

平成 26 年度 学士論文

**黙声認識における発話文章特徴量による
認識結果の校正**

平成 26 年 2 月 19 日

学籍番号 07106005

井上 雄介

指導教員 三浦 元喜 准教授

九州工業大学大学 工学部 電気工学科

電子通信システムコース

目次

第1章	はじめに	5
1.1	黙声認識	5
1.2	背景	6
1.3	目的	6
第2章	関連研究	8
2.1	筋電位情報を用いた黙声認識システム	8
2.2	黙声認識の問題点	8
2.3	黙声認識における五十音認識	9
2.4	子音情報からの単語予測	10
2.5	確率モデルによる文章の特徴量推定	10
第3章	文章校正システムの作成	12
3.1	文章校正システムの流れ	12
3.2	黙声認識における変化特徴量	13
3.3	形態素解析	14
3.4	文章破壊	14
3.5	文章校正	14
第4章	校正実験	16
4.1	実験方法	16
4.2	文章データ	16
4.3	評価方法	16
第5章	実験結果	18
5.1	文章破壊と文章校正	18
5.2	未知語	24
5.2.1	校正の精度に対する未知語の影響	24

5.2.2	未知語と補正率	25
5.2.3	未知語の種類	26
第 6 章	結論	27
6.1	実験のまとめ	27
6.2	今後の課題	27
第 7 章	謝辞	29

概要

近年音声認識技術は急速な発展を遂げており、主に一般的な場面ではスマートフォンの操作の一つとして、大きく活躍している。手を利用することができない場面において、音声認識技術は高い利便性を発揮する。例えば料理中にレシピを見たり、運転中の際に音楽プレーヤーで楽曲を再生したりなど、手がふさがっている中で、スマートフォンの操作が要求される場面は往々に見られる。一方、音声を伴わない発話認識技術、すなわち黙声認識技術が利用されることは音声認識技術の利用状況と比べると非常に少ない。その主な理由の一つとして、黙声認識の精度の低さが挙げられる。我々は一般的な黙声認識技術の普及を目標に、発話文の単語特徴量に基づいて、認識結果の校正する方法を考えた。形態素解析エンジン Mecab を利用し平仮名化した文章を、プログラム内で特徴量に基づき変化させたところ、元のまま正しく出力された単語の割合は約 13.7%であった。続いて、変化した文章を発話文の単語特徴量に基づき校正を行った。交差検証法により校正実験を行った結果、正しく出力された単語の割合は約 61.3%であり誤って認識された単語の約 55.2% を校正することができた。

第1章 はじめに

近年、スマートフォンの普及により、音声認識技術が利用される機会が以前と比べて多くなった。キーボード入力を用いた操作と比較すると、音声認識システムにおける音声入力インタフェースの特徴は二つある。

一つは発声という日常的な行動による直観的な操作ができるという点と、素早く目的までたどり着けるといふ点である。しかし声を出せない状況、例えば図書館や映画館などの公共の場においては、周囲への配慮や秘匿性に関する問題が発生し、利用することが憚られてしまうという場合もある。

1.1 黙声認識

この問題に対して、音声情報を伴わない音声認識技術の研究が進んでいる。一般的にこの技術は「黙声認識技術」と呼ばれている。この技術は音声情報ではなく、画像認識を利用して口の動きを読み取るもの、すなわち読唇を行うものや、皮膚に電極を装着し、筋肉の収縮による筋電位の変化を読み取るものなどいくつか手法がある。この技術を用いることで、音声認識技術の利用制限の幅を広くすることができるため、黙声認識技術の発展は音声認識技術の発展に繋がることになると思われる。

しかし、現在の研究では文章単位で認識を行うまでには至らず、母音、もしくは限られた単語の認識にとどまっているという現状である。現在、実際に音声認識のように利用されることはほとんどなく、実用に至るまでは更なる認識精度の向上と誰でも簡単に利用するための環境が必要となる。

しかし、黙声認識において認識率を上げることは容易ではなく、また、大掛かりな準備が必要な手法を以てしても認識精度を大きく上げる手法といったものは決定付けられてはいない。ここで、本論文ではどうすれば認識率を上げられるかではなく、現在認識した結果を用いて文章を正しい日本語として校正す

る手法について考えた。

1.2 背景

この研究を行った背景として、筋電情報を用いた音声認識においては、母音の認識は子音と比べてかなり容易であるということがある。なぜかという、日本語母音は口唇の形状にそれぞれ特徴があり判別しやすいが、子音認識においては口唇周辺だけでなく舌の動きも大きく関わってくるからである。そういった点から、口の動きを読み取る黙声認識における認識結果はある特徴を持っているのではないかと考えた。例えば、母音'a'を持つ発音は、認識結果が誤っている場合でも母音'a'を持つ発音と認識されやすいのではないかとということである。

