

機関番号：13904

研究種目：若手研究（B）

研究期間：2008～2010

課題番号：20700156

研究課題名（和文）

6階層モデルに基づくマルチモーダル対話システムのフレームワーク構築と実装

研究課題名（英文）

Construction and implementation of MMI system based on 6-layerd model

研究代表者

桂田 浩一（KATSURADA KOUICHI）

豊橋技術科学大学・国際交流センター・講師

研究者番号：80324490

研究成果の概要（和文）：

本研究ではマルチモーダル対話システムの6階層モデルを情報処理学会試行標準委員会の標準として取り纏め、公開を行った。また、6階層モデルに基づいたマルチモーダル対話システムのプロトタイプとして、一般的なWebブラウザからアクセス可能なマルチモーダル対話システムのプロトタイプを開発した。

研究成果の概要（英文）：

In this study we proposed a 6-layer model for multimodal interaction systems and published it as a standard under the Information Technology Standards Commission of Japan. We also developed a multimodal interaction system based on the model which has a feature that users can access to the system through a usual web browser installed in their PCs.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2008年度	800,000	240,000	1,040,000
2009年度	1,200,000	360,000	1,560,000
2010年度	1,200,000	360,000	1,560,000
年度			
年度			
総計	3,200,000	960,000	4,160,000

研究分野：マルチモーダル対話

科研費の分科・細目：情報学・知覚情報処理・知能ロボティクス

キーワード：マルチモーダル対話，6層モデル，標準化，webベースシステム

1. 研究開始当初の背景

音声認識／合成や顔画像認識／合成，ジェスチャ認識といった多様な入出力を統合したマルチモーダル対話システムの開発が盛んに行われている。ここ数年の間にAT&TのMATCH，DFKIのSmartkomをはじめとして，実用システムを目指した計画も多数進められてきた。国内においても，音声認識，音声合成，および顔画像合成を利用したGalateaツールキットが15大学の共同プロジェクトとして開発され，その一部がフリー

のツールとして公開されるなど，マルチモーダル対話システムの利用は徐々に拡大しつつある。研究代表者もこれまで，モダリティ拡張性の高いマルチモーダル対話システム，および対話記述言語XISLを提案すると共に，Galateaプロジェクトとその継続事業であるISTCにおいて，対話制御モジュールとプロトタイプングツールを開発し，マルチモーダル対話システムの普及に助力してきた。

2. 研究の目的

マルチモーダル対話システムは多様な入出力を統合制御するため、対話制御、入力統合、出力生成、音声認識...といった非常に多数のモジュールを内包する。このため、マルチモーダル対話システムの内部をどのように設計するか、すなわちシステムのフレームワークについても W3C を始めとする機関から多様な提案がされてきた。しかしながら、未だにモジュールの基本設計や API レベルで標準となるものが提案されていないのが現状である。本研究では、ISTC および情報処理学会・試行標準委員会のメンバーと連携を取りつつ、国内での標準となるマルチモーダル対話システムのフレームワークを提案し、多くの研究者が共通に利用できる参照システムを開発することを目標とする。

3. 研究の方法

本研究では、6 階層モデルの各層の仕様策定と階層間 API の文書化、およびこれらに基づいたマルチモーダル対話システムの実装を行う。このうち、仕様策定と文書化については平成 20 年度から平成 21 年度にかけて行った。また、システムの開発は平成 20 年度から開始し、平成 22 年度までの間に行った。

4. 研究成果

まず、平成 20 年度には 6 階層モデルの各階層の詳細と、階層間 API に関して、以下の内容を検討した。

- 各層に関して、標準の記述言語を決定する予定であったが、フレームワークの自由度を高めるため、特に標準言語は設けないことにした。
- 階層間の情報授受について、ソケット通信等を利用した通信のメッセージ/イベントをまとめた。
- 以上の検討項目を文書にまとめ、情報処理学会の試行標準として公開の準備作業を行った。また、情報処理学会試行標準委員会のひとりのメンバーが W3C の f2f ミーティングにおいて本取り組みに関する報告を行った。

