

## 汎用PC上で利用された音声対話システムによる音声収集と評価

原 直<sup>†</sup> 宮島千代美<sup>†</sup> 伊藤 克亘<sup>††</sup> 武田 一哉<sup>†</sup>

† 名古屋大学大学院 情報科学研究科 〒464-8603 名古屋市千種区不老町1  
†† 法政大学 情報科学研究科

あらまし 実際の利用環境に近づけた被験者実験を行うために、WWW 経由でカスタマイズ可能な音声対話システム及びそのデータ収集システムを構築した。本システムのユーザは自分のPCに音声認識システムをインストールし利用する。不特定多数のユーザに対応するために、インターネット上に用意されたリモートサーバ上で各ユーザが本システムのカスタマイズすることが可能である。さらに、本システムを利用する各ユーザのPCで収録された音声データはインターネットを通してリモートサーバに送信される。本システムを利用することにより、複数のユーザによる実環境における音声認識システム利用時データを収集することが可能になる。本論文では実際に本システムを使用したフィールドテストを行いデータ収集及び分析を行った。インターネット上に公開して2ヶ月間で59時間のデータが収集され、そのうち約5時間41分のデータ(11351個)が音声区間として検出されていた。認識実験の結果、検出された音声データからシステムに対して発話されていた音声データ4716発話を評価用データとした場合の単語正解率は66.0%であり、各ユーザ毎に教師無しMLLR適応を行うことで単語正解率は70.5%に向上した。

キーワード 音声データ収集, 音声対話システム, 楽曲検索システム, WWWアプリケーション

## Speech data collection and evaluation by using a spoken dialogue system on general purpose PCs

Sunao HARA<sup>†</sup>, Chiyomi MIYAJIMA<sup>†</sup>, Katsunobu ITOU<sup>††</sup>, and Kazuya TAKEDA<sup>†</sup>

† Graduate School of Information Science, Nagoya University

†† Graduate School of Information Science, Hosei University

**Abstract** We developed a user customizable speech dialogue system and a framework for automatic speech data collection in field experiments over the Internet. Users can download and install the speech dialogue system onto their own PCs and customize the system on a remote server for their own use. The speech data recorded on their PCs are transferred to the remote server through the Internet. The system enables us to collect speech data spoken by many users with wide variety of acoustic environments. During a two-month field test, we obtained 59 hours of recorded data including 5 hours and 41 minutes detected as speech, which corresponds to 11351 speech segments. The word correct rate for the 4716 speech utterances spoken to the dialogue system was 66.0%, which was improved to 70.5% after applying unsupervised MLLR for each user.

**Key words** speech data collection, spoken dialogue system, music information retrieval system, WWW application

### 1. はじめに

音声認識システムの性能はモデル学習時とシステム利用時の環境が異なるとき大きく劣化する。利用環境を特定できないPC上の音声認識の性能を向上させるためには、多様な実環境下でPCに入力される音声を大量に収集し、音響モデルを構築することが有効と考えられる[1]。従来研究として、音声認識を用いた実運用システムによる様々な音響環境下での収録が行われてきた。例えば、電話を介した天気情報システム[2]や、市

民向けの多目的コミュニケーションホールに常設された館内案内システム[3][4]などが挙げられる。このように従来研究では、音声認識システムを特定の室内で運用し固有の音響環境下で収録された音声为主要な収集対象とされてきた。しかし、一般家庭等でユーザが音声認識を利用する場合には、音響環境(室内伝達回数や雑音など)だけではなくPC及び周辺機器の利用環境(使用するマイクの種類、マイクと口との距離や位置など)の変動も起こると考えられる。

