

報告番号	※甲	第	号
------	----	---	---

主 論 文 の 要 旨

論文題目 多様な音響環境下における音声データ収集システムの構築と課題未達成に着目した対話品質の評価
氏 名 原 直

論 文 内 容 の 要 旨

音声認識技術の応用事例は日々増え続けており、法廷書き起こし文作成や衆議院会議録作成など公的機関における活用がなされている。また、通常の生活においてもカーナビゲーションシステムで音声認識による“アイズフリー・ハンズフリー”のナビゲーションを実現するために用いられている。しかし、音声認識システムには認識誤りなど不可避の問題も多いことから、どんな利用環境下でもユーザビリティの高い音声インタフェースを実現することは困難であり、音声インタフェース普及の阻害要因となっている。そこで、本論文では、音声対話インタフェースを様々な環境下で利用した際に起こりうる問題点を明らかにし、そのような問題を自動的に検出する手法を確立することにより、音声対話インタフェースの性能向上と普及に貢献するための研究を行った。

まず、多様な利用環境における音声対話インタフェースの利用状況を把握するために、インターネット上のユーザを対象とした音声収集サーバを開発し、このサーバと連携するクライアントとして楽曲検索・購入のための音声対話インタフェースを開発した。音声対話インタフェースの利用課題として、多くのユーザに実験への関心を持ってもらうために、World Wide Web (WWW) 上の楽曲検索サイトに登録されている膨大な楽曲からユーザが欲しい曲を購入するために音声対話インタフェースを利用する課題とした。この実験課題では購入対象となりうる楽曲数は約50万曲もあり、全てを認識対象とした場合、認識辞書の語彙数が膨大になってしまい認識率が低下してしまうと考えた。そこで、本システムではサーバ上のインタフェースで、ユーザが事前に好みのアーティストを数人選ぶことで、そのアーティストに関連したアルバム名や楽曲名のみを含んだ音声認識辞書を自動で作成する機能を実装した。これにより認識対象語彙は総数50万語から平均して約5000語程度に制限することができた。また、ユーザが発話の前にインタフェース上のボタン

を押下する仕組みを備えることで、不要な発話に反応してシステムが誤動作することを防いだ。そして、クライアントシステムには、システムの起動時と終了時にサーバと通信を行い、サーバで作成された個々の音声認識システム用辞書データの取得や、システム利用時に収録された音声や対話ログなどを自動的にサーバにアップロードする機能を実装した。これらの仕組みにより、認識率の低下を避けながら不特定多数のユーザが各ユーザに適応した音声認識システムを利用可能になったとともに、利用時の音声データをインターネットを経由して容易にサーバ上に集めることが可能となった。システムの予備調査としてフィールドテストを行い、32名による約6500発話の音声データを収集するとともに、実際に利用したマイクロホンなどの情報をアンケートにより収集した。このフィールドテストによって本システムを用いる事で様々な環境における多数のユーザを対象とした音声対話インタフェースの利用実験が可能となることが示された。

続いて、開発した音声対話インタフェースを改良してより多数のユーザを対象とする実験を行った。この実験ではユーザ自身が所有する楽曲を検索対象とした。そのため、ユーザが既に知っているはずのアーティスト名、アルバム名、楽曲名だけが認識対象となり、システムに登録されていない単語である未知語の発声はほぼ無い。そして、検索対象をWWWサイトの情報からPC内の情報とすることで、必ずしもインターネットに接続していない状態でも音声対話システムを利用することができるようにした。こうした変更とインタフェースの改善などによりユーザビリティの向上を図ると共に、様々な環境でのシステム利用が可能になった。この音声対話システムを用いてインターネット上での公開実験による大規模なデータ収集実験を行った。この収集実験によって、1359名の被験者による総計488時間の利用データを得た。また、十分にシステムを利用したユーザと判断された500名については、システムに関するアンケート調査も行い、主観的な評価データも同時に得られた。自動的に検出された音声区間について、雑音・音声などのタグ付け、書き起こしなどを行い、コーパスとしての整備を行った。さらに、本実験システムの可搬性を示すために、実車走行時の音声対話システム利用実験も行った。この実験では音声対話を行いながらどの程度運転に不安を感じていたかを調査し、本システムを利用しながらの運転は通常の運転時よりもやや不安に感じていたことがわかった。

最後に、収集された大量のデータを用いた音声対話システムの自動評価手法について研究を行った。一般の音声認識システムの評価では音声認識率が用いられることが多いが、音声対話システムの場合は統合システムとしての総合評価指標であるユーザ満足度や課題達成率などが用いられる。しかし、これらの指標は手作業で付与された書き起こしやタグなどを必要とするため、コストが大きい。従って、自動評価が行えることが望ましい。

まず、音声認識システムの認識精度をユーザが推定した評価値を、本論文では主

観認識率と呼び、満足度を示す心理指標の一つとしてとらえた。分析によって、主観認識率は客観認識率よりも満足度への関係が強いことを示した。この知見を元にしてユーザの利用環境やプロフィール情報からユーザ満足度を推定するモデルをベイジアンネットワークによって構築した。そして、5段階のユーザ満足度の推定実験を行うことで、ユーザプロフィールを利用する有効性を示し、若干の誤差を許容すれば73.2%と高い推定精度が得られる事を示した。

そして、音声対話システムではそのシステムとユーザのやりとりに課題達成の可否に関する重要な情報が含まれていると考え、この情報から課題達成の検出を行う手法について研究を行った。具体的には、ユーザとシステムの間でなされた楽曲検索対話を抽象化されたタグの系列であるとみなし、そのタグ系列をN-gramとしてモデル化を行うことで、ある対話が楽曲検索を行うことができた対話かできなかった対話かを識別する手法を提案した。まず、タグの種類としてシステム発話のみのタグ系列を用いた場合、ユーザ発話のみのタグ系列を用いた場合、両者を用いた場合の3者について、N-gram確率モデルに基づく識別実験を行った結果、両者を用いたタグ系列をN-gramとして扱う有効性が示された。そして、データ収集の事後処理に利用可能なオフライン検出性能の評価実験と、システム実行時の応用が可能な対話途中でのオンライン検出性能の評価実験を行った。その結果、オフライン性能として約87%の識別率が得られ、オンライン性能としてはユーザの4回目の発話までを用いることで、約78%の識別率が得られた。