

**科学研究費助成事業 研究成果報告書**

平成 27 年 6 月 5 日現在

機関番号：14603

研究種目：基盤研究(C)

研究期間：2012～2014

課題番号：24500256

研究課題名(和文)MOSAIC：自然で双方向的な対話処理モデル

研究課題名(英文)On The Use of Non-Verbal Speech Sounds in Human Communication

## 研究代表者

CAMPBELL Nick (Campbell, Nick)

奈良先端科学技術大学院大学・情報科学研究科・教授

研究者番号：50395109

交付決定額(研究期間全体)：(直接経費) 4,100,000円

研究成果の概要(和文)：本研究は土台となる伝達意図の処理に注目し、どのように音声処理を行うかに関して取り組む。我々は発話から、状況により変わる話し手の意図の予測、あるいは解釈の支援のために、発話のノンバーバル情報の解釈を組み込んだシステムを開発することを目的としている。自然対話の内容は、発話の意味を伝達することと同時に、社会的なつながりにも大いに寄与している。このような社会的な交流を認識することには、音声情報や話し方(対話中では何を言ったかよりも重要な単位)といったものが不可欠である。我々は音声処理に、発話内容と並んで、話し方情報を含めることにより、より価値のある発話の表現と理解を実現することを目標としている。

研究成果の概要(英文)：This study focused on the processing of non-verbal information to facilitate the transfer of speaker intentions that underlie the spoken text of a social conversational interaction. The contents of natural dialogue, while conveying the semantic meaning of the utterance, also contribute greatly to social ties and informal bonding.

To be aware of such social exchanges, it is essential to process voice information and speaking style (perhaps more important than what was directly said in the conversation). We developed audio processing, alongside speech content information, that is targeted at achieving a more likely interpretation and understanding of speech-interaction.

Prediction of the intentions of the speaker varies depending on the circumstances and the history of the conversation, but reliable paralinguistic signals allow us to develop a system facilitating the interpretation of nonverbal information speech, and to moderate the words according to this extra non-verbal information.

研究分野：言語科学

キーワード：ノンバーバル 音声発話 情動情報 センシング 技術

## 1. 研究開始当初の背景

今現在、高齢化社会にむけてロボットの研究が多くなり様々な業者が対話型ロボットを研究中である。人間とコミュニケーションをするロボットは対話音声モデルを使用する。現在の音声処理技術つまり音声認識、音声合成、音声対話モデルはよく検討されているが人間センシングの部分つまり人間相手の理解、参加程度を判断する技術が必要となる。本研究では実音声対話から、語彙的情報とともに参加者の意図的対話等を示す社交的信号を処理する。

平成12年から18年、独立行政法人科学技術振興機構(JST)の助成による研究「表現豊かな発話音声のコンピュータ処理システム」において世界一音声大規模コーパスの収録を行った。このコーパスから対話インタラクション情報の分析を可能にした。本分析結果は語彙的情報のみではなく、ノンバーバルコミュニケーションの重要性、詳細な声の使い方等の利用をモデル化可能にし、発話者の対話意図・参加程度等のセンシング技術の開発を可能にした。平成16年から平成18年の総務省戦略的情報通信研究開発推進事業(SCOPE)の助成により「ヒューマンコミュニケーションの「場」が読めるロボットの研究開発(041307003)」の研究において、マルチモーダルつまり音声+ビデオを収録、音声対話コーパスを作成し、声の調子と共に微妙な動作情報を含めて分析を行った。その後平成19年から22年には科学研究費補助金研究(基盤研究(B))において、「音声コミュニケーションにおける、ノンバーバル発話の研究」(“On the Use of Non-Verbal Speech Sounds in Human Communication”)という課題名で非語彙情報(社交的面)の音声部分に着目した研究をすすめた。そして今回の「MOSAIC:自然で双方向的な対話処理モデルの研究」においてはその非語彙情報の社交的面から心理的情報に着目するに至った。この全てのコミュニケーションモデルにより、より豊かな対話処理システムが作成可能となった。

## 2. 研究の目的

本研究は発話の言語的内容のみならず、土台となる伝達意図の処理に注目し、どのように音声処理を行うかに関して取り組む。我々は発話から、状況により変わる話し手の意図の予測、あるいは解釈の支援のために、発話のノンバーバル情報の解釈を組み込んだシステムを開発することを目的としている。本研究の位置は、音声合成と音声対話インターフェースの開発。つまり従来音声合成は音声放送技術のみで相手の反応(情報の受け方)を反応する機能はなかった。本研究成果を利用すると従来、音声合成は双方向的なインタラクティブなコミュニケーション装置

