

## 科学研究費助成事業 研究成果報告書

平成 26 年 6 月 23 日現在

機関番号：94301

研究種目：若手研究(A)

研究期間：2011～2013

課題番号：23680019

研究課題名(和文) 韻律・声質の動的特徴および形態素・品詞を考慮した発話意図認識システムの構築

研究課題名(英文) Development of a speech act identification system considering prosodic, voice quality and linguistic information

研究代表者

石井 カルロス寿憲 (Ishi, Carlos Toshinori)

株式会社国際電気通信基礎技術研究所・知能ロボティクス研究所・室長

研究者番号：30418529

交付決定額(研究期間全体)：(直接経費) 19,300,000円、(間接経費) 5,790,000円

研究成果の概要(和文)：音声，EGG，言語情報，発話スタイル，発話意図を含む自然対話音声データベースを構築した。氣息音発声において新たな音響特徴量を提案した。構築したデータベースを利用し，多くの発話意図を伝達する単音節の感動詞(うん，あー，おー，やー，はー，等)を取り上げ，形態素と発話スタイル(音調と声質)との関連構造を解明した。感動詞の音例集を構築し，Web上で公開した。発話意図の識別性を評価し，提案した音響特徴の効果を示した。感動詞の繰り返し発話(うんうん等)に対し，繰り返しの数と音調と機能的役割に関連がみられた。質問系発話において，文末音調が上昇しないものが多くみられ，対人関係における韻律の使い分けも示された。

研究成果の概要(英文)：We constructed a natural conversational dialogue speech database, including linguistic, prosodic and paralinguistic information. A new acoustic parameter was proposed for breathy voice quality. Monosyllabic interjection utterances (such as "un", "oh", "haa") were extracted from the database, and the relations between speaking style and speech acts were analyzed. A set of interjections expressing different meanings were made available in the web. Automatic identification of speech acts were also evaluated, indicating the effectiveness of the proposed acoustic features. Repeated interjection utterances (such as "unun") were also analyzed, indicating relation between the number of repetitions and the paralinguistic functions. Question-type utterances were also analyzed, indicating relations between the phrase final intonation and the inter-personal relationship.

研究分野：総合領域

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：知覚情報処理 音声情報処理 パラ言語情報処理 韻律情報処理 言語情報処理 声質特徴 感情音声  
 自然対話音声

1. 研究開始当初の背景

従来の対話システムの殆どは、図1の上部のブロックに示すように、音声のスペクトル特徴を用いて発話内容（単語列または音素列）のみを認識する「機械的な」コミュニケーションに留まっている。しかし、エージェントやロボットなどに対話システムを応用する際、サービスに応じて、より自然な人間らしい対話コミュニケーションが求められ、「え」のような感動詞や、平叙文の「今日.」、考え中の「今日...」、疑いの「今日??」、などのように、言葉が持つ意味のみならず、その言い方によって伝達される発話意図を正しく理解する必要がある。

対話システムが、より完全な音声理解を行うため、従来の音声認識技術によって出力される発話内容と並列して、図1の中部のブロックに示すように、音声に含まれるイントネーションなどの韻律特徴と氣息音などの声質特徴で表現される発話スタイルによって、意図・態度・感情などのパラ言語情報（総称して「発話意図」）を認識する構造が効果的である。なお、喜怒哀楽の感情を対象とした国内外の研究は多いが、対話理解に重要な発話意図全体を考慮したものは少ない。

提案者はこのような構造機構を実現するため、これまで声質に関する音響特徴や発話意図との関連構造の研究開発を進めてきた。「え」と「うん」の特定の感動詞で評価を行った結果、発話スタイルと発話意図の間に関連はみられたが、その関連構造が他の感動詞（「あ」、「あら」、「あれ」、「はい」、「や」、「わ」など）に直接対応するとは限らない。共通した部分もあるが、例えば音調が同じ下降上昇型でも、残念の「あーあ」や、否定の「うん」のように、発話スタイルと発話意図の関連は一途ではなく、感動詞の種類により、関連する発話スタイルの種類も表現可能な発話意図の項目も異なる場合が多い。従って、発話スタイルのみならず、形態素などの言語情報も発話意図の識別に考慮する必要がある。そこで、本提案では、形態素や品詞などの言語情報も導入し、図1の下部に示すように、形態素や品詞の違いによる、発話意図と発話スタイルの種類と関連構造の解明に取り組む。

