

対話支援のための感情音声合成システムの試作と評価

飯田 朱美*¹ 伊賀 聡一郎*² 樋口 文人*³ ニック キャンベル*⁴ 安村 通晃*³

A Prototype of a Speech Synthesis System with Emotion for Assisting Communication

Akemi Iida,*¹ Soichiro Iga,*² Fumito Higuchi,*³ Nick Campbell*⁴ and Michiaki Yasumura*³

Abstract — This paper reports on a prototype of a speech synthesis system which can convey users' emotion (joy, anger and sadness) by using emotional speech database with CHATR, a concatenated speech synthesis system at ATR-ITL. Target users of this system are people with speaking difficulties. Perceptual Experiments were conducted with university students and the target users to evaluate the identification rate for the speaker-intended emotion, intelligibility, preference and so forth. The results proved to be positive for all experiments conducted. This system can be extended to a communication aid when equipped with input assistive device for the target users.

Keywords : Speech synthesis with emotion, Emotional speech database, User interface, Assisting Communication, Perceptual Experiments

1. はじめに

近年、音声処理技術の進歩とともに合成音声の品質も向上し、意思伝達支援として、コミュニケーションエイドや合成音声システムが実用化されるようになったが^[1]、より豊かな表現力のためには、感情情報の付与が重要である^[2]。著者らはこれまで音声を用いたユーザーの対話支援を目的として、感情音声データベース（コーパスとも呼ばれる^[3]）を作成し、それを基に感情表現可能な合成音声（感情合成音声）を作成する手法を提案している^{[4][5]}。本研究では、新たに男声の感情音声データベースを作成し、男声女声の喜び、怒り、悲しみの各感情を含んだ合成音声を作成できるようにし、対話支援を目的とした音声合成システム Chatako を試作した。本システムはノート PC 上で実現されており、話すことが困難な人々の意思伝達装置としての応用が可能である。

本稿では、2.で試作システムの設計指針について、3.で本システムの仕様について述べ、4.で試作システムの評価について報告し、最後に今後の課題と展望について述べる。

2. 試作システム Chatako の設計指針

本システムは感情音声合成を実装し、その対話支援としての有効性を確認するための試作システムである。

対象ユーザーは、主に話すことが困難な人々の利用を想定している。例えば、ALS（筋萎縮性側索硬化症）、筋ジストロフィーなどの神経筋難病^[6]などの患者が挙げられる。神経筋難病の場合、徐々に全身の筋肉の萎縮と筋力低下を起し、その過程で呼吸筋麻痺のため、気管切開を行い人工呼吸器を装着する。気管切開を受けた場合、話すことが困難になることが多い。また、脳血管障害、脳腫瘍、脳損傷、脊髄損傷、脳性麻痺の患者も脳の損傷を受けた部位によって、話すことが困難になる場合がある^[7]。

ユーザーインターフェースの設計においては、上記のようなユーザーを想定し、簡単な動作で操作できるように配慮したが、本システムの実際の利用にはユーザーの障害に合わせた補助入力装置の利用が必要である。本試作システムでは補助入力装置の提供は行わない。

本システムでは、Shaver が挙げる 4 つの最上位感情（喜び、怒り、悲しみ、驚き）のうち^[8]、他感情との切り分けが困難と思われる驚きを除いた喜び、怒り、悲しみの 3 感情を合成のための感情音声として試験的に選定した^[4]。

3. 試作システム Chatako

ここでは、本システムのシステム構成ならびに各構成要素について説明する。

3.1 システム構成

本システムは Microsoft Windows 95/98 上で動作し、感情音声合成部と、話者や感情の選択やテキスト入力操作を行うユーザーインターフェース部から成っている。その構成を図 1 に示す。

感情音声合成部には、Windows で動作する ATR 音声

*1: 慶応義塾大学 SFC 研究所

*2: 日本学術振興会特別研究員 (COE)

*3: 慶応義塾大学政策・メディア研究科

*4: (株)ATR 音声言語通信研究所

*1: Keio Research Institute at SFC, Keio Univ.

*2: Research Fellow of the Japan Society for the Promotion of Science(COE)

*3: Grad. School of Media and Governance, Keio Univ.

*4: ATR Spoken Language Telecom. Research Labs.

