

感情認識及び感情創生に基づく知的学内案内ロボットの構築

任福継^{1*}, 鈴木基之¹

Development of Campus Navigation Robot Based on Affective Computing

by

Fuji Ren, Motoyuki Suzuki

Recent years, a huge amount of information is available through the internet, and many information retrievers have been developed. However, these retrievers only show retrieved results without hearty communication.

In this paper, an intelligent campus navigation robot is developed. It recognizes a user's utterance using a speech recognizer, and retrieves information from a knowledge database. Finally, the agent makes an appropriate answer from retrieved results, and give it to the user. In order to communicate with a user warmly, the agent also recognizes user's emotion from a voice and a facial expression, and the agent represents its own emotion using voice and behaviour.

Keywords: Intelligent campus guide system, Mental state transition network, Affective computing.

1 はじめに

インターネットの普及により誰もが容易に莫大な情報に触れる機会が拡大している。しかし、情報検索システムによる一辺倒な情報提供、あたかも理解したかのように感じてしまうカット＆ペースト的知識の収集が空虚な知識体系の構築を助長している。また、デバイスの進歩によりハード面から人間をサポートできる機械やロボットが開発されているが、その知能部分の開発の遅れが問題視されている。未だ、人と共に生活できるような人間身溢れるものには達していないのが現状である。

一方、大学の現状を考えると、業務の効率化や予算削減などの観点から、大学で働く職員の数も年々減少傾向にある。その反面、大学見学や地域への大学の開放などこれまで以上に大学に求められている役割は大きくなってきている。そのような状況下において、これまで通りに大学職員が見学者や地域住民に付きっ

りて丁寧に対応することはもはや不可能な状況であるといえる。

そこで本研究では、人の気持ちを理解し、円滑なコミュニケーションを行うことができる知的学内案内ロボットの構築を行う。学内案内システムを含むロボットとの対話は、一般的に無機質なものと思われがちであるが、人が行う受付業務がそうではないように「感情」を積極的に扱うことで、気の利いた、心が通じ合う暖かい対応を実現し、老若男女に自然に受け入れられるロボットの構築を目指す。

具体的には、自然言語処理、音声認識処理、顔表情認識処理などの基礎的な技術と、世界で始めて提案した心的状態遷移ネットワークという人間の心をモデル化する技術を統合することで、人の感情認識ならびに機械の感情創生を実現する手法について研究を行った。この成果を用いることで、人の気持ちを理解し、円滑なコミュニケーションを行うことができる知的学内案内ロボットの構築が可能である。

¹徳島大学大学院ソシオテクノサイエンス研究部
Institute of Technology and Science, The University of Tokushima

*連絡先：〒 770-8506 徳島市南常三島町 2-1
徳島大学大学院ソシオテクノサイエンス研究部

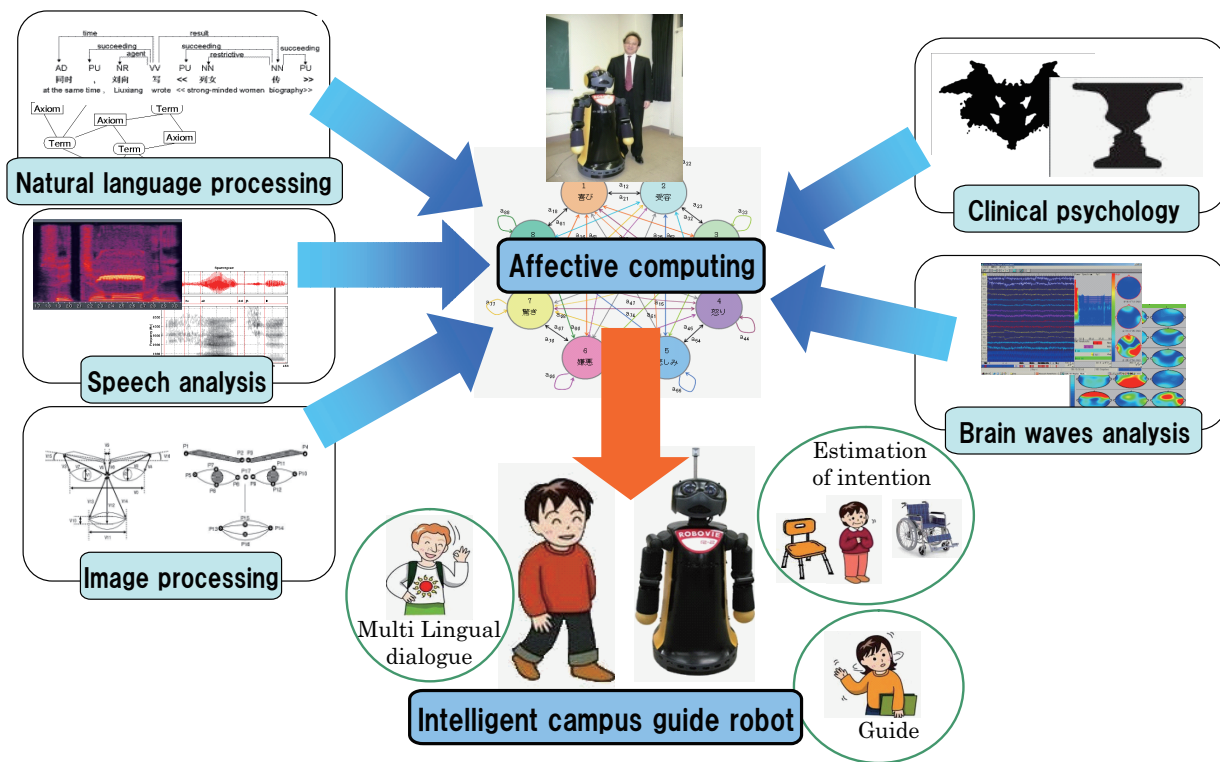


Fig. 1: Overview of the intelligent campus guide system.

