

感情表現が可能な合成音声の作成と評価

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研究は、従来、分析合成方式の音声合成システムにおいて主に行われてきたが、今回我々は、話者の感情が反映された音声コーパスを用いて、感情表現が可能な合成音声を作成する手法を提案する。本研究では、喜び、怒り、悲しみの感情がよく表れている文章を収集・創作して、各感情の音声コーパスをそれぞれ作成した。そして、自然音声直接波形接続型音声合成システム CHATR を用いて、表現したい感情ごとに、その感情に対応する音声コーパスから最適な音声波形素片を選択し、接続することによって、合成音声を作成した。作成した合成音声の聴取実験を行い、有意水準 1% で検定を行ったところ、各感情は有意に判別された。本論文では、感情音声コーパスの設計方法について述べ、感情音声コーパスの朗読音声、および、それをもとに作成した合成音声の評価について報告する。

Design and Evaluation of Synthesized Speech with Emotion

AKEMI IIDA,[†] NICK CAMPBELL^{††} and MICHIAKI YASUMURA[†]

This research attempted to create emotional synthesized speech by designing emotional speech corpora which serve as source database for a concatenated speech synthesis system. Three corpora of emotional speech (joy, anger and sadness) have been designed for the use with CHATR, the concatenated speech synthesis system at ATR. A perceptual experiment was conducted using the synthesized speech generated from each emotion corpus and the result of the perceptual evaluation proved to be significantly identifiable. This paper describes the design strategy of emotional speech corpora and reports on results of perceptual experiments and acoustic analysis of emotional synthesized speech.

1. はじめに

近年、音声合成システムの多くは男声/女声の切替えや声の高さの調節が可能になり、従来の製品と比較すると、合成できる音声の種類が広がり、音韻性も向上した。しかし、依然として、音声合成の品質は人間の音声とはまだ大きな隔りがある。その理由の1つとして、コミュニケーションにおいて重要な役割を担っている感情が合成音声に反映されていないことがあげられる。もし感情表現が可能な音声合成が実現されれば、話すことが困難な障害者のコミュニケーション・エイド¹⁾への応用が期待できる。従来のコミュニケーション・エイドは、登録された音声しか出力できなかったが、最近では音声合成が利用され、任意の文

章を合成できるようになってきている。出力される合成音声ユーザの感情を伝えることができれば、より心の通ったコミュニケーションに近づくと思われる。また、感情表現が可能な音声合成は、高齢者にも有益と思われる。電話の応答メッセージや街中の音声ガイダンスなどに合成音声利用されているが、話者の感情が反映された合成音声情報が読み上げたり、応答したりすれば、高齢者の心理的抵抗感が薄れると考えられる。

感情とその音声表出の研究は、数種の感情の下で発声された音声の基本周波数、継続時間、パワーなどの韻律パラメータ分析を中心に、古くから行われてきており²⁾、それらの研究を基礎として、分析合成法によって、感情を含む合成音声の研究が進められてきた³⁾。市川ら⁴⁾は、平叙文の基本周波数と周波数振幅の実測値を基準として、発話様式の異なる音声(喜び、悲しみ、楽しさ、どなり声など)の値を基準値に対する比率として計算し、それをもとに合成音声を作成して聴取実験を行った。その結果、感情の知覚には、基本周波数が大きく関与することを示した。小林ら⁵⁾は、自然音

[†] 慶応義塾大学大学院政策・メディア研究科
Graduate School of Media and Governance, Keio University

^{††} 株式会社 ATR 音声翻訳通信研究所
ATR Interpreting Telecommunication Research Laboratories

声（喜び、驚き、怒り、悲しみ、平静）の基本周波数パターンと発話速度を比較した。基本周波数パターンに関しては、分析合成法を用いて各感情の合成音声を作成し、アクセント指令とフレーズ指令の強度を比較した。その結果、平均基本周波数は悲しみの音声で最も低く、基本周波数レンジも悲しみの音声で最も狭いことを明らかにした。また、発話速度に関しては、悲しみが怒りより遅いことを示した。北原⁶⁾は、喜び、怒り、悲しみの音声の韻律パラメータを制御した合成音声の聴取評価を行い、韻律特徴分析を行った。

最近では、声質変換については、入力音声をより明瞭性の高い音声に即時的に変換する方法が提案され、試作が行われているが⁶⁾、感情表現が可能な音声合成については、実現が難しく、試作には至っていない。そこで我々は、話者の感情が反映された音声コーパスを用いて、直接波形接続型音声合成システムにより、感情合成音声を作成する手法を提案する。本手法により、感情表現が可能な音声合成の実用化を目指す。

