

加速度センサを用いた声帯振動による母音識別

Japanese Vowel Discrimination by Throat Vibration Using Acceleration Sensor

本田彰吾^{†1, a)} 中沢実¹ 高野佐代子

Shogo Honda^{†1, a)} Minoru Nakazawa¹ Sayoko Takano

喉頭がんや舌がんによる声帯の摘出手術、ALS(筋萎縮性側索硬化症)などが原因でコミュニケーションに難がある患者に対して脳波、顔認識など様々な方法でコミュニケーションを促すデバイスやシステムが研究されている。その中でも、発声時脳波による単音節識別を行う研究がある。しかし、従来の研究では単音節の識別精度は61%であり、脳波計は使用するのに時間がかかるだけでなく、環境によって多くのノイズを拾いやすいという欠点があるため実用化が難しいと考えられる。そこで本研究では、脳波計と比べて比較的測定が容易であり、環境ノイズに左右されない加速度センサによる喉からの振動信号を用いた母音の線形2クラス分類を行った。その結果、「あ」と「い」の識別率は75%、「あ」と「う」では85%、「あ」と「え」では55%、「あ」と「お」では70%となり、加速度センサによって母音の判別ができる可能性を得た。

Devices and systems to facilitate communication with patients who have difficulties in communication due to laryngeal and tongue cancer, ALS (amyotrophic lateral sclerosis), etc. are being studied using various methods such as EEG and face recognition. One of them is the study of monosyllable discrimination by EEG during speech. However, the discrimination accuracy of monosyllables is only 61% in previous studies, and EEG is not practical because it takes a long time to use and is likely to pick up a lot of noise depending on the environment. In this study, a linear two-class classification of vowels was conducted using the throat vibration signal from the accelerometer, which is relatively easy to measure and unaffected by environmental noise, compared to EEG. As a result, the discrimination accuracy of /a/ and /i/ was 75%, 85% for /a/ and /u/, 55% for /a/ and /e/, and 70% for /a/ and /o/, indicating the possibility of discriminating vowels by accelerometers.

1. はじめに *

近年、脳波や顔認識など様々な方法でコミュニケーションを促すデバイスやシステムが研究されている[1][2]。これらの目的は、プライバシーを考慮した公共でのコミュニケーションの促進だけではなく、言葉を発することに難があるALSの患者さんなども対象にしている。その中でも、舌がんの摘出手術を施したことによって舌を失い、舌を使う音節がうまく発声できない患者さんも対象とされる。また、2017年には口腔・咽頭癌の患者数は約22,800人に及んでいる[7]。

そのような舌を失ったことで話すことに難がある患者に適応が考えられる先行研究として、発声時脳波から単音節を識別する試みとして新田ら[3]の研究がある。この研究では、被験者に18種類の単音節を発声させ、その時の被験者の発声時脳波から被験者が実際に発した言葉を推測することができるかを検証している。しかし、結果として単音節の識別精度は61%と実用化するにはまだ厳しい数値である。また、脳波計は測定に時間がかかるだけでなく、環境によって多くのノイズを取得しやすいという欠点がある。

そこで本研究では、脳波計と比べて比較的測定が容易であり、かつ直接頸部から振動信号を加速度センサによって取得することで、環境ノイズに左右されにくい母音の高い識別精度を目指す。これを実現することで、将来的にEEGと併用して、多くの患者さんが利用することが可能になると期待できる。

2. 実験方法

2.1 被験者

頸部からの振動信号測定実験の被験者として、男子大学院生(22)である筆者が自ら測定を行った。

本研究では、ただ1人の被験者に対して実験を行なったが、将来的にはシステムの汎用性を確認するためにも複数名の被験者による実験を予定している。

2.2 実験で用いる単音節

実験で発声する内容は、単音節母音「あいうえお」(/a/, /i/, /u/, /e/, /o/)を用いた。

このような加速度センサを用いた母音の識別は、これまで試みられてなかったため、今回の実験では識別対象を母音のみにした。

*1 情報処理学会

IPSI, Chiyoda, Tokyo, 101-0062, Japan.

