

「高度メディア社会の生活情報技術」

平成11年度採択研究代表者

ニック キャンベル

(㈱国際電気通信基礎技術研究所 人間情報科学研究所 プロジェクトリーダー)

## 「表現豊かな発話音声のコンピュータ処理システム」

### 1. 研究実施の概要

#### 研究の概要

本プロジェクトは高度メディア社会の情報処理技術の枠の中で、発話音声を持つ機能的役割、特に、言語情報以上の発話様式で表現する「声を持つ意味」について解明する。人は文字には表れない情報を声の調子によって、発話意図、態度、感情状態を示す。大規模自然音声対話データベースの収集と分類により、発話様式のバリエーションを分析すると共に、声の表情を含む表現豊かな音声合成や言い方認識技術への応用に向けて研究を行う。

#### 研究成果

現在、約 500 時間程度の自然対話音声を収録済みである。その中、約 200 時間の書き起こしと 75 時間の感情・意図・態度発話様式ラベル付けが出来た。これらのデータの分析結果により、新たな音声合成インタフェースと環境に応じた表現形式の違いに対応するパラメータが示唆された。本プロジェクトにおいて、表現豊かな発話音声は、感情のみならず、発話者による意図・態度・発話行為を含んだ音響情報としてとらえている。実際、話者の感情状態よりも、発話の機能的役割の方が重要であると分かった。自然な発話環境での大規模音声コーパスの分析から得たパラ言語情報・非言語情報を示した新たな韻律パラメータを利用し、さまざまな発話様式を表現できる音声合成技術を提案した。このノウハウは音声合成利用のみならず、双方向的に認識への利用も可能である。つまり、聞こえ方による「言い方認識」の観点から発話者と聞き手との関係が判断できる方法も示唆された。

#### 今後の見通し

コミュニケーション科学、特にパラ言語情報コミュニケーションの分野では、音情報と言語情報との組み合わせにより、新たな技術の開発が可能となる。本プロジェクトで作成した自然な発話環境における大規模音声データベースは、感情や意図をも表現できる音声技術の研究を底上げする、基盤的な研究リソースとなると考えられる。発話様式、対話音声、比較言語など、各研究グループのサブテーマごとの分析結果により、自然対話音声に関するいくつかの韻律パラメータが発見され、今後、データベースを充実し、合成エンジ

ンの改良、ラベル付けの自動化、単位選択手法を含めて、「環境に適切な」、「言い方も制御できる」音声合成を目指す。特に、音素ラベルに依存しない波形接続型音声合成を期待し、文字言語に依存しない入力インタフェースも期待できる。これらの技術は、高度メディア社会の情報伝達に利用され、多くの方々がさまざまな場面で自分の声と自分の言い方で、音声合成を利用できる。ロボットとの会話、情報機器との会話、人と機械との豊かなコミュニケーションが出来るように期待できる。

## 2. 研究実施内容

音声言語グループでは、表現豊かな音声処理の工学的実現を目指し、音声合成技術・音声認識技術が取り扱う対象を話者性、感情、態度といった非言語情報や準言語情報へ拡大する手法を研究すると共に、音声と関連する画像等のメディアを統合した表現豊かなヒューマンインタフェースの実現手法を検討するものである。本年度は前年度までに収録した音声データ（感情をこめた女声ナレーター音声データ、多様な韻律を持つ女性ナレーター音声データ、バイリンガル話者音声データ、発話顔画像つきデータ、機械との対話音声データ）と音声処理手法（混合正規分布モデルと周波数軸伸縮に基づく声質変換技術、音声分析合成系STRAIGHTの素片選択型音声合成器への組み込み）を用いて、合成音声の感情表現制御、合成音声の高品質化、異言語にまたがる話者性の制御手法とその発音評価への適用、対話分析、画像素片選択による発話顔画像つき音声合成手法の研究を進めた。また音声に関連する新たな入力メディアとして非可聴音声を集音する機器を開発した。

