

「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 主幹研究員)

## 「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

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。人は文字に表れない発話意図、態度、感情状態などの情報を声の調子を用いて表す。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。

音声研究の進化はデータベースに依存する。従来研究材料として用いられてきたアナウンサーによる朗読音声は、日常的な一般の人間の発話音声に表れるバリエーションを含まない。そのため、高度メディア社会における一般の人間が使える、一般の人間のための技術を開発するためには限界がある。そこで本研究では、日常の話し方がもつ表現の豊かさを反映する対話音声を、日常自然環境で1000時間収録することを目指した。この自然発話環境において収録される大規模音声データベースは、分野における後続の研究のためのリソースともなる。

### 2. 研究実施内容

現在の音声技術は硬くて、「感情が含まれてない」と言われている。図1が示すように、音声合成は人間の声の変わりに使われている場合が多く、表現豊かな発話様式制御が重要となる。

「感情」を表現できる音声合成技術の開発に向けて自然対話の音声データベースの整備を進め、得られたデータを対象に新しい韻律パラメータの研究を進めてきた。パラ言語情報、非言語情報を数値化するための適切なパラメータを設定することにより、表現形式のバリエーションをかなり説明できることがわかった。

日常対話では、一般の話者は、文字には表れない情報、発話意図、態度、感情状態などを声のバリエーションによって表現する。この発話音声を持つ機能的役割、すなわち言語情報以外の発話様式で表現される「声」が持つ意味について、韻律パラメータとパラ言語情報との関係を追究した。

発話様式と韻律の関係を、主に3つのパラメータ、すなわちピッチ（音声信号の基

本周波数、以下、F0)、声の大きさ (RMSパワー) と発話タイミングによって検討した。さらに、第四の韻律パラメータとして、声質がパラ言語情報の伝達に用いられていることも検証された。図2はその分析用に開発したツールである。



