

パネル討論：アプリケーション開発における音声認識

西村 竜一 和歌山大学
中野 鐵兵 早稲田大学
栗原 一貴 産業技術総合研究所
中臺 一博 (株) ホンダ・リサーチ・インスティテュート・ジャパン / 東京工業大学
吉野 孝 和歌山大学

あらまし 本パネルディスカッションでは、音声認識アプリケーションの開発事例の紹介、アプリ開発に向き合うための提言を関連研究者の方々をお願いした。音声認識アプリケーションの開発を促すことを目的とする。

Panel Discussion

Application Developments of Speech Recognition

Ryuichi Nisimura Wakayama University
Teppei Nakano Waseda University
Kazutaka Kurihara National Institute of Advanced Industrial Science and Technology
Kazuhiro Nakadai HONDA Research Institutes Japan / Tokyo Institute of Technology
Takashi Yoshino Wakayama University

Abstract To induce developments of ASR applications, this panel discussion introduces actual case studies. We also indicate some problems of ASR application developments.

■ イントロダクション

西村 竜一 (和歌山大)

「音声認識の人ってなんでみんな臆病なの？」

ある研究者に言われたことばである。音声認識屋が音声認識を使っていない(使おうとしない)という指摘であった。認識精度を心配するあまり、音声認識の応用に弱腰であることに不満があるらしい。そのときは、その指摘は本当かもしれないと感じた。

今の音声認識は万能ではない。だからといって「使えない技術」ではないと信じている。後藤、緒方らによる「音声認識研究2.0」[1, 2]。そのコンセプトには、わかり易く音声認識のできることを、できないことを世間に示す試みが含まれている。認識率では伝わらない性能の表現に成功すれば、応用・開発に対する機運が高まり、新たなステップへヒントが得られるかもしれない。

「音声認識のアプリケーションを作ってみようよ。」

特に、若手研究者・開発者に呼びかけたい。本パネルディスカッションの企画意図は、音声認識アプリケーションの開発促進である。パネリストのみなさんには、実際の開発事例や、アプリ開発に向き合うために必要な提言をしていただくようお願いした。ぜひ議論に加わって、成

功に向けたヒントを僕と一緒に見つけて欲しい。

次に、パネリストの方々を紹介する。最初に登壇をお願いするのは早稲田大学中野氏である。音声認識を使ったインタフェースの設計指針について述べてもらう。産総研の栗原氏には、音声認識を使ったアプリケーションの例として「音声ペン」「プレゼン先生」などの紹介をお願いした[3, 4, 5]。ここでの論点は「認識率が100%でなくても使っていけるアプリケーション」の可能性である。西村が考える音声認識の有望な応用の一つが、ロボット対話インタフェースである。そこで、ロボット聴覚の研究でご活躍のHRI-JP/東工大の中臺氏に、ロボット聴覚用オープンソースソフトHARKの紹介をお願いした。システム統合技術のノウハウや大切さをお聞きできると期待している。最後に、コミュニケーション技術の研究分野から、和歌山大学吉野氏に研究事例の紹介と音声認識に足りないものを教えて頂く。

以下に、中野氏、中臺氏、吉野氏にご寄稿いただいた原稿を掲載する。本稿の最後では、本パネルディスカッションの後半で予定する議題について述べる。当日は、フロアからのご意見も含めた活発な意見の交換ができることを期待している。

■ アプリケーション開発における音声インタフェース設計指針

中野 鐵兵（早稲田大）

2.1 はじめに

音声認識技術を用いたアプリケーション開発において、専門知識を持たない研究・開発者の参加の敷居を下げるための枠組みが求められている。今日では、音声認識技術に関する研究・開発が広く進んでいるものの、オープンソース開発者や、フリーウェア・シェアウェア開発者と言った一般の開発者が、自身のソフトウェアに音声インタフェースを採用するケースは稀である。もちろん、CTIシステムのように、音声データをはじめから対象としており、商業的にも応用が進んでいる分野も存在する。しかしながら、このような限られた分野の開発者だけでは、音声インタフェースの普及は進まず、従来の規定概念にとらわれない新しい音声認識技術の利用方法も生まれない。このような問題に対して、音声認識技術アプリケーションの開発を困難にする課題について、これまでも様々な分析と報告が行われている [6][7]。本稿では、音声インタフェース設計の困難性に関する問題を取り上げ、良い音声インタフェースの設計の知見を共有し、新規の開発者でも適切な音声インタフェースの設計を可能にする枠組みについて議論を行う。