筋電情報を読み取る手法として、電極を用いることが多い。喋る人物の口唇周辺に複数の電極を張り付けて、口唇周辺にある筋肉の動きを読み取る方法が主である。しかし、筋肉や皮膚には人物ごとに特徴の違いが存在する。例えば同じ言葉話すのに、筋肉の動きは細かくみれば人それぞれであるし、また発汗量も違うため、取得できるデータを別の人物に適用することは容易ではない。そういったデータに関しての適応化が求められるため、黙声認識の認識精度を上げるのは困難である。

1.3 目的

黙声認識の認識精度を上げることには解決すべき課題が非常に多いため、我々は黙声認識の認識方法やデータの適応化などの影響が少ない、認識結果の日本語的校正という手法を提案する。この手法であれば人物ごとの適応化を行わずに、認識結果を結果的により正しく出力できる。

認識結果にある程度の収束が見られるのであれば、結果の文章校正も可能であると考えられる。すなわち認識結果が日本語として意味を成していないとしても、認識結果の特徴量を利用して文章を訂正、補完することができれば、文章そのものの認識を行えるということである。本研究では、発話文章特徴量に基づいて、筋電情報を用いた音声認識に基づいた結果を利用し、文章がどれほどまで正しく復元できるかという調査を目標としている。また同時に、日本語の校正に関する統計とを行った。

以下，音声情報を伴う発声動作を「有声」，伴わないものを「黙声」，と記す．

第2章 関連研究

2.1 筋電位情報を用いた黙声認識システム

「黙声認識」では様々なアプローチでの研究が行われている。例えば、主に研究されている手法の一つとしては筋電位情報を用いた黙声認識システムが挙げられる。

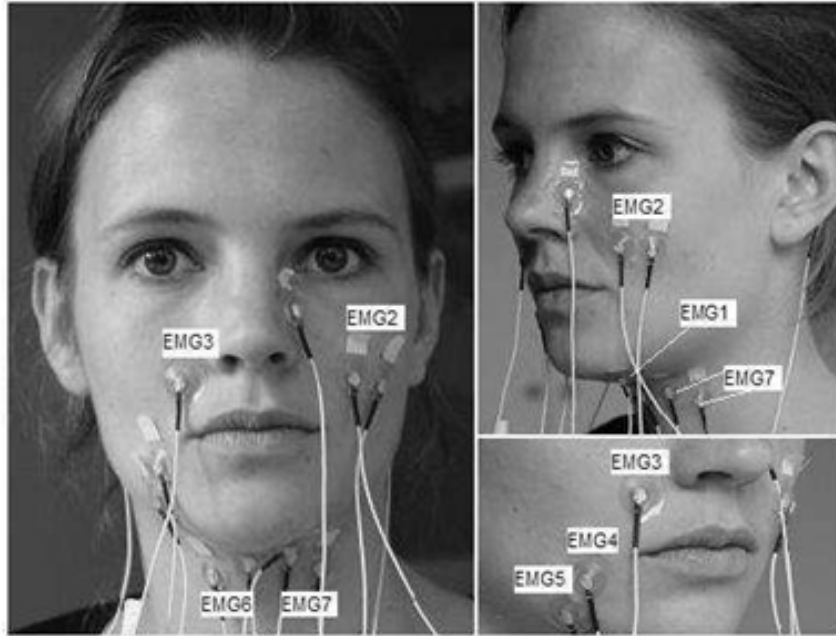
筋電位情報とは、電極を用いて顔の筋肉の動きを信号化したものである。筋肉が動くことにより、任意の複数点において電位差が発生し、その情報から筋肉の動きを読み取るといったものになっている [2.1]。また、どの筋肉の動きを読み取るかという点についても研究が進んでおり、黙声を読み取るための筋電位情報や、連続発声における文字列間境界を読み取るための筋電位情報など、状況に応じて様々な筋電位情報が必要となる。

2.2 黙声認識の問題点

上記に示したように筋電位情報を用いるもの、もしくは画像情報や脳波の信号を用いる手法などがある。しかしこれらはそれぞれ、実用に至るまでの問題点を抱えている。

まず共通していえるのが、認識率の低さである。現在の研究実験では、黙声情報から決まっている数種類のキーワードの内のどのキーワードが発声されているかを特定する、もしくは日本語の母音 (a, i, u, e, o) を一つずつ発声していき認識率を求めるものが多い。これは通常の音声認識と比べた認識率の低さによるものであり、また認識の為の情報量が比較的曖昧であることも原因の一つである。

また、個人における特徴の違いの存在も一つの問題である。脳波、または電極と皮膚との接触部における発汗量は個人差があり、認識率を高めるにはそれらの情報を利用する個人に適応化することが求められる [1]。



