

特徴的口形に着目した日本語発声時の口唇形状変化の コード化表現に関する研究

宮崎 剛†

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

一般に情報処理技術を用いて読唇を行う機械読唇では、カメラ等を用いて口唇とその周辺を含む映像を撮影する。撮影して得られた各フレームの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もとに発話内容を推測するという方法がとられている[1, 2]。

これらの方法では口唇やその周辺を含めた領域の“動き”を対象にし、動きそのものは前の状態からの変化量によって求めているため、相対的な情報である。そのため、語句に関する口唇動作情報を作成するためには実際に発話を行わなければならない。

そこで、本論文では日本語発話時における口形変化の特徴を利用して、語句発話時の口形変化をコードを用いて表現する方法を提案する。

2 日本語の特徴的口形

一般に、日本語を発話する際、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。また、マ行のように唇を閉じて発声する音(以降、両唇音と呼ぶ)では、母音の口形の前に閉唇(唇を閉じた状態の口形)を形成する。さらに、音を発声するときに母音や閉唇とは異なる口形を形成して発声される音も存在する。文献[3]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いる。

3 口形の定義とコード化

本論文では5つの母音 /a/, /i/, /u/, /e/, /o/ の口形と閉唇口形を合わせた6つの口形を“基本口形”と呼ぶこととし、基本口形 B を(1)のように定義する。ここで、 X は閉唇口形であり、 X 以外の記号はその記号に対応する母音の口形を表すものとする。

$$B = \{A, I, U, E, O, X\} \quad (1)$$

A Coding Method of Change in Mouth Shape Based on Peculiar Mouth Shape when Uttering Japanese

†Tsuuyoshi MIYAZAKI ‡Toyoshiro NAKASHIMA

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

‡School of Culture-Information Studies, Sugiyama Jogakuen University

つぎに、母音口形の前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(2)に、終口形 L を(3)に定義する。

$$F = \{I, U, X\} \quad (2)$$

$$L = \{A, I, U, E, O, X\} \quad (3)$$

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“ l ”または初口形と終口形から構成される“ 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”、口形が fl となる音を“複口形音”と呼ぶこととする。

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 口形コード

初口形の記号化表現を“初口形コード”とし、(4)に定義する。同様に、“終口形コード”を(5)に定義する。初口形コードと終口形コードを併せて“口形コード”と呼ぶこととする。

$$C_F = \{i, u, x\} \quad (4)$$

$$C_L = \{A, I, U, E, O, X\} \quad (5)$$

これらの口形コードと文献[3]をもとにして、日本語全ての音を表現したものを表1に示す。ただし、カ行、ハ行、ガ行の各音の口形コードはア行と同じになる。同様に、タ行、ナ行、ラ行、ザ行、ダ行はサ行と同じに、シャ行、チャ行、ニャ行、ヒャ行、リャ行、ギャ行、ジャ行はキャ行と同じに、バ行、パ行はマ行と同じに、ビャ行、ピャ行はミャ行と同じに、ファ行はウァ行と同じになる。そして、ここに示す口形コードを用いることで、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)」と発話したときの口形変化を口形コードを用いて表現するには、それぞれの音に対応する口形コードを表1より抽出する。この例では、“A”、“iA”、“I”となる。そして、これらの口形コードを順につなぎ合

表 1: 日本語の音に対する口形コード

	ア列	イ列	ウ列	エ列	オ列
ア行	あ	い	う	え	お
口形コード	A	I	U	E	O
サ行	さ	し	す	せ	そ
ヤ行	や		ゆ		よ
キャ行	ぎゃ		ぎゅ	ぎえ	ぎよ
口形コード	iA	I	U	iE	uO
マ行	ま	み	む	め	も
ミヤ行	みゃ		みゅ	みえ	みよ
口形コード	xA	xi	xU	xE	xO
ワ行	わ				を
ウァ行	うぁ	うぃ		うぇ	うぉ
口形コード	uA	uI		uE	uO

わせた“ $AiAI$ ”を“口形変化コード”と呼び、この例では、「朝日」と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化していくことを表している。このように、口形変化コードを用いることで、容易に発話時の口形変化を表現することが可能となる。

また、この口形コードを口形節で区切った表現を“口形節コード”と呼ぶこととする。例えば、「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($s = 1, 2, 3, \dots$)とすると、 $c_F(s)$ 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」の例では、 $c_F(1) = \phi$ 、 $c_L(1) = A$ 、 $c_F(2) = i$ 、 $c_L(2) = A$ 、 $c_F(3) = \phi$ 、 $c_L(3) = I$ となる。ここで、“ 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。しかし、全ての語句の口形変化コードがこのように口形コードを順につなぎ合わせるだけで表現できるわけではない。

3.2 口形変形規則

日本語の音では、その音を単独で発声する場合には複口形音であっても、他の音と続けて発声する場合には直前の音との関係により、初口形が出現しなくなる場合がある。さらに、促音(っ)や撥音(ん)は決まった口形がなく、その音の前後の音によって口形が変わってくる。そこで、これら日本語発声時の口形変形規則を以下に定義する。

ここで、口形変形規則は表1を用いて生成された口形変化コードに対して適用することになる。ただし、語句に促音や撥音が含まれている場合、促音や撥音の口形コードは一時的に“*”で表現しておく。

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、 $c_F(s) \approx c_L(s-$

表 2: 生成される口形変化コード例

語句	口形変形規則	
	適用前	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	$E*uO$	E <u>I</u> uO
5 突起(とつき)	$uO*I$	uO <u>I</u> I
6 近所(きんじょ)	$I*uO$	IuO

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。(≈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 $c_F(s+1) = x$ である場合、 $c_L(s) = X$ 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 $c_L(s-1) = A$ または $c_L(s-1) = E$ である場合、 $c_L(s) = I$ 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 $c_L(s-1) = O$ である場合、 $c_L(s) = U$ 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 $c_L(s-1) = I$ または $c_L(s-1) = U$ 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2に語句と生成される口形変化コードの例を示す。

4 まとめ

本論文では、日本語発話時の口形変化の特徴を利用して、語句発話時の口形変化をコードを用いて表現する方法を提案した。日本語全ての音に対する口形コードを示すことにより、口形変形規則の適用が容易になり、語句に対する口形変化コードを生成することが可能となった。

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. D-2, Vol. 73, No. 6, pp.796-803, 1990.
- [2] 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.
- [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.