

眼球動作に基づいた対話支援システムのための 連続画面なぞり入力手法

房 福明¹ 篠崎 隆宏¹

概要: 眼球動作以外のコミュニケーション手段が絶たれた筋委縮性側索硬化症のような患者にとって、眼球動作を用いた対話は非常に重要である。そのため、眼球動作をもとに文字を入力するシステムの研究が行われている。従来入力方式として、文字の形を眼球でなぞる手法が提案されている。しかし、提案されているシステムでは入力時に各文字の区切りを明確にして入力精度を高めるために、一文字を入力するたびに一点を注視し眼球動作を制止させる仕組みとなっている。このため、一文字入力するごとに待ち時間が発生することが避けられず、入力速度が遅い問題がある。そこで本研究では DNN-HMM を応用することで連続的な眼球入力動作に対応した、高精度な連続画面なぞり入力システムを提案する。眼電位信号を用いた評価実験において、提案システムが従来法と比べて入力速度と精度の両面で優れていることを示す。

1. はじめに

眼球動作をもとにしたコミュニケーションは、筋委縮性側索硬化症 (ALS) [1], [2] のような眼球動作を除いて全身の自由運動が失われた患者にとっては非常に重要である。眼球動作を利用した既存のコミュニケーション法として、介助者が透明文字盤を用いて患者の注視する場所の文字をなぞる手法がある [3]。また、介助者の代わりにコンピュータを用いる手法も開発されている。例えば、眼球動作によってカーソルを制御することで画面上のソフトウェアキーボードから文字を選択する手法や画面上の文字を一定時間注視することで文字を特定する手法が挙げられる [4], [5]。また、文字をグループ分けして画面上の円メニューに表示し、視線でグループ、文字の順に階層的に選択する入力手法もある [6]。この手法により視線がメニューの項目を通過するだけですぐに選択できる利点がある。他に、Dasher [7] という連続入力デバイスもある。視線が画面上の右側に表示している文字に近づくと、文字を左に移動し、画面の中心線を超えたら選択を確定する。また、次に入力可能な文字もそのすぐ右側に表示しているため、連続的に選択することができる。上述の階層的な入力手法と Dasher はソフトウェアキーボードの手法のような注視のための待ち時間が要らないため、速い入力速度と滑らかなコミュニケーションが可能という利点がある [6]。

上述のようなコンピュータを用いた手法の共通点はディスプレイを利用していることである。これに対し、ディスプレイが不要な手法も開発されている。このような手法では文字に対応する眼球動作を行うだけで文字を直接に入力する。そして、ディスプレイが要らないため、ポータブルである [8]。例として、Tsai らは“eye-writing”という文字を眼球で描く手法を開発している [9]。認識性能を向上させるために、文字をなぞる終了時に元の位置に戻すことで、文字と文字を区切る。また、眼球動作を検知するためには眼電位 (EOG) [10] という生体信号を用い、眼球動作方向の変化回数の情報を人工ニューラルネットワーク (ANN) [11] に入力して文字を認識、出力する。数字及び四則演算子を入力の評価対象とした場合、27.9%の認識エラー率が得られている。この手法に対し、Lee らは時系列眼電位信号を利用することで、eye-writing の認識性能を向上させる研究を行っている [8]。彼らは時系列信号を認識するために動的時間伸縮法 (DTW) [12] を用いた。26 個のアルファベットと 3 個の機能文字 (スペース、削除とエンター) を入力の対象とした場合、12.6%のエラー率が得られている。しかし、認識タスクを簡単にするために、1 文字あたり 8 秒の固定時間に設定して文字の最終点を 8 秒に達するまで注視することで、文字を区切る。このような仕組みにより入力に待ち時間が生じ、入力速度が遅くなる問題がある。もしこれらの eye-writing システムにおいて上述の Dasher のように待ち時間が要らずに連続的に入力できれば、速い入力と滑らかな対話が期待できる。

