Vol.2019-SLP-128 No.7
2019/7/19

振幅スペクトルと相対位相を用いた想起母音の認識

中川聖一[†1]　堀田順平[†1]　山本一公[†1]

概要：ブレインインタフェースとして，脳波による意思伝達が望まれている．重度の構音障害者など言葉の発声ができない人のために，発声したい言葉の想起により生じる脳波による想起音声の認識が望まれている．我々は複数チャンネルの脳波に対して独立成分分析で雑音成分を除去後，振幅スペクトルと相対位相を特徴パラメータとして、GMMを識別器として用いて，５母音の想起音声に対して約 60%の認識率を得た．また、二母音間の識別に対しては，約72%の識別率を得た．特に、母音/a/と/i/の識別率は 90%であった。脳波による音声認識において独立成分分析と相対位相情報の有効性を示す．

キーワード：想起音声，脳波，母音認識，振幅スペクトル，相対位相，ブレインコンピュータインタフェース

1. はじめに

ブレインインタフェースとして想起音声の脳波による音声認識技術が望まれている．重度の構音障害者など言葉の発声ができない人のためには，発声したい言葉の想起により生じる脳波による想起音声の認識が望まれている．しかし、脳波は雑音の重畳が大きく，音声認識は難しい．SN 比が 0 デシベル程度の通常音声でもヒトでは比較的正しく聴き取れるが，コンピュータでは最新のディープニューラルネットワークを用いた音声認識法でも難しい．これに比べて，SN 比が極めて悪い，雑音に埋もれた想起音声の脳波によって想起された言葉を認識することは困難であることは容易に想像がつく．ヒトが 5 母音アイウエオの想起音声の脳波を比べても，区別は不可能である．

音声想起に基づくインタフェースは，これまでにも多数研究されてきた．主に，fMRI, MEG（脳磁図），EEG などの生体信号を用いている[1][2] ．本研究で使用する EEG は，非侵襲でリアルタイム計測が可能であり，実現できれば社会的需要は計り知れない．この EEG に対しては，母音などの定常な音声の想起や音節などの短時間長の音声の想起を対象とした研究が多い．このような対象では，EEG の時間的変化を捉える必要がなく，複数チャンネルの EEG の想起した時間長にわたる平均振幅エネルギーや事象電位を特徴量として用いて、静的パターン認識問題として扱い，SVM やランダムフォレスト，ニューラルネットワーク法で認識するものが多い．これらの研究で，以下のことが明らかになっている．①想起によって生じる脳波と比べて，人工的および脳に常時発声している雑音の方が大きく，観測されるのは雑音の重畳した音声想起の脳波である．②通常の発声は，音声調音器官を動かして発声するが，その際に生じる脳波は，口唇・舌・咽頭・喉頭などの筋肉の動きによって生じる脳波を捉えている可能性が高く，想起によって生じる脳波とは言い難い．①については，Porbadingk らの研究[3]で，想起音声のデータ収集方法によって認識結果が大きく異なることが明らかになっている．つまり，認識対象の一つであるある単語を 20 回連続して想起し，次に、別の単語を 20 回想起して行くという繰り返しでデータを収集する場合と，認識対象単語を 1 回ずつ順に想起し，これを 20 回繰り返してデータ収集した場合を比較すると，前者はチャンスレベルよりも有意に高い認識精度を示し，後者ではチャンスレベルの認識精度に近い．このことから，前者は，想起音声ではなく，脳波の重畳している雑音の違いを認識している可能性が高く，この種の研究例が多い．②については，神崎卓丸らの研究[4]で，音声調音器官を通して通常に発声した際に生じる 60 チャンネルの脳波を使用して，同期加算平均，ノイズ除去後に RMS の時系列×60 チャンネルの信号に対し KL 展開と複合類似度法という統計的パターン認識法を駆使して被験者 2 名の 18 短音節の認識率が約 53%（5 母音の認識率は約 70%）という高い認識率が得られたと報告している．しかし，これは口唇・舌・咽頭、喉頭等の調音器官の動きによって生じる脳波を捉えていて、想起音声の脳波とは言い難い．一方，調音器官を動かさない場合の想起音声に対しては，18 音節で約 32%（5 母音に対しては約 33%）の認識率に低下すると報告している．