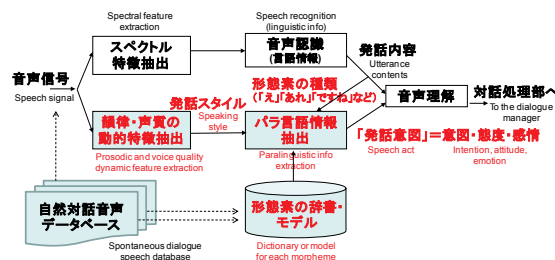


図1. 本提案がねらいとする発話意図認識の構造

2. 研究の目的

本提案では、発話スタイルの音響的表現と

発話意図との関連の2つの問題点を解決する方向で、発話スタイルに韻律および声質の動的特徴を導入し、形態素や品詞などの言語情報も考慮することにより、形態素（あるいは形態素のグループ）毎に、発話スタイルと発話意図を関連付ける「辞書」ないしモデルを構築し、これを用いて発話意図認識システムを実現することを目的とする。

本研究の長期的な目標は、人と機械の間に円滑な対話コミュニケーションを成立させることである。そのために、本申請では、対話音声において、発話された言葉のみでは意図が伝達出来ない場合、その言い方（発話スタイル）によって、伝えたい意図が質問であるのか、断言であるのか、否定的な表現であるのか、驚きの表現であるのかなどの発話意図の種類を認識するシステムを構築する。そのため、発話スタイルと言語情報と発話意図の関連構造を自然対話を分析することにより解明する。

3. 研究の方法

(1) 対話音声データベースの構築

音声および声帯振動のデータと言語情報、発話スタイル、発話意図の情報を含む自然対話音声データベースを構築する。

音声の収集には、各対話者に向けた指向性マイク、ヘッドセットマイクおよびEGG装置 (Electro-glottograph) を用いた。EGG信号には音調に関連するピッチ情報のみならず、声帯の振動様式の違いによる声質の情報が多く含まれ、発話スタイルの記述に役立つ。

研究補助者により、収集したデータベースの発話区間切り出し、書き起こし、およびラベル付与作業を進めた。それぞれの発話において「発話意図」および「談話機能」のラベルを付与した。

(2) 音響特徴の抽出

構築したデータベースを用いて、発話スタイルの表現に用いる、イントネーションなどの韻律および氣息音などの声質に関連する音響特徴の抽出を行った。

イントネーションに関連するF0の抽出には、声帯振動を直接測定したEGGの波形を用いる。音声波形ではなく、EGG波形を用いることで、高精度のF0抽出が可能となる。

声質には、氣息性を含んだ発声、りきみ発声、パルス的なフライ発声など、さまざまな種類が存在し、提案者がこれまで独自に提案してきたパラメータや手法を用いて自動抽出を行う。ただし、音声波形の代わりにEGG波形を利用できる場合は、高精度の検出が可能となる。また、抽出された音響特徴および声質の区間情報は、発話スタイルと発話意図の関連を求めるための使用に限らず、声質の特徴抽出を改善するための正解データとしても使用した。

また、「うんうん」のように感動詞が連続するものと、長い「うーん」との識別にお

いて、ピッチの下降度合い、あるいは F0 の軌道のみでは識別がつかない場合、スペクトル成分の動的特徴を導入することにより、異なる発話意図が識別される閾値を求めた。

### (3) 言語情報の解析

本提案では、発話意図や談話機能の表現に重要となる品詞に着目し、発話を「感動詞」；「フレーズが助詞・助動詞で終わる発話」；「接続詞」；「言い詰まりや言い淀み」などに分類し、形態素や品詞、修飾の関係などの言語情報およびそれに伴うイントネーションと、談話機能や発話意図の表現との関連について解析を行った。

### (4) 発話スタイルと発話意図との関連構造の解明およびモデルの構築

発話スタイルと発話意図との関連構造を解明する。形態素または形態素のグループ毎に、韻律・声質の動的特徴による発話スタイルと発話意図・談話機能との関連構造の辞書またはモデルを構築する。「発話意図」の中には、「肯定」や「聞返し」のようにカテゴリカル（離散的）なものもあれば、「驚き」や「怒り」などのように度合いで表すものもある。また、複数の意図・態度・感情が同時に表現されることも考慮する。

