

携帯端末用の音声情報案内システム開発に向けた ネットワークサービスの検討

原 直^{1,a)} 川波 弘道^{1,b)} 猿渡 洋^{1,c)} 鹿野 清宏^{1,d)}

概要: 本論文では携帯端末用の音声情報案内システム開発に向けたネットワークサービスソフトウェアの開発について検討する。このソフトウェアは2002年11月よりコミュニティセンターに設置されたたけまろくんシステムを基礎としている。本システムは、音声認識、対話管理、テキスト音声合成を統合したシステムである。WWWサイト開発者にも使いやすいような、Software-as-a-Service (SaaS) を目指すとともに、熟練した開発者や研究者にも使いやすいパッケージとしての展開を検討する。

キーワード: 音声認識, 音声対話, インターネットサービスソフトウェア

Design of a network service for developing a speech-oriented guidance system used on mobile computers

SUNAO HARA^{1,a)} HIROMICHI KAWANAMI^{1,b)} HIROSHI SARUWATARI^{1,c)} KIYOHIRO SHIKANO^{1,d)}

Abstract: In this paper we propose a novel speech service software for speech-oriented guidance systems. This software has been developed based on *Takemaru-kun* system, that is implemented at a community center since Nov. 2002. It is consisted of several modules, such as, Automatic Speech Recognition, Dialog Management, Text-to-Speech, Internet browser, and Computer Graphic Agent. This software and toolkit is plan to be freely distributed. It will be used as the speech service software as Software-as-a-Service (SaaS) for WWW site developers, and also used for an upgrade system of our system for advanced developers or researchers.

Keywords: Speech recognition, Spoken dialogue, Internet service software

1. はじめに

音声対話システムの開発には大量の音声対話データが必要となるが、そのデータはできる限り利用時と同等の対話課題によって収集されていることが望ましい [1]。システムの開発とデータ収集とは同時に行われるべきだが、そのためには少量のデータからシステムを構築し、データが収集されるにつれて容易に更新可能なシステムである必要がある。サーバ・クライアント型のシステム構成はその1つの解であり、最も扱いが容易なシステム構成としてはWorld-Wide-Web (WWW) アプリケーションとしての構成が挙げられる。

音声認識システムや音声対話技術を用いたWWWアプリケーションに着目した研究は多く存在している。WebGALAXY [2] は音声入力に電話回線を用いてはいるが、Graphical User Interface (GUI) としてWWWインタフェースを利用したシステムを1997年に提案している。Nishimuraら [3] は1行のJavaScriptを追加するだけで、既存のWWWページに音声認識機能を追加することができるw3voice^{*1}を開発している。GruensteinらのWAMI toolkit [4] も同様にJavaScriptによる音声インタフェースを追加することが可能であり、このツールキットを用いた多数のWWWアプリケーションに関する報告がなされている(例えば、文献 [5], [6] など)。また、AT&Tによる音声マッシュアップフレームワーク (speech mashup framework) [7] や AT&T WATSON の音声 API^{*2}などの

¹ 奈良先端科学技術大学院大学
8916-5 Takayama-cho, Ikoma, Nara 630-0101, Japan

a) hara@is.naist.jp

b) kawanami@is.naist.jp

c) sawatari@is.naist.jp

d) shikano@is.naist.jp

^{*1} w3voice: <http://w3voice.jp/>

^{*2} AT&T WATSON(SM) Voice Recognition Technology & Speech API: <http://www.research.att.com/projects/WATSON/library/multimedia/watson>

取り組みもおこなわれている。

携帯端末での利用に着目した研究として、Personal Digital Assistant (PDA) が利用されていた頃には、PocketSphinx [8] や Pocket Julius [9] などが開発されてきた。これらは携帯端末内で認識処理も行っていたが、ネットワーク回線の高速化が進みスマートホンの普及が広まってきた現在においては、サーバ上での音声認識がしばしば利用されている。特に Voice Search タスクでの利用が増えてきており、Goog-411[10]、Google Search by Voice[11]、[12] や Microsoft Bing [13] などのサービスやアプリケーションも公開されている。また、Apple の「Siri」、NTT Docomo の「しゃべってコンシェル」、Yahoo の「音声アシスト for Android」など、音声対話サービスとしての実用化が進められている。

