

# 多様な利用環境における楽曲検索音声対話システムのフィールドテストと評価

原直<sup>†</sup>, 宮島千代美<sup>†</sup>, 伊藤克亘<sup>‡</sup>, 北岡教英<sup>†</sup>, 武田一哉<sup>†</sup>

<sup>†</sup> 名古屋大学大学院情報科学研究科, <sup>‡</sup> 法政大学情報科学研究科

## 1 はじめに

音声認識技術は手軽に利用可能でハンズフリーな入力手段として注目されており, 電話越しの応対システム [1][2] や公共施設における案内システム [3] などに利用されている。これらのシステムは一つの音声認識システムを多数のユーザが利用する構成となっている。しかし, 近年ではカーナビゲーションシステムに音声機能が搭載されるなどユーザ毎にシステムを所有する例が増えているが, 利用時の音響環境は一定ではなくてしまうために単一の認識システムでは必ずしも十分な音声認識性能を得ることができない。

そこで本研究ではインターネットを介したカスタマイズ機能を有した楽曲検索音声対話システム *MusicNavi* を開発した。本システムはインターネットを介して利用者の音声データを収集する機能と音声認識システムに必要なデータである音響モデルや音声認識辞書などをシステムが自動的にアップデートする機能を備えている。これらの機能により一般 PC ユーザによるインターネットを介した音声対話システム利用実験を行うことが可能となる。本論文では, 楽曲検索音声対話システム *MusicNavi* の構成と本システムを用いた様々な利用環境におけるフィールドテストについて述べる。

## 2 楽曲検索音声対話システム

楽曲検索音声対話システムは図 1 のように構成されている。*MusicNavi* はユーザの PC にインストールされるクライアントプログラムであり, 音声インタフェースを通して, ユーザ自身の PC に入っている楽曲ファイルを検索・再生することができる Windows 2000/XP 専用のソフトウェアである。*MusicNavi* は必要に応じて *MusicNavi* を管理するためのリモートサーバとインターネット回線を通じてデータの送受信を行う。

### 2.1 実験データ収集機能

*MusicNavi* はユーザ毎にあらかじめユーザアカウントを発行しており, このユーザアカウントによってユーザを識別している。*MusicNavi* を起動する際, ユーザはまずアカウント情報を入力する。*MusicNavi* は与えられたアカウント情報を元に *MusicNavi Server*

にログインし, プログラムのアップデートパッチをダウンロードし, プログラムのアップデートを行う。この機能により, 全てのユーザが常に最新の *MusicNavi* を容易に使用することができる。

システムの動作ログとしては FLAC (<http://flac.sourceforge.net/>) による可逆圧縮された音声 (16 bit, 16000 Hz, 1 ch) ファイル, 音声区間の切り出し及び認識結果を元に生成されたラベルファイル, 音声インタフェースの動作ログファイルがユーザの PC 内に蓄積される。

メインインタフェースを終了すると最後にログアップロードプログラムが起動し, ユーザの PC に蓄積された *MusicNavi* 利用時の各種ログが *MusicNavi Server* に送信・蓄積される。

### 2.2 楽曲検索機能

ユーザは *MusicNavi* の起動時にユーザの所有する楽曲を含むディレクトリを指定する。*MusicNavi* は指定されたディレクトリに含まれる楽曲ファイルを走査し, 楽曲プレイリストを作成する。プレイリストの構築には楽曲ファイルに含まれている ID3 タグなどのメタ情報を利用する。

### 2.3 音声対話機能

音声認識エンジンには Julian 3.5.2 を用いた。音響モデルは, CSRC の標準日本語音響モデル [4] より, 状態数 3000 (基底分布 129 状態), 性別非依存, 64 混合, PTM triphone モデルを用いた。単語認識を行う文法を用いており, 音声認識用の辞書は, “制御コマンド辞書” と “プレイリスト辞書” に大別される。制御コマンド辞書には音楽の再生を止めるための「停止」や現在読み上げているリスト中の曲を選択するための「それ」などが含まれる。“プレイリスト辞書” は, 起動時に作成された楽曲プレイリスト内のアーティスト名, 楽曲名を認識するために動的に生成された辞書である。読み仮名は音声合成器の出力を元に与えられる。雑音棄却用のモデルとして [3] の GMM

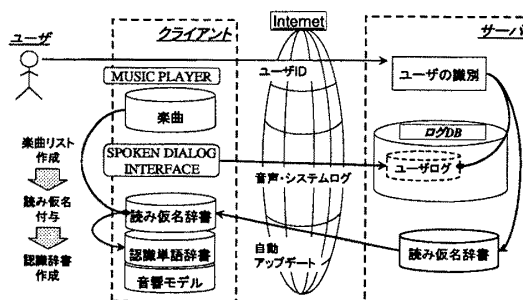


図 1: MusicNavi Client と Server の関係図

Evaluation of training effects by long-term use of a spoken dialogue interface

<sup>†</sup> Sunao HARA, Chiyomi MIYAJIMA, Norihide KITAOKA, Kazuya TAKEDA, Graduate School of Information Science, Nagoya University

<sup>‡</sup> Katsunobu ITOU, Graduate School of Information Science, Hosei University

U ヒライケン  
 S アーティスト 平井堅 を検索しますか？  
 U はい  
 S アーティスト 平井堅 を検索します。  
 S 平井堅の曲が 30 個見つかりました。POP STAR,  
 大きな古時計, ... (以下, 楽曲名を読み上げる)  
 U ポップスター  
 S 平井堅 の POP STAR を再生します。

