

「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

Nick Campbell

(㈱国際電気通信基礎技術研究所 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」

(EXPRESSIVE SPEECH PROCESSING)

1. 研究実施の概要

本研究では、5年間の研究期間内に1000時間分の「意図」、「態度」、「感情」などを含む音声データベースの構築(自然対話約70%、スタジオ収録音声10%、特定用途用20%)を目標としている。本年度は、昨年度に引き続き、自然対話音声、スタジオ収録、特定用途音声データ収集を行った。収集した音声データを付録に記す。自然対話音声の収録においては、マイクや録音機材の比較検討を経て、本格的な収録を開始した。スタジオで収録された音声は自然音声波形接続型音声合成システム用データベース、声質変換用データベースの音声データベース用に加工され、これらを用いた応用システムの試作を行った。また、システム統合グループでは、本研究を円滑に進める上での分析ツールの作成を行なった。

2. 研究実施内容

(1) 音声言語グループ(鹿野清宏・奈良先端大)

本年度は、表現豊かな音声合成処理研究に用いる音声データベースの設計・収録を行うとともに、その基礎技術の一つである声質変換技術の改善とその異言語間にわたる適用の有効性の調査を昨年度から継続して進めた。また波形選択型音声合成システム CHATR に分析合成方式 STRAIGHT を組み込んだ合成器の整備については、幅広い韻律の音声を高品質に実現するデータベース設計を行い実装した。以下詳細を示す。

作成したデータベースは以下の通りである。合成対象とする情報の拡大の観点から、(1)日英バイリンガル話者による日英読上げ音声データベース、(2) ナレーターによる感情をこめた朗読音声データベース、(3) 音声同期した顔画像データベース、(4) 受付案内機器への質問音声を、また合成音声の品質改善の観点からナレーターによる韻律的に異なる9種の発話セットを設計、収録した。

研究成果としては、基礎技術である声質変換技術の改良として従来の混合正規分布モデルの適用に加え周波数軸伸縮を併用することで合成音声の自然性の改善を実現した。また応用研究として、収録したバイリンガル話者音声に対して声質変換技術を適用し異言語間にわたる話者性抑圧の実験を行いその有効性を確認した。その成果は投稿論文としても採録された。また韻律的に

多様化した複数の音素バランスデータベースを用い波形接続型音声合成システム CHATR に分析合成手法 STRAIGHT の技術を援用することで高品質かつ柔軟な話速に対応できる音声合成手法を開発した。

(2) 対話音声グループ (伝康晴・千葉大)

発話冒頭における語句の再開始を対話音声文法の観点から分析し、これらがインタラクションの調整のための話し手の意図を伝える文法装置であることを示した。話し手は、言いかけた語を中断する際に末尾音を引き伸ばすことで、その発話全体の産出に通常より困難が伴うことを事前告知していることがわかった。

談話構造、とりわけ談話境界と、その近辺で出現するフィラーの頻度・種類・音響的特徴との対応関係を、講義、学会講演、模擬講演を対象に調べた。発話の即興性が高いほど、談話境界の深さとフィラーの出現頻度・種類との間により密接な対応関係が観察された。

自閉症者を対象とした対話収録を実施した。対話を成立させるコミュニケーション能力と、韻律の異常の度合いについては個人差が激しいが、数名(二例の成人症例を含む)において顕著な傾向が観測された。

(3) 発話様式グループ (芦村和幸、科学技術振興事業団研究員・ATR)

本研究チームは音声合成のための韻律分析とモデル化、および、そのために必要な表現豊かな発話形式に関する様々な韻律を含む自然音声データベースの構築を行なっている。

- ・ 前年度検討を行なったデータ収集方法にもとづいて、250 時間程度の音声データ収録を行なった。収録された音声データについては、随時、知覚的および音響的な分析の上で、発話態度、意図、感情、タイミング情報などのラベル付与を行なっている
- ・ 上記データベースの分析結果をもとに、日常会話における発話タイミングの制御モデルについて検討した。また、フォルマント情報にもとづいて、「声の固さ(breathiness)」を分析した。
- ・ データ収集用録音機材としては、主に DAT を用いているが、被験者に負担をかけない収集方法として、MD を用いた録音およびデータ分析手法について検討した。

