

# 単語視認時の脳波信号を用いた未知語知覚検出

## Detection of Visually-presented Unknown Words based on Electroencephalography

笹倉 隆史<sup>1\*</sup> サクリアニ サクティ<sup>1</sup> グラム ニュービッグ<sup>1</sup> 戸田 智基<sup>1</sup> 中村 哲<sup>1</sup>  
Takafumi Sasakura,<sup>1</sup> Sakriani Sakti,<sup>1</sup> Graham Neubig,<sup>1</sup> Tomoki Toda,<sup>1</sup> Satoshi Nakamura<sup>1</sup>

<sup>1</sup> 奈良先端科学技術大学院大学

<sup>1</sup> Nara Institute of Science and Technology

**Abstract:** The appearance of unknown words often disturbs communication. Most spoken dialog systems deal with unknown words that are uttered by the user, but which are not covered by the system's vocabulary. In this paper, we focus on detecting unknown words from the user side in which the system utterance is unknown to the user. In particular, we develop a classifier based on Electroencephalography (EEG) signal from user's brain waves, including the use of absolute power and Event-Related Desynchronization (ERD) features. The results show that we could detect the characteristics of brain waves at the time of unknown word perception significantly better than the chance rate.

### 1 はじめに

対話において、相手に伝えたい情報は共通であってもそのための発話には様々な候補が存在する。人間同士のコミュニケーションでは、単語単位での同義語の存在や、言い回し、省略や参照などを用いて状況に合わせた発話を生成しており、これは円滑な情報伝達には必要不可欠である。特に、話題に対する相手の知識に応じた語を用いて発話を構成することは重要であると考えられる。例えば、医師であれば医療関係者との会話には専門用語を用い、患者との会話にはより一般的な語を用いる。従って、システムと人間との対話でも同様にユーザの語彙力に応じた対話を行うことが望ましい。本研究ではユーザに適した応答文生成を行う対話システムの構築のために、情報伝達の破綻を検出することを目指す。

対話破綻が生じる主な原因の1つとして、未知語の出現が挙げられる。対話システムではシステムに対して未知な語がユーザ発話に含まれる場合の扱いが重要であり、この問題に対処する研究が行われている [10, 2]。一方で、ユーザに対して未知な語がシステム発話に含まれる場合も考慮されるべきである。本稿では後者に着目し、システムの出力した語がユーザに対して未知である状態を検出する。システム発話に含まれる未知語をユーザの振舞から検出することができれば、言い

換え等の対処が可能になる。更に、検出した未知語から現在の話題に対するユーザの知識レベルを推定し、対話制御や応答文生成に利用することが期待される。

ユーザの振舞からシステム発話に含まれる未知語知覚を検出することは困難である。これを実現する方法としては、ユーザの視線や表情といった非言語情報を取得する、あるいはより直接的にユーザの明示的な発話を得るというものがまず考えられるであろう。しかしながら、未知語を知覚したときに必ずこれらの言語・非言語情報をユーザが発信するとは限らない。また、未知語知覚を明示的に表すことは円滑な対話を少なからず阻害し、負担をユーザに課すこととなる。人間同士のコミュニケーションではこれらを軽減するために、未知語を知覚したとしても前後の文脈などからその語の意味を推定したり、その語が重要でないと判断した場合は未知のまま対話を続けるなどの対処が行われている。ここで、語の重要度に対して発話側と聞き手側とでずれがある場合に問題が生じうる。医師の発話に含まれる重要な語が患者にとって未知語であるにも関わらず、重要でないと患者が判断してしまった場合、致命的なミスに繋がる可能性がある。このような情報伝達の過失を回避するためには、聞き手側にとって未知の語の中から発話側にとって重要な語を発話側が選択し、補足することが望ましい。

上記の問題に対処するために、本研究では生体情報の1つである脳波情報に着目する。生体情報は自律神経が支配する指標であり、人間が無意識に発している信号である。そのため、自身での制御が困難であり、