想起音声の認識研究では，2 母音の識別や 3 音韻クラス分類の研究が多い. Iqbal らは、Daslla と同じデータベース[5]を用いて、30 個の想起母音の EEG の単純な平均と分散の統計量を学習し、2 母音/a/と/u/を約 95%で識別している[6]。Riaz らは声を出さずに発声器官を動かして発声した時(silent speech)の脳波と想起音声時の脳波による母音認識の比較を行っている[7] ．音声認識で用いられる MFCC を特徴量として，SVM, HMM, kNN 法の識別器を比較し，HMM と kNN 法で, 2 母音間の識別で両脳波に対して約 75%の識別率を得ている．Sarmiento らは，口形を想起した想起音声でなく(without the imagination of the movement of the oral

†1 中部大学
Chubu University

cavity)，心的に想起した（mental state）想起音声で，Nicolet EEG 測定器を用い，左側頭部の 21 チャンネルの 2～16Hz のスペクトラムを用いて広舌母音/a,o/と中舌母音/e/のグループと狭舌母音/i,u/の 2 クラス識別を SVM で行い，84－95%の識別率を得ている[8]．Matsumoto らは，日本光電製の Neurofax EEG-1100 の 128 チャンネルのうち 19 チャンネルを用いて 1000Hz サンプリング後 250HZ にダウンサンプリングし，1 秒間の 5 母音の想起音声に対して，CSP(Common Spatial Patterns[9])による統計的処理に基づく特徴量と RVM-G(Relevance Vector Machine with Gaussian kernel)を用いて，2 対の母音の識別において約 79%の識別率を得ている[10]．Rojas らは，EMOTIV 社製の EMOTIVE-EPOC を使用してサンプリング周波数 128Hz の 6 チャンネルの脳波を用いて，Blackman-Turky 変換で周波数分析後, Simbolic Aggregate Approximation という方法で、6×6 個の幾何的特徴を抽出し、SVM (Support Vector Mavhine)でスペイン語の/a/と/e/の 2 クラス識別で約 85%の識別率を得ている[11]．Morooka らは，Guger 社製の Mobilab+を用いて 5 母音と無音の想起脳波をサンプリング周波数 250Hz，8 チャンネルで測定し，母音対（例：a-i）の認識実験を行っている．各母音を 3 秒間想起し，ビープ音等を含め，1 サンプル 6 秒の測定時間で 10 サンプルを 1 セットとして 3 セット測定した[12]. 3 秒の想起音声区間から，フレーム長さ 0.2 秒，フレーム周期 0.1 秒で 29 フレームを抽出し、各フレームから平均値，分散，標準偏差，歪度の 4 つの統計量, 計 4×8＝32 次元の特徴パラメータを使用して SVM の認識器で, 3 名の被験者の想起音声に対して，約 70%の母音対の認識率を得ている．Nguyen らは，64 チャンネル，256Hz サンプリング後，8H～70Hz の帯域フィルタ通過波形（60Hz のノッチフィルタで電気ノイズ除去も行っている）に対して，独立成分分析後 CSP パターン（統計的特徴抽出）と RVM(Relevance vector machine)を使用して，3 母音/a,I,u/に対して約 49%の識別率を得ている[13].

D'Zmura らは、/ba/ と /ku/ を 3 種類のリズム（/ba/.../ba/…/ba/…/ba/や/ba/ …/ba//ba//ba/, /ba/./ba/./ba/）で想起した脳波を用いて合計 6 カテゴリの識別実験を行っている。Electrical Geodesics 社の 128 チャンネルの脳波計を用いて眼球や首、こみかみに近い 18 チャンネルを除くなどの前処理後、ヒルベルト変換で得られるエンベロープを用いて、ベータ波(13-30Hz)で約 70%の識別率を得ている（アルファ波(8-13Hz)で約 50%、ガンマ波(30-80Hz)で約 55%）[14]. Brigham らは D'Zmura と同じデータを用いて、128 チャンネルから、独立成分分析でランダムさの尺度を表す Hurst exponent を用いて雑音コンポーネントを除いた後、PCA 分析で主成分の脳波を抽出し、線形予測係数を用いて NN 法で/ba/と/ku/を約 60%で識別している（被験者独立）[15]。

