Vol.2010-EC-17 No.1
2010/8/23

くちびるエンタテインメント

竹山峻平[†]　佐藤俊樹[†]　野嶋琢也[†]

人はその口唇形状を，発話や食事，感情表現などのために自らの意志に基づいて変化させることが可能である．近年，この口唇形状を利用した読唇システムなどが研究されている．しかしながら，多くの研究は「発話」という機能に重点を置いており，その他の口の持つ機能に注目しているものは少ない．そこで本研究では口唇形状認識インタフェースの異なる可能性を検討するため，口唇の形状・動きをインタフェースとしたエンタテインメントシステムを提案する．

Entertainment with the lip

Shumpei Takeyama[†]　, Toshiki Sato[†]
and Takuya Nojima[†]

People often change their shape of lips especially when they speak. Then, many researchers focus on recognizing the shape of the lips for “lip reading”. However, the shape of lips are also changed when people eat, and when they express their emotion. Furthermore, lips are the part of human body that have high degree of freedom of motion. Then, in this research, we propose to use the shape of lips as human machine interface. In this report, we describe on the entertainment system that utilize shape of lips.

1. はじめに

人はその口唇形状を，発話や食事，感情表現などのために自らの意志に基づいて変化させることが可能である．近年の画像処理技術などの向上に伴い，この口唇形状を認識し，インタフェースとして利用する研究が多数行われてきた．特に人間が発話する際には，その口唇形状は発話内容に応じて特徴的な変化を起こすことが知られており，この特徴を利用して口唇形状に基づいた発話認識は古くから多数の研究が行われてきた[1]．また，発話時の口唇形状情報には，発話に関する情報のみならず個人に特徴的な要素も含まれており，これを利用した個人認証の研究なども行われている[2]．

しかしながら，口は発話のみならず食事や感情表現など，多様な場面で広く活用される身体部位である．そしてその形状は，高い自由度で，自らの意志に基づいて自由に変形させることができるという特長を有している．そこで本研究では，くちびるの形状変化，ならびに頭部全体を含むその動きに着目した，エンタテインメントシステムの開発を目指す．

2. システム提案

2.1 関連研究

口唇形状の認識は，主に読唇や個人認証，福祉などの分野で研究が行われてきた．例えば斎藤らは，口唇形状の時間変化を軌道として表現することで単語読唇を行うトラジェクトリ特徴量を提案している[3]．トラジェクトリ特徴量は，単語発話中の各時点の口唇形状の面積とアスペクト比の二つの特徴量を平面上に投射し，投射された各点を結ぶことでその時間変化を軌道として表現したものである．単語によってこの軌道は異なるため，求めたトラジェクトリ特徴量に対して二次元 DP マッチングを適用することで単語認識が可能となる．この手法を利用することで，音声認識が困難な雑音の多い環境であっても読唇による発話認識が可能となる．また加藤らは，口唇形状のリアルタイム認識による項目選択式の意思伝達システムを提案している[4]．加藤らのシステムでは母音発声時の口唇形状を 5 つ以下の項目に対応させており，選択したい項目に対応する母音発声時の口唇形状を取ることで，入力が可能となっている．一方 Lyons らは，口唇形状とキー操作を併用してテキスト入力を行うシステムを提案している[5]．このシステムにおける日本語入力では，キー操作により子音を指示し，口唇形状により母音を指示することでテキスト入力を行う．携帯電話などの一般的な携帯機に搭載されるキーボードは非常に小型であるため，通常は PC のキーボードのようなスムーズなテキスト入力は難しい．しかしこのシステムでは指と口の動作を併用

[†] 電気通信大学大学院情報システム学研究科
Graduate School of Information Systems,
The University of Electro-Communications

することで，ローマ字入力に必要な子音と母音を同時に選択することが可能となっている．さらに母音のキーを選択する際には，対象となる母音を発音する場合と同じ口唇形状をすれば十分であることから，直感的でスムーズな文字入力が可能となっている．