2.2 音声インタフェース設計の課題

音声インタフェース設計の困難さは、その多様性にある。例えば、同じ音声インタフェースと言っても、機能選択のための音声インタフェースであったり、ディクテーションやデータ入力のための音声インタフェース、または音声対話システムであったりと、それぞれ全く違ったインタフェース設計が必要となることがある。また、GUIアプリケーションと比較して、基本的な入出力方式が定義されておらず、標準的な使い方もないという致命的な問題がある。一般的に GUI の場合、全ての GUI アプリケーションは以下の 3 つの入出力を用いて説明できる。

- スクリーンデバイスによる情報提示
- ポインティングデバイスによる項目選択
- キーボードデバイスによるデータ入力

もちろん、キーボードがソフトウェアキーボードになったり、テンキー入力になったりと差異は存在する。ポインティングデバイスに関しても、マウスではなくタッチパネルの利用を前提とするケースや、近年ではマルチタッチを採用したインタフェースも選択肢となりうる。しかしながらいずれの場合も基本的な概念は同一である。それに対して音声インタフェースの場合、このように抽象化してインタフェースを表現することが出来ない。

また、もうひとつの特殊性として、音声認識システムの存在が無視できないという点がある。これは、音声認識システムが単体として動作可能なライブラリとして利用可能になるのではなく、語彙や言語モデル、音響モデ

ル、それらの変更による認識精度の変化、さらに場合によっては物理的なマイクロフォンの数や配置、信号処理方式まで含めて設計の対象としなくてはならないという点である。GUI 設計で例えると、テキストとして入力されるデータは常に正確ではなく、その精度はインプットメソッドの構成や入力方式によって大きく変化するため、それらの影響を念頭においたインタフェース設計が求められるというのに等しい。さらに、システムリソースや処理時間を無視することができないため、稼働プラットフォームやアーキテクチャ設計も念頭に置かなくてはならない。

2.3 パタン・ランゲージを用いた設計指針の記述と知見の共有

前節で述べた課題を考えると、音声インタフェース設計に関しては、モデルを単純に抽象化して設計論を検討するのは非常に困難だということがわかる。そこで本節では、パタン・ランゲージ [8] を用いて音声インタフェースの設計指針を記述することを検討する。パタン・ランゲージとは、ある状況下で繰り返し発生する問題と、熟練者によって得られる解決策のセットであるパターンの集合であり、特定の分野で発生する複数の問題に対して一般的で抽象的な解法を提供する。全てのパターンには名前が付けられ、通常どのパターンも同一のフォーマットで記述される。また、それぞれのパターンはその問題の前提条件や解法によって新たに発生する問題によって互いに関係を持ち、全体の問題領域に対して最適な解法として体系立てる。

例えば音声認識アプリケーション設計のためのパタン・ランゲージの場合、問題をマルチモーダルインタフェース設計として捉え、問題の前提条件として、ポインティングデバイスの有無、キーボードデバイスの有無によってパターンをそれぞれ記述することができる。また、孤立単語認識を利用した Command and Control や、大語彙連続音声認識を利用した音声対話システム等、様々なアプローチをそれぞれ別の問題に対する別のパターンとして記述する。さらにそこから発生する別の問題に対する解法をさらにそれぞれ別のパターンとして記述することで、音声認識アプリケーション設計のための共通の解法として提供することが可能となる。また、解法をパターンとしてまとめそれを知見として共有することで、特定の特徴をもったインタフェースに明確な名前が定義されるようになり、開発者間のコミュニケーションコストの低下が見込まれる。

