

スマートフォン向け口唇動作映像自動生成アプリケーションの試作

宮崎 剛 †

山本 富士男 †

中島 豊四郎 ‡

† 神奈川工科大学情報学部情報工学科

‡ 椋山女学園大学文化情報学部文化情報学科

1 はじめに

著者らは、日本語読唇術の技能を持つ人の読唇方法をモデル化して、機械読唇を実現するための研究を進めてきた [1]。その過程で、日本語の仮名から、その語句を発話する際に順に形成される基本口形（日本語の母音 5 口形と閉唇口形を指す）の列を生成する方法を提案した [2]。これを、“口形順序コード” (Mouth Shapes Sequence Code) と呼ぶこととし、この口形順序コードを用いて、任意の日本語語句発話時の口唇動作映像を再現する方法について提案した [3]。これにより、コンピュータに日本語の仮名を入力すれば、その語句を発話する口唇動作映像を自動的に生成することが可能となった。この自動生成は、読唇術の修得、特に聴覚障害者向け読唇術トレーニングのアプリケーションへの応用が期待できる。一方、今日、スマートフォンに代表される携帯端末は、その優れた携帯性と操作性により、種々の分野で利用されている。

そこで、本論文では、聴覚障害を持つ人や高齢者を対象に、スマートフォン上で動作する口唇動作映像の自動生成アプリケーションを試作し、パソコンと比べて処理能力やリソースに制約があるスマートフォンへの適用の方法と問題点について考察する。

2 口唇動作映像自動生成アプリケーション

本研究では、OS に Android を搭載したスマートフォン上に搭載するアプリケーションの開発を行う。本アプリケーションの構成を図 1 に示す。また、その開発環境を表 1 に示す。

表 1: アプリケーション開発環境

OS (PC)	Mac OS X 10.7.2
Android SDK	R16
Java	JDK 1.6
Android NDK	R7
C コンパイラ	gcc 4.2.1
画像処理ライブラリ	OpenCV 2.3.1 for android

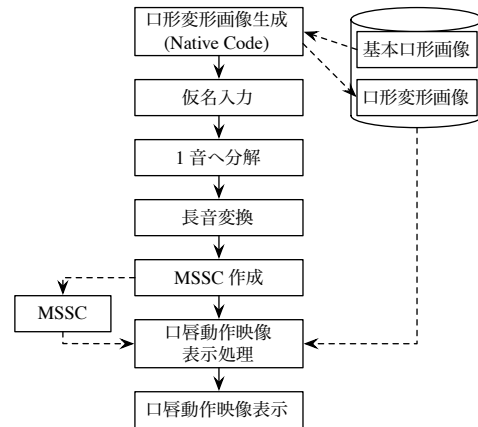


図 1: 本アプリケーションの構成

本アプリケーションは、起動後、図 1 に示すように口唇動作映像を表示する際に必要な全ての口形変形画像を、基本口形画像から生成する。なお、この生成部分では、複雑な画像処理に相当の時間を要するためネイティブコードで作成し、処理の短縮を図っている。ここで、画像処理には OpenCV ライブラリを使用している。そして、その他の部分は Java で作成し、ネイティブコードへは JNI (Java Native Interface) を通じてアクセスする。

全ての口形変形画像生成後、ユーザによる語句（仮名）の入力が行われる。入力された語句は 1 音毎に分解されたあと、長音変換が行われる。長音変換とは、日本語の仮名表記の際、オ列の後の長音は“う”と表記されるため、この仮名を長音に変換する処理である。例えば、“包丁”は“ほうちょう”と仮名表記されるが、発声する際は“ほーちょー”となるため、“ほうちょう”と入力された仮名を“ほーちょー”へ変換する。そして、文献 [2] で提案した方法を用いて、仮名から口形順序コード (MSSC) を生成する。生成された MSSC に従い、基本口形画像と本アプリケーション起動時に生成された口形変形画像を用いて、口唇動作映像を表示する。

本アプリケーションの表示画面を図 2 に示す。図 2 左の画面がアプリケーション起動後の画面（メイン画面と呼ぶ）で、同図右の画面が仮名入力画面である。メイン画面は、上部に口唇動作映像を表示する領域があり、この領域は高速描画を実現するため、SurfaceView と呼ばれる描画領域で構成している。メイン画面の開始ボタンをタッチすることにより仮名入力