本研究では、喜び、怒り、悲しみの感情がよく表れている文章を収集・創作して、各感情を反映する音声コーパスをそれぞれ作成した。そして、それらを用いて、ATR 通信翻訳研究所の自然音声直接波形接続型音声合成システム CHATR⁷⁾により、表現したい感情ごとに、その感情に対応する音声コーパスから最適な音声波形素片を選択し、接続することによって、表現したい感情を含む合成音声を作成した⁸⁾。次章では、感情音声コーパスの設計方法について述べ、3章では、収集・創作したテキストの評価、および、感情音声コーパスの音響的特徴と評価について報告する。そして、4章では、感情合成音声の作成方法、および、感情合成音声の音響的特徴や評価について報告する。

なお、本論文では、感情音声を話者の感情が反映された肉声、感情合成音声を話者の感情が反映された合成音声とする。また、感情音声合成とは、感情合成音声を音声合成システムを用いて作成することを指す。そして、感情テキストコーパスとは、収集・創作したテキストに修正を加えた朗読用のテキストコーパス、感情音声コーパスとは、その感情テキストコーパスを話者に朗読させ、収録した音声コーパスのことを指す。

2. 感情音声コーパスの設計方法

2.1 本研究で対象とする感情

本研究では、以下に述べる Shaver⁹⁾の上位感情のうち、喜び、怒り、悲しみの感情を扱った。Shaverらは、感情タイプの境界は曖昧だとしながらも階層化を行い、135の感情サブタイプを25種類に分類した。そ

表1 感情テキストコーパス
Table 1 Details of text corpus.

感情	テキスト数(編)	文章数	音素数
喜び	12	461	40916
怒り	15	495	39171
悲しみ	10	426	31840

れらをさらに、喜び、怒り、悲しみ、驚きの4感情にまとめた。本研究では、驚きは喜び、怒り、悲しみと共に起る場合が多く、切り分けが難しいと考えられるため、対象外とした。各感情には下位感情が多く含まれるが、本研究では、それらの判別が目的ではないため、下位分類は追求しなかった。

2.2 規模

CHATR で使用する音声コーパスの理想的な規模は日本語の場合、約 30,000~50,000 音素であり、この規模であれば、日本語の音素環境と韻律環境をおおむね包含できることが経験的に分かっている¹⁰⁾。これに従い、本研究では約 30,000 音素を目標にした。各感情テキストコーパスの規模を表1に示す。本論文では、まとまりのある話題の開始から終了までを便宜的にテキストと呼ぶ。すなわち、テキストとは、新聞の場合は1社説、または、1コラム、日記の場合は1エッセイを指す。収集したテキスト長は、平均 1600 文字程度(A4紙1枚程度)であった。朗読者に特定の感情を一定時間持続させるため、複数人の対話形式ではなく、モノローグ形式のテキストを収集・創作した。

2.3 タスク設定/音素バランス

従来の音声コーパスは、タスク(発話内容)が設定されたものや、音素バランスが考慮されたものが多かった¹¹⁾。本研究では、汎用性を持たせるためにタスク設定を行うことはせず、各感情テキストコーパスには喜び、怒り、悲しみのそれぞれの感情を含むテキストを収集した。また、音素バランスとは、音素や音節の出現頻度や結合頻度がコーパス内で均衡を保っているかの指標であり、コーパスを設計するうえで重要だが、今回は、CHATRでの感情音声合成に必要な規模の感情音声コーパスを作成することが優先課題であったために、音素バランスについては考慮しなかった。

2.4 内容

従来の感情音声の研究では、研究目的が複数の感情間の音響のパラメータの比較に重点が置かれたため、異なる感情のもとで同一文章を発声させることが一般的であった¹²⁾。しかし、本研究では、朗読者が各感情音声コーパスを朗読しているときに、その感情を抱きやすいテキストを収集することに重点を置き、まず、コーパス設計者が、テキストの原作者の喜び、怒り、悲

表2 各感情音声コーパス例文

Table 2 Example sentences from each emotional corpus.