続いて、平成 21 年度には 6 階層モデルの標準化とプロトタイプシステムの開発に関して、以下の項目を実施した。

- 平成 20 年度に検討したマルチモーダル対話システムの 6 階層モデルについて、各階層の役割、および階層間 API を文書化した。また、この内容を情報処理学会試行標準委員会に提出し、平成 22 年 2 月 3 日に第 12 番目の試行標準として公開した。

- 上記の試行標準に従ったマルチモーダル対話システムとして、PC 上の一般的なブラウザで動作する web ベースのマルチモーダル対話システムを試作した。このシステムは申請者が Galatea プロジェクトおよび音声対話技術コンソーシアム (ISTC) において開発を進めてきた Galatea ツールキットをベースにしており、実写真から作成した擬人化エージェントとのマルチモーダル対話を web ブラウザ上で行うことができる。ブラウザ側の処理はデファクトスタンダード (Flash, Java Applet, JavaScript 等) のみを利用して実現しているため、一般的な web ユーザが特別なソフトウェアやプラグインをインストールすることなくマルチモーダル対話を体験できる。6 階層モデルに準拠することによりシステムの構成変更に対する柔軟性が増したことから、モダリティや対話記述言語を切り替えることが容易になった。

最後に、平成 22 年度は参照システムの構築の一環として、マルチモーダル対話システムのプロトタイプ開発に関して、以下の各項目を実施した。

- 多様な端末で動作するマルチモーダル対話システムの実現を目指して、スマートフォン上で動作するシステムを試作した。このシステムは平成 21 年度に作成した PC 上の一般的なブラウザで動作するマルチモーダル対話システムをベースに構築した。
- ユーザとシステムとのシームレスなマルチモーダル対話を実現するために、対話エージェントが端末間を自由に行き来するためのシステムのアーキテクチャとプロトコルを検討した。また、プロトタイプを作成して、PC とスマートフォンの間に対話エージェントが移動することを確認した。
- 開発者が容易にマルチモーダル対話システムを開発できるよう、Web アプリケーションの開発にしばしば用いられる JavaScript 用の対話 API を作成した。これによりエージェントによる発話、エージェントの移動、ユーザの発話受付やマルチモーダル入力等が JavaScript プログラム内で自由に記述できるようになった。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計1件)

