発話映像から導出した特徴的口形の機械読唇への適用評価

宮崎 剛 †　　中島 豊四郎 ‡

† 神奈川工科大学 情報学部 情報工学科

‡ 椙山女学園大学 文化情報学部 文化情報学科

1 はじめに

現在，音声認識を補完して発話内容の認識率を向上させたり，聴覚障害者とのコミュニケーションを支援したりする目的で，情報処理技術を用いて読唇を行う“機械読唇”に関する研究が進められている [1, 2]．これらの研究では，発話中の口唇やその領域の動きに着目しており，発話期間の口唇の動きを時系列に数値化している．しかし，これらの方法では発話語句を認識する際に必要な語句のデータは予め発話した映像からでなければ取得できないという問題があった．そのため，認識対象語句を追加や変更するたびにデータをとる必要があった．

そこで，本研究では機械読唇におけるこの問題を解決するために，発話中の口形に着目する方法を提案する．実際に，読唇技能保持者は発話中の口形に着目して読唇を行っており [3]，著者らはこの方法を利用して発話時の口形変化を記号（口形コード）によって表現する方法を提案した [4]．この方法により，語句から発話時の口形変化を導出することが可能となった．

本論文では，発話映像から口形変化を導出し，得られた口形変化のデータが，著者らが提案した方法に適用でき，発話中の口形変化を利用した機械読唇が実現可能か否かを評価する．

2 特徴的口形の導出

文献 [4] では，日本語発話時の特徴的な口形として 6 つの口形（母音 5 口形＋閉唇口形）を提案した（以降，基本口形と呼ぶ）．ここでは，その提案を基に発話映像からこれら基本口形の導出を行う．そのための方法として，基本口形をテンプレート画像としたテンプレートマッチングを採用する．発話映像の各フレームに対して基本口形とのテンプレートマッチングを行い，その類似度を時系列に取得する（図 1）．そして，取得した類似度から発話期間における基本口形を導出する．

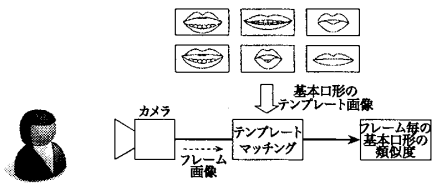

図 1: テンプレートマッチングの概要

3 実験

発話映像から基本口形を導出するための実験は，図 2 に示すように，家庭用ディジタルビデオカメラ（DV カメラ）と映像信号変換器，コンピュータ（PC）を接続する．そして，被験者を DV カメラで撮影し，その映像を PC に取り込み，画像処理を行う．語句を発話する際，発話の前後では唇を閉じた状態にする．

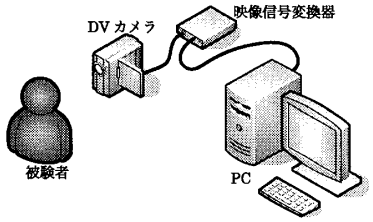

図 2: 実験の概略図

実験で発話した語句と口形変化コードを表 1 に示す．また，それぞれの実験で得られた各基本口形の類似度のグラフを図 3，図 4 に示す．グラフでは縦軸に類似度を，横軸にフレーム番号 $n(\geq 0)$ をとる．

表 1: 実験用発話語句とその口形変化コード

実験番号	発話語句	口形変化コード
実験 1	厚木 (あつぎ)	AUI
実験 2	海老名 (えびな)	ExIA

4 考察

実験 1 では，$n = 0$ から $n = 9$(図 3 中 (a)) まで基本口形 X の類似度が最も高く，値の変化も少ない安定した状態が続いている．これは，発話前の閉唇口形の状態が続いていた期間と解釈できる．この期間は他の基本口形の類似度も安定している（以降，この期間を“口形安定期間”と呼ぶ）．その後，$n = 9$ から $n = 13$(図 3 中 (b)) にかけて各基本口形の類似度に大きな変化が見られ，$n = 13$ から基本口形 A に対する類似度が高くなり，$n = 20$(図 3 中 (c)) まで口形安定期間が続いている．ここで，$n = 9$ から $n = 13$ の期間は口形が閉唇口

