

## 科学研究費助成事業 研究成果報告書

平成 29 年 6 月 8 日現在

機関番号：33907

研究種目：若手研究(B)

研究期間：2013～2016

課題番号：25730117

研究課題名(和文)音声認識における誤認識要因の解明およびシステム利用のための訓練方法の確立

研究課題名(英文) Analysis of factors of misrecognition and study of training methods for use of speech recognition system

研究代表者

柘植 覚 (Tsuge, Satoru)

大同大学・情報学部・准教授

研究者番号：00325250

交付決定額(研究期間全体)：(直接経費) 3,300,000円

研究成果の概要(和文)：本研究は音声認識システムを広く一般に普及させることを最終的目標として掲げ、その前段階として申請期間において『音声認識の誤認識要因の解明』及び『音声認識システム利用のための訓練方法の確立』に関し研究を進めた。『音声認識の誤認識要因の解明』は雑音下数字音声認識において誤認識になりやすい点として、発話区間検出誤りによる無音区間の誤認識が大きく影響することが解った。また、特定の発声者が長期に渡り複数回音声認識器を使用した場合の誤り傾向の分析を行なった。『音声認識システム利用のための訓練方法の確立』に関しては最終的には確立までは至らなかったが、複数の音声認識システムを構築することによりその礎を築いた。

研究成果の概要(英文)：The main theme of my study is the wide spread of speech recognition system. In this study, we focused on "analysis of factors of mis-recognition" and "training method for speech recognition system users". For analysis of factors of mis-recognition, I investigated the factors of mis-recognition under the noisy environment using CENSREC1 speech database. From this investigation, one of mis-recognition caused by the error of voice activation detection. In addition, I collected the speech data when the speaker uttered the speech recognition system for long term. Using these speech data, I investigated the factors of mis-recognition. On the other hand, I unfortunately have not constructed the training system for speech recognition system user. But, I have constructed some speech recognition system, which are a navigation system on potable system, a introduction system of university on personal computer, and so on.

研究分野：音声情報処理

キーワード：音声認識 音声インターフェース 音声認識システム訓練方法 誤認識要因の解明

## 1. 研究開始当初の背景

人間のコミュニケーション手段の一つである音声をコンピュータなどのコミュニケーションに利用するためには、音声認識技術を用いた音声インターフェースが必要である。音声認識技術は、現在に至るまで長年に渡り研究され、カーナビゲーションや携帯電話のインターフェースとして搭載されるようになってきた。しかし、利用者がそれらの音声認識システムを積極的に使用しているとの話はほとんど聞かれず、未だ音声はコンピュータなどのインターフェースとして定着していない。この原因の一つとして、音声認識精度が利用者の要望に十分応えられていないことが挙げられる。このことを解決するため、多くの研究者が音声認識技術を研究し、日進月歩で音声認識精度が向上している。しかしこれらの研究は新しい技術を研究し音声認識精度を向上させることが主であり、誤認識した発声に対する詳細な分析などの誤認識に着目した研究はあまり行われていない。そこで、誤認識発声の分析を行い、誤認識を引き起こす根本的な原因を解明する必要があると思った。誤認識解明ができれば、それらを改善する手法の研究にも結び付き、今後のこの分野の研究に大きな発展が期待できる。さらに、システム開発においても、誤認識の可能性が高い発声を使用しない対話制御を行うことで誤認識を低減し、システム全体の満足度が向上すると思われる。

一方で、音声認識精度が利用者の要望を満たしていない原因は、技術が不十分であること以外に、利用者がシステムを理解していないため、システムの性能を十分に発揮できないことが考えられる。実際、音声認識システム開発者などがシステムを使用した場合、高精度に音声を認識するが、一般利用者が同じ環境で同一システムを利用した場合にはまったく認識しないと言う場面を数多く経験している。これは、開発者などはシステムが認識しやすい発声を経験上理解しており、このような発声を行うための発声訓練を知らず知らずのうちに積んでおり、高精度な認識を行う方法を修得しているからだと考えられる。現在までは、音声認識技術を用いた音声インターフェースの利点の一つである「利用者が特別な知識や訓練を必要とせず利用可能である」ことが制約となり、利用者へシステムの理解を促すなどの利用者へ負担を強いる研究はあまり行われていない。広く一般に普及し使用されているインターフェースであるキーボードを初見から素早く正確にタイピングできる人は皆無であり、利用するためには何らかの訓練を行っている。何らかの訓練が必要であったとしても、有益なインターフェースであれば利用者に受け入れられる。そこで、本研究では『音声認識システム利用のための訓練手法の確立』を目指す。