2.4 音声認識アプリケーション開発の為のパタン・ランゲージ

実際に作成したパターンの例を示す。実際にはこのような例を数十から数百作成し、それらを分類し、体系づけることでパタン・ランゲージとしてまとめられる。

片手持ちゲームコントローラ

運転中や移動中でも利用可能なシステムを構築するために、音声入力を用いた効率的なインタフェースを構築したい。この場合、ポインティングデバイスやキーボードのような視覚フィードバックが必要なデバイスの利用は困難であるが、完全なハンズフリー環境を必要としている訳ではなく、簡単な操作も可能である。

問題

- 音声入力の不正確性に起因する、ユーザビリティの低下にどう対処すべきか？
- 音声認識に要する処理の遅延を如何に抑えるべきか？

解決策

- 複数のボタンを備えた、片手で操作可能な小型コントローラを利用し、敏速かつ正確性が要求される操作をコントローラで行い、音声入力の不正確性を補完する。
- 音声入力、コントローラでは入力操作が困難な個所で使用する。
- コントローラとしては、ゲームで使用されているような一般的なボタン(上下左右と26個程度の押しボタン)を備えたものとし、その操作の習得に練習をほとんど必要としないものとする。
- ボタンの数やボタンに対する機能の割り当ては、標準的なものがあればそれに従う。

結果

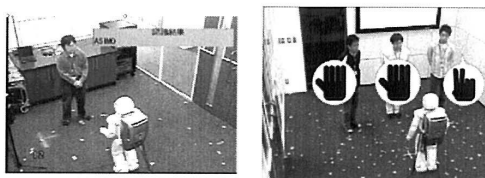
- 発話タイミングの指定ややり直し発話等、音声入力に関する処理をコントローラで、データ入力のみを音声入力できると、必要に応じて処理を分割できる。
- 必ずボタン操作が必要となるため、ハンズフリー音声認識の利用はできなくなる。
- コントローラの操作が複雑になったり標準的な操作と異なると、ユーザビリティが大きく低下する。
- コントローラに慣れていないユーザの場合、コントローラの操作ミスにより、期待通りの操作ができなくなる。

関連するパターン

- 操作に関する音声フィードバックと組み合わせることで、ユーザはアイズフリー環境での操作の正確性を向上することができる。
- コントローラと音声入力を使った効率的なインタフェースとして、GUI との整合性の高いインタフェースを設計したい場合には、Select&Voice[9] が有用である。
- メニューベースのインタラクションを用いたインタフェースを構築する場合、メニューの移動と基本操作をコントローラに割り当て、Flexible Shortcuts[10] と併用することで効率的な音声操作が可能となる。

2.5 むすび

音声認識アプリケーション開発を困難にする問題として、音声インタフェースの設計に関する問題を取り上げ議論を行った。この音声インタフェースの抽象化が困



a) 音楽雑音下認識 b) 口じゃんけん判定
図 1: ロボットにおける音声認識

難であるという問題に対し、ボタン・ランゲージを用いた音声インタフェースの設計指針の記述方式についてのべ、その具体例を示した。今後は、既知の見聞からパターンの抽出を行い、パターン・ランゲージとしてまとめる予定である。

■ ロボット聴覚用オープンソースソフト HARK の紹介

中臺 一博 (HRI-JP/東工大)

3.1 はじめに

音声認識を用いるアプリケーションの例として、ホームロボットやサービスロボットを取り上げたい。こうしたロボットは、日常環境で人とコミュニケーションを行い、家事、子守、介護、エンターテインメントなど様々な面で人をサポートすることが期待されている。つまり、ロボットでの聴覚処理は重要な課題である。実際、「ロボット聴覚」研究 [11] という日本発の研究分野として、ロボティクス分野を中心に研究が行われている。しかし、ロボットはそれ自体が雑音源であること、周囲の雑音や反響が動的に変化すること、同時発話認識が必要な場合があること、入力音響信号が音声とは限らないことなど、音声認識への入力だけをとっていても様々な問題が山積している。こうした問題に音響モデルへの雑音適応など既存の音声認識技術だけでは対応することは難しい。我々は、マイクロホンアレイを用いた音源定位、追跡、分離などの前処理を導入し、さらに、ミッシングフィーチャ理論によって前処理と音声認識を統合することにより、上述の問題を解決する試みを行ってきた。結果として、同時発話認識や高雑音環境に耐えうるロボット聴覚システム HARK を開発した(図 1)。HARK は、本年度、研究用途の使用がフリーなオープンソースソフトとして公開を開始した¹ ので以下に簡単に紹介する。