勿論, 想起音声による 5 母音の識別研究も行われている．Aguila らは EMOTIVE　EPOC+TM 脳波計を用いて 5 母音の認識を行っている[16]. サンプリング周波数は 128Hz で，50Hz と 60Hz を除去した 14 チャンネルの脳波を使用している．これから，線形特徴（平均、分散、標準偏差）と非線形特徴（近似エントロピー、フラクタル次元、Hurst 指数）を抽出して，各母音 60 個の想起音声の 35 個を学習データに使用してニューラルネットワークを学習して 5 母音の認識を行い，10 人の平均で約 68%の認識率を得ている．この結果は筆者らの知る限り，5 母音の認識率としては世界最高水準である．しかし，認識率を比較するときには，学習に使用するデータ数など実験条件に注意を要する．

以上の研究例からもわかるように、脳波のチャンネル間の位相差を瞬時周波数を用いる検討はあるものの[17]、想起音声の EEG の信号処理として位相スペクトラムを積極的に利用したものはない[18].

本研究では，5 母音 (/a/，/i/，/u/，/e/，/o/) を想起した際の EMOTIV－EPOC で非侵襲的に計測した脳波 (EEG 信号) から，フーリエ変換で得られる 8Hz～24Hz, 32Hz, 40Hz に相当する低次 X(X=3，4，5) 次元での相対位相特徴および振幅スペクトルで特徴抽出を行い，混合ガウス分布 (GMM：Gaussian Mixture Model) を用いて，各母音かつ各チャネルでの認識率を求めることで，脳波から抽出する特徴パラメータについて検討することを目的とした．また，脳波に重畳する雑音を除去するための独立成分分析の有効性についても検討し，日本語 5 母音の想起音声に対して約 60%の認識率を得ることができた.

2. 脳波測定

2.1 測定装置

脳波測定器はオーストラリアの Emotiv System 社製の EPOC を用いた．この測定器はヘッドセット型の脳波計である. 頭部への電極配置の規格である国際 10-20 法とは若干異なるものの，数は 14 個と比較的多く頭部を広くカバーしている (図 1). また, サンプリング周波数は 128Hz と設定されている.

EMOTIVE-EPOC は無線のヘッドセット型であるため，装着が容易である．また，洗浄液や食塩水を十分に染み込ませたスポンジであれば，導電状態が比較的短く，装着作業を含めて 15 分から 20 分程度で利用できることが多い. もちろん，髪の量が多い場合には，スポンジが頭皮に触れにくいことに加え，髪が液を吸ってしまうため，髪の掻き分けや液の補充などで時間がかかる．しかし，それは他の脳波計でも同様であり, Emotiv EPOC が専用のジェルではなく，被験者の不快感が少ないコンタクトレンズの洗浄液や食塩水を利用する点で, より装着の敷居が低いといえる. 研究用のヘッドセットには生データの取得のためのソフトウェア Emotiv TestBench も用意されている．このソフトウェアは，各電極の導電状態と計測した脳波のリアルタイム表示および保存ができる．また，このソフトウェアを用い

ることで脳波計正しく測定できる場所に設置されているかを目で見ることができる.

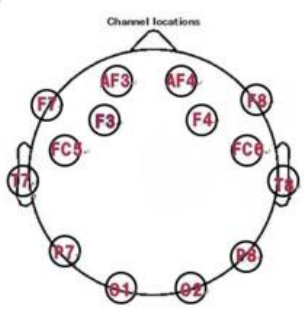

図 1　Emotiv　EPOC の電極配置

2.2 測定手順

被験者には実験室の中でピンクノイズ(1000kHz)を提示後に母音 1 つを想起する課題を行った．被験者は，両耳にヘッドホンを装着し，消灯した部屋の中で椅子に座り，安静した状態を保ちながらピンクノイズ音を聴取した後、開眼状態で母音を想起した（閉眼状態ではアルファ波が生じやすく心理状態の測定では閉眼状態が望ましいが、ベータ波やガンマ波をスペクトル解析する場合は開眼状態でよいと考えられる）．計測したデータは，母音/a/，/i/，/u/，/e/，/o/の 5 クラスで，それぞれ 20 サンプルずつ計測した．以下のセッションで想起してもらう．

母音の計測セッション

/a/→/i/→/u/→/e/→/o/→/a/→/i/→/u/→/e/→/o/→/a/→/i/→/u/→/e/→/o/→/a/→/i/→/u/→/e/→/o/→/a/→/i/→/u/→/e/→/o/