日本の音声対話研究においても、双方向音声案内デジタルサイネージのためのツールキットである MMDAgent [14] が公開されており、特定の地域に設置されるデジタルサイネージとしての特性を生かした、単一エージェントによるマルチドメイン音声対話を実現している。また、複数の言語理解モデルを用いるためのツールキット MLUTK [15] なども開発されており、こちらも単一エージェントでの利用を志向したツールキットである。

本研究では研究・開発用途において自由に利用可能な特定地域の情報案内サービスを可能とする音声対話サーバとそれらサーバ間の連携を可能とするツールキットの開発を行う。現在、爆発的に普及が進んでいるスマートホンなどの携帯端末では、容易に位置情報を取得可能であり、現在位置に応じた情報案内サービスが求められており、本システムによって様々な拠点においてサービスが提供されることを想定している。本報告では携帯端末用のクライアント・サーバ型音声対話ツールキットの概要を報告するとともに、サーバ間連携可能な音声対話システムを応用したサービスについて検討を行う。

2. 「たけまるくん」音声対話開発キット

たけまるくんは奈良県生駒市のコミュニティセンターに2002年11月より設置され、約10年間にわたり一般市民に利用され続けている音声対話による情報案内システムである [16]、[17]。本システムを活用して、2006年3月より近鉄学研北生駒駅にキタちゃん及びキタロボを設置し改良を続けてきた [18]、[19]。さらに、2007年7月から10月にかけて平城遷都1300年祭での案内に特化したたけまるくんを設置してきたように [20]、[21]、たけまるくんシステムを用いる事で短期間のうちに、異なる情報案内システムを容易に構築することができる。

この「たけまるくん」システムは「キタちゃん・キタロボ」の開発に際して、ツインタアレクによって Ruby 言語を用いて大幅に改修され [22]、「Tankred: A Dialogue System Toolkit written in Ruby」と呼ばれており、現在もメンテナンスされている*3。このツールキットは、音声認識 (ASR)、対話管理 (DM)、テキスト音声合成 (TTS)、インターネットブラウザ、コンピュータグラフィックスエー

ジェント (CGA) という5つのモジュールから構成されている。ASR モジュールにはオープンソースソフトウェアの音声認識システム Julius [23] が用いられている。DM モジュールには対話ドメインの識別機能と返答生成機能が含まれる。TTS モジュールには商用の音声合成ソフトウェアが利用されている。インターネットブラウザはオープンソースソフトウェアの Galeon *4 が用いられている。CGA モジュールはブラウザを利用した Flash アニメーションの表示を行う。このツールキットは Ruby1.8 がインストールされた Debian GNU/Linux 4.0 (etch) *5 上で開発・運用が行われているが、他の Linux ディストリビューションや Mac OS X などでも動作することが確認されている。

2.1 用例ベース応答生成

このツールキットでは、Q&A データベースに基づく用例ベースの応答生成が用いられている。この用例ベース応答生成は、ユーザからの質問例とシステム応答のペア (質問応答対) を質問応答データベース (QADB) として保持し、INMS (Intersection Normalization by Maximum Size) 類似度を用いた最近傍法により入力に最も類似した質問例を一つ選択する手法である [24]、[25]。以下に INMS 類似度の算出と用例選択の手順を示す。

用例データベースには M 個の質問応答対が登録されており、その用例の集合を $Q = \{Q_1, Q_2, \dots, Q_M\}$ とし、音声認識による N -best 仮説候補群を $\mathcal{H} = \{H_1, H_2, \dots, H_N\}$ とする。 $\mathcal{W} = \{w_1, w_2, \dots, w_K\}$ は出現しうる全単語の集合とする、すなわち $\forall H_n \subseteq \mathcal{W}$ かつ $\forall Q_m \subseteq \mathcal{W}$ である。入力仮説群 \mathcal{H} とある1つの用例 Q_m の類似度 $s(\mathcal{H}, Q_m)$ を以下の式で表す。

$$s(\mathcal{H}, Q_m) = \frac{\sum_{w_k \in \mathcal{W}} \min \left(\frac{1}{N} \sum_n c_k(H_n), c_k(Q_m) \right)}{\max \left(\frac{1}{N} \sum_n \ell(H_n), \ell(Q_m) \right)} \quad (1)$$