となる。ロボットの対話システムはその機能を含むと、そのロボットが出した音声は理解しているか言い直しが必要かなどの判断が利用可能となる。

## 3. 研究の方法

多くの音声研究はタスクに依存するデータを利用してソフト作成やデータモデル化を行う。その結果、ビジネス用の対話は十分処理可能であるがビジネス以外の社交的人間の音声コミュニケーションの特徴データがない。その問題を克服するため、我々のデータ収集手法(タスクに依存しない方法)が心理的情報パラ言語的情報つまり人間同士の気持ちのやり取りを発達し、日常対話にも応用するデータとなる。収録結果はそのままではデータにならず、人間の判断による情報付与のうえでコーパスとなる。このコーパスは統計的手法による学習材料であり、物理的データ(収録データ)とともに各場面のラベルによるデータと理解との相関関係が可能となる。上記の統計処理による特徴のモデル化から、新たな相手の判断が可能な対話音声技術の開発を行う。

## 4. 研究成果

主な研究成果はノウハウであり、学会論文、図書などで示され、それとともに新分野を作成するに至った。それは会議など、意見交換の場面で他の研究者との交流による我々の考え方やデータの利用により広がる。もちろん、ソフト作成、対話インターフェースの開発を行うが、我々はメーカーではなくプロトタイプのみを示し評価する。

今回その研究の延長として本研究課題において、発話者・参加者の心理的情報を具体的に計るため、バイオメトリックセンサーを利用した。アイルランドのダブリン大学と共同で、“d-ans”データの収録を、シャノン・ヘニング氏と行った。これまで参加者の参加態度(心理情報等)を手動でアノテーションしデータを構築していたが本研究では、バイオメトリックセンサーにより同様な結果が得られるかどうかの確認をするためデータ収集・分析を行った。米国マサチューセッツ工科大が開発したQセンサーの技術を導入し、対話インタラクションをマルチモーダル技術に収録しながら、発話者のバイオメトリックセンシングを同時に行った。

国際シンポジウム Speech Prosody で我々が提案する social prosody (社交的韻律情報)のテーマで第7回本年ダブリンにて行った。また、国内は神戸大学で「日本語音声コミュニケーション研究のこれまでとこれから」というワークショップを開催した。本研究で作成したD-ANSコーパスはLREC 2014 Reykjavik

(Iceland)と INTER SPEECH 2014 (Singapore)で紹介され、高く評価された。本研究結果による、EU のロボットとの対話プロジェクトに参加可能となり、これらの影響は強いと期待出来る。最後に3月開催された国立情報学研究所による NII 湘南会議 “The Future of Human-Robot Spoken Dialogue: from Information Services to Virtual Assistants” (NII Shonan Meeting Seminar 059)に招待され発表を行った。

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 8 件)

Catharine Oertel, Fred Cummins, Jens Edlund, Petra Wagner, Nick Campbell  
D64: a corpus of richly recorded conversational interaction Multimodal User Interfaces 無 Volume 7 2013 p19-p28

Vaughan, B., Han, G. J., Gilmartin, E & Campbell, N  
Designing and Implementing a Platform for Collecting Multi-Modal Data of Human-Robot Interaction.  
Acta Polytechnica Hungarica 無 9 2012 p7-p17

Carlos Toshinori Ishi, Nick Campbell  
Analysis of acoustic-prosodic features of spontaneous expressive speech  
Revista de Estudos da Linguagem 無 12 2012 p38-p49

Tanaka Hiroki and Nick Campbell  
"Classification of social laughter in natural conversational speech"  
Journal of Computer Speech & Language 無 Vol 28(1) 2013 p314-325  
<http://dx.doi.org/10.1016/j.csl.2013.07.004>

De Looze, C. and Schere, S and Vaughan, B and Campbell  
Investigating automatic measurements of prosodic accommodation and its dynamics in social interaction Speech Communication 無 58 2013 p11-34  
<http://dx.doi.org/10.1016/j.specom.2013.10.002>

Catharine Oertel, Fred Cummins, Jens Edlund, Petra Wagner, Nick Campbell  
Refeed Original Articles D64: A corpus of Richly Recorded Conversational Interaction Journal on Multimodal User Interfaces 無 volume 7, issue 1-2 2013

pp19-28  
<http://dx.doi.org/10.1007/s12193-012-0108-6>

Francesca Bonin, Nick Campbell, Carl Vogel,  
Time for Laughter, Knowledge-based systems, 71, 2014, p15 - 24 2014 Notes: [<http://dx.doi.org/10.1016/j.knosys.2014.04.031>]

Céline De Looze, Stefan Scherer, Brian Vaughan, Nick Campbell,  
Investigating automatic measurements of prosodic 2014 accommodation and its dynamics in social interaction, Speech Communication, 58, (1), 2014, p11 - 34