## 2. 研究の目的

本研究では、音声認識システムの認識精度向上を目指し、『誤認識要因の解明』および『音声認識システム利用のための訓練方法の確立』に関する研究を行う。音声認識システムが誤認識する発声に対し、音響的観点および言語的観点から詳細に分析し、誤認識の要因を明白にする。その後、誤認識を回避する簡易でかつ効果的な訓練方法の確立を目指す。訓練方法は、ユーザ・エクスペリエンスの観点を取り入れ、発声方法や発声内容の観点から研究を進める。この訓練方法が確立できれば、音声インターフェースの普及を促進できる。

## 3. 研究の方法

本研究の目的である『音声認識における誤認識要因の解明』を達成するため、音響的観点および言語的観点より音声を分析し、誤認識となる音響的特徴および言語的特徴を解明する。音響的観点においては音声認識の特徴量を二次元へ射影し誤認識空間の解明、言語的観点においては、音声認識システムに適した発声内容などの解明を行う。これらの分析結果を基に『音声認識システム利用者のための発声訓練手法』を確立する。短時間で効率的に訓練を行うために、訓練に適した発声内容や発声速度を強制する手法などの検討を行い、実際に訓練システムを開発する。

## 4. 研究成果

音声認識における誤認識要因の解明  
不特定話者音声認識実験における誤認識要因の調査として、背景雑音や入力系の際の影響が少ない既存の音声データベースを用い、不特定話者音声認識実験を行い、誤認識要因を調査した。実際には、認識に使用する言語モデルを様々なレベル(単語、音節、音素など)で構築し音声認識実験を行い、言語制約が引き起こす誤認識要因を調査した。この結果より、一部の音素に対する尤度が著しく低下する場合には、単語レベルでの誤認識を生じさせる原因となっていることがわかった。これらの改善策として、裕度のフロアリングなどが考えられるが、実際にどの程度有益なのかの実験までは至らなかった。さらに、雑音下音声認識用評価基盤「CENSREC-1」を用いた音声認識実験を行い、特定の数字の音声特徴が雑音の特徴と類似していることがわかり、それらの数字が音声認識率を低下させていることがわかった。また、認識を行う前の発話区間検出が誤認識に影響をしていることが明確になった。雑音と類似する音声特徴量に関しては、現在使用している特徴量以外の情報を追加することによる回避が考えられる。また、発話区間検出はに関しては、継続的に検討することとなる。また、数名の特定話者が半年に渡り音声入力を行い、その誤り傾向を分析し誤認識要因の解明を試みた。誤認識要因の一つとして、話者が音声認

識システムに精通しておらず、認識可能な語彙が把握していないことに起因する音声認識誤りが存在することを明確にした。また、一般に音声認識システムの評価基準として用いられる単語正解率に対し、正解系列とのマッチング部に改善の余地があることが解った。

音声認識器利用者のための訓練手法の確立前段階として、文法規則に基づく不特定話者連続単語音声認識システムの開発として大同大学案内システムを構築し、音声収録の環境を整えた。本システムは、文法規則に基づき構築されているため、利用者からの音声入力がある程度規定することとなる。このシステムを利用することにより利用者のシステムへの熟練度の低さが誤認識を引き起こす原因の一つであることがわかった。音声収録用のシステムおよび不特定の場所での音声認識システムを使用するため、携帯端末上で動作するナビゲーション機能を有した音声認識システムを構築した。構築したシステムを利用し、様々な状況における音声の収集を行なった。さらに、二足歩行ロボットに音声対話システムを構築することにより実施検証を行った。二足歩行ロボットに音声対話システムを搭載するにはロボットの持つ特有の制限により対話システムに制限がかかり、納得いくシステム作成には至らなかった。科学研究費補助金として補助を受けた期間は2016年度で終了するが、補助期間で明確となった問題点などを継続して研究していく予定である。