ここで、 $c_k(\cdot)$ は文中の単語 w_k の出現回数、 $\ell(\cdot)$ は文に含まれる単語数である。分子は入力と用例の一致単語数の合計を意味しており、分母は入力単語数と用例単語数のいずれかが大きい値を表しており、完全に一致する場合に $s(\mathcal{H}, Q_m) = 1.0$ となる。この INMS 類似度に基づいて、最適な用例 \hat{Q}_m を選択する。

$$\hat{Q}_m = \arg \max_{Q_m \in Q} s(\mathcal{H}, Q_m) \quad (2)$$

このように選ばれた用例 \hat{Q}_m に対応する応答がシステムの応答となる。

2.2 質問応答データベースの構築

QADB の構築の際には、書き起こし文に対して応答文を手で与える事で様々なタスクに対応可能な質問応答システムを構築することができる。一般的には書き起こし文の数が多いほど、システムが対応できる文章が増えるため応答正解率は向上することが期待される。書き起こし文の元

*3 「Tankred」とは古ドイツ語で「the thinking adviser」の意味。

*4 <http://galeon.sourceforge.net/>

*5 <http://www.debian.org/>

となる利用データが少ない場合には、書き起こし文と認識結果を混合した質問例を用いたり [25], N-gram に基づいて対話例を生成して質問例とする手法 [26], [27] などが有効であると報告されている。

また、応答文生成に関する研究として、ユーザに質問応答対を追加させることで応答正解率を向上させる試み [28] や、応答文生成を質問文から応答文への翻訳と見なして統計的機械翻訳を応用した研究 [29], Twitter から得られた情報を QADB 拡張に用いた研究 [30] なども行われている。

ツインタレク [31] によって、「たけまるくん」システムの頑健性とポータビリティの高さは示されているが、一方で学習データ数の増加に対する性能向上の飽和についても指摘している。また、応答正解率による性能評価シミュレーションを通して、1) 質問応答データベース, 2) 音響モデル, 3) 言語モデル, の順でシステム性能に強く関わっていることを指摘している。

クラウドソーシングを活用したアノテーションによる音響・言語モデルの改善を行った緒方らの PodCastle に関する実験 [32], [33] においても、音響モデルの改善による音声認識率向上への寄与は、言語モデルの改善よりも大きいことが報告されている。一方で、言語モデルの改善に伴う未知語解消がユーザの満足度に影響しており、アノテーション付与のモチベーション維持に効果的であることも示唆している。

ユーザにとって最終的に見えているシステムの応答に着目すれば、「たけまるくん」では情報案内の文章であり、PodCastle では音声認識結果の文章である。従って、システム応答の最も表層部に相当するモデルの改善はユーザの満足度向上のためには極めて重要であるといえる。

2.3 タスク判別

DM モジュールでは、a) 連続音声認識 (LVCSR) に基づく Q&A, b) 記述文法音声認識に基づく Q&A, c) 二段階の Q&A, という 3 種類の対話機能を備えている。システム応答の際には、ASR モジュールの認識スコア、すなわち音響モデルの対数尤度と言語モデルの対数尤度の重み付き和、に従って最善とされる対話機能が選択される。たけまるくん、キタちゃんはこの対話機能として、a) として前節で説明した QADB を用いた質問応答を行い、b) として簡単な数値計算 (例: 『 3×5 は?』に 『15 です。』と答える。) を行うことができる。また、c) として、WWW からの検索機能を実装しており、具体的にはまず文法規則により検索語 (例: 『検索開始』) を認識すると、次に認識した文章を WWW 検索エンジンによって検索する機能である。これはやや冗長に見えるかもしれないが、ユーザからの入力は何らかの対話タスクに従った発声なのか WWW 検索のための発声なのかを明示的に分けられるため、誤動作を防ぐことができると考えられる。

3. 音声情報案内サービスツールキット

本報告では、前節の「たけまるくん」システムを WWW サービス向けに拡張し、Software-as-a-Services として運用可能なツールキットを構築した。ASR, DM, TTS はコア

モジュールとしてサーバ側に実装して、それ以外のテキスト表示, CG エージェント表示, 音声入力, 音声出力などインタフェースに関するモジュールはクライアント側に実装するものとした。