†1 金沢工業大学 工学研究科 情報工学専攻, ロチェスター工科大学 情報工学専攻

Presently with Kanazawa Institute of Technology and Rochester Institute of Technology MS Computer Engineering.

a) c6000589@planet.kanazawa-it.ac.jp

*1 <https://www.atr-p.com/products/TSND121.html>

*2 <https://invensense.tdk.com/wp-content/uploads/2015/02/MPU-6000-Datasheet1.pdf>

*3 <https://www.atr-p.com/products/Altima.html>

*4 https://tomari.org/main/java/audioapi/audio_sanyousho.html

将来的には、この母音識別をさらに発展させ、母音・子音を識別対象にする予定である。

2.3 データ測定

振動信号、及び音声信号は被験者が単音を発声している時に収録された。

振動信号は TSND121（小型無線多機能センサ）に内蔵された加速度センサ (MPU-6000) によって、Z 軸加速度をサンプリング周波数 1000Hz、加速度レンジ±2G で図 1 のように設置して、測定した (*1)。加速度センサのその他詳細が記させたデータシートの抜粋 (*2) を図 2 に示す。

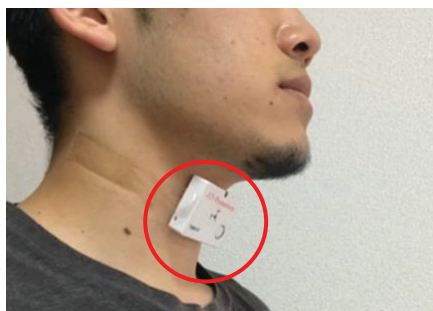


図 1 加速度センサの設置位置

PARAMETER	CONDITIONS	MIN	TYP	MAX	UNITS	NOTES
ACCELEROMETER SENSITIVITY						
Full-Scale Range	AFS_SEL=0		±2		g	
	AFS_SEL=1		±4		g	
	AFS_SEL=2		±8		g	
	AFS_SEL=3		±16		g	
ADC Word Length	Output in two's complement format		16		bits	
Sensitivity Scale Factor	AFS_SEL=0		16,384		LSB/g	
	AFS_SEL=1		8,192		LSB/g	
	AFS_SEL=2		4,096		LSB/g	
	AFS_SEL=3		2,048		LSB/g	
Initial Calibration Tolerance			±3		%	
Sensitivity Change vs. Temperature	AFS_SEL=0, -40°C to +85°C		±0.02		%/°C	
Nonlinearity	Best Fit Straight Line		0.5		%	
Cross-Axis Sensitivity			±2		%	
ZERO-G OUTPUT						
Initial Calibration Tolerance	X and Y axes		±50		mg	1
	Z axis		±50		mg	
Zero-G Level Change vs. Temperature	X and Y axes, 0°C to +70°C		±35		mg	
	Z axis, 0°C to +70°C		±60		mg	
SELF TEST RESPONSE						
Relative	Change from factory trim	-14		14	%	2
NOISE PERFORMANCE						
Power Spectral Density	@10Hz, AFS_SEL=0 & ODR=1kHz		400		μg/√Hz	
LOW PASS FILTER RESPONSE						
Programmable Range		5		260	Hz	
OUTPUT DATA RATE						
Programmable Range		4		1,000	Hz	
INTELLIGENCE FUNCTION INCREMENT						
			32		mg/LSB	

図 2 加速度センサのデータシート

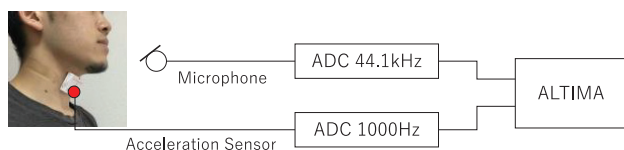


図 3 各信号測定の実験フロー

一般にヒトの声は、喉頭にある声帯の開閉によって生じる音波を音源とする。喉頭は首の中央前方に位置する甲状軟骨（喉仏）およびその周辺軟骨からなる器官である。甲状軟骨前方内壁に声帯前部が付着し、左右 1 対の披裂軟骨に声帯後部が付く[4]。そのため、本実験では図 1 のように甲状軟骨側壁にセンサを設置し、声帯の開閉によって振動が生じることを想定し加速度を計測した。

音声信号は、SONY Dynamic Microphone F-720 を用いてサ

ンプリング周波数 44.1kHz で測定した。最後にこれらの取得したデータを TSND121 専用のソフトウェア「ALTIMA」(*3) に記録した。そのセットアップを図 3 に示す

