 3 本の手法がより優れていると考える．

次に，特徴量の算出処理では，算出処理(ⅱ)と比較し，算出処理(ⅰ)の誤検出率は小さいことが示された．したがって，提案する口裂判定画素の位置を垂線 P_1〜P_3 の 3 本，口唇形状変化の判定を算出処理(ⅰ)による 3 フレーム間差分とする手法が高検出率かつ誤検出の少ない結果となり，比較した 12 通りの条件の中で最も発話フレーム検出に有用であることが明らかになった．

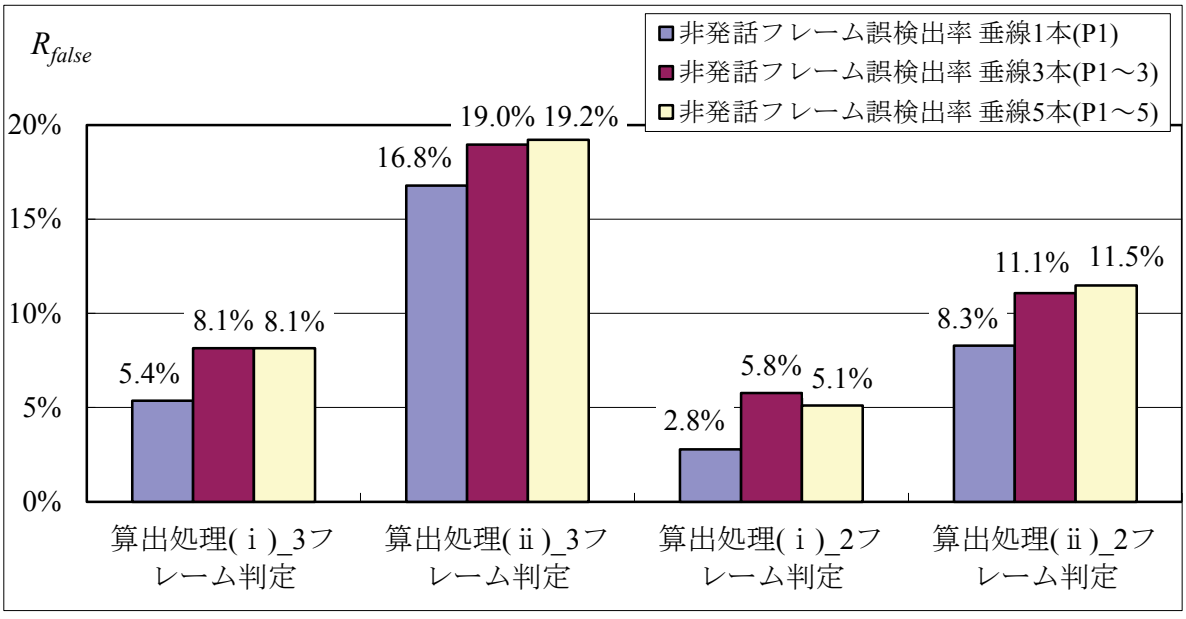


図 2.18　非発話フレーム誤検出率 R_{false}

Akita University

(3)各発話内容における発話フレーム検出に関する考察

発話フレーム検出に用いたデータセット B（2.2.3 項参照）の発話内容である Word1 から Word3 の 3 語それぞれの検出率を図 2.19～図 2.21 に示す．Word1「オオダテケンシロウ」の検出率は約 95.3%，Word2「カゲヤマヨウイチ」の検出率は約 97.8%，Word3「アキタウメコ」の検出率は約 95.8%となり，発話内容ごとの検出率に顕著な違いは見られなかった．次に，各発話内容における未検出フレーム数の算出を行ったところ，1 単語あたりの平均的な未検出フレームは約 0.9～約 1.9 フレームであった．また，口裂判定垂線 3 本，3 フレーム間差分による検出では 98.5%～100%の検出率が得られ，1 単語あたりの平均的な未検出フレーム数は 0～約 0.6 フレームとなった．したがって，Word1 から Word3 いずれの発話内容についても極めて高い検出結果が得られたことがわかる．

次に，未検出フレームを調べたところ，①母音“ウ”，“オ”の部位において検出率が低下しやすい傾向，②特定の母音の繋がりにおいて検出率が低下しやすい傾向を認めた．この時の口唇画像を目視により確認したところ，口を窄める状態または薄く開いた状態であることを認めている．なお，これらの検出失敗事例については 2.6.2 項で詳細に考察する．

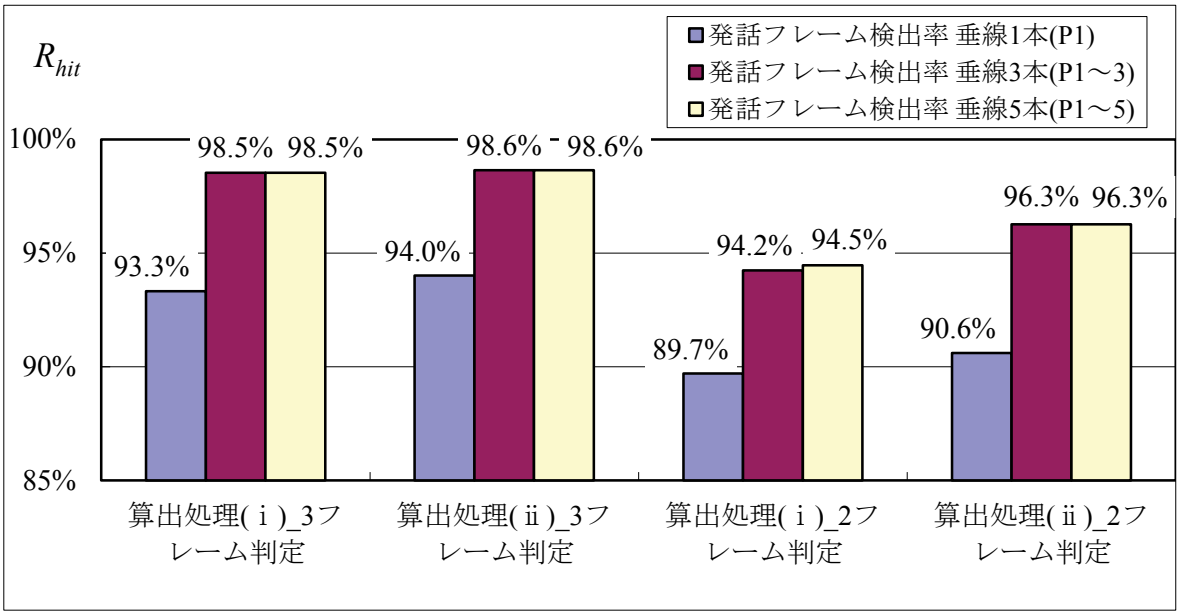


図 2.19　発話内容ごとの検出率
（Word1：オオダテケンシロウ）

Akita University

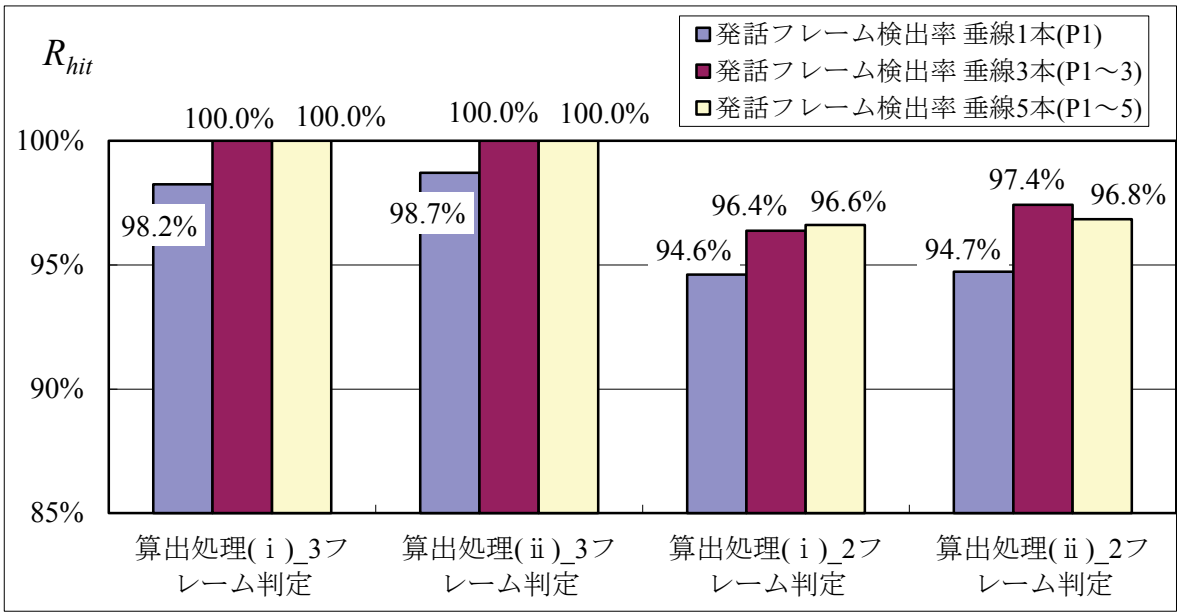


図 2.20　発話内容ごとの検出率
（Word2：カゲヤマヨウイチ）

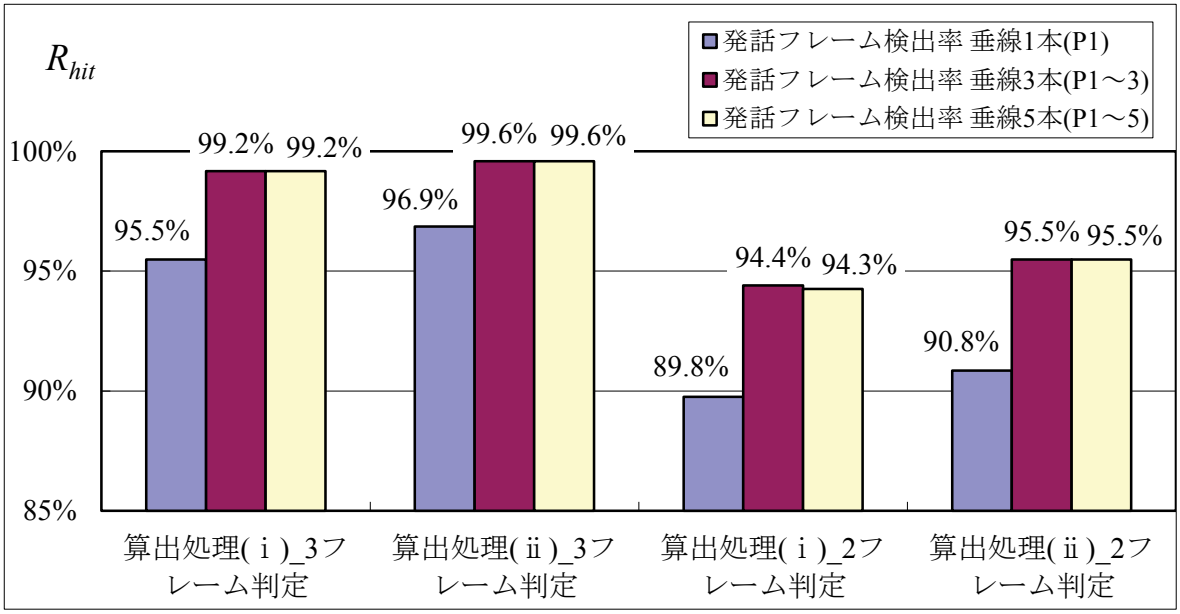


図 2.21　発話内容ごとの検出率
（Word3：アキタウメコ）

Akita University

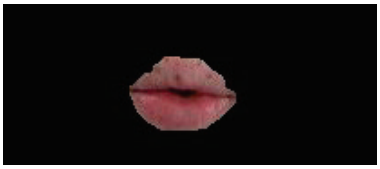
2.6.2 検出失敗事例に関する考察

(1)未検出事例に関する考察

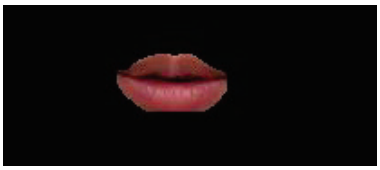
発話フレームの未検出例を図 2.22 に示す．図 2.22(a)は“ウ”の発音時における口唇の状態を示したものである．この画像は目視で「口裂なし」と判定されるが，自動検出を行った結果では「口裂あり」と判定された．また，図 2.22(b)は“オ”の発音時における口唇の状態である．この場合も(a)と同様に目視では「口裂なし」，自動検出では「口裂あり」と判定されている．すなわち，口裂の誤判定が未検出の一因であると考えられる．このため，図 2.22 のような発音状態の口唇画像について，L*および a*に着目した特徴解析を行い，口裂候補画素判定に用いるしきい値に関してさらに検討する必要がある．

(2)誤検出事例に関する考察

発話フレームの検出の各処理過程におけるデータ調査の結果，誤検出の主な要因は①口裂の誤判定, ②形状変化の誤判定であった. 要因①の例を図 2.23 に示す．この口唇画像は，目視では「口裂あり」と判定されたものである．口裂判定垂線を 1 本として口裂有無を判定した場合には誤判定は生じず，3 本以上としたときに誤判定が生じる事例を多数認めた．この事例では，口裂判定垂線 3 本のうち 1 本（5 本の場合は 1～2 本）において，口裂候補画素が検出できなかった場合に誤判定が生じている．したがって，誤検出を減少させるためには，複数の垂線上で口裂判定を行う場合の最適な判定方法について更なる検討が必要である．



(a) 未検出事例 1（被験者：Ex1-id028）



(b) 未検出事例 2（被験者：Ex1-id026）

図 2.22 未検出事例

Akita University

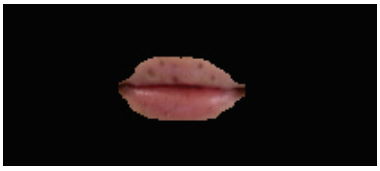


図 2.23 誤検出事例（被験者：Ex1-id028）

次に，要因②の形状変化による誤判定の場合，発話区間終了直後の数フレームで発生するケースの多いことが明らかとなった．この場合には，発生位置の特定が可能であるため，誤検出されたフレームをノイズとして除去する処理を施すことで誤判定の減少が可能と考える．

2.7 まとめ

本章では，個人識別や発話認識のインタフェース実現のための発話区間自動推定を目的とし，口唇形状自動抽出法により得られた口唇の色彩情報（L*a*b*値）および形状情報（口裂の有無）に着目した発話フレーム検出に関する検討を行った．さらに，母音の影響を考慮した発話内容を設定し，被験者 5 名の発話データについて発話フレームの自動検出を行った．得られた成果を以下にまとめる．

(1) 口を開けた状態，閉じた状態における色彩情報解析を行った結果，L*および a*推移が開閉状態を表す有用な特徴であることを明らかにした．

(2) 色彩情報（L*および a*）に着目した口裂有無の判定処理およびフレーム間差分による口唇形状変化の判定処理を用いた発話フレーム検出法は，発話フレームを高精度（約 91.4%～約 99.4%）に検出可能であることが明らかとなった．

(3) 上唇結節周辺の 3 本の口裂判定垂線（P_1～P_3）による口裂判定，算出処理(i)による 3 フレーム間の形状変化判定処理を用いた発話フレーム検出法が高検出率かつ誤差の少ない手法であることを明らかにした．

Akita University

第2章　文献

(1)佐藤，景山，西田：「口唇の動き特徴を用いた非接触コマンド入力インタフェースの提案」，電学論 C, Vol.129, No.10, pp.1865-1873 (2009)

(2)Y. Huang, H. Dohi, M. Ishizuka: “Man-Machine Interaction Using a Vision System with Dual Viewing Angles”, IEICE Trans. INF. & SYST., Vol.E80-D, No.11, pp.1074-1083 (1997)

(3)齋藤, 小西:「トラジェクトリ特徴量に基づく単語読唇」, 信学論, Vol.J90-D, No.4, pp.1105-1114 (2007)

(4)中西，寺林，梅田：「インテリジェントルームのための DP マッチングを用いた口唇動作認識」，電学論 C, Vol.129, No.5, pp.940-946 (2009)

(5)二宮，坂，前野，根木，宮島，森，北坂，末永：「音声と画像の統合によるドライバの発話区間検出」，映情学誌, Vol.62, No.3, pp.435-441 (2008)

(6)山口，浜田：「音声と口唇縦線画像を融合した発話区間検出法」，信学技報, HIP2007-161, pp.13-18 (2007)

(7)日本色彩学会編：「新編　色彩科学ハンドブック（第 3 版)」,東京大学出版会 (2011)

(8)白澤，三浦，西田，景山，栗栖：「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,Vol.60,No.12,pp.1964-1970 (2006)

(9)黒田，渡辺：「HSV 表現法に基づく顔画像の唇抽出法」，日本機械学会論文集 C 編, Vol.61, No.592, pp.150-155 (1995)

Akita University

第 3 章　口唇局所領域の形状解析に基づいた顔画像のグループ化手法

3.1 はじめに

口唇の有する「行動的特徴」に着目した従来研究として，発話に伴う口唇形状の時系列変化や口形パターンに着目した個人認証・識別，コマンド識別・読唇などが挙げられる[(1)－(5)]．これらのシステムは音声を伴わずに利用可能であるため，静寂を要する環境や雑音環境下など，様々な場面で利用可能である．しかしながら，口唇形状の時系列変化などは，発話慣れや体調・心情変化などの影響を受けるため[(6)]，得られるデータにはあいまいさが存在する．また，システム利用者の増加に伴い，(1)口唇の動き特徴の類似する利用者も増加すること，(2)登録情報の類似するケースが増加すること，ならびに(3)登録データとの照合処理の負荷が増大することが予想される．このため，コマンド識別・発話認識の精度向上に寄与するアルゴリズムの開発は重要な課題である．

一方，口唇は「形状」や「色」といった身体的特徴を有し，独特の形状を有するいくつかの局所部位（図 1.1 参照）により各個人の口唇が形成されている[(7)]．したがって，これらの局所形状の特徴に基づいて，利用者を幾つかの形状グループに分類できれば，認証・認識を行う上で対象データの絞り込みが可能になるため，口唇の動き特徴を利用したコマンド識別・発話認識システムの信頼性向上に寄与すると考える．口唇形状に着目した利用者のグループ化処理を応用したシステムの例を図 3.1 に示す．想定するシステムでは，データベース上の登録者情報を口唇形状グループごとにまとめた形で保管し，口唇形状の分類処理結果に基づき，形状類似度の高いグループから順次，照合処理を行う．すなわち，口唇領域の抽出処理後からコマンド識別・発話認識のデータ照合処理の前までに，口唇形状に着目した照合対象の絞り込みを行うことで，口唇の動き特徴が類似するデータを照合する際の優先順位を設定することができるため，認識精度の向上および照合処理量の軽減が可能になると考える．

身体的特徴としての口唇形状に着目した従来研究としては，Travieso 氏らによる顔と口唇のバイモーダル認証[(8)]，Sforza 氏らによる口唇形状の加齢変化や性別による差異に関する研究[(9)]などが行われている．Travieso 氏らの研究は顔と口唇のバイモーダルによる認証であり，口唇の局所形状に着目した特徴量抽出は行っていない．また，Sforza 氏らの研究は口唇形状に着目した認証対象者の絞り込み手法に関する検討はされておらず，形状計測には特殊な機器である電磁式 3 次元ディジタイザを必要とするなどの課題を有している．また，口唇の局所領域化では，口唇の領域分割法[(10)]が挙げられるが，口唇の局所部位に着目した形状分類は行われていない．

以上のように，一般的なビデオカメラで取得した口唇画像をコマンド識別・発話認識へ応用するという観点から，口唇局所部位の形状解析，ならびに分類によ

Akita University

る対象者絞り込みを行う手法，特に口裂（Oral fissure）形状を主眼点として行われた研究は，筆者らの調査した範囲では見当たらない．

そこで本章では，発話に伴う口唇の動き特徴を利用したコマンド認識，発話内容認識技術の認識精度向上を目的とし，口唇の局所部位における形状解析および解析結果を用いた識別対象者（ユーザ）のグループ化に関する検討を行った．

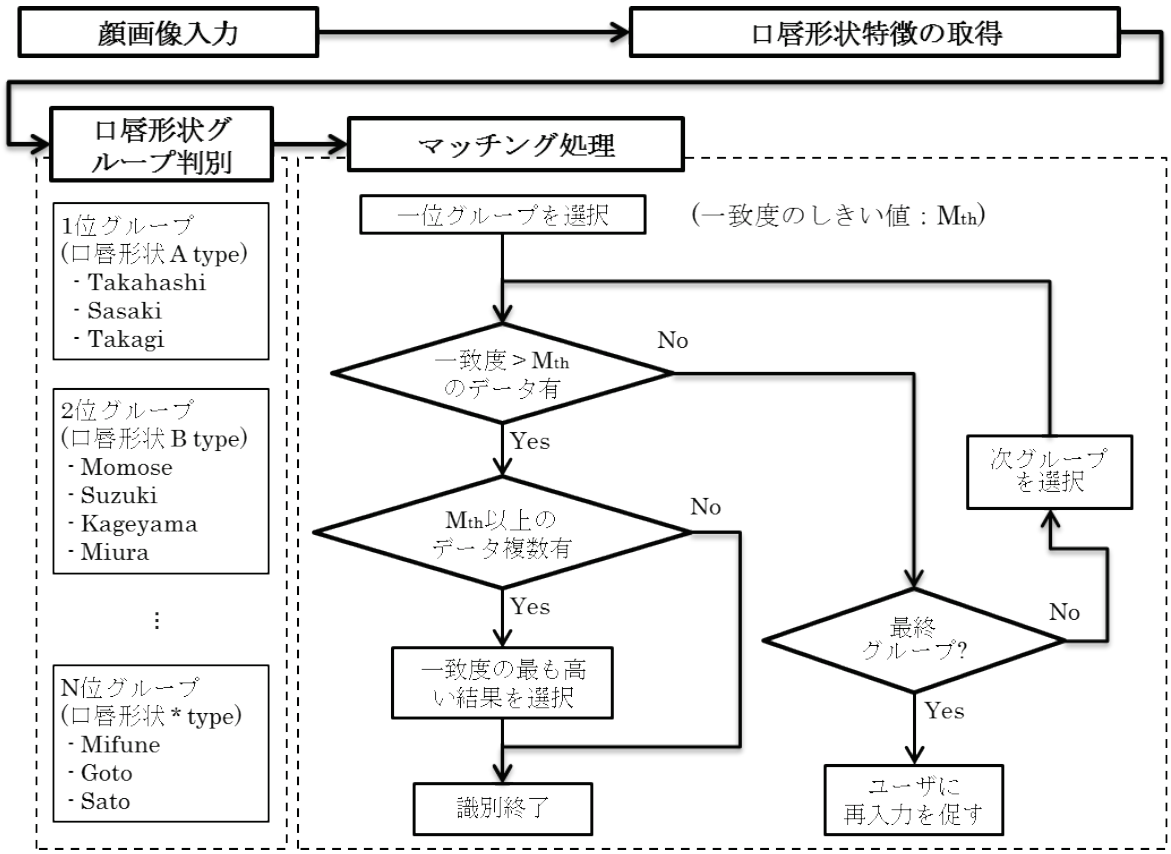


図 3.1　グループ化を応用するシステムの処理イメージ

Akita University

3.2 使用データ

3.2.1 データ取得環境

口唇形状は発話に伴って時系列的に変化するため, 口唇形状に基づいた分類には, 非発話時の閉口状態が適すると考える．そこで本研究では，3CCD ビデオカメラ（SONY 製 : DCR-VX2100）を用い, 閉口状態の被験者の顔正面を約 5 秒間撮影し，解像度 320×240 画素の顔動画像を取得した．なお，撮影時に発話動作は行っていない.