3.1 Tankred on Rails: 音声情報案内サーバ

サーバシステムは WWW サービス開発にしばしば用いられる Ruby on Rails *6 で実装した。本研究ではこれを *Tankred on Rails (ToR)* システムと呼ぶ。システムの流れ図を図 1 に示す。クライアントはユーザから入力された質問リクエスト音声のデータファイル (PCM, 16bit, 16kHz) を *ToR* システムの特定の URL に HTTP リクエストとして送信する。*ToR* システムでは、Julius [23] による音声認識を行い、その N-best 認識結果を用いた用例ベースの応答文生成 (2.1 節) を行う。応答文に基づいて TTS モジュールで応答音声ファイルが生成される。クライアントには認識結果と応答文章が応答音声ファイルの URL と共に HTTP レスポンスとして送られる。なお、音声ファイル, 認識結果, 応答結果などは端末 ID と利用時間を付与したログとして、サーバに蓄積される *7。

TTS モジュールには、オープンソースソフトウェアの Open JTalk *8 とそのバックエンドシステムである hts_engine [34] *9 を利用している。従って、ユーザもしくは開発者が音声合成用の音響モデルを所有しているならば、エージェントの声を変更することも可能となる。

また、このソフトウェアでのインターネット通信は HTTP プロトコルのみを用いている。従って、Apache httpd の mod_proxy *10 などのリバースプロキシソフトウェアを利用することも可能である。

なお、音声対話機能としては前節の「たけまるくん」と同様に 3 種類の機能が実装されているが、対話機能の追加・変更も容易に可能である。

3.2 iTakemaru: 携帯端末用の音声情報案内システム

図 1 における HTTP Request の生成と HTTP Response の解釈が可能で音声情報案内クライアントとして iTakemaru を開発した (図 2)。これは Objective-C 言語で開発されており、iPhone4S 上で動作する。*ToR* システムのクライアントであり、「たけまるくん」同等の対話が可能である。

本実装では、Press-to-Talk, すなわちボタンを押している間の音声を収録する手法を採用した。Press-to-Talk による音声入力は、始端と終端を確実に与える事できるため、入力音声を 1 つのファイルとして容易に扱うことができることや自動音声区間検出で不可避の検出誤り問題が起こらないというメリットがある。しかし、画面を押し続けるという動作がユーザにとって必ずしも使いやすいとは言えず、現在の携帯端末の音声認識機能によく用いられてい

*6 <http://rubyonrails.org/>

*7 後の章で説明する iTakemaru では端末 ID として “UDID (Unique Device Identifier)” ではなく “UUID (Universally Unique Identifier)” を利用している。