これを 4 セッション繰り返すことで各母音の 20 サンプルを計測する．1 サンプルの計測にかかる時間は 5 秒であり，計測開始 2 秒後にヘッドホンから被験者に想起してもらう母音の音を音声呈示したのちに，ピンクノイズを聞かせ，測定中の無関係な他の母音に影響されないように脳へのリセットをかけて,その後呈示された母音の音を約 2 秒間想起してもらう．この流れを模式的に表した図を図 2 に示す．また，脳の疲労を抑えるために，2 セッションごとに 5 分程度の休憩を挟んでいる．

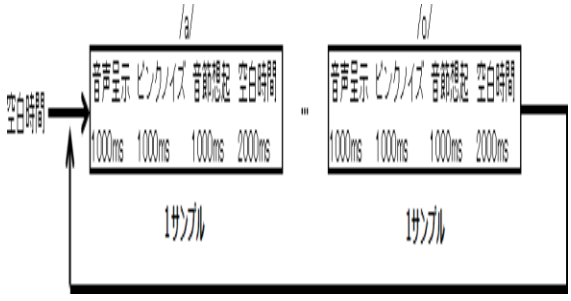

図 2　/a/から/o/までの計測手順

3. 特徴パラメータ

3.1 独立成分分析

独立成分分析 (ICA：independent cpmponent analysis) は，観測信号が統計的に独立な成分の線形和であるという仮定のもと，N チャンネルの観測信号 x(t)から混合前の信号 s(t)を推定する方法である．これらの関係は、行列 $A \in R^{N \times N}$ によって次のように表される．

$$x(t) = As(t) \quad (1)$$

ここで，$s(t) \in R^{N \times N}$ は統計に独立である信号源である．ただし，実際に s(t) は観測できない．

このモデルのもとで，$W = A{-}1$ となるような W を A に関する知識なしに見つける問題が ICA である．具体的には，

$$y(t) = W\,x(t) \quad (2)$$

の要素が互いに独立になるような行列 W を見つけることになる．W を分離行列，y(t) のそれぞれの要素を独立成分と呼ぶ．ICA のモデルでは，y(t) と s(t) は完全に一致している必要はなく，振幅の任意性と，チャンネル同士の入れ替えを許容している．

いま W が求まったとして，y(t) の中には，眼電図に由来する独立成分が含まれているはずである．また，筋電や脈拍などに起因するアーティファクトも独立成分としてあらわれる．アーティファクトに起因する成分 (複数あってもよい) を $y_i(t) = 0$ と置き換えて，$\tilde{y}(t)$ とする．アーティファクト除去した脳波は，

$$\tilde{x}(t) = W^{-1}\tilde{y}(t) \quad (3)$$

として得られる。本実験では、オープンソフトウエアである EEGLAB[19]を使用した。

なお、Thammasan らは、EEG による音楽の感情認識で、眼球運動ノイズの除去のために、眼球に近い AF3 と AF4 を含めて ICA 分析を行っている[20]。本稿では、３か所の隣接した３チャンネルを用いて ICA 分析を行った。

3.2 振幅スペクトル

脳波は時間変化を追うのに適しているので，周波数解析の際は，通常短時間フーリエ変換による時間周波数スペクトログラムを算出する．短時間フーリエ変換の際に用いる窓関数は，通常用いられるハミング窓を用いた．

音声認識のための信号分析の目的は，与えられた信号を生成した調音フィルタの性質を信号より推定することにあり，信号の周波数領域における表現がその基礎を与える．音声から連続する数十 ms 程度の時間長の信号区間を切り出し，切り出された信号が定常確率仮定に従うと仮定して, スペクトル解析を行う. すなわち, 与えられた信号 x(n) に長さ N の分析窓を掛けることで信号系列 Xω(m;k) を取り出す．添え字 k は信号の切り出し位置に対応している．すなわち，k を一定間隔 T で増加することで，定常とみな

される長さ N の音声信号系列 Xω(n) (n = 0, …, N − 1) が間隔 T で得られる．この処理はフレーム化処理と呼ばれ，N をフレーム長，T をフレーム間隔と呼ぶ．また，フレーム化処理を行う窓関数 ω(n) としては，ハミング窓を用いる。フレーム化処理によって得られた音声信号系列の短時間フーリエスペクトルは，離散フーリエ変換 (DFT) で求められる．複素数表現で得られるスペクトラムの振幅値を振幅スペクトラムと呼び、位相部分を位相スペクトラムと呼ぶ。