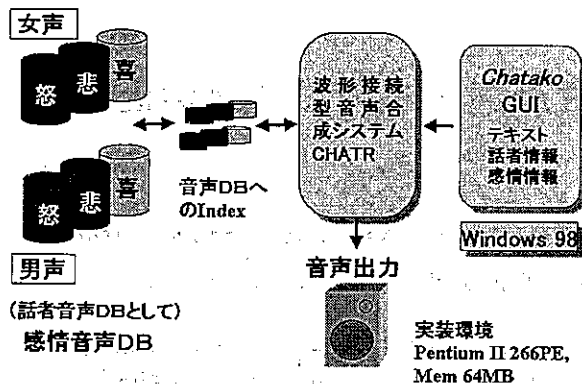


図1 本システムの構成
Fig.1 The System Configuration

翻訳通信研究所の波形接続型音声合成システム CHATR Windows版を利用した。CHATRの話者データベースとして男性、女性それぞれの喜び、怒り、悲しみの合計6セットの感情音声データベースを装備した。ユーザーインタフェース部はインタプリタ型スクリプト言語である Tcl/Tk 8.1を用い実装した^[9]。

3.2 感情音声合成部

音声波形を合成する方式には、波形編集方式、分析合成方式、ターミナルアナログ方式、声道アナログ方式などがある^[10]。このうち、分析合成方式は、1960年代から韻律パラメータ分析を中心として、感情とその音声表出の研究に用いられている^{[11][12]}。本研究では、波形接続型音声合成システムが、読み上げ音声をそのままデータベースとして用いることに着目し、話者の感情が反映された音声データベースを作成し、感情表現が可能な合成音声を作成する。本稿では、この提案方式を感情音声データベースに基づく波形編集方式と呼ぶ。

3.2.1 音声合成システム CHATR

本研究では CHATR を音声合成エンジンとして用いた。CHATRは、大規模な自然音声波形データベースから、音声合成単位の音声波形を前後の音素環境、韻律的情報などの情報とともに切り出し、波形辞書として蓄積する。そして、音声合成時には、音韻環境がテキストの音韻処理結果と最も合致する波形を選択して接続する^[13]。CHATRのテキスト音声合成処理の流れを図2に示す^[14]。CHATRでは自然音声波形そのものを用いるため、人間味のある音声という観点での品質は高い。

3.2.2 感情音声データベースの作成手法

感情音声データベースの作成にあたっては、まず、読み上げ用の感情テキストデータベースを作成した。収集にあたっては、朗読者にできるだけ、特定の感情を一定時間持続させるために、複数人の対話形式ではなく、モノログ形式のものを収集することにした。そして、まとまりのある1話題を単位として、原著者の喜び、怒り、悲しみの感情がよく表れていると感じられた文章を、既

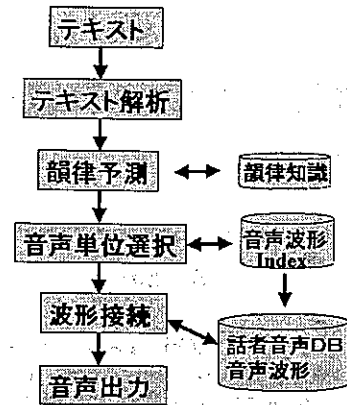


図2 CHATRにおける音声合成処理の流れ
Fig.2 Speech Synthesis Method of CHATR

に出版されているものから収集した。収集したテキストを、極力朗読者が原著者と同じ感情のもとで朗読できるように、テキストに感情語の補完などの修正を加えた。

各音声データベースの規模は、喜び: 461文章、40,916音素、怒り: 495文章、39,171音素、悲しみ: 426文章、31,840音素であった。朗読された音声は DAT に録音し、16kHz、16bit でデジタル化した。現在までに女性話者1名と、今回新たに収録した男性話者1名分の喜び、怒り、悲しみの3種類、合計6セットの感情音声データベースを作成した。

3.2.3 感情音声データベースの音響的特徴

各感情音声データベースを構成する音素ごとに、基本周波数、パワー持続時間を求め、それぞれの話者の感情音声データベースごとに平均と標準偏差を求めた。その結果、平均基本周波数は男声女声ともに悲しみ(男声 124.1Hz、女声 242.9Hz)、喜び(男声 149.8Hz、女声 256.6Hz)、怒り(男声 154.2Hz、女声 262.4Hz)の順で高くなり、有意水準5%で分散分析を行った結果、三者は有意に異なった。これは先行研究を裏付ける結果である^[11]。一方、平均継続時間は、女声は、怒り<喜び<悲しみの順で長くなったが、男声は、悲しみ<喜び<怒りの順であった。またパワーは、女声は、3感情とも同レベルであったが、男声は、怒り<喜び<悲しみの順であり、この2特徴においては、2者間に一致した傾向は見られなかった。

3.3 ユーザーインタフェース部

ここでは、画面レイアウトならびに画面の各構成要素について説明する。

3.3.1 画面レイアウト

本システムの画面レイアウトを図3に示す。画面レイアウトとしては、中央にテキスト入力ウィンドウを配し、その右側に話者アイコン(男女)、感情アイコン(怒り、喜び、悲しみ)を配置した。画面下にはコマンドボタンを配置し、ユーザーは、入力したテキストの操作や合成方