*8 <http://open-jtalk.sourceforge.net/>

*9 <http://hts-engine.sourceforge.net/>

*10 <http://httpd.apache.org/>

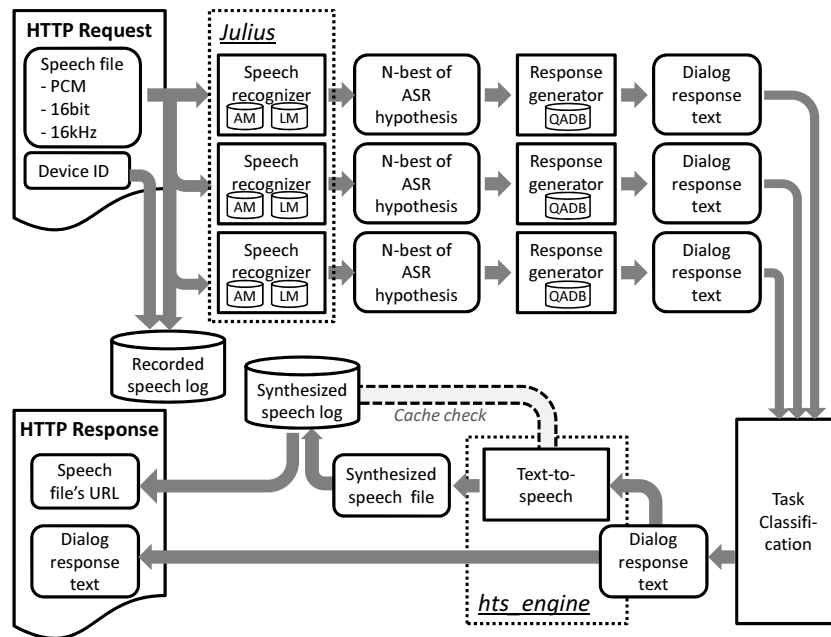


図 1 サーバシステムの流れ図
 Fig. 1 A flowchart of server system

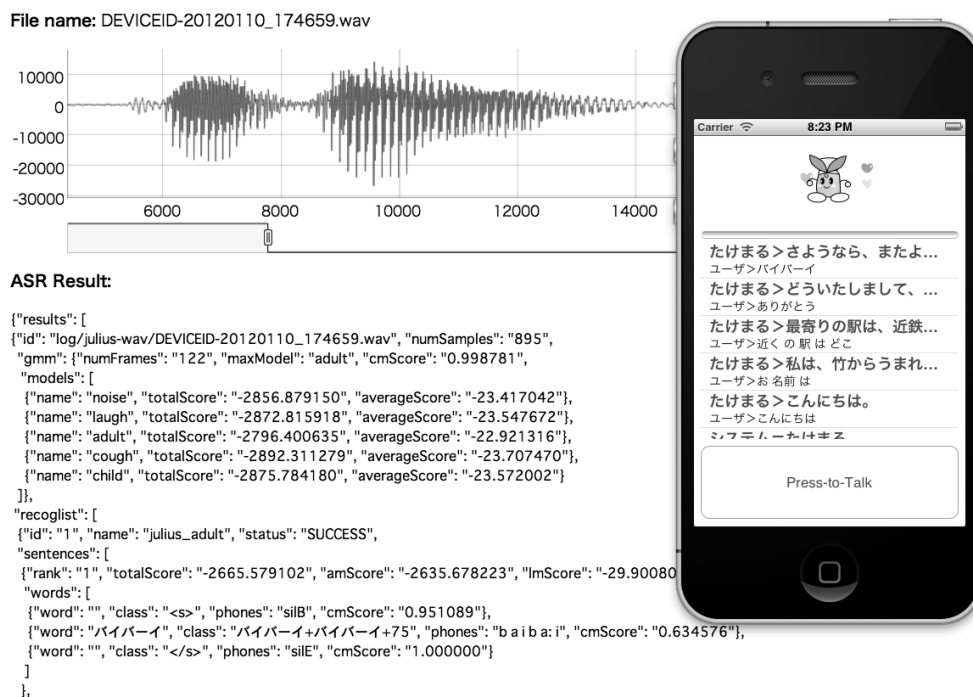


図 2 iPhone4S で動作する音声情報案内システム iTakemaru (右図) と収録データの管理画面 (左図)

Fig. 2 An speech-oriented guidance system for iPhone 4S (iTakemaru; right) and administrator's view of recorded data (left).

る，始端のみを与える Push-to-Talk が，あるいは自動的に非音声のみを棄却する手法 [35] などを実装し，その性能を評価することが必要になるが，これは今後の課題とする．

4. マルチエージェント対応の音声情報案内サービス

前節の音声情報案内サービスシステムをさらに拡張し，

マルチエージェント化を行う．システムの流れ図を図 3 に示す．システム全体は 1 つのメインエージェントと複数のサブエージェントから構成される．クライアントはメインエージェントと対話を行っており，メインエージェントはサブエージェントとのサーバ・サーバ間通信も行う．メインエージェントの応答データにはどのサブエージェントに對話データが委譲されたかを記載しておくことで，クライアントは適宜サブエージェントからも情報を得ることがで

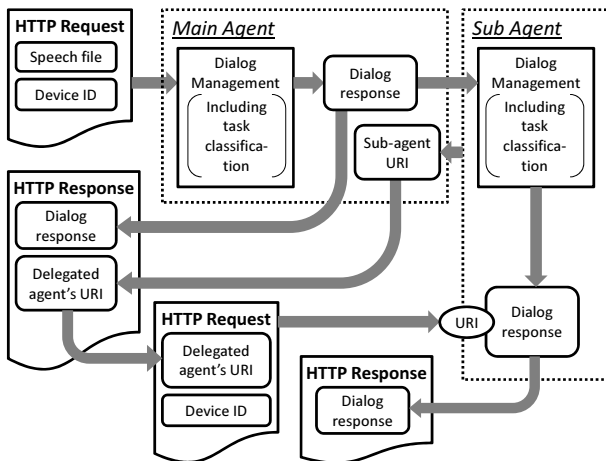


図 3 マルチエージェント対応のサーバ・サーバ間通信の流れ図
 Fig. 3 A flowchart of Server-to-Server connection handling multiple agents



