スピーチ支援のための広頚筋隆起センシングに基づく良い開口での母音発話認識

Recognition and Feedback of Vowel Utterance with a Good Mouth Shape Based on Sensing Platysma Muscle Bulging

5116E015-6　西村　幸泰　　指導教員　橋田　朋子　准教授

NISHIMURA Yukihiro　　Assoc.Prof. HASHIDA Tomoko

概要：スピーチは言語的伝達能力と非言語的伝達能力から評価されており，特に開口はこの両方を補助するという重要な役割を持つ．開口は母音発話の際に主に行われる．開口の中でも，話者の明瞭な発音を促し表情を豊か見せることができるような開口を本研究では良い開口と定義する．良い開口であるかどうかは首にある広頚筋という筋肉の隆起から推測できると筆者らは考えた．そこで本研究では，良い開口での母音発話の支援を目指し，良い開口での母音発話を認識するシステムを提案する．具体的には，広頚筋の隆起をフォトリフレクタで測定し，そのデータを機械学習にかけることによって，良い開口で母音発話が行われているかどうかを判断するシステムを実装する．提案システムの精度計測実験を行い，その結果を報告する．最後に良い開口での母音の発話のフィードバックを返すアプリケーションについて述べる．

キーワード：パブリックスピーチ，プレゼン支援，開口，機械学習

Keywords：public speech, presentation training, mouth shape, machine learning.

１．はじめに

スピーチ支援に関する研究は現在盛んに行われている[1][2]．スピーチ支援の研究の中でもスピーチの評価に関する研究領域がある．パブリックスピーチの分野ではスピーチは言語的伝達能力と非言語的伝達能力の２点から評価されている[3][4]．言語的伝達能力は，発音の明瞭さや抑揚，声の高さ，話す速さなどの音声表現の多様性に基づき評価される．一方，非言語的伝達能力は，身振りやアイコンタクト，表情などの音声以外の表現の豊かさに基づき評価される．これまでのスピーチ支援に関する研究では言語的伝達と非言語的伝達のどちらか一方のみが支援の対象とされてきた．本論文では 話者の開口に着目し，２種類の伝達を共に支援することを目指す．

開口は主に母音を発話するときに行われる．母音の種類によって，明瞭に発話するために適した口の開き具合は異なっている．本研究では，これらの明瞭な母音を発話できる口の開き具合を「良い開口」と定義する．話者が良い開口であるとき，聴衆は口の動きの変化を大きく感じるため，話者の表情が豊かであるという印象を得やすくなる．また聴衆は話者の口の動きがはっきりと見えるため，上手く聞き取れなかった場合であっても話者が何と発言したかを推測しやすくなる．そのため，良い開口は発音を明瞭にするという点で言語的伝達能力を補助するだけでなく，表情を豊かにするという点で非言語的伝達能力をも補助するため，スピーチにおける開口は非常に重要であるといえる．

しかしスピーチを支援するこれまでの研究領域では，話者の開口具合はあまり注目されてこなかった．筆者は学士時代に英語スピーチのサークルに所属しており，全国大会の決勝に数回出場する中で，スピーチにおける開口の重要性に気が付き，良い開口を支援したいと考えた．ここで，筆者の主観では，良い開口ができた際は首の筋肉が上手く使用できている．首には広頚筋という筋肉が存在する．広頚筋は口角を横方向に動かすときと顎を下に引くときに隆起するという特徴を持つ．この性質から良い開口のときは広頚筋が隆起すると推測した．

そこで筆者らは首の筋肉の隆起を測定することにより，良い開口での母音発話が認識できるのではないかと考えた．具体的には表情筋を動かす首の筋肉の隆起をフォトリフレクタで測定し，そのデータを機械学習にかけることによって，良い開口での母音の発話を判断する仕組みを提案する．本稿では提案システムの詳細と精度計測実験の結果，及び良い開口での母音の発話のフィードバックを返すアプリケーションについて述べる．

２．システム

広頚筋を接触型センシングで測定し，良い開口での母音発話を認識するシステムを提案する．バンド状に

配置したフォトリフレクタモジュールを首に装着し，広頸筋の隆起を測定することで発声したどのような開口で母音発話をしたかを認識するシステムを実装した．実装にあたりスピーチ時に話者への装着の違和感を減らし，開口の動作を阻害しないために，デバイスの装着場所として口周りでなく首を選択した．システム構成を図 1 に示す．機械学習には SVM を用いた．

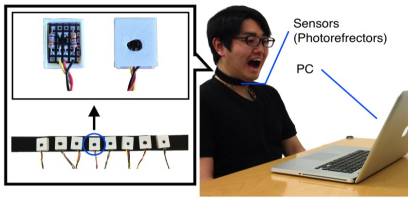

図 1 システム構成

3．精度実験

本システムの精度を調べるために被験者毎と被験者間での認識率を算出した．

【参加者】20 歳代の男性 5 名，女性 1 名で行った．

【認識セット】ニュートラルな表情での無発話，ニュートラルな表情での母音発話 5 種，良い開口での母音発話 5 種の計 11 種類で行った．

【手続き，課題】課題は，実験者が指示した開口での発話をすることであった．実験者が指示した開口での発話とシステムが判定した発話状態の正否を調べた．

【結果】

10 分割交差検証を用いて，各開口での被験者毎の認識率を算出した．分析を行った結果，被験者毎の認識率の平均は全体で 78.5%（SD = 9.8）であり，ニュートラルな表情で 68.7%（SD = 7.4），良い開口で 83.8%（SD = 5.5）であった．開口毎における被験者毎の認識率を表 1 に示す．