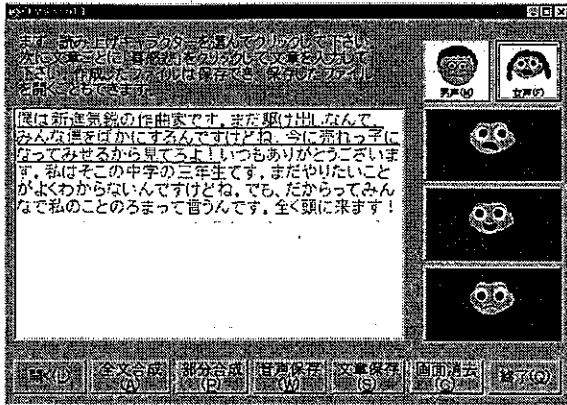


図3 Chatakoの画面レイアウト
Fig.3 User Interface of Chatako

表1 話者感情タグ付きテキスト
Table 1 Text with speaker-emotion tag

TAGMale:happy	僕は新進気鋭の作曲家です。まだ駆け出しなんてみんな馬鹿に思ってるんですけどね。今に売れっ子になってみせるから見てよ！いつもありがどうございませう。私、府中の中学の三年生です。まだやりたいことがよくわからないんですけどね。でも、だからってみんなで私のことをろまってるんです。全く頭にきます！
TAGMale:sad	まだ駆け出しなんなんですけどね。
TAGFemale:angry	全く頭にきます！

法を選択を行う。今回の試作では、画面左から「開く」、「全文合成」「部分合成」「音声保存」「文章保存」「画面消去」「終了」を設けた。これにより、ユーザーは全操作の選択肢を見ることができる。「音声保存」ボタンにより、音声は16kHz、16bitのwav形式で保存される。

3.3.2 話者感情選択とテキスト入力手順

本仕様では、ユーザーは、話者と感情を選択してからテキスト入力を行う。1) ユーザーは合成したい性別の話者を画面右上の話者アイコンから選択する。2) すると、話者アイコンの下に選択した話者の感情アイコンが、上から怒り、喜び、悲しみ(背景色は怒り:赤、喜び:黄緑、悲しみ:青)の順で現れ、ユーザーは表現したい感情を選択する。3) テキスト入力ウィンドウにテキストを入力する。日本語入力は任意の仮名漢字変換ソフトを利用して行い、テキスト修正はカーソル移動により行う。また、ユーザーが文章入力中の任意の時点で話者や感情を変更したい場合には、そこで話者選択や感情選択を行うことによって、新たに設定できる。

3.3.3 話者感情情報の保持と表示

ユーザーが選択した話者や感情の情報はテキストにタグ付けされて、システムが内部的に保持するが、ユーザー自身がタグ情報に関して直接意識することはない。ユーザーが話者と感情の種類を選択すると、テキストウィンドウに入力される文字のスタイルと色が決定される。話者は男性音声に下線付け、女性音声に下線なしという形で表され、感情は感情アイコンと同様、怒りは赤、喜びは黄緑、悲しみは青で表される。またカーソルでテキストを部分的にマークして、再度、話者や感情の選択を行えば、既に入力済みのテキストの話者感情タグを変

更することができる。テキストウィンドウに表示されているテキストとそのタグ情報は「テキスト保存」ボタンを選択することによって、話者感情タグ付きテキストとして保存される。その例を表1に示す。

3.3.4 感情合成音声の作成

入力された話者感情タグ付きテキストの音声合成の方法は利用場面に合わせて3通り実装した。

- **全文合成:** 入力されたテキストを全て音声合成する。まとまった話をする時や保存してあるファイルから音声合成する場合などに利用できる。
- **部分合成:** カーソルマークされたテキスト部分のみ音声合成する。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。
- **改行による合成:** 改行入力時点で音声合成する。改行すると改行直前までのテキストが音声合成される。本機能はユーザーと他の人との実空間での対話支援を想定している。

4. 評価

ここでは、感情音声合成部ならびにユーザーインタフェース部の評価について報告する。

4.1 感情合成音声の評価

ここでは、本システムの感情音声合成部で採用している感情音声データベースに基づく波形編集方式(以下、本方式)で作成した感情合成音声の聴取実験を行った。まず、聞き手の感情判別ならびに文章了解度について確認し、その後、合成音声の好感度および感情がこもっていると感じられる度合(以下、感情認知度)について確認した。

4.1.1 実験1: 感情の判別

この実験では、聞き手が話者の感情を、話者が意図したように受け取ったかどうかを確認する。男声女声の感情合成音声データベースの完成時期が異なるため、実験は2度に分けて行っている。被験者はそれぞれ大学生18人(男性10名、女性8名)である。文章資料としては、過去にCHATRを用いて合成された文章からランダムに5文を選択した。これをもとに音声資料として、喜び、怒り、悲しみの各感情音声データベースを用いて、男声、女声各15文の合成音声を作成し、音声ファイルに保存した。本実験に用いた文章を表2に示す。被験者にはヘッドフォン(SONY MDR-NC20)を着用させ、音量レベルを一定に保った音声を聞かせるようにした。音声資料ファイルをノートPC(Toshiba Dynabook SS3300)上でランダムな順に各文を1回ずつ再生し、話者の感情を3選択肢(喜び、怒り、悲しみ)から強制的に1つ選択させた。

