RK-002

日本語発話時における口形変化のコード化の提案
A Coding Method of Changes in Mouth Shape when Uttering Japanese

宮崎 剛† 中島 豊四郎‡
Tsuyoshi MIYAZAKI Toyoshiro NAKASHIMA

1. はじめに

今日，情報処理技術を用いて読唇を実現しようとする研究が進められている．これらの研究は機械読唇とよばれ，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技術として研究されている．

一般に，機械読唇ではカメラ等を用いて発話時の口唇とその周辺を含む映像を撮影し，発話期間の複数枚の画像 (フレーム画像) を取得する．そして，取得したフレーム画像に対して何らかの画像処理を施し，発話期間の口唇の動きに関する時系列の数的情報 (以降，口唇動作情報と呼ぶ) を算出する．そして，算出された口唇動作情報を基に発話内容を推測するという方法がとられている [1, 2, 3, 4, 5, 6]．

例えば，オプティカルフローを用いた方法 [1, 2] では，口唇動作情報の表現方法として口唇周辺に設定した計測点のフレーム画像間の移動方向と移動距離を用いている．また，口形の特徴量を利用した方法 [3, 4] では，口唇領域のアスペクト比を用いている．このように，これらの機械読唇では主に口唇やその周辺領域の “動き” に着目している．口唇の動きに着目することで，フレーム毎に変化する口唇の情報を取得できる．しかし，動きというものは前の状態からの変化量から求めているため，相対的な情報となる．そのため，従来の研究においては，ある語句を発話するときの口唇の動きに関する情報は，実際に発話を行わなければ得ることができなかった．これでは，認識対象とする語句が多数ある場合等では対応が非常に難しい．

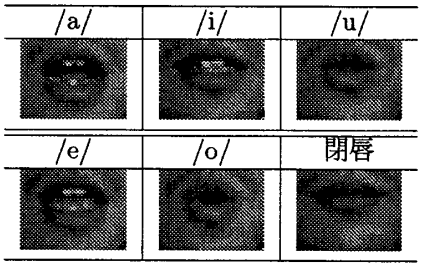
そこで，本論文では口唇の動きではなく，口唇の “口形” に着目する．発話内容の認識に口形を利用した研究 [7] も報告されているが，この研究は音声認識の補完としての利用であり，さらに母音のみの認識にとどまっている．日本語には音を発する際，あるタイミングでいくつかの特徴的な口形が形成される [8, 9]．本論文では単なる母音の口形ではなく，これらの特徴的な口形に着目する．そして，これらの特徴的な口形を計算機上で処理しやすくするためにコードを用いて表現することを提案し，口形のコードを用いて日本語発話時の口形変化の様子を表現する方法について述べる．口形は，動きを対象としてきた研究と比較すると “絶対的” な情報となる．そのため，本論文で提案する方法を用いると実際に発話することなく，ある語句を発話するときの口唇の形状変化の情報を容易に得ることが可能となる．

2. 日本語発話における口形の特徴

日本語を発声する場合，ある音を発声するときの口形はその音の母音の口形になることはよく知られている．

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

表 1: 6 つの口形



また，マ行の音のように唇を閉じてから発声する音 (以降，両唇音とよぶ) も存在している．両唇音では，発声の初期に閉唇口形を形成し，その後母音の口形へ変形させながら発声することで正しい音となる．このように，日本語の音にはその音を発声させるために発声の初期に母音とは異なる口形を形成する音が存在している．日本語の音には，両唇音の他にも発声の初期に/i/や/u/の口形を形成する音が存在している．そして，文献 [8, 9] では日本語の全ての音について，その音を発声するときの口形の組み合わせが示されており，日本語の音は/a/，/i/，/u/，/e/，/o/と閉唇の 6 つの口形 (表 1) の組み合わせで発声されていると示されている．

