

論文 / 著書情報  
Article / Book Information

題目(和文)	眼球動作入力音声対話支援システムのための高効率入力方式の研究
Title(English)	
著者(和文)	房福明
Author(English)	Fuming Fang
出典(和文)	学位:博士(理学), 学位授与機関:東京工業大学, 報告番号:甲第10546号, 授与年月日:2017年3月26日, 学位の種別:課程博士, 審査員:篠崎 隆宏,小林 隆夫,金子 寛彦,杉野 暢彦,熊澤 逸夫
Citation(English)	Degree:Doctor (Science), Conferring organization: Tokyo Institute of Technology, Report number:甲第10546号, Conferred date:2017/3/26, Degree Type:Course doctor, Examiner:,,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	要約
Type(English)	Outline

## 眼球動作入力音声対話支援システムのための 高効率入力方式の研究

眼球動作を除き発話や筆談を含めた自由運動が全て失われた患者では、通常のコミュニケーション手段が使用できなくなる。このため、眼球動作をもとにしたコミュニケーションが不可欠となる。そこで本論文では、眼球動作を入力とした対話支援システムとその効率的な入力方式の実現に取り組んだ。

第1章「序論」では、眼球動作入力をもとにした既存のコミュニケーション支援システムの仕組みとその問題点を概観し、本研究で取り組んだ課題について記述している。眼球動作に基づいたコミュニケーション支援システムは、ディスプレイを用いるかどうかによって2種類分けられる。ディスプレイを用いる場合は、表示している文字を一定時間注視することで選択入力する方式や視線が文字を通過することで選択する方式などがある。これらのシステムはディスプレイを用いるため、携帯性が低くなる問題がある。これに対し、ディスプレイを用いない場合は文字を眼球動作で表現する入力プロトコルを事前に定義し、これに従って文字を直接入力する仕組みが用いられる。この方式はディスプレイを用いないため携帯性が高い利点があるが、既存のシステムでは認識システム側の制約のため各文字を入力した後に入力を一時的に中止する必要がある。このため入力速度が7.5文字/分と遅い問題がある。このようなシステムの入力速度を向上させるために、尾崎らは入力を中止せずに連続的に文字を入力する眼球動作入力音声対話支援システムを提案した。眼球動作を検知には生体信号である眼電位を利用し、隠れマルコフモデル(HMM)を用いた認識システムにより眼球動作を認識する。また、文字入力のためのプロトコルは「上、下、左、右、中」の眼球動作を組み合わせることで定義されている。この仕組みに基づいて試作した連続眼球動作認識システムでは、眼球動作を連続的に行うと隣接する動作が影響しあい眼電位信号が複雑となることから、認識精度が非常に低くなってしまいう問題が存在する。また、入力プロトコルが複雑なため、ユーザにとって非常に覚えにくい点も問題である。このような背景のもと本研究では、連続眼球動作の認識精度の向上と入力プロトコルの覚えやすさの向上について取り組みを行った。

第2章「統計的パターン認識」では、眼球動作入力音声対話支援システムに用いる統計的パターン認識技術について、一般的な説明を行った。主にHMMの基本構成と学習プロセス、N-gram言語モデルと重み付き有限状態トランスデューサ(WFST)に基づいた認識デコーダの基本原則について記述を行った。

第3章「眼球動作入力音声対話支援システム」では、本研究において基本的なフレームワークとなる眼球動作入力音声対話支援システムの構成を説明している。眼球動作入力音声対話支援システムは入力部、認識部と出力部の3つのモジュールより構成される。入力部は眼電位信号を検知して増幅とデジタル化を行い、認識部に入力する。眼電位を検知するために、6個の計測電極を用いた。そのうちの4個は両眼の上下に配置し、残りの2個は両眼の左右に貼り付けた。そして、アース電極とリファレンス電極はそれぞれ額と両眼の間に貼り付けた。認識部では上述の眼球動作をモデル化の単位としてHMMを学習し、WFSTを用いて入力プロトコルを表現し、認識デコーダを用いて入力した眼電位信号を文字に変換する。出力部では認識された文字の音声を作成してスピーカから出力する。健常被験者5人の眼電位データを用いてその5通りの眼球動作を認識した場合、11.9%のエラー率が得られた。しかし、4動作を組み合わせた仮名を認識対象とした場合、仮名ベースの認識エラー率は36.0%と、非常に高い値となった。これは眼球動作をそのままモデル化の単位とした場合、連続入力された眼球動作信号を効果的にモデル化できていないことを示している。