喜び:	全く手足の動かない私にも自分でやれることができたのです。「すごいぞすごいぞ!」「おい、こんなこともできるぞ! 見てるよ、いいか?」「ほら、もう一度やってみるからな! いや、御機嫌だよ、これ!」
怒り:	蕨北町が6月から突然ごみの超細分化を始める。燃えるゴミ、リサイクルできないプラスチックゴミ、缶瓶の類など。はっきり言って何をどう捨てていいのか訳が分からない。それなのに明日からもうスタートするというのだ。町中大混乱だというのに、町のお偉方は一体何を考えているのだ!
悲しみ:	中学2年の7月5日、丈夫だった父が突然この世を去った。その日の国語の時間、何かの偶然だとは思いますが、ちょうど私が父方の亡くなった祖父の話をしている時、父の異変を聞かされた。病棟に迎えに来た親戚と、取るものも取り敢えず我が家へ向かった。

しみが表れていると感じたテキストを新聞¹³⁾、World Wide Web、障害者の自費出版日記などの、すでに出版されているものから収集した^{14)~16)}。そして、収集したテキストを各感情のもとで創作したテキスト(喜び:1編、怒り:5編、悲しみ:1編)と合わせて、表1の規模にした。その結果、各感情音声コーパスを構成するテキストはそれぞれ異なるものとなった。

2.5 テキストの修正

感情テキストコーパスを朗読者が朗読するときに、朗読者が可能な限り、原著者と同じ感情のもとで朗読できるように、テキストに修正を加えた。各感情テキストコーパスの文章例を表2に示す。後述(3.1節)するように、一般的に、読者としてテキストを読んだ場合には、読書後に抱く感情は、原著者と同調するほかに数通り考えられるが、本研究では、朗読者は原著者の立場に立って朗読した。以下に主な修正項目を示す(詳細、および、具体例については付録を参照)。

- 感情を移入しやすくする工夫
- 感情語の補完による感情強化^{17),18)}
- 朗読者の感情を高揚させるリズム・テンポ作り

2.6 収録とデジタル化

従来の感情音声の研究では、同一文章を数種の感情のもとで発声することが多く、明確に感情を発声し分けられるとの理由から俳優やアナウンサーが朗読することが多かった¹²⁾。しかし、専門家に依頼するには費用がかかり、日程の調整なども必要である。本研究では、演技やナレーションの訓練を受けていない一般人でも、感情を込めて朗読できるように感情テキストコーパスを設計した。そして、すべての感情テキストコーパスを、一般人の女性話者1人が防音室で同一日に朗読した。DATに収録された音声は16kHz、16bitでデジタル化し、1文章ずつに分割した。

感情テキストコーパスの設計方針を以下にまとめる。

- (1) 種類:喜び、怒り、悲しみの感情を含む音声コーパスをそれぞれ作成。

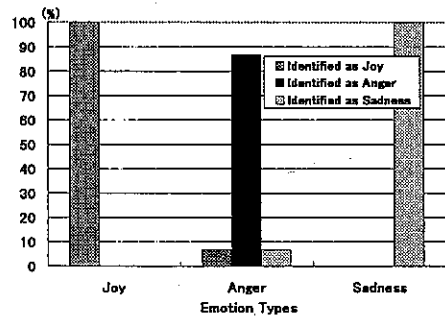


図1 感情テキストコーパスの評価

Fig. 1 Evaluation of text corpus.

- (2) 規模:それぞれの感情につき、約30,000音素。
- (3) タスク設定:汎用性を考慮し、行わない。対象とする感情を設定。
- (4) 音素バランス:特には考慮しない。
- (5) 内容:コーパスごとに異なる。原著者の感情がよく表れているテキストを採用。
- (6) 修正:できるだけ原著者と同じ感情を引き出し、それを持続させ、高揚させるために行う。
- (7) 朗読者:一般人が朗読練習を行ったうえで行う。

3. 感情音声コーパスの評価

3.1 感情テキストコーパスの評価

本研究で作成した喜び、怒り、悲しみの各感情テキストコーパスを構成するテキストが、文字情報の段階で、内容的にそれぞれの感情を表しているかを確認する実験を行った。表1に示すように各コーパスのテキスト数は、喜び:12編、怒り:15編、悲しみ:10編で合計37編である。これらを大学生74人に、1人2編ずつ読んでもらい、原著者が各テキストを書いたときの感情と思われるものを喜び、怒り、悲しみから3者択一してもらった。怒りのテキスト2回答以外はコーパス設計者の感情分類と一致した(図1)。

また、被験者自身は各テキストに対して、どんな感

表3 朗読音声の音響的パラメータ

Table 3 Acoustic characteristics of emotional human.