以前に著者らはディスプレイを用いない eye-to-speech

¹ 東京工業大学
Tokyo Institute of Technology, Yokohama, Kanagawa 226-8502, Japan
<http://www.ts.ip.titech.ac.jp>

対話支援システムの提案を行った [13]。このシステムは眼球動作の「上、下、左、右、中」を組合せることで仮名に対応させた入力プロトコルを定義し、これに従って眼球を連続的に動かすことで文字を連続的に入力することが可能であった。眼球動作を検知するために眼電位を利用し、認識するために隠れマルコフモデル (HMM) [14] ベースの音声認識技術を用いた。また、眼球動作のコンテキストによる影響を抑制するために著者らはコンテキスト依存眼電位モデルの提案も行った [15]。これを GMM-HMM にモデル化し、さらに N-gram 言語モデル [16] を組合せた場合、0.9%の仮名認識エラー率が得られている。このシステムに使用する眼球動作は単純であるが、1 仮名を表すために 4 動作を組合せる必要があり、さらに全部で 48 通りがあるため、全部習得するのがユーザにとっては大変であった。

そこで、本研究では連続画なぞり入力手法を提案し、従来の eye-writing システムに対して入力速度を向上させる。そして、提案法を eye-to-speech システムに組み込むことで、入力プロトコルの習得容易性を向上させる。提案法では、文字を基本的なコンポーネント (例えば、画やアルファベットなど) に分解し、眼球でこれらのコンポーネントを連続的になぞることにより文字を入力する。このような連続入力により入力速度を向上させる。また、文字と文字を構築する画及び入力順番の対応関係を辞書に定義すれば、提案法はユーザのもとで使用できる手書きの書き方と書き順と同じように設定することができる。このため、ユーザが提案法を実装しているシステムをすぐに使い始められる。本研究では日本語を入力の対象として提案法を実装し、12 種類のカタカナ画及び対応する眼球ジェスチャーを定義する。また、複雑な連続眼球ジェスチャーをモデル化するために、本研究では DNN-HMM を用いる。実験では Lee らの eye-writing 入力手法より提案法が高精度で速い入力速度であることを示す。

2. ベースラインシステム

前章に説明したとおりに Tsai ら [9] と Lee ら [8] は類似している eye-writing システムを開発している。Lee らは眼電位の時系列信号を用いることで、認識の性能を向上させた。そこで、本研究はより性能が高い Lee らのシステムをベースラインとした。彼らのシステムについて以下に簡単に紹介する。

彼らは 26 個のアルファベットと 3 個の機能文字 (スペース、削除、エンター) を入力対象とした。手書きとは違い、眼球でなぞる場合は全ての眼球トレースが記録されてしまう。このため、なぞりの順番が異なると、同じ文字であっても対応する眼電位信号が異なる。例えば、「t」のような複数個の画で構成するアルファベットが挙げられる。また、全てのアルファベットを手で書けるが、眼球で描く場合ではなぞりにくい部分がある。例えば、「i」と「j」のドット

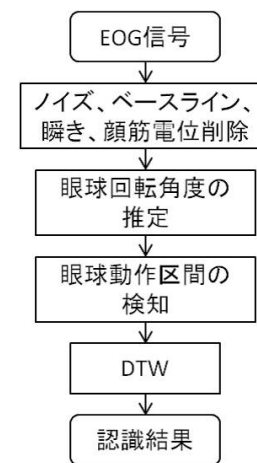


図 1 ベースライン eye-writing システムの信号処理と認識の流れ

が挙げられる。このため、その 3 個の機能文字を含め、12 個のアルファベットのなぞり方やなぞり順番を明確に定義した。これ以外のアルファベットは手書きと同じように定義した。アルファベットのなぞり方の一例として、「i」と「j」のドットをまるに拡大して入力する。そして、入力を行う際に 1 文字あたり 8 秒の固定時間に設定して、眼球は「中」から文字をなぞって終了点を 8 秒に達するまでにそのまま注視する。