第4章「認識精度の向上」では、連続入力した眼球動作の認識精度を向上させる提案手法について記述を行った。連続的に動かした眼球動作から観測した眼電位信号は前後眼球動作の影

響を受け、同じ眼球動作であってもコンテキストによって信号の特性が変化する。従来のように眼球動作をそのままモデル化する場合では、異なる眼電位の特性が同じモデルに混ざってしまうため、モデルの精度が低くなる。そこで、本研究はコンテキスト依存眼動素を提案する。コンテキスト依存眼動素はコンテキストに従って眼電位の特性を分け、眼電位特性ごとにモデル化を行う。この方法では同じモデルには同じ特性の信号しか含まれないため、モデル化精度の向上が期待される。さらに認識精度をより向上させるために、**N-gram** 言語モデルを利用する。**N-gram** モデルは文章の自然さをモデル化することができ、認識処理において文章の自然さの評価を手掛かりにすることで認識性能を向上できると期待される。実験では、健常被験者5名を対象に仮名入力を行う眼電位データを収録した。従来の眼球動作のコンテキストを考慮せずに眼球動作をそのままモデル化の単位としたコンテキスト独立眼動素と比べ提案するコンテキスト依存眼動素モデルを用いることで、仮名認識エラー率を **36.0%** から **1.3%** まで大幅に削減することができた。そして、**N-gram** を組み合わせた場合はさらに **0.9%** まで削減することができた。

第5章「入力プロトコルの覚えやすさと速度の向上」では、入力プロトコルの覚えやすさと入力速度を向上させる手法について説明した。入力プロトコルの覚えやすさと入力速度を向上させるために、本研究は縮約文字表を提案し、入力プロトコルを自動的に生成する。この手法はまずユーザが既に知っている文字表（例えば 50 音表）を活用し、表の行と列に動作を対応づけて行と列の動作を組み合わせることで文字を指定する。さらに記憶の必要な動作数を減らし入力速度を向上するため、隣接の複数の行や列を同じブロックにマージし、同じ動作に対応付けることで文字表を縮約する。この仕組みでは入力に曖昧性が存在し同じ動作列は複数の文字列に対応することになるが、**N-gram** を利用して入力の履歴を考慮することで曖昧性を抑制することが可能である。入力速度を向上させるために、行や列の併合とは別に **NULL** 動作も用いた。**NULL** 動作とは何も動作を行わないことで、待ち時間もない仮想動作である。**NULL** 動作をマージしたブロックに対応させると、このブロックに関連する文字の入力に必要な眼球動作数を削減できるため、入力速度を向上させることができる。マージ方法と **NULL** 動作の指定方法によって、縮約文字表の覚えやすさと曖昧性および入力速度が変わるため、これらの要素を同時に考慮して最善な縮約文字表を探索し、得られたプロトコルをユーザに推薦する手法を提案した。プロトコルの習得の容易性を評価するために、8名の被験者による評価実験を行った。実験において被験者にはできるだけ正しく入力できるように入力プロトコルを覚えるよう指示をした。マージと **NULL** 動作を用いなかった 50 音表をベースライン入力プロトコルとして、それらを用いた縮約文字表プロトコルの性能を評価した。その結果、ベースライン入力プロトコルを覚えるために 5.1 分掛かったのに対し、縮約した 50 音表を覚えるために掛かった時間は 2.4 分と短くなり、縮約文字表を用いることで覚えやすさが大幅に向上することが確認された。また、健常被験者1人により入力速度を評価した。試作仮名入力プロトコルと比較した場合、縮約文字表を用いた入力速度は 17.4 仮名/分から 31.8 仮名/分と大幅に向上した。

第6章「まとめと課題」では、本研究のまとめと解決すべき課題を記述した。本研究は眼球動作入力音声対話支援システムを実現するため、まずコンテキスト依存眼動素を提案し、高精度な連続眼球動作入力を実現した。次に、入力プロトコルの覚えやすさと入力速度を向上させるために、縮約文字表を提案した。今後の課題としては、より高精度な眼球動作入力の実現とより覚えやすい入力プロトコルの開発が挙げられる。また、提案法では計算時間に関わらず、認識アルゴリズムに起因する出力遅延が存在する。滑らかなコミュニケーションを実現するため、出力遅延の抑制が必要である。また、実際の患者を対象としたシステムの評価も今後の課題である。