

筋電信号を用いた無発声音声認識 —一定常状態における母音の識別—^{*1}

真鍋 宏幸 平岩 明 杉村 利明^{*2}

NTTドコモ マルチメディア研究所^{*3}

1 はじめに

通常の音声認識技術では、発話者が発声した音響信号からスペクトル包絡やスペクトルの微細構造を抽出する等の方法で、発話情報を検出している。しかし発声による音響信号を信号処理する音声認識では、当然のことではあるが発声による音響情報が入力されない限り、発話情報を検出することができない。そして音響情報を利用している為に、周囲の騒音や雑音の影響を受けやすいという問題点もある。またオフィスや図書館等の静けさが要求される場において、発話することが周囲の人々の迷惑となるために、通常の音響信号を用いた音声認識では利用することができなかった。近年急速に普及した携帯電話に関して、電車内等の公共機関内での利用自粛が求められている。この原因は、ペースメーカー等への電磁波的影響の問題だけでなく、発話による周囲の人々への迷惑というマナー的な問題があるものと考えられる。

そこで本稿では、発声することなく音声認識可能なインタフェースとして、筋電信号（EMG:electromyography）を用いた無発声音声認識について提案を行う。EMGを用いて動作の認識を行う研究は従来から行われており、平岩らは手首の少数チャネル皮膚表面EMGを、ニューラルネットで認識し、10自由度の指関節角度を認識している^[1]。また、Sugieらは口周辺の筋から測定されるEMGを測定し、それぞれのEMGに対して閾値との交差回数をカウントす

ることによって、母音の認識を行い、平均で60%程度の認識率を得ている^[2]。

本稿では、EMGを用いた無発声音声認識の提案を行い、及びその有効性を確かめるために行った実験の結果について述べる。

2 認識システム概要

発話者が発声する動作を考えると、有発声／無発声に関わらず、口周辺の筋が協調して動作していることを確認することができる。このため、発声時の口周辺の筋の活動を測定することができれば、無発声でも音声認識が行えることが考えられる。本稿で提案する無発声音声認識システムでは、発声時の動作に大きく寄与していると考えられる1：口輪筋、2：大頬骨筋、3：顎二腹筋の3つの筋のEMGを測定する。

そして、各筋から得られるEMGをそれぞれ1つのチャンネルとし、各チャンネルの平均振幅を算出する。そこで算出された平均振幅を3層ニューラルネット（以下、NN）に与え、無発声時における5母音の音声認識を行う。Fig.1に筋の配置とシステムの構成を示す。

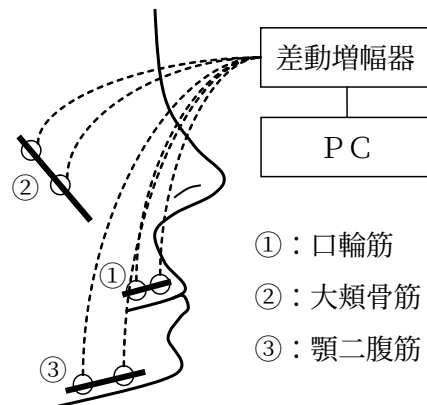


Fig.1: 筋の配置とシステム構成

3 実験概要

本稿で提案する認識システムの有効性を確認

^{*1} Speech recognition with EMG

-Vowel discrimination in a steady state-

^{*2} Hiroyuki Manabe, Akira Hiraiwa, Toshiaki Sugimura

^{*3} NTT DoCoMo Multimedia Laboratories

するために実験を行った。EMGの測定はノイズの影響を除くためにシールドルーム内で行った。Fig.1 に示す筋の皮膚上に銀塩化銀電極を貼り、安静時（非発声時）及び“a”、“i”、“u”、“e”、“o”の5母音を無発声で発声した時の6状態におけるEMGを測定した。本実験では口の形状に変化がない定常的な母音発声時のEMGを測定するものとした。表面電極に誘発されたEMGは、差動増幅器（ローカットフィルタ時定数：1ms、ハイカットフィルタ：330Hz、ゲイン：1000倍）で増幅してからサンプリングレート1kHz/16bitでAD変換し、PCに取り込んだ。なお、それぞれの状態について5秒間の発声を10回繰り返すこととした。

PCに取り込んだEMGは400ms毎に切り出し、平均振幅として二乗平均平方を計算した。そして、口輪筋・大頬骨筋・顎二腹筋に対応した3チャンネル分のEMGの二乗平均平方を入力層3・中間層6・出力層5のNNに与える。5つの出力層は“a”、“i”、“u”、“e”、“o”にそれぞれ対応しており、例えば“a”が発声された場合には“a”に対応した出力層のみが発火するものとし、安静時には全ての出力層が発火しないものとして、NNの学習を行った。学習に用いる教示データはPCに取り込んだEMGの中から、半分を用いて行い、残り半分のEMGをテストデータとして認識実験を行った。なお、NNの学習回数は3000回とした。

4 実験結果

NNにテストデータを与えたときの認識結果をFig.2に示す。認識は、それぞれの母音に対応した出力層からの値がある閾値以上の場合にその母音であると認識し、全ての出力層がある閾値以下の場合には安静時であると認識するものとした。なお状態を認識できなかった場合にはerrorとして処理をした。この結果から、“i”、“u”以外の状態についてはほぼ100%認識でき

		認識結果						
		安静時	"a"	"i"	"u"	"e"	"o"	error
状態	安静時	100	0	0	0	0	0	0
	"a"	0	100	0	0	0	0	0
	"i"	0	0	73	0	27	0	0
	"u"	0	0	0	82	0	17	1
	"e"	0	0	0	0	100	0	0
	"o"	0	0	0	1	0	99	0

Fig.2：認識結果

ていることがわかる。また“i”、“u”についても70%以上の認識率が得られている。

次にFig.2に示す結果が再現できるか確認するために、NNに与える初期値をランダムに与えて、学習及び認識のサイクルを1000回繰り返した実験を行った。その結果、Fig.2に示す結果とほぼ同様な結果が毎回得られることを確認した。

さらに、EMGを用いた無発声音声認識がノイズの多い電車内でも行えるかどうか確認するために、JR山手線の動力車内におけるEMGの測定実験を行った。その結果、電車内においてもEMGが測定できることを確認した。特に、電極とアンプが一体となったアクティブ電極を用いた場合には、ほとんどノイズの乗らないEMGを測定することができた。

5 まとめ

本稿ではEMGを用いた無発声音声認識の提案を行い、その有効性を確認するための実験を行った。その結果、“i”、“u”以外の母音についてはほぼ100%の認識率を得ることができた。また、電車内においてもEMGが測定できることを確認し、EMGを用いた無発声音声認識が電車内等のノイズの多い環境でも利用できることの可能性を示すことができた。

参考文献

- [1] 平岩, 内田, 下原, 曾根原, “筋電操作ハンドの制御のための皮膚表面筋電信号のニューラルネットによる認識,” 計測自動制御学会論文集, Vol.30, No.2, pp.216-224, 1994.
- [2] N. Sugie, K. Tsunoda, “A speech prosthesis employing a speech synthesizer,” IEEE Transactions on Biomedical Engineering, Vol. BME-32, No.7, pp.485-490, 1985.