そこで本研究では一般家庭やオフィスなど多様な利用環境に

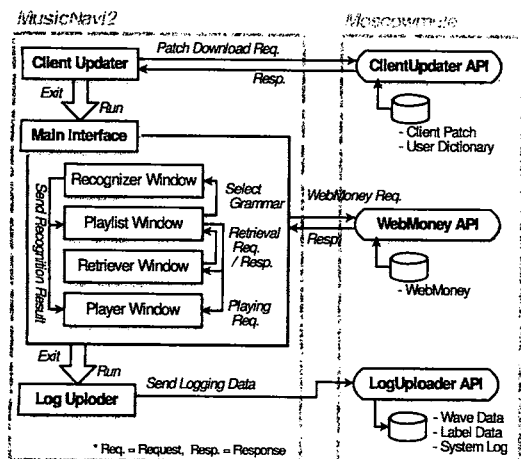


図1 MusicNavi2 と Moscovmule の関係図

において被験者実験及び音声収集を行うことを目的に、WWW を利用した音声対話システム及びそのデータ収集システムを構築した。本システムのユーザは自分の PC に音声認識システムをインストールし利用する。不特定多数のユーザに対応するために、インターネット上に用意されたリモートサーバ上で各ユーザが本システムをカスタマイズすることが可能である。さらに、本システムを利用する各ユーザの PC で収録された音声データはインターネットを通してリモートサーバに蓄積される。本システムを利用することにより、不特定環境における不特定多数の音声認識システム利用データを収集することが可能になる。

本システムを利用したフィールドテストでは、実験への参加やデータの送信は全てユーザの自由意志に任せている。多くの人に何度も実験に参加してもらい、効率的に実験をすすめるために、ユーザの関心が高いタスクとして音声認識による楽曲検索再生システムを構築し、二ヶ月にわたるフィールドテストを行った。本システムでは楽曲を無料でダウンロード可能な仕組みを実装することにより、実験参加者を広く募った。

本報告では、まず本収録システムの概要及びフィールドテストの概要を述べ、次にフィールドテストにより実際に収録された音声データの分析結果について述べる。

## 2. インターネットを介した被験者実験システム

前節で述べた音声認識システムを用いて音声収集実験システムを構築した。本実験システムはインターネットを介した楽曲検索のための音声対話インタフェース“MusicNavi2”及び実験データ収集サーバ“Moscovmule”から構成されている。

MusicNavi2 と Moscovmule の関係を図 1 に示す。MusicNavi2 はユーザの PC にインストールされるシステムであり、Moscovmule は MusicNavi2 を管理するためのリモートサーバで動作するシステムである。MusicNavi2 は必要に応じてインターネット回線を通じて Moscovmule とデータの送受信を行う。

表 1 MusicNavi2 の推奨動作環境。

OS	Windows 2000/XP
CPU	800 MHz 以上
メインメモリ	256 MB 以上
空き HDD 容量	1 GB 以上
音声入力	全二重対応サウンドボード マイクロフォン
ネットワーク	インターネット回線 (1Mbps 以上)

### 2.1 MusicNavi2: インターネットを介した楽曲検索のための音声対話インタフェース

MusicNavi2 はユーザの PC 上で動作するプログラムである(図 2)。ユーザは MusicNavi2 の音声インタフェースを通して、ユーザ自身の PC の HDD に入っている楽曲ファイルを検索して再生したり、WWW 上のサービス Mora から楽曲を検索し、試聴及び購入できる。楽曲購入時の支払いにはプリペイド型電子マネーである WebMoney<sup>(注1)</sup>を利用している。MusicNavi2 の推奨動作環境を表 1 に示す。

MusicNavi2 はアーティスト名やコマンド語の単語認識を基本とした音声認識インタフェースを持つ。検索対象の楽曲はユーザの PC に CD のリッピングなどにより取り込まれた楽曲ファイルとインターネット上の楽曲検索サービスに存在する楽曲ファイルである。インターネット上の楽曲検索サービスとして、(株)レーベルゲート運営のオンライン楽曲検索サービス Mora<sup>(注2)</sup>を利用している。

Mora に登録されている楽曲は約 50 万曲であり、全てのアーティスト名や楽曲名を音声認識辞書に登録した場合に音声認識率が低下してしまい、音声認識を用いた楽曲検索システムとして機能しない恐れがある。そこで、Mora の楽曲はユーザが事前にリモートサーバ上で選択したアーティストのみを検索対象とした。このユーザによる事前選択を行うことで楽曲の検索対象をシステム側が制限することなく、音声認識率の低下を防いだ。フィールドテストの結果、ユーザによるアーティストの事前選択機能により、音声認識に用いる辞書は総語彙数約 50 万語から平均約 5000 語程度に制限された。