- [1] [1] Yurie Iribe, Hiroaki Nagaoka, Kouichi Katsurada and Tsuneo Nitta: "Web-Based Lecture System Using Slide Sharing for Classroom Questions and Answers", International Journal of Knowledge and Web Intelligence, Vol.1 No. 3/4, pp. 243-255 (2010-7).
- [学会発表] (計49件)
- [1] Masashi Kimura, Takayuki Onoda, Yurie Iribe, Kouichi Katsurada and Tsuneo Nitta: "One-Model Speech Recognition and Synthesis Based on Articulatory Movement HMMs", Proc. of NCSP11, pp. 392-395 (2011-3).
- [2] Tsuneo Nitta, Takayuki Onoda, Masashi Kimura, Yurie Iribe and Kouichi Katsurada: "One-Model Speech Recognition and Synthesis Based on Articulatory Movement HMMs", Proc. of INTERSPEECH2010, pp. 2970-2973 (2010-9).
- [3] Ryuichi Fukui, Kouichi Katsurada, Yurie Iribe and Tsuneo Nitta: "Facial Expression Mimicking System", Proc. of ICPR2010, pp. 3776-3779 (2010-8).
- [4] Yurie Iribe, Hiroaki Nagaoka, Kouichi Katsurada and Tsuneo Nitta: "Web-based Lecture System using Slide Share for Questions and Answers in the classroom", Proc. of ICCE2009, pp. 638-642 (2009-12).
- [5] Kouichi Katsurada, Akinobu Lee, Tatsuya Kawahara, Tatsuo Yotsukura, Shigeo Morishima, Takuya Nishimoto, Yoichi Yamashita and Tsuneo Nitta: "Development of a Toolkit for Spoken Dialog Systems with an Anthropomorphic Agent: Galatea", Proc. of APSIPA09, pp. 148-153 (2009-10).
- [6] Kouichi Katsurada, Shigeki Teshima and Tsuneo Nitta: "Fast Keyword Detection Using Suffix Array", Proc. of INTERSPEECH2009, pp. 2147-2150 (2009-9).
- [7] Mohammad Huda, Hiroaki Kawashima, Kouichi Katsurada and Tsuneo Nitta: "Distinctive Phonetic Feature (DPF) Based Phoneme Recognition Using MLNs and Inhibition/Enhancement Network for Noise Robust ASR", Proc. of NCSP09, 1PM2-K3 (2009-3).
- [8] Kouichi Katsurada, Teruki Kirihata, Masashi Kudo, Junki Takada and Tsuneo Nitta: "A Browser-based Multimodal Interaction System", Proc. of ICMI'08, pp. 195-196 (2008-10).
- [9] Mohammad Nurul Huda, Kouichi Katsurada and Tsuneo Nitta: "Phoneme Recognition Based on Hybrid Neural Networks with Inhibition/Enhancement of Distinctive Phonetic Feature (DPF) Trajectories", Proc. of INTERSPEECH2008, pp. 1529-1532 (2008-9).
- [10] 桂田 浩一, 李晃伸, 河原達也, 四倉達夫, 西本卓也, 山下洋一, 新田恒雄: "擬人化エージェントとの音声対話を実現するツールキットの開発", HAI シンポジウム 2008, 1C-1 (2008-12).
- [11] 澤田 心太, 桂田 浩一, 入部 百合絵, 新田 恒雄: "音声ドキュメント高速検索におけるクエリ分割とマッチング手法の比較検討", 日本音響学会 2011 年春季研究発表会講演論文集, 3-5-18 (2011-3).
- [12] 小野田 高幸, 入部 百合絵, 桂田 浩一, 新田 恒雄: "調音運動 HMM による少量音声サンプルからの音声合成", 日本音響学会 2011 年春季研究発表会講演論文集, 3-7-10 (2011-3).
- [13] 佐々木 俊, 木村 優志, 入部 百合絵, 桂田 浩一, 新田 恒雄: "調音特徴を用いた確率モデルに基づく VAD", 日本音響学会 2011 年春季研究発表会講演論文集, 1-5-14 (2011-3).
- [14] 荒木 厚太, 小野田 高幸, 桂田 浩一, 新田 恒雄: "調音運動 HMM に基づく高精度音素認識の検討", 日本音響学会 2011 年春季研究発表会講演論文集, 1-5-2 (2011-3).
- [15] 澤田 心太, 桂田 浩一, 入部 百合絵, 新田 恒雄: "高速音声ドキュメント検索における検索クエリ分割手法およびマッチング手法の比較評価", 第 5 回音声ドキュメント処理ワークショップ講演論文集, (2011-3).
- [16] 小野田 高幸, 桂田 浩一, 新田 恒雄: "調音運動 HMM 音声合成における調音と音源の改良", 日本音響学会 2010 年秋季研究発表会講演論文集, 2-Q-9 (2010-9).
- [17] 澤田 心太, 桂田 浩一, 新田 恒雄, 入部 百合絵, 手島 茂樹: "大規模音声ドキュメントからの高速キーワード検索法の提案とその評価", 日本音響学会 2010 年秋季研究発表会講演論文集, 2-9-10 (2010-9).
- [18] 森 拓郎, 入部 百合絵, 桂田 浩一, 新田 恒雄: "調音特徴に基づく非母語話