このシステムの信号処理及び認識の流れを図 1 に示す。最初は眼球動作を反映する眼電位信号の縦と横の 2 チャンネルを取得し、128Hz までにダウンサンプリングした。次に 160ms のメディアンフィルタを用いてノイズを除去した。その後、直流成分をカットして瞬きと顔筋電位を除去した。また、眼電位信号を検知するための電極の貼り付ける位置を変えると、観測した信号は電極の位置を変える前と変わってしまう問題が生じる。この問題を解決するために、彼らは眼球の回転角度を推定した。具体的には、まずは基準となる 9 箇所を注視することで、校正データを収録した。このうち 1 箇所の回転角度は 0° であり、残りの 8 箇所は縦と横方向それぞれに 14° であった。次に、縦と横方向の回転角度と校正データを 3 次元平面になるように変換行列を推定し、この変換行列を用いて入力データから眼球の回転角度を推定した。最終的に、眼球動作区間を検出し、DTW を用いてテンプレートと照合することで認識を行った。

3. 連続画なぞり入力手法

提案の連続画なぞり入力手法はまず文字を基本的なコンポーネント (画やアルファベットなど) に分解し、このようなコンポーネントを眼球で連続的になぞることで文字を入力する仕組みである。提案法を実装するために本研究では日本語のカタカナを入力の対象と仮定する。詳細については以下に述べる。

カタカナを入力するために本研究では図 2 に示す 12 種

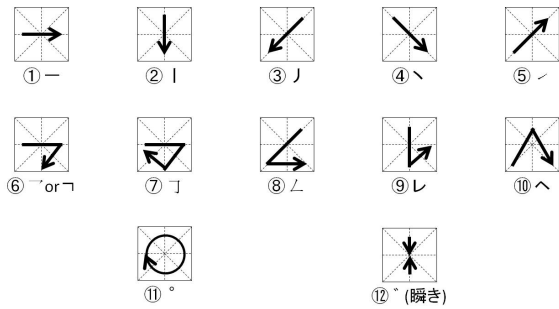


図 2 カタカナ画と眼球ジェスチャーの対応関係

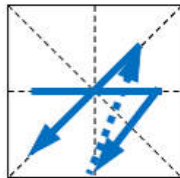


図 3 「ア」の画を連続的になぞる時の眼球トレース。点線は画と画の接続トレース

類の画を用いる。また、画に対応する眼球ジェスチャーは画の形に類似するように設計した。ただし、濁点を表すために 1 回瞬きを用いる。半濁点を表すために時計回りか反時計回りのまるをなぞることを用いる。例えば、これらの画を利用して「ア」を入力するためには 6 番と 3 番の画を用いる。「ウ」は 2 番、2 番と 6 番の画で表す。「ウ」に位置が異なる「縦棒」が 2 つあるが、入力時は文字中の画の位置を考慮せず同じ位置で上から下へ 2 回なぞる。また、画を連続的になぞる際に眼球は止めずに今の画の最終点から次の画の開始点へ直接に移動する。例えば、図 3 に「ア」をなぞる時の眼球トレースを示す。図に示すとおりに実線は画に対応するトレースであり、点線は画と画の接続眼球ジェスチャーのトレースである。そして、文字を表す画及び画の入力順番を制限しない。例えば、「ツ」は 4 番、4 番と 3 番または 2 番、2 番と 3 番の画を用いて表すことが可能である。「ヒ」は 3 番と 9 番または 9 番と 3 番の順番で入力が可能である。これを実装するために文字と画及び画の順番の対応関係を辞書に定義する。このような仕組みにより、ユーザはもともとと使用できる手書きの書き方と書き順に従って入力ができ、システムをすぐに使い始められると考えられる。しかし、入力した同じ画の系列は複数の出力系列に対応する可能性がある。例えば「マ」と「ス」は同じ画と画順で表している。本研究では、この問題を解決するために N-gram を利用して自然な文章を出力する。