出典: Lena Maier-Hein, Florian Metze, Tanja Schultz, and Alex Waibel
SESSION INDEPENDENT NON-AUDIBLE SPEECH RECOGNITION USING
SURFACEELECTROMYOGRAPHY p.2

図 2.1: 筋電位情報取得の様子

これらの問題に対しての手法として、情報処理方法の変更や、装置に対する工夫などがとられているが、本研究では黙声音声認識における認識率の低さという問題に対して、発話文章における特徴量を用いた文章校正というアプローチを提案する。

2.3 黙声認識における五十音認識

黙声認識での認識実験結果は様々であるが、ここでは日本語の五十音それぞれを一音ずつ認識した実験に着目する。大内は隠れマルコフモデル (Hidden Markov Model : HMM) とサポートベクトルマシン (Support Vector Machine : SVM) を組み合わせた認識手法を用いて、筋電位計測における五十音の黙声認識を行っている [2]。結果として、母音 (あ行) のみの認識を行った場合はそれぞれ 90%以上の認識率であったが、五十音全ての認識を行った場合、母音に

関しての認識率は平均 72.0%であり、それ以外の音は「ら行」が最も平均的に高く 50.0%，最も低いもので「た行」は平均 8.0%という認識率であった。

この結果から、母音と子音との認識率に大きな差があることが見て取れる。筋電位における子音の黙声認識は非常に重要な課題であり、高い障壁である。本研究ではこの認識結果を用いて文章を一音ずつ変化させた。また、文章の校正における予測単語候補を絞るための特徴量としても利用した。

2.4 子音情報からの単語予測

佐藤らは子音からの単語予測を用いた入力インターフェースによる、入力速度の向上を図った [3]。佐藤らは変換候補の数を絞る為に HMM と PPM (Prediction by Partial Match) を用いた言語モデルを利用した。また、辞書データに単語同士の関連度を追加することにより、更なる絞り込みを行った。しかし、文脈が切り替わったり利用できない状態にも対応する必要がある為、ユーザの利用情報も併用し、もっともらしい候補を選出することにした。

この結果から、教師あり学習を行わずにに変化した文字列から単語を予測するのは非常に難しいことが見て取れる。本研究では、単語の予測に関しては教師なし学習を行い、文字列の文字数や変化特徴量、そして前後の品詞情報に基づき行うものとした。

2.5 確率モデルによる文章の特徴量推定

白井らは文章全体の名詞の割合を X ，動詞の割合を V ，形容詞類の割合を A ，接続詞類の割合を C とし、多項式係数ベクトル w ，ガウス事前分布 μ, σ を含む確率モデルを作成した。小説、話し言葉、特許、新聞の各ジャンルに与え、それぞれにガウス分布パラメータと多項式ベクトルの学習をさせ、文章の品詞分布特徴量からジャンルを推定した [4]。このように文法や使われる語句にはある程度の特徴量があるため、発話文章に存在する特徴量からどれほどまで単語を予測することができるかを本研究では調査した。

$$\mu = \frac{1}{n} \sum_{n=1}^N x_n, \quad \sigma = \frac{1}{n} \sum_{n=1}^N (x_n - \mu)^2 \quad (2.1)$$

$$w_{i0} = \frac{n \sum_{k=1}^n x_k y_k - \sum_{k=1}^n x_k \sum_{i=k}^n y_k}{n \sum_{k=1}^n x_k^2 - \left(\sum_{k=1}^n x_k \right)^2} \quad (2.2)$$

$$b = \frac{n \sum_{k=1}^n x_k^2 \sum_{k=1}^n y_k - \sum_{k=1}^n x_k y_k \sum_{k=1}^n x_k}{n \sum_{k=1}^n x_k^2 - \left(\sum_{k=1}^n x_k \right)^2} \quad (2.3)$$

第3章 文章校正システムの作成

3.1 文章校正システムの流れ

本研究では誤って出力された文章をどのように校正するか、またどれほど正しく校正できるかを目的としている。校正するための文章は、大内の研究結果[2]に基づき確率的に変化させることで、システム内で作成するものとした。大内は非常に精度の高い黙声認識システムでの五十音別認識を行い、有用な結果を残しているため、こちらから信頼性の高い特徴量が得られると考えた。

我々は、Java上で文章校正を行うためのシステムを作成した。ここではシステムの流れについて説明を行う。初めに元となる文章の形態素解析を行う。形態素解析には形態素解析エンジン Mecab を利用し、それにより文章の読み込み、形態素解析、文章の平仮名化を行う。