\*連絡先：奈良先端科学技術大学院大学  
〒630-0192 奈良県生駒市高山町 8916-5  
E-mail: sasakura.takafumi.sl4@is.naist.jp

ユーザの意識的な判断に左右されない情報が取得できると考えられる。コミュニケーションの際、多くの生体情報は提供されないが、専用の計測器を用いることで取得することができる。本研究では、未知語知覚時のユーザの認知的な状態を直接的に推定するため、脳波計 (Electroencephalograph: EEG) を用いて脳波情報を取得し、知覚した語が未知語か否かの識別を試みる。

## 2 脳波とその利用方法

脳波検査 (Electroencephalography: EEG) は脳活動から生じる頭皮上の電位変化を電気生理学的に計測することである [6]。EEG は頭皮上に複数の電極を配置し、2箇所以上の電極から電位差を計測することによって脳から生じる電気信号を記録する。対話に関連した研究として、Sridharan ら [9] は、NeuroDialog と呼ばれる対話システムを構築した。NeuroDialog は対話中のユーザの EEG に基づいた予測モデルを利用しており、この研究によって脳波情報からシステムの音声認識誤りを推定することが可能であることが示された。これに対し、本研究では未知語知覚という形で生じる情報伝達の失敗を検出することに着目する。

知覚した語が未知語か否かを判断する際、その語を自身の記憶と照合するという処理を行っていると考えられる。Sederberg ら [8] は、単語の記憶タスク中の脳波において、後に思い出せる名詞を記憶しているときの  $\theta$  及び  $\gamma$  周波数帯のパワーが、思い出せない名詞のときと比較して有意に高かったと報告している。また、Sauseng ら [7] は、矢印を用いた記憶照合タスクの結果が、入力された視覚情報と事前に格納された情報との間でのマッチングの成否によるものであると述べている。更に、Klimesch [3] の研究では、 $\theta$  及び  $\alpha$  周波数帯での脳波の振動が認識及び記憶のパフォーマンスを特に反映していることを示している。本研究では、未知語を知覚した時点の各周波数帯のパワーの定常的な状態を用いた識別と、未知語知覚に伴って生じるパワーの変化を用いた識別を行う。後者は事象関連脱同期 (Event-Related Desynchronization: ERD [5]) と呼ばれ、参照区間と比較した標的区間の各周波数帯のパワーの減少率を表す。参照区間 ( $P_r$ ) は標的区間 ( $P_t$ ) 以前の時間から実験者によって選択され、ERD は以下の式で計算される。

$$ERD = \frac{P_r - P_t}{P_r}. \quad (1)$$

計算された特徴量に対して、識別性能を向上させるために特徴量選択を行う。特徴量選択には、実験者が選択した特定の部位や周波数帯のみを用いる手法と全特徴量から統計的手法によって選択するものがある。こ

こで、一般的な脳波データの性質として、個人差が大きいということが知られており、立岡ら [11] の研究では、同一のタスクであっても実験参加者によって異なった特徴が統計的手法によって選択されている。本研究では、タスクに対する脳活動の個人性に対応するために統計的手法を用いて特徴量選択を行う。

実際には、計測した EEG データからパワーや ERD といった特徴量を算出した後、各特徴量がそれぞれ平均 0、標準偏差 1 となるように正規化し、特徴量選択を行った。使用する特徴量は以下の前向き部分変数集合選択法 [1] を用いて選択した。

1. 変数サイズ 1 の部分集合を全特徴量に対して生成する (初期化)。
2. 各部分集合に対し、以下の式で与えられる  $J$  の値を計算する。

$$J = \frac{S_B}{S_W} \quad (2)$$

このとき、 $S_B$  と  $S_W$  は以下の式で計算される。

$$S_B = \sum_j^L N_j (m_j - M)^t (m_j - M) \quad (3)$$

$$S_W = \sum_j^L \sum_i^{N_j} (x_i - m_j)^t (x_i - m_j) \quad (4)$$

ここで、 $L$  はクラス数を表す。 $N_j$  はクラス  $j$  に含まれる部分集合の数を表し、 $M$  は各クラスの部分集合の平均ベクトルの加重平均、 $m_j$  はクラス  $j$  の部分集合の平均ベクトル、 $x_i$  は  $i$  番目の部分集合ベクトルを表す。