例えば 512 点で DFT を行えば、256 個の周波数に対して離散スペクトラムが得られ、それぞれを周波数ビンと呼ぶ．フーリエ変換した複素表現 X′(k) が音声のスペクトル表現として最も一般的に用いられる．音声信号の音素的特徴は主として調音フィルタの振幅フィルタの振幅伝達特性に含まれている．したがって，音声認識においては，音声信号の振幅スペクトル, あるいはその 2 乗値であるパワースペクトルが注目すべきスペクトル表現である．

3.3 相対位相

相対位相は話者認識のために提案された優れた特徴量である[21][22][23]．MFCC は振幅スペクトル $|X(\omega)|$ のみを用いた特徴量であり，位相スペクトル $\theta(\omega)$を無視しているため，その位相スペクトル $\theta(\omega)$を話者の特徴量として用いるが，音声の位相スペクトル $\theta(\omega)$は位相ラッピングの問題のため，ほとんど雑音のようなものになる．また，同じ音声波形であってもフレームの切り出し位置によって位相スペクトル $\theta(\omega)$ が変動してしまう問題があり，話者の特徴抽出には適さない．

この問題への対処として，ある基準とする周波数成分の位相を一定と解釈し，他の周波数における位相を相対的に求める手法が提案された．基準周波数 ω_b のときの位相値を θ_bと一定であるとすると，

$$|X(\omega_b)| \times e^{j\theta(\omega_b)} \quad (4)$$

$$\rightarrow \quad |X(\omega_b)| \times e^{j\theta(\omega_b)} \times e^{j(\theta_b - \theta(\omega_b))} \quad (5)$$

となる．これを基準に他の周波数成分 ω における位相を求めると，

$$|X(\omega)| \times e^{j\theta(\omega)} \times e^{j\frac{\omega}{\omega_b}(\theta_b - \theta(\omega_b))} \quad (6)$$

となる．ここで位相成分のみ注目してみると，

$$\tilde{\theta}(\omega) = \theta(\omega) + \frac{\omega}{\omega_b}(\theta_b - \theta(\omega_b)) \quad (7)$$

が得られる．基準位相 θ_b は任意であるので $\theta_b = 0$，とすると次式が得られる．

$$\tilde{\theta}(\omega, t) = \theta(\omega, t) - \frac{\omega}{\omega_b}(\theta(\omega_b, t)) \quad (8)$$

相対位相は，全体の基準位相をもとに各フレーム内の相対位相値を計算するため，異なるフレーム間の位相を統一的に比較することができる．位相は 2πn の周期で同じ位相となることと，-π と π との不連続性を解消する目的で θ の代わりに、対応する座標値に変換した {cosθ, sinθ} を位相特徴として用いる．

4. 認識実験

4.1 使用データと特徴パラメータ

被験者男性 2 名 (21 歳の YZ と 22 歳の WB) で, 被験者毎に 2.2 節で示した測定手順に従い測定した 5 母音 (/a/, /i/, /u/, /e/, /o/) 各 20 回の計 100 個の測定データを使用した．

サンプリング周波数は 128Hz，GMM の混合数は 8，1 母音あたりの区間は、ピンクノイズ提示直後からの 1500ms の 192 サンプル点，フーリエ変換は窓シフト長 62.5ms(8 サンプル点), 窓長 125ms(16 サンプル点) で行った．1 想起音声当たりのフレーム数は 23 であり，各母音 20 個のうち 15 個(各母音 345 フレーム=15 個× 23 フレーム) を学習データとし，残りの 5 個 (各母音 115 フレーム=5 個× 23 フレーム) をテストデータとした. 今回の研究では，特徴パラメータとして振幅スペクトルと相対位相特徴を用いた. 今回の研究ではフーリエ変換 (FFT) を 16 点で行い，振幅スペクトルでは 0 次の直流成分を除くために低次の 3(1-3)，4(1-4)，5(1-5) 次元を用いて特徴抽出を行った．相対位相特徴では，振幅スペクトルに対応する低次の位相スペクトル 1- 3，1-4，1-5 次元の相対位相を座標値の sin， cos に変換して合計 6，8，10 次元を用いて特徴抽出を行った. つまり, 使用した周波数帯 域は 3 次元で 8Hz～24Hz，4 次元で 8Hz～32Hz，5 次元で 8Hz～40Hz である．