次に、読み込んだ文章から辞書を作成する。辞書には単語毎に読み、発音、頻度の情報を持たせる。読みと発音の違いについては、例えば「遠い」、「当然」という二つ言葉の場合、前者の読みは「トオイ」、後者は「トウゼン」となり、二つの言葉には「トオ」と「トウ」の違いが表れる。しかし、発音はどちらも「トー」となり、実際に発声する場合の同一性を持たせる。

次に、基となる文章を変化させる。文章は黙声認識における認識結果の特徴量を基に、文字列をそれぞれ一音ずつ確率的に変化させる。

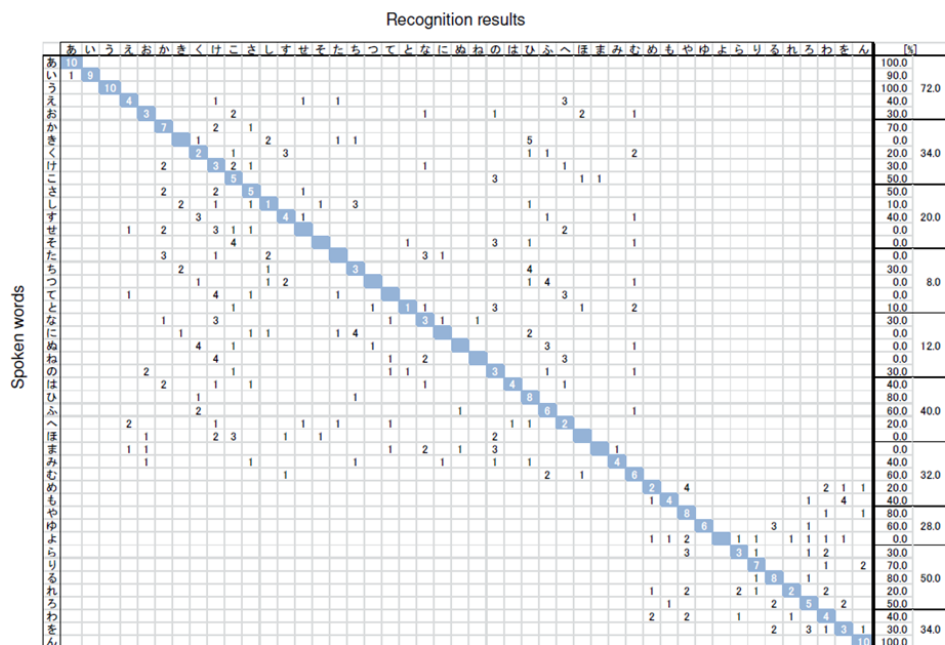
特徴量については3.2において詳しく記述する。

最後に変化させた文章を校正する。文章の校正については、変化させた際と同様の特徴量を基に行う。校正の予測単語については、システム上で作成した辞書から文字数とハミング距離によって候補を選出する。また、単語の更なる絞込みとして認識結果の特徴量を用いることとした。例えば、「あ」が「ん」と認識される可能性が0だとするならば、変化後の文字列において「ん」を持つ箇所において、「あ」を持つ単語は候補から除外するものとした。

以降、話を簡単にするため、入力する元の文章を「原文」、原文を変化させた結果の文章を「破壊文」、変化文を校正した結果の文章を「校正文」とする。

3.2 黙声認識における変化特徴量

黙声認識において認識結果がどのように変化するかという特徴量については、大内の研究結果における五十音の認識率に基づくものとした [2]。大内は五十音を一文字ずつ発声した場合の筋電位情報に基づく黙声認識を行い、それぞれの音がどのように認識されたかを調査した。



出所: Hybrid HMM SVMを利用した筋電位に基づく日本語黙声認識 p.35

図 3.1: 黙声認識システムにおける五十音の認識結果

本論文ではその結果を基に、文章データを擬似的に黙声認識結果とするようなプログラムを作成した。

3.3 形態素解析

形態素解析とは文章中の語句を品詞毎に分類し、それが意味を持つ最小単位の言葉まで分解することである。例えば、「私は学校へ行きます」という文章の場合、

私 (名詞) /は (助詞) /学校 (名詞) /へ (助詞) /行き (動詞) /ます (助動詞)

という形に区分化される。しかし、日本語として複雑な文章を完璧に形態素解析することは困難である。複雑な文章の一つとして頻繁に挙げられる例が、「すもももももものうち」という文章である。一般的な解釈としては

すもも (名詞) /も (助詞) /もも (名詞) /も (助詞)
/もも (名詞) /の (助詞) /うち (名詞)

となるが、日本語における機械学習を用いた単語境界の判別は非常に困難であり、機械的にこういった文章を正しく確実に形態素解析できるわけではない。本研究では、認識結果の特徴量に基づいた文章校正を目的としているため、形態素解析を行った際の誤りに関しては修正を行わないものとする。