#### 2.1.1 音声による検索の手順と対話例

本システムでは、Mora で行う検索手順を参考に検索手順を決定した。図 3 に検索手順を示す。

以下に、アーティスト名からの検索対話例を示す (U はユーザの発話、S はシステムの発話である)。

- 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 平井堅が見つかりました。  
 S 平井堅のアルバムを検索します。  
 S 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)

(注1) : WebMoney: <http://www.webmoney.jp/>

(注2) : 音楽ダウンロード・メガサイト Mora: <http://mora.jp/>

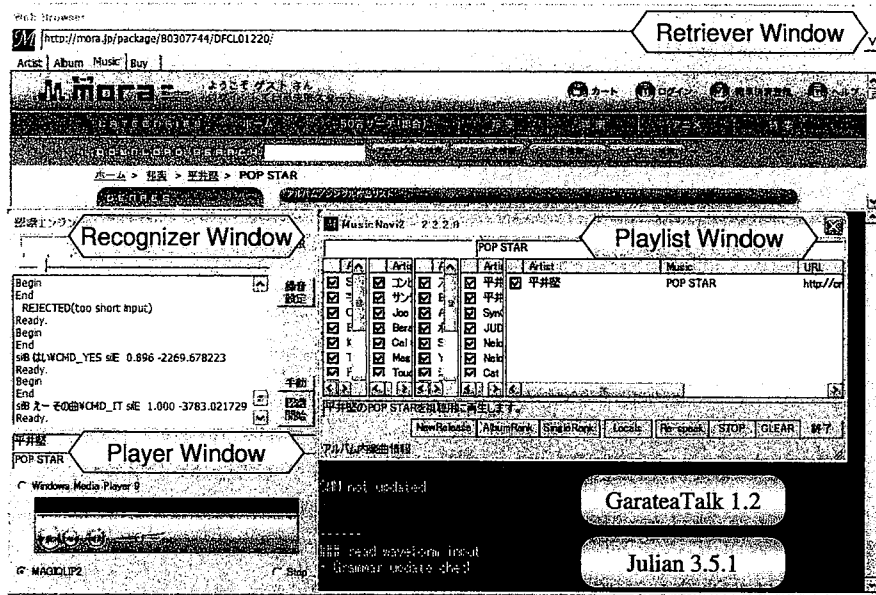


図2 MusicNavi2の動作画面。MusicNavi2は楽曲検索部（Retriever Window）、音声認識部（Recognizer Window）、楽曲リスト管理部（Playlist Window）、楽曲再生部（Player Window）の4つのウィンドウと音声認識器（Julian 3.5.1）、音声合成器（GarateaTalk 1.2）の2つの外部モジュールで構成される。

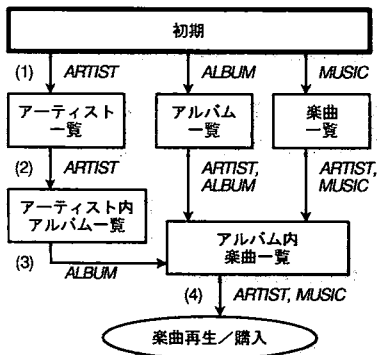


図3 MusicNavi2による検索手順

- U ポップスター
- S アルバム POP STAR を検索します。
- S 1曲見つかりました。平井堅のPOP STAR.
- U ソノキョク
- S 平井堅のPOP STARを試聴用に再生します。
- U コウニュウ（購入手続きの開始）

### 2.1.2 音声認識エンジンの仕様

音声認識エンジンには Julian 3.5.1 を用いた。音響モデルは、CSRC の標準日本語音響モデル [5] 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“ユーザ辞書”と“PC 内楽曲辞書”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。「ユーザ辞書」はサーバ上でアーティスト選択を行うことで、ユーザ毎に自動作成される辞書である。“PC 内楽曲辞書”は、ユーザ PC 内の楽曲を検索する場合を想定して、MusicNavi2 起動時に動的に生成される辞書である。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 [4] の GMM モデル（adult, child, cough, laugh, noise）を使用した。

### 2.1.3 音声合成エンジンの仕様

音声合成エンジンには GalateaTalk 1.2 [6] を基本として、利用の簡易化、動作の安定化、合成速度の向上、合成音量の調整機能付加などを施したライブ러리版を作成し、利用した。合成用の辞書は GalateaTalk 付属の UniDic-1.1.2 を元にアーティスト名、アルバム名、楽曲名として使用されている単語を 3057 語追加した辞書を使用した。これにより、フィールドテストで良く選ばれていたアーティスト上位 100 名に関するアーティスト名、アルバム名、楽曲名の読みは全てカバーされる。