一方エンタテインメント分野においては，任天堂社の Wii[6]のような，身体運動をインタフェースとして採用する動きが近年活発になっている．身体運動をインタフェースとして利用することで，直感的で理解しやすいゲーム操作が可能になる，適切な運動を誘発することが可能になるといった利点が挙げられる．例えば佐藤らは，テーブルトップ上での多人数エンタテインメントシステム“PAC-PAC”を開発した[7]．このシステムでは，天井に設置された高速度カメラによりテーブルトップ上にかざされた手を撮影し，その位置および形状を認識する．この際，テーブル上の手でつまむ動作（親指と人差し指で輪を作る動作）を繰り返すことで弾を発射することが出来る．なお，手の位置と画面上の弾が発射される位置は対応している．複数人での認識も可能であり，互いにこの弾を撃ちあう等して遊ぶことが出来る．水上らは，視線情報を利用したリアルタイムゲームシステムを開発した[8]．視線計測装置をインタフェースとして利用することで，日常生活で誰もが行っている「見る」という行為をそのまま入力としたエンタテインメントを考案している．

2.2 口唇形状を利用したインタフェース

本研究では，くちびるの形状変化，ならびに頭部全体を含むその動きに着目した，エンタテインメントシステムの開発を目指す．口唇形状は高い自由度で変形させることが可能であるが，今回はゲームとしての単純さを追求して，開閉動作に着目したシステムを構築した．本システムにおける特徴は以下の 2 点となる．

- 口唇の位置をポインタとする
- 口唇の動き(開閉)によるコマンド入力

本システムでは口唇を画像で認識し，カメラ画像内における口唇位置をポインタとして利用している．口唇は人体頭部にしか存在しないことから，これをポインタとして採用した場合，ポインティングのために全身運動が要求されることとなる．例えば画面の右上隅にポインティングするためには，身体そのものを右上隅へ移動させなければならない．これにより，自然な形でユーザの全身運動を誘発することが可能になると考えられる．また，今回試作したシステムは，特に口の開閉に着目したものとなっている．入力可能な動作を口の開閉のみに限定することによって，直感的に操作可能な，シンプルなエンタテインメントシステムとして構成することが可能であると考えられる．

次章ではこの考えに基づいて構成された試作システム，“パクつくライフ”について，その構成を述べる

3. 試作システム

3.1 口唇形状を利用したゲーム：「パクつくライフ」

今回我々は，食事の際の口唇動作，すなわち開ける・閉じるの二つの動作に注目したゲーム“パクつくライフ”を制作した．

図 1 にゲーム画面を示す．“パクつくライフ”は，金魚が餌を食べる行動をモチーフとしたゲームである．図 1 画面上にはプレーヤの位置を示す口の映像と，色分けされた餌が表示されている．プレーヤは金魚の立場になり，連続的に投下される餌を食べるゲームとなっている．プレーヤは身体運動により口唇位置を移動させることで，画面上に表示された口の画像の位置を制御する．そして口の画像を餌に重ねた状態で口を開閉させることで，画面内に表示された餌を「食べる」ことが可能となる．本ゲームでは，所定の制限時間内に「食べる」ことのできた餌の数や種類に応じた得点が得られるようになっており，取得できた点数を競うゲームとなっている．また，ゲーム性を高めるために，以下のようなルールを設定した．

- 同時に複数の餌を食べると獲得できる得点が増加する
- 餌の種類は 3 色存在しており，赤＞黄＞緑の順に得点が高く，出現確率が低い
- サイズが大きいほど得点が高い

そのため，高得点を獲得するためには得点の高い餌を優先して，可能ならばまとめて食べる必要がある．しかし，餌は投下されてから一定時間で沈んでしまうため，食べる順番を適切に判断することが高得点を獲得する鍵となる．

図 1 ゲーム画面

3.1 システム概要

図 2 に試作したシステムの概要図を示す．本システムでは，被験者の顔を正面に設置したカメラで撮影し，被験者の顔周辺の画像を取得する．その後，取得した画像に対して画像処理を施し，口唇形状およびその位置を取得する．そして，取得した口唇形状情報および位置情報をゲームの入力として利用する．