3.4 文章破壊

校正前の文章データとして、黙声認識で認識された場合に出力されたような文章が必要になる。ここで本研究では、黙声認識における結果のシミュレーション用プログラムを準備した。このプログラムに文章を入力することで全文を平仮名化し、文字を一音ずつ確率で変化させることで黙声認識的出力を実現させた。また、Mecabによって形態素解析された文章は最小単位まで分割されているので、単語毎について境界の考慮は行わないものとした。

3.5 文章校正

変化させた出力文章は基本的に日本語として不完全であるため、正しい語句を選びだし校正しなければならない。語句のもっともらしさ、すなわち出力文字列との類似度を測る尺度として、ハミング距離を用いる。

ハミング距離は、ある文字列を別の文字列に変更する際の置換回数を表す。ハミング距離の大きさを基準にして、もっともらしい単語かどうかの判断を行う。

その際、問題となるのが単語境界の判別である。出力文字列は日本語として意味を成さず、どこまでが一つの単語なのかを判別することは困難であるため、ここでは校正前文章における文節が既知であるものとする。

文節とは日本語を自然な発音によって区切った最小単位であり、自立語、もしくは自立語と一つ以上の付属語からなる。文節は、黙声認識の実験を行う際に被験者が意識をすることで、容易に判別することができる。

しかし、形態素解析を行うだけでは文節毎に区切られている文章とは異なるため、システム上では一文字の形態素を直前の単語につなげることにより、文節に区切られた文章として表現するものとした。

第4章 校正実験

4.1 実験方法

我々は3章で述べたシステムを用いて校正実験を行った。実験の方法としては、同程度の文字数を持つ17個のデータセットの文章をシステムに流し込み、その中の一つをテストデータとし、その他の残りを辞書データとした。それらのデータセットで学習と校正実験を行った。その次に、先程の辞書データセットの一つをテストデータとし、先程のテストデータは学習データセットの一つとして扱うといったような交差検証法を用いて実験を行った。スタージェスの公式(4.1)に基づいて、一つのテストデータにつき5回の校正実験を行い、それを全てのデータセットについて行った。

$$k = 1 + \log_2 x \quad (4.1)$$

4.2 文章データ

実験をするにあたって、データセットとして利用する文章が多量に必要となる。黙声認識においては、いわゆる書き言葉よりも喋り言葉を認識する場面が多くなるため、実際に喋った言葉の文章をデータセットとして利用するのが望ましいと考えた。そのため本研究では、多様な資源データを持つWebサービスである、Wikisourceを利用した。Wikisourceはオープンコンテンツ形式のWebサービスであり、発話文章データとしてそこに掲載されている内閣総理大臣所信表明演説の原文を、計16人分利用した。

4.3 評価方法

文章の校正に関しては、テストデータにおける原文と破壊文との間で異なる単語数、また原文と校正文との間で異なる単語数の割合、つまり破壊された単

表 4.1: 発話文章データ毎における単語数

	単語の合計 (語)	単語の種類 (語)	未知語の合計 (語)	未知語の種類
文書 1	2276	1358	374	346
文書 2	2592	1460	344	328
文書 3	2955	1549	380	344
文書 4	1709	1045	233	215
文書 5	3036	1692	463	436
文書 6	2234	1306	372	349
文書 7	2805	1704	595	563
文書 8	2198	1323	329	313
文書 9	2038	1294	587	537
文書 10	4347	2321	1011	912
文書 11	3693	2080	848	753
文書 12	2067	1298	337	319
文書 13	3322	1905	715	535
文書 14	1708	1028	338	308
文書 15	1381	874	333	299
文書 16	1479	910	343	319

語がどれだけの割合で校正されたかという点で評価を行った。また今回の手法では、校正する際の単語を既知の文章データから精選している。そのため、その文書にしか現れないような未知語に関しては校正することは不可能である。よって、文書内の未知語とその数をシステム上で抽出し、実質上校正不可能な単語、つまり未知語を除いた単語がどれだけ校正できたかという点についても調査することとした。

第5章 実験結果

我々が作成したシステムにおいて校正実験を行った結果及び考察を以下に述べる。

5.1 文章破壊と文章校正

本実験において、文章を破壊した際文書中の単語のうち、どれだけの単語が違う意味の単語、もしくは意味を成さない文字列になったかを示す割合を「変化率」、また校正した際、文書中の単語のうち、どれだけの単語が元の単語に校正されたかを示す割合を「校正率」、そして変化した単語のうち、どれだけの単語が元の単語に校正されたかを示す割合を「補正率」と表す。

黙声認識に基づいた認識結果の特徴量を基に文章を破壊した結果、文書全体における平均変化率は約 13.6% という結果になった。次に、その文章を校正した結果、文書全体における平均校正率は約 61.3% であった。また、文書全体における平均補正率は約 55.2% という結果であった。中でも文書 2 における平均補正率が最も高く、60.8% であり、文書 9 における平均補正率 45.8% が最も低い結果となった。