2 知的学内案内ロボットの開発

2.1 開発の概要

本研究で開発する知的学内案内ロボットの開発概要イメージを Fig. 1 に示す。

我々の研究室では、言語処理技術、音声情報処理技術、画像解析技術という従来の工学的研究分野の他に、心理学的分野の研究や医学の分野である脳波分析技術など非常に幅広い研究を手がけ、それぞれにおいて研究成果をあげている。

言語処理技術では、独自に開発した Super Function 技術による高精度の機械翻訳 [1-3, 27] や人間が普段用いている自然な言語表現の意味を理解する技術を開発している [6, 9, 18, 19, 23, 24, 29-31]。それらの応用として、文章の書き手の感情を判断する技術の開発に成功している。

音声情報処理技術では、音声認識技術やその技術を応用した音声によるセキュリティシステムの開発を行っている [7, 8, 11, 12, 20, 22, 26, 32]。また、画像処理技術では、人間の自然な顔表情からその人の感情を認

識する技術を開発している [14, 21, 25]。これらの技術は、知的なロボットの構築に必要不可欠な技術である。

その他、心理学や脳波に関する知見は、人間とコミュニケーションを行うロボットを開発するために有効に活用できると考えられる [4, 5, 15, 17, 28]。

これらの技術を要素技術とし、100 名以上の被験者からのデータを基に人間の心の状態をモデル化した心的状態遷移ネットワークを組み合わせることで、人間の気持ちを理解できる機能を実現している [10, 13]。これにより、これまで多数開発されている案内ロボットでは実現できなかった円滑なコミュニケーションを行うことができる知的なロボットの構築に成功した。

2.2 動作イメージ

Fig. 2 に実際に構築した知的学内案内ロボットとの会話イメージを示す。Fig. 2 の上の図では、研究室の場所を聞いた人に対して、「C 棟の 2 階」という単なる場所だけではなく、その場所がどこにあるのかを丁寧に回答している。通常の対話システムにおいては、

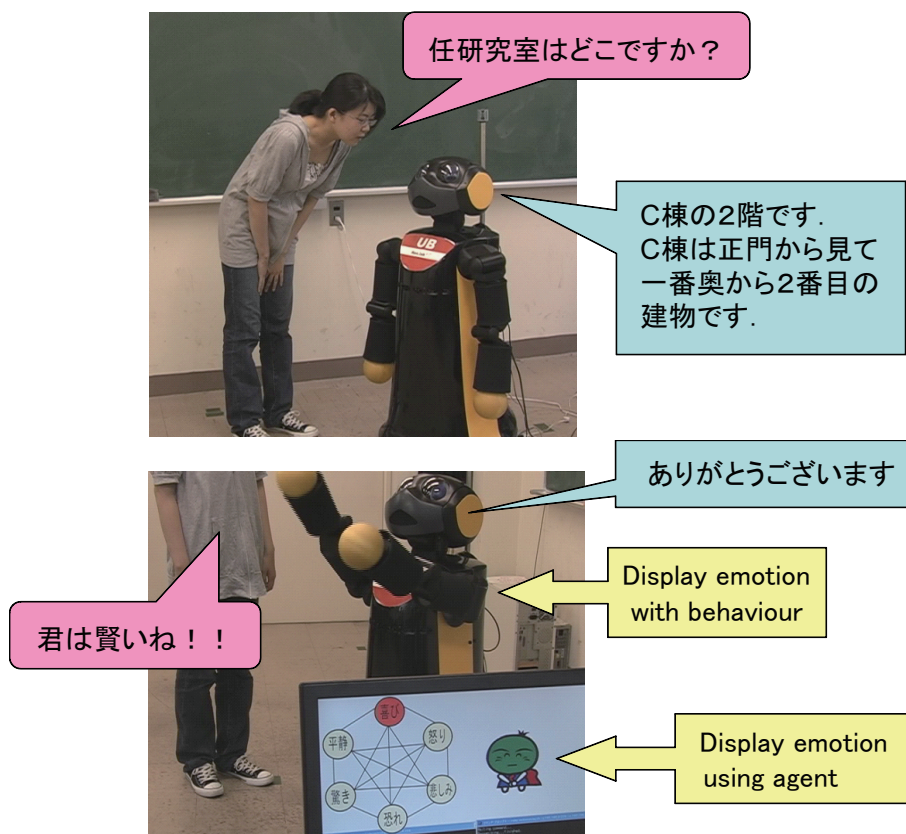


Fig. 2: Example of a dialogue with the root.