3. 口形の定義

本論文では，文献 [8, 9] で示されている 6 つの口形を日本語発話時の特徴的口形と考える．そこで，これら 6 つの口形を “基本口形” と呼ぶこととし，基本口形 B を式 (1) のように定義する．ここで，X は閉唇口形であり，X 以外の要素はそのアルファベットに対応する母音の口形を表すものとする．

$$B = \{A, I, U, E, O, X\} \tag{1}$$

つぎに，日本語発声時の初期に形成される口形を “初口形”，発声した音の母音に相当する口形を “終口形” と呼ぶこととし，初口形 F を式 (2) に，終口形 L を式 (3) に定義する．

$$F = \{I, U, X\} \tag{2}$$

$$L = \{A, I, U, E, O, X\} \tag{3}$$

ここで，終口形は日本語全ての音で形成されるが，初口形は音によっては形成されない場合もある．そのため，初口形を $f(f \in F)$，終口形を $l(l \in L)$ としたとき，日本語の音を発声するときの口形は，終口形のみの “l” または初口形と終口形から形成される “fl” のどちらかとなる．本論文では，日本語の発声時に口形が l となる音を “単口形音” と呼び，口形が fl となる音を “複口形音” と呼ぶこととする．ただし，複口形音を形成する f と

表 2: 複口形音を形成する初口形と終口形の組み合わせ

f	l
I	A
I	E
U	A
U	I
U	E
U	O
X	A
X	I
X	U
X	E
X	O

l の口形の組み合わせは表 2 に示す 11 通りのみとなる [8, 9]．また，l や fl のことを 1 つの音を発声する際の口形であることから，“口形節” と呼ぶこととする．

3.1 口形変化コード

初口形と終口形の各口形に対するコードを “口形コード” として定義する．初口形に対する口形コードを “初口形コード” とし，式 (4) に C_F として定義する．同様に，“終口形コード” を式 (5) に C_L として定義する．

$$C_F = \{\mathtt{i}, \mathtt{u}, \mathtt{x}\} \tag{4}$$

$$C_L = \{\mathtt{A}, \mathtt{I}, \mathtt{U}, \mathtt{E}, \mathtt{O}, \mathtt{X}\} \tag{5}$$

そこで，これらの口形コードと文献 [8, 9] より，日本語全ての音を発声するときの口形がコードによって表現することが可能となる (表 3)．

このように，日本語全ての音の口形がコードによって表現することができるため，日本語の語句を発話する際の口形変化を口形コードを用いて表現することが可能となる．例えば，「朝日 (あさひ)」という語句を発話するときの口形変化を口形コードを用いて表現するには，それぞれの音に対する口形コードを表 3 より抽出する．この例では，“A”(あ)，“iA”(さ)，“I”(ひ) となる．そして，これらの口形コードを音の順に連結させた口形コードの列 “AiAI” を “口形変化コード” と呼ぶ．この例では，「朝日」を発話する際に口形変化コードの左から右の順に口形が変形していくことを意味している．

また，この口形変化コードを口形節で区切ったコード列を “口形節コード” と呼ぶこととする．例えば，先の「朝日」の例では，口形節コードは “A/iA/I” と区切られ，左から順に “第 1 口形節”，“第 2 口形節”，“第 3 口形節” となる．

ここで，口形変化コード中の口形コードを示すものとして c_F と c_L を定義する．s を口形節番号 ($s = 1, 2, 3, \ldots$) とするとき，$c_F(s)$ とは第 s 口形節の初口形コードを示し，$c_L(s)$ は第 s 口形節の終口形コードを示す．先に述べた「朝日」の例における各口形節の初口形コードと終口形コードは表 4 の通りになる．ここで，表 4 中の初口形コード “ϕ” は，その口形節の音が単口形音であるため初口形コードが存在しないことを意味する．

