母音形状を用いたモバイル端末向けコマンド入力手法

澤田　佳樹[1,a)]　高橋　伸[2]　田中　二郎[2]

概要：本研究では発話時の口唇形状の情報に着目し, 携帯端末の操作を行う手法を提案する. 携帯端末のフロントカメラを用いて, 口唇形状を認識する. そして発話した際と同様の口唇の形状から, 母音を推定することによりコマンド入力を行い, アプリケーションの操作などの操作を行うことができる. 本稿ではモバイル端末において, 口唇形状の情報から機械読唇の技術を用いて母音を推定する方法の検討を行う. ユーザは, 実際に発話することなく口語的に携帯端末にコマンドを入力することができる. そのため, 音声入力のように直感的な入力が可能であり, 入力の際にユーザは各操作ごとのジェスチャを覚える負担を軽減することができる.

An Intaraction Technique with Vowel Input Using Lip Shape from Mobile Camera

Yoshiki Sawada[1,a)]　Shin Takahashi[2]　Jiro Tanaka[2]

Abstract: We propose an input method to operate mobile devices by using users' lips. By processing live image from the device's front camera, it recognizes a shape of the lip and estimate a represented vowel. Estimated vowels can be used as commands to manipulate applications running on the devices. The commands are combinations of vowels of names of these operations such as the names of the applications to launch. This method allows users to input intuitively and does not require users to remember gestures for each commands. In this paper, we describe the method to recognize vowels from image of image including the lip in detail and also designs of the prototype system that supports a shortcut to launch applications with the proposed method.

1. はじめに

近年では携帯端末の高機能化が進み, 多様なアプリケーションを操作することが可能である. それに伴い, 音楽プレイヤーや, ウェブブラウジングなど, 複数のアプリケーションを並行して利用する場面が増加している. 現在の携帯端末では, そういったアプリケーションを起動するなどといった操作は, タッチパネルで行うことが主流となっている. 各アプリケーション毎のアイコンをホーム画面に配置しておき, そのアイコンをタップすることによってアプリケーションを起動するのが一般的な操作となっている. しかし, 例えばブラウザを開いている際にメールを起動するといった場面では, ブラウザ画面から, 一度ホーム画面に戻り, メールアイコンをタップするといった手間が生じることが考えられる. また, 携帯端末の大画面化に伴い, タッチパネルのすべての範囲に片手でタップ操作を行うことは困難になりつつある. そういった問題を解決するには, 従来の操作手法を拡張する必要がある.

そういった手間を必要としない入力拡張手法としては,iPhone の Siri[2] のような音声入力による操作といったことが考えられる. 音声により, アプリケーション名をそのまま入力に用いることができるので, より直感的に画面上にアイコンの存在しないアプリケーションの切り替えを行うことができる. しかし, 音声入力は周囲の雑音に影響を受けやすい技術であり, また発声を行った際の, 周囲からの視線など, 公共の場における抵抗感 [3] などが指摘されてい

[1] 筑波大学大学院システム情報工学研究科コンピュータサイエンス専攻
Department of Computer Science, Graduate School of Systems and Information Engineering, University of Tsukuba
[2] 筑波大学システム情報系
Faculty of Engineering, Information and Systems, University of Tsukuba
[a)] sawada@iplab.cs.tsukuba.ac.jp

る．そういった, 音声入力を補完する技術として, 口唇から情報を読み取り, 音声推定を行う研究は長く行われてきた．これは機械読唇と呼ばれる技術で, 周囲の雑音に影響されることなく音声に関する情報を取得することができる．実際に発声することなく, 情報を取得できるため, 発声が困難な環境においても発話内容の推定に有効であるといわれている．

そこで, 本研究ではそういった口唇形状の情報に着目し, 携帯端末の操作を行う手法を提案する．提案手法においては, 携帯端末のフロントカメラを用いて, 口唇形状を認識する．そして発話した際と同様の口唇の形状から, 母音を推定することによりコマンド入力を行い, アプリケーションの選択などの操作を行う．これにより, ユーザは発話した際と同様の直感的な入力を行うことができると同時に, 発声の必要がないことから, 公共の場における抵抗感などを軽減した操作を行うことができる．また, コマンド名をそのまま発話することから, 各操作ごとのジェスチャをユーザは覚える必要がない．また, 他の入力, 例えばタッチパネルにタップすることなく, アプリケーションの切り替えなどのショートカット操作を行うことが可能になる．

2. 関連研究

機械読唇を用いた携帯端末の文字入力手法としてLYONSら [4] は, 口唇の形状から母音を推定し, その母音と子音をキーで指定することによって日本語を入力する手法を提案している．従来の携帯端末の文字入力の母音入力部分を口唇形状を利用して行う．それにより, 指の操作の負担を軽減することができる手法である．