そこで、音響特徴抽出部で得られた韻律および声質に関する音響特徴を用いて、発話意図との関連構造のモデル化を検討する。

### (5) 発話意図認識システムの構築および評価実験

最終的な対話システムでは、音声認識を通して発話内容を出力する必要があるが、本提案では、発話内容を既知と想定した場合の発話意図認識システムを構築し、発話スタイルと発話意図との関連構造の辞書・モデルを用いて、発話意図の理解度を評価する。

## 4. 研究成果

### (1) 対話音声データベースの構築

音声および声帯振動のデータと言語情報、発話スタイル、発話意図の情報を含む自然対話音声データベースを構築した。男女計 31 名のデータ収集を行った。1 対話当たりおよそ 10 ~ 15 分程度で、合計 25 時間のデータが揃った。

収集したデータベースを用い、それぞれの発話に「発話意図」および「談話機能」のラベルの付与を実施した。3 名の被験者間の一致度を分析し、ラベルの統一を進めた。言語学の専門家も含め、議論を重ねた結果、これまでのラベルを「発話行為」、「心的態度」、「形態素的表現」、「談話機能」の種類にラベルを整理した。

### (2) 音響特徴の抽出：NBP (Normalized Breathiness Power)

声質に関する音響特徴分析においては、パ

ラ言語情報を多く伝達する「りきみ」と「氣息音」に焦点を当て、後者に関しては国際会議 Interspeech2011 で論文が採択された。

図 2 に氣息音の強度を定量化するために提案した音響パラメータ (NBP: Normalized Breathiness Power) の求め方を示す。このパラメータは過去に提案した F1F3syn を利用したものであり、F1F3syn では氣息音の有無を表現するが強度は表現できない欠点を補ったものとなる。

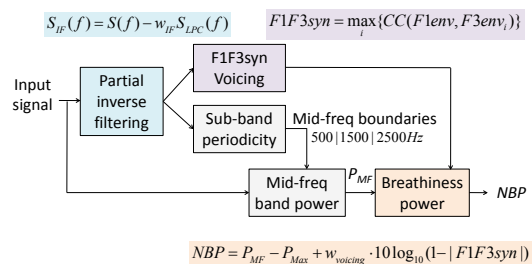


図 2. 提案した氣息音発生の音響特徴 (F1F3syn および NBP)

図 3 および図 4 にそれぞれ F1F3syn および NBP による、氣息音を含む区間 (br/wh, br/wh voiced) と氣息音を含まない区間 (modal, rough, fricative, aspirated, no label) の識別性を示している。

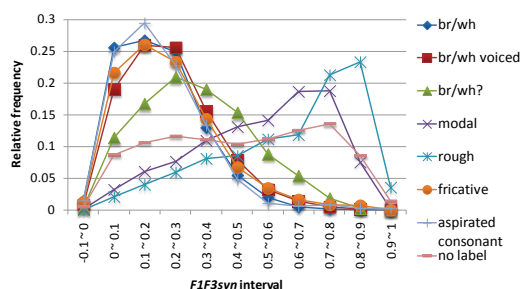


図 3. F1F3syn による氣息音の識別性

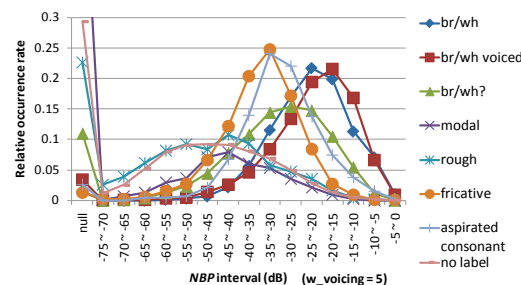


図 4. NBP による氣息音の識別性

図 3, 4 より、主に (fricative, aspirated, no label) の区間において、NBP により、識別性が高められたことが読み取れる。

図 5 には NBP の計算における重みづけパラメータを変化させた場合の効果を示しており、ベストの条件で、氣息音区間の検出率 87% 程度と挿入誤り率 12% 程度の結果が得られた。