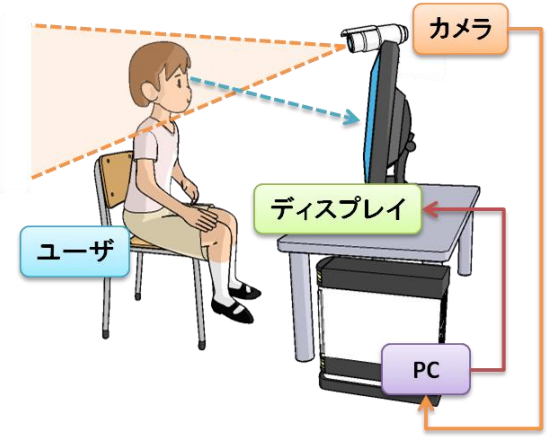

図 2　システム概要図

3.2 処理の流れ

図 3 にシステムの処理の大まかな流れを示す．まず，ゲームの開始前に初期設定を行い，その後ゲーム本編を開始する．ゲーム中はカメラが撮影した画像に対して口唇検出処理を行い，被験者の口唇位置を取得する．更に口唇形状判定処理により被験者の口が開いた状態か，閉じた状態かの判定を行う．そうして得られた口唇位置と口唇形状をゲーム処理部分に入力として与え，ゲーム処理を行い，その結果をディスプレイに出力する．これら一連の処理をフレーム毎に繰り返すことで，ゲームは進行していく．なお，開発は Visual Studio 2008（C++）にて行い，OpenCV[9]および SDL[10]を利用した．

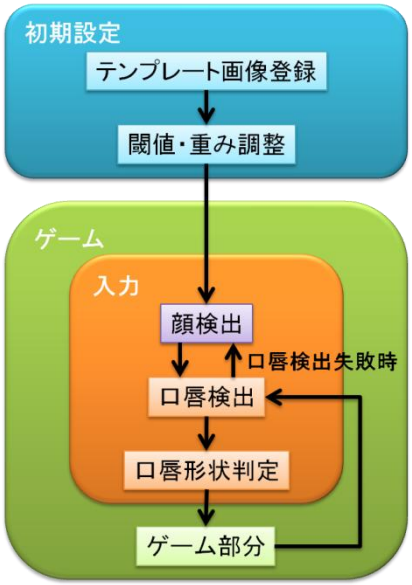

図 3　処理の流れ

3.1 初期設定

この試作システムでは，画像情報により口唇形状認識を行うため，環境の変化や被験者間の個人差により，その精度に大きな影響を受ける恐れがある．そこで，実際のゲームを開始する前に初期設定を行うことで，精度の低下を抑制している．初期設定は主として，テンプレート画像の登録，二値化閾値および口唇形状判定の重み調整の二点である．

3.1.1 テンプレート画像登録

今回は口唇検出にテンプレートマッチングを利用しているため，あらかじめテンプレート画像を用意する必要がある．そこで，初期設定時にテンプレート画像の設定を行う． 図 4 にテンプレート画像設定時の画面を示す．テンプレート画像は，口を開いた状態と閉じた状態の 2 通りを設定する．それぞれの状態で口唇周辺の領域を選択することでテンプレート画像の設定は完了する．なお，簡便のためにあらかじめデフォルトのテンプレート画像も用意してあり，この操作は必須ではない．ただし省略した場合は，環境や個人差の影響により，口唇の認識精度が低下する可能性がある．