携帯端末の操作といったところに焦点を当てた関連研究として,LUI[5] という研究がある．これは, 口唇の形そのものを操作に割り当てることにより操作を行う手法である．例えば, マップのアプリケーションを操作する際には, 口を開くといったジェスチャを拡大, 口を閉じるといった操作を縮小といったように, 口唇の形そのものをジェスチャとして扱う．しかし, ユーザはそういったジェスチャを記憶する必要があるといった問題点が挙げられる．

公共の場におけるタッチパネルによる操作にユーザは抵抗感を感じないことが考えられる．そんなタッチパネルを用いた携帯端末のショートカット手法としては, タッチジェスチャのような手法が考えられる．そんなタッチジェスチャに関して,Poppinga ら [1] によると, ユーザはアプリケーション名の頭文字をジェスチャとして設定する傾向がある．そうした場合, ユーザは頭文字の同じアプリケーションには, 別のジェスチャを割り当てる必要があり, さらにそのジェスチャを覚える必要がある．

これらの研究に対して本研究では, 口唇形状から, 母音の推定を行い, それを携帯端末の操作に割り当てる．発話時と同様の口唇の形状を読み取るため, ユーザは各操作ごとのジェスチャを記憶する必要がなく, 指の操作を必要としない入力が可能になる．

3. 口唇形状から推定した母音列によるコマンド入力

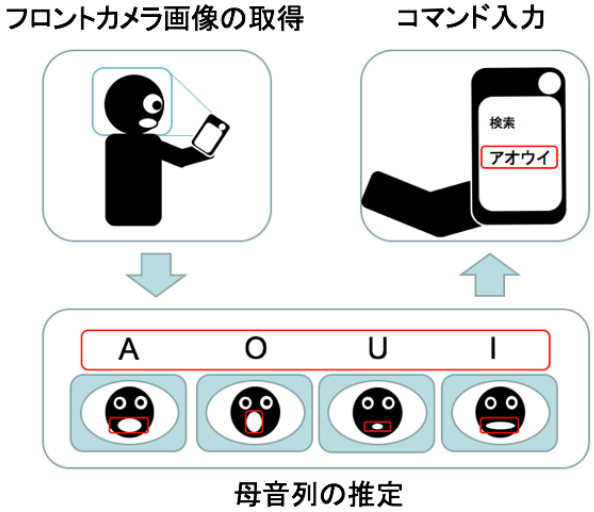

図 1 提案手法：概要

提案手法における概要を図 1 に示す．本手法では, 携帯端末のフロントカメラを用いて, ユーザが発話した際の口唇形状を認識する．その形状変化から母音列を推定し, それをコマンドとして入力に用いることで端末を操作する手法である．

3.1 口唇形状を用いた母音列の推定

本手法では, 母音を発話した際の口唇形状を携帯端末のフロントカメラを用いて認識を行う．そして, 口唇形状から母音を認識した際の表現として, 口形コード法（宮崎ら [6]）という既存手法を用いる．これは, 単語発声時の口の形を認識し,「あ, い, う, え, お (a,i,u,e,o)」, それに加えて口を閉じた「ん (x)」の６つのクラスに分類し, 単語を推測する手法である．口形コード法では, 日本語五十音を発声時の初口形 (i,u,x) と終口形 (A,I,U,E,O,X) を認識してそれを合わせて口形コードを生成する．例えば,「ま」は発声した際の初口形「x」と発声し終わった際の終口形「A」により「xA」という口形コードで表すことができる．しかし, 単語を発声した際の口唇の変化というのは, 口形コードのみでは表現することができない．

口形コード法では, そういった口唇の変化を口形変化コードを用いて表現する．口形変化コードは, 口形コードを連結して生成されるものである．例えば,「あかり」という単語に対して, 口形コードは「A,A,I」となる．だが,「あかり」という単語の口唇の動きは,「あ」と「か」の発声の間には変化がなく, 実際の口唇の動きは「A,I」となる．このように口形コード法では, 実際の口唇の動きを口形コードで読

み取り, それを連結規則に則り生成される口形変化コードを用いることで, 発声内容を表現することができる. 本手法では, このように口唇の形状変化を認識し, 取得された口形変化コードを母音列と呼ぶことする. 例えば, 図 2 に示されるように,「ブラウザ」というコマンドに対しては, 発声した際の口唇形状の変化を認識し, 口形コード法に従い,「xU,iA,U,iA」といった母音列の推定を行う. こうして推定された母音列を携帯端末への入力とし, その母音列からコマンドの推定を行う.