図 4 iTakemaru-Multi: マルチエージェント対応のクライアントソフトウェア。メインエージェントの「たけまるくん」、サブエージェントの「たけまるくん 遷都 1300 年祭版」と「キタちゃん」が利用されている (左図)。サブエージェントはさらに詳細な情報を表示することも可能である (右図)。
 Fig. 4 iTakemaru-Multi: A client software handling multiple agents.

きる。

この仕組みの利点としてはクライアント・サーバ間の通信よりも、サーバ・サーバ間通信は高速かつ安定していると思えることができるため、演算速度、メモリ容量、回線速度に制限が多い携帯端末向けのシステムに有効であると考えられる。また、クラウド・コンピューティングを活用すれば、クライアントはメインエージェントの存在するサーバとだけ通信しているにも関わらず、大量のサブエージェントと対話しているように見せることも可能になるだろう。

4.1 iTakemaru-Multi: マルチエージェント対応の音声情報案内クライアント

実際のシステム動作の具体例を図 4 に示す。図 4 左図では、メインエージェントとしてたけまるくんを用いており、ユーザは基本的にたけまるくんと対話を行っている。たけまるくんはサブエージェントとして、近鉄の駅に詳しい「キタちゃん [19]」と遷都祭のイベント会場に詳しい「たけまるくん～遷都 1300 年祭版～ [21]」と通信を行うことができる。この対話例において、たけまるくんは『近くの駅 = 学研北生駒駅』という情報は知っているが、その駅についての詳細を答える知識は持っていない。しかし、「キタちゃん」はユーザと「たけまるくん」との対話内容を知っており、たけまるくんの提案した『学研北生駒駅』は自分の方が詳しいと判断したため、ユーザに吹き出しによる割り込み通知を行っている。図 4 右図はユーザがその吹き出しをタップすることで切り替わった画面であり、ユーザは『学研北生駒駅』についてのより詳細な情報を、「キタちゃん」から得ることができる。

5. まとめと今後の課題

本研究では音声情報案内サービスソフトウェアの開発を行った。そして、本ツールキットを用いて様々な地域に専用のエージェントが存在し、それらと相互通信しながら対話するための仕組みを提案した。この音声対話機能はインターネット経由で利用可能であると同時に、開発者向けのツールキットを用いた実装も可能であり、今後ドキュメントの整備とツールキットの配布を行っていく。

様々な開発者や開発機関によって多数のエージェントが作成された場合に、それぞれの機関に音声データなどが収集されてしまう。これらのデータは音響モデルや言語モデルの改善に利用可能であるが、収集データを共有する仕組みを備えることで、より高度なモデル改善が可能となるため、今後の課題として検討する必要があるだろう。

謝辞 たけまるくんを初めとする音声対話システム開発に携わってきた奈良先端科学技術大学院大学 音情報処理学研究室の歴代の卒業生皆様に感謝すると共に、たけまるくんの基礎を作り上げた西村竜一博士 (和歌山大学・助教) と Ruby によるシステム化を行った Dr. Tobias Cincarek には心より感謝申し上げます。なお、本研究の一部は、戦略的創造研究推進事業「共生社会に向けた人間調和型情報技術の構築」(JST/CREST) の援助を受けて行われた。

参考文献

- [1] Möller, S. and Skowronek, J.: Quantifying the impact of system characteristics on perceived quality dimensions of a spoken dialogue service, *Proc. of EUROSPEECH 2003*, pp. 1953-1956 (2003).
- [2] Lau, R. and et al.: WebGALAXY - Integrating Spoken Language and Hypertext Navigation, *Proc. of Eurospeech-97*, pp. 883-886 (1997).
- [3] Nisimura, R., Miyake, J., Kawahara, H. and Irino, T.: Development of Speech Input Method for Interactive VoiceWeb Systems, *Proc. of Human-Computer Interaction, Part II*, Springer, pp. 710-719 (2009).
- [4] Gruenstein, A., McGraw, I. and Badr, I.: The WAMI