また、測定は実験に支障が出ないほどの騒音レベル下で行われた。

2.4 前処理

図 4 に時系列における母音「あいうえお」発声時の振動信号、図 5 には各母音の加速度信号を拡大した図を示す。この比較図より、各母音の波形に違いが見られ、それぞれの特徴が加速度センサによって取得できていることがわかる。

振動などの時系列データを識別する場合、それらを周波数帯域に変換してから取り扱うことが多いため、ここでは識別する前に必要な処理について述べる。

- 0.3 秒間の振動信号を、各母音から切り取った。
- 切り取った各信号に対して、ハミング窓を 300 点で掛け、周波数帯域に変換させるために時系列データから AR モデルのパラメータを推定するユールウォーカー法を用いてスペクトル解析を行った。
- スペクトル解析された各信号に見られる第一高調波と第二高調波を推定した (図 6)。そして、これを識別に用いる特徴量とした。

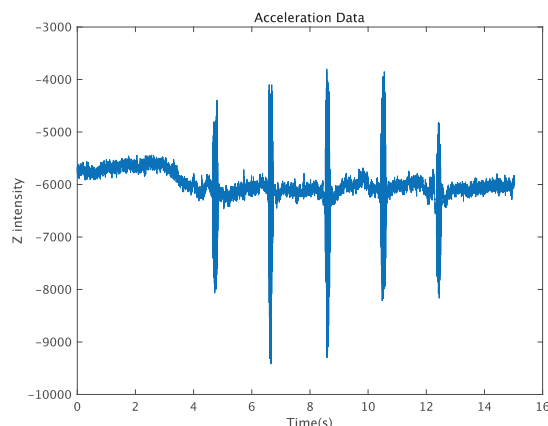


図 4 /a/, /i/, /u/, /e/, /o/ の振動信号

3. 実験手順

3.1 測定について

2.2 で述べたように、実験で発声する内容は、単音節母音「あいうえお」 (/a/, /i/, /u/, /e/, /o/) を用いた。各母音の発声の音高さが揃うように、各母音の発声前に目安となる複合音(*2)を被験者に聴いてもらった上で収録を行った。これを 10 回繰り返した。以下の図 7 に測定手順を示す。