### 2.1.4 音声及び動作ログの送信

音声インタフェースを終了すると自動的に Moscovmule へのデータ送信ソフトウェアが起動する。ユーザが何らかの理由で利用時の音声を Moscovmule に送信することを望まない可能性を考慮し、本ソフトウェアの“送信”ボタンが押された場合のみ、音声インタフェース利用時のログを送信する。送信されるログは、音声ファイル、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（図 4）の 3 種類である。

音声ファイルは、録音された音声（16bit、16kHz、1ch）を

```

22:09:37.911: AlbumList: Focused
22:09:53.033: Julius: Result: silB Best%20Friend\MUSIC silE 0.333 -3817.45898
22:09:53.033: Julius: Keyword: Best Friend,2
22:09:53.033: StateManager: AddRequest: Best Friend,2
22:09:53.063: StateManager: ProcessRequest: Best Friend,2
22:09:53.063: State: action: CSearchResultAlbumState
22:09:53.063: SpeechManager: stopSpeechAll.
22:09:53.494: Browser: Navigate: 2: URL[http://mora.jp/package/80311316/VICL-35255/] POST[]
22:09:53.494: State: next: CSearchResultAlbumState -> CSelectFromAlbumState
22:09:53.494: State: WordSlot: Kiroro,Best Friend,Best Friend

```

図 4 動作ログファイルの記録例。動作ログは MusicNavi2 の内部イベント発生毎に、時刻、イベントの種類、追加情報の順に記録される。このログファイルの示す意味は次の通り。(1 行目) リストを選択した後、(2 行目) 「Best Friend」と認識、(3 行目) 楽曲に関するキーワードとして「Best Friend」を抽出し、(4 行目) 対話状態管理モジュールにキーワードを送信、(5 行目) キーワードの処理を開始する。(6 行目) 対話状態管理モジュールの現在の状態に応じた処理として、(8 行目) WWW ブラウザに対してキーワードを元に URL をリクエスト、(9 行目) 最後に、対話状態を次の状態に移し、(10 行目) 現在のキーワードスロットが更新された。

FLAC [7] により可逆圧縮して記録される。ラベルファイルは、音声始端時間、音声終端時間、GMM による音声判別結果、音声認識結果の 4 種類が各行に記録される。時間は秒単位で出力され、分解能は 10[msec] である。認識結果には Julian の出力がそのまま記録される。動作ログファイルは、ユーザがインタフェース内のボタンやリスト内の項目をクリックした情報、音声認識結果と認識結果から抽出したキーワード、抽出キーワードに基づくインタフェース内の状態遷移情報がユーザ PC の内部時計から得た時刻 (ミリ秒単位) と共に記録される。動作ログファイルを見ることでユーザがインタフェース利用時に行った行動を追跡することができる。

## 2.2 Moscovmule: 実験データ収集サーバ

Moscovmule は Apache2.0, PHP5.1, MySQL4.1 を利用した WWW アプリケーションサーバである。ユーザが Moscovmule にアクセスする際には、ユーザの PC にインストールされている WWW ブラウザを用いる。動作確認は Windows XP がインストールされた PC 上で、Internet Explorer 6.0 及び Firefox 1.5 を使用し、正常に動作することを確認している。ユーザはユーザ ID とパスワードを用いてログインすることで MusicNavi2 システムのダウンロード、音声認識辞書のカスタマイズ、アンケートへの回答、メールフォームによる意見の送信を行うことができる。

### 2.2.1 ユーザ情報の管理

MusicNavi2 のユーザはフィールドテストへの参加や音声認識辞書のカスタマイズを行うためにユーザアカウントを作成する必要がある。ユーザアカウントは、Moscovmule 上の指定の WWW ページにアクセスして実験の同意書を読みフォーム上の必要項目に記入することで作成される。

フィールドテストではユーザに楽曲の無料ダウンロード機能を提供する。少数のユーザが大量のユーザアカウントを作成して無料ダウンロード機能を不当に利用することを防ぐために、ユーザアカウント作成後に電子メールによる登録確認を行い、重複した電子メールアドレスの登録は不可能とした。ユーザアカウントのパスワードを再発行する場合にも、ユーザアカウン