3.2 HARK の紹介

HARK は、Honda Research Institute Japan Audition for Robots with Kyoto University の略称であり、中世の英語で listen の意味を持つ。HARK は、データフロー指向ソフトウェアプログラミング環境 FlowDesigner [12] で動作するモジュール群として実装されている。FlowDesigner はそれ自体がオープンソースであり、信号処理に必要なベクトル、行列、論理演

¹ 11/17 に、京大、12 月に韓国で HARK 講習会を行う予定。

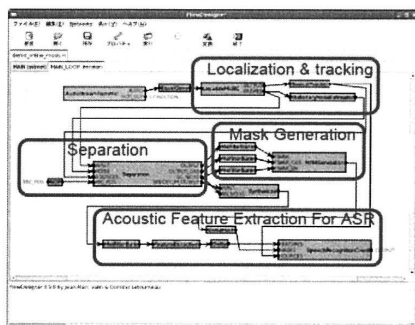


図 2: ロボット聴覚システムの例

算、デジタルフィルタ、GMM、HMM、ネットワーク通信、I/Oといったモジュールが予め用意されている。また、共有オブジェクトベースのモジュール統合を行っているため、モジュール間通信がポインタを引数とした関数呼び出で実現されており、遅延が少ないといった特長を備えている。

HARKの開発言語はC++で、現状ではLinux上で動作する。モジュールをGUI環境で組み合わせてプログラミングを行うことにより、フレキシブルにロボット聴覚システムを構築することができる。ALSAベースのサウンドカードなど多くの多チャンネル音響信号入力デバイスをサポートしており、音源定位(MUSIC)、追跡、音源分離(遅延和ビームフォーマ)、音声認識特徴量抽出(スペクトル特徴量)、ミッシングフィーチャマスク生成、音声認識(Julius/Julianへのパッチとして供給)などロボット聴覚に必要なモジュールを一通り用意している。HARKを用いて構築したロボット聴覚システムの一例を図2に示す。HARKの詳細については、文献[13, 14]を参考されたい。また、ソフトウェア、サンプルデータ、マニュアルはHARKのサイト²からダウンロード可能である。なお、同様にFlowDesigner上で動作するロボット聴覚システムManyEars³の音源定位(steered ビームフォーマ)、追跡(パーティクルフィルタ)、分離(Geometric Source Separation)もHARKで利用可能である。

HARKの公開の目的は、1) ロボット聴覚研究への貢献、2) 他分野へのツールとしての提供、3) HARKの高機能化・安定化である。ロボット聴覚で必要とされる前処理は、提供している定位・追跡・分離だけではなく、残響抑圧、音源同定、発話区間検出など多岐にわたっている。これらの一つだけでも大きなテーマであるが、HARK公開によって、一部の機能に特化して研究を行っている研究者にもシステムとしての有効性を検証できるプラットフォームを提供することができる。また、ロボティクスはそれ自身が異分野融合型の研究分野であり、分野の異なる研究者にユーザとして使ってもらえるツールを提供できることは大きな貢献である。コアなユーザ

² <http://winnie.kuis.kyoto-u.ac.jp/HARK/>

³ <http://manyyears.sourceforge.net/>

からバグ報告やパッチ提供を受けることによってツールの安定化を図ることができることは、オープンソースとして公開する大きなメリットである。

3.3 おわりに

アプリケーションを構築する際には様々な技術を統合するシステム統合技術が重要である。各技術を独立に扱って単純に接続するだけではシステム統合はうまくいかない。お互いの技術をよく知り、有機的に統合する必要がある。このためには、Julius/Julianのような音声認識システムだけではなく、それを取り巻く音響処理や対話処理といった部分のオープンソース化も重要な課題であろう。こうした意味でもHARK公開は一石を投じることができると考えている。