撮影時の照明などのデータ取得環境は 2 章と共通であり（2.2.1 項参照），本章に限定したデータ取得条件を以下にまとめる.

・被験者：20 代から 40 代の 106 名（Ex2-id001～Ex2-id106）.

・撮影回数 : Ex2-id001～Ex2-id052 は 2 回（1 週間以上の間隔をおいた後，2 度目のデータを取得），Ex2-id053～Ex2-id106 は 1 回のみ（解析にのみ使用）

・状態：上下の歯を噛み合わせ，上唇と下唇が軽く接触する状態.

ビデオカメラにより取得した動画像を 30fps の時系列顔画像に変換し，口唇形状自動抽出法 [1] を用いて時系列口唇画像を取得した.

3.2.2 解析用データ

取得した時系列口唇画像において，口唇形状が良好に抽出された画像を被験者ごとに無作為に 5 枚選定（被験者 106 名×5 枚＝530 枚）し，口唇形状解析用データセットとした．なお，Ex2-id001～Ex2-id052 については，初回の撮影で得られた画像データを使用した.

3.2.3 分類実験用データ

分類実験は 52 名の被験者（Ex2-id001～Ex2-id052）を対象に実施し，登録データ用動画像の撮影（初回撮影）から 1 週間以上の間隔を開け，分類実験データ用動画像の撮影を行った．解析用データセットと同様に，口唇形状が良好に抽出された画像を被験者ごとに 5 枚無作為に選定し，これを 1 枚ずつ振り分け，52 枚の口唇画像からなるデータセット（52 名×1 枚=52 枚）を 5 セット（データセット 1～5）作成した．なお，分類実験データにおける Ex2-id001～Ex2-id052 は，解析用データの Ex2-id001～Ex2-id052 と同一被験者である.

Akita University

3.3 口唇の形状特徴

3.3.1 形状特徴と局所領域分割

口唇領域は局所部位（図 1.1 参照）ごとに特有な形状を有している[7]．これら局所部位の形状は個人ごとに相違が認められ，特に上唇と下唇の厚さ，ならびに口裂が描く曲線の形状は相違が顕著であり，目視でも形状の違いを認識可能である．また，口唇の縦幅と横幅の比率も個人ごとの相違が大きく，上記部位と同様に目視でも認識可能である．そこで本研究では，「口唇の縦幅と横幅の比率（以後，アスペクト比と呼ぶ)」，「上唇・下唇の厚さ」，「口裂形状」に着目し，その形状特徴の抽出を試みた．

着目した形状の特徴量を抽出するためには，各着目部位のサイズを計測する必要がある．着目部位の形状特徴の中で，アスペクト比は口唇の縦幅および横幅から算出される値であり，特徴量取得のための計測点は一意に決定する．しかしながら，上唇・下唇の厚さ（縦幅），ならびに口裂形状は測定箇所に依存する．例えば，口角付近で計測した口唇の厚さとキューピッド弓付近で計測した口唇の厚さは大きく異なる．そこで，口唇を 3 つの矩形領域に分割し，それを局所領域 A～C として各部位の計測点を決定した．図 3.2 に示すように，局所領域 A は口裂全体（口裂画素すべて）を包含する最小の矩形領域であり，局所領域 A～C の中心的な領域である. 局所領域 B は局所領域 A の上底から上唇輪郭までを包含する矩形領域, 局所領域 C は局所領域 A の下底から下唇輪郭までを包含する矩形領域である．口唇領域を上記 3 矩形領域に分割し，各矩形領域の上底および下底に局所部位形状を計測するための基準点を設け，上唇・下唇の厚さおよび口裂の形状を計測した．

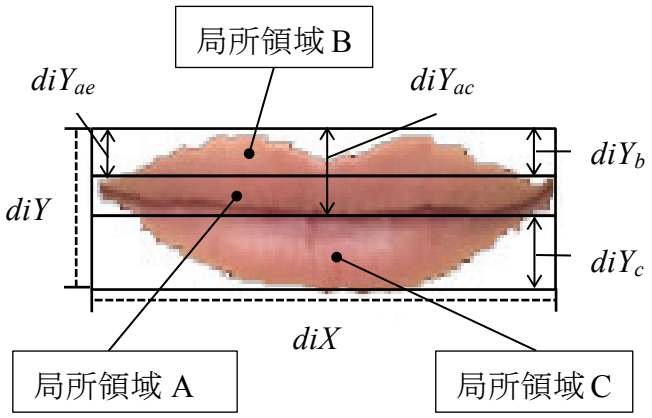


図 3.2　口唇の局所領域と計測部位

Akita University

3.3.2 口唇領域の明度情報

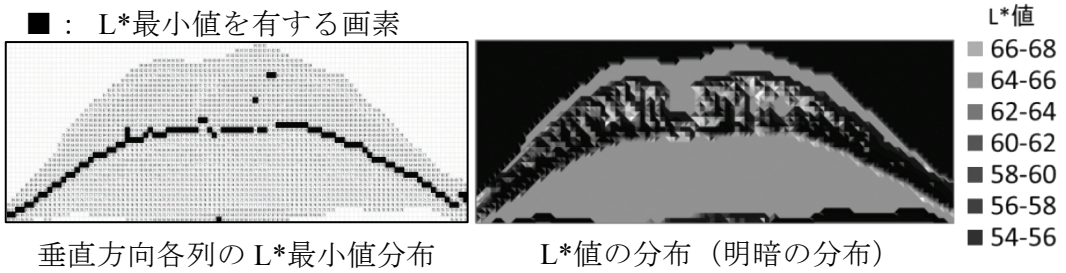
局所領域 A~C に分割するためには，領域 A の基準である口裂を口唇画像から良好に抽出する必要がある．口裂は上唇と下唇の境界線であり，裂け目の形状を有する．このため，一般の照明環境下において，口裂は周囲組織と比較し，明度の低い部位となっている．そこで，口裂周辺には陰影が生ずるものと仮定し，口唇領域における明度情報の解析を行った．具体的には，解析用データセットにメディアンフィルタ[11]を施した「平滑化画像」，ならびにフィルタ処理を施していない「口唇原画像」の両画像を対象とし，CIE-Lab[12][13]の明度値 L*の口唇領域における分布，垂直方向画素列における最小値の分布を調査した．

図 3.3(a)に口唇原画像における各画素の L*値の分布例（Ex2-id027）を示す．図 3.3(a)(b)の左図は口唇領域の各画素に基づいたメッシュ図であり，着色部分は垂直方向各列における L*の最小値を示す．また，右図は各画素の L*値に基づいて口唇領域の陰影を表した図である．なお，右図では同値の画素を同色面としており，口唇外周の暗い領域は背景部分（L*=0）である．口裂領域の明度値は被験者ごとに異なるものの，L*の最小値は口裂およびその近傍に存在することが極めて多く，口裂とその近傍以外に存在する事例はほとんど認められなかった．このことから，口裂近傍以外において L*の最小値を有する画素は，口裂の特徴とは関連しないノイズ画素と見做し，「飛び画素」と定義した．全被験者の口裂領域における L*の範囲は 96.3～40.6，垂直方向各列における L*の最小値は 82.7～40.6 の範囲であった．さらに，垂直方向各列における L*の最大値と L*の最小値の差は 53.6～0.9（平均値 23.6）であった．口角付近は全体的に明度が低く，口角から 2 列目までの領域では L*の最大 - 最小差が著しく小さくなり，その差が 2.0 以下と極めて小さい被験者の存在比率は，全被験者中約 5.8%であった．また，全体的に飛び画素数の多かった 4 名の被験者（Ex2-id027, Ex2-id028, Ex2-id033, Ex2-id034）を対象として口唇横幅に対する飛び画素の発生率を調査した結果，口唇原画像における発生率は平均 1.4%であった．

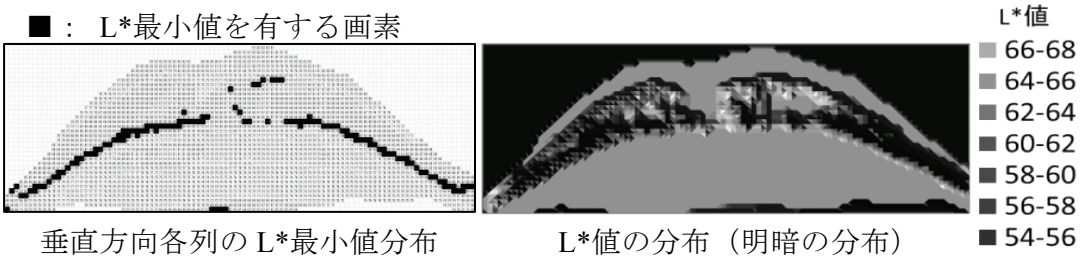
次に，図 3.3(b)にメディアンフィルタによる平滑化処理を行った場合の口唇画像における L*値の分布例を示す．口唇原画像の場合と同様に，口裂付近に明度の最小値が多く認められた．しかしながら，その出現頻度は大きく低下し，画素間の明度差が小さくなる傾向を示した．全被験者の口裂領域における明度値の範囲は 94.6～43.4 であり，口裂付近の画素は 82.5～43.4 の範囲であった．また，垂直方向各列における L*の最大値と L*の最小値の差は 37.0～0.02（平均値 18.8）であり，明暗差の小さい傾向が口唇原画像よりも顕著に認められた．特に，口唇右端の L*の最大 - 最小差が 2.0 以下である被験者の存在比率は，全被験者中約 28.8%となり，口唇原画像と比較して大幅に増加している．なお，被験者 4 名（Ex2-id027, Ex2-id028, Ex2-id033, Ex2-id034）における飛び画素の発生率は平均 2.6%となり，口唇原画像を用いた場合の約 2 倍となった．

Akita University

以上の結果より，口裂画素は周辺画素と比較して明らかに明度値が低いこと，ならびにメディアンフィルタ処理画像と比較して口唇原画像は口裂の明暗情報を明確に有することが明らかとなった．そこで，以後の検討には口唇原画像を用い，その L*値に着目して口裂の抽出を行った．なお，抽出処理の詳細は 3.5.1 項で述べる．



(a) 原画像



(b) メディアンフィルタ処理を加えた画像

図 3.3 口唇領域の L*値分布（Ex2-id027）

Akita University

3.3.3 特徴量の算出

データの取得が簡便であり，加えて形状分類に有意な特徴量と考えられることから，局所領域 A～C の上底・下底を基準とし，エッジ情報である口唇輪郭および口裂画素の座標値から特徴量算出のためのベースとなる値として *diY*, *diX*, diY_b, diY_c，diY_{ae}，diY_{ac}（図 3.2 参照）を用いた．*diY* は口唇全体の縦幅，*diX* は口唇全体の横幅の画素数であり，diY_b，diY_c は矩形領域 B および矩形領域 C の縦幅の画素数である．また，diY_{ae} は領域 B の上底から口角までの縦幅を表す画素数，diY_{ac} は領域 B の上底から口裂中央点までの縦幅を表す画素数である．なお，口角の縦座標が口裂画素中で最も高い位置にある場合，diY_{ae}＝diY_b となり，最も低い位置にある場合は diY_{ae}＝(*diY* - diY_c)となる．diY_{ac} も同様である．

上唇および下唇の厚さは口角付近が最も薄く，口唇の中間付近で最大の厚さとなる．また，上唇と下唇の境界線である口裂は，上唇結節の大きさや歯並びなどの影響を受け，被験者ごとに多様な形状となる．このため，上唇および下唇の厚さは水平方向の計測位置によって大きく異なる．そこで，局所領域 B，C の縦幅 diY_b，diY_c を用いて厚さを表す特徴量を算出するが，diY_b，diY_c は画素数であるため，被験者 - カメラ間の微小な距離変動やズームの度合いによって変化する．この影響を軽減するため，口唇の縦幅 *diY* との比（R_{by}，R_{cy}）を算出し，上唇および下唇の厚さを表す量として用いた．すなわち，R_{by}，R_{cy} を(3.1)，(3.2)式で算出し，得られた R_{by} と R_{cy} の差 R_{by} - R_{cy} を上唇と下唇の厚さを比較する特徴量とした．

$$R_{by} = \frac{diY_b}{diY} \qquad \cdots\cdots\cdots(3.1)$$

$$R_{cy} = \frac{diY_c}{diY} \qquad \cdots\cdots\cdots(3.2)$$

Akita University

次に，口角と口裂における水平方向の中間点を計測座標とし，口裂が描く弧の方向とその度合いを口裂形状特徴とした．口裂形状における特徴量算出には局所領域 B の上底から右の口角までの縦幅 diY_{ae}，局所領域 B の上底から口裂中央点までの縦幅 diY_{ac} を用いた．口裂に関しても，被験者 - カメラ間の微小な距離変動などの影響を軽減するため，diY との比 R_{ae}，R_{ac} を(3.3)，(3.4)式で算出し，得られた R_{ae} と R_{ac} の差 R_{ae}- R_{ac} を口裂の凹凸形状を表す特徴量とした．

$$R_{ae} = \frac{diY_{ae}}{diY} \qquad \text{…….(3.3)}$$

$$R_{ac} = \frac{diY_{ac}}{diY} \qquad \text{…….(3.4)}$$

アスペクト比は口唇全体の縦幅 diY と横幅 diX を用い, (3.5)式により算出した．特徴量 R_{xy} は口唇の縦横比率を示し，対象の口唇が横方向と縦方向のどちらに長いかを示す量を意味する．

$$R_{xy} = \frac{diY}{diX} \qquad \text{…….(3.5)}$$

Akita University

3.4 特徴量による形状分布解析と分類カテゴリ

3.4.1 形状特徴の分布

解析用データ 530 枚（106 名×5 枚）を用い，口唇の厚さに関する特徴量 R_{by}-R_{cy}，口裂形状に関する特徴量 R_{ae}-R_{ac}，アスペクト比に関する特徴量 R_{xy} の分布を調査した．なお，分布の作成にあたり，画像の分解能を考慮し，4 通りの階級間隔（0.050，0.020，0.010，0.005）で検討を加えた．その結果，階級間隔を 0.010 とした場合に各特徴量の傾向を最も良好に表したため，階級間隔 0.010 で分布図を作成し，以後の検討を行った．

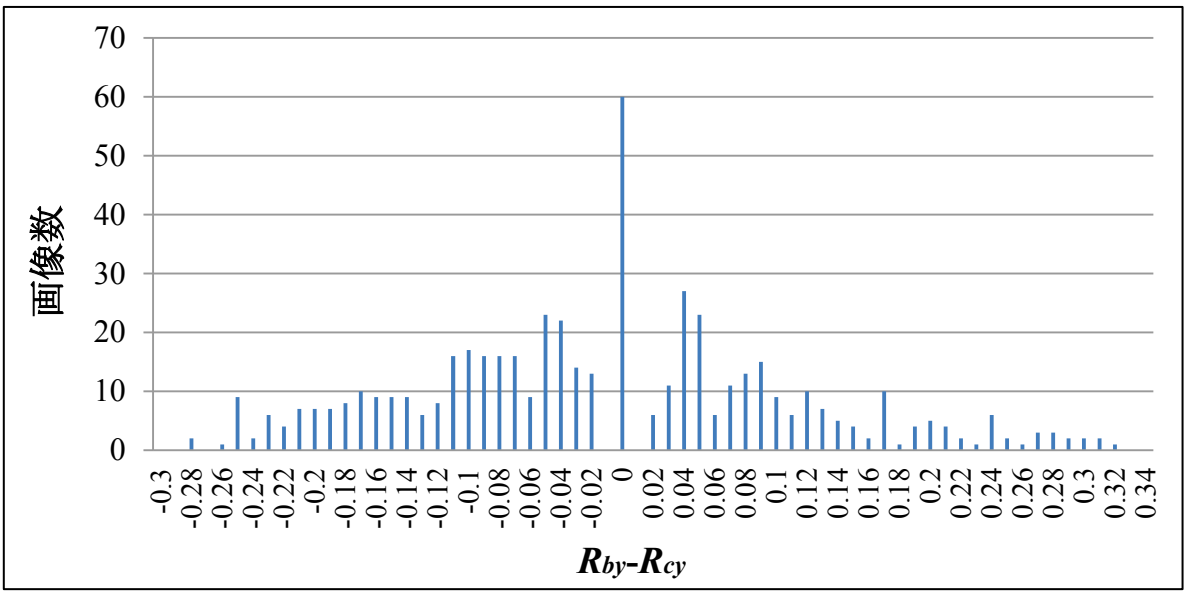
(1) 口唇の厚さに関する解析

図 3.4(a)は口唇の厚さ特徴量 R_{by}-R_{cy} における口唇画像の頻度分布である．分布より，R_{by}-R_{cy}=0 の頻度が突出していることがわかる．また，R_{by}-R_{cy}=0 を挟んで低値側と高値側に度数の集中する領域が存在し，三峰性を有する分布を形成していることがわかる．なお，口唇画像データの分解能に起因し，-0.017～0 および 0～0.019 の範囲における頻度は 0 となっている．具体的には，解析用データセット全 530 枚において *diY* 値は 66 ピクセル以下であるため，$-0.0152<R_{by}-R_{cy}<0$, $0<R_{by}-R_{cy}<0.0152$ の範囲の値を有するデータは存在しない．このため，単峰性ではなく三峰性の形状を呈する分布となっている．分布において R_{by}-R_{cy}=0 が全体の最頻値であり，全体の 13.2%（530 枚中 70 枚）を占めている．また，全体の平均値が-0.013 であることから，口唇の平均的な形状は上下唇の厚さが同等程度であることが推測される．

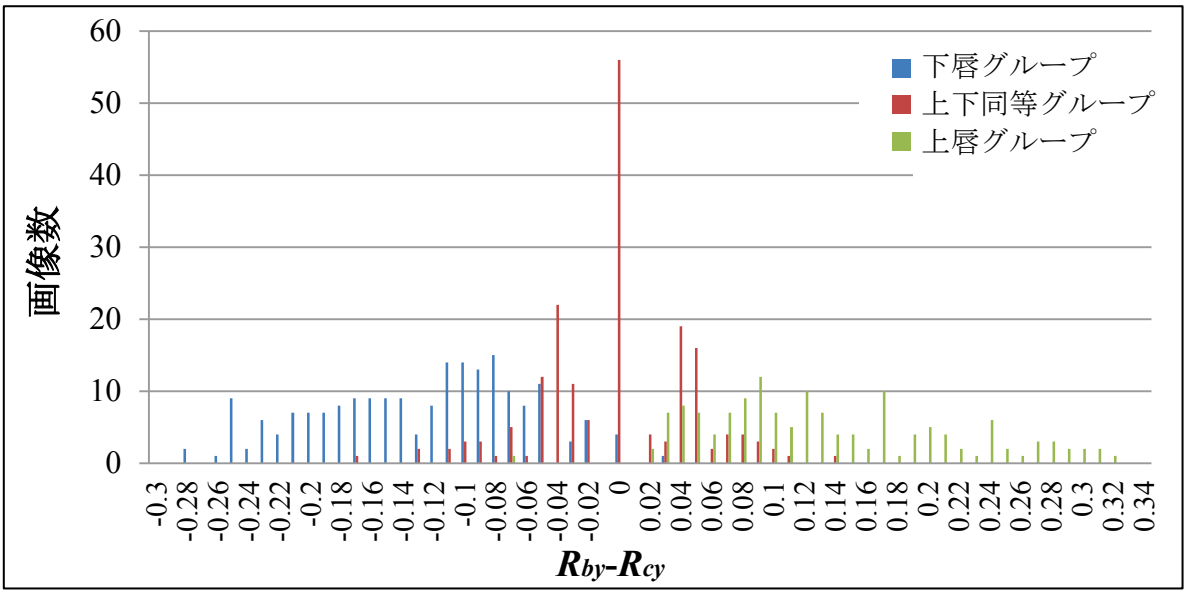
したがって，本研究のデータ取得条件下では，口唇の厚さ形状を，下唇が厚い（R_{by}-R_{cy}<0），上下同等の厚さ（R_{by}-R_{cy}=0），上唇が厚い（R_{by}-R_{cy}>0）の 3 つの形状グループに大別可能であることを示している．図 3.5 に各形状グループの代表的な例を示す．