Vol.2010-EC-17 No.1
2010/8/23

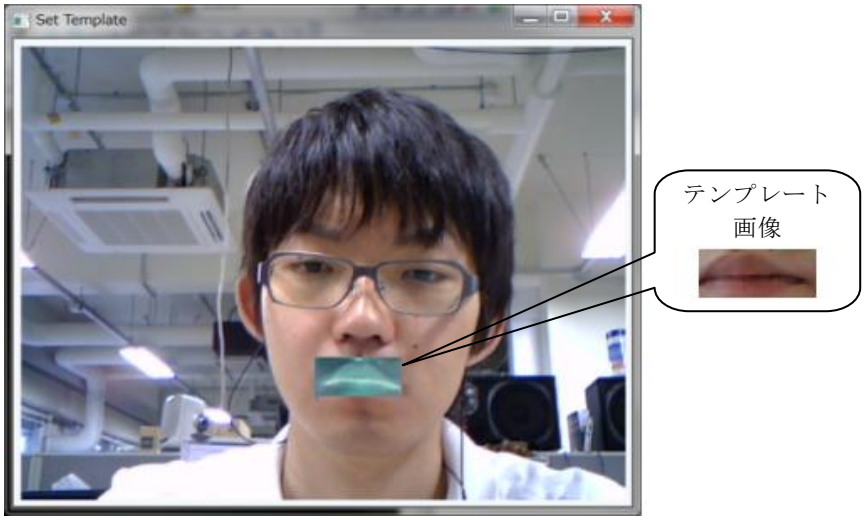

図 4　テンプレート画像設定画面（左）および取得したテンプレート画像（右）

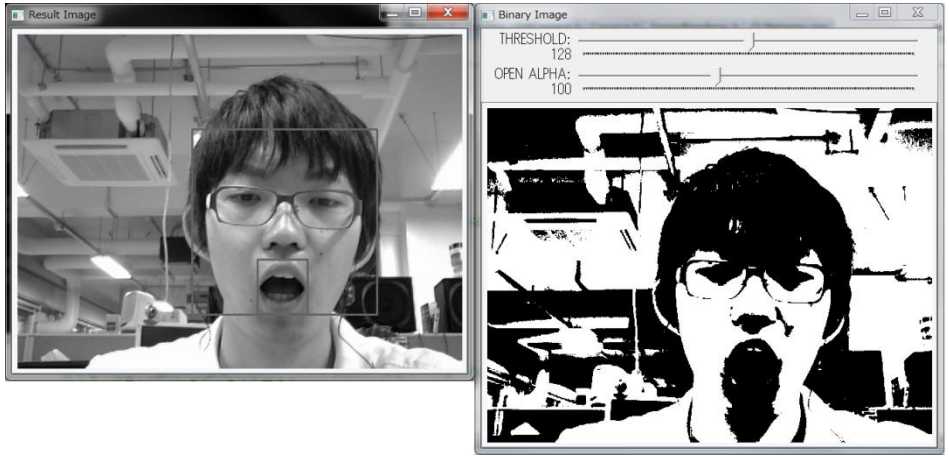

図 5　二値化閾値と口唇形状判定の重み調整画面
（右ウィンドウ上部のスライダによって調整を行う）

3.1.1 閾値・重み調整

テンプレートマッチングは二値化画像において処理が行われるため，その二値化閾値によって精度が左右される．そこで，その環境下において検出精度が最適となる閾値にて二値化処理を行うため，初期設定時に調整出来るようにしている．また，口を開いた状態と比較して口を閉じた状態ではテンプレートマッチングにより算出される類似度が高くなる傾向がある．口を閉じた状態は比較的一様であるのに対し，口を開いた状態はひらき具合が一定ではないことがその原因として考えられる．この試作システムでは各画像についてテンプレートマッチングを行い，算出された類似度を比較することで口唇形状の判定を行っている．そのため，二つの状態間で検出成功時の類似度の差が大きいと口唇形状の正常な判定が困難となる可能性がある．そこで，この試作システムでは口唇形状判定時に口を開いた画像との類似度に重み付けを行うことで，判定精度を向上させている．最適となる重みは被験者ごとに異なるため，この重みも初期設定時に調整できるようにしている．図 5 に二値化閾値と重みの調整画面を示す．

3.2　顔検出

本ゲームではプレーヤの操作入力として，口唇形状および口唇位置を利用している．そのためまず，画面内における口唇位置を検出する必要がある．しかしカメラで取得した画像には，プレーヤとなる被験者以外にも背景などの異物が含まれている．この状態の画像に対してそのまま口唇検出を施した場合，誤検出等の問題が生じる恐れがあることから，あらかじめ顔検出を行うことで，口唇検出処理を適用する領域を限定することで誤検出の可能性を低減した．