表 5.1: 文書 1 における実験結果

	変化率 (%)	校正率 (%)	補正率 (%)
1 回目	14.5	64.4	58.3
2 回目	14.1	65.2	59.5
3 回目	13.8	65.3	59.7
4 回目	13.9	65.4	59.8
5 回目	14.4	64.5	58.5

表 5.2: 文書 2 における実験結果

	変化率 (%)	校正率 (%)	補正率 (%)
1 回目	14.2	66.0	58.3
2 回目	13.7	66.0	60.3
3 回目	12.9	66.9	62.0
4 回目	14.1	66.2	60.7
5 回目	13.5	65.5	60.2

表 5.3: 文書 3 における実験結果

	変化率 (%)	校正率 (%)	補正率 (%)
1 回目	12.9	63.9	58.3
2 回目	12.7	63.8	58.6
3 回目	13.0	64.9	59.6
4 回目	13.2	64.0	58.6
5 回目	13.0	64.0	58.6

表 5.5: 文書 5 における実験結果

	変化率 (%)	校正率 (%)	補正率 (%)
1 回目	12.7	63.8	58.5
2 回目	13.0	64.3	59.0
3 回目	13.7	65.0	59.4
4 回目	13.8	64.0	58.2
5 回目	12.9	64.7	64.7

表 5.4: 文書 4 における実験結果

	変化率 (%)	校正率 (%)	補正率 (%)
1 回目	12.3	67.2	62.6
2 回目	13.2	67.1	62.1
3 回目	12.6	65.6	60.6
4 回目	12.2	66.1	61.4
5 回目	12.1	65.6	60.9

表 5.6: 文書 6 における実験結果

	変化率 (%)	校正率 (%)	補正率 (%)
1 回目	13.3	62.1	71.7
2 回目	13.2	62.3	72.5
3 回目	13.7	62.2	72.0
4 回目	14.1	63.7	74.2
5 回目	13.4	62.1	71.7

表 5.7: 文書 7 における実験結果

	変化率 (%)	校正率 (%)	補正率 (%)
1 回目	11.7	59.8	56.3
2 回目	14.0	59.6	57.4
3 回目	13.9	58.3	56.2
4 回目	13.0	59.3	57.8
5 回目	13.5	59.1	56.3

表 5.8: 文書 8 における実験結果

	変化率 (%)	校正率 (%)	補正率 (%)
1 回目	15.0	65.9	54.4
2 回目	13.8	65.5	53.0
3 回目	15.4	65.7	51.5
4 回目	13.9	65.7	59.3
5 回目	15.0	65.5	52.7

表 5.9: 文書に 9 おける実験結果

	変化率 (%)	校正率 (%)	補正率 (%)
1 回目	15.8	55.4	47.0
2 回目	16.0	53.9	45.1
3 回目	15.2	54.2	46.0
4 回目	15.2	54.1	45.9
5 回目	16.0	53.8	45.0

表 5.10: 文書 10 における実験結果

	変化正解率 (%)	校正率 (%)	補正率 (%)
1 回目	13.8	58.1	51.4
2 回目	13.8	58.2	51.4
3 回目	14.0	57.5	50.6
4 回目	14.1	57.8	50.9
5 回目	14.0	58.0	51.1

表 5.11: 文書 11 における実験結果

	変化率 (%)	校正率 (%)	補正率 (%)
1 回目	13.3	57.1	50.5
2 回目	13.9	57.5	50.6
3 回目	13.8	56.6	49.6
4 回目	13.7	58.4	51.8
5 回目	13.3	57.5	51.0

表 5.12: 文書 12 における実験結果

	変化率 (%)	校正率 (%)	補正率 (%)
1 回目	14.3	63.4	57.3
2 回目	14.2	63.1	57.0
3 回目	14.4	62.8	56.6
4 回目	14.5	63.3	57.1
5 回目	14.4	64.0	58.0

表 5.13: 文書 13 における実験結果

	変化率 (%)	校正率 (%)	補正率 (%)
1 回目	13.3	58.5	52.1
2 回目	13.7	58.5	52.2
3 回目	13.5	59.0	52.6
4 回目	13.2	58.4	52.1
5 回目	14.4	59.9	53.1

表 5.14: 文書 14 における実験結果

	変化率 (%)	校正率 (%)	補正率 (%)
1 回目	12.3	60.4	54.9
2 回目	12.6	60.9	55.3
3 回目	12.3	61.0	55.5
4 回目	12.2	61.1	55.7
5 回目	12.6	61.9	56.4