3.2 ラベリング

信号処理によって得られた各母音の第一高調波と第二高調波に正解ラベル (1:あ, 2:い, 3:う, 4:え, 5:お) を割り当てた。

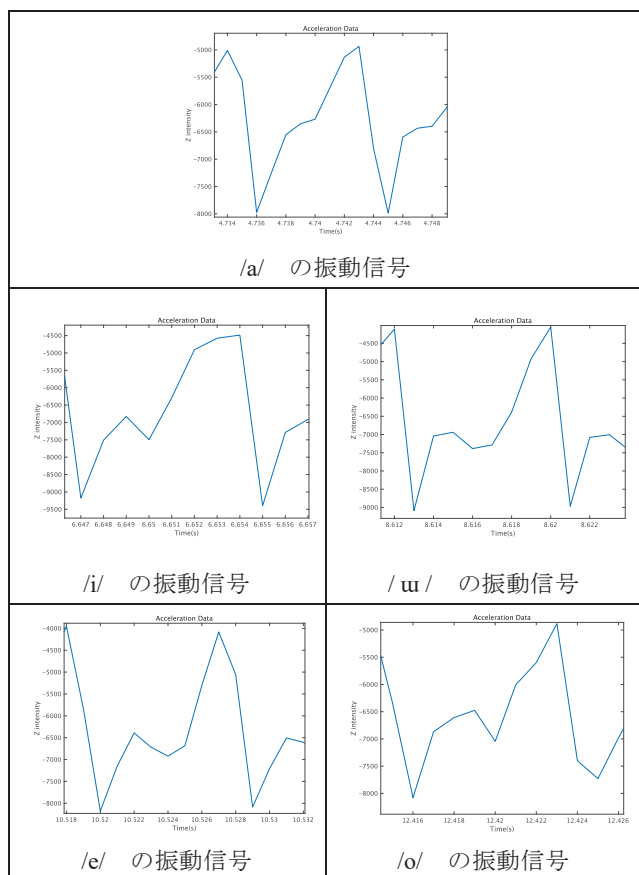


図5 各母音の加速度信号の拡大図

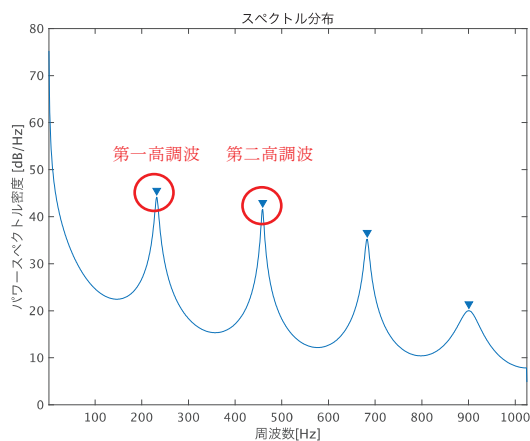


図6 母音「あ」発生時の声帯振動スペクトル分布



図7 測定手順

3.3 識別モデル

実験により得られたデータは、合計で50サンプル（各母音10サンプル）であった。

ここで本研究の目標である、各母音が振動信号によって識別可能かどうかを検証するために、今回は識別パターンとして「あ」とそれ以外の母音の振動信号における2クラス分類にした。これらの取得したデータから各母音の特徴量を特徴空間で比較したものを図8～11に示す。また図中では、前処理で得られた第一高調波と第一高調波をそれぞれF1, F2と表す。

図8の特徴空間ベクトルの比較図から、各母音の特徴に偏りが見られ、線形識別可能であると考えたため、入力パターンに対するマージン最大化によって線形識別を行うSVM(Support Vector Machine)を識別器として用いた。