顔検出には，Haar-like 特徴を利用したオブジェクト検出手法を用いている[11]．このオブジェクト検出手法では，あらかじめ Haar-like 特徴と呼ばれる白黒パターンの位置と組み合わせによって定義された検出窓を用意しておく．特定領域内の画像の明度パターンと検出窓のパターンが類似している場合，その領域は高い特徴量を持つとする．この検出窓により画像中をスキャンすることで，オブジェクトの検出を行う．実際の検出に際しては，OpenCV 標準の分類器カスケードを利用した．図 6 に顔領域と口唇検出対象領域を示す．

なお，実際にゲームを行う環境，すなわち立位か着座か，画面の大きさ，カメラ画角，カメラとプレーヤの距離などの条件によって，プレーヤの頭部運動量は大きく変化すると考えられる．例えばはカメラ画像のほとんど全ての領域をプレーヤの顔が占めるという状況の場合，毎フレーム顔検出を行う必要性は低くなる．このような場合には計算量低減のため，ゲーム開始時のみ顔検出を行い，検出された顔領域の下半分を口唇検出の対象領域として定義するといった手法が考えられる．

図 6　顔領域と口唇検出対象領域

3.3　口唇検出

本試作システムでは簡単のため，テンプレートマッチングを利用した以下の手法により，口唇位置ならびに状態の検出を行った．まず，口唇検出対象領域において，閉じた状態の口唇テンプレート画像を用いてテンプレートマッチングを行う．更に，開いた状態の口唇テンプレート画像でも同様にテンプレートマッチングを行う．その結果，各テンプレート画像との最大類似度およびその座標を判定可能となる．続いて，口を開いた画像との最大類似度に重み付けを行った上でそれぞれの最大類似度を比較し，どちらの口唇形状の状態であるかを判定する．これらの処理を通じて口唇形状およびその座標を取得し，ゲーム状態に反映させている．なお，最大類似度がどちらとも一定値以下であった場合，口唇領域を見失ったと判定して，顔検出処理を再度実行するようになっている．

3.4　口唇形状判定

前述のテンプレートマッチングを利用した口唇検出により，口唇形状および座標を取得することが出来る．しかしながらこの口唇検出処理は特定の 1 フレームのみを対象としており，誤検出が生じる恐れがある．そこで本試作システムでは，直近の数フレームと比較することでその検出精度を高めている．

たとえば，口唇検出処理によって口が閉じた状態と判定された場合であっても，すぐには口が閉じた状態と確定せず，直前の 2 フレームとの比較を行う．現在のフレームを含めた連続する 3 フレームにおいて，口が閉じた状態と判定され続けていた場合のみ，口が閉じた状態であると確定する（図 7）．比較するフレーム数が増えると，その分口唇形状の変化がゲームに反映されるまでのタイムラグが大きくなるため，今回は直近 3 フレームの比較とした．また，口唇の座標が直前のフレームにおける座標と比較して急激に変化をしていた場合，誤検出であると判定して検出結果を無効としている．

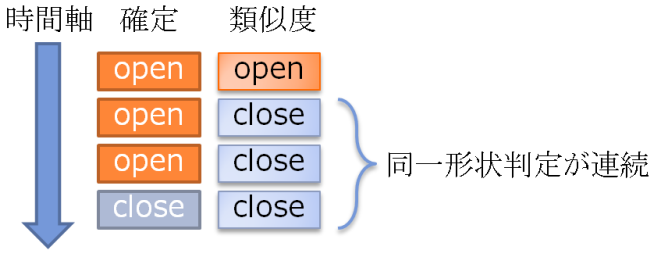

図 7　口唇形状の確定処理

3.5　ゲーム部分

カメラの画像を元に口唇形状および位置を取得すると，それらの情報をゲーム部分に入力として引き渡す．なお，口唇位置は口唇領域の上辺中央の座標である．人間が口を開く際には下顎が大きく動くため，口唇領域は下に引き伸ばされる形となる．そのため，口唇領域の中心あるいは下部の座標とした場合，口を開閉するだけで大きく座標が変化してしまう．また，口の開閉に伴い口唇領域の幅が変化するが，中心線は大きく動くことはない．以上の理由により上辺中央の座標を採用している．