4.2 認識方法

今回の研究では，認識法として GMM(Gaussian mixture model) を用いた. GMM は複数の正規分布の重み付け和で確率分布を表現する手法であり，音声から抽出した特徴量をモデル化することができる．母音 k のモデル λ_k は、m 番目のガウス分布に対して平均ベクトル、共分散行列、m 番目のガウス分布の混合重みのパラメータ集合 $\{\mu_{km}, \Sigma_{km}, \omega_{km}\}$ の組みで表され，学習用のデータからこれらのパラメータを推定する．

パラメータの学習は，最尤（Maximum Likelihood: ML）推定に基づいておこなわれる．すなわち想起母音 k の脳波データから抽出した T フレームの特徴ベクトル系列 X = $(x_1, x_2, \cdots, x_T)$が観測されたとき母音モデル λ_k による対数尤度を最大にする問題となる．

$$\hat{k} = \arg\max_{\lambda_k} \log P(X|\lambda_k) \quad (9)$$

最尤パラメータ $\hat{\lambda}_k$ を求めるために，尤度の期待値を最大化する方向にパラメータを更新する EM アルゴリズムがよく用いられる．

認識は, 入力特徴パラメータ系列(15 フレーム)に対して，各母音ごとの GMM で尤度の和を求め，最大の尤度となる

モデルの母音を認識結果とする

4.3 認識結果

(a) 5 母音の認識

5 母音の想起音声の認識実験結果を表 1 と表 2 に示す．表 1 は独立成分分析前の脳波による認識結果の被験者 2 人の平均値を，表 2 は隣接 3 チャンネルの脳波を独立成分分析した後，第 1 主成分（分散の大きい成分）の脳波による音声認識結果を示す．

まず，独立成分分析前の振幅スペクトルと相対位相の比較では、大きな差はなく，チャンスレベルの 20%を少し上回る程度である．チャンネル別では，AF3, O2 による認識率が高い．特徴パラメータを併用することで，最大 30～40%の認識率が得られた．

独立成分分析後の脳波では, いずれの 3 チャンネルを使用しても大きな認識率の改善が見られた．そのなかでも{AF4, F4, FC6}の 3 チャンネルが振幅スペクトルや相対位相，被験者二人に対して安定して良かった．特徴パラメータを併用することにより, 被験者 WB で 60%, 被験者 YZ で 56%, 二人の平均で 58%の認識率が得られた．これは，学習に使用したデータが各母音 15 個の想起母音であることを考えると，Aguila らの結果に匹敵すると考えられる．

表 1　5 母音の認識率（被験者二人の平均）[%]

(a)　振幅スペクトラム

X次元の振幅スペクトル	チャンネル名						
	AF3	F3	F7	FC5	T7	P7	O1
X=3	26	24	32	20	28	18	12
X=4	28	18	26	20	24	24	16
X=5	34	20	18	20	18	20	16
X次元の振幅スペクトル	チャンネル名						
	O2	P8	T8	FC6	F4	F8	AF4
X=3	24	16	22	20	14	20	18
X=4	24	16	22	24	18	22	26
X=5	22	16	26	18	24	18	16

(b)　相対位相

X次元の相対位相特徴	チャンネル名						
	AF3	F3	F7	FC5	T7	P7	O1
X=3	24	24	16	20	10	6	12
X=4	32	14	26	22	22	24	12
X=5	30	16	16	24	18	18	16
X次元の相対位相特徴	チャンネル名						
	O2	P8	T8	FC6	F4	F8	AF4
X=3	26	20	8	16	22	22	20
X=4	22	22	14	24	30	28	22
X=5	26	20	10	18	26	24	18

(c) 振幅スペクトラムと相対位相の併用

次元数	チャンネル名						
	AF3	F3	F7	FC5	T7	P7	O1
X=3	32	28	36	28	28	18	20
X=4	40	28	32	28	28	28	20
X=5	24	32	20	24	20	24	24
次元数	チャンネル名						
	O2	P8	T8	FC6	F4	F8	AF4
X=3	32	24	30	22	22	34	28
X=4	26	26	26	30	34	28	28
X=5	34	22	26	26	36	24	24

表 2　独立成分分析後の 5 母音の認識率[%]