システム統合は重要な技術である割にはノウハウ的な側面が多く、その方法論を研究として成立させることが難しい。しかし、アプリケーションという観点から音声認識技術を考え、音響信号処理やマルチモダリティ統合の導入、対話システムなどを利用した音声認識エラー回復などを通じた統合的なシステム構築、および個々の技術の性能ではなく、システム全体としての性能向上を目指すシステム統合的な研究が増えいくことを期待したい。このためには、音声認識、音響信号処理、対話処理のみならず様々な分野間の相互理解を深めることが必要であろう。

■ 多言語音声認識機能を使いたい!

吉野 孝 (和歌山大)

4.1 多言語音声認識機能を使いたい!

私たちは、2007年、多言語医療受付対話支援システムの開発のために音声認識機能の利用を試みた。実際に使ってみると、様々な不満が出てきており、それは本質的には未だに解決していない。具体的な不満は下記である。

1. 現在の音声認識機能は、使用場面や話題、使用単語数など、ある程度範囲を限定するという条件下において、やっと利用が可能である。
2. 上記の限定された条件でさえ、条件を揃えるには大きな(それも非常に大きな)コストを要する。

私たちが開発している多言語医療受付対話支援システムM³は、医療従事者と外国人患者との対面における対話を支援するシステムである[15]。タッチパネルで利用するシステムであり、病院内の受付での利用を想定している。図1に問診用の画面の例を示す。この画面では、画面をタッチすることで症状の伝達が可能である。

開発中のシステムを医療従事者にデモしたところ、おおよその機能については満足してもらえた。しかし、計算機を用いた支援にもかかわらず、次の2点の機能が含まれていない点について、大きく失望された。

- 音声で操作できない
- 音声で回答が得られない



図 1: 多言語医療受付対話支援システム M³ の問診画面

多言語医療受付対話支援システム M³ は、数多くの対話用例があるため、その対話に必要なデータを探すために、音声で検索を行いたいという要望があった。また、対話支援システムであるため、その選択した対話文を発話して欲しいという要望があった。

そこで早速、多言語音声認識機能と多言語音声合成機能の導入の可能性について調べた。日本語の音声認識機能に関しては、あるオープンソースのシステムが有望そうであった。また、市販の音声認識ソフトウェアも候補として上がった。但し、両者とも、多言語への対応は難しくそうであった。音声合成機能に関しては、いくつか可能性のある市販のソフトウェアが見つかった。

まず、オープンソースの音声認識ソフトウェアを試したが、文章の音声認識はかなり難しいことがわかった。次に、市販の音声認識ソフトウェアについても検討したが、若干精度は良かったものの、同様に文章の音声認識はかなり難しいことがわかった。そこで、単語を認識することによって、検索支援を行うことにした。図 2 は、単語の認識による検索結果画面の一例である。「保険証」と発話したときに表示された結果である。「保険証」と発話すると、「保健所」と誤認識されることが多い。また、現時点では「多言語」の音声認識機能は利用できないため、日本人利用者側のみ利用が可能である。本システムは、実験的に京都市立病院に導入されているが、現時点では、この検索機能はほとんど利用されない機能となっている。

音声合成機能に関しては、さらに困難な状況であった。多言語の音声合成機能の開発会社に問い合わせたところ、1言語の音声合成機能に対して、かなりの料金(大学の1研究室で支払うことが難しい程度)の金額が必要なのことがわかった。また、対応言語の数にも限界があった(多言語医療受付対話支援システム M³ は5言語への対応が必要)。そこで、w3voice[16, 17]を用いて、音声合成ではなく、実際の人の音声をそのまま利用することにし、そのための音声収集ソフトウェアの開発を行った。音声収集という方法のために、即座に多様な音声へ対応することは難しいが、多言語には容易に対応するこ