画を連続的になぞる場合、観測した眼電位信号は複雑になり、認識も難しくなる。そして、眼球ジェスチャーが複雑であり、画と画の接続トレースはコンテキストによって異なるため、頑健なモデルと認識手法が必要となる。コンテキスト依存眼電位モデルは精密であるが、大量な学習データが必要となる [15] ため、データ量が少ない場合に

は利用することが困難である。そこで、本研究では眼球ジェスチャーのコンテキストを考慮せず、表現能力が高い DNN-HMM を利用して画に対応する眼球ジェスチャーをモデル化する。そして、画と画の接続眼球ジェスチャーをノイズとして画にモデル化する。

眼電位信号から文字列へ変換するために、本研究では重み付け有限状態トランスデューサ (WFST) [17] をもとにした認識デコーダを用いる。デコーダに使用する WFST は HMM 状態から HMM への WFST, HMM からカタカナへの WFST (画と文字の変換ルール) と N-gram を表現する WFST を合成したものをを用いる。そして、DNN の入力層は眼電位信号の特徴量ベクトルに対応し、softmax を活性化関数とした出力層は HMM 各状態の出力確率を表す。眼電位特徴量ベクトルは眼電位信号及びその 1 回と 2 回微分を用いる。具体的には 4.3 節に説明する。

4. 実験

4.1 眼電位データの収録

眼電位信号を検知するため、左目の上、下、左と右に 1 つずつの信号電極を配置し、両眉毛の間にリファレンス電極を 1 つ貼り付けた。これにより横と縦方向の合計 2 チャンネルの眼電位信号を収録した。収録用のデバイスは BlueGain (Cambridge Research Systems Ltd.) というアンプを用いた。BlueGain はバッテリーで駆動し、Bluetooth を経由して眼電位信号をリアルタイムに PC に送信する。本研究では、2 種類の眼電位データを収録した。1 つ目は「孤立画なぞりデータ」であり、ベースラインと比較するためのデータである。2 つ目は提案法の入力速度と認識性能を評価するための「連続画なぞりデータ」である。また、健常者 6 人を被験者とし、このうち 1 人が女性、残り 5 人が男性であった。

孤立画なぞりデータは 3 章に記述した 12 種類の画を入力対象として収録した。入力はベースラインシステムの入力の仕組みに従って、1 回の入力あたり 8 秒の固定時間を設定した。また、被験者ごとに各種の画を 10 回ずつ収録した。ただし、いくつかの画は 9 回や 11 回収録した。眼球の回転角度を推定するため、各被験者ごとの校正データを収録した。

連続画なぞりデータは学習データとテストデータから構成される。学習データは GMM-HMM と DNN-HMM を学習するためのデータであり、テストデータは認識評価を行うためのデータである。学習データは 150 単語 (カタカナの長さが 3 つ以内) とし、5 回ずつ収録した。ただし、いくつかの単語は 3 回から 8 回収録した。テストデータは 25 文章 (カタカナの長さは 20 個以内、合計 386 仮名) に従って、1 回ずつ収録した。これらのデータを収録するための単語と文章は日本語話し言葉コーパス (CSJ) [18] から選択したものである。テスト文章の一例として、「ウサギガ

表 1 GMM-HMM と DTW の画認識エラー率 (%)

| 認識手法 | 被験者 ID | | | | | | 平均 |
|-----------|--------|-----|------|------|------|-----|------|
| | A | B | C | D | E | F | |
| DTW | 11.3 | 6.3 | 40.4 | 26.4 | 10.8 | 5.9 | 16.9 |
| HMM | 2.5 | 3.3 | 13.1 | 15.6 | 4.5 | 5.0 | 7.3 |
| HMM+deriv | 2.5 | 2.5 | 9.9 | 10.6 | 4.5 | 1.7 | 5.3 |