図1：コミュニケーション支援の立場からの発話音声コーパスの収集と感情音声の分析

### Snapshot of our Acoustic Analysis Visualisation tool (Tcl/Tk)

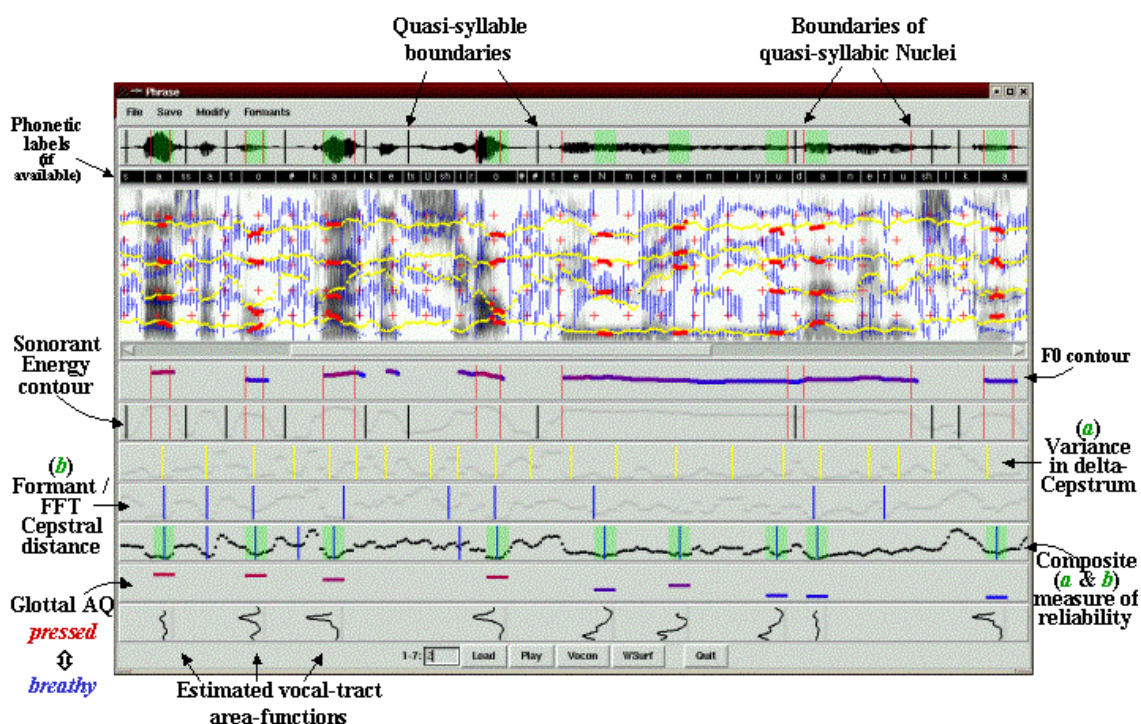


図2：音声波形データからの韻律特徴抽出と声質の分析

本ソフトは入力対話音声波形から自動的に韻律情報を抽出し、発話音声のデータ行列を出力する。

図3は、30歳代の女性話者が発話した100時間の対話音声の声質パラメータ（NAQ）とF0の平均値を示す。声質は「緊張音」（pressed）「平静音」（modal）「氣息音」（breathy）の範囲に納まって、相手に対する親しさや丁寧度と関連している。すなわち声質は相手、対話環境、話者の状態によって変化する。右は、左に示された家族に対する発話をメンバーごとに分析した結果である。二つのパラメータの振る舞いから発話者と各家族メンバーとの人間関係が伺える。

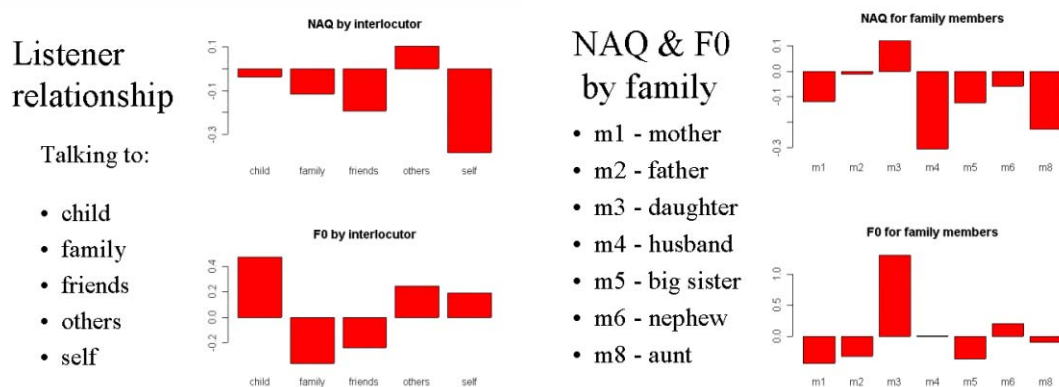


図3：相手による発話様式の違い

相手グループ（左）および家族構成員（右）ごとのNAQ（声の固さ、上段）およびF0（声の高さ、下段）。発話者の相手に対する気の使い方によってF0と声質が変化する。

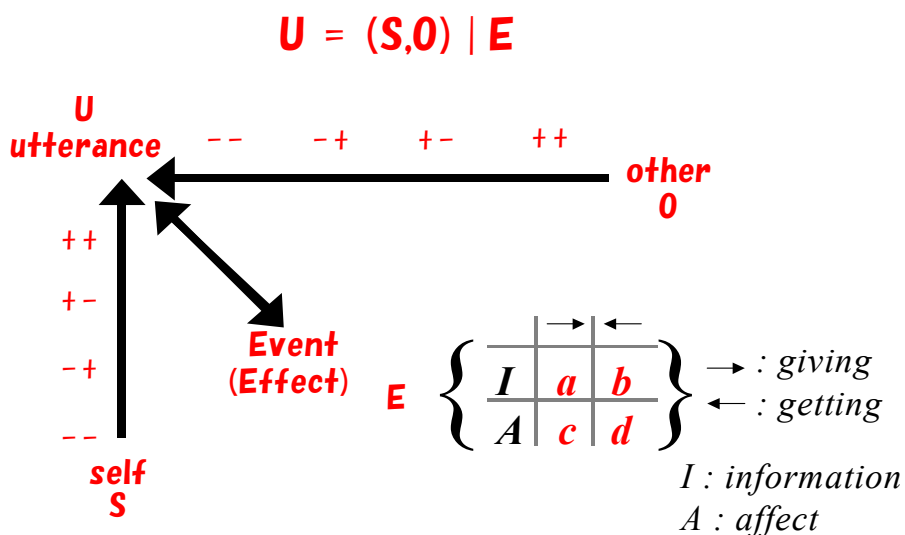
図4は、上述の発話様式情報を考慮する新たな波形接続型音声合成手法を示す。小規模音声データを利用して合成したターゲット音声波形を元に、大規模自然発話データベースから、韻律特徴フィルターを通して候補となる音声単位を決定する。音声波形の基本単位はフレーズであり、発話の自然性が保たれる。