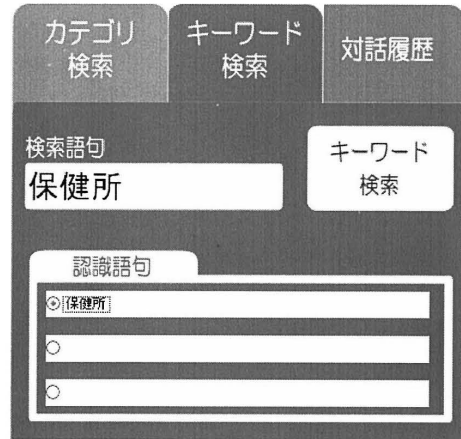


図 2: M³ のキーワード検索画面

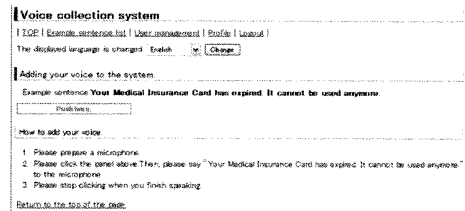


図 3: 音声収集ソフトウェアの画面

とが可能となった [18]。図 3 に開発した音声収集ソフトウェアの画面を示す。Web ブラウザ上で予め指定された文を読み上げることで、音声の収集を行うことができる。

4.2 音声認識機能に対する不満

音声認識機能は、未だに十分には利用されていない。最近では、多くの OS にも標準機能として装備されているものの、一般的な機能ではない。計算機の利用者は、様々な情報を入力するためにキーボードやマウスを用いる必要がある。計算機への情報入力装置として、音声認識は基本的な機能になってもおかしくないと思われるが、現時点では、計算機に情報を入力するためには、キーボードを使わざるを得ない。

このような基本的なインタフェースにもかかわらず、なぜこんなに精度がでないのだろうか? 音声認識機能が高精度に動作しないと、計算機への指示には、キーボードやタッチパネルが必要となる。いわゆるユビキタスな世界は現実のものとはならない。いつまでたっても人と計算機との対話は実現できない。

音声認識の研究者に音声認識機能の精度の向上について相談したことがある。ある程度の精度向上は可能という回答ではあったが、様々な工夫、いわゆる職人芸的な工夫が必要であり、単純な方法による精度向上は困難なことがわかった。また、多言語への対応についても相談したが、実現可能ではあるが、それを実現するためには

膨大な量の対話コーパスの収集（つまり、収集のための膨大な予算）が必要であることを告げられた。

最後に、音声認識機能に対する疑問を掲げる。

1. 音声認識機能は計算機を利用するための基本機能ではないのか？つまり、音声認識機能は重要と思われたが、実はこれだけ利用されていないのは、もともと必要とされていない機能（技術）であるからなのか？
2. 現在、限られたドメインでは十分な音声認識結果が得られるらしい。もし、大きな予算を投入すれば、汎用に近い音声認識機能は可能なのか？それともまだブレークスルーが必要なのか？人間の言葉を理解するには、まだまだかなりの時間を要するのか？あるいは、音声認識を完璧に行う、つまり、人間並みに行うことは、そもそも夢なのか？
3. 音声認識に関する技術は、未だ黎明期ではないのか？音声認識には、膨大な量のデータが必要と言われている。このような時期こそ、多言語コーパスなどのデータを多くの研究者がもっと自由に入手できるような環境の整備が必要なのではないか？膨大なデータの収集が可能な潤沢な資金を持つ、限られた研究グループや企業のみが音声認識の研究をするのではなく、多くの研究者が様々なアイデアを自由に試せるような環境を用意すべきではないのか？

今後の多言語音声認識技術の進展を、切実に待ち望んでいる。

■ 音声認識アプリ開発の促進に向けた課題

西村 竜一（和歌山大）

最後に、本パネルディスカッションの中で取り上げる予定の課題について述べる。ここで述べることは、目新しいことでは無く、歴史を繰り返しているのご批判もあるだろう。しかし、まだ解決を要する問題であり、議論は有益だと考える。