- 者の英語発音評価”，日本音響学会 2010 年秋季研究発表会講演論文集，3-1-3 (2010-9).
- [19] 木村 優志，小野田 高幸，入部 百合絵，桂田 浩一，新田 恒雄：“調音運動に基づくワンモデル音声認識合成への CELP 適用”，2010 年度人工知能学会全国大会論文集，1J1-OS13-2 (2010-6).
- [20] 澤田 心太，木村 優志，桂田 浩一，新田 恒雄：“情景と音声言語の混在情報から得た部分空間に基づくタスク推定の改良”，2010 年度人工知能学会全国大会論文集，1J1-OS13-10 (2010-6).
- [21] 福井 竜一，桂田 浩一，入部 百合絵，新田 恒雄：“AAM を利用した表情成分抽出に基づく表情模倣システム”，電子情報通信学会技術研究報告，PRMU2009-322，pp.513-518 (2010-3).
- [22] 工藤 正志，桂田 浩一，入部 百合絵，新田 恒雄：“柔軟なモジュール切替えを可能にした Web ベース MMI システムの開発”，情報処理学会第 72 回全国大会論文集，5U-2 (2010-3).
- [23] 福井 竜一，桂田 浩一，入部 百合絵，新田 恒雄：“AAM を利用した表情の模倣”，情報処理学会第 72 回全国大会論文集，4Y-5 (2010-3).
- [24] 長岡 紘昭，入部 百合絵，桂田 浩一，新田 恒雄：“調音特徴に基づく音素単位での英語発音誤り検出と発音評価”，情報処理学会第 72 回全国大会論文集，1U-10 (2010-3).
- [25] 手島 茂樹，桂田 浩一，入部 百合絵，新田 恒雄：“大規模音声データベースに対する高速なキーワード検索”，日本音響学会 2010 年春季研究発表会講演論文集，3-6-7 (2010-3).
- [26] 下田 和樹，入部 百合絵，桂田 浩一，新田 恒雄：“調音特徴抽出技術を用いた日本語発音学習”，日本音響学会 2010 年春季研究発表会講演論文集，3-7-6 (2010-3).
- [27] 木村 優志，澤田 心太，桂田 浩一，新田 恒雄：“情景と音声言語の混在情報から得た部分空間に基づくタスク推定”，日本音響学会 2010 年春季研究発表会講演論文集，3-6-3 (2010-3).
- [28] 福井 竜一，桂田 浩一，入部 百合絵，新田 恒雄：“顔まねエージェントシステム”，インタラクション 2010，SA-08 (2010-3).
- [29] 長岡 紘昭，入部 百合絵，桂田 浩一，新田 恒雄：“英語発音学習のための調音特徴抽出と発音評価”，情報処理学会研究報告 2010-SLP-80，No. 13 (2010-2).
- [30] 手島 茂樹，入部 百合絵，桂田 浩一，新田 恒雄：“大規模音声ドキュメントを対象とした高速キーワード検索システムとその評価”，情報処理学会研究報告 2010-SLP-80，No. 11 (2010-2).
- [31] 工藤正志，桂田 浩一，入部 百合絵，新田 恒雄：“階層型アーキテクチャに基づいた Web ベース MMI システムの開発”，電子情報通信学会技術研究報告，SP2009-146，pp.351-356 (2010-1).
- [32] 荒木 雅弘，西本 卓也，桂田 浩一，新田 恒雄：“階層的 MMI アーキテクチャに基づくプラットフォーム実装方法の検討”，情報処理学会研究報告 2010-SLP-78，No. 5 (2009-10).
- [33] 手島 茂樹，桂田 浩一，新田 恒雄：“Suffix Array を用いた高速なキーワード音声検索システム”，日本音響学会 2009 年秋季研究発表会講演論文集，1-R-26 (2009-9).
- [34] 福井 竜一，桂田 浩一，新田 恒雄：“AAM を利用した顔画像処理における表情成分と個人成分の分離”，平成 21 年度電気関係学会東海支部連合大会講演論文集，0-207 (2009-9).
- [35] 工藤 正志，桂田 浩一，入部 百合絵，新田 恒雄：“MMI6 階層モデルに準拠した Web ベース MMI システムの開発”，FIT2009 情報科学技術フォーラム，E-039 (2009-9).
- [36] 手島 茂樹，桂田 浩一，新田 恒雄：“Suffix Array を用いた高速なキーワード検索”，情報処理学会研究報告 2009-SLP-77，No. 3 (2009-7).
- [37] 新田 恒雄，武井 匠，木村 優志，桂田 浩一：“調音運動 HMM に基づくワンモデル音声認識合成”，情報処理学会研究報告 2009-SLP-77，No. 4 (2009-7).
- [38] 溝口 勇太，木村 優志，桂田 浩一，新田 恒雄：“Q 学習を用いた協調行動のための戦略獲得”，2009 年度人工知能学会全国大会論文集，1F2-OS7-11 (2009-6).
- [39] 澤田 心太，木村 優志，入部 百合絵，桂田 浩一，新田 恒雄：“情景と音声言語の混在情報から得た部分空間に基づくタスク推定”，2009 年度人工知能学会全国大会論文集，1F2-OS7-7 (2009-6).
- [40] 武井 匠，木村 優志，桂田 浩一，新田 恒雄：“調音特徴に基づく 1-model 音声認識-合成”，2009 年度人工知能学会全国大会論文集，1F2-OS7-6 (2009-6).
- [41] 手島 茂樹，桂田 浩一，新田 恒雄：“Suffix Array を用いた音声文書の高速全文検索”，日本音響学会 2009 年春季研究発表会講演論文集，2-5-2 (2009-3).
- [42] 工藤 正志，桐畑 輝樹，高田 淳貴，桂田 浩一，新田 恒雄：“ウェブブラウザをインターフェースとしたマルチモーダル対話システム”，インタラクシ

