

感情表現が可能な音声合成を用いたコミュニケーションエイドの試作

飯田朱美* Srinivas Desirazu† 伊賀聡一郎‡ Nick Campbell† 安村通晃‡

*国際医療福祉大学国際医療福祉総合研究所

〒107-0062 港区南青山 1-1-1 新青山ビルディング西館 22 階

† (株) ATR 音声翻訳通信研究所

〒619-0288 京都府相楽郡精華町光台 2-2

‡ 慶応義塾大学環境情報学部

〒252-8520 藤沢市遠藤 5322

E-mail: a-iida@irihw.ac.jp

あらまし 話者の感情が反映された音声コーパスを ATR の波形接続型音声合成システム CHATR の話者データベースに利用することにより、感情のこもった合成音声を作成することができる。本研究ではこの技術を利用して感情表現が可能なコミュニケーションエイド *Chatako* の試作を行った。本研究では、ユーザーは感情を表すアイコンを選択することにより、喜び、怒り、悲しみといった感情情報を含んだテキストを入力することができる。これにより、ユーザーは複雑な操作なしにテキストに感情情報を記したタグを付与して、ユーザーの表現したい感情がこもった音声を合成することができる。アイコン選択のみで表現したい感情を入力できることから、障害者のみならず、パソコンに不慣れな人や高齢者にも利用可能である。

キーワード コミュニケーションエイド, 感情, 音声合成, 音声コーパス, 話者感情タグ

A Prototype of Communication Aid using Speech Synthesis with Emotion

Akemi Iida* Srinivas Desirazu† Soichiro Iga‡ Nick Campbell† Michiaki Yasumura‡

*International Research Institute of Health and Welfare, International University of Health and Welfare
Shin-Aoyama Bldg. 22F, 1-1-1 Minami-Aoyama, Minato-ku, Tokyo 107-0062 Japan

† ATR Interpreting Telecommunication Research Laboratories,
2-2 Hikaridai Seika-cho, Soraku-gun, Kyoto 619-0288 Japan

‡ Faculty of Environmental Information, Keio University
5322 Endo Fujisawa, Kanagawa 252-8520

E-mail: a-iida@irihw.ac.jp

Abstract This paper reports on a prototype system of a communication aid that can convey users' emotion using emotional speech corpora as source database for CHATR, a concatenated speech synthesis system developed at ATR-ITL. With this system, users can input texts with emotional information such as joy, anger and sad by selecting an icon which expresses emotion. By doing so, users can add tags which carry emotional information and can synthesize emotional speech of their preference. Since this system can be handled with simple operation as just selecting an icon, it can be used not only by people with disabilities but also by the elderly people.

Key words Communication Aid, Emotion, Speech Synthesis, Speech Corpora, Speaker-Emotion Tag

1 はじめに

「障害者にもできるだけノーマルな生活状態に近い生活をつくりだすこと」というノーマライゼーション理念は1995年にデンマークで発祥し、その後、我が国でも紹介されたが、福祉先進国に比べ、我が国のノーマライゼーションは遅れていた[1]。しかしここ数年、我が国でも障害者のための物理的・心的バリアの排除、地域での共生や社会的自立の促進に目が向けられるようになり、現在、政策的、技術的支援が進められている。

コミュニケーション支援機器も障害者の社会参加にかかせない機器であり、様々なものが開発されている。音声で話すことが困難な人のためのコミュニケーションエイドもその一つである。コミュニケーションエイドには非電子デバイスやVOCAと呼ばれる携帯用意思伝達装置もある。これらの多くは会話に想定される音声を録音して登録しておき、必要時にユーザーが対応したボタンを押す装置が多かった。しかし、最近では音声合成システムを利用したパソコン搭載型のコミュニケーションエイドも増えてきている[2]。