- Toolkit for Developing, Deploying, and Evaluating Web-Accessible Multimodal Interfaces, *Proc. of ICMI 2008*, pp. 141–148 (2008).
- [5] McGraw, I., Gruenstein, A. and Sutherland, A.: A Self-Labeling Speech Corpus: Collecting Spoken Words with an Online Educational Game, *Proc. of Interspeech 2009*, pp. 3031–3034 (2009).
- [6] McGraw, I., Lee, C., Hetherington, L., Seneff, S. and Glass, J.: Collecting voices from the cloud, *Proc. of LREC 2010*, pp. 1576–1583 (2010).
- [7] Fabbriozzi, G. D., Okken, T. and Wilpon, J. G.: A Speech Mashup Framework for Multimodal Mobile Services, *Proceedings of ICMI 2009* (2009).
- [8] Huggins-Daines, D., Kumar, M., Chan, A., Black, A. W., Ravishankar, M. and Rudnick, A. I.: PocketSphinx: A Free, Real-Time Continuous Speech Recognition System for Hand-Held Devices, *Proc. of ICASSP 2006*, pp. I-185–I-188 (2006).
- [9] 原 直, 河口信夫, 武田一哉, 板倉文忠: 汎用音声認識エンジン Julius/Julian の PDA への移植と性能評価, 情報処理学会研究報告, SLP-45-22, Vol. 2003, No. 14, pp. 131–136 (2003).
- [10] Bacchiani, M., Beaufays, F., Schalkwyk, J., Schuster, M. and Strope, B.: Deploying GOOG-411: Early Lessons in Data, Measurement, and Testing, *Proc. of ICASSP 2008*, pp. 5260–5263 (online), DOI: 10.1109/ICASSP.2008.4518846 (2008).
- [11] Schalkwyk, J., Beeferman, D., Beaufays, F., Byrne, B., Chelba, C., Cohen, M., Kamvar, M. and Strope, B.: Google search by voice: A case study, *Advances in Speech Recognition: Mobile Environments, Call Centers and Clinics* (Neustein, A., ed.), Springer, chapter 4, pp. 61–90 (2010).
- [12] Schuster, M. and Nakajima, K.: Japanese and Korean voice search, *Proc. of ICASSP 2012*, pp. 5149–5152 (2012).
- [13] Levit, M., Chang, S., Buntschuh, B. and Kibire, N.: End-to-end speech recognition accuracy metric for voice-search tasks, *Proc. of ICASSP 2012*, pp. 5141–5144 (2012).
- [14] 李 晃伸, 大浦圭一郎, 徳田恵一: 魅力ある音声インタラクションシステムを構築するためのオープンソースツールキット MMDAgent, 電子情報通信学会技術研究報告, NC2011-51, SP2011-96, Vol. 2011, pp. 1–6 (2011).
- [15] 竹内誉羽, 中野幹生, 森祥二郎, 駒谷和範: 音声対話システムにおいて複数の言語理解モデルの利用を容易にするツールキット: MLUTK, 情報処理学会研究報告, Vol. 2011-SLP-89, No. 28, pp. 1–6 (2011).
- [16] 西村竜一, 西原洋平, 鶴身玲典, 李 晃伸, 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, 電子情報通信学会論文誌, Vol. J87-D-II, No. 3, pp. 789–798 (2004).
- [17] Nishimura, R., Lee, A., Saruwatari, H. and Shikano, K.: Public Speech-Oriented Guidance System with Adult and Child Discrimination Capability, *Proc. of ICASSP 2004*, pp. I-433–I-436 (2004).
- [18] 川波弘道, 木田 学, 早川直樹, ツインツアレクトピアス, 北村任宏, 加藤智之, 鹿野清宏: 駅構内音声案内システム「キタちゃん」「キタちゃんロボット」の開発, 電子情報通信学会技術研究報告, SP2006-14, pp. 19–24 (2006).