(4) コミュニケーション支援サブグループ (安村通晃、慶応義塾大)

- ・ 昨年度に引き続き、表情豊かな音声合成を応用したコミュニケーション支援システムの音声データベースの整備、および、入力補助機能の充実を図った。日本語システムに加え、英語対応システムを実装すべく英語 GUI とリスト表示部分を拡張した。本研究の成果を国際学会、および、国際ワークショップで発表した。
- ・ 1999 年までに作成した大規模感情音声コーパス(男女各 1 名の怒り、喜び、悲しみ)の音響的特徴(F0、音素長、パワー、フォルマント)の分析を行った。同話者の平静音声(音韻バランス文章セット読み上げ)を収録し、聴取実験により、4 音声有意向に聞き分けられることを確認した後、1999 年に作成した 3 感情音声合成システムに平静音声を付加しシステムを拡張した。本研究の成果をまとめた論文が国際ジャーナル誌に採録が決定した。
- ・ 音声収録:(1)屋外自然音声の収録を行った。話者に DAT にヘッドセットマイクを装着してもらい、牧場、森林、スーパー、郵便局などで収録を行った。(2)音韻バランス文章セット男性 2 名、

女性1名を収録した。

(5) 意味構造グループ(定延利之、神戸大学国際文化学部助教授)

本年度は発話態度の分類に本格的に着手した。発話態度は、韻律だけでなく、文末形式と相まって表現されるという考えのもと、暫定的解答として以下15類の発話態度を抽出するに至った:「不本意な承認」「本意の承認」「請け合いa」「請け合いb」「伝言伝え」「おどし」「感動a」「感動b」「程度強調a」「程度強調b」「落胆・失望」「突き放しa」「突き放しb」「ひとりつぶやき」「ぼやき」。

以上の発話態度を抽出する過程で、主に以下3点を明らかにした。第1点、フォネーション情報はこれまで軽視されてきたが、日本語においては、いくつかのネイティブアメリカン諸語と同じく(あるいはそれらよりも)、フォネーション情報が発話態度に関与する。第2点、アクセントとイントネーションの関係はこれまで考えられてきたような相互独立の関係とはかぎらない。話し手の態度を強く表すイントネーションの中には、アクセントと共存せずアクセントを消去するものがある。第3点、イントネーションの使い分けが文法的区分と対応する例がある。

本年度は、自然な音声データに付与すべきタグを考案するための準備的作業もおこなった。言語学にかぎらず、情報処理工学、社会学、行動学等、多方面の領域で利用可能なデータベースを構築するため、タグの考案を2方式で並行しておこなうことに決め、作業を着手した。第1の方式は、CRESTのデータベースを利用しそうな研究者が所属すると思われるさまざまな学会に、大規模なアンケート調査をおこない、それを集計・分析して、「多方面で望まれているタグ」を割り出すという方式である。この方式でのタグ検討を進めるために、13年度には数十の学会を選定し、リストアップした。第2の方式は、神戸グループ内部でさまざまな言語事実に考慮しながら、理想的なタグを見出すという方式である。この方式でのタグ検討を進めるために、13年度には、タグの多層構造(談話構造層、隣接ペア層、文法構造層、プラス音声レベルとして、イントネーション層、相互作用層、さらにジェスチャー層)の検討作業に入った。具体的には、(1)先行研究の調査、(2)タグの可能性の探求、(3)パイロットスタディー(実際にタグをつけてみる)、(4)(3)の問題点の検討、(5)必要なタグの広域調査、(6)(4)、(5)に基づく新たなタグの付与、(7)データの取り方の検討、(8)新たなデータで(6)を繰り返し行う、というサイクルで作業を始めた。

自然なデータ収集を進める過程で浮上してきた問題も洗い出した。まず、プライバシー問題は、現行の同意書では解決しきれないことが判明した。現行の同意書は、インフォーマントのプライバシー問題は解決するが、インフォーマントが録音データの中で話題にした(たとえば悪口を言った)人物のプライバシー問題は、解決しないからである。また、インフォーマントが未成年である場合や、音声だけでなく画像までも収録する場合、現行の同意書での対応には限界が生じることを洗い出し、先行する国研データベースを参考にしつつ、対応を検討し、それらの問題を解消するための法的検討に着手した。

