

K-025

口形順序コードを用いた発話映像自動生成方法

An Automatic Generation Method of Speaking Pictures used Mouth Shape Order Codes

宮崎 剛[†]
Tsuyoshi MIYAZAKI

中島 豊四郎[‡]
Toyoshiro NAKASHIMA

1. はじめに

聴覚に障害を持つ人とのコミュニケーション手段として、一般的には“手話”や“筆談”、“読話(あるいは口話)”が考えられる。しかし、手話で正しくコミュニケーションをとるためには特別な訓練や練習を必要とし、筆談は意志の伝達に時間がかかってしまう。

一方、読話、または口話は、話者の口唇の動きを見ながら発話内容を理解する。ただし、そのためには、聴者側に読唇術の技能が必要となる。しかし、聴覚障害者が読唇術を習得すれば多くの健聴者と円滑なコミュニケーションをとることが可能となり、コミュニケーションを取ることができる相手の幅も広がる。

読唇術を身につけるためのトレーニングは、ビデオ映像を利用した訓練が主となっている。ビデオ映像による学習は、実際に発話している映像を用いているため対面に近い状況を作ることができ、かつ、繰り返し何度も再生することもできるため、有効な教材である[1]。しかし、予め撮影された映像であるため、任意の語句の発話に対応することができない問題がある。

このような問題を解決するために、聴覚障害者の読話訓練のための動画像を生成する研究が行われている[2]。この研究では、キーボードから仮名入力した語句の発話の動画像を生成する。日本語の音節に対して42種類に口形を分類し、あらかじめ作成しておいたそれらの口形画像を入力された仮名の音節に合わせて連続して表示させる。しかし、日本語の口形変化は1つの音節からだけでは確定できず、音の前あるいは後に来る音も含めて判断しなければならない場合が存在しているが[1]、これらについては考慮されていないため必ずしも正確な口形変化が表現されているとは言えない。

そこで、本論文では任意の語句の発話時の口形変化を正確に表現する映像の生成方法について述べる。著者らはこれまでに、読唇技能保持者の知見を論理化して口形を計算機上で扱う方法を提案した[3]。日本語の5つの母音の口形と閉唇口形を基本口形として定義し、さらに基本口形をコード化して口形順序コードを提案した。この、口形順序コードは日本語の音と口形変化の特徴を利用して任意の語句(仮名)から生成することも可能であるため、この方法を発話映像の生成に利用する。

2. 口形順序コード

口形順序コードは、初口形の口形コードと終口形の口形コードの連続で表現される。ここで初口形とは、“マ”や“サ”、“フ”のように発話の初期に形成される母音の口形とは異なる口形のことである。一方、終口形は

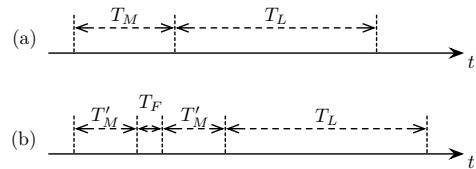


図 1: 発話時の口形変形時間と基本口形形成時間

母音に相当する口形のことである。口形コードは $C = \{A, I, U, E, O, X, -\}$ として定義する。口形コードはそれぞれ先頭から順に、ア口形からオ口形、閉唇口形に対するコードと最後は初口形が形成されない場合に対応する。

口形順序コードは、語句を発話する際に順に形成される基本口形を表現したものである。例えば、“厚木(アツギ)”の口形順序コードは“-A-U-I”であり、“海老名(エビナ)”の口形順序コードは“-EXI-A”である。口形順序コードでは、奇数番目の口形コードは初口形を、偶数番目の口形コードは終口形を表現するため、厚木では初口形は形成されず、海老名では2音目の“ビ”で初口形が形成されることがわかる。

3. 口形変形映像生成方法

日本語を発話する場合の口形変化は、ある基本口形から別の基本口形への変形の繰り返しである。このことは、口形順序コードも同様である。従って、発話時の1つの音に関する時間には、前の基本口形から次の基本口形へ口形が変形する時間と基本口形を形成している時間の2種類がある。発話時の時間軸におけるそれぞれの時間を図1に示す。なお、図1(a)は初口形を形成しない場合、図1(b)は初口形を形成する場合であり、 T_M 、 T_M' は口形変形の時間、 T_L は終口形を形成している時間、 T_F は初口形を形成している時間を表す。

そこで、口形順序コードから発話映像を生成する処理の流れを図2に示す。図2中の $code(p)$ は、口形順序コードの p 番目のコードを表し、“口形変形”は前の基本口形から次の基本口形へ T 時間かけて変形する処理を表し、“口形表示”は T 時間 c に対する基本口形を表示する処理を表す。口形変形では、各基本口形の画像をデータベースに格納しておき、変形元口形と変形先口形の2枚の画像を画像処理することで口形の変形を実現する。