- ン 2009, C15 (2009-3).
- [43] 手島 茂樹, 桂田 浩一, 新田 恒雄: “Suffix Array を用いた音声文書の高速検索”, 第 3 回音声ドキュメント処理ワークショップ講演論文集, pp. 27-32 (2009-2).
- [44] 田村 哲嗣, 桂田 浩一, 伊藤 慶明: “マルチモーダル情報処理の研究動向”, 電子情報通信学会技術研究報告, SP2008-72, pp. 13-18 (2008-11).
- [45] 桐畑 輝樹, 工藤 正志, 高田 淳貴, 桂田 浩一, 新田 恒雄: “ウェブブラウザ上で動作可能なマルチモーダル対話システム”, 情報処理学会研究報告 2008-SLP-73, pp. 35-40 (2008-10).
- [46] 山下 洋一, 李 晃伸, 河原 達也, 四倉 達夫, 西本 卓也, 桂田 浩一, 新田 恒雄: “音声対話技術コンソーシアム (ISTC) の活動成果報告”, 情報処理学会研究報告 2008-SLP-73, pp. 47-52 (2008-10).
- [47] 木村 優志, 作元 佑輔, 田口 亮, 桂田 浩一, 岩橋 直人, 新田 恒雄: “発話シーンの共有信念に基づく推定とその評価”, 日本認知科学会第 25 回大会発表論文集, pp. 318-319 (2008-9).
- [48] 木村 優志, 作元 佑輔, 田口 亮, 桂田 浩一, 岩橋 直人, 新田 恒雄: “人間 - ロボット間の共有信念に基づく発話場面の推定”, 2008 年度人工知能学会全国大会論文集, 3E3-07 (2008-6).
- [49] 溝口 勇太, 田口 亮, 木村 優志, 土井 岡 伴哉, 桂田 浩一, 新田 恒雄: “知

的エージェント学習実験プラットフォームの構築”, 2008 年度人工知能学会全国大会論文集, 3E3-08 (2008-6).

[図書] (計 0 件)

[産業財産権]

○出願状況 (計 1 件)

名称: 音声検索装置および音声検索方法
発明者: 桂田 浩一, 手島 茂樹, 新田 恒雄
権利者: 桂田 浩一, 手島 茂樹, 新田 恒雄
種類: 特許
番号: 特願 2009-44842
出願年月日: 2009 年 2 月 26 日
国内外の別: 国内

[その他]

ホームページ等

<http://www.itscj.ipsj.or.jp/ipsj-ts/ts0012/toc.htm>

<http://www.vox.tutkie.tut.ac.jp/html/Introduction/pdf/press.pdf>

6. 研究組織

(1) 研究代表者

桂田 浩一 (KATSURADA KOUICHI)
豊橋技術科学大学・国際交流センター・講師
研究者番号: 80324490