感情	基本周波数 (Hz)		継続時間 (ms)		パワー (*)	
	平均	標準偏差	平均	標準偏差	平均	標準偏差
喜び	256.6	52.9	64.8	31.4	6.8	0.6
怒り	262.4	57.3	66.1	28.6	6.8	0.6
悲しみ	242.9	40.0	73.4	31.8	6.8	0.6

(*単位: arbitrary unit)

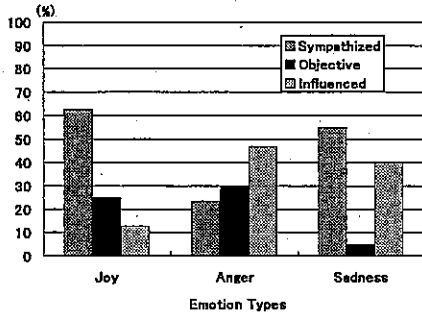


図2 テキスト読書後の読者の抱いた感情
Fig. 2 Emotion of text readers after read.

情を抱いたかを自由形式で回答してもらった。回答結果は、以下の3つに分類できた。1) 原著者と同調, 2) 客観的, 特別な感情を抱かない, 3) テキストの表す感情が向けられた者としての感情を抱く(怒りの文章の場合は怒られている気分になる)。その結果を図2に示す。喜びと悲しみでは1)と回答したものが多かったが、怒りは3)と回答したものが多かった。本実験の結果を受けて、朗読者は怒りのコーパスについてはいっそう原著者の立場に立って朗読するよう心がけた。

3.2 感情音声コーパスにおける朗読音声の音響的特徴

先行研究では、基本周波数⁴⁾は感情判別への寄与率が高く、継続時間⁵⁾も、悲しみを他の2感情から識別するのに寄与率の高いパラメータであると報告された。このことを確認するために、本研究で作成した感情音声コーパスにおける朗読音声の音響的特徴を分析した。

各感情音声コーパ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、感情音声コーパスごとの平均と標準偏差を求めた(表3)。その結果、以下が確認できた。1) 平均基本周波数は、悲しみ(242.9 Hz)、喜び(256.6 Hz)、怒り(262.4 Hz)の順で高くなり、有意水準5%の分散分析の結果、3者は有意に異なっていた。2) 平均継続時間は、悲しみ(73.4 ms)が、喜び(64.8 ms)、怒り(66.1 ms)よりも長かった。分散分析では、悲しみは、他の2感情と有意に異なっていたが、喜びと怒りには差が認められなかった。3) パワーは、3感情とも平均6.8、標準偏差0.6であった。

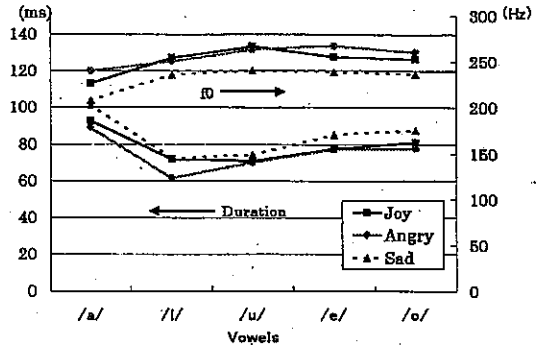


図3 母音のピッチと継続時間
Fig. 3 Duration and pitch of vowels.

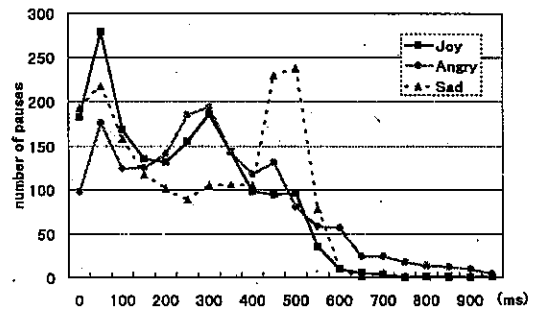


図4 ポーズの継続時間
Fig. 4 Duration of pauses within a sentence.