本システムの一般性を調べるため，6 人の被験者から収集した実験データを全て合わせて学習とテストを行い，被験者間の認識率を算出した．

被験者間の認識率を表 2 に示す．被験者間の認識率の平均は全体で 31.4%（SD = 20.1）であり，ニュートラルな表情で 30.3%（SD = 19.8），良い開口で 21%（SD = 21.0）であった．

表 1　被験者毎の認識率

		Neutral Mouth Shape						Good Mouth Shape					
	Actual / Predict	normal	a[%]	i[%]	u[%]	e[%]	o[%]	a[%]	i[%]	u[%]	e[%]	o[%]
Neutral Mouth Shape	normal	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	a	0.3	58.3	0.0	7.0	11.0	17.2	4.5	0.0	0.0	0.0	0.0
	i	0.0	1.7	80.8	0.0	0.0	2.7	1.7	3.7	1.7	0.0	1.7
	u	0.0	6.2	4.3	67.2	11.8	2.3	0.0	0.0	1.7	0.0	0.0
	e	1.7	5.8	2.5	4.5	71.7	3.3	1.7	0.0	0.0	2.2	0.0
	o	0.0	13.2	0.0	2.7	4.7	65.8	5.3	0.0	3.3	0.0	1.7
Good MouthShape	a	0.0	4.7	1.7	0.0	0.0	3.2	84.8	1.7	1.7	1.7	0.7
	i	0.0	3.8	0.0	0.0	0.0	0.0	0.0	84.8	0.0	3.3	8.0
	u	0.0	0.0	0.0	4.0	0.0	0.0	0.0	3.3	82.8	3.2	6.7
	e	0.0	0.0	0.0	1.7	2.3	0.0	1.7	1.0	1.7	92.0	0.0
	o	0.0	0.0	1.7	0.0	0.0	5.0	0.0	9.3	9.2	0.0	74.8

表 2　被験者間の認識率

		Neutral Mouth Shape						Good Mouth Shape					
	Actual / Predict	normal	a[%]	i[%]	u[%]	e[%]	o[%]	a[%]	i[%]	u[%]	e[%]	o[%]
Neutral Mouth Shape	normal	44.2	3.3	0.0	25.8	1.7	0.0	0.0	0.0	16.7	8.3	0.0
	a	0.0	10.5	7.3	6.3	5.5	3.3	45.2	0.0	16.7	0.0	0.0
	i	0.0	0.0	48.0	3.8	26.3	1.5	10.3	8.3	0.0	0.0	1.7
	u	2.8	0.0	2.0	28.8	19.2	5.7	26.7	0.0	11.0	1.7	0.0
	e	0.0	0.2	5.0	28.2	35.7	0.0	14.3	0.0	16.7	0.0	0.0
	o	5.7	1.2	2.8	10.2	21.2	11.0	27.7	0.0	19.5	0.8	0.0
Good Mouth Shape	a	3.3	3.5	5.0	2.0	13.5	1.8	47.5	0.0	16.7	6.7	0.0
	i	0.0	0.0	0.2	18.0	7.0	0.0	6.5	42.7	10.7	0.0	15.0
	u	25.3	0.0	0.0	8.0	17.0	1.7	4.2	15.5	6.2	20.7	1.5
	e	3.7	0.0	6.7	19.7	35.2	1.7	0.0	3.3	18.2	3.3	8.3
	o	11.7	0.0	2.5	17.5	5.7	0.0	1.7	18.2	10.0	10.0	21.8

4．アプリケーション

話者の発話した音素のうち母音部分のみを認識し，良い開口で発話されたかどうかをディスプレイでフィードバックするアプリケーションを制作した．図 2 に実装したアプリケーションを示す．使用方法は次の通りである．このフィードバックは実時間で行われる．

(1) まずユーザは机の前の椅子に座り，首の位置に本システムを装着する．
(2) 各母音にてニュートラル表情で無発声の状態とニュートラルな表情での開口と良い開口を本システムに学習させ，それらの教師データを作成する．
(3) ユーザは机上に配置されたディスプレイに向かってスピーチの原稿を読む．
(4) ユーザはディスプレイに表示された母音発話認識の結果を見る．認識の結果は良い開口のときにのみ母音発話を判定し，ニュートラルな表情のときは母音を判定しない．例えば良い開口で/ki/と発声した場合，良い開口での/i/と判定される．

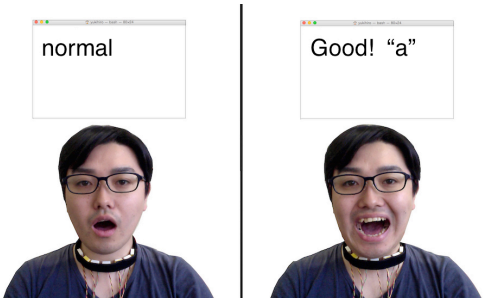

図 2　フィードバック

参考文献.

[1] 張鑫磊，味八木崇，暦本純一：“WithYou：音声認識を用いたインタラティブシャドーイングコーチ”，インタラクション 2016，2016.

[2] H. Trinh , R. Asadi , D. Edge , T. Bickmore, RoboCOP: A Robotic Coach for Oral Presentations , Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, v.1 n.2, p.1-24, June 2017

[3] Swai Johar. Emotion, Affect and Personality in Speech. Springer, p.1-6, 2016.

[4] Rodrigues IG , “Verbal and nonverbal signals in face-to-face interaction: a theoretical framework for a holistic micro-analysis (The example of a parenthesis)” . Interacting Bodies, Lyon, 15-18 June 2005 .