感情が正しく判別された割合は、男声では、喜び52%、怒り51%、悲しみ74%であった。女声についての実験

表2 実験1に使用した文章

Table 2 Sentences used in Experiment 1

1.	チャターは色々な声でしゃべることのできる新しい音声合成のシステムです。
2.	あー、疲れた。
3.	あれー、風邪引いたみたい。
4.	「おまえ、人間は顔やないで」というところを「おまえの顔は人間やないで」と言うてしもた。あほやなー。
5.	ある日、お婆さんが川で洗濯をしていると、川上から大きな桃がドンブラコ、ドンブラコと流れてきました。

表3 実験2で使用した文章と合成音声

Table 3 Sentences used in Experiment 2

文章の種類	文章	Group A	Group B
1. Uncommon	X) あきら君は山桜とかき鍋と言いました。	Emo-M Sad	Dist-M
2. Number	X) シドニーと東京の時差は1時間ですので、今は朝の9時30分です。	Emo-F Angry	Dist-F
3. Emotion	X) わーい、やっと来るんだね。頑張ってきて良かった。	Emo-M Joy	Dist-M
4. Uncommon	Y) 山本さんは懐中電灯とクレパスと言いました。	Dist-F	Emo-F Joy
5. Number	Y) 8/15の日曜日から8/16の月曜日、1泊2日です。	Dist-M	Emo-M Angry
6. Emotion	Y) あー、疲れた。初めてなので、とても不安です。	Dist-F	Emo-F Sad

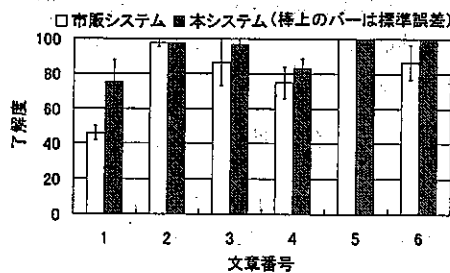


図4 実験2: 文章ごとの文節了解度

Fig. 4 Phrase Intelligibility per sentence

結果は報告済みであるが、喜び51%、怒り60%、悲しみ82%であった^[4]。有意水準1%で検定を行ったところ、女声同様、男声も3感情の合成音声において、感情は有意に判別された。

本実験の結果から、聞き手は、男声女声の両合成音声において、話者の感情を話者自身が意図したように受け取ったとすることができる。

4.1.2 実験2: 文章了解度

次に、市販の音声合成システム(以下、市販システム)を比較対象としながら、本方式で聞き手が文章の内容を聞き取れるかどうかを確認する。被験者は12名(男性7名、女性5名)を各6名ずつA、Bの2群に分けた。文章資料としては、数文節から成る文章で、数字を含む文章(Number)、共起が連想されにくい名詞の組合せを含む文章(Uncommon)、感情語を含む文章(Emotion)をそれぞれ2文章、合計6文章を使用した。これらを表3のようにX、Yの2群に分けた。

音声資料として、文章資料をもとに本システムと市販システム上で合成音声を作成した。本実験の比較対象となる市販システムはPC上で動作するテキスト入力型(Text-to-Speech)音声合成システムである。合成に際しての諸設定は次の通り。抑揚は9レベル中の7、音量は8レベル中の1、声の高さは8レベル中の5、スピード

は9レベル中の6とした。これらの設定は実験者が最も自然と感じるレベルに基づいている。男声女声の設定は本方式で作成した感情合成音声に合わせた。

文章の違いが合成音声システムの了解度評価に影響を及ぼさないように、3文章ずつ、A群とB群とで文章と合成システムの組合せを反転させた。すなわち、A群用には、X群の3文章は本方式(Emo)で、Y群の3文章は市販システム(Dist)で作成し、B群用にはその逆とした。作成した合成音声は両システムの音声保存機能を用いて、16kHz、16bitのwavファイルとして保存した。

本方式には男女の別に加えて、さらに感情の別があるが、全ての組合せを被験者に呈示することは、被験者の負担が大きくなりすぎる。そのため、本方式の感情合成音声については、被験者が全ての感情ならびに男女の音声を同数ずつ聞くように保ちながら、文章と感情・性別の組合せを限定したものに留めた。尚、被験者の負担を軽減するために行った、文章と感情・合成音の性別の組合せの限定については、同一文章に対する了解度の予備実験を行い、男女、喜び、怒り、悲しみの感情の間で有意差がないことをt検定および分散分析を用いて確認した。

各被験者に対して、これらの文章をランダムな順で2回ずつ呈示し、全文章を書き取るよう指示した。内容書き取りの開始時については被験者の任意とした。実験機材ならびに実験環境は実験1に準じる。