各母音ごとの平均と標準偏差を求めた結果も全音素を対象とした場合とほぼ同様であった(図3)。次に、文中ポーズの継続時間の分布を調べた。図4に階級間隔を50 msに設定した度数分布を示す。3者とも正規分布ではなく、喜び、怒りは全体的に右下がりの傾斜を成したが、悲しみは、50~100 ms間と500~550 ms間にそれぞれ大きなピークがあり、M字型の形状を示した。変動係数は、喜び(1.2 ms)、怒り(1.1 ms)に比較して、悲しみ(0.8 ms)は小さく、悲しみの音声コーパスの文中ポーズ長は、短い場合と長い場合とに集中し、他の2感情よりもばらつきが小さいことが認められた。

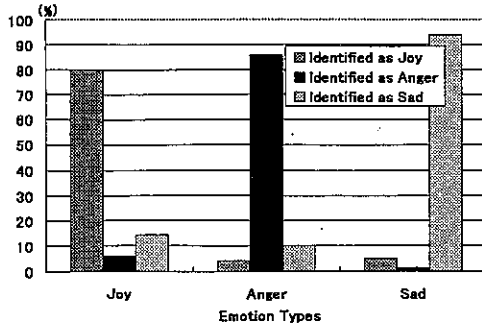


図5 朗読音声の聴取実験結果

Fig. 5 Evaluation of emotional human speech.

3.3 感情音声コーパスにおける朗読音声の評価

作成した感情音声コーパスの朗読音声を大学生 29 人に聞かせ、感情を判別してもらった。聴取実験では、文脈の影響を少なくするために、すべての感情音声コーパスの中からランダムに 1 文ずつ音声を再生し、被験者 1 人に 50 文ずつ朗読音声の表す感情を選択してもらった。被験者は 1 文提示されるたびに喜び、怒り、悲しみの中から 3 者択一した。3 種類の感情音声コーパスの合計は 1382 文であり、1 文につき 1 回答を得た。結果は、感情が正しく判別された確率は、喜び 80%、怒り 86%、悲しみ 93%であった(図 5)。有意水準 1% で検定を行ったところ、有意に判別された。

4. 感情合成音声の作成と評価

4.1 本研究で採用した音声合成システム

本研究では、自然音声直接波形接続型音声合成システム CHATR^{7),10),19)}により、感情が変化することにより、異なる感情音声コーパスを用いて、合成音声を作成した。CHATR は音声波形の生成にあたっては、信号処理は行わずに、大規模音声コーパスから最適な音声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、以下に述べる音素単位選択処理により、韻律をとともう合成音声を作成した。

● 音声コーパスの特徴抽出過程

音声コーパス中の全音素について、音素記号、基本周波数、継続時間、パワーを統計的に計算し、音声コーパスに記録する。

● 音響的、韻律的特徴の重み係数決定過程

重み係数とは、どのような音素環境のときにどのパラメータを優先してコスト計算を行うかを示したものである。音声コーパス全体の韻律パラメータの分布をもとに自動計算し、重み係数を音声コーパスに記録する。

● 単位選択コストを用いた音素単位選択処理

合成したい音素との差を表す target cost と、隣接音素単位間での不連続性を表す concat cost の和が最小になるような音素系列を音声データベースから探索する。今回の感情音声合成の実験では、この処理における重み付けを、target cost では、1) 先行、後続音素を含めた当該音素のケプストラム距離、2) 基本周波数、3) 継続時間の順で、concat cost では、1) 当該音素と隣接音素のケプストラム距離、2) 基本周波数の順で設定した。

4.2 感情音声合成に使用した文章

過去に CHATR を用いて合成された文章の中から、ランダムに 5 文を選び、本研究で作成した喜び、怒り、悲しみの各感情音声コーパスを用いて、CHATR により合計 15 文の合成音声を作成した。音声合成に使用した文章を表 4 に示す。

4.3 感情合成音声の音響的特徴

本研究で作成したすべての感情合成音声の各音素について、韻律パラメータを分析し、喜び、怒り、悲しみの感情ごとに、平均と標準偏差を求めた(表 5)。その結果、基本周波数、および、継続時間については、感情音声コーパスにおける朗読音声の音響的特徴の分析結果とほぼ同様の結果を得た。また、パワーについては、感情音声コーパスの音響分析では、感情差が認められなかったが、感情合成音声では、平均は、怒り(5.3)が喜び(6.2)や悲しみ(5.6)より小さかった。逆に標準偏差は怒り(2.2)が喜び(1.2)や悲しみ(1.5)よりも大きく、変動係数も怒りが最も大きかった。