これらのシステムで利用されている合成音声はこの十年で処理速度も速くなり、目覚しい進歩を遂げた。特に価格面においては、¥10,000以下のものも販売されるようになり、一般ユーザーにも気軽に購入でき、今日スクリーンリーダーなど視覚障害者支援に活用されている。しかしながら、一般に合成音声の品質は話し口調が単調で、人間の音声とはまだ大きな隔りがある。筆者らはその理由として、従来の合成音声はコミュニケーションにとって重要な「感情」が十分に反映されていないことが一因であると考えている。感情を伝えたり共有することは人にとって大きな喜びである。感情は顔の表情に表れると同様に音声にも表れる。この音声の感情表現によって、同じ言語情報でも受け取られ方が異なり、時には言語情報以上に送り手の心

の状態を受け手に伝えることができる。

昨年、筆者らはATRの波形接続型音声合成システムCHATRと、話者の感情が反映された音声コーパスを用いることによって、感情表現が可能な合成音声を作成した[3]。本稿では、この技術を応用して試作したコミュニケーション・エイドChatakoについて報告する。

2 コミュニケーションエイド Chatako

本システムは話すことが困難な人のコミュニケーション支援を目的としたパソコン搭載型のコミュニケーションエイドである。本システムの特徴は男性、女性のそれぞれの音声について喜び、怒り、悲しみの3感情の音声を出力出来ること、および、ユーザーが感情を表すアイコンを選択することにより、テキストに話者と感情情報を記したタグ（話者感情タグ）を付け、合成時にそのタグ情報に沿って、該当する感情データベースをロードし、ユーザーの表現したい感情を担った合成音声を出力することにある。

2.1 システムイメージ

本研究のシステムイメージは、話すことが困難なユーザーが本システムを携帯型パソコンに搭載し、自由に街に出て、街の人と触れ合い、社会参加が出来るというものである(図1)。最近の携帯パソコンは800g以下のものが市販されるようになり、ユーザーは重量に対する負担感なく手軽に携帯できる。パソコンにオンスクリーンキーボードを搭載し、プッシュスイッチなどの1入力マウスを使って、本システムを利用することを想定している。

また、本システムをメール送受信機能と連動させれば、実空間のコミュニケーション支援にとどまらず、遠隔コミュニケーションツールとして応用することもできる。

2.2 システム構成

本システムの構成を図2に示す。本システム

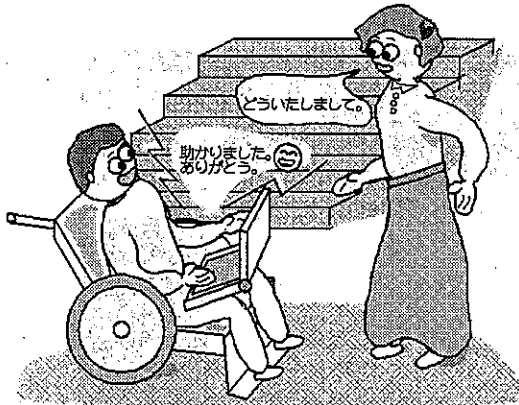


図1 システムイメージ

は Windows 95/98 上で動作する。音声合成システムは同 OS で動作する ATR 音声翻訳通信研究所の波形接続型音声合成システム CHATR Windows 版を使用した。CHATR の話者データベースとして男性、女性の喜び、怒り、悲しみの合計 6 つの音声コーパスを用いた。ユーザーとのインタフェース部分は TCL/TK 8.1 で記述した。

2.3 音声合成システム CHATR

本システムで採用した音声合成システム CHATR は音声波形の生成にあたっては、信号処理を行わずに大規模音声コーパスから最適な波形素片を選択し接続する。この方式により、韻律制御を直接行うことなく、音声単位選択処

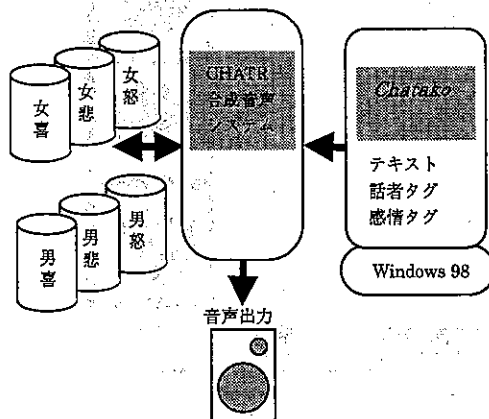


図2 システム構成

理により、韻律を伴う合成音声を生産できる。本システムでは感情が変化することに異なる感情音声コーパスを用いることによって、各感情音声を合成した。CHATR の詳細な合成手法については Campbell 他[4]、声村[5]に詳しい。