意味構造グループでは、まず、音声に関わる韻律や音色により、聞き手に伝わる印象、意味の認識にどのような影響が現れるか考察した。具体的には、「りきみ」と呼ばれる現象や延伸や言い淀みあるいは、感動詞や一音節形容詞の韻律情報の考察を行なった。この対照研究として、声調言語である中国語の一音節形容詞に応答表現、感動詞に関して検討を行なった。音韻レベルの考察としては、長音、または二重母音が、何らかの話者の意図や感情と結びついていることに関して準備的考察を行なった。また、反復などの発話スタイル、イントネーションが音声要素に与える影響、聴覚印象に与える影響に関して検討した。さらに、これまで多く論じられている旧情報・新情報の理論を音声の枠組みから捉えた実験を行なった。その他に、比較言語文化の観点から、ヨーロッパ諸語とアジア諸語の「呼びかけ」表現の音声的観察、日本語と中国語の卓立（プロミネンス、フォーカス）は音響的観察から、「卓立強調」の表現意図に寄与している要因を検討した。日本語の習得程度と韻律情報の理解には関係があるのか、という応用的な観点から教育への実験的アプローチも行なった。さらには最終目標の一つである発話態度のグルーピングに関する基礎的研究も行った。

対話音声グループでは、「あの」、「ええと」などのフィラーは、単に話者が言い淀んでいることを知らせるだけでなく、それらを含む談話の構造についての情報を担っているという指摘が先行研究でなされている。日本語のフィラーにも談話構造についての情報の

1つである談話境界を知らせる働きがあるかどうかを、談話境界とそれ以外の境界近辺のフィラーの種類や頻度を比較することによって調べた。その結果、多くの聞き手によって境界があると認定された句境界ほど、その近辺にフィラーが出現する比率は高く、とりわけ「ええと」の出現率の上昇が著しかった。この結果から、フィラー、特に「ええと」は、談話境界をマークする特徴の1つであることが示唆された。使用されるフィラーの種類や頻度には個人差の大きいことが指摘されている。そこで、各種フィラーの使用頻度パターンによって、話者タイプを分類することを試みた。同時に、各種フィラーの使われ方に話者の違いを超えた一般性がないかどうかを、対応分析を行った。引き伸ばしは、発話頭のフィラーで生じることが最も多く、ついで発話頭の語で生じることが多い。また、長い発話（情報量の多い発話）ほど、発話冒頭部で引き伸ばしの生じることが多い。これらのことは、発話冒頭部のなるべく早い地点で時間を稼いで、文の構造を整えようという発話産出ストラテジーの現われと考えられる。

発話様式グループでは、発話様式の基本研究を行っている。自然対話音声データとその書き起こしデータ、および発話意図・発話行為ラベルとの統計的関係を分析することにより、自動ラベル付与ソフトウェアの開発、および発話様式に対応する音声特徴量抽出アルゴリズムの検討を行った。本年度は新たな「声質判断」アルゴリズムを開発し、発話環境と声質の関係を明らかにした。この声質判断方法については、特許申請中である。また、近未来「ロボットの声」になる音声対話システム・インタフェースに関する検討、および合成音声の評価も行った。声の「カミ」による意味の違いに関する研究を開始するとともに、ピッチの変化や声の強さを音声波形から自動抽出するアルゴリズムの開発を行った。

比較言語文化グループでは、話者態度や感情が豊富に含まれる相互性のある発話音声データベースを作成した。平成13年度から比較すると、データ収集技術はWizard-of-Oz手法により改良された。怒り、喜び、失望をコントロールし、被験者の表現や発話スタイルを収録することが可能となった。この発話様式実験ソフトウェアは日本人被験者においても使用できるためシステム統合グループでのデータ収集においても利用された。これらの音声データの分析のために、ICPでは、生態学的データ収集（心音、血圧、呼吸速度、ガルバニー電気反応など）を行い、医学的に感情や態度を統計的データによって分析することに着手した。また、発話様式グループによって構築された音質抽出ソフトウェアを利用し、発話スタイルとストレスや喜びの影響による音質変化の比較検討を行っている。