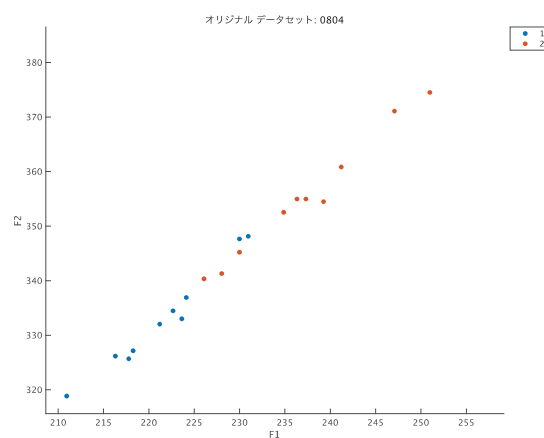


図8 「あ」と「い」特徴空間ベクトル

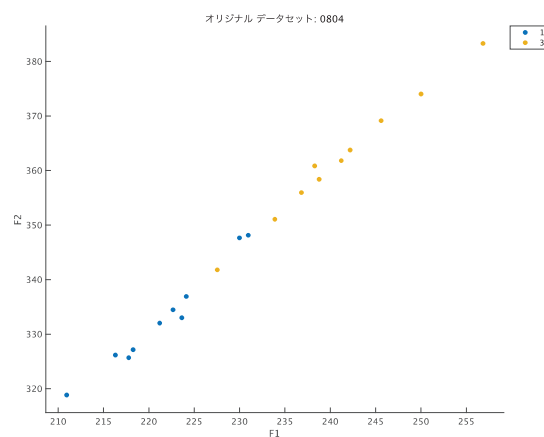


図9 「あ」と「う」特徴空間ベクトル

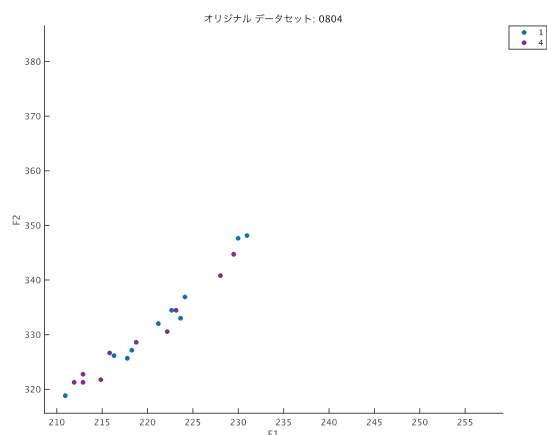


図 10 「あ」と「え」特徴空間ベクトル

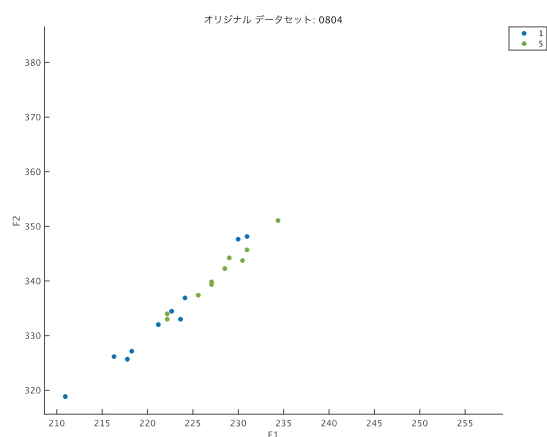


図 11 「あ」と「お」特徴空間ベクトル

今回の実験では、少量のデータしか取得できなかった為、全ての入力パターンを訓練とテストで用いる 5 分割交差法を用いた。これらの処理を MATLAB2020a が提供しているアプリケーション「分類学習器」によって行なった。

4. 実験結果

上記の識別方法で述べた識別器 (SVM: Support Vector Machine) を使って、各母音の 2 クラス分類を行った。その結果、「あ」と「い」は識別率 75%、「あ」と「う」では 85%、「あ」と「え」では 55%、「あ」と「お」では 70% という結果となった (表 1)。

表 1 各音節の識別精度の結果

パターン	識別精度[%]
「あ」 vs 「い」	75
「あ」 vs 「う」	85
「あ」 vs 「え」	55
「あ」 vs 「お」	70

5. 考察

本研究では、加速度センサを用いて頸部から各母音の振動信号を取得し、第一、第二高調波を抽出し比較した結果、各音韻に特徴が見られた。さらに、これらを線形識別器 SVM (Support Vector Machine) によって 2 クラス分類した結果、「あ」と「い」では 75%、「あ」と「う」では 85%、「あ」と「え」が 55%、「あ」と「お」では 70% の識別結果を記録した。

「あ」と「え」、「あ」と「お」の識別率が低い原因として、音声の第一、二フォルマントにおける分布 [5] の位置関係が若干類似していること、つまり周波数に大きな違いがないことが挙げられる。もしくは、学習に用いたサンプル数が少ないことが挙げられる。

この精度をさらに向上させ、識別できる単音節の種類を増やすことで、脳波計と比べて測定に時間を掛けずに、かつ環境ノイズに影響されない単音節の識別が可能となる。

6. 終わりに

今回の研究では母音に限定した識別を行ったが、今後の研究の取り組みとしては、母音と子音の識別を目指す。また、識別精度が低かった「あ」と「え」の識別精度の改善に努める。

さらに、地声裏声また母音の判別に関する研究動向 (表 2) として、音声における地声裏声や母音の識別は可能であるが、それ以外の対象として挙げられる、喉頭の接触程度を電流によって測定する EGG (Electroglottography) [6] では、地声と裏声の判別はできると知られているが、母音識別は検証されていない。本研究では、この EGG では困難と考えられる母音の判別を加速度センサによって判別できる可能性を得た。今後の研究として、加速度センサにおける地声裏声の判別の可否を検証する。

表 2 各手法における地声裏声/母音の判別の可否

	地声裏声	あいうえお
EGG	○	?
加速度	?	?
音声	○	○

参考文献

- [1] B. Denby, T. Schultz, K. Honda, T. Hueber, J. M. Gilbert, and J. S. Brumberg, “Silent speech interfaces,” *Speech Commun.*, vol. 52, no. 4, pp. 270–287, Apr. 2010, doi: 10.1016/j.specom.2009.08.002.

- [2] N. Kimura, M. C. Kono, and J. Rekimoto,
“Sottovoce: An ultrasound imaging-based silent
speech interaction using deep neural networks,”
Conf. Hum. Factors Comput. Syst. - Proc., 2019,
doi: 10.1145/3290605.3300376.

- [3] 神崎卓丸ほか：発話時と想起時の脳波による日
本語短音節の比較，日本音響学会春季研究発表会，
2017

- [4] 榊原健一，“発声と声帯振動の基礎，” 日本音響学
会誌, pp. 73--79, 2015.

- [5] 田淵行則，前川喜久雄，壺菌晴夫，本多清
志，白井克彦，中川聖一，岩波講座 言語の
科学. 2，岩波書店，1998

- [6] 石毛美代子ほか：Electroglottography，音声
言語医学，1996

- [7] 国立がん研究センターがん情報サービス「がん
登録・統計」（全国がん登録）