文章了解度の評価手法には、聞き取った内容を被験者に書き取らせ、音節明瞭度や単語了解度を使って定量的に評価する手法や^[15]、被験者に数分程度の各種の文章を聞かせ、内容についての質問をし、その正解をもって評価する手法などが^[16]報告されているが、本実験では、文節(文を実際の言語音声として出来るだけ多く区切った最も短い1区切り)という単位^[17]での正解をもって文章了解度を算出した。

対話支援のための感情音声合成システムの試作と評価

表4 実験3で使用した文章
Table 4 Texts used in Experiment 3

男喜び:わーい、やっとなるんだね。頑張ってきて良かった。
男怒り:全くどう考えても危ないのに、誰がそんなことをしたんだ。
女悲しみ:あー、疲れた。夜も眠らずに絵を描き続けて、気がついたら朝になっていたんだ。
3感情:(男喜び)僕は新進気鋭の作曲家です。(男悲しみ)まだ駆け出し なんでみんな僕をばかにするんですけどね。(男怒り)今に売れっ子になって みせるから見てろよ!(女喜び)いつもありがとうございます。私はその中学の 三年生です。(女悲しみ)まだやりたいことがよくわからないんですけどね。 (女怒り)でも、だからってみんなでのるまで言うんです。全く頭に來ます!

図4に示す通り、文章ごとの了解度は全文章において、本方式の方が回答率は高かった。被験者12人全員の全回答の平均は本方式92.1%、市販システム81.9%で、5%有意水準でt検定を行った結果、両者には有意差は見られなかった。また、文章の種類という観点から正解率を見ると、本方式、市販システムともに、Uncommon<Emotion<Numberの順で高くなっている。

本実験の結果により、本方式により作成した感情合成音声は市販システムの合成音声と同程度の文章了解度を有していると言える。また、本実験と同じ手順で、感情移入を特別に考慮していない音声データベースを用いてCHATRで合成した音声との了解度比較実験も合わせて行ったところ、平均は本方式が92.5%、CHATRが97.5%で、こちらについても、5%有意水準でt検定を行った結果、有意差は見られないという結果が得られ、本方式の感情合成音声はCHATRで他の音声データベースを利用して合成した場合と同程度の文章了解度が得られることがわかった。

4.1.3 実験3: 主観評価

ここでは、本方式の感情合成音声に対する聞き手の好感度、および、感情認知度を確認する。被験者は身体に障害のない大学生143名(男性69名、女性74名)である。文章資料としては、喜び、怒り、悲しみの感情が文脈に表れている文各1文と3感情が混合した複数文からなる文章を用いた。本実験に用いた文章を表4に示す。

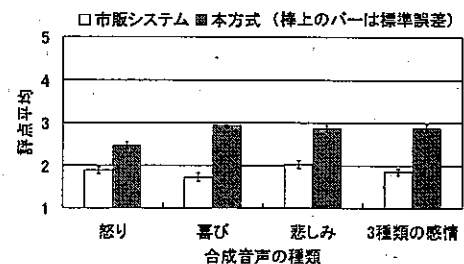
音声資料は上記の文章資料を基に試作システムと市販システムを用いて合成音声を作成した。本実験で比較対象となる市販システムは4.1.2で採用した市販システムと同社製の音声合成LSI搭載のコミュニケーションエイドである。設定レベルは実験者の聴覚的印象において市販システムの出力と同レベルにした。抑揚は5レベル中の4、音量は8レベル中の3、声の高さは8レベルの5、スピードは4レベル中の3である。男声女声の設定は本方式で作成した感情合成音声に合わせた。作成した合成音声は本方式のものについては試作システムで音声保存し、市販システムの合成音声はマイク(SONY F-V610)を通して16kHz、16bitのwavファイルとしてノートPCに保存した。これをノートPCにスピー-

カー(SONY SRS-38)を接続して被験者に聞かせた。実験者の聴覚的印象において、4.1.2の市販システム合成音声と比較して劣化はない。

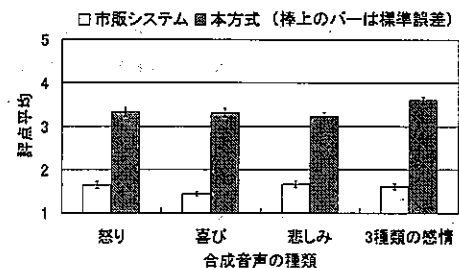
各文章ごとに二つの方式による合成音声をランダムに順番を変えて、被験者全員に聴取させた。その後、本方式と市販システムで作成した合成音声のそれぞれについて、好感度、感情認知度を、5段階(5=非常に良い、3=どちらでもない、1=非常に悪い)で評価させた。

評価の結果に対して、評定尺度間の等間隔性を仮定して^[15]、各合成音声に対する評定の平均と標準誤差(標本平均の標準偏差)を求めた。その結果、図5に示す通り、好感度ならびに感情認知度の双方において、評点平均は本方式の方が市販システムよりも高かった。