ゲーム部分では，入力された口唇情報に応じて処理を行っていく．ゲーム開始時の口唇位置をゲームフィールドの中心座標として，その初期値と入力された口唇位置の差分によりゲームフィールドにおける座標を決定する．この座標と口唇形状を元に金魚の口を描画する．口唇形状が開いた状態から閉じた状態に遷移した際に，金魚の口領域内に存在していた餌を「食べた」と判定し，得点を加算する．

4. 考察

本システムでは，口の位置ならびに開閉動作がそのままゲームの操作入力となっており，直感的で理解しやすいゲームを実現することができた．しかし現状のシステムは簡易的な構成のため，ゲーム画面も小さく，プレーヤは画面の前に着座した状態からほとんど動くことができない．そのため画面周辺部の餌を食べにいく場合，上半身を動かすのではなく，頭部の向きを変えるだけで食べようとしてしまう場合があることが判明した．一方，今回構築したシステムでは口唇形状の認識にテンプレートマッチングを利用しているため，テンプレートと同じ状態，すなわちカメラに対して正対した状態からずれてしまうと，認識精度が極端に低下する問題がある．そのため，特に画面周辺部の餌を食べに行く際に認識精度が落ちる傾向が見受けられた．また，現状では認識速度も十分であるとは言えないことから，認識速度改善，ならびに認識の安定性向上が必要であると考えられる．

5. おわりに

本研究では，くちびるの形状変化，ならびに頭部全体を含むその動きに着目した，エンタテインメントの提案を行った．具体的には口唇の位置ならびに開閉動作の認識が可能なシステムを構築し，画面上に投下された餌を自らの口の動作をつかって食べるゲーム，“パクつくライフ”を試作した．

今後は認識速度の向上を目指すとともに，よりダイナミックな動きが可能なシステムの構築を目指す．現状のシステムでは画面が小さく，プレーヤの動作が制限されてしまうという問題が指摘されている．プレーヤの動きは大きくある方が望ましいと考えられることから，ゲーム画面の拡大などにより，プレーヤが自由に動けるようなシステムの構築を目指す，

参考文献

1) 石井雅樹, 佐藤和人, 西田 眞, 景山陽一: 時系列口唇画像を用いた読唇のための特徴抽出と唇の動き解析, 電子情報通信学会論文誌 D, vol.119, no.4, pp.465-472, 1999.
2) 白澤洋一, 三浦 信, 西田 眞, 景山陽一, 栗栖怜史: 口唇の動き特徴を用いた個人識別に関する検討, 映像情報学会誌, vol.60, no.12, pp.1964-1970, 2006.
3) 斎藤剛史, 小西亮介: トラジェクトリ特徴量に基づく単語読唇, 電子情報通信学会論文誌. D, 情報・システム, J90-D(4), pp.1105-1114, 2007.
4) 加藤友哉, 斎藤剛史, 小西亮介: リアルタイム口部形状認識を利用した意思伝達システム, 電子情報通信学会技術研究報告. TL, 思考と言語, 107(433), pp.99-104, 2008.
5) Lyons, M.J., Chi-Ho, C., and Nobuji, T.: Mouthtype: Text Entry by Hand and Mouth, Proc. Conference on Human Factors in Computing Systems, pp.1383-1386, 2004.
6) Wii, http://www.nintendo.co.jp/wii/
7) Toshiki, S., Haruko, M., Hideki, K., and Kentaro, F.: An Augmented Tabletop Video Game With Pinching Gesture Recognition, ACM SIGGRAPH ASIA 2008 artgallery: emerging technologies, pp.38-38, 2008.
8) 水上 明, 伊藤毅志: 視線を用いた新しいエンターテインメント. 情報処理学会研究報告. EC, エンタテンメントコンピューティング, 2008(26), pp.23-28, 2008.
9) Open Source Computer Vision Library, http://sourceforge.net/projects/opencvlibrary/
10) Simple DirectMedia Layer, http://www.libsdl.org/
11) 奈良先端科学技術大学院大学: OpenCV プログラミングブック, 毎日コミュニケーションズ (2007).