(6) 比較言語文化グループ(Veronique Auberge ICP Grenoble(仏)教授)

本年度は信頼性のある感情音声の収集に研究の目的を置き、話者の自然な感情を引き出す方法の検討を行い、収集に着手するとともに独自のラベリングソフトの開発を行った。

新しい試みとして、予期しない出来事に対する話者の戸惑いとそれに伴う感情の誘引方法につ

いての検討を行った。その結果、Wizard of Oz 方式を採用した話者の感情音声の収集が有効と考え、この方法に基づく音声収集のためのアプリケーション開発に着手した。場面としてはフライトシミュレータ、言語教室、子供部屋を設定し、対話方式で被験者の発話を収集する。現在、フライトシミュレータプログラムが完成し、それを用いた音声収集に着手している。本研究で開発したアプリケーションは多言語環境での利用が可能である。

その他、従来の音素バランス文セットを用いた感情音声収集にも取り組んでいる。対象言語はフランス語、英語、ロシア語、対象感情は、喜び、満足、怒り、不安、驚き、疑念、苛立ち、嫌悪、緊張、平静である。また同文セットを用いて、態度音声の収集も行っている。

さらに、声質ラベリングに対応できる独自のラベリングソフト EDIPROSO を開発した。開発環境には Matlab を使い、Mac および Windows PC で稼動する。

(7) システム統合グループ(ニック キャンベル・ATR)

表現豊かな発話形式の研究に必要な各種ツールの研究開発を行なった。また、本プロジェクトの目指す音声技術の応用に関する予備的実験を実施した。

研究用ソフトウェアの開発：

- ・ 発話単位切出し：長時間の連続発話データから、一発話分を認識し切出す。
- ・ 発話様式ラベリング：音声に含まれる緊張感や声の固さを測定する。
- ・ 音声合成インタフェース：音声合成の入力方法として、アイコンを用いて、テキストおよび発話様式を簡単に指定する。

応用に関する実験：

- ・ Java サブレット技術を利用して、音声合成に対して、Web から情報を取得できるインタフェースを作成した。
- ・ 上記インタフェースを応用して、携帯電話を端末として音声合成処理を行うシステムを実験的に構築した。なお、この音声合成システムにおいては、実験的に、i-appli を通して感情パラメータを音声合成の選択基準に組み込んだ。
- ・ 「音声合成サービス」の例として、音声合成と言語翻訳を統合した実験システムを構築し、技術的可能性を確認した。

3. 研究実施体制

音声言語グループ

グループ長：鹿野清弘(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：高品質音声分析合成系による音声表情 (Expressive Speech) の制御

対話音声グループ

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

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

コミュニケーション支援サブグループ

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

グループ長: 定延利之(神戸大学国際文化学部助教授)

研究項目: 発話理解・発話表現・パラ言語マッピング

比較言語文化グループ

グループ長: Veronique Auberge (ICPGrenoble(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

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、プロジェクトリーダー)

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. 研究成果の発表

論文発表

- キャンベル ニック、発話音声の特徴: 音声文法からの観点、pp 161-182, in **音声と文法 III**, くらしお出版、2002.
- 小川暁夫, 2002, 「非人称構文と人称構文の境界: 対照類型的研究」 In: *Grammatische Kategorien aus sprachhistorischer und typologischer Perspektive.*, ドイツ Muenchen: iudicium, pp.75-90.
- 朱春躍, 2001.10, 「北京語会話の音脱落」, 『新世紀的現代語音学——第五届全国現代語音学学術会議』, 中国: 清華大学出版社, pp.204-207.
- 立石浩一, 2001.9, 「最適性統語論がしていることとしていないこと」, 『言語』9月号第30巻第10号, pp.50-55.
- 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.
- Morlec, Y., G. Bailly, and V. Aubergé (2001) *Generating prosodic attitudes in French: data, model and evaluation*. *Speech Communication*, **33**(4): p. 357--371.
- Rilliard, A. & Aubergé, V., forthcoming, *Prosody diagnostic using a reiteration paradigm* in Special Issue of the *Journal of Speech Processing* (TTS Edinburgh WS).

特許出願

なし