コミュニケーション支援サブグループでは、コミュニケーション支援システムの設計の上で必要な談話ラベル階層の検討を行った。いくつかの提唱された談話構造を参考にしながら、ATRグループで収集された女性話者音声の分析に基づき、独自の談話ラベル階層を提案した。そして、提案談話ラベルを用い、ALS患者とその家族の日常会話データについて、ラベル付与に着手した。また、前述の女性話者音声データの中から約140の相槌を抽出し、それらについて、意図（肯定、反射的、ターン保持）と音響的特徴（韻律、および、声質）との関連性について分析し、有効な結果を得た。昨年度までにデータベース作成、ならびに、システム試作が完了したALS患者本人の音声を用いたコミュニケーション

支援システムについて学術論文を執筆した（本年4月に採録決定）。図5は患者宅で試作したシステムの使い方を患者に説明している様子である（本人の同意を得て掲載）。

システム統合グループでは、自然対話音声データベースの作成、音声データのテキスト書き起こし、およびラベル情報付与（図1）に加え、音声合成ソフトウェアやデータベース検索ソフト（図2）の研究を行っている。平成14年度は、約500時間の音声データ、200時間のテキスト書き起こし、および75時間のラベル情報付与により研究分析材料を整えた。さらに、「tap2talk(図3、テキスト入力が必要で、キーワードのボタン操作のみで合成)」音声合成のための単位選択インタフェースを開発した（図3、図4）。この方式では、対話音声データベースに含まれる発話単位を、目的に応じて、適宜再利用して対話合成音声を作成する。また、いわゆる「TTS(テキスト入力からの音声合成)」ではなく、音声信号を入力として、新たな合成音声を作成する、声質変換的な音声合成単位選択方法について提案した（図6）。中間的な特徴パラメータの抽出にあたって、朗読音声データベースの音声を利用した上で、最終的な合成音声は、自然対話音声データベースから単位選択する。この音声合成単位選択方法については、特許申請中である。