次に，口唇の厚さ 3 形状それぞれの分布を求めた．解析データは被験者あたり 5 枚の口唇画像を有するため，それぞれの画像間で特徴量にばらつきが存在する．そこで，分布のモード値である R_{by}-R_{cy}=0 の画像を有する被験者を対象に，被験者ごとの平均値を算出した．得られた平均値群の範囲は-0.060～0.062 であったため，R_{by}-R_{cy} の平均値が-0.060 未満である者は「下唇が厚い被験者群（下唇グループ）」，-0.060 から 0.062 の者は「上下同等の厚さである被験者群（上下同等グループ）」，0.062 を超える者は「上唇が厚い被験者群（上唇グループ）」に属すると仮定し，106 名をそれぞれの被験者群に振り分け，その特徴量分布を求めた．その結果（各形状の被験者群における R_{by}-R_{cy} 分布）を図 3.4(b)に示す．また，各群における統計量（平均値，標準偏差，モード値，最大値，最小値）を表 3.1 にまとめる．なお，106 名の振り分け結果は，上唇が厚い被験者群 29 名（割合 27.4%），上下同等の被験者群 37 名（割合 34.9%），下唇が厚い被験者群 40 名（割合 37.7%）となった．

Akita University



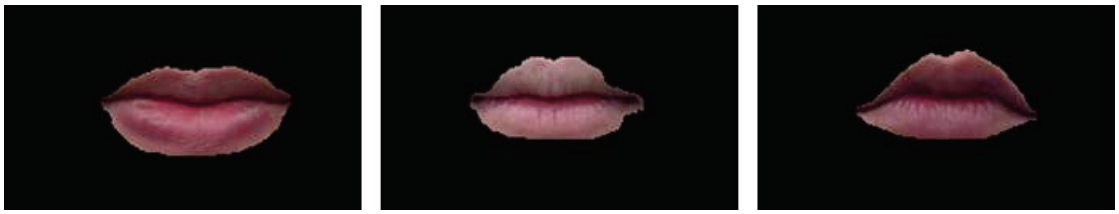
(a) 全被験者



(b) 3 グループに分割

図 3.4 上唇・下唇の厚さ特徴の分布（R_{by}-R_{cy}）

Akita University



(a)下唇グループ　　(b)上下同等グループ　　(c)上唇グループ

図 3.5　代表的な形状（口唇の厚さ）

表 3.1 上唇・下唇の厚さ特徴における統計量

	下唇グループ	上下同等グループ	上唇グループ
平均値	-0.128	-0.002	0.134
標準偏差	0.066	0.050	0.079
モード値	-0.250	0	0.125
最大値	0.032	0.143	0.321
最小値	-0.282	-0.174	-0.073

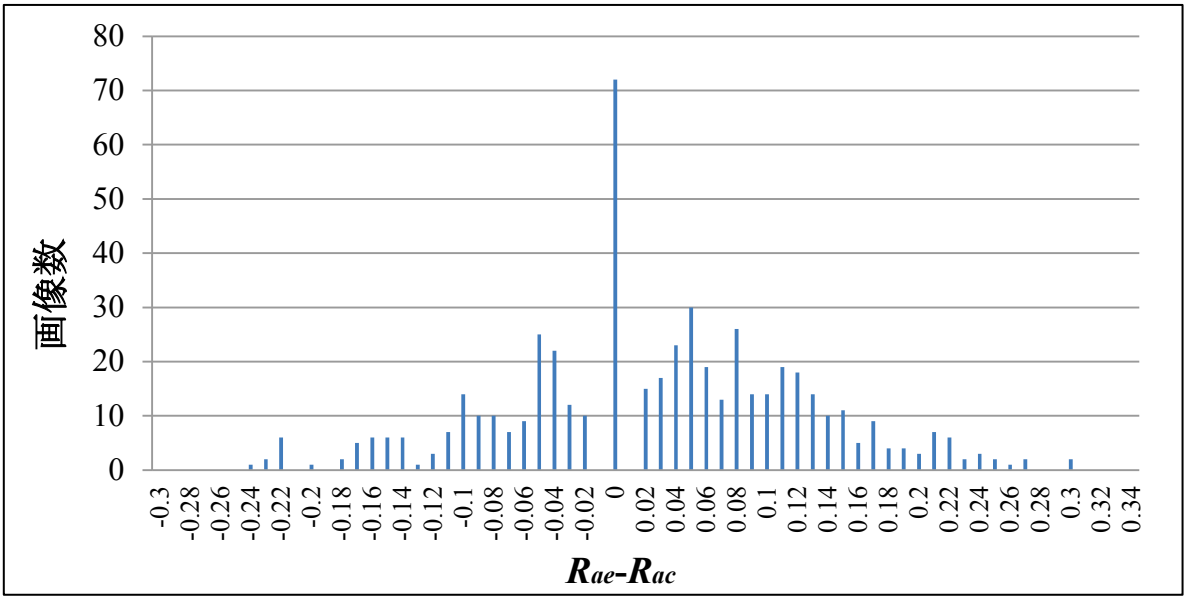
Akita University

(2)口裂形状に関する解析

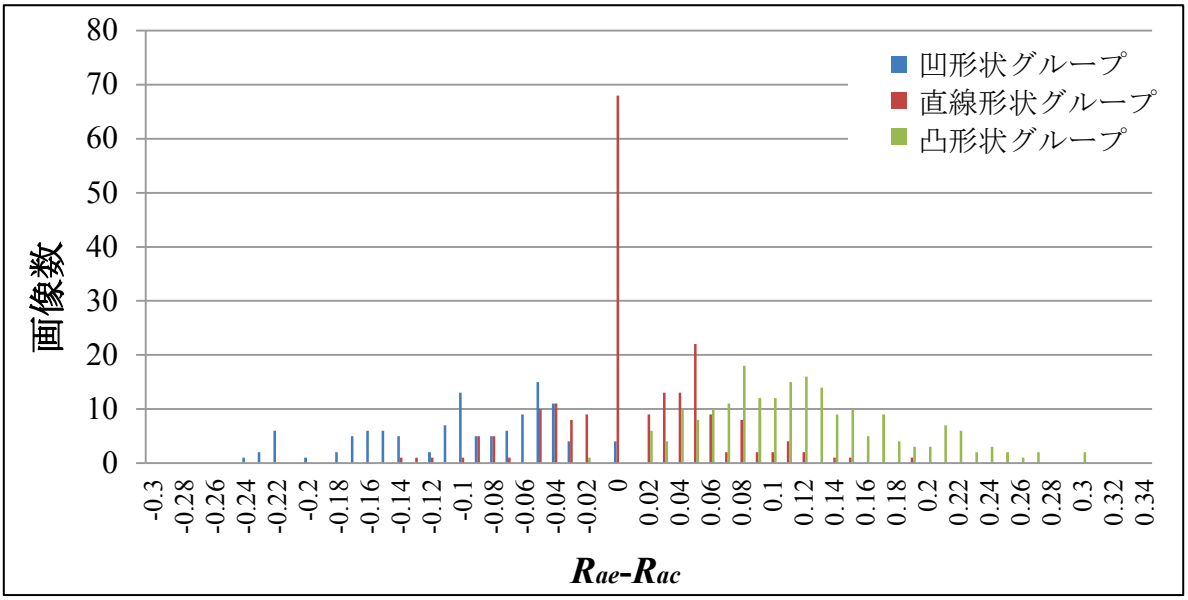
口裂形状特徴量 R_{ae}-R_{ac} における口唇の頻度分布を図 3.6(a)に示す．R_{ae}-R_{ac} 分布は，R_{ae}-R_{ac}=0 に該当する口唇の標本数が突出するなど，R_{by}-R_{cy} 分布と類似した様相を呈しており，R_{ae}-R_{ac}=0 を挟んで低値側および高値側に標本数の多い領域が形成されている．なお，R_{ae}-R_{ac}=0 近傍の頻度が 0 となる理由は，R_{by}-R_{cy} の場合と同様である．分布において R_{ae}-R_{ac}=0 が全体の最頻値であり，全体の 13.6%（530 枚中 72 枚）を占めている．この結果は，口裂形状を，凹形状（R_{ae}-R_{ac}<0），直線形状（R_{ae}-R_{ac}=0），凸形状（R_{ae}-R_{ac}>0）の 3 つの形状グループに大別可能であることを示唆している．図 3.7 に各形状の代表的な例を示す．

次に，口裂形状 3 形状それぞれの分布を求めた．口裂形状においても，口唇の厚さ解析と同様に，特徴量 R_{ae}-R_{ac}=0 の画像を有する被験者を対象として被験者ごとの平均値を算出した．得られた R_{ae}-R_{ac} の平均値群の範囲-0.068～0.078 に基づき，R_{ae}-R_{ac} の平均値が-0.068 未満である者は「凹形状の被験者群（凹形状グループ）」，-0.068 から 0.078 の者は「直線形状の被験者群（直線形状グループ）」，0.078 を超える者は「凸形状の被験者群（凸形状グループ）」に属するものと仮定し，106 名をそれぞれの被験者群へ振り分けた．得られた R_{ae}-R_{ac} 分布を図 3.6(b)に示す．また，R_{by}-R_{cy} 分布と同様に各群における統計量（平均値，標準偏差，モード値，最大値，最小値）を表 3.2 にまとめる．なお，106 名の振り分け結果は，凸形状の被験者群 41 名（割合 38.7%），直線形状の被験者群 42 名（割合 39.6%），凹形状の被験者群 23 名（割合 21.7%）であった．

Akita University



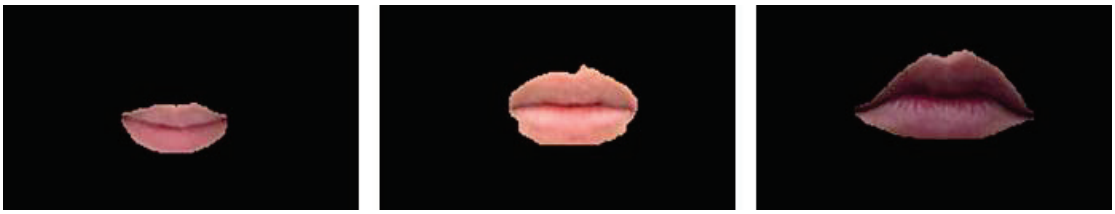
(a) 全被験者



(b) 3 グループに分割

図 3.6 口裂形状特徴の分布（R_{ae}-R_{ac}）

Akita University



(a)凹形状グループ　　(b)直線形状グループ　　(c)凸形状グループ

図 3.7　代表的な形状（口裂形状）

表 3.2 口裂形状特徴に関する統計量

	凹形状グループ	直線形状グループ	凸形状グループ
平均値	-0.099	0.011	0.119
標準偏差	0.059	0.051	0.061
モード値	-0.222	0	0.083
最大値	0	0.188	0.303
最小値	-0.242	-0.136	-0.024

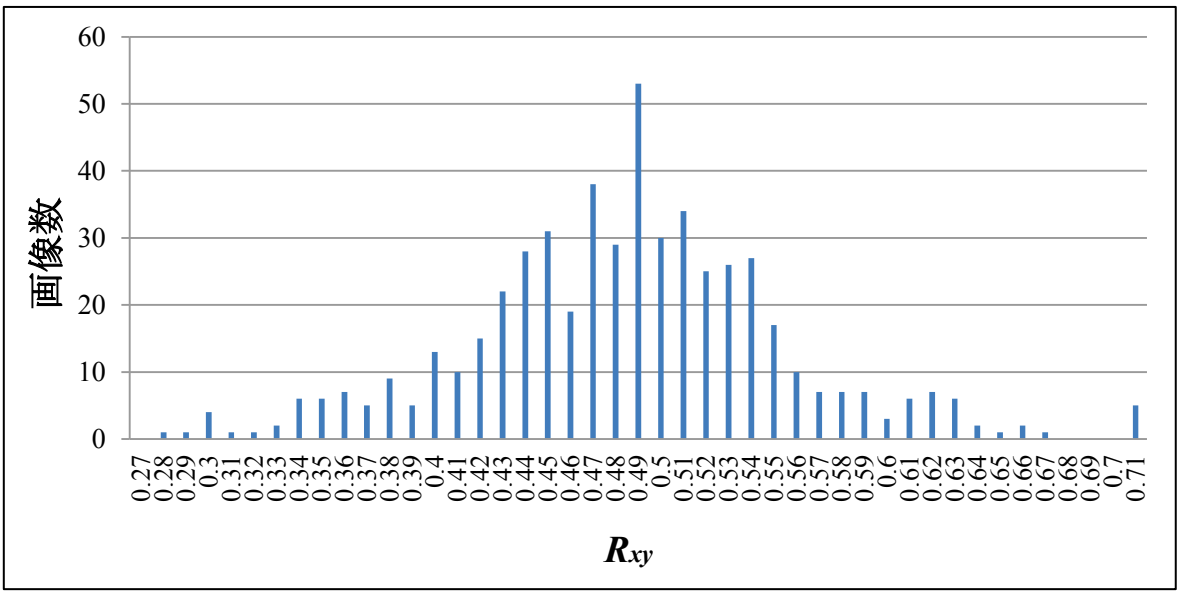
Akita University

(3)アスペクト比に関する解析

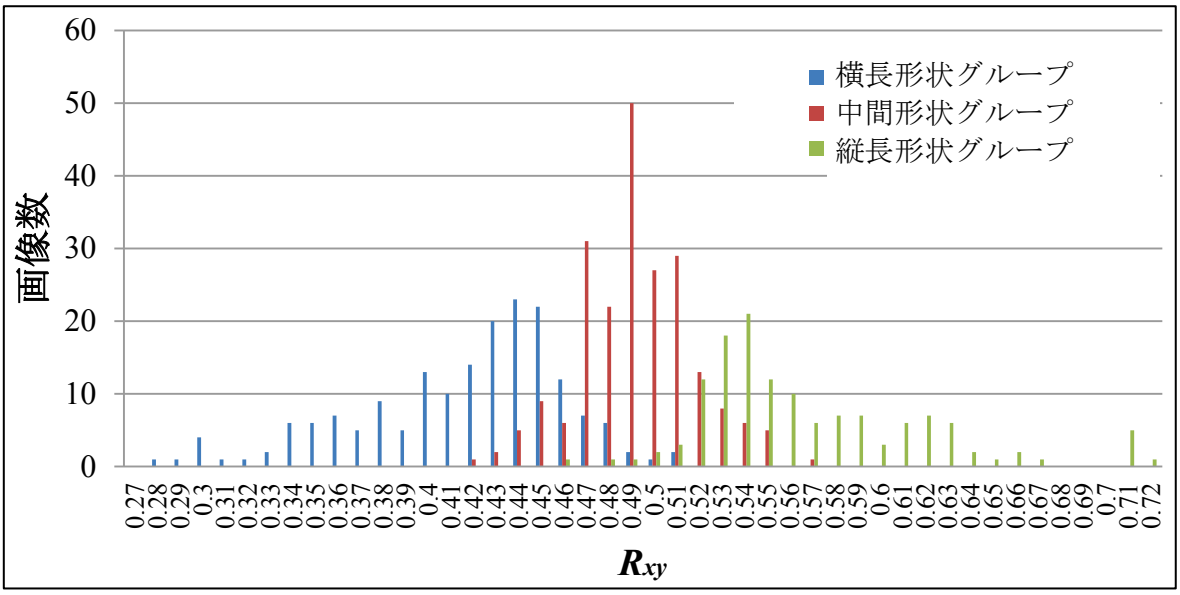
図 3.8(a)は R_{xy} の分布を示しており，標本数の最も多い R_{xy}=0.490 を中心として低値側，高値側に分かれ，全体的には正規分布の形状を有している．R_{xy} の平均値が約 0.486，モード値は 0.490 であることから，R_{xy}=0.490 は中間的なアスペクト比であり，横長形状の口唇（R_{xy}<0.490），中間的形状の口唇（R_{xy}=0.490），縦長形状の口唇（R_{xy}>0.490）の 3 つの形状グループに大別可能である．図 3.9 に各形状グループの代表例を示す.

アスペクト比における 3 形状の各分布作成も R_{by}-R_{cy} や R_{ae}-R_{ac} と同様に, モード値（R_{xy}=0.490 ± 0.005）の画像を有する被験者を対象とし，被験者ごとの平均値を算出した．得られた平均値群の範囲は 0.467～0.525 であったため，R_{xy} の平均値が 0.467 未満である者は「横長形状の口唇を有する被験者群（横長形状グループ)」，0.467～0.525 である者は「中間的形状の被験者群（中間形状グループ)」，0.525 を超える者は「縦長形状の口唇を有する被験者群（縦長形状グループ)」に属すると仮定し，106 名を各被験者群に振り分け，各群における特徴量分布を求めた．得られた特徴量分布を図 3.8(b)に示す．また，表 3.3 に各群における統計量（平均値，標準偏差，モード値，最大値，最小値）をまとめる．各被験者群への振り分け結果は, 横長の被験者群が 23 名（存在割合 21.7%), 標準の被験者群が 59 名（存在割合 55.7%)，縦長の被験者群が 24 名（存在割合 22.6%）であった.

Akita University



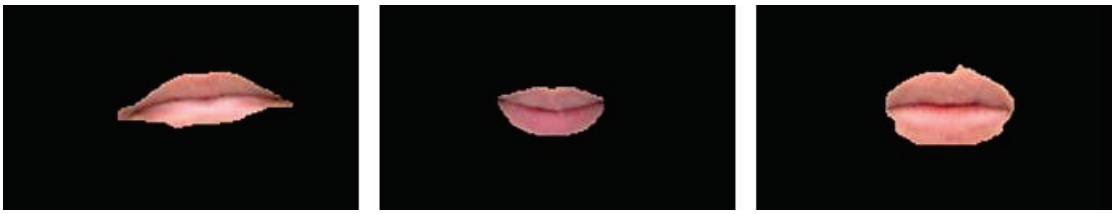
(a) 全被験者



(b) 3 グループに分割

図 3.8 アスペクト比の分布（R_{xy}）

Akita University



(a)横長形状グループ　　(b)中間形状グループ　　(c)縦長形状グループ

図 3.9　代表的な形状（アスペクト比）

表 3.3　アスペクト比に関する統計量

	横長形状グループ	中間形状グループ	縦長形状グループ
平均値	0.415	0.490	0.565
標準偏差	0.046	0.025	0.051
モード値	0.429	0.500	0.533
最大値	1.000	0.553	0.718
最小値	0.278	0.410	0.460