聞かれた事に対して直接的な回答のみを行うことがほとんどであるが、それでは不十分であり、「不親切」な回答であると受けとられる場合がある。この例の場合は、研究室の場所を聞く人は、そもそもその場所がどこにあるのかも知らないだろう、という推測のもと、追加の情報を自ら提供している。

Fig. 2 の下の図では、人の「ありがとう」という感謝の言葉に対して、嬉しいことをジェスチャーを交えて表現している。このように、ロボットも人間と同じように心、感情を持つことで、ユーザは自然にロボットに対して親近感を抱き、抵抗感を低減させる効果が期待できる。また、このように行動、回答することにより、ロボットが「心」を持っていることを人間に伝えることもできると考えられる。

さらに、Fig. 2 の下の図には、4 章で説明するエージェントによりロボットの感情が表現されている。これは、例えば Web の HP 上での学内案内など、実体のロボットが使用できない環境を想定し、デジタルの

世界でもロボットとの対話と同じような効果を得るために開発したものである。

3 知的学内案内ロボットの構成

Fig. 3 に本研究で構築した知的学内案内ロボットの構成を示す。

ユーザは知的学内案内ロボットと音声を通じて会話を行うことができる。入力された音声は、音声認識されテキスト情報に変換される。そのテキスト情報を解析し、学内案内に関する話をしている場合は、EDB や大学の HP から構築した学内案内用のデータベースを検索し情報を抽出する。一方、学内案内に関する話をしていない場合には、その他の雑談をしていると判断し、雑談用のデータベースを利用して雑談に必要な知識を抽出する。抽出された情報は合成音声によりユーザに伝えられる。

この際、ユーザが入力した発話内容と音声の特徴を

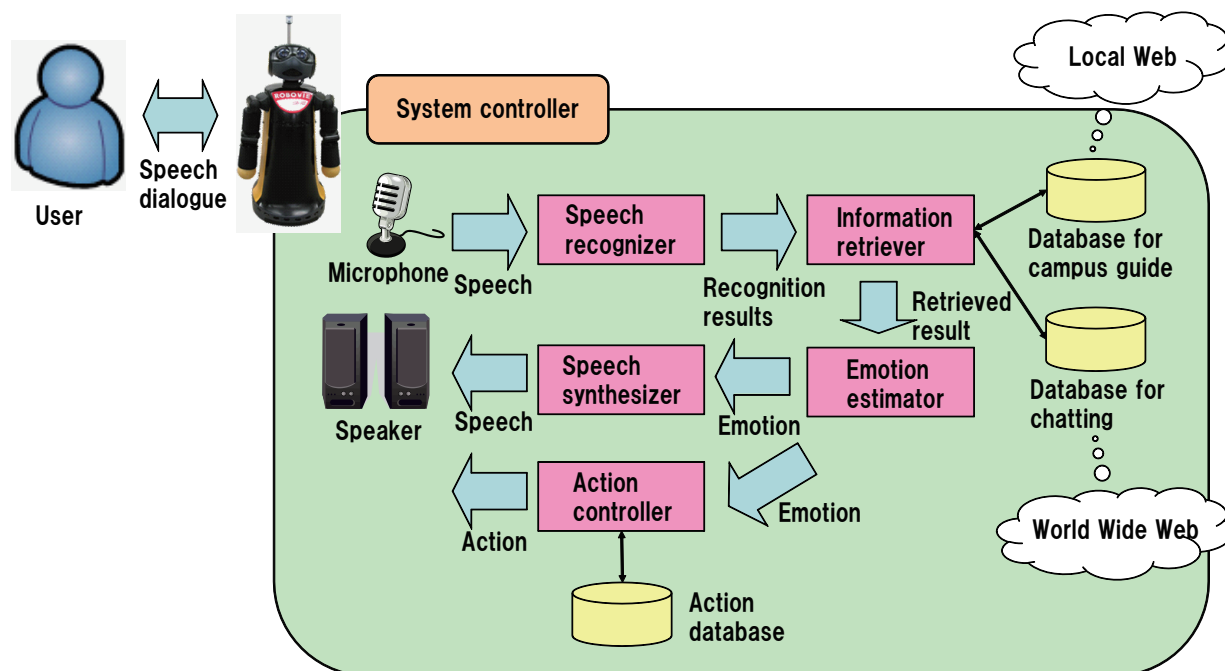


Fig. 3: Block diagram of the intelligent campus guide system.

解析することで、ユーザが抱いている感情、気持ちを推定し、ユーザの状況に適する音声で返答を行う。また、ユーザが抱いている感情、気持ちに対してロボットも感情を生起し、生起した感情をジェスチャー、行動としてアウトプットする。例えば、ユーザが怒っているときには、怒りを静めるように寂しそうな音声でやさしく返答すると共に、悲しいことを体の動きで表現する。

本研究で構築した知的学内案内ロボットで重要である技術について以降で詳細に説明する。

4 心的状態遷移ネットワークとエージェント

人の気持ちを理解し、円滑なコミュニケーションを行うためには、人間の抱いている感情を推定する能力と共に、自分も相手と同じように感情を持つことが必要不可欠であると考えられる。人間の抱く感情に関する研究は多数行われており、数多くの見解が示されているが、定説は存在しない [33-37]。そこで本研究では、Ekman が提唱 [38] した人間が顔表情で表現できる「喜び」「怒り」「悲しみ」「恐れ」「驚き」「平静」の6つを人間が抱く感情と定義し、これらのうちの