3. ステップ 2 で計算した  $J$  のスコアが最大となる部分集合を選択する。
4. ステップ 3 で選択した部分集合に、それに含まれない特徴量 1 つを追加し新たな部分集合とする。
5. スコア  $J$  の最大値が下がるまでステップ 2 から 4 までを繰り返す。

## 3 実験設定

本研究では、未知語知覚時のユーザの状態を直接的に推定するため、単語知覚時の脳活動から、その語が未知語か否かの識別を試みた。本稿では刺激を視覚的に行うことで実験参加者が単語を知覚する時点を明確にした。本実験の目的は、単語視認時に計測した EEG データからその単語が未知語か既知語かを識別できるかを検証することである。目的を達成するために、それぞれのクラスのデータ量が十分に取得でき、尚且つ未知か既知か以外の要因が実験結果に影響を及ぼさないように配慮し、実験を設定した。

表 1: 提示単語例

単語親密度	
4 (高)	1 (低)
ヨソモノ	ワラバイ
ドロボウ	ラシャメン
ホケンジョ	ザイカタ
ソトガワ	ドカヒン
オヤモト	オシタジ

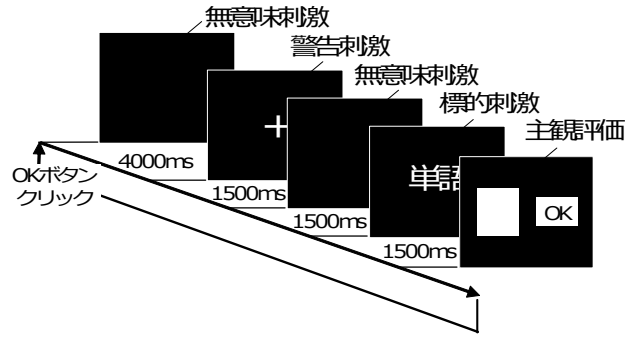


図 1: 実験の手続き

### 3.1 実験参加者

右利きの男子学生 (23 - 24 歳) 6 名が実験に参加した。全参加者が日本語を母国語とし、正常、あるいは正常に矯正された視力を有していた。実験参加者は精神医学的、神経学的な病気やアルコール依存によって視覚に異常が生じた経歴を持たない。記録した EEG データに含まれるアーチファクトを除去した結果から、6 名中 5 名分の EEG データを解析対象として用いた。

### 3.2 刺激提示

視覚刺激用の単語は、未知か既知か以外の要因が結果に与える影響を極力減らすように選出した。そのため実験には 4 モーラからなる日本語名詞 300 語を視覚刺激として提示した。提示する単語は親密度別単語了解度試験用音声データセット 2007 (FW07) [4] から選出した。これらの語は 4 段階の親密度に基づいてランク付けされており、実験には親密度最小と最大の語のみを用い、親密度 1 と 4 の単語をそれぞれ 150 語ずつ使用した。親密度 2 や 3 の単語を除いたのは、実験参加者にとって既知か未知かの判断が難しいような状況を回避し、片方のクラスにデータ数が偏ることを防ぐためである。刺激提示に使用した親密度 1 と 4 の単語例を表 1 に示す。

実験以外の刺激を遮断するために実験は薄暗い防音室で行った。実験参加者は 27 インチのディスプレイの前に 120 cm 程度離れて配置された椅子に楽な姿勢で座り、画面を見ながらマウス操作を行った。画面の背景は黒で視覚刺激として提示する単語は画面中央に白い文字で表示した。