Development of an Application Software to Generate Lips Movement Images Automatically for Smartphones

†Tsuayoshi MIYAZAKI †Fujio YAMAMOTO

‡Toyoshiro NAKASHIMA

†Department of Information and Computer Sciences, Kanagawa Institute of Technology

‡School of Culture-Information Studies, Sugiyama Jogakuen University

表 2: 実験に使用したスマートフォンの仕様

	機種 1	機種 2
製品名	Xperia arc SO-01C	GALAXY S II SC-02C
CPU	MSM8255 1GHz	S5PC210 1.2GHz デュアルコア
RAM	512MB	1GB
ディスプレイ解像度	480×854 ピクセル	480×800 ピクセル
OS	Android 2.3.4	Android 2.3.5



図 2: 本アプリケーションの画面

表 3: 画像生成の平均時間 (秒)

	機種 1	機種 2
口形変形画像 1 枚	1.86	1.14
全ての口形変形画像	57.79	35.28

画面に切り替わり、仮名入力後に OK ボタンをタッチすることによりメイン画面に戻り、口唇動作映像が表示される。口唇動作映像の表示方法は、文献 [3] で提案した方法を利用する。

3 口唇動作映像生成実験

本アプリケーションを用いて、口唇動作映像の生成実験を行った。ここで、本実験に使用した 2 台のスマートフォンの機種の仕様を表 2 に示す。

実験の結果、口形変形画像の生成と口唇動作映像の表示は正しく処理でき、口形変形の際のフレームレートも 30fps を実現できた。また、生成する口唇動作映像のサイズを 480×360 ピクセルとし、生成する口形変形画像数を 31 とした場合の画像 1 枚の生成と全ての画像の生成に要した時間を表 3 に示す。表 3 における平均時間は、本アプリケーションを 5 回起動した際の平均の時間である。

表 3 より、機種 1 と機種 2 の CPU 性能の違いによる画像生成時間の差が確認できた。機種 2 は機種 1 に比べて約 1.6 倍の処理性能であったが、画像 1 枚の生成に 1 秒以上かかっている。従って、現時点では、口

唇動作に合わせてリアルタイムに口形変形画像を生成することは、実用上問題があることがわかった。また、今回の実験では、生成する口形変形画像の数を 31 としたが、これ以上増やした場合に、機種 1 においてメモリ不足が発生してアプリケーションが不正終了するケースがあることも確認できた。これらより、スマートフォン上で口唇動作映像を生成する際には、予め口形変形画像を生成する必要があり、さらにその画像数についても、機種のリソースに応じて考慮する必要があることがわかった。

4 まとめ

本論文では、口唇動作映像を自動生成するスマートフォン向けのアプリケーションを試作し、その動作や処理性能について実験を行った。その結果、現在のスマートフォンは、パソコンに比べて CPU の処理性能も低く、ハードウェアリソースも十分ではないため、実用上問題があることがわかった。しかし、全ての画像が生成できれば、口唇動作映像は正しく表示され、フレームレートについても十分な結果が得られた。今後、スマートフォンのハードウェア性能の向上が十分に予想されることを考慮すれば、近い将来、処理時間の問題も解決できると考えられ、本アプリケーションの実用化の可能性も十分にあると言える。

参考文献

- [1] Tsuyoshi, M., Toyoshiro, N. and Naohiro, I.: A Detection Method of Basic Mouth Shapes from Japanese Utterance Images, *Proceedings of the 14th International Conference on Human-Computer Interaction, Part I, HCI 2011, LNCS 6761*, pp. 608–617 (2011).
- [2] 宮崎剛, 中島豊四郎: 日本語発話時の特徴的口形のコード化と口形変化情報表示方法の提案, *電気学会論文誌 C*, Vol. 129, No. 12, pp. 2108–2114 (2009).
- [3] 宮崎剛, 中島豊四郎: 口形順序コードを用いた発話映像自動生成方法, 第 9 回情報科学技術フォーラム (FIT2010) 講演論文集, 第 3 分冊, pp. 671–672 (2010).