感情を抱いているかを推定できるモデルを提案した。

実際には、100名以上の被験者に対して大規模なアンケート調査を実施し、その結果を解析することで人間の心の動きをモデル化することに成功した [10, 13]。本研究では、この独自に開発した心のモデルを心的状態遷移ネットワークと呼んでいる。

Fig. 4 にイメージ図を示す。前述した6つの感情からはそれぞれ別の5つの感情に遷移することができる。アンケート解析からこの遷移にかかるコストを算出することで、人間のように感情豊かな対応ができるようになっている。

また、Fig. 4 には、6つの感情に対応するエージェントを作成している。これは、3章で述べたように、ロボットがいない環境下において使用するデジタルエージェントである。今回は、徳島県のマスコットである「すだちくん」をモチーフに作成した*。

5 音声の誤認識に頑健な情報検索手法

3章で説明したように、本研究では、ユーザは音声によりロボットと会話を行う。しかし、ロボット自身

*本研究を本学主催のエンジニアリングフェスティバルで展示した際には、県職員の方から好評でした。

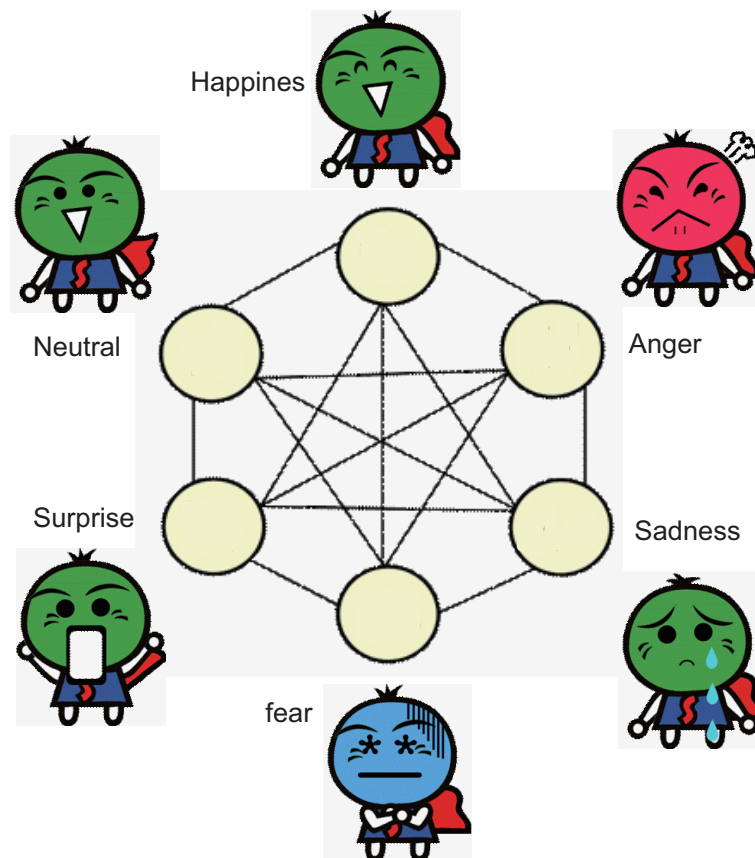


Fig. 4: Mental state transition network and agent “Sudachi”.

が発生するノイズや環境音の影響を受けて、人間の音声を正しく聞き、解析することは非常に困難である。実際、オフィスのような比較的静かな環境下においても音声認識ソフトが誤認識することからもその難しさは理解できる。

そこで本研究では、音声認識誤りが起こることを想定し、多少誤った情報が入力された場合にも適宜解釈し、誤りを補完できる情報検索手法を提案した。Fig. 5 に提案手法の概要を示す。

ユーザから入力された音声は音声認識技術により、テキスト情報に変換される。この際に前述した原因により誤った情報に変換されることがあるため、もっともらしい認識結果を複数出力し、統計処理を行うことで誤認識を訂正する。具体的には、もっともらしい認識結果を複数出力し、出力結果に含まれる単語の認識結果としての信頼度 [42, 43] を算出する。算出された単語信頼度を考慮した上で、データベース中の想定質問文との類似度を計算する。これにより、誤りを含む

認識結果から、高精度な音声検索を実現している。

この手法は、実際人間が行っている行動と同じであるといえる。例えば、うるさい環境で人と会話を行う場合、我々は聞き取れなかった部分を相手に聞き返したり、また、聞き取れなかった部分を前後関係から推測して相手の話している内容を復元しようとする。本研究で提案する手法では、もっともらしい認識結果を複数出力するが、これは聞き間違いそうな単語、つまり発音が似ている単語を複数出力している。そして、その出力に比較的是っきりと聞こえた単語（信頼度の高い単語）に重点を置いて、聞き取れなかった部分とあわせて検索を行う。つまり人間で例えると、何度も聞き返し、その中で多く聞き取ることができた単語に重点を置いて、相手に返答をするのと同じ効果を得られると考えられる。