図 2: MusicNavi との対話例

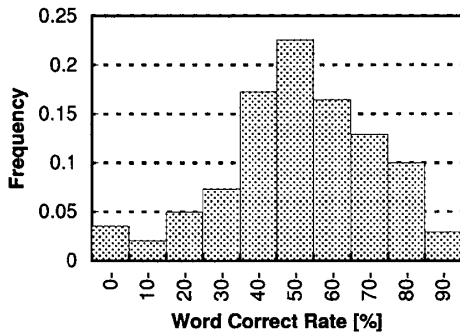


図 3: 各被験者の平均単語認識率分布

(adult, child, cough, laugh, noise) を使用した。

音声合成エンジンには GalateaTalk 1.2 [5] を用いた。合成用の辞書は GalateaTalk 付属の UniDic-1.2.0 を基本にアーティスト名, アルバム名, 楽曲名に含まれる形態素を適宜追加した。

MusicNavi との対話例を図 2 に示す。

### 3 フィールドテスト

本研究では MusicNavi システム利用時のデータ収集実験としてインターネットリサーチによるユーザーアンケート調査と車運転時における利用調査を行った。

#### 3.1 インターネットによる利用調査

インターネットリサーチは 2007 年 6 月 2 日から 2007 年 7 月 23 日まで行った。被験者は 16 歳から 39 歳までの合計 500 名 (男性 311 名, 女性 189 名) である。全ての被験者は, MusicNavi を利用して「5 曲以上再生した」かつ「20 回以上アーティスト名や曲名が認識されたか 40 分以上利用した」という二つの条件を満たしている。本報告では検聴済みである 341 名のデータを用いて分析を行った。

平均単語認識率の分布を図 3 に, アンケート項目である「音声認識機能に対する満足度 (1 が非常に不満, 5 が非常に満足の 5 段階評価)」に対する回答の分布を図 4 に示す。平均認識率は 57.5%, 平均満足度は 3.15 であった。80% 以上の認識率であった被験者は全体の 12.9% である一方, 音声認識機能に対する満足度が 4 以上の回答をしたユーザは 44.3% であることから, 客観評価としては十分な音声認識性能を得られなかったがユーザの主観評価では認識性能にほぼ満足したシステムであったと考えられる。

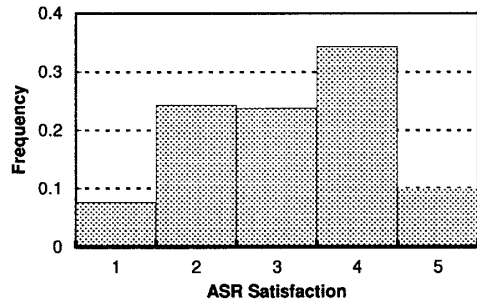


図 4: 各被験者の音声認識機能に対する満足度分布

#### 3.2 実車運転時における利用調査

ドライビングセンシングカー [6] を用いて, 市街地や高速道路で車を運転しながら MusicNavi を利用する実験を行った。被験者は 20 歳から 64 歳までの合計 72 名 (男性 45 名, 女性 27 名) である。実験後のアンケートにより運転中にどの程度不安を感じていたかについて 5 段階 (普段と変わらない場合が 1, 走行することが辛いと感じた場合が 5) の回答を得た。平均評価値は通常の運転時が 1.72, MusicNavi 利用時は 2.55 となっており, MusicNavi を利用しながら車を運転することに不安を感じていたことがわかる。

### 4 まとめと今後の課題

本論文ではインターネットを介してカスタマイズ可能な楽曲検索音声対話システム MusicNavi の構成について述べた。本システムを用いたフィールドテストとしてインターネットでのアンケート調査を行った結果, 44.3% のユーザが音声認識機能に満足しているシステムであることが分かった。しかし, 本システムを用いた自動車運転時における利用実験の結果, ユーザは運転時に何もしない場合に比べてシステムの利用に不安を感じていた。従って, 本システムを自動車内での利用を想定して実装する場合には, 音声対話システムとしてさらなる機能改善が必要である。

#### 謝辞

本研究の一部は文部科学省リーディングプロジェクト「e-Society 基盤ソフトウェアの総合開発」によるものである。

#### 参考文献

- [1] Zue et al., "JUPITER: A telephone-based conversational interface for weather information," IEEE Trans. Speech Audio Processing, vol.8, pp.85-96, Jan. 2000.
- [2] 駒谷他, "音声対話システムにおける適応的な応答生成を行うためのユーザモデル," 信学論, vol.J87-D-II, no.10, pp.1921-1928, 2004.
- [3] 中村他, "実環境音声情報案内システムにおける環境雑音および不要発話の識別," 信学技報, SP2003-172, pp.13-18, Jan. 2004.
- [4] 河原他, "連続音声認識コンソーシアム 2002 年度版ソフトウェアの概要," 情処学研報, SLP-48-1, pp.1-6, 2003.
- [5] Kawamoto et al., "Open-source software for developing anthropomorphic spoken dialog agent," Proc. of PRICAI-02, pp.64-69, Aug. 2002.
- [6] Miyajima et al., "On-going data collection for driving behavior signal," Proceedings of 2007 Biennial on DSP for in-Vehicle and Mobile Systems, July 2007.