表 2 連続画なぞり法の入力速度 (カタカナ/分)

| | 被験者 ID | | | | | | 平均 |
|------|--------|------|------|------|------|------|------|
| | A | B | C | D | E | F | |
| 入力速度 | 34.0 | 31.0 | 24.4 | 29.9 | 19.9 | 28.1 | 27.9 |

ピョンピョンハネルホド」が挙げられる。また、これらのデータはカタカナの画を連続的になぞって収録したものである。そして、学習データを収録する前に各被験者は 60 単語を 5 回ずつ、テストデータを収録する前に 1 回ずつ正しく入力できるように練習した。

4.2 提案法とベースラインの比較

本節では提案法とベースライン法の認識方法及び入力速度を比較する。認識性能を比較するため、孤立画なぞりデータを用いた。また、信号処理の流れはベースラインに従って行った。ただし、収録設備のサンプリング周波数の違いにより、本研究では 125Hz にダウンサンプリングした。また、本研究は瞬きを用いて濁点を表すため、瞬きの削除処理を行わなかった。そして、文字と文字を区切った場合の認識が比較的に簡単なため、GMM-HMM を用いた。HMM の状態数は 4、混合数は 16 に設定した。学習ツールは HTK [19] を用いた。評価は HMM と DTW とともに 10 クロスバリデーションを用いた。

認識の結果を表 1 に示す。DTW を用いた場合は画の認識エラー率が平均で 16.9% であった。これに対し、HMM を用いた場合では 7.3% と大幅に削減された。これは DTW の性能が単一のテンプレートだけで決定し、HMM が複数の信号からより頑健なパラメータを学習したためであると考えられる。そして、眼電位信号及びその 1 回と 2 回微分の情報を考慮すれば、エラー率をさらに 5.3% までに削減できた。眼電位信号の微分情報は認識性能の向上に有効であることが分かった。

また、表 2 に連続画なぞりのテストデータから観測した入力速度を示す。ベースラインでは文字ごとに固定の 8 秒に設定し、注視で各文字を区切る入力の仕組みであったため、入力速度は 7.5 アルファベット/分の非常に遅い入力速度であった。これに対し、提案法は眼球動作を止めずに画を連続的に入力したため、平均入力速度は 27.9 カタカナ/分であり、ベースラインより大幅に改善した。そして、提案法では一番速い入力速度は 34.0 カタカナ/分となり、遅い場合でも 19.9 カタカナ/分であった。

表 3 GMM-HMM と DNN-HMM のカタカナ認識エラー率 (%)

| 認識手法 | 被験者 ID | | | | | | 平均 |
|---------|--------|------|------|------|-----|-----|------|
| | A | B | C | D | E | F | |
| GMM-HMM | 9.3 | 11.9 | 10.1 | 11.1 | 8.3 | 9.3 | 10.0 |
| DNN-HMM | 3.6 | 9.1 | 8.8 | 5.2 | 3.1 | 6.7 | 6.1 |

4.3 DNN-HMM と GMM-HMM の比較

DNN-HMM と GMM-HMM の認識性能を比較するために連続画なぞりデータを用いた。計算量を削減するために、収録した眼電位信号を 50Hz までにダウンサンプリングした。そして、眼電位信号の各チャンネルの 2 サンプル及びその 1 回と 2 回微分の $12(=(2+2+2) \times 2CH)$ 次元のベクトルを特徴量ベクトルとして使用した。GMM-HMM は 4 状態 16 混合に設定し、HTK を用いて学習した。DNN を学習するために、まず GMM-HMM を学習し、このモデルを用いて学習データに状態情報を付けた。次にこの学習データから DNN のパラメータを学習した。また、DNN の入力眼電位特徴量の前後 7 フレームのコンテキストを考慮して合計 180 次元のベクトルを用いた。DNN の構造は 180-500-250-52(full connected) に設定し、隠れ層の活性化関数は ReLU とした。DNN-HMM の状態遷移確率は GMM-HMM の値をそのまま用いた。そして、DNN の学習データの偏りを削除するため、本研究では学習データの平均を 0 に、分散を 1 に正規化して主成分分析 (PCA) [20] により次元相関削除を行った。また、テストデータを正規化するため、学習データの平均と分散を用いた。次元相関を削除するためにも学習データと同じ固有ベクトルを用いた。DNN の過学習を抑制するために隠れ層のユニットを 20% の確率でランダムにドロップアウト [21] した。N-gram はカタカナをモデル化の単位とした 5-gram に設定し、テスト文章を除いた CSJ テキスト文章を用いて学習した。学習には SRILM [22] を用いた。WFST の合成演算を行うために OpenFst [23] を用いた。