表 5.15: 文書 15 における実験結果

	変化率 (%)	校正率 (%)	補正率 (%)
1 回目	12.0	55.5	49.4
2 回目	12.2	56.7	50.7
3 回目	12.4	55.5	49.3
4 回目	12.2	56.6	50.5
5 回目	12.7	57.1	50.8

表 5.16: 文書 16 における実験結果

	変化率 (%)	校正率 (%)	補正率 (%)
1 回目	14.0	59.1	52.4
2 回目	15.1	59.1	51.8
3 回目	14.5	58.6	51.5
4 回目	13.8	57.9	51.2
5 回目	14.8	59.1	59.1

表 5.17: 文書毎における実験結果の平均

	平均変化率 (%)	平均校正率 (%)	平均補正率 (%)
文書 1	14.1	64.9	58.2
文書 2	13.7	66.1	60.8
文書 3	12.9	64.1	58.8
文書 4	12.5	66.3	61.5
文書 5	13.2	64.3	58.9
文書 6	13.5	62.6	56.8
文書 7	13.2	59.2	53.0
文書 8	14.6	65.7	59.8
文書 9	15.6	54.3	45.8
文書 10	14.0	57.9	51.1
文書 11	13.6	57.4	50.7
文書 12	14.4	63.3	57.2
文書 13	13.6	58.9	52.4
文書 14	12.4	61.1	55.6
文書 15	12.3	56.3	50.1
文書 16	14.4	58.8	51.8

文章を校正する際、辞書内の単語に含まれていない単語、つまり未知語が多ければ多いほど校正率は低くなる傾向にある。文書 8 における、文書内の単語に対する未知語の割合は 15.0% であるのに対し、文書 15 における未知語の割合は 24.1% であった。

5.2 未知語

5.2.1 校正の精度に対する未知語の影響

当然のことながら、文書ごとに未知語の数は異なっている。今回の実験において文書内における未知語の割合は、少ないものだと文書 3 の 12.9% であったが、最も多い文書では文書 9 の 28.8% であった。

本研究において、未知語への対策は行っておらず、文章破壊の際に未知語が破壊された場合、元の単語に戻ることはない。よって、校正できた単語数の割合に関して、未知語の数を考慮した結果についても算出した。文書 k 内の単語

の数を w_k (語), 未知語の数を u_k (語), 校正できた単語の数を t_k (語) としたとき, 未知語を除く文書内の単語のうち, 校正できた単語の割合 x_k は

$$x_k = \frac{t_k}{w_k - u_k} \quad (5.1)$$

となる. このように未知語を考慮した場合の補正率 x の平均は 74.1% であった.

5.2.2 未知語と補正率

未知語が多ければ多いほど補正率は低くなる傾向にある. ここで, 未知語と補正率の関係について調べてみたところ, 二つの数値の散布図 5.1 からある程度の線形性が見られた.

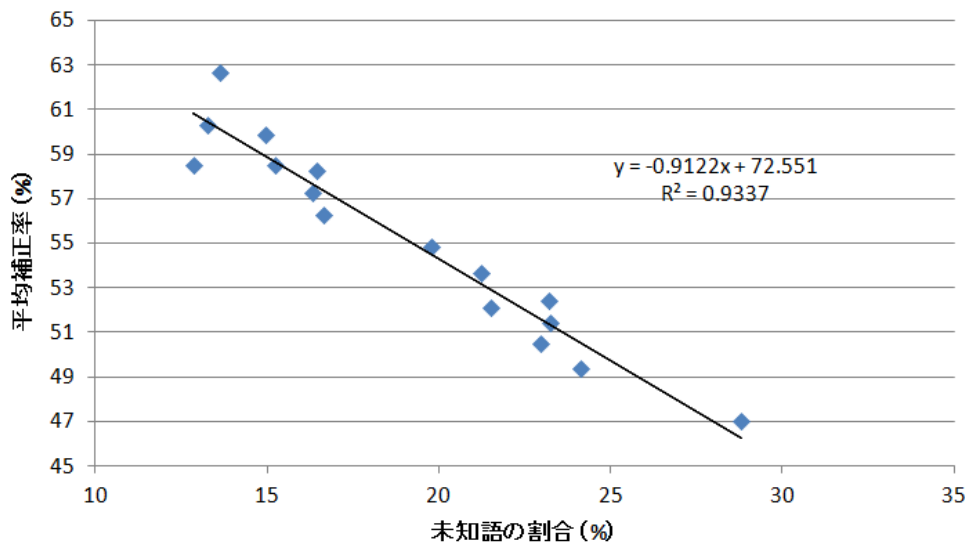


図 5.1: 未知語の割合と平均補正率の相関図

最小二乗法 5.2, 5.3 によって直線の式を確認したところ, $a \approx -0.87$, $b \approx 0.88$ となり, 直線の式は $y \approx -0.87a + 0.88$ となった.