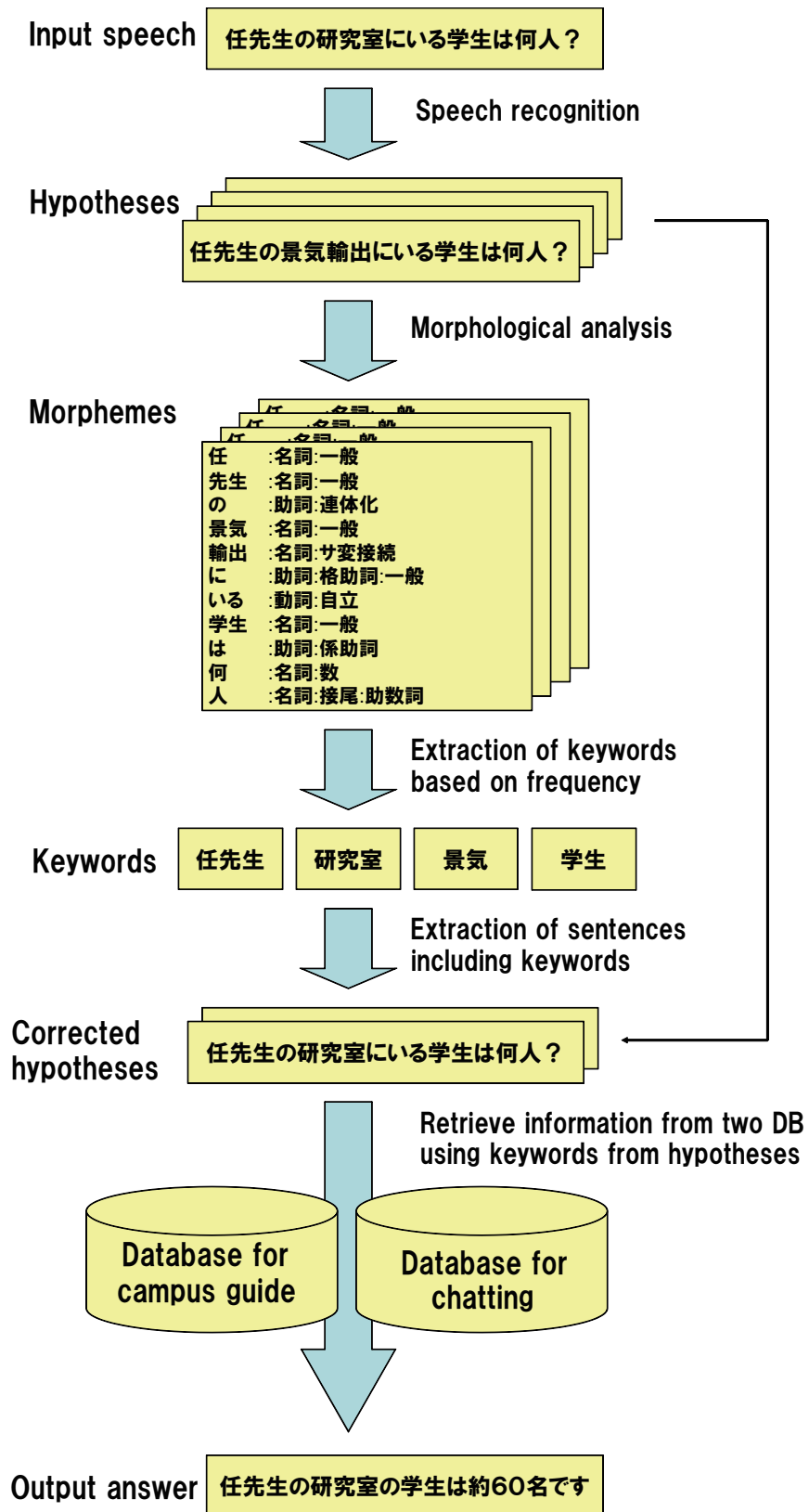


Fig. 5: Flowchart of the information retrieval system dealing with a misrecognition of a speech.