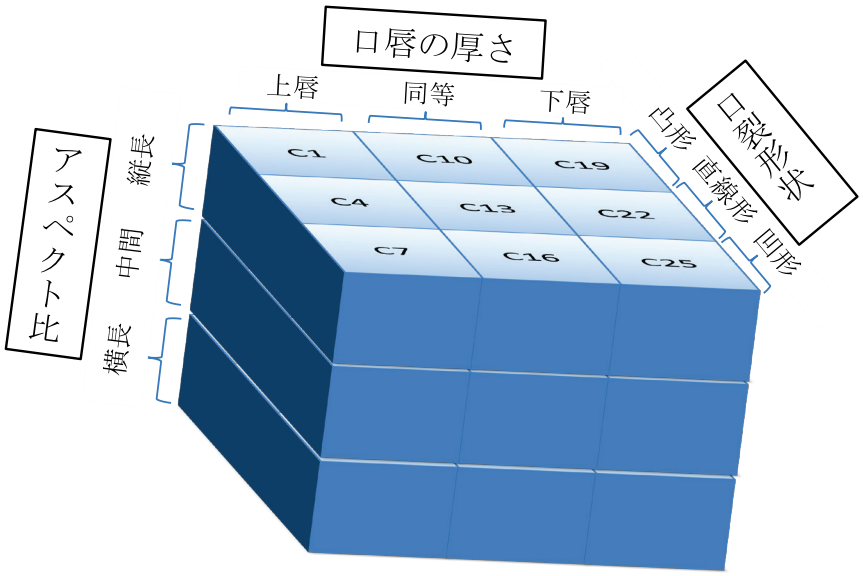
Akita University

3.4.2 形状カテゴリの定義

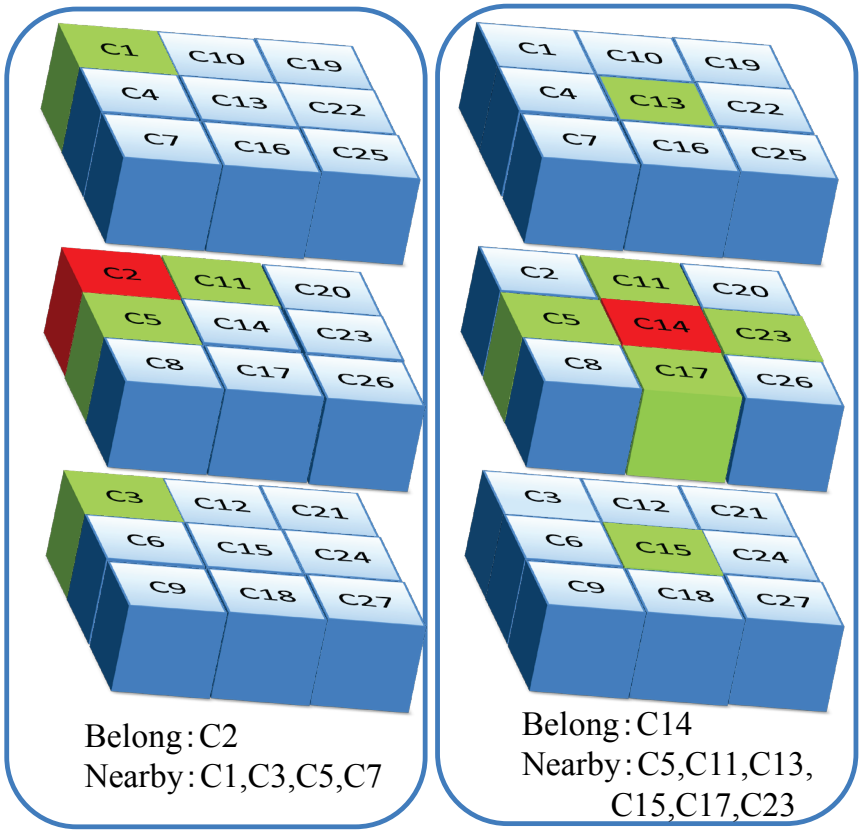
形状分布の解析から，上唇・下唇の厚さでは R_{by}-R_{cy}＝0 を形状の基準値として「上唇が厚い」，「上下同等の厚さ」，「下唇が厚い」の 3 クラス，口裂形状では R_{ae}-R_{ac}＝0 を形状の基準値として「凸形状」，「直線形状」，「凹形状」の 3 クラス，アスペクト比は R_{xy}=0.490 を形状の基準値として「縦長形状」，「中間形状」，「横長形状」の 3 クラスにそれぞれ分類可能という結果が得られた．そこで，各着目部位の形状クラス数に基づき，図 3.10(a)に示す口唇の形状カテゴリ C1～C27 を形成した．C1～C27 において，近傍に存在するカテゴリはお互いに類似した形状を有し，図 3.10(a)で示される位置関係はその類似度合いを表している．互いに面が接するカテゴリ同士は，特に類似した形状を持つ関係であるため，これを本研究では「隣接カテゴリ（Nearby Category）」と定義した．具体例を図 3.10(b)に示す．図 3.10(b)左側の場合，口唇形状の所属カテゴリ C2 と面で接している C1，C3，C5，C11 の 4 カテゴリが C2 の 隣接カテゴリである．一方，右図では，所属カテゴリ C14 と面で接している C5，C11，C13，C15，C17，C23 の 6 カテゴリが C14 の隣接カテゴリとなる．なお，所属カテゴリの位置によって隣接カテゴリの数は異なり，その範囲は 3～6 である．また，3.4.1 項における各局所形状の振り分け結果（各 3 クラス）に基づいた 106 名の所属カテゴリを表 3.4 に示す．各局所形状クラスにおいて存在割合の大きいクラス同士の組み合わせである C3，C14，C23 の 3 形状で全体の 1/3 を占める結果となった．最も割合の大きい C14 は，形状カテゴリの中心に位置する形状（上下同等の厚さ，口裂は直線形状，中間的アスペクト比）であり，多くの被験者が中間的な口唇形状を有していることがわかる．

以上の解析結果に基づいて，次節以降では 52 名の被験者を対象に口唇形状の分類を試みた．さらに，マッチングデータの絞り込みに関する有用性についても検討を加えた．

Akita University



(a) カテゴリの位置関係（類似度合）



(b) 隣接関係の例

図 3.10 口唇形状カテゴリ（27 カテゴリ）

Akita University

表 3.4　解析データ（Ex2-id001～Ex2-id106）の所属カテゴリ

カテゴリ	被験者数[名]	割合[%]
C1	5	4.7
C2	3	2.8
C3	12	11.3
C4	2	1.9
C5	2	1.9
C6	2	1.9
C7	0	0
C8	0	0
C9	0	0
C10	3	2.8
C11	8	7.6
C12	2	1.9
C13	7	6.6
C14	13	12.3
C15	6	5.7
C16	2	1.9
C17	1	0.9
C18	1	0.9
C19	0	0
C20	0	0
C21	1	0.9
C22	6	5.7
C23	11	10.4
C24	7	6.6
C25	1	0.9
C26	5	4.7
C27	6	5.7

Akita University

3.5 局所形状に着目した顔画像のグループ化法

3.5.1 口唇領域分割・特徴量算出処理

口唇分割処理では，口唇原画像の L*値に基づいて口裂を抽出し，得られた口裂および口唇輪郭の上下左右端を基準点として口唇領域を局所領域 A～C に分割する．処理の流れを以下①～⑧にまとめる．

①口唇画像を 2 値化し，口唇輪郭および口唇領域左右端の座標を取得する．

②口唇領域の L*値を取得し，垂直方向各列における L*最小値を口裂候補画素として抽出する．

③口唇輪郭および口裂候補画素に対し，アフィン変換による回転処理を施し，顔の傾きに対する補正を行う．回転の中心は口裂の中点であり，回転角は口裂候補画素の左右端（口角画素）の垂直方向座標が同一となる角度とする．

④左端から口裂候補画素を走査し，同一列上に最小値画素が複数存在した場合，それらの垂直方向座標を比較して前列の口裂候補画素の最近傍である画素を選択する．次に，再度左端から口裂候補画素を走査し，垂直方向に 2 画素以上のギャップを有する部分（不連続部分）を検出する．不連続部分前後の口裂候補画素の垂直座標を参照し，そのギャップが 3 画素以下である場合，その中間点に不連続部分の画素を移動する．一方，3 画素以上である場合には，さらに 1 画素右方向の口裂候補画素を参照して不連続点を補正し，口裂画素を得る．なお，口裂候補画素が 3 画素以上連続してギャップを生ずる場合には，その画像データを棄却する．

⑤口裂画素を包含する最小の矩形を領域 A とする（図 3.11 参照）．

⑥領域 A の上底を一辺とし，上唇輪郭を包含する最小の矩形領域を領域 B と設定する．（図 3.2 参照）．

⑦領域 A の下底を一辺とし，下唇唇輪郭を包含する最小の矩形領域を領域 C と設定する．（図 3.2 参照）．

⑧特徴量のベース値を取得し，(3.1)～(3.5)式を用いて特徴量 R_{xy}，R_{by}-R_{cy}，R_{ae}-R_{ac} を算出する．

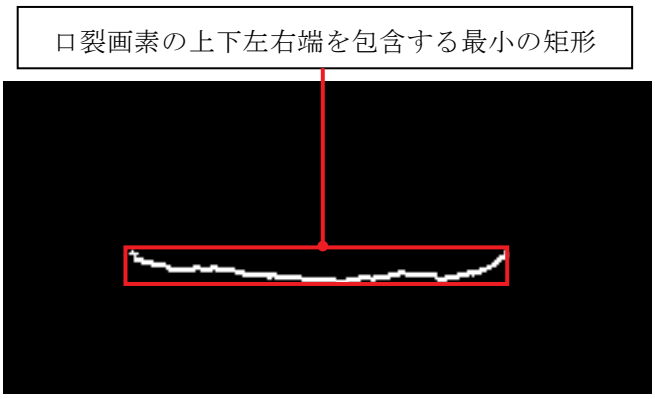


図 3.11 口裂抽出結果および領域 A の設定

Akita University

3.5.2 局所形状クラスの判定処理

(1) メンバーシップ関数の設定

カメラで取得された口唇画像には, 顔の向きの微小変動やそれに伴う口唇部位への照度変動の影響が存在するため, 同一被験者であっても得られる特徴量は不定である. そこで本研究では, 口唇画像データから得られる各局所形状の特徴量はファジィ集合を形成すると仮定し, その分類を行うために, 図 3.12 に示す三角型のメンバーシップ関数 A_{i1}, A_{i2}, A_{i3} を設定した（i=1～3, 1:上下唇の厚さ R_{by}-R_{cy}, 2:口裂形状 R_{ae}-R_{ac}, 3:アスペクト比 R_{xy}）. なお, 各入力値（R_{by}-R_{cy}, R_{ae}-R_{ac}, R_{xy}）に対して出力された帰属度が最大である形状クラスを判定結果とした.

メンバーシップ関数の設定値は 3.4.1 項の分布解析結果（図 3.4, 図 3.6, 図 3.8 参照), ならびに統計量（表 3.1～3.3 参照）に基づき, 以下のように設定した.

μ_{11}：上唇が厚い群（上唇グループ）の平均値
μ_{12}：上下同等群（上下同等グループ）のモード値
μ_{13}：下唇が厚い群（下唇グループ）の平均値
μ_{21}：口裂凸形状群（凸形状グループ）の平均値
μ_{22}：口裂直線形状群（直線形状グループ）のモード値
μ_{23}：口裂凹形状群（凹形状グループ）の平均値
μ_{31}：アスペクト比縦長群（縦長形状グループ）の平均値
μ_{32}：アスペクト比中間群（中間形状グループ）の平均値
μ_{33}：アスペクト比横長群（横長形状グループ）の平均値

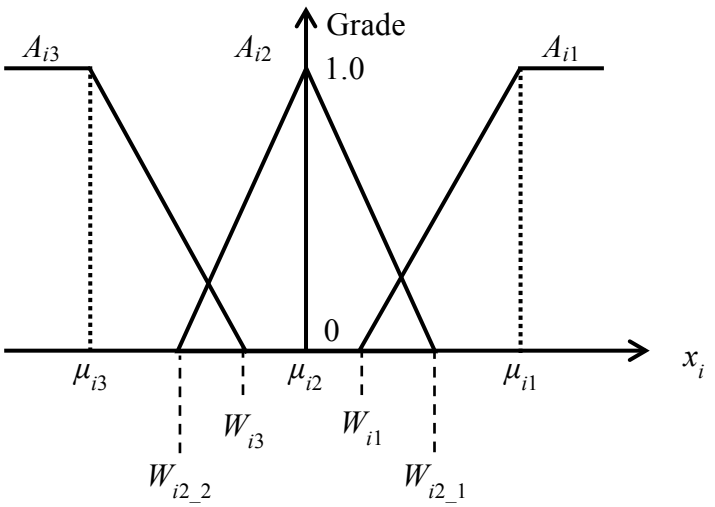


図 3.12 三角型メンバーシップ関数

Akita University

各群の平均値付近では他群の存在割合が極めて低いことから，上記数値に各群の平均値を用いた．一方，3.4.1 項で述べたように，画像解像度の影響のため，上下同等群および口裂直線形状群は，モード値（R_{by}-R_{cy}＝0 および R_{ae}-R_{ac}=0）の度数が突出し，その近傍の度数は極めて低い分布となった．そこで，μ_{12} および μ_{22} にはモード値を採用した．なお，メンバーシップ関数の幅の位置 W_{ij} は，(3.6)～(3.9)式を用いて算出した．

$$W_{i1} = \mu_{i1} + k_{i1} \times \sigma_{i1} \quad \cdots\cdots\cdots(3.6)$$

$$W_{i2\_1} = \mu_{i2} + k_{i2\_1} \times \sigma_{i2} \quad \cdots\cdots\cdots(3.7)$$

$$W_{i2\_2} = \mu_{i2} - k_{i2\_2} \times \sigma_{i2} \quad \cdots\cdots\cdots(3.8)$$

$$W_{i3} = \mu_{i3} + k_{i3} \times \sigma_{i3} \quad \cdots\cdots\cdots(3.9)$$

ここで，σ_{ij} は各形状群の標準偏差（表 3.1～表 3.3 参照），k_{ij} は関数の幅 W を調整する係数である．図 3.13 に示すように，隣り合う形状群同士で頻度の差が最も小さい階級において，それぞれのメンバーシップ関数（A_{i1} と A_{i2} ならびに A_{i2} と A_{i3}）が互いに Grade=0.5 で交わるように，k_{ij} の初期値 $k0_{ij}$ を設定する．次に，($k0_{ij}$ - 1)≦ k_{ij} ≦($k0_{ij}$ + 1)までの範囲を最小刻み 0.01 で調整し，各メンバーシップ関数の幅を決定する．このとき，各関数は必ず幅を有するものとし，$\mu_{i1} > W_{i1}$，$\mu_{i2} < W_{i2\_1}$，$\mu_{i2} > W_{i2\_2}$，$\mu_{i3} < W_{i3}$ の満たす範囲で k_{ij} を調整した．

なお，本章の実験で用いた各係数は k_{11}=0.162，k_{12_1}=2.00，k_{12_2}=1.60，k_{13}=2.36，k_{21}=1.61，k_{22_1}=2.33，k_{22_2}=1.55，k_{23}=1.66，k_{31}=1.51，k_{32_1}=3.00，k_{32_2}=2.53，k_{33}=1.88 である．

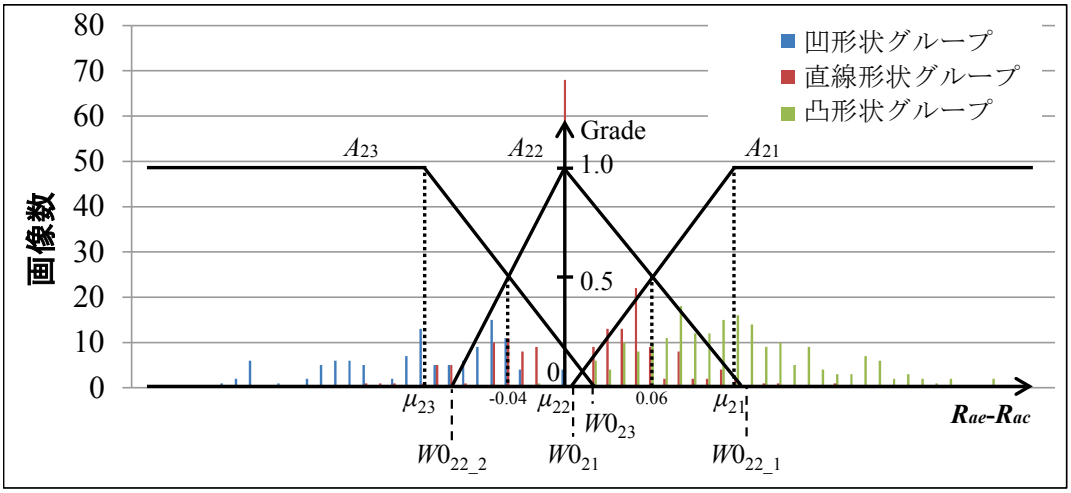


図 3.13 係数の初期値設定例（口裂形状での例）

Akita University

(2) 正規分布型メンバーシップ関数との比較

メンバーシップ関数の形状を決定するにあたり，図 3.14 に示す正規分布型のメンバーシップについても検討を加えた．正規分布型では，各形状群の平均値ならびに分散を用いて確率密度関数を決定するため，μ_{12} および μ_{22} にはそれぞれの形状群における平均値を用いた．

$$f_{ij}(x_i) = \frac{1}{\sqrt{2 \times \pi \times \sigma_{ij}{}^2}} \exp\left\{ \frac{(x_i - l_{ij} \times \mu_{ij})}{2 \times \sigma_{ij}{}^2} \right\} \qquad \cdots\cdots\cdots\cdots(3.10)$$

また，局所形状特徴量 R_{by}-R_{cy}，R_{ae}-R_{ac}，R_{xy} の各形状クラス群における確率密度関数の最大値を $\max f_{ij}(x_i)$ とし，正規分布型のメンバーシップ関数 NA_{ij} を生成する．

$$NA_{i1}(x_i) = \begin{cases} \dfrac{f_{i1}(x_i)}{\max f_{i1}(x_i)} & (x_i \le \mu_{i1}) \\ 1.0 & (x_i > \mu_{i1}) \end{cases} \qquad \cdots\cdots\cdots\cdots(3.11)$$

$$NA_{i2}(x_i) = \frac{f_{i2}(x_i)}{\max f_{i2}(x_i)} \qquad \cdots\cdots\cdots\cdots(3.12)$$

$$NA_{i3}(x_i) = \begin{cases} 1.0 & (x_i \le \mu_{i3}) \\ \dfrac{f_{i3}(x_i)}{\max f_{i3}(x_i)} & (x_i > \mu_{i3}) \end{cases} \qquad \cdots\cdots\cdots\cdots(3.13)$$

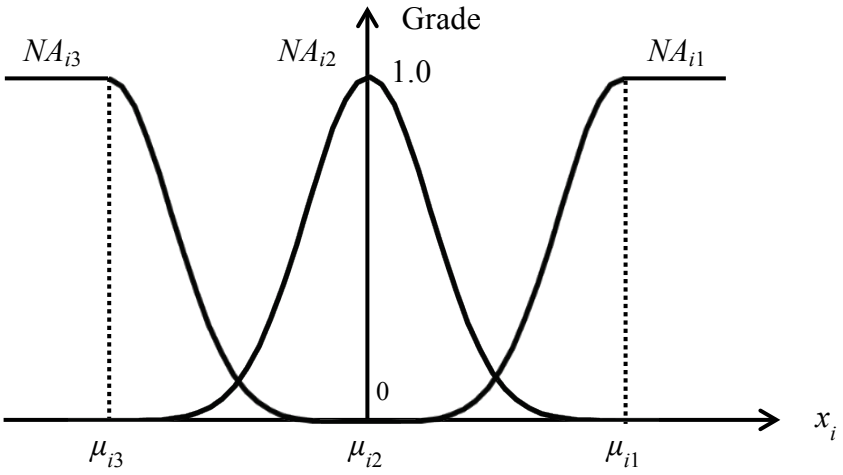


図 3.14 正規分布型メンバーシップ関数

Akita University

正規分布型 NA_{ij} では，形状特徴解析から得られた標準偏差 σ_{ij} を基準として，係数 l_{ij} を用いてメンバーシップ関数の形状に調整を加えた．係数 l_{ij} は 0.5 から 3 まで 0.1 刻みで調整し，最も良好な分類結果が得られた l_{11}=0.90，l_{12}=2.4，l_{13}=2.35，l_{21}=1.55，l_{22}=2.00，l_{23}=2.72，l_{31}=1.00，l_{32}=2.55，l_{33}=1.60 を採用した．

これらメンバーシップ関数の選択にあたり，データセット 1～5 を用いて，正規型メンバーシップ関数と三角型メンバーシップ関数の分類精度を比較した．その結果，正規分布型メンバーシップ関数と比較し，三角型メンバーシップ関数は第 3 位までの一致率が平均で 15.4%良好であり，第 4 位までの全ての一致率においても平均で 10.0%良好であった．したがって，以後の検討には三角型メンバーシップ関数を用いた．

3.5.3 口唇形状の分類処理

各局所形状クラスの判定結果に図 3.10 に示す組合せルールを適用し，1 位カテゴリ（BC：Belong Category）を出力した．また，図 3.4～3.10 で示したように，口唇形状の分布は多様であるため，同一の形状カテゴリに属する口唇群であっても，それぞれが有する形状は一律でない．例えば，C10 に属する被験者群（上下同等の厚さの口唇形状を有する）において，R_{by}-R_{cy} 値が上唇の厚い形状クラスに極めて近い被験者の口唇形状は，カテゴリ C10 と C1 との境界付近に属することになる．このように，カテゴリの境界付近に存在する口唇の場合，動画像取得時の僅かな姿勢変動などに起因し，隣接カテゴリへ誤分類される頻度が高くなることが推測される．そこで本研究では，局所形状クラス判別の際に算出した各形状クラスの帰属度に基づき，隣接カテゴリに NC1（Nearby Category 1：2nd カテゴリ），NC2（3rd カテゴリ），NC3（4th カテゴリ）の順位付けを行った．この NC1，NC2，NC3 を用いて，上記要因の誤分類が発生した場合にも複数の隣接カテゴリから順次，データ照合を行うことを可能にする．

Akita University

3.6 分類実験および考察

被験者 52 名（Ex2-id001～Ex2-id052）を対象とし，登録データと分類実験用データ 5 セットの分類結果の比較を行い，3 つの局所形状特徴を用いた口唇形状分類の有用性について検討した．

3.6.1 登録データ生成

解析用データにおける Ex2-id001～Ex2-id052 の口唇画像（各被験者 5 枚, 計 260 枚）を用いて提案手法による分類を実施し，各被験者の口唇形状の登録データを生成した．登録データ生成では，被験者ごとに口唇画像 5 枚分の平均帰属度を用いて，各局所形状クラスを判定した．最後に，各局所形状クラスの判定結果から各被験者が属する第 1 位カテゴリをそれぞれ求め，それを登録データ（正解データ）とした．