本実験の結果から、本方式による感情合成音声を用いて、実際に感情を表現することが可能であり、好感度も高いことが確認できた。



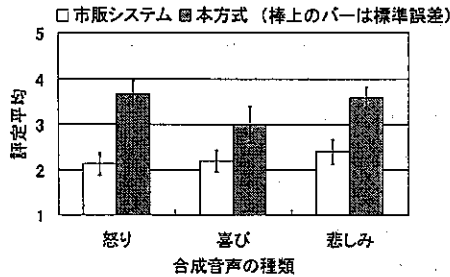
(a) 好感度



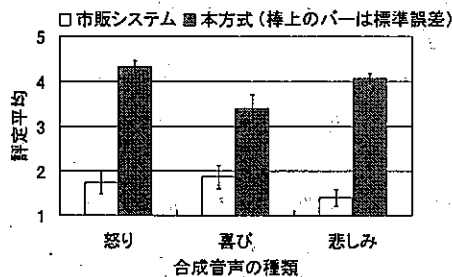
(b) 感情認知度

図5 実験3: 主観評価結果(大学生)

Fig. 5 Result of Experiment 3



(a) 好感度



(b) 感情認知度

図6 実験4: 主観評価結果 (ALS患者)
Fig.6 Result of Experiment 4

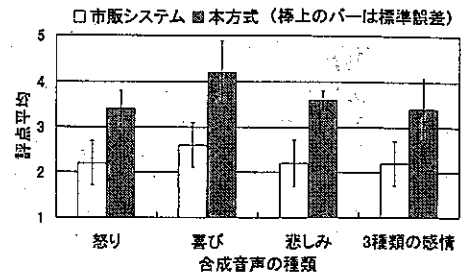
4.1.4 実験4: 対象ユーザーの評価

ここでは、対象ユーザーの本方式の感情合成音声に対する評価を行う。被験者は46歳～61歳のALS患者5名(男性3名、女性2名)である。文章資料は表4の文章を利用し、そのうち、3感情が混合した文章については1文章ずつに分け、合計9文章、各感情は3文章ずつとなった。

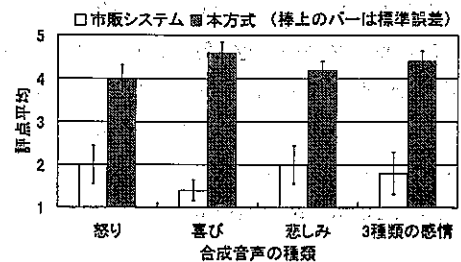
音声資料は本方式の合成音声ならびに4.1.3で使用したコミュニケーションエイドの出力音声を使用した。3感情混合文章の合成音声は1文ずつに切り分けて使用した。実験方法は1名ずつ直接対面方式とし、合成音声の出力は実験3と同様の機材を使用した。患者の回答は実験者が書き取った。各文章ごとに合成音声の表示順を変えて示し、各合成音声ごとに、感情判別、了解度、好感度、感情認知度について回答を得た。了解度については、実験者が各文章を大きなフォントで打ち出した紙を呈示しながら、聞き取れなかった単語を確認する形をとった。

本方式により感情が話者の意図したように受け取られたかについては、実験の結果、喜び66.6%、怒り93.3%、悲しみ86.6%という結果になった。有意水準1%で検定を行ったところ、それぞれの感情合成音声において、感情は有意に判別された。

了解度については、本方式が91.7%、市販システムが85.2%の正解率であったが、5%有意水準でt検定を行っ



(a) 好感度



(b) 感情認知度

図7 実験5: 主観評価結果 (視覚障害者)
Fig.7 Result of Experiment 5

た結果、両者に有意差は見られなかった。好感度、感情がこもっていると感じる度合いについても、図6に示す通り、本方式の評点平均の方が市販システムよりも高かった。

本実験では感情が文脈から受け取れる文章を用いたため、感情判別実験の判別結果には少なからず、文脈の影響はあるが、皮肉など特殊な意図の下以外では、話者の感情は文脈と一致することが多いため、本実験結果は有効と思われる。文章了解度の実験結果についても、4.1.2のような綿密な実験ではないが、直接対面方式で実験者が確認を行いながら、聞き取りを行ったため、有効と考える。

4.1.5 実験5: 視覚障害者による評価

本試作システムの対象ユーザーではないが、日常音声合成システムを使用する視覚障害者の評価を得る。被験者は20～40歳代の視覚障害者5名(男性3名、女性2名)である。文章資料および音声資料は実験3と同様のものを使用した。音声資料は、男性1人については実験4と同様の直接対面方式で呈示したが、他の4人についてはWeb上に音声ファイルを呈示し、質問事項と選択肢を記載したアンケートに電子メールで回答させた。

本実験の結果を図7に示す。実験3ならびに実験4と同様、好感度、感情認知度の双方において、本方式の感情合成音声の評点平均の方が市販システムの出力音声よりも高かった。