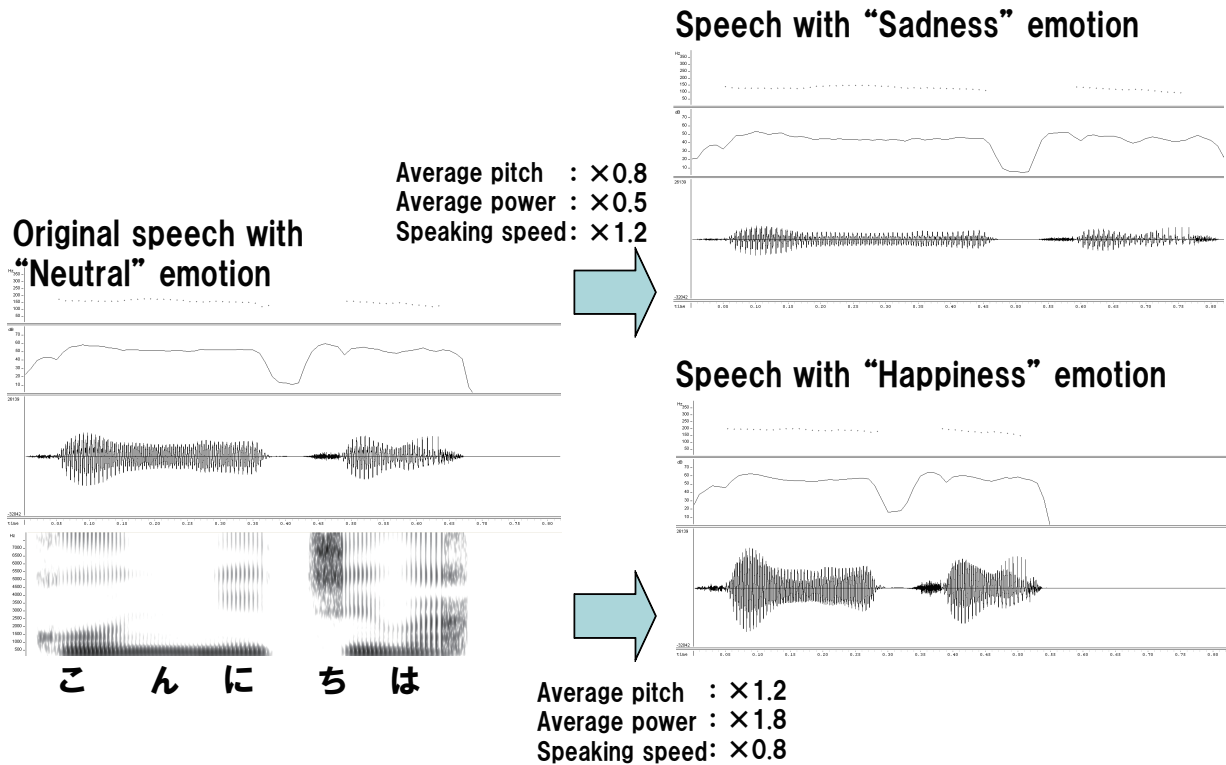


Fig. 6: Example of the emotional speech synthesis.

6 感情を表現するための音声合成手法

コミュニケーションを行うためには、相手に情報を伝える必要がある。本研究においてもこの情報を伝える技術は非常に重要である。相手が伝えたい情報を得る技術は5章ですでに述べており、本章では、逆にロボットからユーザに情報を伝えるための技術について説明する。本研究では、ユーザに情報を伝える手段として音声を用いた。機械が音声を発生するためには、人間のような声帯を持っていないことから、デジタル的に音声波形を作成し人間のような声を生成する必要がある。このような技術を音声合成と呼ぶ。

音声合成の身近な例としては、カーナビゲーションシステムの音声案内が挙げられる。しかし、現存するシステムで生成される音声はまだ機械的であり、いわゆる「ワレワレハ、ウチュウジンダ」を想像させるような人間にとって不自然な音声であることは否定できない。特に人間は抑揚に敏感であり、平坦な音声を聞くとどうしても機械的な音声だと認識してしまう傾向にある。

そこで本研究では、人間との円滑なコミュニケーショ

ンができるロボットの開発を目指しているため、人間のように抑揚のある音声を生成する手法を提案した。具体的には、複数の被験者の発話音声を収録し、クラスタリング手法を用いることで、それぞれの感情のときの音声の特徴を解析した [16, 41]。なお、本研究で使用した音声特徴は「基本周波数」、「音量」、「発話速度」の3種類である。結果のイメージを Fig. 6 に示す。例えば、感情が「平静」のときの音声特徴を基準にした場合、「悲しい」感情を表現するためには「基本周波数」を0.8倍、「音量」を0.5倍、「発話速度」を1.2倍にする、また、「嬉しい」感情を表現するためには、「基本周波数」を1.2倍、「音量」を1.8倍、「発話速度」を0.8倍にするというルールを獲得することができた。

このように音声に抑揚をつけることができると、例えば「ありがとう」という言葉であっても、感謝の意を込めた「ありがとう」や皮肉を込めた「ありがとう」など表現豊かにコミュニケーションを行うことが可能になる。

7 今後の展開

7.1 学内案内ロボットの今後

本研究で構築した学内案内ロボットはプロトタイプであり、実用化に際しては、まだまだ解決すべき課題は多い。高精度な音声認識、より高度で複雑な感情まで対象とした感情認識と感情創生、より自然で表現力豊かな音声合成や動作の設計など、個々の要素技術について更なる研究開発が必要である。

また、学内案内ロボットとしても更なる発展を遂げる必要がある。具体的には、音声の自動翻訳技術を組み合わせた多国語化を検討している。近年は様々な分野での国際化が必須となっており、大学においても多数の留学生や外国からの訪問者が日常的に在籍している。こうした人達に対応するためには、多国語を理解し、また話すことができる案内ロボットになることが望ましい。

こうしたニーズに対し、個々の言語ごとにシステムを開発するのではなく、入出力部に自動機械翻訳技術を導入することで、効率的な多言語化を目指す。単純に機械翻訳技術を導入しただけでは、その品質の低さから実用に耐えないものとなることは容易に想像される。そのため、機械翻訳技術の精度向上、翻訳誤りに頑健な意味解析や情報検索法の開発などを行う必要があると思われる。

7.2 開発された技術の応用

本研究では、感情認識と感情創生に注目し、人の気持ちを理解し、円滑なコミュニケーションを行うことができる知的学内案内ロボットの構築した。本研究で用いた技術は多岐に渡っている。これらの技術は知的学内案内ロボットにのみ適用できる限定的なものではなく、様々な応用が期待できる。特に人間が直接関わるといったシステムへの応用に有効に機能すると考えられる。