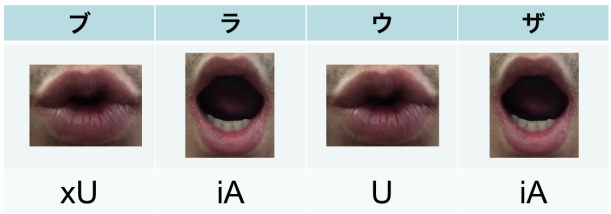

図 2 「ブラウザ」発声時の母音列

3.2 母音列から推定されたコマンドによる携帯端末操作

本手法では, 母音列をコマンドとして入力する際には, ユーザは発声を実際にする必要がなく, コマンド名を発声するかのように口唇を動かすことで入力が可能であり, コマンド入力に必要な口唇のジェスチャをユーザは記憶する必要がない. 携帯端末における操作名を口形変化コードに変換し, そのコードと認識された母音列を対応させることにより, 入力と端末の操作を関連付ける. つまり, 操作名から生成された口形変化コードと, 口唇形状より認識された母音列（口形変化コードで表現されたもの）の認証を行うことによって, 操作を行う. 「ブラウザ」コマンドを例にとると, アプリケーション名から生成された母音列をブラウザの起動といった操作のコマンドに割り当てておくことにより, ユーザは,「ブラウザ」というアプリケーション名を発話するように口唇を動かすことによって, 端末の状態に関わらずショートカット操作を行うことができる (図 3). このように, 操作の名前をそのまま発話的に扱い, コマンド入力とすることで, 操作内容の指定が重複することもなく, ユーザは操作を行うことが可能になる.

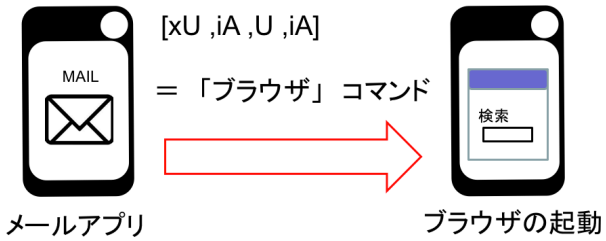

図 3 ショートカット操作

3.3 利用シーン

本手法を利用する例としては, アプリケーションの切り替えを行う際のショートカット操作といった利用を考えている. 電車のつり革につかまっている際, ユーザはスマートフォンを片手でしか扱うことができない. そういった場面において, タッチパネルをタッチしてアプリケーションの起動を行うには, アイコン位置まで指が届かないなど, いくつかの障害が発生する. そういった際に本手法を利用することで, 音声入力のように, 発話的にアプリケーションの切り替えを行うことができる. また, ユーザは実際にコマンド名を発話する必要がないので, 周囲からの視線などの抵抗感を感じることなく操作を行うことができる. このように, つり革につかまっている状態でもユーザは直感的に携帯端末を操作することができ, 端末の状態に関わらずアプリケーションの切り替えを行うことができる.

4. 母音列によるコマンド入力システム：設計

本研究では, 提案手法を用いた母音列によるコマンド入力から, アプリケーションの切り替えといったショートカット操作を行う機能の実装を行う. 現在, 口唇形状から母音列を認識する手法の検討を行い, プロトタイプを作成中である. 開発言語は,Objective-C,OpenCV for iOS で行い,iOS アプリケーションとして iphon5s にて実装を行う.

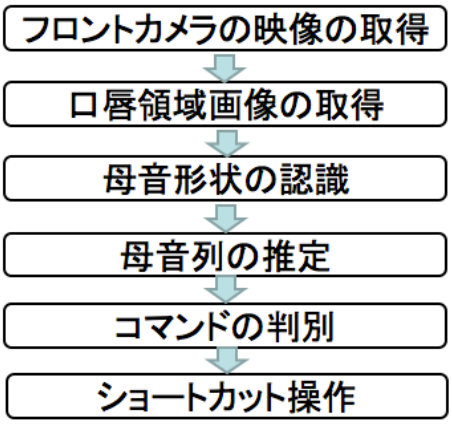

図 4 処理の流れ

プロトタイプの処理の流れを図 4 で示す. 端末のフロントカメラにて映像の取得を行い, その映像から OpenCV を用いて口唇領域の抽出を行う. その口唇領域から, 母音形状, つまり母音を発声した際の口唇の形の認識を行い, 母音列の推定を行う. 最終的に, その推定された母音列から, どのコマンドが指定されたか判別を行い, ショートカット操作を行う.