ト情報のメールアドレスにパスワード再発行の手順を説明するメールを送信している。

### 2.2.2 音声認識用辞書のカスタマイズ

辞書として登録するアーティスト名、アルバム名そして曲名は Mora で日本のポップミュージック (J-POP) として分類されていた曲を主に取得した。Moscovmule にはアーティスト 4533 名、アルバム 18643 個、楽曲 135775 曲分の取得データが手動で付与した読み仮名と共に登録されている。

Moscovmule において認識用辞書を作成するためには、(1) 基本辞書の選択、(2) アーティストの選択、(3) 選択内容の確認、(4) 辞書の作成、を行う必要がある。

#### (1) 基本辞書の選択

辞書の作成の際にはすべて 1 から作る「新規作成」以外に、「既存辞書からの作成」によりあらかじめ用意されたアーティストセットを選択した状態から辞書の作成を行うことができる。「既存辞書からの作成」では主に二種類のアーティストセットを用いる。一方は過去にユーザ自身が作成した辞書に基づくアーティストセットであり、もう一方はこれまでに全てのユーザから選ばれたアーティストの上位 20 名、50 名、100 名に相当するアーティストセットである。

#### (2) アーティストの選択

大量のアーティストから選択するための手法として、「50 音順の一覧表からの選択」及び「名前の一部による検索」の二通りの選択方法が用意されている。

#### (3) 選択内容の確認

ユーザは前段階で選ばれたアーティストを確認し問題が無ければ次の段階へ移動する。

#### (4) 辞書の作成

前段階までに選ばれたアーティストを元に辞書を Moscovmule 内に作成する。辞書を作成する際には選択されたアーティスト名だけではなくそのアーティストのアルバム名、曲名をすべて辞書に単語として登録する。

以上の手順で作成された辞書は、Moscovmule 内に保管されており、MusicNavi2 を起動する際に自動的にダウンロードさ

表 2 2ヶ月間に収録されたデータ概要

概略	
登録ユーザー数	78 人
音声認識辞書作成を行ったユーザー数	48 人
データアップロードを行ったユーザー数	32 人
総試聴曲数	801 曲
総購入曲数	260 曲
MusicNavi2 の総起動時間	3540 min.
発話検出区間の総時間	341.5 min.
発話検出区間の総数	11,351 発話
GMM による雑音識別性能	
音声棄却率 (FRR)	9.8 %
雑音受入率 (FAR)	29.8 %
MusicNavi2 に対する発話の分析	
音声発話数	6,335 発話
発話区間検出誤り	30.3 %

れる。

### 2.2.3 アンケート

フィールドテスト中に行ったアンケートは、ユーザの PC 及びネットワーク環境に関するアンケートと使用したマイクの種類や使用方法についてのアンケートである。アンケートは収録データの分類に利用する予定である。

## 3. 提案システムを用いた不特定多数の環境における音声収集実験

2006 年 2 月 1 日から 2006 年 3 月 31 日の間、楽曲の無料ダウンロード機能を備えた本システムをインターネット上で試験的に公開しデータの収集を行った。

収集したデータの概要を表 2 に示す。登録ユーザの 62% が音声認識辞書の作成を行い、登録ユーザの 41% がデータのアップロードまで行っていることが分かる。辞書作成などの初期導入に手間がかかった、ユーザの思うようにシステムを動かすことが出来なかった等がデータのアップロードを行わなかった理由と考えられる。

GMM による音声棄却機能により音声区間を雑音区間として検出した率 (False Rejection Rate; FRR) 及び雑音区間を音声区間として検出した率 (False Acceptance Rate; FAR) はそれぞれ全体の 9.8%、29.8% であった。FAR がやや高めとなっているが、認識結果の自動送信モードにおける“認識開始”ボタンの押下されていない場合の入力を無視すると、FAR は 15.1% となる。また、発声区間が極端に短い場合、入力音声が大きすぎてクリッピングが起きている場合、呼吸による風切り音が含まれている場合などに誤って雑音と検出されている。これらの誤認識は GMM 学習時とシステム利用時の音響環境の不整合が原因と考えられる。