3.6.2 評価基準

提案手法の分類精度を評価するため，以下に示す①～③の評価基準を定めた．
①登録カテゴリと BC が一致：最良の結果と評価．
②NC1，NC2，NC3 が登録カテゴリと一致：隣接カテゴリに分類され，認証対象の絞り込みに有用な結果が得られたと評価．
③NC3 までの圏内に一致カテゴリ無し：異なる形状に分類．分類結果不良と評価．

3.6.3 分類結果に関する検討

実験に用いた分類実験用データセット 1～5 セット（各被験者 1 枚，各セット 52 枚）を入力データとし，口唇形状の分類実験を実施した結果を表 3.5 に示す．BC（1 位カテゴリ）と登録データとの一致率は 38.5%～50.0%であり，高い一致率は得られなかったものの，隣接 NC3 までの範囲では 80%以上の一致率が得られた．このことから，多くの被験者において，撮影時の姿勢変動などに起因する口唇形状の変位が生じたと考えられる．また，その変位の範囲は登録カテゴリの周辺に集中したと推測される．なお，各段階での一致率は，NC1 までの範囲（全 27 カテゴリ中 2 カテゴリの範囲）において 55.8%～69.4%，NC2 までの範囲（全 27 カテゴリ中 3 カテゴリの範囲）において 73.6%～82.8%，NC3 までの範囲（全 27 カテゴリ中 4 カテゴリの範囲）において 80.8%～86.6%である．

次に，データセット 1～5 における一致率の平均を調査したところ，NC1 との一致率は 20.8%，NC2 との一致率は 14.6%，NC3 との一致率は 5.0%となり，NC1>NC2>NC3 の順に一致率の高いことが認められた．また，データセット 4 で NC3>NC2，データセット 3 で NC2=NC3 という結果であったものの，全体的には NC1>NC2>NC3 の傾向を示すことが認められた．上記結果から，隣接カテゴリの適

Akita University

用と帰属度を参照した順位付け（NC1～NC3）は，照合対象データの絞込みに有用であることを示唆している．

一方，実際の使用状況を想定した場合，発話前後の顔画像（口を軽く閉じた状態の顔画像）を複数枚取得可能である．そこで，データセット 1～5 をまとめた統合データセット（データセット 6）を作成し，データセット 1～5 と同様に分類実験に用いた．データセット 6 では，口唇画像 5 枚の平均帰属度に基づいて局所形状クラス分類・形状カテゴリ判別を行った．その結果，データセット 1～5 と比較し，各段階で高い一致率が得られ，4 位カテゴリまでの範囲におけるトータルの一致率は 88.5%となった．この結果は，複数枚の画像を用いることで，提案手法の分類精度向上が可能であることを示唆している．なお，30 フレーム毎秒の動画撮影における 5 フレームは 0.17 秒程度であるため，発話の前後に軽く口を閉じる条件下ならば，発話の前後において複数枚の閉口状態画像を安定的に取得可能である．

表 3.5　登録データとの一致率（提案手法）

	BC と一致	NC1 と一致	NC2 と一致	NC3 と一致	Total
データセット 1	40.4%	25.0%	13.5%	1.9%	80.8%
データセット 2	46.2%	23.1%	13.5%	3.8%	86.6%
データセット 3	44.2%	19.2%	9.6%	9.6%	82.6%
データセット 4	38.5%	17.3%	25.0%	5.8%	86.6%
データセット 5	50.0%	19.2%	11.5%	3.8%	84.5%
Average (データセット 1-5)	43.8%	20.8%	14.6%	5.0%	84.2%
データセット 6	46.2%	25.0%	15.4%	1.9%	88.5%

Akita University

3.6.4 k-means 法による分類結果

提案手法の有用性を検討するため，代表的な教師無し分類として広く用いられている k-means 法[14][15]との比較を行った．入力データは登録データおよびデータセット 1～6 の特徴量（R_{xy}，R_{by}-R_{cy}，R_{ae}-R_{ac}）である．なお，クラスタ数は形状カテゴリと同数の 27，初期値はランダム，繰り返し回数 100000 回，分類の試行回数は 30 回である．

表 3.6 に試行 30 回中最も高い一致率を得た結果を示す．なお，各データセットにおける被験者（Ex2-id001～Ex2-id052）のクラスタが登録データと同じクラスタに分類された場合に一致と判定し，近傍 NC1～NC3 は対象データと各クラスタ中心とのユークリッド距離によって決定した．k-means 法による分類結果と比較し，提案手法は NC3 までの一致率（Total）が約 11.4%～19.3%高いという結果が得られた．さらに，BC～NC2 までの平均一致率においても提案手法が良好な分類結果を示した．また，初期値条件をランダムとしているため，k-means 法では試行ごとに分類結果が変動する．試行 30 回における第 1 カテゴリと登録データとの平均一致率は 33.1%～34.7%（データセット 1：33.1%，データセット 2：34.6%，データセット 3：32.9%，データセット 4：33.9%，データセット 5：34.5%，データセット 6：34.7%）であり，各試行における変動幅は約 11%～20%程度であった．一方，提案手法は試行ごとの結果に変動は生じないアルゴリズムであるため，提案手法は k-means 法と比較して安定した分類結果が得られた．

表 3.6　登録データとの一致率（k-means 法）

	BC と一致	NC1 と一致	NC2 と一致	NC3 と一致	Total
データセット 1	38.5%	13.5%	9.6%	7.7%	69.3%
データセット 2	44.2%	17.3%	5.8%	5.8%	73.1%
データセット 3	38.5%	23.1%	5.8%	3.8%	71.2%
データセット 4	44.2%	13.5%	3.8%	5.8%	67.3%
データセット 5	38.5%	13.5%	13.5%	1.9%	67.4%
Average (データセット 1-5)	40.8%	16.2%	7.7%	5.0%	69.7%
データセット 6	42.3%	13.5%	11.5%	3.8%	71.1%

Akita University

3.6.5 絞り込みに関する検討

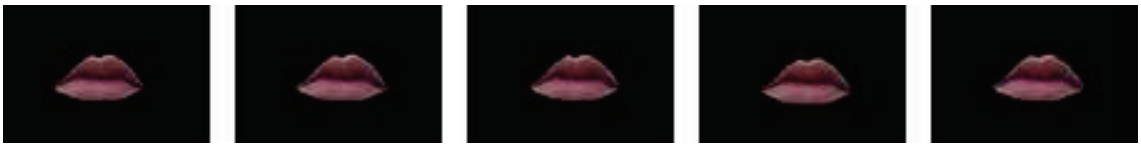
提案手法ではデータセット 6 において，88.5%（52 名中 46 名）の精度で一定の範囲内に絞り込みが可能であった．そこで，BC から NC3 までに形状が一致した被験者 46 名について，各カテゴリに属する登録データ数に基づき，そのマッチング候補数を求めたところ，平均約 6.1 名（全登録者数 52 名，約 11.7%）という結果が得られた．このことは，提案手法によってマッチング候補数が低減されたことを示している．残りの 6 名については絞り込み不能であるため，マッチング候補は 52 名となる．これを考慮した場合においても，マッチング候補数は平均 11.4 名（約 21.9%）に低減されている．次に，k-means 法による分類結果では，BC から NC3 までに形状が一致した被験者 35 名において，絞り込み範囲内に存在するマッチング候補の平均は約 4.8 名（約 9.2%）であった．この結果では k-means 法の方がマッチング処理時の探索範囲が小さいが，絞り込み不能の被験者を考慮した場合，平均的なマッチング候補数 18.4 名（約 35.4%）となり，提案手法の方が良好となる．

上記の結果から，グループ化による絞り込みが成功した場合には，以後に続くマッチング処理の負荷を低減することが可能であると考える．さらに，提案手法は形状分類の精度が比較的良好であるため，全体的な効果は高いことがわかる．

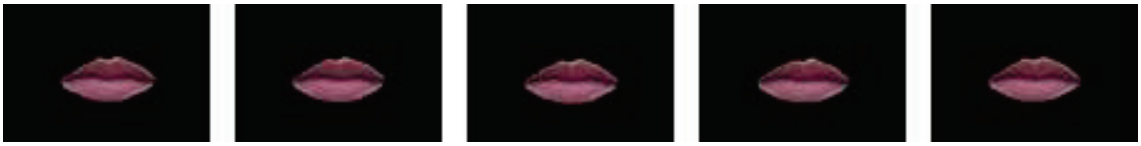
3.6.6 分類不良となった事例

データセット 1～5 の全てにおいて，NC3 までの圏内に一致カテゴリが無いと評価された被験者は 3 名（Ex2-id031，Ex2-id037，Ex2-id043）である．図 3.15 に Ex2-id031 の口唇画像を例示する．上記 3 名の被験者は，①カメラに対する顔の角度（特に上下角）が異なる，②閉口時の力み具合が異なることなどに起因し，登録データと実験データで口唇形状が大きく異なっていた．また，口唇形状の面では，③上唇，下唇ともに厚く，その厚さが同等程度である（いわゆるタラコ唇），④口裂が直線形状に近い（Ex2-id037，Ex2-id043）点が全体的な傾向として認められた．これに対し，全てのデータセットにおいて 1 位カテゴリと登録カテゴリが一致した被験者は 8 名である．一例として，Ex2-id012 の口唇画像を図 3.16 に示す．一致率の高い被験者の口唇画像は形状の変動や，上下方向の傾きの少ない傾向が認められる．したがって，被験者が安定した姿勢でカメラに正対することが可能であれば，より安定した分類結果が得られると考える．

Akita University

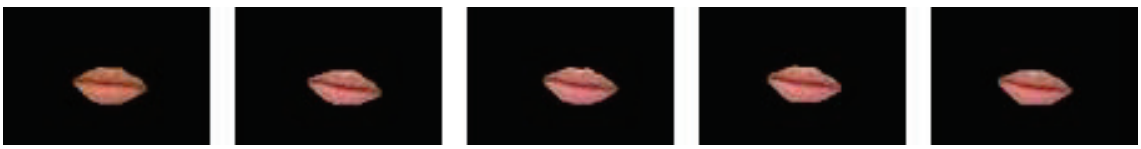


(a) 登録データの画像

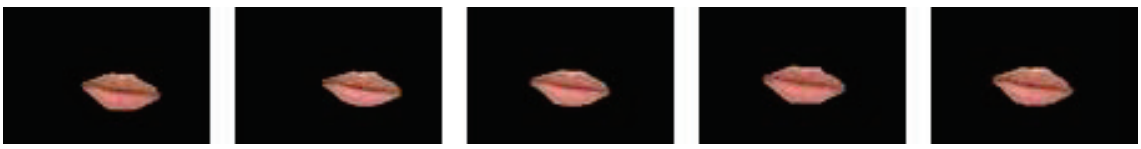


(b) 分類実験に用いた画像

図 3.15 分類失敗事例（Ex2-id031）



(a) 登録データの画像



(b) 分類実験に用いた画像

図 3.16 分類成功事例（Ex2-id012）

Akita University

3.7 まとめ

本章では，口唇形状に着目したグループ化法および照合対象データの絞り込み手法の開発を目的とし，口唇の局所部位に着目した形状解析を行い，抽出した特徴量を用いた形状分類処理について検討を加えた．得られた成果を以下にまとめる．

(1) 口唇のアスペクト比（R_{xy}），上唇・下唇厚さ特徴（R_{by}-R_{cy}），口裂形状特徴（R_{ae}-R_{ac}）は口唇形状の分類および各被験者のグループ化に有用な特徴量となることを明らかにした．

(2) 明度値 L*の垂直方向分布に着目することで，口裂を良好に抽出可能であることを明らかにした．

(3) 局所形状特徴に基づく形状カテゴリの生成，ならびに隣接カテゴリ NC1～NC3 の適用は，対象者のグループ化による照合データの絞り込みに有用であることが示唆された．

Akita University

第 3 章 文献

(1)白澤，三浦，西田，景山，栗栖：「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, Vol.60, No.12, pp.1964-1970 (2006)
(2)佐藤，景山，西田：「口唇の動き特徴を用いた非接触コマンド入力インタフェースの提案」，電学論 C, Vol.129, No.10, pp.1865-1873 (2009)
(3)中西，寺林，梅田：「インテリジェントルームのための DP マッチングを用いた口唇動作認識」，電学論 C, Vol.129, No.5, pp.940-946 (2009)
(4)齋藤，小西：「トラジェクトリ特徴量に基づく単語読唇」，信学論，Vol.J90-D, No.4, pp.1105 -1114 (2007)
(5)渡邊，西：「口部パターン認識を用いた日常会話伝達システムの研究」，電学論 C, 124-C, 3, pp.680-688 (2004)
(6)景山，安東，西田：「発話に伴う口唇の動き特徴を用いた心情変化の検出」，電学論 C, Vol.131, No.1, pp.201-209 (2011)
(7)小野，飯塚，吉竹：「口腔外科学（第 6 版）」金芳堂, 2002.
(8)C. M. Travieso, J. Zhang, P. Miller, J. B. Alonso, and M. A. Ferrer: “Bimodal biometric verification based on face and lips”, Neurocomputing, Vol.74, pp.2407-2410 (2011)
(9)C. Sforza, G. Grandi, M. Binelli, C. Dolci, M. D. Mendes, and V. F. Ferrario: “Age- and sex-related changes in three-dimensional lip morphology”, Forensic Science International, Vol.200, pp.182.e1-182.e7 (2010)
(10)S. Lucey, S. Sridharan and V. Chandran: “Adaptive mouth segmentation using chromatic features”, Pattern Recognition Letters, Vol.23, pp.1293-1302 (2002)
(11)高木，下田：「新編　画像解析ハンドブック」東京大学出版会 (2004)
(12)日本色彩学会編：「新編　色彩科学ハンドブック（第 3 版）」,東京大学出版会 (2011)
(13)日本規格協会：「JIS ハンドブック 61　色彩」，日本規格協会(2011)
(14)田中(編著)：「画像処理応用技術」，工業調査会(1989)
(15)安居院，長尾：「画像の処理と認識」，昭晃堂(1992)

Akita University

第 4 章　発話に伴う口唇の動き特徴と発声の関連に関する検討

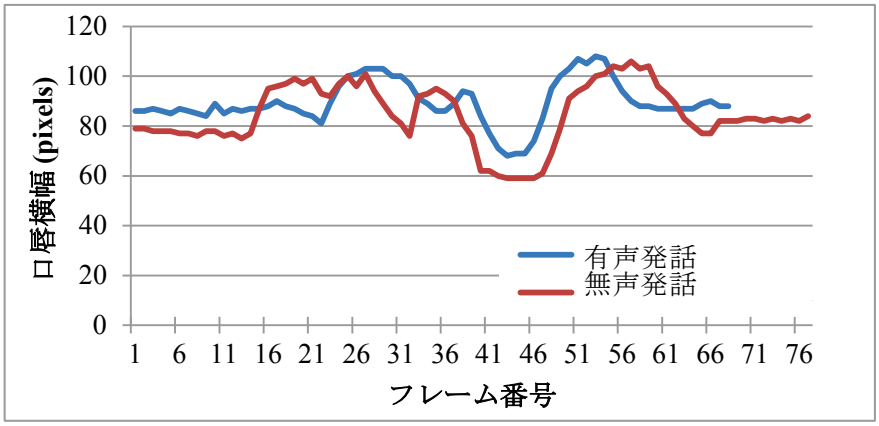
4.1 はじめに

第 2 章および第 3 章では，発話に伴う口唇の動き特徴を応用するための「発話区間自動推定」および非発話時の「口唇形状特徴に着目したグループ化法」について検討を加えた．これらの要素技術に加え，発話に伴う口唇の動き特徴を応用するシステムを構築するためには，「行動的特徴」[(1)-(3)]である口唇の動き特徴の変動を考慮したコマンド識別・発話認識技術が必要である．特に，口唇の動き特徴は，発声時のみでなく，発声をしない状況下でも取得可能であることから，発声の有無に起因する口唇の動き変動を考慮することは極めて重要である．また，この発声の有無に関する要素技術は，口唇の動き特徴の利用環境を広げることにつながると共に，音声認識との併用においても重要である．

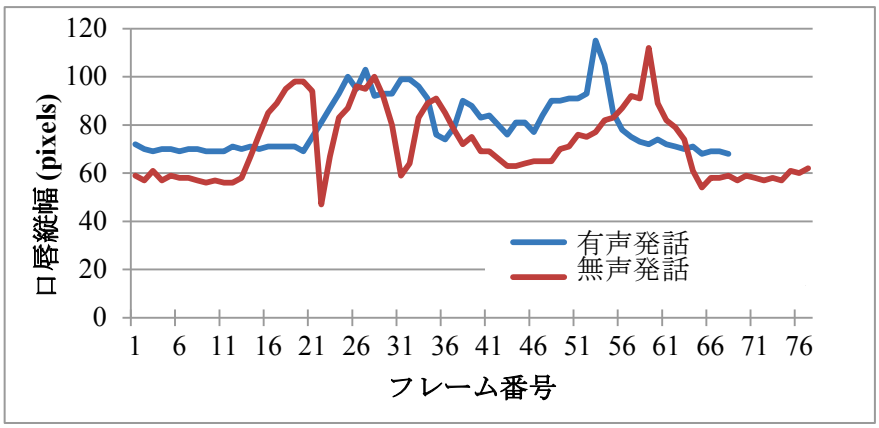
発話は，発声を伴う通常の発話（以後，有声発話と呼ぶ）と発声を伴わない発話（以後，無声発話と呼ぶ）に分けられる．有声発話は，音声による意思伝達を目的として行われる一般的な発話形式である．一方，無声発話は，視覚情報（いわゆる「ジェスチャー」）による意思伝達を目的に行われ，日常一般的には使用頻度の低い発話形式である．発話に伴う口唇の動き特徴は，これら 2 種類の発話のいずれにおいても同等に取得可能である．しかしながら実際には，図 4.1 に示すように，同一被験者が同一の内容を発話した場合の発話データにおいても，横方向や縦方向の口唇の動き，さらには発話の長さも異なる事例が多く認められた．したがって，口唇の横方向および縦方向の動き特徴を用いたコマンド識別手法を無声，有声いずれの発話においても利用可能とするためには，発声の有無と口唇の動き特徴の関連について明らかにする必要がある．しかしながら，筆者の調査した範囲において，口唇の動き特徴と発声の関連について検討された例は未だ存在しない．

そこで本章では，発声の有無と口唇横幅・縦幅ならびに発話区間の変動との関連について定量的な調査を行った．さらに，発声と動き特徴変動の調査結果に基づいた有声発話と無声発話の判別について検討を加えた．

Akita University



(a) 口唇横幅の推移例



(b) 口唇縦幅の推移例

図 4.1 口唇の動き特徴変動の例（発話内容：アキタウメコ）

Akita University

4.2 使用データ

4.2.1 データ取得の流れ

被験者 7 名(Ex3-id001～Ex3-id007)が同一の発話内容（以後，コマンドと呼ぶ）を「無声」および「有声」で発話した動画を CCD ビデオカメラ(SONY 製：DCR-VX2100)を用いて撮影した．データ取得に用いたコマンドは一人当たり 6 種類であり，その選定理由など詳細については次項（4.2.2 項）にて述べる．撮影時の照明などのデータ取得環境は 2 章および 3 章と共通であり（2.2.1 項参照），本章に限定した動画像取得条件のみを以下にまとめる．

- 発話動画像を撮影する前に 5 分間安静にする．
- 各コマンド 6 回の発話において，発話間に 30 秒間のインターバルをとる．
- 同一コマンドを「無声」と「有声」それぞれ 6 回ずつ発話する．

図 4.2 に示すように，以下の 4 ステップで発話動画像を取得した．なお，被験者には撮影開始前に十分な時間的余裕をもって実験室に入室してもらい, 実験の流れを示した図ならびに発話コマンドのリストを渡し, 発話データ取得の流れについて詳細な説明を行った．

Step (1) 被験者に 5 分間の安静を与える．
Step (2) 各コマンドを無声で 6 回発話させ，被験者の無声発話動画像を取得する（同一コマンドを 6 回無声で発話した後，30 秒の間隔をあけて次のコマンドの無声発話を開始）．
Step (3) 無声での発話終了後，被験者に 2 分間の休憩を与え，無声発話の影響を軽減（無声発話時の口唇動作イメージを除去）する．
Step (4) 指定されたコマンドを有声で 6 回発話させ，被験者の有声発話動画像を取得する．

なお, データ取得における発話回数については, 一度に 6 回までの発話であれば, 心理的負担・ストレスが小さいという調査報告 [4] に基づいて設定した．

Akita University

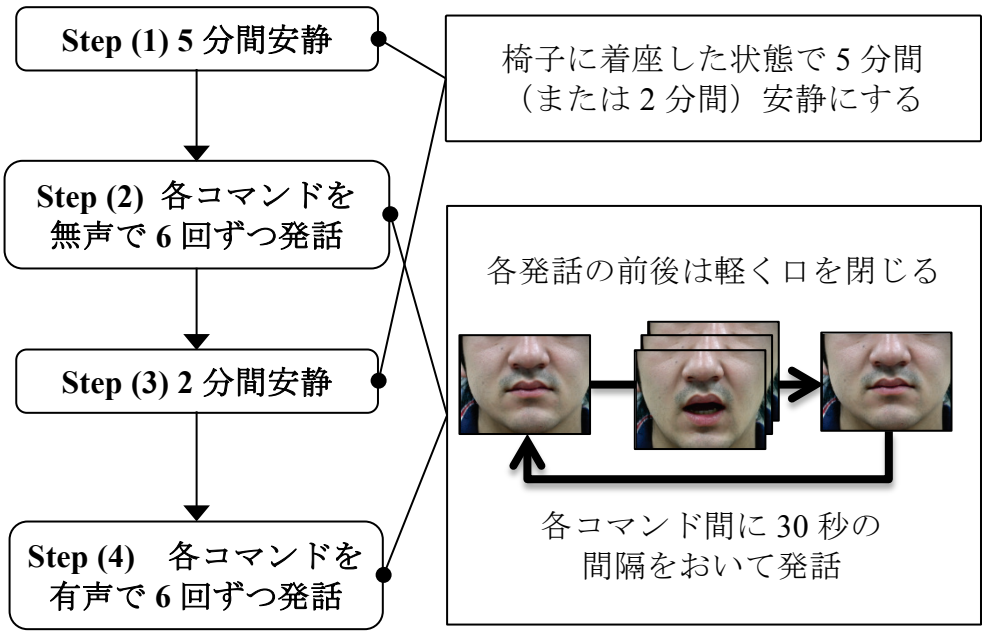


図 4.2　データ取得の流れ

Akita University

4.2.2 使用コマンドの選定

データ取得時に発話するコマンドとして 6 種類の単語を選定し，コマンド A～F と定義した.