表 3 に GMM-HMM 及び DNN-HMM のカタカナ認識エラー率を示す。GMM-HMM の平均エラー率は 10.0% であったことに対し、DNN-HMM では 6.1% であった。これは DNN の方が表現能力が高いためであると考えられる。画と画の接続眼球トレースはコンテキストによって異なり、これらのトレースをノイズとし、画のモデルを学習したため、表現能力が高いモデルを用いた方がノイズに強いと考えられる。この理由により、DNN-HMM を利用することで 39.0% の相対エラー削減率を達成でき、より高精度の連続画なぞり入力を実現した。

5. まとめと課題

本研究では連続画なぞり入力手法を提案した。提案法では文字の基本コンポーネント (画やアルファベットなど) を連続的に眼球でなぞることで入力を行う。従来の

eye-writing システムの文字を区切って入力を行うような入力の仕組みに対し、提案法では今の画の終了点から次の画の開始点に直接移動し、画を連続的になぞることで入力速度を向上させる仕組みであった。また、提案法はユーザのもともとの手書きの知識を利用することで、提案法を実装した入力システムをすぐに使い始められるという利点もあった。

本研究では、カタカナを入力の対象として提案法を実装した。また、カタカナを入力するため、12種類の画及びこれに対応する眼球ジェスチャーを定義した。同じ画で構成されるカタカナでは区別できない場合があったため、本研究ではN-gramを利用して自然な文章を出力した。そして、眼球動作を検知するため、眼電位を利用した。複雑な眼球ジェスチャーをモデル化するために表現能力が高いDNN-HMMを用いた。眼球動作のコンテキストによって画と画の接続眼球ジェスチャーが異なるが、コンテキスト依存のモデルを用いる場合には大量な学習データが必要となるため、このような接続眼球ジェスチャーの信号をノイズとし、画のモデルに学習した。

実験は健常者6人を被験者として行った。従来の文字の境界をはっきり区切るeye-writingシステムをベースラインとして提案法と比較した。孤立画なぞりデータを用いて認識性能を評価する時、ベースラインのDTWを用いた場合は16.9%の画認識エラー率であり、提案法に使用したHMMを用いた場合は7.3%のエラー率であった。ベースラインより高い認識性能であることを示した。また、連続画なぞりデータを用いて入力速度を評価した際、入力速度は27.9カタカナ/分であったため、ベースラインの7.5アルファベット/分より速い入力速度であることを確認した。そして、DNN-HMMを用いて連続画なぞりデータを認識した場合は6.1%のエラー率を実現し、従来のGMM-HMMの10.0%から36.0%の相対エラー削減率を達成できた。

しかし、現在の提案法を実装しているシステムでは計算時間に関わらず、デコーディングによる出力遅延が発生している。より滑らかなコミュニケーションには速い入力速度に加えて、低遅延出力も必要となるため、今後はより低い出力遅延を実現できるように提案法を改善する必要がある。