(a)　振幅スペクトラム

次元数	被験者WB			被験者YZ		
	AF3,F3,FC5	AF4,F4,FC6	O1,O2,P7	AF3,F3,FC5	AF4,F4,FC6	O1,O2,P7
X=3	44	44	40	52	40	36
X=4	56	56	44	40	48	52
X=5	52	52	48	44	44	48

(b) 相対位相

次元数	被験者WB			被験者YZ		
	AF3,F3,FC5	AF4,F4,FC6	O1,O2,P7	AF3,F3,FC5	AF4,F4,FC6	O1,O2,P7
X=3	40	48	40	24	36	36
X=4	44	44	48	36	56	52
X=5	40	56	44	44	56	52

(c) 振幅スペクトラムと相対位相の併用

次元数	被験者WB			被験者YZ		
	AF3,F3,FC5	AF4,F4,FC6	O1,O2,P7	AF3,F3,FC5	AF4,F4,FC6	O1,O2,P7
X=3	48	56	48	52	48	48
X=4	60	56	48	44	56	56
X=5	56	60	52	44	56	52

(b) 二母音間の識別

表 2(c)の手法で, 二母音間の識別をした結果が表 3 である．二母音の識別率は約 72%であり，特に，/a/と/i/の識別率は 90%，/u/と/0/の識別率は 100%と従来の研究結果を上回っている．一方、60%の認識率の母音ペア（/a/と/e/、/a/と/o/、/i//と u/）もあり、まだ安定性に欠ける。詳細に結果を分析してみると、母音間の尤度の差よりもサンプル間の尤度の差の方が大きいものが数多く見られた。

(c) 被験者独立の 5 母音の認識

通常、想起音声の認識は、被験者依存で行う場合が多い。これは想起音声の脳波が被験者に依存していることによる。本実験では、被験者が 2 名であるので、1 名の想起音声を学習に使用し、他の 1 名の想起音声を認識した。結果を表 4 に示す。表より、振幅スペクトルと相対位相の特徴量による差は小さく、約 36%の母音認識率が得られており、チャンスレベルの 20%よりは高い認識率であった。

表 3　二母音間の識別結果[%]
(対角右上:被験者 WB、対角左下:被験者 YZ)

ペア	/a/	/i/	/u/	/e/	/o/
/a/	—	100	80	60	60
/i/	80	—	60	80	60
/u/	60	60	—	60	100
/e/	60	60	100	—	60
/o/	60	60	100	80	—

表 4　被験者独立の想起 5 母音の認識率

Feature parameter	WB → YZ			YZ → WB		
	AF3, F3, FC5	AF4, F4, FC6	O1, O2, P7	AF3, F3, FC5	AF4, F4, FC6	O1, O2, P7
Spectrum (X=4)	28	36	32	28	32	36
Spectrum (X=5)	28	32	28	32	36	28
Phase (X=4)	32	28	36	36	36	40
Phase (X=5)	28	32	32	32	28	32

5. むすび

本稿では、唇や舌などの発声器官を動かさないで音声を想起した際に観測される複数チャンネルの脳波に対して、独立成分分析で雑音成分を除去後，振幅スペクトルと相対位相を特徴パラメータとして、5 母音の認識実験を行った。

GMM を識別器として用いて，５母音の想起音声に対して約 60％の認識率を得た．また、二母音間の識別に対しては，約 72％の識別率を得た．被験者独立の条件でも認識実験を行い、チャンスレベルよりも高いに認識率を得ることができた。脳波による音声認識において独立成分分析と相対位相情報の有効性を示した。