- [19] Kawanami, H., Takeuchi, S., Torres, R., Saruwatari, H. and Shikano, K.: Development and operation of speech-oriented information guidance systems, *Kita-chan and Kita-robot*, *Proc. of APSIPA-ASC 2011* (2011).
- [20] 平井良佑, 久保慶伍, 木佐木雄介, 川波弘道, 猿渡 洋, 鹿野清宏: 遷都 1300 年祭会場における音声情報案内システムの運用と発話データの分析, 日本音響学会講演論文集, 3-P-24 (2011).
- [21] 川波弘道, 久保慶伍, 木佐木雄介, トーレスラファエル, 鹿野清宏: 展示会場での運用のための音声情報案内システム「たけまるくん」のデータベース拡張, 日本音響学会講演論文集, 3-10-8 (2011).
- [22] Cincarek, T.: Selective Training for Cost-Effective Development of Real-Environment Speech Recognition Applications, PhD Thesis, Nara Institute of Science and Technology (2008).
- [23] Lee, A. and Kawahara, T.: Recent Development of Open-Source Speech Recognition Engine Julius, *Proc. of APSIPA-ASC 2009*, pp. 131–137 (2009).
- [24] Takeuchi, S., Cincarek, T., Kawanami, H., Saruwatari, H. and Shikano, K.: Question and Answer Database Optimization Using Speech Recognition Results, *Proc. of Interspeech 2008*, pp. 451–454 (2008).
- [25] 竹内翔太, 川波弘道, 猿渡 洋, 鹿野清宏: 書き起こし文と認識結果を混合した質問例を用いる用例ベース応答生成の検討, 日本音響学会講演論文集, 3-6-8, pp. 107–108 (2010).
- [26] 吉見孔孝, 南角吉彦, 李 晃伸, 徳田恵一: 音声対話システムのための N-gram に基づくキーワードを制約とする文生成手法の検討, 音響学会講演論文集, 1-R-29, pp. 209–210 (2009).
- [27] 平野隆司, 南角吉彦, 李 晃伸, 徳田恵一: 双方向探索に基づく N-gram を用いたキーワードからの文生成, 日本音響学会講演論文集, 2-P-40, pp. 211–212 (2011).
- [28] 福田敏則, 吉見孔孝, 南角吉彦, 李 晃伸, 徳田恵一: ユーザ生成型音声対話コンテンツに基づく音声情報案内システムの構築, 音響学会講演論文集, 1-R-30, pp. 211–212 (2009).
- [29] Nishimura, K., Kawanami, H., Saruwatari, H. and Shikano, K.: Investigation of Statistical Machine Translation Applied to Answer Generation for a Speech-Oriented Guidance System, *Proc. of APSIPA-ASC 2011* (2011).
- [30] 別所史浩, 原田達也, 國吉康夫: リアルタイムクラウドソーシングと Twitter 大規模コーパスを利用した対話システム, 情報処理学会研究報告, Vol. 2012-SLP-91, No. 13, pp. 1–6 (2012).
- [31] ツインツアレクトピアス, 川波弘道, 木田 学, 猿渡 洋, 鹿野清宏, 西村竜一, 李 晃伸: 「たけまるくん」実環境音声案内システムのデータベース整備と「キタちゃん」へのポータビリティの検討, 情報処理学会研究報告, 2006-SLP-64, Vol. 2006, pp. 173–178 (2006).
- [32] 緒方 淳, 後藤真孝, 江渡浩一郎: PodCastle: ポッドキャストをテキストで検索, 閲覧, 編集できるソーシャルノーションシステム, WISS2006 論文集 (2006).
- [33] 緒方 淳, 後藤真孝: PodCastle: ポッドキャスト音声認識のための集合知を活用した言語モデル学習, 情報処理学会研究報告, Vol. 2010-SLP-80, No. 10, pp. 1–6 (2010).
- [34] Zen, H., Nose, T., Yamagishi, J., Sako, S., Masuko, T., Black, A. W. and Tokuda, K.: The HMM-based speech synthesis system (HTS) version 2.0, *Proc. of ISCA SSW6*, Bonn, Germany, pp. 294–299 (2007).
- [35] 真嶋温佳, トーレスラファエル, 川波弘道, 原 直, 松井知子, 猿渡 洋, 鹿野清宏: 音声情報案内システムにおける Bag-of-Words を特徴量とした無効入力への棄却, 情報処理学会研究報告, Vol. 2012-SLP-92, No. 7, pp. 1–6 (2012).