本実験の参加者は正面の画面に表示された語を視認し、その単語についての質問にマウス操作で回答した。実験の具体的な手続きを図 1 に示す。まず最初に 4000 ms の無意味刺激 (暗転) を提示し、次に警告刺激として十字のマーク (+) を画面中央に 1500 ms 表示した。これは画面中央に視線を固定し、眼球運動を減らすた

めでである。次に再び無意味刺激を 1500 ms 提示し、続いて標的刺激として、300 語から無作為に抽出した語 1 つを片仮名で 1500 ms 提示し、最後に主観評価画面を表示した。実験参加者は提示された語に対し、「既知語」「未知語」「刺激提示を見逃した」の 3 項目のうち 1 つを選択し、OK ボタンをクリックして次の試行に移った。以上の試行を全単語が表示されるまでの合計 300 回に渡って繰り返した。

参加者には OK ボタンをクリックしてから主観評価画面が表示されるまでの間の脳活動を計測している旨を伝え、主観評価画面で随時休憩をとりながら実験を進めるように促した。また、実験の前に本実験で使用する 300 語以外の単語で数回の試行を練習した。更に、計測中は体動や瞬きなどを極力抑えるように促すなど、実験参加者への負担軽減と EEG データへのアーチファクトの混入を防ぐための事前教示を行った。

### 3.3 脳波計測

脳波計測には Brain Products 社製の BrainAmp を使用した。脳波計の電極は銀・塩化銀電極を用い、EEG キャップを用いて頭皮上に固定した。信号は頭皮上の 29 チャンネルの電極から取得し、接地電極は両耳朶、参照電極は鼻尖に 2 つ配置した。頭皮上の電極の配置は図 2 に示す。EEG データは、信号の SN 比を向上させるため、各電極のインピーダンスは専用のペーストを用いて 5 kΩ 未満に下げたのち、1000 Hz のサンプリング周波数で取得した。

収録した EEG データは、筋電などのアーチファクトを取り除くため、40 Hz 以下のローパスフィルタを用いて高周波成分をカットした。更に、振幅絶対値が 80 μV 以上の成分を含む試行を瞬きや体動などがあつたとみなし解析対象から除外した。また、主観評価において刺激提示を見逃したと回答している試行に関しても脳波解析の対象から除外した。

視覚刺激として単語が提示された時点から幅 1024 ms (1024 pt) を抽出したデータ系列に対し、幅 256 ms を

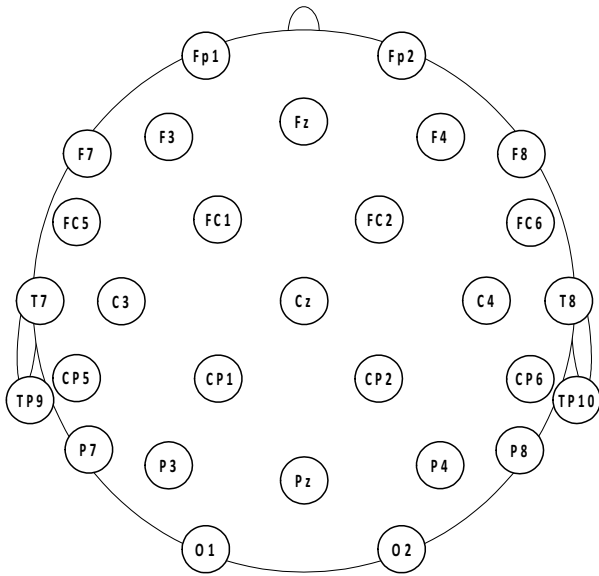


図 2: 頭皮上の電極の分布

50 % のシフト幅で切り出し、解析対象とした。切り出したデータ系列に対し、ハミング窓を用いて 1024 pt の高速フーリエ変換 (FFT) を用いてパワースペクトルを算出した。また、パワーの変動 (ERD) を計算するために、警告刺激提示時点からの 1024 ms に対しても同様の処理を行い、基準とした。以上の処理によって算出された値を用いて、各周波数帯のパワー及び ERD を計算した。