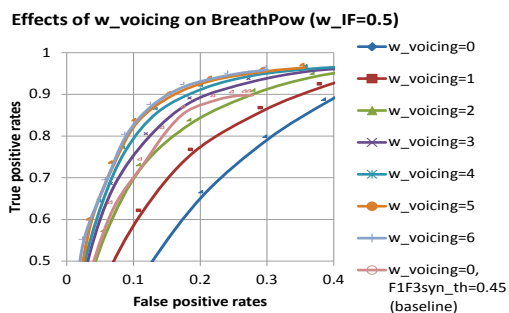


図 5. NBP による氣息音の識別率

### (3) 単音節の感動詞：形態素と発話スタイル (音調と声質) と発話意図との関連構造の解明および発話意図認識の評価

項目 (1) の自然対話音声データベースを用いて、感動詞発話をもたらす発話意図を分析した。過去に分析した「うん」と「ええ」に加え、「おー」、「おん」（「うん」のくだけた発声）、「わー」、「やー」の感動詞において、音調や声質の特徴と、発話意図の関連を調べ、国際会議 Interspeech2011 で発表した。

また、パラ言語情報を多く伝達する感動詞において、自然発話データベースから得られたサンプルを分析して感動詞の音例集を構築し、日本語学習者の発音教育への活用を目指している。このテーマについても国際会議（日本語教育国際研究大会 2012）で発表し、高い評価を得た。要望も多く、感動詞の音例集を Web ページにも公開した。

更に、これまで取り上げていなかった感動詞（「あー」、「おー」、「はー」、「へー」、「ほー」）について、発話意図と発話スタイルの関連を調べ、これまで開発してきた音響特徴を用いて発話意図の識別性の評価を行い、国際会議 Speech Prosody2012 に論文が採択された。

図 6 に「おー」、「おん」、「わー」、「やー」の感動詞をもたらす発話意図の分布を示す。感動詞の種類によって、表現される発話意図に偏りがあることが分かる。音調ではこれらを識別することが難しい結果も得られ、これらの感動詞の発話意図の識別には、感動詞の種類（形態素情報）が重要であることが示唆される。

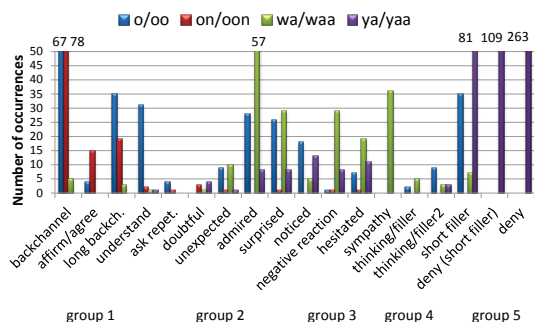


図 6. 感動詞「おー」、「おん」、「わー」、「やー」が伝達する発話意図の分布

ただし、図 7 に示すように、氣息音 (a, w) やりきみ発声 (p, h) の声質を含む発話は、感動詞「わ」および「や」に多く現れ、これらが伝達する「感心」、「驚き」、「否定的反応 (非難・嫌悪など)」、「同情」、「否定」など、強い感情や態度の表現に多く出現する結果が得られた。

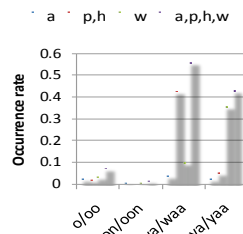


図 7. 感動詞「おー」、「おん」、「わー」、「やー」における声質の分布

図 8 に、「あー」、「おー」、「はー」、「へー」、「ほー」の感動詞において言語情報と韻律特徴を用いた発話意図認識の結果を示す。この結果より、「聞き返し」、「驚き」と「同情」は認識精度が 50% 以下となったが、「聞き返し」と「驚き」の発話には共通した発話スタイルが存在し、「同情」は「感心」や「理解」と共通した発話スタイルが存在するため、これらの識別には他の文脈を考慮する必要がある。

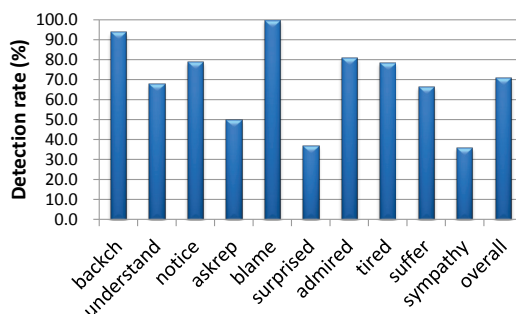


図 8. 単音節の感動詞における言語情報と韻律特徴を用いた発話意図認識の結果

図 9 に識別にイントネーションに関連する韻律特徴のみを使った場合 (F0move & dur)、氣息音の音響パラメータを加えた場合 (+breathiness)、りきみ発声の音響パラメータを加えた場合 (+pressed) と、感動詞の形態素情報も加えた場合 (+interjection id) の識別結果を示す。