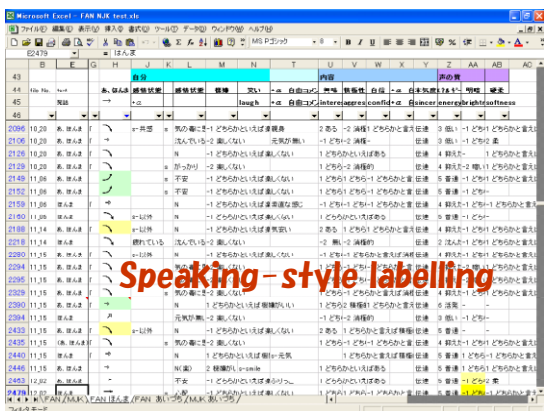


図 1

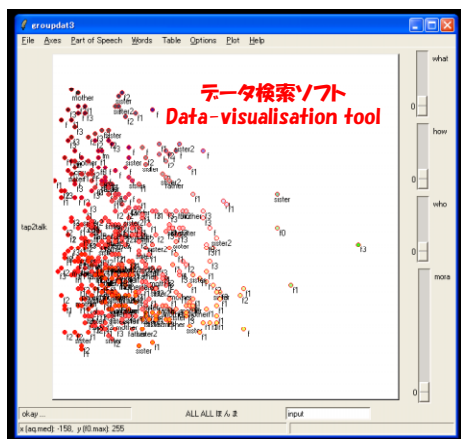


図 2



図 3



図 4



図 5

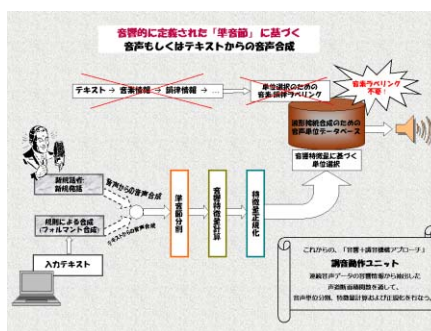


図 6

### 3. 研究実施体制

#### 音声言語グループ

研究分担グループ長：鹿野清宏（奈良先端科学技術大学院大学 情報科学研究科、教授）

研究項目：音声合成技術

#### コミュニケーション支援サブグループ

研究分担グループ長：安村通晃（慶応義塾大学環境情報学部教授/政策・メディア研究科委員）

研究項目：音声インタフェース技術の研究

#### 意味構造グループ

研究分担グループ長：中川正之（神戸大学国際文化学部、教授）

研究項目：発話理解・発話表現・パラ言語マッピング

#### 対話音声グループ

研究分担グループ長：伝康晴（千葉大学文学部行動科学科認知情報科学助教授）

研究項目：音声文法理論・音声表現のカテゴリ化

#### 発話様式グループ

研究分担グループ長：芦村和幸（科学技術振興事業団、研究員）国際電気通信基礎技術研究所

研究項目：韻律と声質における識別のモデル化

#### 比較言語文化グループ

研究分担グループ長：Veronique Auberge（ICP Grenoble（仏）教授）

研究項目：言語依存と個別性の比較聴覚実験

#### システム応用グループ

研究分担グループ長：ニック キャンベル（国際電気通信基礎技術研究所、主幹研究員）

研究項目：音声データ収集・ツール作成・音声合成・音声認識インタフェース

### 4. 主な研究成果の発表（論文発表および特許出願）

#### （1）論文発表（国内）

○ ニック キャンベル、音声合成の観点から見た言語音声の特徴、大修館書店「言語」10

月号, pp 52-61, 2002.10

- ニック キャンベル、音声合成、共立出版AI辞典第2版、pp86-87
- 定延利之 2002(編著) 『「うん」と「そう」の言語学』. 東京: ひつじ書房. (執筆  
者: 串田秀也・黄麗華・井上優・定延利之・坊農真弓・富樫純一, 以上7名) (執筆  
部分: 序文, 「「うん」と「そう」に意味はあるか」 (pp. 75-112.))
- 中川正之 2002 「中国語の動詞をどう捉えるか」, 大修館書店『月刊言語』Vol. 31-No. 12,  
pp. 74-79.
- 中川正之・朱春躍・澤田浩子 「韻律パターンと語の意味」, 中国言語文化研究会『CHINESE  
LANGUAGE & CULTURE 』, NO. 3, pp. 1-34. 2002
- 坊農真弓 2002 「プロソディからみた「うん」と「そう」」, ひつじ書房『「うん」  
と「そう」の言語学』, pp113-126.
- 定延利之 2003 「書評論文: 宮岡伯人『「語」とはなにか: エスキモー語から日本語  
を見る』」, 日本語文法学会『日本語文法』, 第3巻1号, pp. 135-146.
- 澤田浩子・朱春躍・中川正之 「連体修飾における文法と音声」, 日本語文法学会『日本  
語文法』第3巻1号, pp.100-116. 2003

## (2) 論文発表 (国際)

- Mokhtari P., Campbell N., “Automatic Measurement of Pressed/Breathy Phonation  
at Acoustic Centres of Reliability in Continuous Speech” IEICE TRANS. INF.&SYST.,  
Vol. E86-D, No. 3 pp574-582, March 2003
- Douglas-C. E, Campbell N., Cowie R., Roach P., “Emotional speech: Towards a new  
generation of databases” Speech Communication 40(2003) pp33-60
- Douglas-C. E., Cowie R., Campbell N., “Speech and Emotion” Speech Communication  
40 (2003) pp1-3
- 松本恵美子・定延利之 2002 「日本語の韻律における「りきみ」と、日本語学習者の  
理解度」, Department of Japanese Studies, The Chinese University of Hong Kong  
and Society of Japanese Language Education (ed.), Quality Japanese Studies and  
Japanese Language Education in Kanji-Using Areas in the New Century, Hong Kong:  
Himawari Publishing Company, pp. 455-461.
- Mikiko Mashimo (NAIST), Tomoki Toda (NAIST), Hiromichi Kawanami (NAIST), Kiyohiro  
Shikano (NAIST), Nick Campbell (ATR/NAIST/CREST), “Cross-language Voice  
Conversion Evaluation Using Bilingual Database”, IPSJ Journal, Vol. 43, No. 7,  
pp. 2177-22185, July 2002
- Iida, A., Higuchi, F., Campbell, N., Yasumura, M. “A Corpus-based Speech  
Synthesis System with Emotion,” Speech Communication. Vol. 40/1-2 pp. 161 - 187.

## (2) 特許出願

H14年度特許出願件数: 1件 (研究期間累積件数: 1件)