### 3.4 識別器

算出されたパワーあるいは ERD の値を素性として EEG データの識別器を生成した。識別は主観評価結果に基づいて決定した「未知」「既知」の 2 クラスに対して分類を行った。各クラスのサイズを等しくするためにデータ数の少ない方のクラスに合わせてデータ数を調整した。また、脳活動の特徴は個人によって大きくことなるため、識別器は各実験参加者ごとにそれぞれ構築した。各実験参加者のデータに対して特徴量選択を行い、選択された特徴を識別器の素性として 2 クラス分類に利用した。識別器にはサポートベクターマシン (SVM) を用い、カーネル関数として動径基底関数 (RBF) を使用した。

## 4 結果

### 4.1 主観評価結果

主観評価結果を表 2 に示す。既知語に振り分けられた総数は 776 語、未知語に振り分けられた語は 687 語

表 2: 親密度に対する主観評価結果

単語数 [語]	単語親密度		合計	
	4 (高)	1 (低)		
主観 評価	既知	720	56	776
	未知	11	676	687
	見逃し	19	18	37

となった。結果は 5 名の参加者に対するものであり、提示刺激は実験参加者 1 名あたり単語親密度 1 と 4 の単語を 150 単語ずつ提示した。よって提示した総単語数は 1500 語となった。主観評価の結果、高親密度の語が既知語となった割合は 96.0%、低親密度の語が未知語となった割合は 90.1% となった。

### 4.2 識別結果

5 名の実験参加者のそれぞれ EEG データに対して 10 交差検定を行った平均識別率、及び二項検定による P 値を表 3 に示す。各クラスで等しくなるようにデータ数を調整したため、チャンスレートとなるランダムでの識別率は 0.500 となる。まず、各周波数帯のパワーを特徴量として識別器を構築した場合、チャンスレートに対して二項検定を行った結果、B, C, E の 3 名で有意水準 5% で有意差 (\*), A と D の 2 名で有意傾向 (+) が認められた。次に、パワーの変化 (ERD) を特徴量として構築した識別器では、チャンスレートに対して二項検定を行った結果、全実験参加者で有意水準 5% で有意差 (\*) がみられた。特に B, C, D, E に対しては有意水準 1% で有意差 (\*\*) が認められた。更に、パワーと ERD によるそれぞれ 2 つの識別器を比較すると、全実験参加者に対して ERD での識別率がパワーでの識別率を上回り、二項検定では A 以外の実験参加者に有意傾向 (+) が確認された。

### 4.3 特徴量選択結果

各実験参加者の EEG データから算出した各周波数帯のパワー及び ERD に対してそれぞれ前向き部分変数集合選択法を用いて特徴量選択を行った。5 名の実験参加者のうち、B と E の 2 名の EEG データに対して選択された主な特徴量を図 3 に示す。図の左が B、右が E の特徴量を表す。緑色は  $\theta$  波 (4~7 Hz) に属する周波数帯、赤色は  $\alpha$  波 (8~13 Hz) の周波数帯、青色は  $\beta$  波 (14~25 Hz) の周波数帯に含まれる特徴がその部位から選択されたことを示している。また、その部位が二重円かつ部位名に下線が記されたものは ERD

表 3: 識別結果

実験参加者	識別率		P 値 (二項検定)			データ数
	パワー	ERD	ランダム vs パワー	ランダム vs ERD	パワー vs ERD	
A	0.520	0.550	0.074 <sup>+</sup>	0.048 <sup>*</sup>	0.101	100
B	0.550	0.573	0.018 <sup>*</sup>	0.005 <sup>**</sup>	0.071 <sup>+</sup>	220
C	0.508	0.558	0.048 <sup>*</sup>	0.009 <sup>**</sup>	0.066 <sup>+</sup>	260
D	0.505	0.570	0.056 <sup>+</sup>	0.008 <sup>**</sup>	0.079 <sup>+</sup>	200
E	0.527	0.588	0.034 <sup>*</sup>	0.001 <sup>**</sup>	0.066 <sup>+</sup>	260