対話支援のための感情音声合成システムの試作と評価

被験者にヒアリングした結果、ドキュメント読み上げに合成音声を使用する場合は感情よりもスピードが重要視されるが、友人や親戚などとの遠隔対話支援としては、感情合成音声も有用であることがわかった。

4.2 ユーザーインタフェース部の評価

ここでは、対象ユーザーによるシステム構成および合成音声における優先順位に関する主観評価を行った。

4.2.1 システム構成と構成要素の評価

構成要素および画面レイアウトについて対象ユーザーの評価を得ることを目的とした。被験者は、実験4の被験者である ALS 患者 5 名である。評価項目は次の通りである。

- 機能は満たしているか。
- 画面レイアウトはわかりやすいか。
- 画面は好ましいか。
- 感情表現可能という点は良いか。
- 感情の選択方法はわかりやすいか。
- 感情の表現方法はわかりやすいか。

試作システムでは補助入力装置の提供を行わないため、今回の評価は被験者がシステムを操作するのではなく、実験者が説明しながら、操作を見せた。評価は5段階主観評価とし、被験者の回答は実験者が書き取った。

評価項目の評点平均を表5に示す。評点平均は全評価項目において、5点中の4以上であった。

評価結果から試作システムは対象ユーザーの要求を概ね満たしていることがわかった。ユーザーからは追加機能として次の要望が挙げられた。1) クリップボードの内容を再生できるようにすること、2) 保存した音声を本システムから呼び出せるような登録ボタンの装備、3) テキスト入力と登録ボタンとの連動、4) 保存内容の検索システムの装備などである。また、本システムの画面レイアウトは年齢や知能に合わせて、何種類か用意してほしいという要望もあった。

表5 ユーザーインタフェースの評価
Table 5 Evaluation on User Interface

評価項目	評点平均
機能は満たしているか	4.6
画面レイアウトはわかりやすいか	4.8
画面は好ましいか	4.2
感情表現可能という点は良いか	4.0
感情の選択方法はわかりやすいか	4.6
感情の表現方法はわかりやすいか	4.6

4.2.2 重視すべき特性

本試作システムにおいて、意思伝達支援を目的とした場合に、重視される特性を探るために主観評価を行った。被験者は実験4と同じ被験者である ALS 患者 5 名である。次の各項目について、重要度を5段階で答えてもらった。評価方法は4.2.1と同様である。

- 意味内容の正確な伝達

- 感情表現
- 即時性

評価の結果は、5点満点で、意味内容の正確な伝達が4.8、感情表現が4.4、即時性が3.4であった。

これにより、「意味内容の正確な伝達」が最重要項目ということを確認できた。その上で、「感情表現」も要望が高かった。即時性については、登録ボタンなどを本システムに追加していくことでユーザーのニーズに対応できる予定である。これまでの実験で被験者からは指摘はないが、市販システムがほぼ即時的に合成音声を出力するのに対して、今回の試作システムは5秒ほどの時間を有するなど、処理に時間がかかっている。これについては、実装上の問題であり、改善できる見通しである。被験者からは、怒りの音声も要望がある一方、怒りは他者と共有できない感情のため、むしろおもしろい音声を使って、怒りの文章を読み上げられるようにしてほしいという要望もあった。

5. 今後の課題と展望

試作システムでは大規模な感情音声データベースを参照しているが、これらを音韻ならびにアクセントパターンのバランスを取りながら、小規模化すること、および、現在参照している複数のデータベースを各感情に対応したタグを音素単位に付与することにより、1つのデータベースにまとめることなどが課題として挙げられる。試作システムの応用として、次の2つのシステムを検討している。

意思伝達装置

今回の対象ユーザーの評価を反映しながら、本システムを補助入力装置と組合せることで、話すことが困難な人々の意思伝達装置として実装する。本システムにユーザーが日常使う表現を録音音声や合成音声を使って登録し、テキスト入力との併用を可能にする。

遠隔対話支援

視覚障害者や高齢者の遠隔対話支援への応用が考えられる。例えば、本システムを電子メール送受信機能と連動させ、受信者が送信者の送信した感情話者タグ付きテキストを受信し、システムに読み上げさせることができる。この場合のユーザーインタフェース部については、視覚障害者や高齢者の利用を念頭に新たに設計する。

6. おわりに

話者の感情が反映された音声データベースを波形接続型音声合成システムの音声コーパスとして用いることにより、感情合成音声を作成できる。本研究ではこの技術を利用して、対話支援を目的とした感情音声合成システムを試作した。本システムで作成した感情合成音声を感情の判別、文章理解度、好感度、感情認知度などの側面

から評価した結果、その有効性を確認できた。また、対象ユーザーのユーザーインタフェース部の評価においても有用性が認められた。今後は今回の対象ユーザーの評価を反映しながら、話すことが困難な人々の意思伝達装置として実装する。また、長期的には、視覚障害者や高齢者の遠隔対話支援への応用も検討していく予定である。