まずは、カーナビゲーションシステムや銀行のATMに代表される機器のインタフェース部に応用ができると考えられる。カーナビゲーションシステムや銀行のATMは非常に便利なものであり、一度慣れてしまうと必要不可欠なものとして扱われてる。しかし、この「慣れる」という行為が高齢者や機械に抵抗感のある人にとっては、非常に敷居の高いものである。そこで、

本研究で開発した技術を応用することにより、老若男女に自然に受け入れられるシステムを構築できると考えられる [39,40]。

また、医療や介護の分野では、現在人手不足が深刻化している。将来、人間に代わりロボットが介護や看護をしてくれる時代が来るかもしれない。その際には、本研究で開発した技術は非常に有効に機能し、そのような世の中の実現に寄与できることを期待している。

8 まとめ

本研究では、人の気持ちを理解し、円滑なコミュニケーションを行うことができる知的学内案内ロボットの構築を行った。具体的には、自然言語処理、音声認識処理、顔表情認識処理、脳波解析処理などの基礎的な技術と、世界で始めて提案した心的状態遷移ネットワークという人間の心をモデル化する技術を統合することで、人の感情認識ならびに機械の感情創生を実現する手法について研究を行った。

本研究で開発した技術は、カーナビゲーションシステムや介護ロボットなど様々な分野、システムに応用されることが期待できる。ロボットが一般家庭にも普及し、ロボットと人間が共存する次の時代の到来に寄与できることを期待している。

参考文献

- [1] 篠山, 黒岩, 任: Super-Function に基づく日英機械翻訳における日付・時間表現の抽出, 電気学会論文誌 C, 128(8), 1342-1350 (2008).
- [2] M. Sasayama, F. Ren and S. Kuroiwa: Automatic Super-function Extraction for Translation of Spoken Dialogue, International Journal of Innovative Computing, Information and Control, 4(6), 1371-1382 (2008).
- [3] L. Mi, X. Luo and F. Ren: Chinese-Japanese Translation of Causative Sentences Using Super-function Based Machine Translation System, International Journal of Innovative Computing, Information and Control, 4(4), 915-926 (2008).

- [4] K. Osaka, S. Tsuchiya, F. Ren, S. Kuroiwa, T. Tanioka and L. C. Rozzano: The Technique of Emotion Recognition Based on Electroencephalogram, *Information*, 11(1), 55–68 (2008).
- [5] K. Osaka, T. Tanioka, S. Ueno, C. Kawanishi, T. Tada, S. Kuroiwa and F. Ren: Empathetic Understanding as Caring in Nursing Using Electroencephalographic Data as Evidence, *International Journal for Human Caring*, 12(1), 7–16 (2008).
- [6] K. Matsumoto, F. Ren, S. Kuroiwa and S. Tsuchiya: Emotion Estimation Algorithm Based on Interpersonal Emotion Included in Emotional Dialogue Sentences, *Lecture Notes in Artificial Intelligence*, 4827, 1035–1045 (2007).
- [7] S. Kuroiwa, S. Tsuge, M. Kita and F. Ren: Speaker Identification Method Using Earth Mover’s Distance for CCC Speaker Recognition Evaluation 2006, *International Journal of Computational Linguistics & Chinese Language Processing*, 12(3), 239–254 (2007).
- [8] 原, 黒岩, 田中, 柘植, 任, 獅々堀, 北: 符号化音声認識のための合成音声を用いた不特定話者音響モデルの適応法, *電子情報通信学会論文誌 D*, J90-D(9), 2541–2549 (2007).
- [9] 松本, 三品, 任, 黒岩: 感情生起事象文型パターンに基づいた会話文からの感情推定手法, *自然言語処理*, 14(3), 239–271 (2007).
- [10] H. Xiang, P. Jiang, S. Xiao, F. Ren and S. Kuroiwa: A Model of Mental State Transition Network, *IEEJ Trans. EIS*, 127(3), 434–442 (2007).
- [11] 柘植, 黒岩, 獅々堀, 任, 北: 分散音声認識における実時間周波数特性正規化手法, *情報処理学会論文誌*, 48(2), 900–908 (2007).
- [12] 黒岩, 森, 高階, 柘植, 任: スペクトルの微細構造を考慮した風雑音除去手法, *電子情報通信学会論文誌 A*, J90-A(1), 1–12 (2007).
- [13] H. Xiang, S. Xiao, F. Ren and S. Kuroiwa: A Mind Model for an Affective Computer, *International Journal of Computer Science and Network Security*, 6(6), 62–69 (2006).
- [14] J. Ma, M. Suzuki and F. Ren: Speaker Detection Method for Autonomous Robot in Complex Communication Environment Based on Image Processing, *Proc. ICAI 2008*, 349–354 (2008).
- [15] S. Tsuchiya, K. Osaka and F. Ren: A Method of Emotion Judgment in Empathetic Understanding, *Proc. ICAI 2008*, 130–133 (2008).
- [16] M. Adachi, S. Tsuchiya and F. Ren: Emotion Inference Method Based on Word’s Meaning and Utterance Features, *Proc. ICAI 2008*, 138–141 (2008).
- [17] K. Osaka, S. Tsuchiya, F. Ren and T. Tanioka: A Analysis of Empathetic Understanding Using Relation Between Electrode and Frequency, *Proc. ICAI 2008*, 116–121 (2008).
- [18] K. Mishina, S. Tsuchiya and F. Ren: Comparison Between the Human Emotion Transfer Ratio and the Similarities of Emotion, *Proc. ICAI 2008*, 126–129 (2008).
- [19] A. Hakamata, F. Ren and S. Tsuchiya: Human Emotion Model based on Discourse Sentence for Expression Generation of Conversation Agent, *Proc. IEEE NLP-KE 2008*, 235–242 (2008).
- [20] K. Obayashi, S. Tsuge, M. Fukumi, S. Tsuchiya, R. Sumitomo, F. Ren and S. Kuroiwa: A Study of speaker identification using phoneme-information, *Proc. NCSP’08*, 164–167 (2008).
- [21] P. Jiang, J. Ma, Y. Minamoto, S. Tsuchiya, R. Sumitomo and F. Ren: Orient video database for facial expression analysis, *Proc. 10th IASTED International Conference Intelligent Systems and Control*, 211–214, (2007).
- [22] T. Hirai, S. Tsuge, S. Kuroiwa and F. Ren: A speech emphasis method for noise-robust speech

