HOKKAIDO UNIVERSITY

Title	Study on Optimal Spoken Dialogue System for Robust Information Search in the Real World [an abstract of dissertation and a summary of dissertation review]
Author(s)	徐, 昕
Degree Grantor	北海道大学
Degree Name	博士(工学)
Dissertation Number	甲第12405号
Issue Date	2016-09-26
Doc URL	https://hdl.handle.net/2115/63361
Rights(URL)	https://creativecommons.org/licenses/by-nc-sa/2.1/jp/
Type	doctoral thesis
File Information	Xin_Xu_review.pdf, 審査の要旨

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

学位論文審査の要旨

博士の専攻分野の名称　博士 (工学)　　氏名　徐 昕

審査担当者　主　査　教　授　宮永 喜一
副　査　教　授　齊藤 晋聖
副　査　教　授　大鐘 武雄
副　査　准教授　筒井 弘

学位論文題名

Study on Optimal Spoken Dialogue System for Robust Information Search in the Real World
（実環境下におけるロバスト情報検索のための最適音声対話システムに関する研究）

本論文は, 実環境下における, ロバスト情報検索のための音声対話システムに関する研究開発について述べ, 音声認識の導入も考慮した上で全体システムの性能評価を行っている。

近年, 音声認識や自然言語処理技術の高度化に伴い, 実用性の高い音声対話システムが普及する傾向が強まっている。Apple の iPhone に搭載された Siri が大きなセンセーションを引き起こし, 日本でも複数の商用サービスが開始され, 特に情報検索をアシスタントするエージェント系のサービスが多い。しかしながら, 現状の音声対話システムでは, 習慣的かつ継続的な利用が難しいとの課題があり, 原因としては, ①話しかけたいと思わせる魅力的なキャラクタ存在しない②音声認識の誤りによる検索や対話の失敗③他のモダリティに対する優位性や差別化を示せていないと分析された。

本研究は上記諸課題を解決するため,3 つの手法:①対話の継続利用促進にゲーミフィケーション理論を活用し, 音声対話インタフェイスを備えた育成ゲームを設計。②音響距離を基づく 2 パス検索アルゴリズムを使った音声認識結果の音素列とデータベースの高速照合手法を提案, 最も適切の回答候補を提示。③非言語的な情報を活用し, 話者の感情を認識して応答を合わせる仕組みを実環境でも実現するための雑音環境における高精度音声ピッチ抽出技術を提案した。

本研究テーマにおける博士論文の構成としては, 以下のような章立てと記載内容を計画している。
「第 1 章. イントロダクション」　第 1 章では, 本研究における社会的な背景, 研究対象, 研究動向及び論文本体の章立てを紹介している。
「第 2 章. 対話システムの構造について」　第 2 章では, 対話システムの要素技術とそれぞれの発展概要を整理し, 代表的な商用対話システムを紹介している。更に, 現状の対話システムの課題を分析している。
「第 3 章. 継続利用される対話システムの設計について」　第 3 章では, ユーザが受け入り安い, 継続的に使用する対話システムを設計する方法を中心に議論を展開している。まず, 人間中心設計 (HCD) とゲーミフィケーション理論を実践した対話システムの設計プロセスを紹介し, 対話システムの概況を説明した。そして設計された対話システムを用いたゲームシナリオにおける主観評価実験について述べている。最後に商用の利用者ログを分析した結果を解説し, 継続利用される実態を実証した。ゲームシナリオに対する主観評価実験では被験者による高い共感評価得られた。実際の 3,000 人以上の商用利用者のログを分析した結果では Daily active rate は 23% 以上になり,95% 以上の利用者は音声入力を利用した実績を検証出来た。

「第 4 章. 情報検索における対話応答方策について」　第 4 章では, 情報検索タスクにおける適切な対話情報を提示するために, 音声検索および質問選択の技術を中心に議論を展開している。まず, 現状の認識結果の音素列とデータベースを照合する音声検索技術において, 既存技術の全般を紹介した。そして, 音素混同行列 (phonetic confusion matrix) に基づく音響距離を使用した DP ベースの 2 パス検索アルゴリズムを提案し, 提案法の詳細を解説している。間違えた歌詞クエリによる歌詞検索のタスクにおいて従来技術との比較検討を行った。さらに検索情報を明確させ, 意思決定支援として, 情報ゲイン値による質問選択の提案法を解説している。間違えた歌詞クエリによる歌詞検索のタスクにおいて, 提案手法は, 全データベースに対する DP マッチングによる検索と同程度の検索精度のまま, 処理時間は従来の高速 DP マッチング処理より 87% 以上削減できることを示した。
「第 5 章. 対話中の非言語情報の利用について」　第 5 章では, 非言語情報 (パラ言語) の利用による対話システム構築ついての議論をしている。まず, パラ言語の担う役割を整理し, 対話においてパラ言語に注目することの重要性について述べた。特に音声に含まれるピッチ情報は, 発話者の態度・感情などの認識に大きな役割を果たしていることを示した。本研究で目指している実環境での対話システム運用に於いては, ピッチ抽出の耐雑音性が求められる。ここでは, 雑音ロバスト音声処理技術「Running Spectrum Filtering(RSF)」を利用したロバストなピッチ抽出手法を提案し, 従来ピッチ抽出技術との比較検討を行った。
「第 6 章. まとめ」　最後の第 6 章では, 本研究で得られた成果をまとめ, 更に今後拡張あるいは継続すべき研究課題を整理している。

以上より, 本論文では, 実環境下におけるロバスト情報検索のための音声対話システムに関する研究開発について検討し, 有効な対話システムを提案し, その性能評価を行った。これにより次世代の情報化社会に必要とされる新しい音声対話技術の開発を行った。

これを要するに, 筆者は, 実環境下における高性能な音声対話システムの提案・開発・評価を行った。これにより, 音声対話技術に関する多くの有益な知見を得ており, 情報科学・工学の分野に貢献するところ大なるものがある。

よって筆者は, 北海道大学博士 (工学) の学位を授与される資格あるものと認める。