2.4 感情音声データベース

今回は男性、女性の喜び、怒り、悲しみの 3 種類、合計 6 セットの音声コーパスを話者データベースとして用いた。これらの感情音声コーパスの作成にあたっては、それぞれの感情がよく表れている文章を収集・創作してテキストコーパスを作成し、男性 1 名、女性 1 名が感情を込めて朗読したものを収録した。各データベースの規模は、文章数は約 400 文、音素数では約 30,000 音素である。作成した合成音声の感情を学生に聴取実験により判別させた。そして、有意水準 1% で検定を行ったところ、それぞれの感情は有意に判別され、この手法で作成した合成音声は感情を伝えることができることを確認した[3]。

2.5 Chatako のユーザーインタフェース

本システムの初期画面を図 3 に、話者選択後の画面を図 4 に示す。ユーザーインタフェース設計では、次の点に配慮した。

- 1) 画面レイアウトはシンプルにする。
- 2) 機能・ボタンは必要最小限に留める。
- 3) ユーザーの操作も最小限に留める。
- 4) 感情を扱うに相応しい配色やイメージの工夫をする。

2.5.2 話者アイコンと感情アイコン

男性と女性のアイコンを画面の右上に配置した。そして、怒り、喜び、悲しみの感情選択アイコンをその下に配置した。感情選択アイコンの背景色は怒り：赤、喜び：黄色、悲しみ：青とした。初期画面では背景色のみ表示されておの背景色は怒り：赤、喜び：黄色、悲しみ：青