- recognition by using repetitive phrase, Proc. IEEE NLP-KE 2007, 140–144 (2007).
- [23] K. Hisazumi, S. Tsuchiya, S. Kuroiwa and F. Ren: Extraction of the term which has an explicit relation using Coincidence Frequency, Proc. IEEE NLP-KE 2007, 178–183 (2007).
- [24] K. Mishina, S. Tsuchiya, S. Kuroiwa and F. Ren: An Emotion Similarity Calculation Using N-gram Frequency, Proc. IEEE NLP-KE 2007, 160–165 (2007).
- [25] J. Ma, S. Tsuchiya, S. Kuroiwa, F. Ren and Y. Lei: The New Image Processing Method in Expression Recognition System, Proc. IEEE NLP-KE 2007, 134–139 (2007).
- [26] S. Tsuge, K. Seida, M. Shishibori, K. Kita, F. Ren, M. Fukumi and S. Kuroiwa: Analysis of Variation on Intra-Speakers Speech Recognition Performances, Proc. IEEE NLP-KE 2007, 387–392 (2007).
- [27] M. Sasayama, F. Ren and S. Kuroiwa: Automatic Super-Function Extraction for Translation of Spoken Dialogue, Proc. IEEE NLP-KE 2007, 141–148 (2007).
- [28] K. Osaka, T. Tanioka, L. C. Rozzano, S. Ueno, K. Matsumoto, C. Kawanishi, S. Kuroiwa, S. Tsuchiya and F. Ren: Electroencephalograph Estimation Method of Measuring “Empathic Understanding”, Proc. IEEE NLP-KE 2007, 514–519 (2007).
- [29] T. Yamada, S. Tsuchiya, S. Kuroiwa and F. Ren: Classification of Facemarks Using N-gram, Proc. IEEE NLP-KE 2007, 322–327 (2007).
- [30] M. Shinomiya, F. Ren, S. Kuroiwa and S. Tsuchiya: Extracting the Opinions of News Articles based on Emotionally Laden Words, Proc. IEEE NLP-KE 2007, 262–267 (2007).
- [31] S. Tsuchiya, F. Ren, S. Kuroiwa, H. Watabe and T. Kawaoka: A Semantic Information Retrieval Technique and an Evaluation for a Narrow Display Based on a Association Mechanism, Proc. IEEE NLP-KE 2007, 209–214 (2007).
- [32] S. Kuroiwa, M. Takashina, S. Tsuge and F. Ren: Artificial Bandwidth Extension for Speech Signals using Speech Recognition, Proc. INTER-SPEECH 2007, 2501–2504 (2007).
- [33] 齋藤勇:感情と人間関係の心理, 川島書店 (1986).
- [34] 福井康之:感情の心理学, 川島書店 (1990).
- [35] 九鬼周造:「いき」の構造, 岩波書店 (2001).
- [36] リタ・カーター:脳と心の地形図, 原書房 (1999).
- [37] スーザン・グリーンフィールド:脳の探求, 無名舎 (2001).
- [38] P. Ekman: Expression and the nature of emotion, Approaches to Emotion, Erlbaum, 319–344 (1984).
- [39] F. Ren: Affective Information Processing and Recognizing Human Emotion, Electronic Notes in Theoretical Computer Science, 225(2009), 39–50 (2009).
- [40] F. A. Mohamed and F. Ren: GA, MR, FFNN, PNN & GMM based Models for Automatic Text Summarization, Computer Speech and Language, 23(1), 126–144 (2009).
- [41] Atsushi Sasaki, Masashi Adachi, Motoyuki Suzuki and Fuji Ren : Influence on Emotional Impression of Voice by Changing Prosodic Features, Proc. IEEE NLP-KE, pp.445–451 (2009)
- [42] 李晃伸, 河原達也, 鹿野清宏: 2パス探索アルゴリズムにおける高速な単語事後確率に基づく信頼度算出法, 情報処理学会研究報告, 2003-SLP-49-48 (2003)
- [43] Akinobu Lee, Kiyohiso Shikano, and Tatsuya Kawahara: Real-time word confidence scoring using local posterior probabilities on tree trellis search, Proc. ICASSP2004, Vol.I, pp.793–796 (2004)