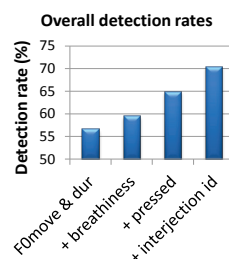


図 9. 単音節の感動詞における言語情報と韻律特徴を用いた発話意図認識の結果

この結果より、韻律特徴、声質特徴、言語情報を用いる効果が示される。但し、識別率70%を改善するには、音響特徴の改善は考えられるが、文脈を考慮することが最も効果的と考えられ、今後の課題として残される。

#### (4) 質問系発話：句末の形態素と音調と発話意図との関連構造の解明

項目(1)の自然対話音声データベースを用いて、言語情報と発話スタイルと発話意図の関連を解析した。主に、質問系発話において、句末音調が上昇することが一般的とされる概念に対し、自然発話では文末音調が上昇しないものが数多く観測され、その理由を解明するため、質問の種類を「情報要求」、「真偽要求」、「確認」、「反復的質問」、「相槌的質問」などに分類し、句末音調と句末の形態素によって分析を行った。その研究成果をまとめた論文は国際会議 *Interspeech2013* に採択された。また、対人関係における韻律の使い分けに関して分析を発展させた成果も国際会議 *Speech Prosody2014* で論文が採択され、雑誌論文(音声研究)にもまとめた。

図10に、自然発話データベースに現れた質問系発話の種類(YN質問、意見要求、情報要求、質問系相槌、提案、自己質問、同意要求、曖昧な質問)における句末音調(上昇調、リセット調、平坦調、リセット下降調、下降調)の分布を示す。

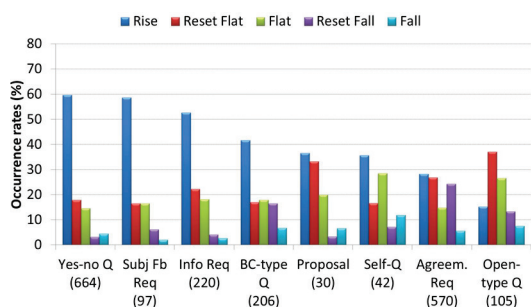


図10. 質問系発話の種類における句末音調の分布

この結果より、対話相手に情報を求める度合いが高いほど、句末音調は上昇調になりがちであることが導かれる。

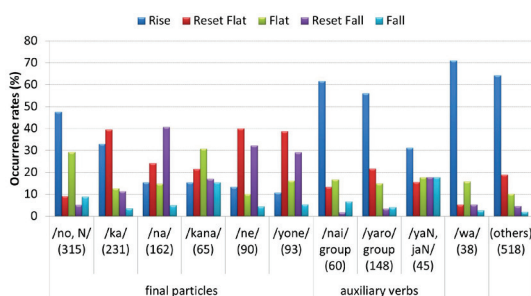


図11. 質問系発話の句末の形態素における句末音調の分布

図11には、句末の形態素別(「の、ん、か、な、かな、ね、よね、ない、やる、やん、わ、それ以外)で句末音調の分布を示す。

この結果より、相手に応答を求める意図を本質的に持つ形態素(「な、よね、ね、かな、か)では、上昇調の頻度が低くなる傾向があることが導かれる。

図12に、対人関係による影響(気を遣っている、気を遣っていない)の結果を示す。この結果より、対話相手に気を遣った場合、上昇調の出現率が下がる結果が得られた。

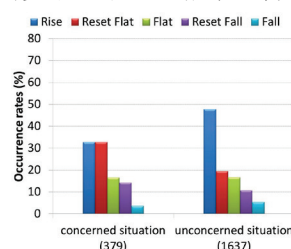


図12. 質問系発話の対話相手への気遣いにおける句末音調の分布

#### (5) 感動詞の繰り返し発話：繰り返しの数と音調と発話意図との関連構造の解明

項目(3)では、感動詞がもたらす発話意図において、単音節のもの(「ええ、うん、はい、へー、やー、わー)に焦点を当ててきたが、最終年度では、自然発話データベースから「うんうん」や「うんうんうん」のような繰り返し発話される感動詞の音響的特徴および機能的役割の分析を行い、日本音響学会で発表し、国際会議 *Speech Prosody 2014* にも論文が採択された。