・コマンド A :「“被験者自身の氏名”」
・コマンド B :「アキタウメコ」
・コマンド C :「オオダテケンシロウ」
・コマンド D :「カゲヤマヨウイチ」
・コマンド E :「ジョウホウコウガッカ」
・コマンド F :「シャシンシュウ」

各被験者に共通である 5 種類のコマンドの母音数は，コマンド B は 6 つ（a/ki/ta/u/me/ko），コマンド C は 9 つ（o/o/da/te/ke/n/shi/ro/u），コマンド D とコマンド E は 8 つ（ka/ge/ya/ma/yo/u/i/chi， jo/u/ho/u/ko/u/ga/kka），コマンド F は 4 つ（sya/shi/n/syu）である．選定基準および各コマンドとの関連を表 4.1 に示す.

被験者 7 名が無声および有声でコマンド A～F それぞれを 6 回ずつ発話した 72 本の発話動画像を 1 日分として取得し, 1 名あたり 216 本（72 本×3 日）, 合計 1512 本の発話データ取得を実施した．得られた発話動画像を 30fps の時系列静止画像に変換し，これを使用データとした.

表 4.1 使用コマンドおよびコマンド選定基準一覧

選定基準	使用コマンド					
	コマンド A：“被験者自身の氏名”	コマンド B：a/ki/ta/u/me/ko	コマンド C：o/o/da/te/ke/n/shi/ro/u	コマンド D：ka/ge/ya/ma/yo/u/i/chi	コマンド E：jo/u/ho/u/ko/u/ga/kka	コマンド F：sya/shi/n/syu
親しみのある言葉	○	×	×	×	どちらでもない	×
5 つの母音全てを含む	×	○	○	○	×	×
撥音“ん”を含む	×	×	○	×	×	○
破裂音を含む	×	○	○	○	○	×
長母音を含む	どちらでもない	×	○	○	○	○
コマンドの母音数	5 to 7	6	9	8	8	4

Akita University

4.3 特徴量抽出処理

4.3.1 前処理および口唇特徴計測

本章では，発話に伴う口唇の縦幅および横幅に基づく口唇の動作特徴量を解析対象としている．そこで，特徴量抽出のための前処理として，2 章ならびに 3 章と同様の過程で顔画像から口唇画像を抽出し，口唇の横幅（*diX*）ならびに口唇の縦幅（*diY*）を取得した．具体的には，発話動画像を 30fps の時系列静止画像へ変換し，得られた時系列の顔画像に対して L*a*b*表色系[5]を用いた色彩情報に着目した口唇抽出処理[6]を施した．次に，図 4.3 に示す口唇の横幅（*diX*），口唇の縦幅（*diY*）を計測し，さらに口唇輪郭に内接する矩形の面積（*S*），ならびに口唇のアスペクト比を算出した．最後に，得られた各値に対し，(4.1)～(4.4)式を用いて発話区間の初期フレームを基準とした口唇サイズの正規化処理を施し，口唇の動き変動特徴量のベースとなる raX_i，raY_i，raS_i，A_i を算出した．なお，i は発話区間のフレーム番号を表す．

$$raX_i = \frac{diX_i}{diX_1} \qquad \text{.....(4.1)}$$

$$raY_i = \frac{diY_i}{diY_1} \qquad \text{.....(4.2)}$$

$$raS_i = \frac{diX_i \times diY_i}{diX_1 \times diY_1} \qquad \text{.....(4.3)}$$

$$A_i = \frac{diX_i}{diY_i} \qquad \text{.....(4.4)}$$

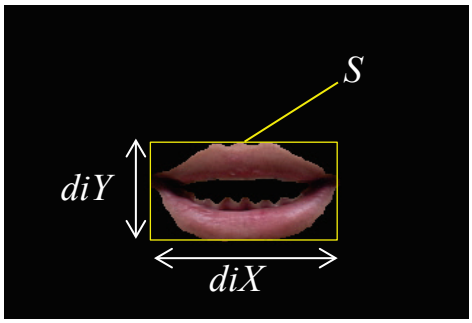


図 4.3 着目部位

Akita University

4.3.2 特徴量の算出

図 4.1 に示すように，発声は口唇縦幅や口唇横幅に加え，発話区間の長さにも影響を与えている．そこで本章では，口唇の動き推移（raX_i，raY_i，raS_i，A_i）と発話区間のフレーム数（FN）に着目し，発声の有無に起因する口唇の動き特徴変動を解析した．raX_i，raY_i は口唇横幅および縦幅の変動を直接的に表し，raS_i，A_i は全体的な開口量と動き方向を表す．すなわち，これらのフレーム間差分の累積値は，各発話における口唇の動作量を示す指標となる．そこで，(4.5)～(4.8)式を用いて各動き特徴量におけるフレーム間差分の累積値を算出し，それを無声発話と有声発話における口唇動作量特徴とした．FN は，各データにおける発話区間のフレーム数であり，発話区間はオペレータの手動により抽出した．

$$VraX = \sum_{i=1}^{FN-1} \left| raX_{i+1} - raX_i \right| \qquad \text{.....(4.5)}$$

$$VraY = \sum_{i=1}^{FN-1} \left| raY_{i+1} - raY_i \right| \qquad \text{.....(4.6)}$$

$$VraS = \sum_{i=1}^{FN-1} \left| raS_{i+1} - raS_i \right| \qquad \text{.....(4.7)}$$

$$VA = \sum_{i=1}^{FN-1} \left| A_{i+1} - A_i \right| \qquad \text{.....(4.8)}$$

無声発話と有声発話における特徴量の増減を比較するため，(4.9)式を用いて有声発話データに対する無声発話データの増加量を算出した．

$$R_k = \frac{100 \times (K_{non-vocalized} - K_{vocalized})}{K_{vocalized}} \qquad \text{.....(4.9)}$$

ここで，$K_{vocalized}$ は有声発話データにおける 5 つの特徴量（$VraX$，$VraY$，$VraS$，VA，FN）の何れかであり，$K_{non\text{-}vocalized}$ は $K_{vocalized}$ に対応する無声発話データにおける 5 つの特徴量である．

Akita University

4.4 口唇の動作量変化に関する検討

4.4.1 発話区間の変動

発話フレーム数 *FN* の 3 日間の平均増加率 R_{F_Ave} を表 4.2 に示す．各被験者の各取得日における各コマンド（6 データ）の平均 *FN* を用いて取得日ごとの増加率を算出し，その 3 日間の平均値から全 42 例(被験者数 7 名×コマンド数 6 種類)の R_{F_Ave} 算出した．全被験者，全コマンドの平均は 13.3%となり，全体的には無声発話のフレーム数が増加する結果となった．また，全 42 例中 29 例において，R_{F_Ave} が正の値を有し，その割合は約 69%であることからも，全体的には増加傾向を示していることがわかる．コマンド別の平均値を見ると，被験者全員に共通する 5 種類のコマンドでは，コマンド B の値が大きく，コマンド F の値が小さい．コマンド B「a/ki/ta/u/me/ko」は母音が「アイアウエオ」と上下方向に大きな開口動作が必要な「ア」や「エ」を多く有している．これに対し，コマンド F「sya/shi/n/syu」は母音数が少ない上に伸ばし音や撥音「ん」で構成されている．このことが増加率に差を生じる一因となったと推測される．

次に，被験者別に見た場合では，Ex3-id001，Ex3-id002，Ex3-id004，Ex3-id005 の 4 名は増加傾向を，Ex3-id003，Ex3-id006，Ex3-id007 の 3 名は微減の傾向を示し，傾向の異なる 2 群に分かれる結果となった．2 群それぞれにおける R_{F_Ave} の平均を求めたところ，増加傾向の被験者群（Ex3-id001，Ex3-id002，Ex3-id004，Ex3-id005）では 25.2%，微減傾向の被験者群（Ex3-id003，Ex3-id006，Ex3-id007）では-2.5%という結果となった．また，増加傾向の被験者群では R_{F_Ave} が負の値となる事例は認められず，負の値となった 13 例はすべて Ex3-id003，Ex3-id006，Ex3-id007 の 3 名のデータである．以上の結果を考慮し，4.4.2 項以降の各特徴量についても傾向の異なる 2 群に関して考察を加えた．

表 4.2 発話フレーム数の変化

	R_{F_Ave}[%]						
	コマンド 1	コマンド 2	コマンド 3	コマンド 4	コマンド 5	コマンド 6	全コマンドの平均
Ex3-id001	12.8	25.2	23.1	18.5	27.9	11.2	19.8
Ex3-id002	44.0	46.9	43.7	56.8	23.9	18.8	39.0
Ex3-id003	24.7	-1.8	-3.3	-6.5	-9.5	-7.3	-0.6
Ex3-id004	41.8	28.6	18.8	21.3	10.2	7.9	21.4
Ex3-id005	13.5	28.1	14.3	21.6	33.7	11.9	20.5
Ex3-id006	12.2	-3.9	-1.3	-10.3	2.6	-11.7	-2.1
Ex3-id007	-19.4	-4.1	-10.5	2.5	-0.9	3.8	-4.8
全被験者の平均	18.5	17.0	12.1	14.8	12.6	5.0	13.3
id001,id002,id004, id005 の平均	28.0	32.2	25.0	29.5	23.6	12.5	25.2
id003,id006,id007 の平均	5.8	-3.3	-5.0	-4.8	-2.6	-5.1	-2.5

Akita University

4.4.2 口唇横幅および縦幅の変動

VraX と *VraY* は各発話データにおけるフレーム間差分の累積値であり，大きくはっきりとした口唇動作を繰り返す発話データほど値が増加する特徴量である．発話フレーム数 *FN* と同様に，無声発話と有声発話の 3 日分の平均増減率を算出した結果 R_{VraX_Ave} を表 4.3，R_{VraY_Ave} を表 4.4 にそれぞれ示す．

口唇横幅の変動に関する増加量 R_{VraX_Ave} も，発話フレーム数と同様に全体的に無声発話の値が大きい傾向を示し，全 42 例中 36 例とほぼ全ての事例で正の値を有する結果となり，その平均は 27.7%であった．被験者別にみると，全ての被験者が増加傾向を示していることがわかる．コマンド別に半数のデータが負の値である Ex3-id006 の場合においても，全コマンドの平均では無声発話の *VraX* が 11.4%程度大きい結果となった．特に，増加率が大きい Ex3-id005 と EX3-id007 は R_{F_Ave} の平均が 50%以上という値を示したが，現状では個人差以外の要因は特定できていない．また，発話フレーム数で大きな違いが現れた 2 群については，それぞれの平均が 27.3%と 28.2%となり，2 群間で増減傾向に大きな違いは認められなかった．

次に，口唇縦幅の変動に関する平均増加量 R_{VraY_Ave} では，R_{VraX_Ave} と比較して負の値を有する事例が増加する結果となった．R_{VraY_Ave} において，正の値を有するのは全 42 例中 31 例であり，その平均値は 16.8%である．被験者別に見た場合，Ex3-id006 と Ex3-id007 においてデータの半数以上が負の値，すなわち無声発話の口唇動作量が小さいという結果が得られ，Ex3-id006 では全体の平均においても減少（-5.8%）した．R_{VraX_Ave} の検討と同様にフレーム数で異なる傾向を示した 2 群に分けた場合，フレーム数増加群（Ex3-id001, Ex3-id002, Ex3-id004, Ex3-id005）における R_{VraY_Ave} の平均は 26.5%，フレーム数減少群（Ex3-id003，Ex3-id006，Ex3-id007）における R_{VraY_Ave} の平均は 3.6%と大きな違いが現れた．このことから，左右方向よりも上下方向動作量が発話フレーム数との相関が強いものと推測される．また，Ex3-id005 や Ex3-id007 の結果に見られるように，*VraX* と *VraY* は個人差を顕著に表出する傾向にあるものと考えられる．なお，R_{VraX_Ave} と R_{VraY_Ave} 両データを合わせた結果では，正の値を有する事例は全 42 例中 28 例であり，その割合は 67%程度であった．

以上の結果より，*VraX* の *VraY* は被験者ごとに増減の傾向が大きくばらつくため，口唇の縦幅および横幅の直接的な増加量のみに着目した無声発話と有声発話の判別は困難であると推測される．

Akita University

表 4.3 *VraX* の 3 日間の平均増加量

	$R_{VraX\ Ave}$[%]						
	コマンド 1	コマンド 2	コマンド 3	コマンド 4	コマンド 5	コマンド 6	全コマンドの平均
Ex3-id001	-2.2	30.1	22.3	26.5	35.5	19.9	22.0
Ex3-id002	17.2	12.5	13.2	14.1	5.6	14.1	12.8
Ex3-id003	45.4	27.8	10.2	0	1.7	-4.1	13.5
Ex3-id004	4.3	1.0	1.4	18.7	-5.6	10.0	5.0
Ex3-id005	52.2	27.2	49.9	45.2	132.4	109.8	69.5
Ex3-id006	17.0	-9.0	-13.9	-26.1	38.7	61.5	11.4
Ex3-id007	44.6	27.4	76.7	38.0	87.2	84.7	59.8
全被験者の平均	25.5	16.7	22.8	16.6	42.2	42.3	27.7
id001,id002,id004, id005 の平均	17.9	17.7	21.7	26.1	42.0	38.5	27.3
id003,id006,id007 の平均	35.7	15.4	24.3	4.0	42.5	47.4	28.2

表 4.4 *VraY* の 3 日間の平均増加量

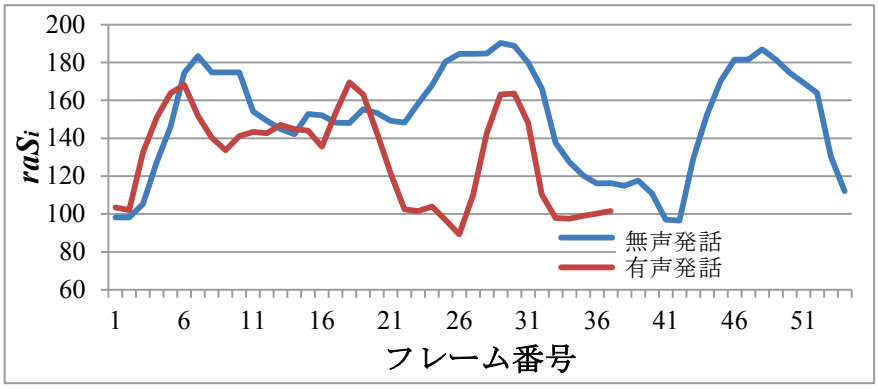
	$R_{VraY\ Ave}$[%]						
	コマンド 1	コマンド 2	コマンド 3	コマンド 4	コマンド 5	コマンド 6	全コマンドの平均
Ex3-id001	15.8	18.9	-1.0	19.8	37.8	40.1	21.9
Ex3-id002	26.7	22.5	53.3	37.6	21.8	36.1	33.0
Ex3-id003	57.5	24.5	21.3	1.5	6.0	-7.2	17.3
Ex3-id004	4.6	12.8	15.2	-3.4	19.7	23.8	12.1
Ex3-id005	-6.2	20.0	67.8	14.5	102.8	35.1	39.0
Ex3-id006	14.8	-41.8	4.0	-38.0	28.1	-1.8	-5.8
Ex3-id007	-6.1	-8.2	16.9	-15.8	29.9	-15.9	0.1
全被験者の平均	15.3	7.0	25.4	2.3	35.2	15.7	16.8
id001,id002,id004, id005 の平均	10.2	18.6	33.8	17.1	45.5	33.8	26.5
id003,id006,id007 の平均	22.1	-8.5	14.1	-17.4	21.3	-8.3	3.9

Akita University

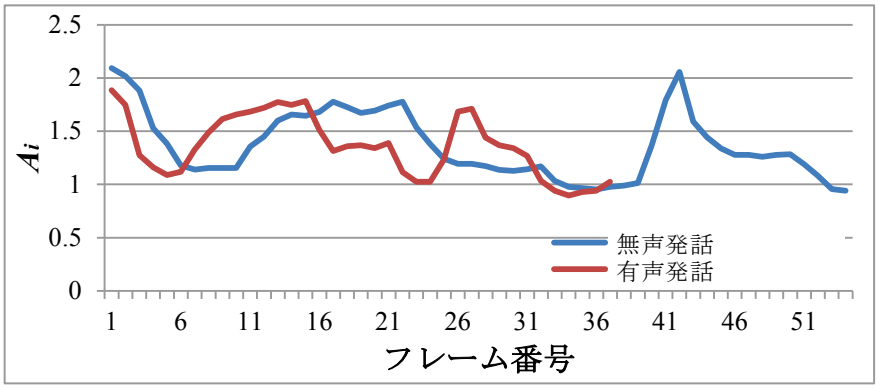
4.4.3 面積（raS_i）およびアスペクト比（A_i）の変動

図 4.4(a)(b)に raS_i と A_i の推移例を示す．有声発話時の推移グラフは最大値と最小値の幅が無声発話と比較して小さく，滑らかな推移を有する傾向が認められる．特に，母音を多く含むコマンドBおよびコマンドEではその傾向が顕著であった．

次に，表 4.5，表 4.6 に無声発話の 3 日間の平均増加量 R_{VraS_Ave} と R_{VA_Ave} の算出結果を示す．R_{VraS_Ave} および R_{VA_Ave} は R_{VraX_Ave} や R_{VraY_Ave} と比較して正の値を有する割合が高い結果となった．R_{VraS_Ave} の結果では，R_{VraX_Ave} や R_{VraY_Ave} の結果と比較して増加率自体は小さいものの，全 42 例中 37 例が正の値となり，その割合は 88.1%と 9 割弱を占めた．コマンドごとの平均では，全てのコマンドにおいて無声発話時の *VraS* が大きくなる結果が得られた．また，被験者別においても，全被験者において無声発話時の *VraS* が大きくなる結果が得られ，発話フレーム数増加群の R_{VraS_Ave} 平均は 23.7%，フレーム数減少群の R_{VraS_Ave} 平均は 8.2%であった．2 群に分けた場合の傾向は R_{VraY_Ave} と類似しており，*VraS* には上下方向の変動に関する特徴も包含されていることがわかる．



(a) 矩形領域 raS_i の推移



(b) アスペクト比 A_i の推移

図 4.4 raS_i ならびに A_i の推移例
（被験者 Ex3-id002，コマンド B の 2 回目の発話，1 日目）

Akita University

R_{VA_Ave} の算出結果では，全 42 例中 39 例（約 93%の割合）において無声発話時に *VA* が増加する結果が得られた．コマンドごとの平均では，*VraS* と同様に全てのコマンドにおいて無声発話時の *VA* が大きくなる結果が得られ，発話フレーム増加群では 24 例中 23 例において R_{VA_Ave} が正の値となり，その平均値は 23.6%，発話フレーム減少群においても 18 例中 16 例が正の値となり，その平均値は 11.5%となった．また，R_{VA_Ave} におけるコマンドごとの平均は 14.4%～22.8%であり，他の特徴よりも数値的に安定した結果が得られた．

表 4.5 *VraS* の 3 日間の平均増加量

	$R_{VraS\ Ave}$[%]						
	コマンド 1	コマンド 2	コマンド 3	コマンド 4	コマンド 5	コマンド 6	全コマンドの平均
Ex3-id001	7.5	20.6	3.8	19.8	38.5	37.0	21.2
Ex3-id002	19.3	22.9	49.4	30.3	22.2	34.6	29.8
Ex3-id003	43.9	23.3	-0.4	-4.2	3.0	1.2	11.1
Ex3-id004	0.7	11.0	9.5	-3.3	19.3	9.0	7.7
Ex3-id005	-0.2	17.1	37.9	0.7	120.9	39.6	36.0
Ex3-id006	10.5	7.0	3.0	4.5	5.1	0	5.0
Ex3-id007	0.4	7.6	-2.9	16.7	23.2	5.9	8.5
全被験者の平均	11.7	15.6	14.3	9.2	33.2	18.2	17.0
id001,id002,id004, id005 の平均	6.8	17.9	25.2	11.9	50.2	30.1	23.2
id003,id006,id007 の平均	18.3	12.6	-0.1	5.7	10.4	2.4	8.2