表 3: 日本語の音に対する口形コードの一覧

		ア列	イ列	ウ列	エ列	オ列
ア行	音	あ	い	う	え	お
	口形コード	A	I	U	E	O
カ行	音	か	き	く	け	こ
	口形コード	A	I	U	E	O
サ行	音	さ	し	す	せ	そ
	口形コード	iA	I	U	iE	uO
タ行	音	た	ち	つ	て	と
	口形コード	iA	I	U	iE	uO
ナ行	音	な	に	ぬ	ね	の
	口形コード	iA	I	U	iE	uO
ハ行	音	は	ひ	ふ	へ	ほ
	口形コード	A	I	U	E	O
マ行	音	ま	み	む	め	も
	口形コード	xA	xI	xU	xE	xO
ヤ行	音	や		ゆ		よ
	口形コード	iA		U		uO
ラ行	音	ら	り	る	れ	ろ
	口形コード	iA	I	U	iE	uO
ワ行	音	わ				を
	口形コード	uA				uO
ガ行	音	が	ぎ	ぐ	げ	ご
	口形コード	A	I	U	E	O
ザ行	音	ざ	じ	ず	ぜ	ぞ
	口形コード	iA	I	U	iE	uO
ダ行	音	だ	ぢ	づ	で	ど
	口形コード	iA	I	U	iE	uO
バ行	音	ば	び	ぶ	べ	ぼ
	口形コード	xA	xI	xU	xE	xO
パ行	音	ぱ	ぴ	ぷ	ぺ	ぽ
	口形コード	xA	xI	xU	xE	xO
キャ行	音	きゃ		きゅ	きぇ	きょ
	口形コード	iA		U	iE	uO
シャ行	音	しゃ		しゅ	しぇ	しょ
	口形コード	iA		U	iE	uO
チャ行	音	ちゃ		ちゅ	ちぇ	ちょ
	口形コード	iA		U	iE	uO
ニャ行	音	にゃ		にゅ	にぇ	にょ
	口形コード	iA		U	iE	uO
ヒャ行	音	ひゃ		ひゅ	ひぇ	ひょ
	口形コード	iA		U	iE	uO
ミャ行	音	みゃ		みゅ	みぇ	みょ
	口形コード	xA		xU	xE	xO
リャ行	音	りゃ		りゅ	りぇ	りょ
	口形コード	iA		U	iE	uO
ギャ行	音	ぎゃ		ぎゅ	ぎぇ	ぎょ
	口形コード	iA		U	iE	uO
ジャ行	音	じゃ		じゅ	じぇ	じょ
	口形コード	iA		U	iE	uO
ビャ行	音	びゃ		びゅ	びぇ	びょ
	口形コード	xA		xU	xE	xO
ピャ行	音	ぴゃ		ぴゅ	ぴぇ	ぴょ
	口形コード	xA		xU	xE	xO
ウァ行	音	うぁ	うぃ		うぇ	うぉ
	口形コード	uA	uI		uE	uO
ファ行	音	ふぁ	ふぃ		ふぇ	ふぉ
	口形コード	uA	uI		uE	uO

表 4: 「朝日」に対する各口形節の口形コード

s	$c_F(s)$	$c_L(s)$
1	ϕ	A
2	i	A
3	ϕ	I

しかしながら，全ての単語や文章の口形変化コードがこのように口形コードを順に連結させるだけで生成できるわけではない．

3.2 口形変化コード生成規則

日本語の音では，その音を単独で発声するときには複口形音であっても，語句として発話されるときなど他の音と続けて発声するときには直前の音との関係により，初口形が形成されなくなる場合がある．また，単口形音も直前の音の口形によっては，直前の音の口形に吸収されてしまう場合もある．さらに，促音 (っ) や撥音 (ん) は特定の口形がなく，その音の前後に来る音によって口形が変化する．そこで，これら日本語発話時の口形変形規則を，文献 [8, 9] を基に定義し，単純連結された口形変形コードから発話時の口形変化に対応した口形変形コードを生成する方法を述べる．ただし，単純連結の口形変形コードを生成する段階で促音や撥音の口形コードは確定することができないため，これらの口形コードは一時的に "*" で表現しておくこととする．

口形変形規則 1 $s > 1$ なる $c_L(s)$ に対し，$c_L(s) = c_L(s-1)$ かつ $c_F(s) = \phi$ である場合，$c_L(s)$ は $c_L(s-1)$ に吸収される．

口形変形規則 2 $s > 1$ なる $c_F(s)$ に対し，$c_F(s) \simeq c_L(s-1)$ である場合，$c_F(s)$ は $c_L(s-1)$ に吸収され，$c_F(s) = \phi$ となる．($\simeq$ は口形コードに対する口形が等しいことを意味する．)