まず、音響特徴の分析より、「うんうん」発話の識別には、音声認識で用いる MFCC のようなスペクトル特徴の変化のみでは不十分であり、基本周波数の変化も考慮すべきである結果が得られた。図13bではスペクトル変化が少なく、図13dでは基本周波数の変化が少ない例をそれぞれ示す。

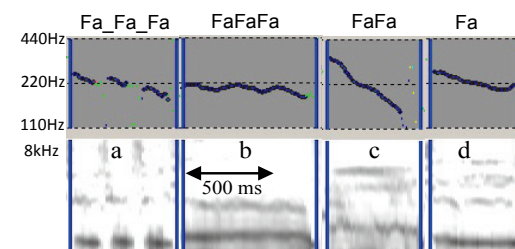


図13. 「うんうん」発話の基本周波数パターン(上図)とMFCC平滑化スペクトログラム(下図)の例

図14に「うんうん」発話において、「うん」の繰り返しの数における発話意図(相槌、肯定、同意、理解、興味)の分布を示す。

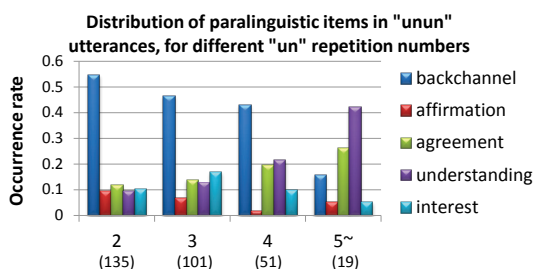


図1 4. 「うんうん」発話の「うん」の繰り返しの数における発話意図の分布

図1 4の結果より、繰り返しの数が増えるに応じて、単なる相槌から興味・同意・理解の意図が含まれるものが増える傾向が導かれる。

## 5. 主な発表論文等

[雑誌論文] (計 1 件)

- ① 波多野博頭, 石井カルロス寿憲, 木曾美耶子 (2014). 人関係が質問発話末の音調に与える影響 - 日本語自然対話データに基づく分析 - . 音声研究 (採択決定)

[学会発表] (計 14 件)

- ① Ishi, C. T., Ishiguro, H., and Hagita, N. (2011). "Improved acoustic characterization of breathy and whispery voices," Proc. of The 12th Annual Conference of the Intl. Speech Communication Association (Interspeech' 2011), 2965-2968.
- ② Ishi, C. T., Ishiguro, H., and Hagita, N. (2011). "Analysis of acoustic-prosodic features related to paralinguistic information carried by interjections in dialogue speech," Proc. Interspeech' 2011, 3133-3136.
- ③ Ishi, C. T., Hatano, H., Hagita, N. (2012) "Extraction of paralinguistic information carried by mono-syllabic interjections in Japanese," Proceedings of The 6th International Conference on Speech Prosody (Speech Prosody 2012), 681-684.
- ④ 新井 潤, 波多野博頭, 石井カルロス寿憲 (2012) "感動詞の音例集作成のためのラベルセットの構築", ICJLE2012 (日本語教育国際研究大会) .
- ⑤ Hatano, H., Kiso, M., and Ishi, C. (2013) "Analysis of factors involved in the choice of rising or non-rising intonation in question utterances appearing in conversational speech," Proc. 14th Annual Conference of the International Speech Communication Association (Interspeech 2013), 2564-2568, August, 2013.
- ⑥ Ishi, C., Hatano, H., and Kiso, M.

(2014). "Acoustic-prosodic and paralinguistic analyses of "uun" and "unun", " Proc. of the 7th international conference on Speech Prosody 2014, pp. 100-104, May, 2014.

- ⑦ Hatano, H., Kiso, M., and Ishi, C. (2014). "Interpersonal factors affecting tones of question-type utterances in Japanese," Proc. Speech Prosody 2014, pp. 997-1001, May, 2014.

以上査読付き国際会議 7 件, その他, 国内学会 7 件.

[図書] (計 0 件)

[その他]

感動詞の音例集

<http://www.irc.atr.jp/~xinjing/Ver1/01.html>

## 6. 研究組織

(1) 研究代表者

石井カルロス寿憲 (ISHI, Carlos Toshinori)

株式会社国際電気通信基礎技術研究所・知能ロボティクス研究所・音環境知能研究室・室長

研究者番号: 3 0 4 1 8 5 2 9

(2) 研究分担者

(3) 連携研究者