Applicability of Peculiar Mouth Shapes to Lipreading and Its Evaluation

†Tsuyoshi MIYAZAKI ‡Toyoshiro NAKASHIMA

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

‡School of Culture–Information Studies, Sugiyama Jogakuen University

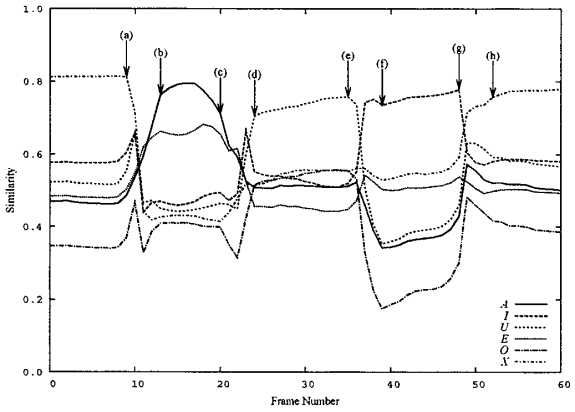

図 3: 実験 1 の類似度変化

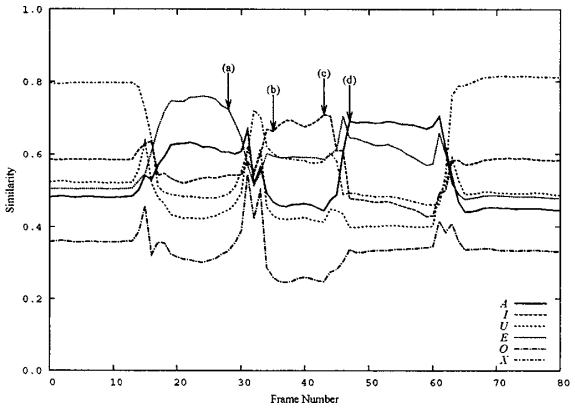

図 4: 実験 2 の類似度変化

形から/a/の口形へ変形している過程であると考えられる（以降，この期間を“口形変形期間”と呼ぶ）．そして，$n = 13$ から $n = 20$ の期間で/a/の口形となっていたことがわかる．その後も同様に口形変形期間と口形安定期間を繰り返している．表 2 に，実験 1 における各口形安定期間の開始フレーム番号と終了フレーム番号，口形安定期間のフレーム数，口形安定期間で類似度が最大となった基本口形を示す．この実験では全ての口形変形期間のフレーム数が 5 となっている．

この結果から，各口形安定期間において，類似度が最大となる基本口形を口形コードに置き換えて時系列に並べると，口形変化コードと同じコード列になることがわかる．ただし，発話開始前と発話開始後の閉唇口形の期間は除く．

同様に，実験 2 の結果を表 3 に示す．この実験では，第 2 の口形安定期間のあとの口形変形期間 (図 4 中 (a) から (b)) のフレーム数が 8 となり，他の口形変形期間のフレーム数 (5 または 6) と比較すると多くなっている．そこでこの期間を見てみると，基本口形 X に対する類似度が上に凸となるグラフ形状を示しており，かつ類似度の最大値は他の口形安定期間での最大類似度基本口形の類似度に近い値を示している．この結果から，この期間に閉唇口形が短期間で出現したと考えることができる．そして，そのあとの口形安定期間の最大類似度基本口形が I であることから，この閉唇口形は“び”の初口形であるととらえることができる．初口形を含めて最大類似度基本口形を順にたどると，口形変化コードで示している口形が順に出現していることが分かる．

表 2: 実験 1 の各口形安定期間と最大類似度基本口形

	1	2	3	4	5
開始フレーム番号	0	13	24	39	52
終了フレーム番号	9	20	35	48	60
継続フレーム数	10	8	12	10	9
最大類似度基本口形	X	A	U	I	X

表 3: 実験 2 の各口形安定期間と最大類似度基本口形

	1	2	3	4	5
開始フレーム番号	0	18	35	47	65
終了フレーム番号	13	28	43	60	80
継続フレーム数	14	11	9	14	16
最大類似度基本口形	X	E	I	A	X

これらの結果から，口形変化を利用した機械読唇の実現の可能性があることが確認できた．

5 まとめ

本論文では，口形に着目した機械読唇を実現するために発話映像から基本口形の導出実験を行った．実験の結果から，基本口形の類似度をもとに発話時の口形を導出することが可能であることを示し，口形変化をもとにした機械読唇の可能性が確認できた．今後は，導出した口形から発話語句を認識する機械読唇システムを構築し，さらなる評価と実験を進めていく必要がある．

参考文献

[1] 間瀬健二, Alex Pentland. オプティカルフローを用いた読唇. 信学論. Vol.J73–D–II, No.6, pp.796–803, 1990.

[2] 齊藤剛史, 小西亮介. トラジェクトリ特徴量に基づく単語読唇. 信学論. Vol.J90–D–II, No.4, pp.1105–1114, 2007.

[3] 読話教材製作・監修委員会（編）. 豊かなコミュニケーションに向けて －読話のためのビデオテキスト－ 家族編. 社団法人全日本難聴者・中途失聴者団体連合会, 東京, 1997.

[4] 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.