収録された検出区間の分類を図 5 に示す。MusicNavi2 に対して発声したと確認された音声区間は検出された全発話区間の 55.8% (6335 個) であった。検出された区間のうち発声の途中に発話区間が始まったり終わったりしている例が全体の 30.3% だった。

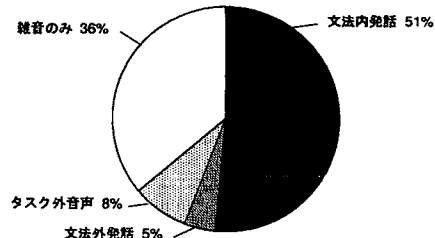


図 5 検出区間の分類

表 3 アンケートの集計結果

Q1. 使用したマイクの種類		Q2. マイクの音声入力スイッチの使用	
ヘッドセットマイク	13	スイッチ有・使用	9
手持ちマイク	4	スイッチ有・未使用	4
PC 内蔵マイク	4	スイッチ無	13
タイピンマイク	3		
その他遠隔・離	2		

マイクに関するアンケートの結果を表 3 に示す。回答者は 26 名である。回答者の半分以上がヘッドセットマイクを使用していることが分かる。

### 3.1 収録音声

収録音声例を図 6 に示す。各ユーザ毎に異なった雑音を含む音声収録されていることがわかる。収録データに含まれていた雑音の種類としてはテレビ音声、呼吸、せき、第三者との発話、合成音声、再生音楽などのマイクロフォンから混入した雑音が含まれていた。マイク端子の抜き差しやマイクに触れた際の音などマイク取り扱い時の不注意による雑音も含まれている。さらに、PC 内部の電氣的ノイズや PC から発生する雑音など環境特有の雑音も収録されていた。

### 3.2 音声認識実験

MusicNavi2 利用時の全発話区間のうち文法に適合していた 4716 発話 (男性 18 名, 女性 5 名) のみを用いて音声認識実験を行った。音響モデルは 3 種類を用いた。一つめのモデルとして、MusicNavi2 で利用しているモデル (GID) を利用した。このモデルは、CSRC の標準日本語音響モデル [5] より、状態数 3000 (基底分布 129 状態)、64 混合、性別非依存、PTM triphone モデルである。二つめのモデルとして、ユーザ登録から得た男女別の情報を元に GID モデルと同等の性別依存モデル (GD) を用いた。三つ目のモデルとして、GID モデルに対して各話者毎に教師無し MLLR 適応したモデル (GID\_MLLR) を用いた。単語辞書は各ユーザが Moscowmule 上で作成し MusicNavi2 で利用した辞書を用いた。文法は MusicNavi2 実行時には動的に切り替えているが、音声認識実験時には全ての文法を常に利用した。

表 4 に各音響モデル利用時の平均単語正解率を、図 7 に各ユーザ毎に算出した単語正解率のヒストグラムを示す。両側 t 検定によると、GID モデルと GID\_MLLR モデルの間には有意水準 1% で有意差があり、GID モデルと GD モデルの間には有意差はない。このことから、音声データ及び認識結果