[学会発表](計 22 件)

Oertel, C., De Looze, C., Campbell, N and Wagner, P  
Identification of cues for the automatic detection of hotspots in natural conversation  
SSPnet-COST 2012 PINK International Conference 2012 年 09 月 08 日 Budapest, Hungary

Han, J., Gilmartin, E., De Looze, C., Vaughan, B., & Nick Campbell  
The Herme Database of Spontaneous Multimodal Human-Robot Dialogues  
Conference on Language Resources and Evaluation (LREC'12) 2012 年 05 月 24 日 Istanbul, Turkey

Stefan Scherer, Georg Layher, John Kane, Heiko Neumann and Nick Campbell  
An audiovisual political speech analysis incorporating eye-tracking and perception data  
Conference on Language Resources and Evaluation (LREC'12)  
2012 年 05 月 24 日 Istanbul, Turkey

Bonin, Francesca and Nick Campbell and Carl Vogel  
Laughter and Topic Changes: Temporal Distribution and Information Flow  
3rd IEEE Conference on Cognitive Infocommunications 2012 年 12 月 02 日 Kosice, Slovakia

De Looze, C., Scherer, S., Vaughan, B. & Campbell, N.  
Vers une mesure automatique de l'adaptation prosodique en interaction conversationnelle

29eme Journees d'Etude sur la Parole (JEP 2012) 2012年06月05日 Grenoble, France

Gilmartin, E., Campbell, N  
More than just words: Building a chatty robot  
International Workshop on Spoken Dialog Systems (IWSDS 2012) 2012年11月30日 Paris, France

Francesca Bonin, Nick Campbell, Carl Vogel  
Temporal distribution of laughter in conversation  
Third Interdisciplinary Workshop on Laughter and other Non Verbal Vocalisations in Speech 2012年10月27日 Dublin, Ireland

JingGuang Han, Nick Campbell, Kristiina Jokinen, Graham Wilcock  
Investigating the use of Non-verbal Cues in Human-Robot Interaction with a Nao robot  
IEEE 3rd International Conference on Cognitive Infocommunications (CogInfoCom) 2012年12月05日 Kosice, Slovakia

Hiroki Tanaka, Sakriani Sakti, Graham Neubig, Tomoki Toda, Nick Campbell, Satoshi Nakamura  
Non-verbal cognitive skills and autistic conditions: An analysis and training tool  
IEEE 3rd International Conference on Cognitive Infocommunications (CogInfoCom) 2012年12月02日 Kosice, Slovakia

Francesca Bonin, Ronald Bock, Nick Campbell  
How Do We React to Context? Annotation of Individual and Group Engagement in a Video Corpus, Privacy, Security, Risk and Trust (PASSAT)  
International Conference on and 2012 International Conference on Social Computing (SocialCom) 2012年09月04日 Amsterdam, Netherland

Emer Gilmartin, Celine De Looze, Nick Campbell  
Priming, Timing, and the Phatic Component in Machine-Mediated Dialogue  
The Listening Talker (LISTA 2012) 2012年05月03日 Edinburgh, Scotland

N. Campbell  
Social aspects and speech ability in CogInfoCom systems  
CogInfoCom 2012(招待講演) 2012年12月02日

Kosice, Slovakia

Boeck, Bonin, Campbell, Edlund, De Kok, Poppe, Traum  
"Laughter and Topic Transition in Multiparty Conversation"  
Proceeding of the SIGNAL 2013 Conference 2013年8月22日~2013年8月24日 Metz, France

Francesca Bonin, Celine De Looze, Sucheta Ghosh, Emer Gilmartin, Carl Vogel, Anna Polychroniou, Hugues Salamin, Alessandro Vinciarelli and Nick Campbell  
"Investigating fine temporal dynamics of prosodic and lexical accommodation!"  
14th Annual Conference of the International Speech Communication Association, INTERSPEECH, 2013 2013年8月25日~2013年8月29日 Lyon, France

Emer Gilmartin, Francesca Bonin, Carl Vogel and Nick Campbell  
"Exploring the Role of Laughter in Multiparty Conversation"  
Proceedings(SemDial), semdial 2013, DialDam 2013年12月18日~2013年12月20日 Amsterdam, Netherlands

JinGuan Han, Gilmartin, E., Campbell, N  
"Herme, yet Another Interactive Conversational Robot"  
Affective Computing and Intelligent Interaction (ACII), 2013 2013年9月2日~2013年9月5日 Geneva, Switzerland

Shannon Hennig, Ryad Chellali, Nick Campbell,  
The D-ANS Corpus: the Dublin Autonomous Nervous System Corpus of 2014 Biosignal and Multimodal Recordings of Conversational Speech,  
LREC 2014, Reykjavik, May 28-30 2014, edited by ELRA , 2014