4. 実験

任意の語句から発話映像を生成する実験を実施する。ここでは、任意の語句(仮名)について口形順序コードを生成する[3]。そして、口形順序コードから本論文で提案した方法で発話映像を生成する。なお、本論文で提案する口形変形映像の生成方法では、図1で示した各時間を調整することで、速く発話する映像やゆっくり発話

[†]神奈川工科大学, Kanagawa Institute of Technology
[‡]椋山女学園大学, Sugiyama Jogakuen University

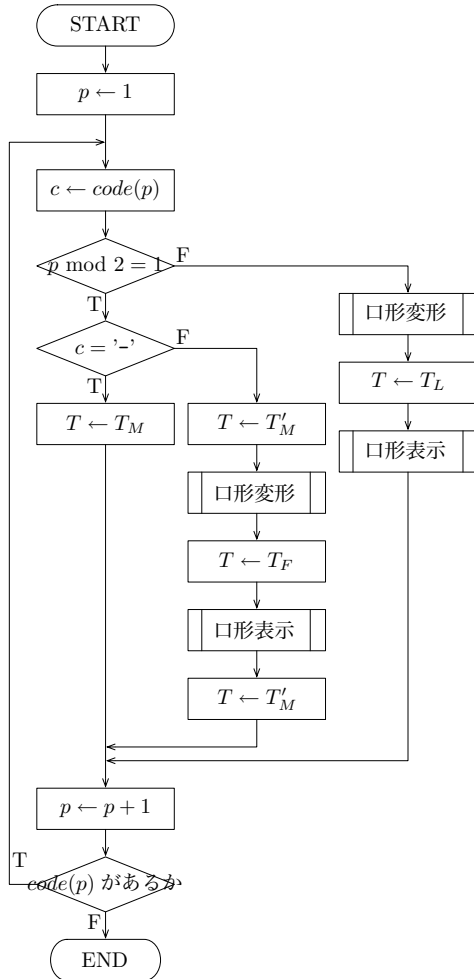


図 2: 口形順序コードから発話映像を生成する処理

する映像を生成することも可能になる。そこで、実験では発話速度の異なる3種類の発話映像を表1に示す時間で生成する。

発話映像を生成する語句とその口形順序コードを表2に、発話映像を生成するためにデータベースに登録した基本口形画像を図3に示す。なお、発話映像の生成において口形を変形させる映像の生成には、CG技術の1つであるモーフィングを利用する。

実験の結果、良好な発話映像を生成することができた。モーフィングによる口形変形例を図4に示す。図4は $T'_M = 125$ の際の画像である。また、口形の変形時間を変更することで速く発話したりゆっくり発話したりする感じが得られた。ただし、一部の映像で歯が二重に見える場合もあったため、モーフィングの方法を改善する必要がある。

5. まとめ

本論文では、任意の語句の発話映像を正確な口形変化で表現する映像の自動生成方法について述べた。日本語発話時の口形変化が、ある基本口形から別の基本口形への変形の連続であることと、口形順序コードを用いることで発話映像を自動生成することができた。また、口形

表 1: 発話速度毎の発話映像生成時間(ミリ秒)

	T_M	T'_M	T_F	T_L
速い	167	125	42	125
標準	167	125	83	208
ゆっくり	208	167	83	375

表 2: 発話映像を生成する語句とその口形順序コード

#	単語	口形順序コード
1	厚木	-A-U-I
2	鉛筆	-E-X-I-U
3	絵の具	-EUO-U
4	川越	-AUA-O-E
5	筆箱	-UIEXA-O
6	消しゴム	-E-I-OXU
7	川崎	-AUAIA-I

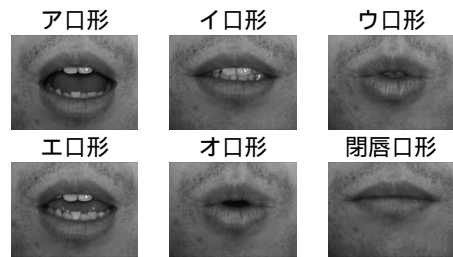


図 3: データベースに登録した基本口形画像

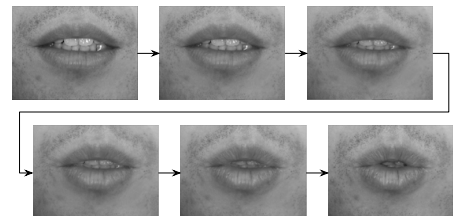


図 4: モーフィングによる口形変形例(イ口形 → ウ口形)

変形時間を変更することで速く発話している映像やゆっくり発話している映像の生成も可能となった。これらの映像は、口話トレーニングの際のインタラクティブな教材の1つとして有効であると考えられる。今後は、実際の口話トレーニングに向けたアプリケーションの開発とその評価が必要である。

参考文献

- [1] 読話教材制作・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.
- [2] 松岡清利, 黒須頭二. 聴覚障害者の読話訓練のための動画プログラム. 信学論, Vol. J70-D, No. 11, pp. 2167-2171, 1987.
- [3] 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. 電学論C, Vol. 129, No. 12, pp. 2108-2114, 2009.