参考文献

- [1] L. P. Rowland and N. A. Shneider, "Amyotrophic lateral sclerosis," *The New England Journal of Medicine*, vol. 344, pp. 1688–1700, 2001.
- [2] T. Kihira, S. Yoshida, M. Hironishi, H. Miwa, K. Okamoto, and T. Kondo, "Changes in the incidence of amyotrophic lateral sclerosis in wakayama," *Amyotroph Lateral Scler Other Motor Neuron Disord*, vol. 6, pp. 155–163, 2005.
- [3] S. Söderholm, M. Meinander, and H. Alaranta, "Augmentative and alternative communication methods in locked-in syndrome," *Journal of rehabilitation medicine*, vol. 33, no. 5, pp. 235–239, 2001.
- [4] K. Yamagishi, J. Hori, and M. Miyakawa, "Development of eog-based communication system controlled by eight-directional eye movements," in *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*. IEEE, 2006, pp. 2574–2577.
- [5] S. Handa and Y. Ebisawa, "Development of head-mounted display with eye-gaze detection function for the severely disabled," in *2008 IEEE Conference on Virtual Environments, Human-Computer Interfaces and Measurement Systems*, July 2008, pp. 140–144.
- [6] M. H. Urbina and A. Huckauf, "Alternatives to single character entry and dwell time selection on eye typing," in *Proceedings of the 2010 symposium on eye-tracking research & applications*. ACM, 2010, pp. 315–322.
- [7] D. J. Ward and D. J. C. MacKay, "Artificial intelligence: Fast hands-free writing by gaze direction," *Nature*, no. 6900, pp. 838–838, 2002.
- [8] K. R. Lee, W. D. Chang, S. Kim, and C. H. Im, "Real-time 'eye-writing' recognition using electrooculogram (EOG)," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. PP, no. 99, pp. 1–1, 2016.
- [9] J. Tsai, C. Lee, C. Wu, J. Wu, and K. Kao, "A feasibility study of an eye-writing system based on electro-oculography," *Journal of Medical and Biological Engineering*, vol. 28, no. 1, p. 39, 2008.
- [10] M. F. Marmor, M. G. Brigell, D. L. McCulloch, C. A. Westall, and M. Bach, "ISCEV standard for clinical electro-oculography (2010 update)," *Documenta Ophthalmologica*, vol. 122, pp. 1–7, 2011.
- [11] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.
- [12] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-26, no. 1, pp. 43–49, 1978.
- [13] F. Fang, T. Shinozaki, Y. Horiuchi, S. Kuroiwa, S. Furui, and T. Musha, "HMM based continuous EOG recognition for eye-input speech interface," in *Interspeech*, 2012, pp. 735–738.
- [14] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," in *Proceedings of the IEEE*, 1989, pp. 257–286.
- [15] F. Fang, T. Shinozaki, Y. Horiuchi, S. Kuroiwa, S. Furui, and T. Musha, "Improving eye motion sequence recognition using electrooculography based on context-dependent HMM," *Computational Intelligence and Neuroscience*, vol. 2016, 2016.
- [16] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, pp. 379–423, 623–656, 1948.
- [17] M. Mohri, F. Pereira, and M. Riley, "Weighted finite-state transducers in speech recognition," *Computer Speech and Language*, vol. 16, no. 1, pp. 69–88, 2002.
- [18] T. Kawahara, H. Nanjo, T. Shinozaki, and S. Furui, "Benchmark test for speech recognition using the Corpus of Spontaneous Japanese," in *Proceedings SSPR2003*, 2003, pp. 135–138.
- [19] S. Young et al., *The HTK book*. Cambridge: Cambridge, U.K.: Cambridge University. Eng. Dept., 2009.
- [20] K. Pearson, "Liii. on lines and planes of closest fit to systems of points in space," *The London, Edinburgh, and*

- Dublin Philosophical Magazine and Journal of Science*, vol. 2, no. 11, pp. 559–572, 1901.
- [21] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting.” *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [22] A. Stolcke *et al.*, “SRILM—an extensible language modeling toolkit.” in *Interspeech*, vol. 2002, 2002, p. 2002.
- [23] C. Allauzen, M. Riley, J. Schalkwyk, W. Skut, and M. Mohri, “OpenFst: A general and efficient weighted finite-state transducer library,” in *International Conference on Implementation and Application of Automata*. Springer, 2007, pp. 11–23.