#### 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計2件)

- (1) 市川 賢, 北岡 教英, 柘植 覚, 武田 一哉, 北 研二: 種々のテキスト検索モデルの頑健性向上による音声ドキュメント検索の高精度化, 情報処理学会論文誌, Vol.56, No.3, pp. 1003-1012, 2015
- (2) 柘植 覚, 大橋 宏正, 市川 賢, 北岡 教英, 武田 一哉, 北 研二: 音声ドキュメント検索における種々の検討および線形補間係数を自動決定する検索質問拡張, 情報処理学会論文誌, Vol.55, No.6, pp.1625-1636, 2014

[学会発表](計5件)

- (1) Satoru Tsuge, Shingo Kuroiwa: Combination method air and bone conducted speech for speaker recognition in i-vector space, 5th Joint Meeting of the Acoustical Society of America and the Acoustical Society of Japan, 2016
- (2) 大須賀 智子, 石本 祐一, 黒岩 眞吾, 柘

植 覚: AWA 長期間収録音声コーパスの公開について, 日本音響学会秋季講演発表会, pp.241-242, 2016

- (3) 三好 真人, 柘植 覚, 武田 一哉: 局所的な音高遷移パターンの言語依存性に関する分析(Language-dependency analysis for local pitch transition), 日本音響学会春季研究発表会, pp. 911-914, 2015
- (4) 三好 真人, 柘植 覚, 武田 一哉: MIDI音源を用いた旋律の言語依存性に関する分析, 音楽シンポジウム, Vol. 2015-MUS-107, No.53, 2015
- (5) Yoko Takahashi, Shingo Kuroiwa, Yasuo Horiuchi, Satoru Tsuge: Missing feature theory for speaker verification with short utterances, Proc. of 2014 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing, 2014
- (6) Satoru Tsuge, Norihide Kitaoka, Kazuya Takeda and Kenji Kita: STD Method Based on Hash Function for NTCIR11 Spoken Query & Doc Task, Proc. of 10<sup>th</sup> NTCIR Workshop Meeting, pp.390-394, 2014
- (7) 北 研二, 松本 和幸, 吉田 稔, 柘植 覚, 北岡 教英, 武田 一哉: ビット演算に基づく高速な音声ドキュメント検索語検出, 音声ドキュメント処理ワークショップ, 2014
- (8) 北岡 教英, 北 研二, 市川 賢, 柘植 覚, 武田 一哉: 種々のテキスト検索モデルの頑健性向上による音声ドキュメント検索の高精度化, 音声ドキュメント処理ワークショップ, 2014
- (9) 王 龍標, 西田 昌史, 柘植 覚, 網野加苗: 話者認識におけるロバストネス, 日本音響学会誌, 69巻7号, pp.357-364, 2013
- (10) Ken Ichikawa, Satoru Tsuge, Norihide Kitaoka, Kazuya Takeda, Kenji Kita: Spoken document retrieval using both word-based and syllable-based document spaces with latent semantic indexing, APSIPA ASC 2013, 2013
- (11) 市川 賢, 北岡 教英, 柘植 覚, 武田 一哉, 北 研二: 単語空間と音節空間を併用した音声ドキュメント検索手法への潜在的意味解析の適用, 第7回音声ドキュメント処理ワークショップ, 2013
- (12) Satoru Tsuge, Ken Ichikawa, Norihide Kitaoka, Kazuya Takeda and Kenji Kita: Spoken Content Retrieval Using Distance Combination and Spoken Term Detection Using Hash Function for NTCIR10 SpokenDoc2Task, Proc. Of 10<sup>th</sup> NTCIR Workshop Meeting, pp.597-603, 2013

[図書](計0件)

〔産業財産権〕

出願状況（計 0 件）

取得状況（計 0 件）

〔その他〕

ホームページ等

## 6．研究組織

### (1)研究代表者

柘植 覚（SATORU TSUGE）

大同大学・情報学部・准教授

研究者番号：00325250

### (2)研究分担者

### (3)連携研究者

### (4)研究協力者