4.1 口唇領域の抽出

取得した口唇領域を表示した様子を図 5 に示す. フロントカメラ映像より得たカラー画像を HSV 変換を行い, 口唇

の色領域を指定することで抽出を行う．色抽出された画像をグレースケール化し，２値化したのち，輪郭点の取得を行う．輪郭点から，直線近似を行うことによって口唇領域の矩形とその面積 (図 5 青枠) を取得したのち，その輪郭点より矩形のアスペクト比（図 5 赤，緑線）の計算を行う．

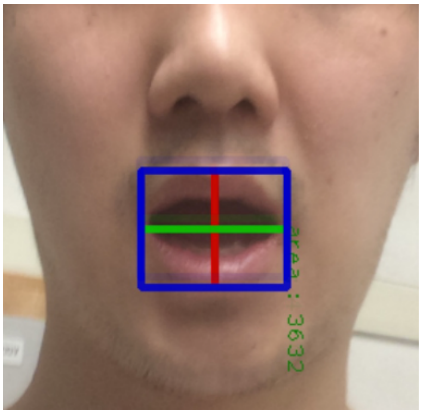

図 5 口唇領域の抽出

4.2 母音形状の認識

母音形状の認識にはフロントカメラにて取得した映像から抽出された口唇領域画像の面積とアスペクト比を用いて行う [4]．これは，母音を発声する際の人間の口唇の動きが横の長さの変化はほとんどなく，縦の長さの変化が大きいといった特徴がもとになっている手法である．判別の為に，閉口形（ん）の口唇領域の面積とアスペクト比を算出し，正規化を行うことによって基準を作る．それに伴って，変化した口唇領域の面積，アスペクト比の変化量を算出し，母音形状を推定することができる．プロトタイプでは，図 5 に示した変化量の特徴をもって，口唇領域より取得した面積，アスペクト比の二つの情報から，あ，い，う，え，お，んの６クラスの分類を行う．

4.3 母音列の抽出

認識された母音形状の羅列を用いて，そこからコマンドと対応付けるための母音列の抽出を行う．母音列の抽出を行うには，フロントカメラより取得される映像から実際に母音を発声しているタイミングを取得する必要がある．本研究では，単語の発声スピードに合わせたリズムをユーザに示す UI を作成し，母音形状の認識を行うタイミングをユーザに示すことによって母音列を抽出することを検討している．そうして推定された母音列から，あらかじめ幾つかのアプリケーションの起動操作のコマンドを割り当てておき，コマンドの認証を行うことによってショートカット操作の機能を実装する予定である．

5. まとめと今後の予定

本稿では，口唇形状を用いた携帯端末の操作手法について提案を行った．提案手法では，携帯端末のフロントカメラを用いて口唇形状を認識し，そこから得られる母音列をコマンド入力とし，携帯端末の操作を行うことができる．特徴としては，音声入力のように直感的な入力が可能であり，入力の際にユーザは各操作ごとのジェスチャを覚える必要がない．また，実際に発話をする必要がないことから，公共の場における抵抗感が音声入力に比べて軽減されるのではないかと考えている．そのプロトタイプとして，フロントカメラ映像から，口唇領域を抽出し，母音形状の認識を行う機能の実装を行った．

今後は，認識された母音形状から，母音列の推定を行い，コマンド入力としてアプリケーションの起動操作を行える機能を実装する予定である．そのプロトタイプを用いて，母音形状の認識率，公共の場における使用時の抵抗感の調査のため，評価実験を行う予定である．また，アプリケーションの切り替え操作以外の応用，例えばタッチパネルと組み合わせた操作などを今後検討していこうと考えている．

参考文献

[1] Poppinga B, Sahami Shirazi A, Henze N, Heuten W, Boll S. Understanding shortcut gestures on mobile touch devices. In Proceedings of the 16th international conference on Human-computer interaction with mobile devices and services, pp. 173-182, 2014.
[2] http://www.apple.com/jp/ios/siri/.
[3] Sami Ronkainen, Jonna Hkkil, Saana Kaleva, Ashley Colley and Jukka Linjama. Tap input as an embedded interaction method for mobile devices. In Proceedings of the TEI'07, pp. 263-270, 2007.
[4] LYONS, Michael J.; CHAN, Chi-Ho; TETSUTANI, Nobuji. Mouthtype: Text entry by hand and mouth. In Proceedings CHI'04 Extended Abstracts on Human Factors in Computing Systems. pp. 1383-1386, 2004.
[5] Maryam Azh,Shengdong Zhao. LUI: lip in multimodal mobile GUIinteraction. In Proceedings of the 14th ACM international conference on Multimodal interaction (ICMI '12), pp. 551-554, 2004.
[6] 宮崎剛, 中島豊四郎. 日本語発話時における口形変化のコード化の提案 第 7 回情報科学技術フォーラム (FIT2008) 講演論文集, 第 3 分冊,pp.55-57,2008.