表 4.6 *VA* の 3 日間の平均増加量

	$R_{VA\ Ave}$[%]						
	コマンド 1	コマンド 2	コマンド 3	コマンド 4	コマンド 5	コマンド 6	全コマンドの平均
Ex3-id001	4.6	20.1	16.4	23.9	29.4	38.1	22.1
Ex3-id002	18.0	14.1	27.4	33.3	12.5	21.9	21.2
Ex3-id003	45.4	17.3	22.4	4.4	0	-8.4	13.5
Ex3-id004	14.4	12.4	5.2	9.2	0.9	24.6	11.1
Ex3-id005	16.4	19.7	58.2	34.7	68.3	43.8	40.2
Ex3-id006	21.3	13.8	8.0	1.6	13.5	-8.0	8.4
Ex3-id007	-0.5	3.2	21.8	14.8	27.5	9.3	12.7
全被験者の平均	17.1	14.4	22.8	17.4	21.7	17.3	18.5
id001,id002,id004, id005 の平均	13.4	16.6	26.8	25.3	27.8	32.1	23.6
id003,id006,id007 の平均	22.1	11.4	17.4	6.9	13.7	-2.4	11.5

Akita University

R_{VraS_Ave} および R_{VA_Ave} の両方が正の値となる事例は全 42 例中 34 例であった．この結果から，*VraX* および *VraY* と比較して，*VraS* および *VA* は口唇全体の動きの傾向を表し，その無声発話時の変動は増加傾向が強いことを示していると考える．したがって，*VraS* と *VA* は無声発話データと有声発話データを判別する有用な指標になる可能性があると考える．

以上の結果は，被験者が無声発話時に口唇を大きく明確に動かしていることを示唆している．これは，無声発話時の口唇動作が被験者のイメージもしくは視覚情報（カメラ映像）に依存し，被験者は口唇動作の明確さに注力しているためと考える．一方，有声発話では発声器官 (7) を動かし，発した音声を認識しつつ発話を行う．これらの要因が複合し，無声発話と有声発話では口唇動作に違いが生じるものと推測される．なお，実際の口唇動作について，被験者 7 名の発話動画像に対する目視調査を行っており，無声発話時は無声発話時と比較し，口唇の開閉動作が明確に行われる傾向にあることを認めている．

Akita University

4.5 口唇の動作量に着目した発声データの判別

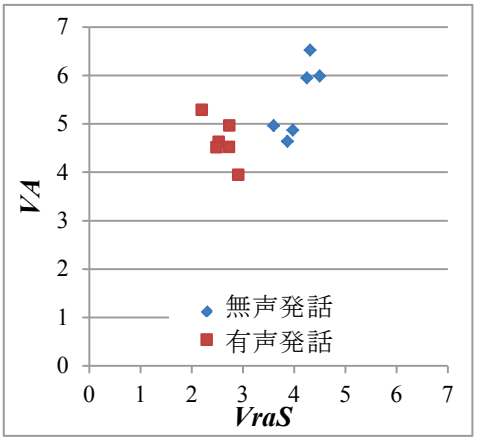
4.4 節では，発声の有無に起因する口唇動作量の変動について解析を行った．本節では，4.4 節で得られた結果を踏まえ，*FN*，*VraS*，*VA* に着目した無声発話データと有声発話データの判別について検討を加えた．

4.5.1 発話データセット

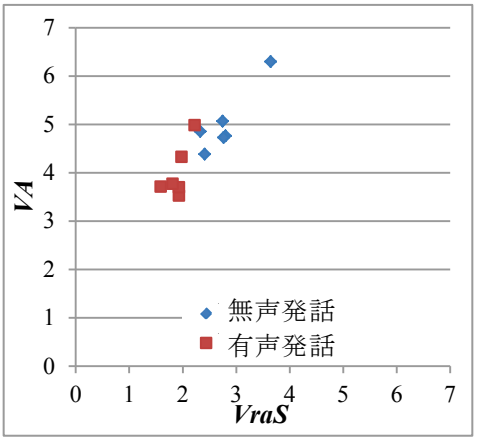
発話に伴う口唇の動きは，発話慣れなどの影響で僅かながら経時的に変動する．このことに起因し，*FN*，*VraS*，*VA* も変化する．*FN*，*VraS*，*VA* の経時的な変動の具体例として，図 4.5 に被験者 Ex3-id002 がコマンド B を発話した 3 日分のデータにおける *VraS*-*VA* の散布図を示す．図 4.5(d)から，*VraS* ならびに *VA* が徐々に小さくなり，3 日目には無声発話データ，有声発話データともに 2.0～4.0 の範囲まで変位したことがわかる．このように口唇の動作量が一定方向に収束する傾向が見える一方，3 日目のデータにおいても無声発話データと有声発話データが混在しておらず，それぞれが口唇動作量に関する特徴を有していることも見てとれる．この傾向は他の被験者についても同様であった．したがって，無声と有声のコマンド入力それぞれの習熟度が同程度であれば，無声データと有声データの判別が可能であると推測される．

そこで，同一取得日かつ同一コマンドの発話データ 12 データ（無声 6 データ，有声 6 データ）を発話データセット U_p と定義し，発話データセットごとに無声データと有声データの判別を試みた．図 4.6 に示すように，発話データセット U_p の総数は，被験者数とコマンド数ならびにデータ取得日の積となり，本研究のデータでは全 126 データ（p=126）である．

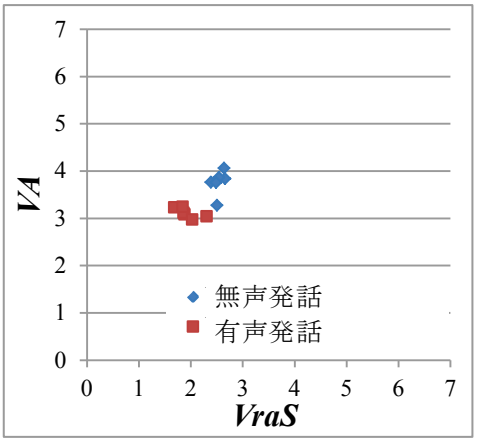
Akita University



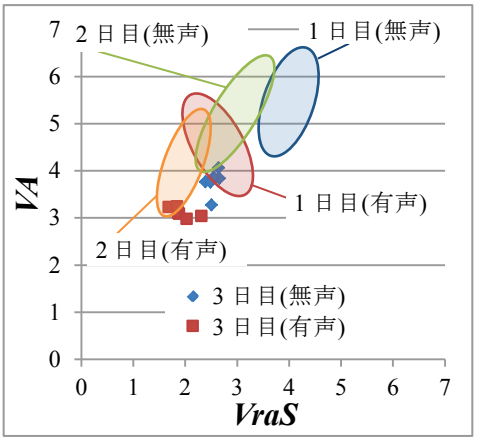
(a) 1 日目の発話データ



(b) 2 日目の発話データ



(c) 3 日目の発話データ



(d) 3 日間の変位

図 4.5 *VraS* と *VA* の散布図（被験者 Ex3-id002, コマンド B）

被験者 ID	コマンド	データ取得日	発話データ番号 *m*
Ex3-id001	A	1	1
			2
			3
			4
			5
			6
			7
			8
			9
			10
			11
			12

発話データセット
$U_p(m)$, m= 1 から 12

無声発話データ

有声発話データ

図 4.6 発話データセット

Akita University

4.5.2 無声，有声判別に関する検討

(1) 2 変数を用いた判別

4.4 節の解析結果に基づき，*VraS* と *VA* が判別の指標になり得るものと仮定し，無声データと有声データの線形判別について検討を行った．本章では，3 つの判別関数を決定し，それぞれの判別率（正解データとの一致率）を評価した．線形判別関数 1（*DF*1）は，*VraS* と *VA* を回帰変数としてデータセット U_p に対して重回帰分析[(8)-(10)]を行い，その結果に基づいて決定した．(4.10)式が線形判別関数 1（*DF*1）であり，係数 A_{p0} ～ A_{p3} は重回帰分析によって得られた値である．

$$DF1_p(m) = A_{p0} + A_{p1} \times VraS_p(m) + A_{p2} \times VA_p(m) \qquad \text{.....(4.10)}$$

多くの発話データにおいて R_{VraS_Ave} と R_{VA_Ave} の両方が正の値を有することから，*VraS* と *VA* が単純な比例関係にあるものと仮定し，(4.11)式の判別関数 2（*DF*2）を用いた判別についても検討を加えた．単純比例であることを前提に傾きを任意の値“-1”と決定し，*VraS*-*VA* 座標において U_p における無声発話 6 データの平均座標と有声発話 6 データの平均座標の中点を *DF*2 が通ると仮定して，*VA* 軸との切片 C を決定した．

$$DF2_p(m) = -VraS_p(m) + C \qquad \text{.....(4.11)}$$

表 4.7 にコマンド別の判別結果を示す. 評価は各データセット U_p (p=1,2,3,…126) における 12 データを一単位として実施し，全 126 データの一致率をコマンドごとにまとめた．判別関数 1 では一致率の平均が 84.3%という結果が得られ，最も一致率が低かったコマンド F の 3 日目のデータにおいても 78.6%であった．

一方，判別関数 2 では全体の平均が 72.4%であり，判別関数 1 と比較して 10%以上低い結果となった．なお，コマンド間においては，判別関数 1，2 共に一致率に大きな差は認められなかった．

次に，表 4.8 に被験者別の判別結果を示す．被験者間では一致率に大きな差が生じていることがわかる．Ex3-id002 と Ex3-id005 は判別関数 1 の結果が特に良好であり, Ex3-id002 については判別関数 2 においても良好な結果が得られている．一方，Ex3-id006 は，判別関数 1，2 のいずれを用いた場合においても被験者 7 名中で最も判別が困難であった．Ex3-id006 の判別関数 1 における 3 日間の平均は 68.5%と全被験者の平均と比較して 15.7%低く, 判別関数 2 に至っては 3 日間の平均が 59.7%と有用性がほとんど認められない結果となった．

Akita University

表 4.7 コマンド別の判別結果（*DF*1，*DF*2）

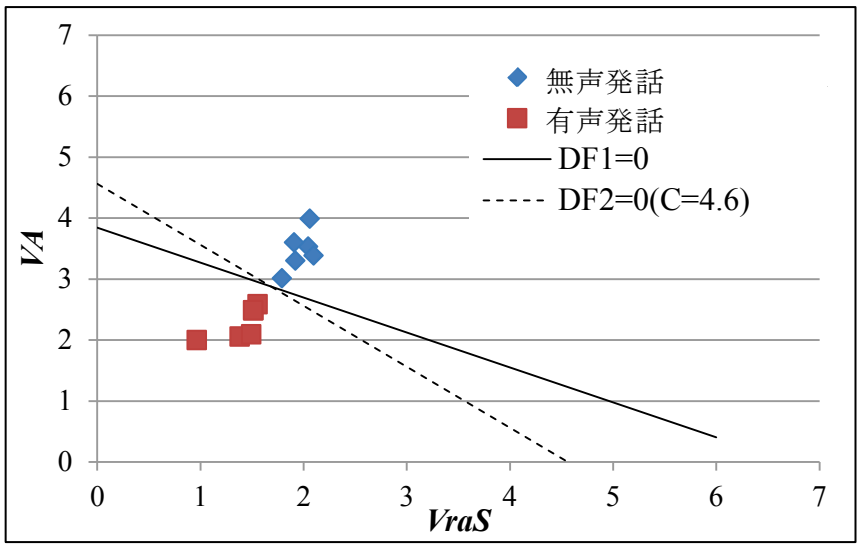
コマンド	データ取得日	判別率 [%]	
		***DF*1**	***DF*2**
A	1 日目	86.9	63.1
	2 日目	86.9	77.4
	3 日目	90.5	71.4
B	1 日目	81.0	71.4
	2 日目	86.9	84.5
	3 日目	82.1	67.9
C	1 日目	81.0	76.2
	2 日目	83.3	75.0
	3 日目	85.7	64.3
D	1 日目	81.0	76.2
	2 日目	81.0	63.1
	3 日目	84.5	67.9
E	1 日目	88.1	79.8
	2 日目	84.5	75.0
	3 日目	89.3	86.9
F	1 日目	82.1	67.9
	2 日目	83.3	70.2
	3 日目	78.6	64.3
全コマンドの平均	1 日目	83.3	72.4
	2 日目	84.3	74.2
	3 日目	85.1	70.4
	3 日間	84.3	72.4

表 4.8 被験者別の判別率（*DF*1，*DF*2）

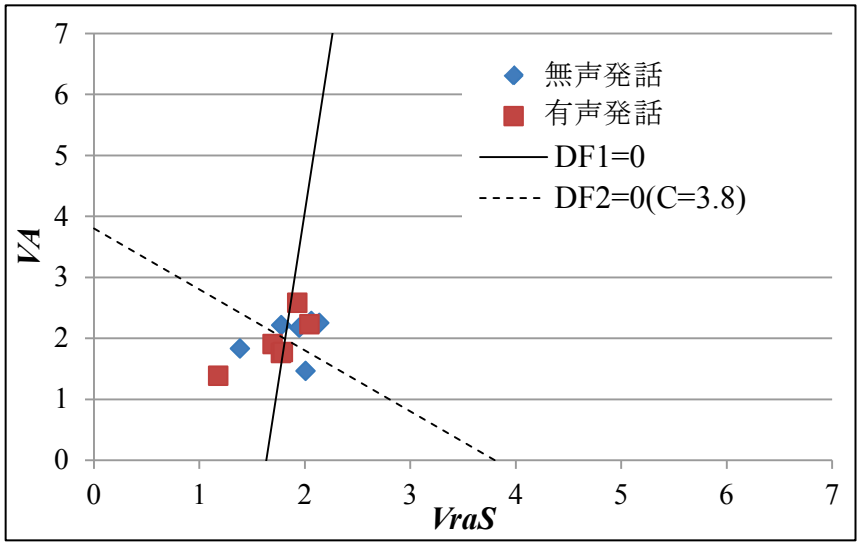
被験者 ID	データ取得日	判別率 [%]	
		***DF*1**	***DF*2**
Ex3-id001	1 日目	84.7	80.6
	2 日目	88.9	88.9
	3 日目	86.1	68.1
Ex3-id002	1 日目	98.6	95.8
	2 日目	91.7	88.9
	3 日目	100.0	100.0
Ex3-id003	1 日目	84.7	65.3
	2 日目	80.6	59.7
	3 日目	84.7	76.4
Ex3-id004	1 日目	79.2	72.2
	2 日目	79.2	68.1
	3 日目	72.2	48.6
Ex3-id005	1 日目	83.3	69.4
	2 日目	95.8	81.9
	3 日目	100.0	75.0
Ex3-id006	1 日目	65.3	55.6
	2 日目	70.8	56.9
	3 日目	69.4	66.7
Ex3-id007	1 日目	87.5	68.1
	2 日目	83.3	75.0
	3 日目	83.3	58.3

Akita University

図 4.7 に Ex3-id005 と Ex3-id006 の判別結果例を示す． Ex3-id005 の発話データは，図 4.5(a)~(c)に示す発話データと同様に，無声発話 6 データと有声発話 6 データの混在割合が小さく，*VraS* と *VA* を回帰変数とした重回帰分析によって線形判別が可能であった．一方，Ex3-id006 の発話では無声発話データと有声発話データが混在しており，発声の有無を判別することが困難であった（図 4.7(b)参照）．このように，無声発話データと有声発話データが混在したデータは Ex3-id003 や Ex3-id004 にも見られるが，その 2 名と比較しても Ex3-id006 は発話データセット 18 セット中 17 セットと突出して多い．したがって，この判別手法では Ex3-id006 のような無声発話と有声発話における *VraS* と *VA* の差異が小さい被験者への対応は困難と考える．



(a) Ex3-id005 における判別例（コマンド C, 3 日目の発話データセット）



(b) Ex3-id006 における判別例（コマンド C, 3 日目の発話データセット）

図 4.7 判別関数 1 および 2 を用いた判別結果例

Akita University

(2) 3 変数を用いた判別

口唇動作に関する特徴量 *VraS* と *VA* を用いた判別では，良好な判別結果の得られない被験者（Ex3-id006）が認められた．一方，発話フレーム数の増加率 R_{F_Ave} を求めた結果（表 4.2 参照）から，各被験者の R_{F_Ave} は正負何れか一方向に変化する傾向を認めている．そこで，発話フレーム数 *FN* を加えた 3 つの変数による判別を試みた． *diX* および *diY* から求められた特徴量である *VraS* および *VA* に加えて発話区間のフレーム数 *FN* を導入するため，3 つ変数それぞれに対して正規化処理を施した．具体的には，発話データセット U_p ごとに(4.12)式を用いて正規化した．

$$Xr_p(m) = \frac{X_p(m) - \overline{X}_p}{\sigma_{Xp}} \qquad \text{.....(4.12)}$$

ここで，X_p は U_p における 3 つの特徴量（$VraS_p$，VA_p ならびに FN_p）を表し，*Xr* はその正規化された値である．

正規化処理された 3 特徴量（$VraSr_p$，VAr_p ならびに FNr_p）を用いて重回帰分析を行い，(4.13)式に示す判別関数 3（*DF*3）を決定した．係数 B_{p0} ～ B_{p3} は $VraSr_p$，VAr_p ならびに FNr_p を用いた重回帰分析から得られた値である．

$$DF3_p(m) = B_{p0} + B_{p1} \times VraSr_p(m) + B_{p2} \times VAr_p(m) + B_{p3} \times FNr_p(m) \qquad \text{.....(4.13)}$$

判別関数 3 による判別結果を表 4.9 に示す．全データセットの平均判別率は 91.8%であり，判別関数 1 と比較して 7.5%判別率が向上している．コマンド別に見てもすべてのコマンドにおいて判別率が向上している（6.0%～10.7%）ことがわかる．被験者別では，判別関数 1 および 2 で良好な結果が得られなかった Ex3-id006 において，10.2%判別率が向上していることに加え，Ex3-id004 についても判別率が 15.7%と大きく向上する結果を得た．

次に，データ取得日別に見ると，1 日目のデータよりも 2 日目，3 日目のデータの方が良好な判別結果を得ている．このことから，経時的な要因によって口唇の動き特徴変化が生じた場合にも，無声発話と有声発話の差異は保持されると考えられる．

以上の結果は，*VraS*，*VA* ならびに *FN* に着目した線形判別法は，同時期に発話された無声データと有声データを判別する上で有用であることを示唆している．

Akita University

表 4.9 判別関数 3（*DF*3）の判別率

コマンド	データ取得日	判別率[%]	被験者 ID	データ取得日	判別率[%]
A	1 日目	90.5	Ex3-id001	1 日目	94.4
	2 日目	96.4		2 日目	98.6
	3 日目	97.6		3 日目	97.2
B	1 日目	79.8	Ex3-id002	1 日目	98.6
	2 日目	94.0		2 日目	98.6
	3 日目	95.2		3 日目	100.0
C	1 日目	91.7	Ex3-id003	1 日目	88.9
	2 日目	91.7		2 日目	90.3
	3 日目	91.7		3 日目	91.7
D	1 日目	94.0	Ex3-id004	1 日目	91.7
	2 日目	91.7		2 日目	95.8
	3 日目	92.9		3 日目	90.3
E	1 日目	91.7	Ex3-id005	1 日目	87.5
	2 日目	95.2		2 日目	100.0
	3 日目	92.9		3 日目	100.0
F	1 日目	85.7	Ex3-id006	1 日目	72.2
	2 日目	92.9		2 日目	80.6
	3 日目	86.9		3 日目	83.3
平均	1 日目	88.9	Ex3-id007	1 日目	88.9
	2 日目	93.7		2 日目	91.7
	3 日目	92.9		3 日目	87.5
	3 日間	91.8			

Akita University

4.6 まとめ

本章では，口唇の動き特徴を用いたコマンド入力および発話認識の精度向上を目的とし，発声の有無に起因する口唇の動き特徴変動に関する解析，および無声発話データと有声発話データの判別法に関する検討を行った．得られた成果を以下にまとめる．

(1)無声発話は有声発話と比較し，口唇の動作量が大きくなる傾向を有することが明らかになった．特に，口唇輪郭を包含する矩形の面積とアスペクト比のフレーム間差分の累積値は無声発話と有声発話の差異を示す有用な指標となる．

(2)発話区間の長さは被験者ごとに増減傾向が異なり，無声発話時に発話区間が長くなる群と短くなる群に大別可能であった．また，データ取得日が異なっても，増加群，減少群ともにその傾向は保持されることが明らかになった．

(3)特徴量 *VraS*，*VA* ならびに *FN* の重回帰結果に基づく線形判別法は，同時期に発話された無声発話データと有声発話データを良好に判別可能であることを明らかにした．

Akita University

第 4 章 文献

(1)瀬戸: 「Biometrics Technology in Cyber-Security」, 共立出版(2002)
(2)映像情報メディア学会(編), 半沢(編著) :「バイオメトリクス教科書　原理からプログラミングまで」,コロナ社(2012)
(3)佐藤, 景山, 西田 :「口唇の動き特徴を用いた非接触コマンド入力インタフェースの提案」, 電学論 C, Vol.129, No.10, pp.1865-1873 (2009)
(4)景山, 安東, 西田眞 :「発話に伴う口唇の動き特徴を用いた心情変化の検出」, 電学論 C, Vol.131, No.1, pp.201-209 (2011)
(5)日本色彩学会編:「新編　色彩科学ハンドブック（第 3 版)」, 東京大学出版会(2011)
(6)白澤, 三浦, 西田, 景山, 栗栖 :「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, Vol.60, No.12, pp.1964-1970 (2006)
(7)坂井, 河原 編 :「カラー図解　人体の正常構造と機能（改訂第 2 版)」, 日本医事新報社(2012)
(8)浜本 :「統計的パターン認識入門」, 森北出版(2009)
(9)中村 :「例解回帰分析入門」, 日刊工業新聞社(1982)
(10)竹澤 :「シミュレーションで理解する回帰分析」, 共立出版(2012)
(11)高木, 下田監修:「新編　画像解析ハンドブック」, 東京大学出版会 (2004)