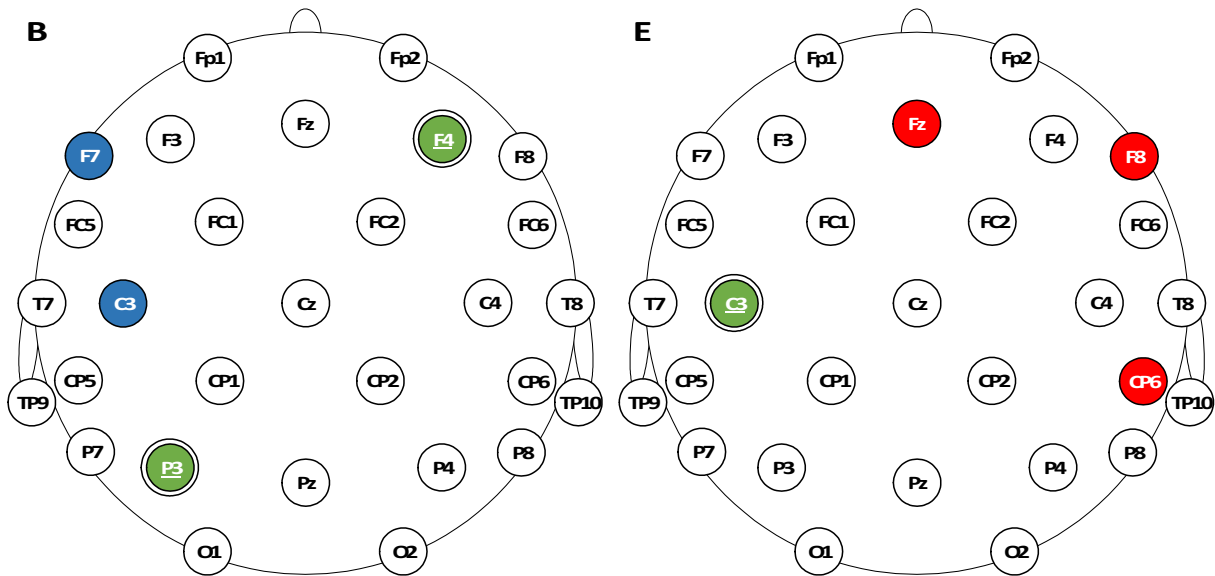


図 3: 選択特徴量例

を特徴量としたときに選択されたものを示し、それ以外はパワーの値を用いたときに特徴量として選択されたものを示す。図より、パワーを特徴量とした場合、Bは左側頭前部と左中心部において $\beta$ 周波数帯が選択され、Eは右側頭部、右側側頭後部、正中前頭部の $\alpha$ 周波数帯が選択された。一方、ERDの場合、Bは右前頭部と左頭頂部、Eは左中心部の $\theta$ 周波数帯が識別に用いられた。また、他の3名でも同様にそれぞれ異なる傾向の特徴量が選択された。

## 5 考察

本実験では、単語親密度に基づいて刺激提示用の単語を選出した。これは、主観評価によって振り分けられたEEGデータが一方のクラスに偏らず、両クラスともに十分な学習データを確保することを目的としていた。単語親密度の高い語は既知語に、低い語は未知語

に振り分けられることを期待して実験設定を行ったが、主観評価結果から全体の93.1%がそのように振り分けられた。従って、各クラスの単語数確保の観点からは実験設定は妥当であったといえる。

識別結果より、各周波数帯のパワー及びその変化を特徴量とした識別器がチャンスレートと比較して有意な性能を示したため、背景脳波の差異を捉えることは重要であるといえる。各周波数帯のパワーとERDのそれぞれを用いた識別器の識別率を比較した結果は、全実験参加者に対してERDでの識別率がパワーでの識別率を上回り、ほとんどの実験参加者に対して有意傾向が確認された。このことから、背景脳波の定常的な特徴のみでは未知語知覚の検出には十分であるとはいえず、ERDのような一時的な変化を捉えた特徴量を利用することが未知語知覚検出には適していると考えられる。