のが文脈における適切さである。図5に示されるような音声データの分析ソフトを用い、同一話者によって様々な発話様式で発話された同一言語内容の発話を分類した結果、同一話者、同一単語であっても音響的な違いによって、何種類もの意味として解釈されることがわかった。

今後の音声合成の基本課題は、利用者がどのような入力でこれらの意味の違いを的確に反映した音声を選択できるかである。図6の携帯電話を用いたインターフェースは一例であり、図7はその基本枠組みを示す。感情情報だけではなく、人間関係、話者意図、態度、気分などの情報によって、出力音声が変化する。

図7：発話内感性情報の枠組み



発話 (U) は、相手 (O) と自分 (S) との関係におけるコミュニケーション・イベント (E) とみなされる。そのイベント (あるいは影響) は発話内容情報 (I) あるいは話者状態情報 (A) のやり取りである。

### 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（科学技術振興機構、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（原著論文）発表

- Akemi Iida, Nick Campbell, “Speech Database Design for a Concatenative Text-to-Speech Synthesis System for Individuals with Communication Disorders”, International Journal of Speech Technology 6, 378-392, 2003
- Nick Campbell, Donna Erickson (2004) “What do people hear? A study of the perception of non-verbal affective information in conversational speech”. Journal of the Phonetic Society of Japan, pp.9-18, Vol.8, No.1.
- ニック キャンベル (2004) 「コーパスベース音声合成技術の動向—大規模音声コーパスによる音声合成」、電子情報通信学会誌Vol.87, No.6, pp.497-500.
- Mokhtari, P. (2003). “Parameterisation and control of laryngeal voice quality by principal components of glottal waveforms”, Journal of the

Phonetic Society of Japan (special issue on phonetics and speech technology), Vol. 7, No. 3, pp.40-54.

- ニック キャンベル (2004) 「声質：パラ言語を持つ第四次元の韻律パラメータ」くろしお出版の「文法と音声 I V」, pp. 25-34.
- Sadanobu Toshiyuki (2004) “A natural history of Japanese pressed voice”. Journal of the Phonetic Society of Japan, pp.29-44, Vol.8, No.1.
- 定延利之 (2004) 「日本語のりきみ」くろしお出版の「文法と音声 I V」, pp. 35-52.
- 定延利之 2003 「体験と知識：コミュニカティブ・ストラテジー」, 『國文學：解釈と教材の研究』, 第48号, 第12号, 東京：學燈社, pp. 54-64.
- 松本恵美子 2004 「オノマトペに使われる尻上がり調イントネーション：準備的考察」, 『大阪樟蔭女子大学 日本語研究センター報告』, 第12号, pp. 41-53.
- 舛田剛志, 戸田智基, 川波弘道, 猿渡洋, 鹿野清宏, “韻律的に多重化した音声データベースの設計と発話速度におけるその評価,”電子情報通信学会論文誌, Vol. J87-D-II, No. 2, pp. 447--455 (2004-2)
- Aubergé V., Cathiard M. (2003), Can we hear the prosody of smile ? Special Issue on Emotional Speech, 40, Speech Communication Review.
- Rilliard A & Aubergé V (2003), Prosody evaluation as a diagnostic process: Subjective vs. objective measurement ». International Journal of Speech Technology, Kluwer Academic Publishers

(2) 特許出願

H15年度特許出願件数：3件 (CREST研究期間累積件数：4件)