謝 辞

本研究のために多くのご支援を頂きました ATR のパトリック・ダビン氏に感謝致します。また福岡県の山口進一氏をはじめ評価に御協力頂いた皆様、および、音声データのラベリングにご協力頂いた皆様に感謝致します。

参 考 文 献

- [1] こころりソースブック: 音声でコミュニケーション出来ない場合; <http://www.ibm.co.jp/kokoroweb/chap16/kkr16d1.html#02>
- [2] 山口進一: パソコンを使いこなそう!; <http://www.isn.ne.jp/kamata/ftp/ftp/jals.html>
- [3] 板橋秀一: 音声コーパス, 情報処理, Vol.38, No.11, pp.1012-1018 (1997).
- [4] 飯田朱美, ニック・キャンベル, 安村通晃: 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).
- [5] 飯田朱美, スリニバス・デジラズ, ニック・キャンベル, 伊賀聡一郎, 安村通晃: 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, 電子情報通信学会第二種研究会資料, WIT99-14, pp.83-88 (1999).
- [6] 神経筋難病情報サービス <http://www.saigata-nh.go.jp/nanbyo/>
- [7] Yorkston, K. M. ed., 伊藤元信監訳 富永優子訳: 拡大代替コミュニケーション入門, 共同医書出版社 (1996).
- [8] Shaver, P., Schwartz, J., Kirson, D., O'Connor, C.: Emotion Knowledge: Further exploration of a prototype approach, *Journal of Personality and Social Psychology*, Vol.52, No.1, pp.1061-86 (1987).
- [9] Welch, B. B., 西中芳幸, 石曾根信訳: Tel/Tk 入門第二版 (1999).
- [10] 広瀬啓吉: 音声合成技術, 情報処理, Vol.38, No.11, pp.984-991 (1997).
- [11] 市川薫, 中山剛, 中田和男: 合成音声の自然性に関する実験的考察, 日本音響学会講演論文集, pp.95-96 (1967).
- [12] 北原義典: 音声における韻律の役割とその応用に関する研究, 東京大学博士論文 (1996).
- [13] 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, 電子通信学会技術研究報告書, SP96-7, pp.45-52 (1996).
- [14] 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, インターフェース, pp.109-118, CQ 出版社 (1998.8).
- [15] 今泉敏: 声の聴覚心理評価; 声の評価法基礎編第二版 (日本音声言語医学会編), 医歯薬出版, 第6章 (1994).
- [16] 北原義典, 武田昌一, 市川薫, 東倉洋一: 音声言語認知における韻律の役割, 電子情報通信学会論文誌 D, Vol.70-D, No.11, pp.2095-2101 (1987).
- [17] 築島裕: 国語学, 東京大学出版会, 第3章 (1983).

(1999年12月6日受付, 2000年4月25日再受付)

著者紹介

飯田 朱美



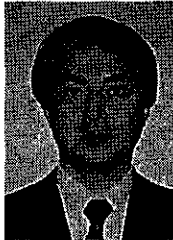
1985年上智大学外国語学部英語学科卒。1986年フロリダ州立大学大学院言語学修士課程修了。(株)NEC情報システムズ, 日本NCR(株)勤務を経て、1999年慶應義塾大学大学院政策・メディア研究科博士課程満期退学。現在、慶應義塾大学SFC研究所所員(訪問)。日本音響学会会員、情報処理学会会員。

伊賀 聡一郎 (正会員)



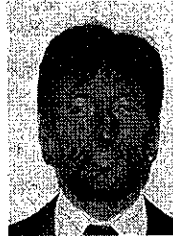
1999年慶應義塾大学大学院政策・メディア研究科博士課程修了。博士(Ph.D)。インタラクティブシステム、インタラクティブアート、デザイン、音楽等に関心をもつ。情報処理学会会員。現在(株)リコー勤務。

樋口 文人



1982年慶應義塾大学大学院工学研究科より工学博士号を取得。現在同大学大学院政策・メディア研究科博士課程在学中。情報処理学会、応用物理学会、IEEE Computer Society、ACM MRS、各会員。

ニック キャンベル



1992年サセックス大学大学院実験心理学博士課程修了。実験心理学(人工知能)博士。1986年イギリスIBM科学研究センター・フェロー、1989年エジンバラ大学主幹研究員、現在、ATR音声言語翻訳通信研究所主幹研究員。日本音響学会会員、ESCA Speech Synthesis SIG Vice Chairman, COCODA Coordinating Committee Member。

安村 通晃 (正会員)



1947年生。1978年東京大学理学系大学院博士課程(情報科学専攻)満了。(株)日立製作所中央研究所主任研究員を経て、現在、慶應義塾大学環境情報学部教授。理学博士。日本ソフトウェア科学会、日本認知科学会、日本教育工学会、情報処理学会、ACM、IEEE Computer Society 各会員。