今回の実験は、被験者が 2 名と少ななかったので、今後は少なくとも 5～10 名程度に増やし、不特定話者に対しての実験も進めていく予定である。

謝辞　脳波の測定に関しては，中部大学工学部ロボット理工学科の稲垣圭一郎講師に協力を得た．

参考文献

[1] M. M. AlSaleh, M. Arvaneh, H. Christensen, R. Moore, “Brain-computer interface technology for speech recognition: a review”, Proc. APSIPA, 5 pages, 2016.
[2] C. Cooney, R. Foli, D. Coyle, “Neurolinguistics research advancing development of a direct-speech brain-computer interface”, iScience, Vol.8. pp.103-125, 2018.
[3] A. Porbadnigk, M. Wesyter, J-P. Calliess, T. Schultz, “EEG-based speech recognition”, Biosignals, pp.376-381, 2009 .
[4] 神崎卓丸：脳波による音節認識、早稲田大学、情報理工学研究科、修士論文、2017.2.
[5] C. S. Dasalla, et al. “Single-trial classification of vowel speech imagery using common spatial patterns”, Neural Netw. 22, 2009
[6] S. Iqbal, M. Shanir, Y. U. Khan, O. Farooq, “Time domain analysis of EEG to classify imagined speech”, Proc. Int. Cof. Computer and Communication Technology, pp.794-800, 2016
[7] A. Riaz, S. Akhtar, S. Iftikhar, A. A. Khan, A. Salman,""Inter comparison of classification techniques for vowel speech imaginary using EEG sensors”, Proc. ICSAI, pp. 712-717, 2014.
[8] L.C. Sarmiento, C.J. Cortes, J.A. Bacca, “Brain computer interface (BCI) with EEG signals for automatic vowel recognition based on articulation mode”, Proc. IEEE ISSNIP Biosignals and Biorobotics Conference, 2014.
[9] G. Townsend, B. Graimann, G. Pfurtscheller. “A comparison of common spatial patterns with complex band power features”, IEEE Trans. Biomedical Eng., Vol.53, No.4, pp.641-651, 2006.
[10] M. Matsumoto, J. Hori, “Classification of silent speech using support vector machine and relevance vector machine”, Applied Soft Computing, Vol.20, pp.95-102, 2014.
[11] D. A. Rojas, O. L. Ramos, "Recognition of Spanish vowels through imagined speech by using spectral analysis and SVM", Jour. Information Hiding and Multimedia Signal Processing, Vol. 7, No. 4, pp.889-897, 2016.
[12] T. Moroaka, K. Ishizuka, N. Kobayashi, "Electroencephalographic analysis of auditory imagination to realize silent speech BCI", Proc. IEEE GCCE, pp.648-651, 2018.
[13] C.H. Nguyen, G.K. Karavas, P. Artemiadis, "Inferring imagined speech using EEG signals: a new approach using Riemannian manifold features", Journal of Neural Engineering, Vol.15, pp.1-16, 2017.
[14] M. D’Zmura, S. Deng, T. Lappas, S. Thorpe, R. Srinivasan, “Toward EEG sensing of imagined speech”, Human-Computer Interaction, Part I, HCII, LNCS 5610, pp.40-48, 2009
[15] K. Brigham, B.V.K.V. Kumar,” Imagined speech classification with EEG signals for silent communication: a preliminary investigation into synthetic telepathy”, Int. Conf. Bioinformatics and Biomedical, 2010
[16] Aguila, H.D. Basilloe, P.V. Suarezt, “Comparative study of linear and nonlinear features used in imagined vowels classification using a backpropagation neural network classifier”, Proc. ICBBB, pp.7-11, 2017
[17] M. Hassan, F. Wendling, “Elecroencephalography source connec-tivity”, IEEE Signal Processing Magazine, pp.81-96,May, 2018
[18] Q. Gui, M.V.R. Blondet, S. Laszlo, Z. Jin, “A survey on brain biometrics”, ACM Computing Surveys, Vo.51, No.6, Feb. 2019
[19] 開一夫、金山範明（編）、河内、松本、宮腰（著）：脳波解析入門、東京大学出版会、2016
[20] N. Thammasan, K. Moriyama, K. Fukui, M. Nu N. Thammasan, K. Moriyama, K. Fukui, M. Numano, “Continuous music-emotion recognition based on electroencephalogram”, IEICE Trans. Inf.&Syst., Vol. E99-D, No.4, pp.1234-1241, 2016
[21] 浅川、中川：MFCC と位相情報を用いた話者認識、日本音響学会春季講演論文集、1-P-17, 2007
[22] S. Nakagawa, L. Wang, S. Ohtsuka,“ Speaker identification and verification by combining MFCC and phase information ”, IEEE Trans TASL, Vol. 20, No.4, pp.1085-1095, 2012
[23] L. Wang, S. Nakagawa, Z. Zhang, Y. Yoshida, Y. Kawakami, ”Spoofing speech detection using modified relative phase information”, IEEE Jour. Selected Topics in Signal Processing, Vol.11, No.4, pp.660-670, 2017