口形変形規則 3 $s > 1$ なる $c_L(s)$ に対し，$c_L(s) = *$ かつ $c_F(s+1) = \mathtt{x}$ である場合，$c_L(s) = \mathtt{X}$ となり，$c_F(s+1)$ は $c_L(s)$ に吸収され，$c_F(s+1) = \phi$ となる．

口形変形規則 4 $s > 1$ なる $c_L(s)$ に対し，$c_L(s) = *$ かつ $c_L(s-1) = \mathtt{A}$ または $c_L(s-1) = \mathtt{E}$ である場合，$c_L(s) = \mathtt{I}$ となる．ただし，口形変形規則 3 の方が優先される．

口形変形規則 5 $s > 1$ なる $c_L(s)$ に対し，$c_L(s) = *$ かつ $c_L(s-1) = \mathtt{O}$ である場合，$c_L(s) = \mathtt{U}$ となる．ただし，口形変形規則 3 の方が優先される．

口形変形規則 6 $s > 1$ なる $c_L(s)$ に対し，$c_L(s) = *$ かつ $c_L(s-1) = \mathtt{I}$ または $c_L(s-1) = \mathtt{U}$ である場合，$c_L(s)$ は $c_L(s-1)$ に吸収される．ただし，口形変形規則 3 の方が優先される．

表 5 に，口形変形規則 n を適用したときに生成される口形変化コードの例を示す．

このように，口形コードを定義し，かつ口形コードに口形変形規則を適用することで，実際に発話を行わずに語句発話時の口形変化の情報を表現することが可能になる．また，生成された口形変形コードは語句発話時の口形変形情報を表現できていることも確認できる．

表 5: 口形変形規則 n の適用後に生成される口形変化コード例

n	語句	単純連結	規則適用後
1	明かり (あかり)	AAI	AI
2	伊勢 (いせ)	IiE	IE
3	コップ	O*xU	OXU
4	エンド	E*uO	EIuO
5	突起 (とっき)	uO*I	uOUI
6	近所 (きんじょ)	I*uO	IuO

4. おわりに

本論文では，日本語発声時の特徴的な口形に着目し，その特徴的口形をコードを用いて表現することを提案した．特徴的口形をコード化することで，日本語全ての音の口形コードを表現することが可能となった．口形をコード化し，日本語全ての音の口形がコード化されることで計算機上での処理が容易となり，かつ口形変形規則を適用することで実際に発話を行わずに語句発話時の口形変化の情報を表現することも可能となった．口形変形規則によって生成された口形変形コードも，いくつかの例を通して語句発話時の口形変形情報を表現できていることが確認できた．

今後は，実際にカメラを用いて本方式の有効性を確認していきたい．その際，発話をする人による口形の個人差と認識との関連についても検討していきたい．

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. 信会論. D-II, Vol. 73, No. 6, pp. 796–803, 1990.

[2] 大槻恭士, 大友照彦. オプティカルフローと HMM を用いた駅名発話画像認識の試み. 信学技報. PRMU, Vol. 102, No. 471, pp. 25–30, 2002.

[3] 李芝, 山崎一生, 黒畑喜弘, 小川英光. 部分空間法による読唇. 信学技報. PRMU, Vol. 97, No. 251, pp. 9–14, 1997.

[4] 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. D, Vol. 90, No. 4, pp. 1105–1114, 2007.

[5] 清田公保, 内村圭一. 口唇周辺画像情報を用いた発話単語認識. 信学論. D-II, Vol. 76, No. 3, pp. 812–814, 1993.

[6] 中田康之, 安藤護俊. 色抽出法と Eigentemplate 法を併用した口の位置検出と読唇処理への適用. 信学技報. PRMU, Vol. 101, No. 303, pp. 7–12, 2001.

[7] 呉簡彤, 田村進一, 光本浩士, 河合秀夫, 黒須顕二, 岡崎耕三. 音声・口形特徴量を併用するニュ-ラルネットを用いた母音認識. 信学論. D-II, Vol. 73, No. 8, pp. 1309–1314, 1990.

[8] 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.

[9] 桜井武志（編）.「話し言葉を読み取ってみませんか？」. 読話塾, 東京, 2004.