Akita University

第 5 章　結論

近年，高機能化の著しい各種情報機器の利用を支援するために，日常一般的な動作である「発話」に着目したインタフェースの研究・開発が行われており，カーナビゲーションなど多くのシステムに応用されている．この「発話」動作には，音声情報に加えて口唇の動きという視覚情報も包含されており，視覚情報を用いたコマンド識別・発話認識も可能である．しかしながら，その実用化には，多くの利用者が共用する状況においても良好な認識を可能とするための要素技術，ならびに自然な発話条件を実現するための要素技術の開発が必要不可欠である．

本論文では，口唇の動き特徴を入力情報としたヒューマンマシンインタフェースの実用化に向けた要素技術について基礎的な検討を加えた．以下に本論文で得られた主な結果を記し，それに引き続いてこれらの工学的意義についてまとめる．

5.1　本論文により得られた主な知見

第 1 章では本研究の目的および本研究に対する筆者の立場を述べた．また，本論文の主題である口唇の動き特徴を用いたコマンド識別・発話認識インタフェースにおいて求められる要素技術について，今日までの研究状況を概観するとともに，本論文の内容について述べた．

第 2 章では，発話区間の自動推定処理について検討を加え，発話時の口唇画像における L*a*b*表色系の色彩情報および口唇形状の時系列変化を特徴量とする発話フレーム自動検出法を提案し，その有用性について検討を加えた．得られた結果を以下にまとめる．

(1) L*a*b*色空間を用いた口唇領域の色彩情報解析を行った結果，口唇の垂直方向における L*および a*推移は，閉口状態の口唇画像を抽出する上で有用な特徴量となることを明らかにした．

(2) 色彩情報（L*および a*）に着目した口裂有無の判定処理およびフレーム間差分による口唇形状変化の判定処理を用いた発話フレーム検出法は，発話フレームを高精度（約 91.4%～約 99.4%）に検出可能であることが明らかとなった．

(3) 上唇結節周辺の 3 本の口裂判定垂線（P_1～P_3）による口裂判定，算出処理(i)による 3 フレーム間の形状変化判定処理を用いた発話フレーム検出法が高検出率かつ誤検出の少ない手法であることを明らかにした．

(4) 3 つの単語を任意のタイミングで発話した場合にも，発話フレームを良好に検出可能であり，発話区間の大部分を推定可能であることを明らかにした．

第 3 章では，利用者の増加を想定した場合においても良好にコマンドの識別を可能にすることを目的とし，口裂などの局所形状について統計的な解析を行い，口唇形状を分類するために有用な特徴量を抽出した．さらに，局所形状解析結果に基づいた顔画像のグループ化法を提案し，その有用性について検討を加えた．得られた結果を以下にまとめる．

(1) 口唇のアスペクト比（R_{xy}），上唇・下唇厚さ特徴（R_{by}-R_{cy}），口裂形状特徴（R_{ae}-R_{ac}）

Akita University

は口唇形状の分類および各被験者のグループ化に有用な特徴量になることを明らかにした．

(2) 明度値 L*の垂直方向分布に着目することで，口裂を良好に抽出可能であることを明らかにした．

(3) 局所形状特徴に基づく形状カテゴリの生成，ならびに隣接カテゴリ NC1～NC3 の適用は，対象者のグループ化による照合データの絞り込みに有用であることが示唆された．

第 4 章では，発声の有無が口唇動作量ならびに発話区間の長さへ与える影響に関して定量的な調査を行い，発話フレーム数および口唇の動作量の差異に着目した有声発話と無声発話の判別手法について検討を加えた．得られた結果を以下にまとめる．

(1) 無声発話は有声発話と比較し，口唇の動作量が大きくなる傾向を有することが明らかになった．特に，口唇輪郭を包含する矩形の面積とアスペクト比のフレーム間差分の累積値は，無声発話と有声発話の差異を示す有用な指標となることを明らかにした．

(2) 発話区間の長さは被験者ごとに増減傾向が異なり，無声発話時に発話区間が長くなる群と短くなる群に大別可能であった．また，データ取得日が異なっても，増加群，減少群ともにその傾向は保持されることが明らかになった．

(3) 特徴量 *VraS*，*VA* ならびに *FN* の重回帰結果に基づく線形判別法は，同時期に発話された無声発話データと有声発話データを良好に判別可能であることを明らかにした．

Akita University

5.2 本論文の工学的意義

以下に本論文の工学的意義について述べる．

(1)口唇の動き特徴を用いたコマンド識別・発話認識インタフェースの利便性向上において，より自然な発話状態での入力をユーザに提供することは重要な課題である．本論文では，口唇領域の色彩情報と時系列形状情報に着目した発話フレーム検出法を提案し，3 つの単語を任意の間隔で発話した場合においても，良好に発話フレームを検出できることを明らかにした．被験者 5 名のデータを用いた評価において，提案手法は 99%以上の精度で発話フレームを検出可能であり，複数の単語を含む発話データにおいても，画像情報のみから良好に発話フレームを検出可能であることを示した．

(2)発話に伴う口唇の動きに着目したコマンド識別手法や発話認識手法については，これまでにも数多くの検討がされていたものの，利用者の増加を想定した検討は十分に行われていなかった．本論文では，上唇・下唇の厚さ，口裂が成す曲線の凹凸方向，ならびに口唇のアスペクト比の 3 種類の形状特徴についての統計的な解析を行い，各形状がそれぞれ 3 クラスに大別可能であることを明らかにした．さらに，口唇の局所形状に基づいて被験者を 27 カテゴリに自動分類するアルゴリズムを提案し，被験者 52 名を対象にした評価実験において，80%以上の精度で登録データおよびその類似形状に分類可能であることを示した．また，最も良好な分類結果となったデータにおいて照合対象者を算出したところ，52 名中の 11.4 名という結果が得られ，照合対象の絞込みが可能であることを明らかにした．

(3) 発話に伴う口唇の動き特徴を用いたインタフェースは，発声の有無に係らず利用可能であるという利点を有するものの，実際には発声の有無に起因して口唇の動き特徴に差異の生じる事例が認められている．しかしながら，発声と口唇の動きの関連についての検討はこれまで行われていなかった．本論文では，発話フレーム数および口唇の動き特徴（縦幅，横幅，面積，アスペクト比）に着目し，各特徴量と発声との関連ついて検討を加えた．その結果，無声時は有声時と比較し，発話区間が長くなる傾向を有すること，発話全体を通した口唇の動作量が大きくなる傾向を有することを明らかにした．さらに，同一取得日の発話データでは，発話フレーム数および口唇動作量に着目した線形判別により，約 92%の精度で無声発話データと有声発話データを判別可能であることを明らかにした．

以上のように，本研究では口唇の動きに着目したコマンド識別システムの構築に関して，種々の基礎的知見を与えることができた．

Akita University

5.3 今後に残された諸問題

最後に今後残された諸問題について述べる．

(1)発話フレーム誤検出の改善

本研究では，口唇の色彩情報および形状の時系列情報に着目した発話フレームの検出を試み，高精度で発話フレームを検出可能であることを示した．しかしながら，誤検出率の改善に関して検討を加えるまでには至っていない．特に，発話終了直後のフレームにおける誤検出，非発話区間における被験者の無意識な微小開口の誤検出の改善は重要な課題である．また，撮影距離や画像分解能が異なる状況下で取得したデータの適用について検討が必要である．

(2)口唇形状分類処理の精度向上および他手法との比較

Ex2-id001～Ex2-id052 の 52 名の被験者を対象とした分類実験において，3 名の被験者が分類不良となった．この要因と推測される顔の上下角変動について検討を加える必要がある．また，蓄積したデータに基づいて形状クラス判別処理の最適な係数を自動設定する手法を開発する必要がある．さらに，ISODATA 法や SOM など複数の手法との比較評価を行う必要がある．

(3)発話慣れなどの経時的な変化に関する調査について

本研究では，同一取得日の発話データを対象として検討を行い，無声発話と有声発話が判別可能であることを示した．判別手法を実際のコマンド識別システムに応用するためには，発話慣れに伴う口唇動作の変化に関する調査を進め，その影響について検討を加える必要がある．

(4)口唇の動き特徴を用いたコマンド識別インタフェースへの適用について

本研究で得られた知見をコマンド識別インタフェースに適用し，システム全体として識別率，誤識別率，再現率などを評価し，さらなる改善について検討を加える必要がある．

Akita University

謝 辞

本研究の遂行並びに本論文の作成にあたって，終始懇切なるご指導とご鞭撻を賜りました秋田大学教授 工学博士 西田　眞 先生，秋田大学教授 博士（工学）景山　陽一 先生に心からお礼申し上げます.

本論文をまとめるにあたり，広い視野から数々の有益なご教示を頂きました秋田大学教授 工学博士 五十嵐 隆治 先生，同教授 博士（工学） 水戸部 一孝 先生，並びに秋田大学理事・副学長 教授 工学博士 玉本 英夫 先生，秋田大学教授 Ph.D. 山村　明弘 先生に深く感謝いたします.

本研究は秋田大学工学資源学部情報工学科西田・景山研究室において行われたものです．本研究の遂行において適切な助言を与えて下さった 博士（工学）石沢 千佳子 先生をはじめ，西田・景山研究室の皆様，卒業生の皆様に心から感謝いたします.

本研究に関して貴重なご助言を頂きました秋田県立大学 博士（工学）石井 雅樹 先生に謝意を表します.

また，大学院博士後期課程への在学について，ご配慮を頂きました秋田大学大学院工学資源学研究科技術部関係各位に厚くお礼を申し上げます.

本研究の一部は，秋田大学平成 23 年度年度計画推進経費，JSPS 科研費 No.24500140，No.24919012 の助成を受けて行われたことを付記し，関係機関各位に厚くお礼申し上げます.

最後に，大学院博士後期課程への入学について理解を示し，在学中の支えとなってくれた家族に心から感謝いたします.

Akita University

本論文の第 2 章は，知能と情報（日本知能情報ファジィ学会誌）掲載論文「髙橋 毅，景山 陽一，西田 眞，若狭 亜希奈，“口唇の色彩情報および形状情報に着目した発話フレーム検出法，”知能と情報（日本知能情報ファジィ学会誌），23 巻, 2 号, 146 頁～156 頁 (2011)」を基に執筆したものであり，本論文の第 3 章は，知能と情報（日本知能情報ファジィ学会誌）掲載論文「髙橋 毅，景山 陽一，西田 眞，“口唇局所領域の形状解析に基づいた顔画像のグループ化手法，”知能と情報（日本知能情報ファジィ学会誌），25 巻, 2 号, 676 頁～689 頁（2013)」を基に執筆したものです.

また，本論文の第 4 章は下記国際会議報告 2 件の内容を基に，さらなる検討を加えて執筆したものです（本論文 91 頁本研究に関連する発表論文, 国際会議(2), (3)参照）.

- T. Takahashi, Y. Kageyama, A. Momose, M. Ishii and M. Nishida, “A Study of the Influence of Vocalization on Lip Motion for Command Input Interfaces,” 2012 International Conference on Fuzzy Theory and Its Applications (iFUZZY 2012), pp.162-167, CD-ROM (2012)
- T. Takahashi, Y. Kageyama, B. Ariuntsengel, A. Momose, M. Ishii and M. Nishida, “Analysis of Lip Motion Due to the Influence of Vocalization,” SICE Annual Conference 2012, TuP01-02, DVD-ROM (2012)

本研究に関連する発表論文

学術論文誌

１　レフェリー制のある学術雑誌

(1) 髙橋 毅，景山 陽一，西田 眞，“口唇局所領域の形状解析に基づいた顔画像のグループ化手法，” 知能と情報（日本知能情報ファジィ学会誌）, 25 巻, 2 号, 676 頁～689 頁（2013）

(2) Y. Kageyama, A. Momose, T. Takahashi, M. Ishii, M. Nishida, A. Mohemmed and N. Kasabov: Analysis of Lip Motion Change Arising due to Amusement Feeling, IEEJ Transactions on Electrical and Electronic Engineering, Vol. 8, No. 5, 538 頁～539 頁

(3) 髙橋 毅，景山 陽一，西田 眞，若狭 亜希奈，“口唇の色彩情報および形状情報に着目した発話フレーム検出法，” 知能と情報（日本知能情報ファジィ学会誌）, 23 巻, 2 号, 146 頁～156 頁 (2011)

２　レフェリー制のない学術雑誌，総説等

(1) M. Ishii, T. Shimodate, Y. Kageyama, T. Takahashi and M. Nishida, “Quantification of Emotions for Facial Expression: Generation of Emotional Feature Space Using Self-Mapping, Developments and Applications of Self-Organizing Maps,” published by In-tech (http://dx.doi.org/10.5772/51136) (2012)

国際会議

(1) S. Sato, T. Takahashi, Y. Kageyama, A. Momose, and M. Nishida, “Method for Identifying Commands Using Lip Motion Features of Utterance,” The 7th Inter. Conf. on Materials Engineering for Resources (ICMR 2013) (accepted)

(2) T. Takahashi, Y. Kageyama, A. Momose, M. Ishii and M. Nishida, “A Study of the Influence of Vocalization on Lip Motion for Command Input Interfaces,” 2012 International Conference on Fuzzy Theory and Its Applications (iFUZZY 2012), pp.162-167, CD-ROM (2012)

(3) T. Takahashi, Y. Kageyama, B. Ariuntsengel, A. Momose, M. Ishii and M. Nishida, “Analysis of Lip Motion Due to the Influence of Vocalization,” SICE Annual Conference 2012, TuP01-02, DVD-ROM (2012)

(4) M. Ishii, T. Shimodate, Y. Kageyama, T. Takahashi and M. Nishida, “ Generation of Emotional Feature Space for Facial Expression Recognition using Self-Mapping,” SICE ANNUAL CONFERENCE 2012 (Akita, Japan), TuP01-08, DVD-ROM (2012)

Akita University

口頭発表

(1) 齋藤 歩, 髙橋 毅, 景山 陽一, 石井 雅樹, 西田 眞, “無声・有声発話における口唇の動き特徴量の経時変化に関する検討,” 平成 25 年度電気関係学会東北支部連合大会講演論文集(USB メモリ), 2E17 (2013)

(2) 佐藤 翔平, 髙橋 毅, 景山 陽一, 百瀬 篤史, 西田 眞, “Generation of Space by Lip Motion Features for Identifying Commands,” 平成25年度電気関係学会東北支部連合大会講演論文集(USB メモリ), 2A04 (2013)

(3) 齋藤 歩, 髙橋 毅, 景山 陽一, 百瀬 篤史, 石井 雅樹, 西田 眞, “発話に伴う口唇の動き特徴における区間分割およびコマンド識別に関する検討(Ⅱ),” 日本素材物性学会平成 25 年度年会, A-22 (2013)

(4) 佐藤 翔平, 髙橋 毅, 景山 陽一, 西田 眞, “撥音の有無に起因する口唇の動き特徴に関する基礎検討,” 平成24年度日本知能情報ファジィ学会東北支部研究会講演論文集,11 頁～13 頁(2013)

(5) 髙橋 毅, 景山 陽一, 西田 眞, “口裂領域に着目した口唇形状特徴抽出および形状分類への応用に関する検討,” 情報処理学会第 75 回全国大会講演論文集(DVD-ROM), 3D-3, 2-55 頁～2-56 頁(2013)

(6) 齋藤 歩, 髙橋 毅, 景山 陽一, 百瀬 篤史, 石井 雅樹, 西田 眞, “発話に伴う口唇の動き特徴における区間分割およびコマンド識別に関する検討,” 情報処理学会第 75 回全国大会講演論文集(DVD-ROM), 3T-5, 2-467 頁～2-468 頁(2013)

(7) 下舘 俊夫, 石井 雅樹, 景山 陽一, 髙橋 毅, 西田 眞, “表情特徴空間の生成における CPN の写像空間に関する検討,” 映像情報メディア学会 2012 年冬季大会講演予稿集(CD-ROM), 12-5 (2012)

(8) 百瀬 篤史, 髙橋 毅, 景山 陽一, 石井 雅樹, 西田 眞, “発話に伴う口唇の動き特徴のばらつきを用いた喜びの感情検出に関する検討,” 映像情報メディア学会 2012 年冬季大会講演予稿集(CD-ROM), 12-4 (2012)

(9) 百瀬 篤史, 髙橋 毅, 景山 陽一, 石井 雅樹, 西田 眞, “発話に伴う口唇の動き特徴を用いた喜びの感情検出,” 平成 24 年度第 1 回情報処理学会東北支部研究会, 講演資料 22 (2012)

(10) 下舘 俊夫, 石井 雅樹, 景山 陽一, 髙橋 毅, 西田 眞, “表情特徴空間を用いた感情の定量的な表現手法に関する検討,” 平成 24 年度第 1 回情報処理学会東北支部研究会, 講演資料 19 (2012)

(11) 髙橋 毅, 景山 陽一, 西田 眞, “口唇局所領域の形状特徴に着目した顔画像のグルーピング,” 平成 24 年度 電気学会 基礎・材料・共通部門大会講演論文集(CD-ROM), Ⅲ-1, 154 頁(2012)

(12) 下舘 俊夫, 石井 雅樹, 景山 陽一, 髙橋 毅, 西田 眞, “顔表情を対象とした感情の定量化手法に関する検討（Ⅱ）,” 平成 24 年度電気関係学会東北支部連合大会講演論文集(CD-ROM), 2D14 (2012)

(13) 百瀬 篤史, 髙橋 毅, 景山 陽一, 石井 雅樹, 西田 眞, “口唇の動き特徴を

Akita University

用いた喜びの感情検出に関する検討,” 平成 24 年度電気関係学会東北支部連合大会講演論文集(CD-ROM), 2D13 (2012)
(14) 百瀬 篤史, 髙橋 毅, 景山 陽一, 石井 雅樹, 西田 眞, “口唇の動き特徴におけるばらつきに着目した喜びの感情検出に関する検討(2),” 平成 23 年度日本知能情報ファジィ学会東北支部研究会講演論文集, 5 頁～8 頁 (2011)
(15) 下舘 俊夫, 石井 雅樹, 景山 陽一, 髙橋 毅, 西田 眞, “顔表情を対象とした感情の定量化手法に関する検討,” 平成 23 年度日本知能情報ファジィ学会東北支部研究会講演論文集, 1 頁～4 頁 (2011)
(16) 髙橋 毅, 景山 陽一, 西田 眞, “口裂抽出を目的とした口唇領域の色情報解析,” 第 16 回日本顔学会大会(フォーラム顔学 2011), P1-07 (2011)
(17) 百瀬 篤史, 髙橋 毅, 景山 陽一, 石井 雅樹, 西田 眞, “口唇の動き特徴におけるばらつきに着目した喜びの感情検出に関する検討,” FIT2011 第 10 回情報科学技術フォーラム講演論文集(DVD-ROM), J-045, 第 3 分冊 639 頁～640 頁 (2011)
(18) 髙橋 毅, 景山 陽一, 西田 眞, “口唇の局所領域形状に着目した個人識別のための口唇形状グループ化法,” FIT2011 第 10 回情報科学技術フォーラム講演論文集(DVD-ROM), J-044, 第 3 分冊 637 頁～638 頁 (2011)
(19) 下舘 俊夫, 石井 雅樹, 景山 陽一, 髙橋 毅, 西田 眞, “自然表情の取得を目的とした情動と心拍の関連に関する検討,” FIT2011 第 10 回情報科学技術フォーラム講演論文集(DVD-ROM), J-035, 第 3 分冊 615 頁～616 頁 (2011)
(20) 下舘 俊夫, 石井 雅樹, 景山 陽一, 髙橋 毅, 西田 眞, “表情表出プロセスによる表情認識を目的とした情動と心拍の関連に関する検討,” 日本素材物性学会 平成 23 年度(第 21 回)年会講演要旨集, A-8, 15 頁～16 頁 (2011)
(21) 百瀬 篤史, 髙橋 毅, 景山 陽一, 石井 雅樹, 西田 眞, “喜びの感情喚起時における口唇の動き特徴に関する検討,” 日本素材物性学会 平成 23 年度(第 21 回)年会講演要旨集, A-7 , 13 頁～14 頁(2011)
(22) 髙橋 毅, 景山 陽一, 西田 眞, “口唇の局所領域における特徴解析と個人識別のためのグループ化,” 平成22年度第2回情報処理学会東北支部研究会, 資料番号 4 (2010)
(23) 髙橋 毅, 景山 陽一, 西田 眞, 若狭 亜希奈, “口唇の色彩情報および形状情報に着目した発話フレーム検出法（Ⅱ）,” 平成 22 年度電気関係学会東北支部連合大会講演論文集, 2D16, 132 頁 (2010)
(24) 髙橋 毅, 景山 陽一, 西田 眞, 若狭 亜希奈, “口唇の色彩情報および形状情報に着目した発話フレーム検出法（Ⅰ）,” 日本知能情報ファジィ学会支部研究部会合同研究会 2010-東北・関東支部＆知的制御・ヒューマンインターフェース研究部会ー, 資料番号 6 (2010)