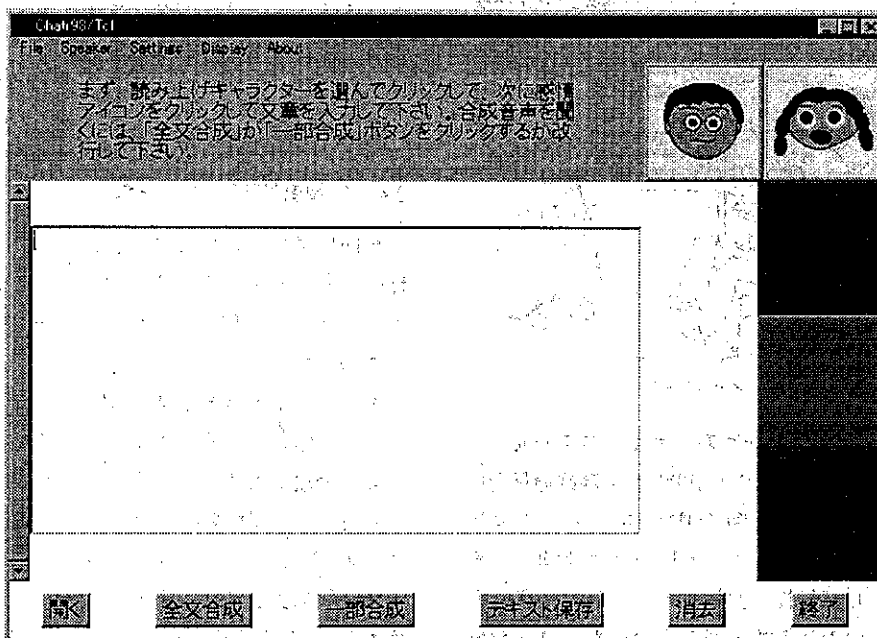


図3 初期画面

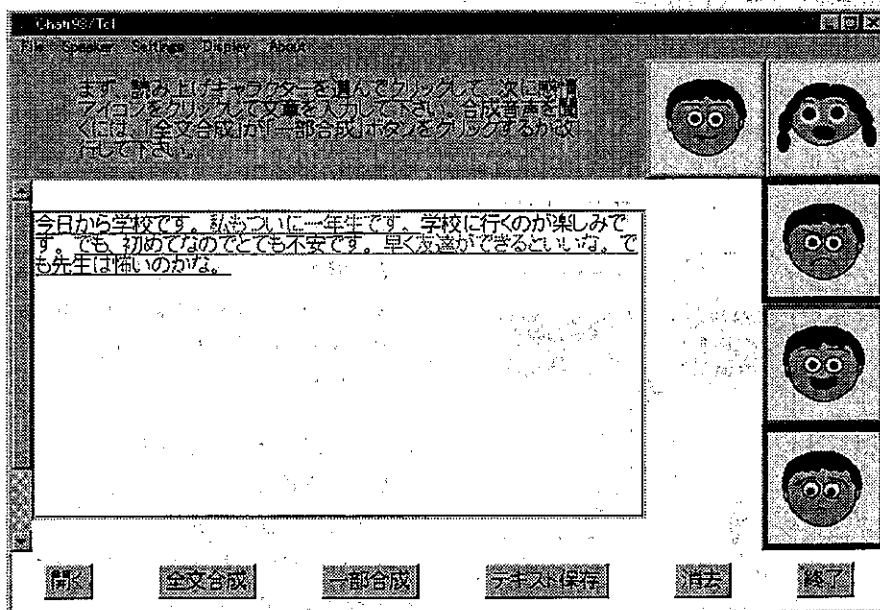


図4 話者選択後の画面

とした。初期画面では背景色のみ表示されており、話者が選択されると、その話者のそれぞれの感情を表すアイコンが表示される。

2.5.3 テキスト入力ウィンドウ

感情選択アイコンの左側にテキストウィンドウを配置し、スクロールバーを付けた。

2.5.4 コマンドボタン

プルダウンメニューはパソコン操作に慣れていない人には使いにくく、操作の回数も増えるため、テキスト入力以外は画面に表示したボタンをクリックすることで全ての処理ができるようにした。機能が多ければ、それだけ実現できるタスクも増えるが、反面、ユーザーにとって学習事項が増える。そのため、今回は作成した合成音声の保存など「あると便利だがなくても良い機能」は設けず、「開く」「全文合成」「一部合成」「テキスト保存」「消去」「終了」を設けた。

2.6 *Chatako* の基本仕様

本システムの基本仕様を操作手順に沿って説明する。

2.6.1 話者選択

男性または女性のいずれかのアイコンをクリックすることで選択できる。今回のキャラクターは *Chatashi* (男) と *Chatako* (女) である。

2.6.2 感情選択

話者が選択されると、各感情選択アイコンには、その話者のそれぞれの感情を表すアイコンが現れるので、ユーザーは合成したい感情をその中から選択する。すなわち話者に男性を選択すると、怒りのアイコンには怒っている男性、喜びのアイコンには嬉しそうな男性、悲しみのアイコンには悲しそうな男性の顔が表れる。

2.6.3 感情テキストの入力

本システムの日本語テキスト入力は Windows 95/98 上で動作する任意の仮名漢字変換ソフト

を利用して入力し、テキスト修正も一般のテキストエディタと同じようにカーソルを移動して行うことができる。現在の仕様では、ユーザーはテキスト入力前に話者選択と感情選択を行う。このユーザーの操作により話者感情タグが設定される。従って、ユーザーが文章入力中に話者や感情を変えたい時、そこで話者選択や感情選択を行うことによって、その都度話者や感情を新たに設定することができる。タグ付けされたテキストは内部的に保持され、ユーザーインタフェース上では色と下線により区別される。話者の区別は男性音声に下線を付けることで区別した。感情の配色は感情選択アイコンと同様、怒り：赤、喜び：黄色、悲しみ：青とした。またカーソルでマークすることにより、既に入力済みのテキストの話者・感情タグを変更することも可能にした。