1. コーパススペースの開発をどう克服すべきか。
アプリ開発のために一からコーパスを構築するのは難しい。最近では Web ネットワークや集合知の利用が進んでいる。どのように集合知を利用すべきか。また、音声認識研究が集合知に求めるものは何だろうか。
2. 使えるツールは十分か。
Julius を筆頭に、諸先輩やワーキンググループ等の活動のおかげで使えるツールの数は増えている。それでもツールは依然として不足しているのだろうか？以前開発されたツールが古くなっているため、再開発が必要であるとの指摘もある。
3. 人間中心のデザインはどうあるべきか。
音声認識は本当に人間に親切なのだろうか。人間中心設計の考えからのインタフェースデザインが必要ではないだろうか。
4. 評価方法はどうかあるべきか。

「アプリ開発は論文にならない。」との意見が多い。研究として「評価」の難しさがその原因の一つだろう。どのような評価方法が考えられるのだろうか。また、開発成果物の品質保証はどのように行うべきなのだろうか。

参考文献

- [1] 後藤 真孝 他, “PodCastle の提案: 音声認識研究 2.0 を目指して,” 情報処理学会研究報告, 2007-SLP-65-7, 2007.
- [2] 緒方 淳 他, “PodCastle の実現: Web2.0 に基づく音声認識性能の向上について,” 情報処理学会研究報告, 2007-SLP-65-8, 2007.
- [3] <http://staff.aist.go.jp/k-kurihara/research/index-ja.html>
- [4] 栗原一貴 他, “音声ペン: 音声認識結果を手書き文字入力で見られる新たなペン入力インタフェース,” コンピュータソフトウェア, Vol.23, No.4, pp.60-68, 2006.
- [5] 栗原一貴 他, “プレゼン先生: 音声情報処理と画像情報処理を用いたプレゼンテーションのトレーニングシステム”, WISS 第 14 回 インタラクティブシステムとソフトウェアに関するワークショップ, pp.59-64, 2006.
- [6] 嵯峨山 茂樹, “音声認識の実用化を阻むもの,” 電子情報通信学会ソサイエティ大会講演論文集, 情報・システム (19950905), pp.312-313, 1995.
- [7] 石川 泰, “UI 設計とユーザビリティ~音声インタフェースの課題~, ” 情報処理学会研究報告, 2007-SLP-68-7-3, pp.35-40, 2007.
- [8] Christopher Alexander, “A Pattern Language,” Oxford University Press, 1977.
- [9] 梅本 暁, 中野 鐵兵, 小林 哲則, “GUI とのアナロジーに基づいた音声インタフェースの提案と評価,” 日本音響学会 2007 年秋季研究発表会, 2-3-5, 2007.
- [10] Teppei Nakano *et al.*, “Design and Formulation for Speech Interface Based on Flexible Shortcuts,” *Inter-speech2008*, pp.2474-2477, 2008.
- [11] K. Nakadai *et al.*, “Active audition for humanoid,” *AAAI-2000*, pp.832-839, 2000.
- [12] C. Côté *et al.*, “Reusability tools for programming mobile robots,” *IROS2004*, pp.1820-1825, 2004.
- [13] 中臺 一博 他, “ロボット聴覚用オープンソースソフトウェア HARK の概要,” *Robomec2008, 1P1-G13*, 2008.
- [14] K. Nakadai *et al.*, “An Open Source Software System For Robot Audition HARK and Its Evaluation,” *IEEE Humanoids 2008. (to appear)*.
- [15] 宮部 真衣, 吉野 孝, 西村 竜一, “病院受付における多言語間コミュニケーション支援システム M³ の開発,” マルチメディア, 分散, 協調とモバイル (DICOMO2007) シンポジウム, pp.355-363, 2007.
- [16] 使ってみよう音声入力・音声認識 Web アプリケーション, <http://w3voice.jp/>
- [17] 西村 竜一, 三宅 純平, 河原 英紀, 入野 俊夫, “音声入力・認識機能を有する Web システム w3voice の開発と運用,” 情報処理学会研究報告, 2007-SLP-68-3, 2007.
- [18] 福島 拓, 宮部 真衣, 吉野 孝, 西村 竜一, 重野 亜久里, “医療分野を対象とした多言語発話収集 Web システム OTOCKER の開発,” 電子情報通信学会, 人工知能と知識処理, AI2007-14, pp.17-22, 2007.