

## 研究課題別事後評価結果

1. 研究課題名 「表現豊かな発話音声のコンピュータ処理システム」
2. 研究代表者名及び主たる研究参加者名（研究機関名・職名は研究参加期間終了時点）  
研究代表者 Nick Campbell ((株)国際電気通信基礎技術研究所 主幹研究員)  
主たる研究参加者  
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)  
中川 正之 (神戸大学 教授)  
Veronique Auberge (ICP Grenoble Professor)  
安村 通晃 (慶応義塾大学 教授)  
伝 康晴 (千葉大学 助教授)

### 3. 研究内容及び成果:

本研究はコンピュータによる音声合成のための新たな情報形式を提案するものである。従来の音声合成技術は主に情報提供という観点に立つもので、人間が日常的に行う双方向的コミュニケーションの利用には不適切である。人の双方向的コミュニケーションには感情を含む必要がある。具体的には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」について解明することが大切である。人は文字には表れない発話意図、態度、感情状態(以下「声の表情」という)などの情報を声の調子によって示す。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った。

発話音声の波形から音素情報、韻律情報、声質の3要素を基本系として、話し手の感情や発話意図などに対応する音声特徴のコンピュータによる自動抽出を行う。さらに、それらを既知の言語構造や個別の発話行為特徴へマッピングすることにより、高次の発話情報処理技術が可能となることを明らかにした。

- (1)大規模音声データベースの設計・収集・分析
- (2)統計的学習によるパラメータ化
- (3)言語構造・パラ言語情報のマッピング
- (4)発話処理システム・コミュニケーション支援の実用化

に関する研究を行い、

- (イ)声情報の役割: 言い方によって意味がどのように変わるか
- (ロ)表現の方法: どのように発話意図を伝えるか
- (ハ)識別の定義: 意味判断の普遍性と言語への依存性

という3つのレベルにおいて音声言語文法の理論化とアルゴリズム化に大きな成果を挙げた。

この技術は音声データから「声を持つ意味」の意味識別の基となる発話様式・話者態度・感情までの情報を抽出・活用するものである。音声合成技術の場合、適切な言い方で「声の表情」を伝える。このために人がどのように声を使って話を伝え、どのように音声情報から発話の「意味」を解釈するか、という人の話し方の研究から意味表現構造を解明した。

本研究の成果である「表現豊かな発話音声のコンピュータ処理システム」の構築はコミュニケーション支援技術をはじめ、人と機械の音声インタフェースのあり方、人に優しい情報交換の手段などを定義し、高度メディア社会における多面的な貢献ができた。さらに、具体的な成果物として得られた音声データベースは比較言語学、比較文化学、実験心理学および音声工学に対して国際的に類を見ない研究材料を提供できた。

双方向的コミュニケーションに必要な感情の表現について、一般には音声合成の次のステップとして声に表情を乗せること、つまり喜怒哀楽を声によって表現することが目標として捉えられた。しかし現実の声が表す意味領域は当然のことながら「感情表現」をはるかに超えている。対話発話音声コーパスの分析によって、発話音声信号のなかでコード化されている情報は言語情報だけでな

く、それと同等にノンバーバルな情報が重要な役割を持つことを明確にした。このノンバーバル情報が示すものは発話者の感情状態にとどまらず、興味や気分などの心的状態と共に聞き手との人間関係や対話環境から生まれる親しさなどの情報を多面的に含んでいる。この情報は言語的音声信号の「上」ではなく、完全独立的な層で伝えられている。各レベルの「声の様子で伝えられる情報」のモデル化や枠組みを制御することが重要であり、そのため単なる「声に表情」だけでなく、多面的コミュニケーションモデルの研究を行った。その結果新たな音声合成モデルや入力インタフェース技術を検討し、大規模対話型音声コーパスからの発話単位選択のアルゴリズムの改良を行った。従来の発話単位は音素であったが、一発話分全体を発話単位としてコーパスから声の様子を制御するため、意図や態度を選択条件として単位選択を行う方式とした。

本研究における音声コーパスの書き起こし結果を分類すると、I-型発話 (information 語彙情報) とA-型発話 (affect 情動情報) の2種類に分ける必要があると判った。I-型発話は書き起こし情報のみ (つまり文字情報) であり、十分な意味を解釈できるが、A-型発話の場合は韻律情報・声質情報がないと意味 (つまり話者意図) を解釈できない場合が多い。この情動情報は辞書に登録されていない単音 (へ、ハア、えー、笑い等) を発話音に載せる情報である。

このような情動情報を表現するため、音声合成の入力に新たなインタフェースの開発を行い、文字入力ではなく意図入力による音声合成手法を提案している。このインタフェースは意図だけでなく話者状態 (気分・興味など) と共に相手関係 (人間関係・対話環境状況など) を入力可能とし、この情報に基づいて目的発話の性質・韻律特徴がフィルタ化され、環境や意図に適切な発話単位がデータベースから選択される。その発話候補は同様なテキスト形式 (文字列) であっても目的対話意味を伝える発話単位である。

本研究成果によって音声信号が伝達する言語情報と情動情報両方の処理が同時にできるようになった。本研究は人間の「声によるコミュニケーション」の理解の上では当然の結果を得たといえるかもしれないが、言語科学・音声工学分野として新たな見方を提案した研究であり、国内外における「Expressive Speech Processing」をテーマとする研究をリードしてきたといえる。

#### 4. 事後評価結果

##### 4 - 1. 外部発表 (論文、口頭発表等)、特許、研究を通じての新たな知見の取得等の研究成果の状況

本研究では発話音声における表現力の豊かさとはなにかを探求し、膨大な音声データの取得、分析、データベース構築を行い新理論を実証的に示した。具体的には日常会話における発話を言語的な意味や発話者の感情だけでなく、興味や気分などの心的状態、聞き手との人間関係や対話環境など多方面から分析し、発話が含むノンバーバルな情報を取り込んでより自然に聞こえる音声合成を可能とした。言語の感性的情報の取り扱いに新しい世界を開いたすばらしい成果といえる。類似の研究はほとんどなく、世界的にもほとんど手がついていない領域の新しい研究分野を立ち上げたという点は高く評価できる。

人の日常会話における自然な発話音声データを収集・分析して作成された膨大な音声データベースは貴重なものであり、今後の音声言語の感性的研究において常に参照されるものとなるだろう。今後この音声データベースがさらに拡充され、音声研究の中核となることを期待する。

研究代表者の構想を中心に7つのサブグループが比較的独立に研究を進めてきたが、もう少しサブグループ間の連携と研究目標の絞込みにむけたコーディネーションがあれば、さらに大きな成果に結びついたであろう。

論文発表は国内31件、海外4件、口頭発表は国内118件、海外103件であり、世界的に新しい研究分野をリードする発表を行ってきた。特許出願は国内5件、海外2件とやや少ないが、今後の研究の展開によっては重要な基本特許となる可能性がある。

##### 4 - 2. 成果の戦略目標・科学技術への貢献

本研究は人間的なコミュニケーションの根幹に挑む研究であり、今後の科学技術に及ぼす影響はきわめて大きい。音声言語処理に新しい技術分野を拓いたことは特筆すべきことであり、今後の

研究の加速が期待される。

4 - 3 . その他の特記事項(受賞歴など)  
特になし