$$a = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} \quad (5.2)$$

$$b = \frac{n \sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i - \sum_{i=1}^n x_i y_i \sum_{i=1}^n x_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} \quad (5.3)$$

また、この直線における決定係数 R^2 は $R^2 \approx 0.88$ であった。

5.2.3 未知語の種類

本実験において発生した未知語は、固有名詞の他に数値や年月、または国の連名における形態素解析の誤りによるものが見られた（例えば日朝関係がニチ/アサ/カンケイと解析されていた）。これらに関してはこちらで辞書を準備するか、もしくは形態素解析ツールに用いられる辞書に書き加えることで半分以上の未知語対策が可能だと考えられる。この問題が解決できれば、さらなる校正精度の上昇を図ることができる。

第6章 結論

6.1 実験のまとめ

本研究では黙声認識技術の普及を目的として、発話文章特徴量に基づいた文章内文字列の変化と校正の実験を行った。特徴量による仮想的な黙声認識結果において、文書内の単語数のうち正しく出力された単語の平均的な割合は13.6%だったのに対し、それらを校正した結果、61.3%まで上昇した。また、誤って出力された単語のうち、校正できた単語の平均的な割合は71.0%であり、全ての結果において6割以上の誤った単語を校正することができた。さらに、未知語を除いた単語のみを考慮した場合、校正できた単語は全ての結果において8割以上であった。

6.2 今後の課題

本研究では筋電位の黙声認識において、発話特徴量を用いた認識結果の校正を行った。結果として、少なくとも未知語を含んだ6割以上の単語を校正することができた。また、少なくとも未知語を含まない単語の8割以上の校正に成功した。

現在一般的に利用されている音声認識システムでは、クラウド上においてコーパスを追加、更新しているものが多々みられる。同様の手法を用いて、黙声認識システムもコーパスを大きくすることで、未知語への対策を行えば、さらなる校正精度の向上が期待できる。しかし、今回の研究では同一ジャンルの発話文章を基にした実験を行ったため、コーパスの追加における過学習の弊害、つまり単語候補の肥大化による選択誤りの増加の考慮はこれからの課題である。

しかし、コーパスのタグにジャンル情報を機械的に付与する研究が行われており、大量のコーパスにそういった情報を付与することができれば過学習の弊害効果を薄めることができるものと考えられる [5]。また、形態素解析の誤りに

対しては、隠れマルコフモデルを利用した解析手法による精度の向上の研究も行われており、加えて、確率モデルによりた品詞分布の特徴量から、文章のジャンル推定を行う研究もされている [6][4]。このような手法を組み合わせることにより、コーパスの増大から得られる恩恵を大きくすることが可能となる。

当面の課題としては、本システムのような校正機能を搭載した黙声認識システムを構築し実装することである。その後、上述したように精度向上を図り、システムにおける適応化や人物ごとの特徴量の関連性なども調べていく必要がある。

第7章 謝辞

本研究を進めるにあたり、丁寧かつ熱心にご指導して頂きました三浦准教授に深謝致します。また輪講や中間発表におきましてご指導して頂きました情報セクションの先生方に深謝致します。加えまして、研究の方針を思索するにあたってご意見を頂きました、三浦研究室と情報セクションの学生方に深謝致します。最後に、学生生活を送るにあたって精神的、また経済的な面においてご支援をして頂きました家族に深謝致します。

関連図書

- [1] Tanja Schultz Lena Maier-Hein, Florian Metze and Alex Waibel. SESSION INDEPENDENT NON-AUDIBLE SPEECH RECOGNITION USING SURFACE ELECTROMYOGRAPHY. In *Automatic Speech Recognition and Understanding, 2005 IEEE Workshop on*, pp. 331–336, November 2007.
- [2] 大内慶久. Hybrid HMM/SVM を利用した筋電位に基づく日本語黙声認識.
- [3] 佐藤大介, 志築文太郎, 三浦元喜, 田中二郎. Popie : フローメニューに基づく日本語入力手法. *情報処理学会論文誌*, pp. 2305–2316, 2006.
- [4] 白井匡人, 三浦孝夫. 確率モデルによる品詞分布の特徴推定. In *Forum on Data Engineering and Information Management*, 2012.
- [5] Robert E. Mercer Hospice Hougbo. An automated method to build a corpus of rhetorically-classified sentences in biomedical texts. In *Proceedings of the First Workshop on Argumentation Mining*, pp. 19–23, 2014.
- [6] 松本 裕治竹内 孔一. HMM による日本語形態素解析システムのパラメータ学習. *情報処理学会研究報告. 自然言語処理研究会報*, pp. 13–19, 1995.