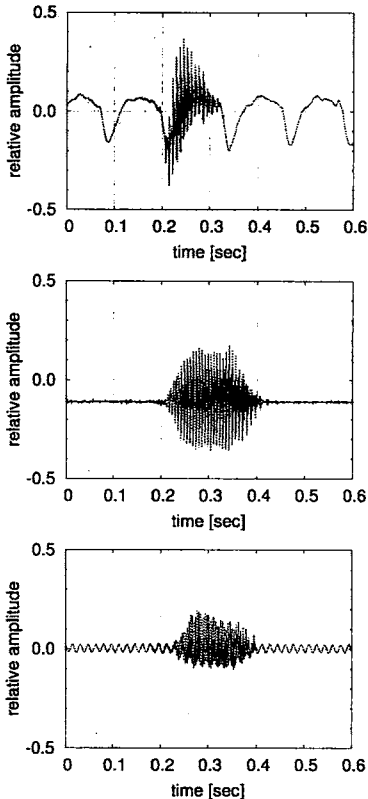


図6 3名による「はい」発音時の音声波形

表4 音響モデル別の単語正解率

音響モデル	単語正解率
GID	66.0 %
GD	64.9 %
GID_MLLR	70.5 %

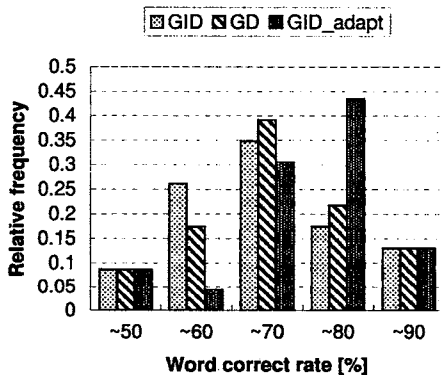


図7 単語正解率のヒストグラム

を Moscovmule に送り、音響モデルの適応を行った後、再度 MusicNavi2 にユーザ毎に適応した音響モデルを利用することで、音声認識システムの単語正解率を増加させることが可能で

あると考えられる。

#### 4. まとめと今後の課題

本論文では、一般の PC ユーザを対象とした被験者実験及び音声収集を行うためのシステムである、インターネットを介した楽曲検索のための音声対話インタフェース MusicNavi2 と実験データ収集サーバ Moscovmule について述べた。

MusicNavi2 による Moscovmule への音声送信機能は、これまでには収録することが難しかったユーザ毎に異なる音響環境で収録された多様な音声の収集を可能とした。また、不特定多数のユーザに対応可能な音声認識インタフェースを備えた楽曲検索システムを構築するためには出来る限り多くのアーティスト名、アルバム名、楽曲名を網羅した大語彙辞書を扱う必要があったが、MusicNavi2 は Moscovmule と連携することでユーザ自身の手で音声認識に用いる辞書をカスタマイズする機能を与え、ユーザ毎に作成された認識辞書の利用を可能とした。

約2ヶ月にわたるフィールドテストにより、データアップロードを行わないユーザの存在、雑音環境の不整合による雑音認識別などの問題点が示唆された。一方で、教師無し MLLR 適応により認識率の向上が可能であることを示した。今後はアップロードされた音声を用いた自動音響モデル適応について実装を行い評価していく必要がある。

なお、本システムを用いたフィールドテストは 2006 年 12 月現在でも以下の URL で運用し、フィールドテスト参加者を常時募集している。

<http://moscovmule.sp.m.is.nagoya-u.ac.jp/>

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。

#### 文 献

- [1] J. Glass, J. Polifroni, S. Senoff, and V. Zuc, "Data collection and performance evaluation of spoken dialogue systems: The MIT experience," Proc. of ICSLP-2000, vol.4, pp.1-4, Oct. 2000.
- [2] V. Zuc, S. Senoff, J.R. Glass, J. Polifroni, C. Pao, T.J. Hazen, and L. Hetherington, "JUPITER: A telephone-based conversational interface for weather information," IEEE Trans. Speech Audio Processing, vol.8, pp.85-96, 2000.
- [3] 西村竜一, 西原洋平, 鶴身玲典, 李兎伸, 猿渡洋, 鹿野清宏, "実環境研究プラットフォームとしての音声情報案内システムの運用," 信学論, vol.J87-D-II, no.3, pp.789-798, 2004.
- [4] 中村敬介, 西村竜一, 李兎伸, 猿渡洋, 鹿野清宏, "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.
- [5] 河原達也, 李兎伸, 小林哲則, 武田一哉, 峯松信明, 伊藤克巨, 山本幹雄, 山田篤, 宇津呂武仁, 鹿野清宏, "日本語ディクテーション基本ソフトウェア (98 年度版)," 音響誌, vol.56, no.4, pp.255-259, 2000.
- [6] S. Kawamoto, H. Shimodaira, T. Nitta, T. Nishimoto, S. Nakamura, K. Itou, S. Morishima, T. Yotsukura, A. Kai, A. Lee, Y. Yamashita, T. Kobayashi, K. Tokuda, K. Hirose, N. Minematsu, A. Yamada, Y. Den, T. Utsuro, and S. Sagayama, "Open-source software for developing anthropomorphic spoken dialog agent," Proc. of PRICAI-02, pp.64-69, Aug. 2002.
- [7] J. Coalsion, "Free Lossless Audio Codec (FLAC)," <http://flac.sourceforge.net/>.