Shannon Hennig, Ryad Chellali and Nick Campbell,  
The D-ANS corpus: the Dublin-Autonomous Nervous System corpus of 2014 biosignal and multimodal recordings of conversational speech,  
The 9th edition of the Language Resources and Evaluation Conference, Reykjavik, Iceland, 26-31 May, ELRA, 2014

invited  
Nick Campbell, Beyond D64 - A fresh look at multimodal dialogue systems from the FastNet point of view - how we are 2014 incorporating different modalities into an interactive speech synthesis system., Multimodality-interdisciplinary perspectives: new questions and new methodologies, in Paris, at INALCO, June the 6th., 2014, UMR SFL – Université Paris 8 et CNR

invited  
Nick Campbell,  
An introduction to the TCD D\_ANS Corpus - a multimodal multimedia biometric corpus of spoken social 2014 interaction, MA3HMI, Singapore, Sept 14th, 2014, Interspeech

21) Francesca Bonin, Carl Vogel and Nick Campbell,  
Social sequence analysis: temporal sequences in interactional conversations , 2014 5th IEEE Conference on Cognitive Infocommunications (CogInfoCom), 5-7 Nov. 2014 ,  
IEEE, 2014, pp403-406  
Notes: [ISBN 978-1-4799-7280-7/14]  
DOI

22) Ya Li, Nick Campbell, JianHua Tao  
Voice Quality: not only about "you" but also about "your interlocutor",  
ICASSP, Brisbane, Australia  
2015 edited by IEEE , 2015

〔図書〕(計 9 件)

D. Gibbon, D. Hirst, and N. Campbell  
Polish Phonetic Association, Poznan  
Rhythm, Melody, and Harmony in Speech  
2012  
p325

Matej Rojc & Nick Campbell (eds)  
Science Publishers, Jersey  
Speech, Gaze, and Affect: concepts of reactive and natural human-machine interaction techniques employing ECAs with personality  
2013  
p500

A., Vinciarelli, A., Hoffmann, R., V.  
(Eds.) [Altmann, U., Oertel, C., & Campbell, N.]  
Springer

Conversational Involvement and Synchronous Nonverbal Behaviour in, Esposito, "Cognitive Behavioural Systems" 2012  
pp343 - 352

Matej Rojc & Nick Campbell  
CRC Press  
Speech Technology and Conversational Activity in Human-Machine Interaction  
2013  
434

Emer Gilmartin, Nick Campbell  
Springer  
More Than Just Words: Building a Chatty Robot / Natural Interaction with Robots, Knowbots and Smartphones.

Matej Rojc, Nick Campbell,  
Converbal Synchrony in Human-Machine Interaction,  
Boca Raton, CRC Press,  
2014, 1 - 434pp 2014

[Emer Gilmartin, Nick Campbell]  
More Than Just Words: Building a Chatty Robot in, Joseph Mariani, Sophie Rosset, Martine Garnier-Rizet, Laurence Devillers ,  
2014 Natural Interaction with Robots, Knowbots and Smartphones,  
New York,  
Springer, 2014,  
pp179 - 185,

Nick Campbell and Shannon Hennig  
Annotating the TCD D-ANS Corpus - a multimodal multimedia monolingual biometric corpus of spoken social interaction in, 2015  
editor(s) Ronald Boeck and Francesca Bonin ,  
Multimodal Analyses Enabling Artificial Agents in Human-Machine Interaction", Heidelberg,  
Springer Verlag, 2015, []

[Nick Campbell and Ya Li]  
Expressivity in Interactive Speech Synthesis; Some Paralinguistic and Nonlinguistic Issues of Speech Prosody for 2015 Conversational Dialogue Systems in, editor(s) K. Hirose, J. Tao ,  
Speech Prosody in Speech Synthesis: Modeling and generation of prosody for high quality and flexible speech synthesis, 2015, pp97 - 107,

〔産業財産権〕  
出願状況（計 件）

名称：  
発明者：  
権利者：  
種類：  
番号：  
出願年月日：  
国内外の別：

取得状況（計 件）

名称：  
発明者：  
権利者：  
種類：  
番号：  
出願年月日：  
取得年月日：  
国内外の別：

〔その他〕  
ホームページ等

#### 6. 研究組織

##### (1) 研究代表者

ニック・キャンベル (Nick CAMPBELL) 奈良  
先端科学技術大学院大学 情報科学研究科  
客員教授

研究者番号：50395109

##### (2) 研究分担者

定延利之 (Sadanobu Toshiyuki) 神戸大学  
国際文化学部 教授

研究者番号：10395022

##### (3) 連携研究者

( )

研究者番号：