2.6.4 感情テキストの合成

入力された話者感情タグ付きテキストの合成方法は利用場面に合わせて3通り用意した。

- 1) 全文合成：入力されたテキストを全て合成。
- 2) 部分合成：カーソルマーク部分のみ合成。
- 3) 改行による合成：改行入力時点で合成。

1) は「全文合成」ボタンをクリックすることで実現できる。作文の読み上げなどまとまった話をする時や一度ファイルに保存してからまた開いて合成する場合などに利用できる。2) は「部分合成」ボタンをクリックすることで実現できる。修正部分の確認や何度も同一内容を音声合成したい時に利用できる。3) では改行すると自動的に改行直前行のテキストが合成される。これにより、本システムはユーザーが実空間で人と対話している際のコミュニケーションエイドとして利用できる。

2.6.5 テキスト保存、テキストオープン

作成した話者感情タグ付きテキストは「テキスト保存」ボタンをクリックすることでテキス

トファイルとして保存でき、「開く」ボタンをクリックすることでロードすることができる。ロードしたファイルの話者感情タグ情報はユーザーインタフェース上ではテキスト入力時と同様に色と下線により区別され、表示される。図5にテキストファイルとして保存された話者感情タグ付きテキストの例を示す。

```
TAGMale:angry 今日から学校です。
TAGMale:happy 私もついに一年生です。
TAGMale:angry 学校に行くのが楽しみです。
TAGMale:sad でも、初めてなのでとても不安です。
TAGMale:happy 早く友達ができるといいな。
TAGMale:sad でも先生は怖いのかな。
```

図5 保存されたテキストファイル

2.6.6 画面消去と終了

「消去」ボタンをクリックすると、テキストボックス内のテキストは消去される。「終了」をクリックすると、*Chatako*を終了する。

2.6.7 評価と改善

実際のフィールド評価に先立って合成速度の評価を行った。最初の実装では、各感情が要求される度に該当する感情データベースをロードし、その部分の合成が終わるとアンロードしていた。しかし、これでは主観的にも非常に遅く感じられ、実利用は難しいと感じられた。そこで、各感情データベースは一度ロードしたら、システム終了までアンロードしないように変更した。その結果、図4の文章では合成に33秒かかっていたところ、19秒に短縮され、処理速度が1.7倍になり、フィールド評価が可能なレベルに達したと考えている。

3 今後の予定

本システムに1入力スイッチ、オンスクリーンキーボードなどの入力補助装置を接続して、実際に話すことが困難な障害者に試してもらい評価を行いたいと考えている。また、今回実装した機能の他に作成した合成音声の保存や再生、

コマンドのキーバインドなども評価時に要望を調査する予定である。

4 おわりに

本稿では感情表現が可能なコミュニケーションエイド *Chatako* について報告した。本研究はATRの波形接続型音声合成システム CHATR と感情が反映された音声コーパスを用いて感情を反映した音声合成を出力できること、および、入力テキストに感情タグを付けることで、ユーザーの表現したい感情音声を合成したことが特徴である。今後はフィールド評価を行い、システムの実用化を目指す。

謝辞

本研究のために多くのご支援を頂きましたATRのPatrick Davin 研究技術員ならびに慶應義塾大学大学院政策・メディア研究科の樋口文人氏に感謝致します。

参考文献

- [1] 福祉士養成講座編集委員会編集: “障害者福祉論”, 中央法規出版 (1999).
- [2] こころリソースブック:
<http://www.jeida.or.jp/document/kokoroweb/>
- [3] 飯田朱美, ニック・キャンベル, 安村通見: “感情表現が可能な合成音声の作成と評価”, 情報処理学会論文誌, Vol.40, No.2, pp.479-486 (1999).
- [4] Campbell, W.N. and Black, A.W.: “CHATR: 自然音声波形接続型任意音声合成システム”, 電気通信学会技術研究報告書, SP96-7, pp. 45-52 (1996)
- [5] 芦村和幸: “マルチメディア時代の音声合成—多言語音声合成システム CHATR”, *Interface*, pp. 109-118 (1998, Aug.)