本稿の実験設定では、全実験参加者を通して特徴量

選択結果に一貫した傾向を確認することはできなかった。これは、単語視認時の脳活動が個人によって異なる傾向があるからであると考えられる。このことから、特定の特徴のみを抽出した一般的な識別器を構築することは困難であるといえる。本研究では、個人ごとに識別器を構築することで脳波の個人性に対処した。

## 6 まとめ

本研究では、単語視認時の脳波データからその語が既知か未知かの識別を行った。実験の結果、各周波数帯のパワーを特徴量として用いた場合とパワーの変化(ERD)を用いた場合の両識別器でチャンスレートよりも高い識別性能が確認された。特にERDを用いた識別器において最も高い識別率を示した。

今後の課題としては、識別性能の向上、対話中などの実環境に近い条件での実験、マルチモーダル対話システムへの導入などが挙げられる。

**謝辞** 本研究の一部は、(独)情報通信研究機構の委託研究「知識・言語グリッドに基づくアジア医療交流支援システムの研究開発」、Microsoft CORE 10 プロジェクト、およびJSPS科研費26870371の助成を受け実施した。

## 参考文献

- [1] N. Hiruma, K. Sagara, Y. Tanaka, H. Takeichi, O. Yamashita, R. Hasegawa, T. Okabe, and T. Maeda. Brain communication : Theory and application. *IEICE Society Conference*, Vol. 94, No. 10, p. 926, 2011.
- [2] A. Kai, Y. Hirose, and S. Nakagawa. Dealing with out-of-vocabulary words and speech disfluencies in an n-gram based speech understanding system. In *ICSLP*, Vol. 2, pp. II-21, 1998.
- [3] W. Klimesch. Eeg alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain research reviews*, Vol. 29, No. 2, pp. 169-195, 1999.
- [4] T. Kondo, S. Amano, S. Sakamoto, and Y. Suzuki. Development of familiarity-controlled word-lists (fw07). *IEICE Society Conference research report*, Vol. 107, No. 432, pp. 43-48, 2008.
- [5] G. Pfurtscheller and A. Aranibar. Event-related cortical desynchronization detected by power measurements of scalp {EEG}. *Electroencephalography and Clinical Neurophysiology*, Vol. 42, No. 6, pp. 817 - 826, 1977.
- [6] S. Regel. *The comprehension of figurative language: electrophysiological evidence on the processing of irony*. PhD thesis, Universitätsbibliothek, 2009.
- [7] P. Sauseng, W. Klimesch, W. R. Gruber, and N. Birbaumer. Cross-frequency phase synchronization: a brain mechanism of memory matching and attention. *Neuroimage*, Vol. 40, No. 1, pp. 308-317, 2008.
- [8] P. B. Sederberg, M. J. Kahana, M. W. Howard, E. J. Donner, and J. R. Madsen. Theta and gamma oscillations during encoding predict subsequent recall. *The Journal of Neuroscience*, Vol. 23, No. 34, pp. 10809-10814, 2003.
- [9] S. Sridharan, Y.-N. Chen, K.-M. Chang, and A. I. Rudnicki. Neurodialog: an eeg-enabled spoken dialog interface. In *ICMI*, pp. 65-66. ACM, 2012.
- [10] S. R. Young. Detecting misrecognitions and out-of-vocabulary words. In *ICASSP*, Vol. 2, pp. II-21. IEEE, 1994.
- [11] 立岡祐也, 吉川大弘, 古橋武, 田中一男.  $\alpha \cdot \beta$  波を用いた思考判別に関する基礎的考察. 日本知能情報ファジィ学会 ファジィ システム シンポジウム講演論文集, Vol. 22, No. 0, pp. 127-127, 2006.