4.4 音声合成システムの単位選択結果

今回の感情音声合成の実験では、4.1 節で述べた単位選択処理において、まず当該、先行、後続音素のケプストラム距離で単位候補を絞り込むため、結果としては、感情音声コーパスの中から 2 音素、3 音素連鎖で選択されることが多かった。CHATR では、もし感情音声コーパス中に合成したい単語とまったく同じ単語が存在し、韻律パラメータ、その中でも特に基本周波数において、合成したい音素の予測値と候補単位の実測値に近い場合には、単語そのものが選択される可能性もある。しかし、target cost, concat cost の両コストにおいて、基本周波数と継続時間も重み付けされており、予測値と実測値に差がある場合には、必ずしも連続する複数の音素が選択されるとは限らない。今回の感情音声合成の実験では、使用した 5 文章には自立語が 39 語あるが、その中で単語ごと選択されたものは、怒り:1 単語、喜び:1 単語、悲しみ:3 単語

表4 感情音声合成に使用した文章
Table 4 Sentences synthesized by CHATR.

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5 感情合成音声の音響的特徴
Table 5 Acoustic characteristics of emotional synthesized speech.

感情	基本周波数 (Hz)		継続時間 (ms)		パワー (*)	
	平均	標準偏差	平均	標準偏差	平均	標準偏差
喜び	243.8	38.8	83.2	84.7	6.2	1.2
怒り	256.2	41.0	81.8	72.6	5.3	2.2
悲しみ	231.1	31.8	93.8	93.1	5.6	1.5

(*単位: arbitrary unit)

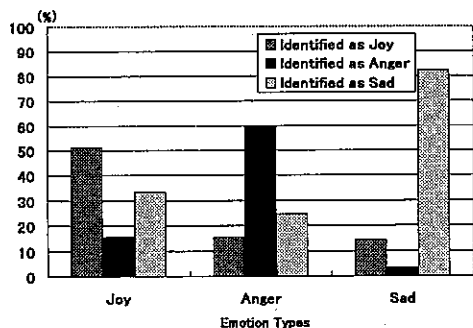


図6 感情合成音声の聴取実験結果

Fig. 6 Evaluation of emotional synthesized speech.

であった。表4の文章4には、「人間」という単語が2回出現し、怒りの音声合成実験では、前者は、その単語そのものが感情音声コーパスから選択されているが、後者は音声コーパス中の別の箇所から4単位選択されている。他の2感情では、どちらも単語ごとには選択されなかった。

4.5 感情合成音声の評価

本方式で合成した感情合成音声15文をランダム提示して、大学生18人に感情を判別してもらった。図6に示すとおり、感情が正しく判別された結果は、喜び51%、怒り60%、悲しみ82%であった。有意水準1%で検定を行ったところ、感情音声コーパスにおける朗読音声と同様に、感情は有意に判別された。被験者からは、CHATRの合成音声は「日本語を母語としない人の日本語のようだ」、「イントネーションが不自然だ」という感想が多かった。

4.6 感情合成音声の聴取実験に関する考察

先行研究では、基本周波数パターンが、感情を識別するのに有力なパラメータであることが示されたが^{(3),(4)}、今回の感情合成音声の聴取実験結果では、基

本周波数パターンが不自然でも、すなわち、人間の音声と異なっても、感情を識別できる場合があることが示唆された。CHATRは、人間の音声波形を直接接続するために、人間の音声そのものを生成できるが、合成したい音素の基本周波数予測エラーにより、イントネーションが不自然に聞こえる場合がある⁽²⁰⁾。これが4.5節で示した被験者の印象の一要因であると考えられる。たとえば、本実験では、表4のテキスト1の「チャター」「システム」「できる」などのアクセントパターンが適切に予測されなかった。前者2単語の頭高型アクセントに対して、合成音声は平板型を、「できる」の起伏型アクセントに対しては、頭高アクセントを予測した。しかし、この文章の合成音声の感情判別率を見ると、聴取実験に使用された5文章の中で最も高い(喜び:61%、怒り:83%、悲しみ:83%)。以上のことから、被験者は朗読者の感情を基本周波数パターン以外の音響的特徴からも判別していると示唆される。先行研究⁽⁵⁾や本研究の音響分析から、継続時間は、悲しみを他の2感情から差別化できるパラメータとして注目できることが確認されているが、今回の合成感情音声の音響分析結果(4.3節)から、その他にも、発話内のパワーの変動にも注目できることがうかがわれた。

5. 今後の課題

本研究では、朗読者1人により感情音声コーパスを収録したが、今後は、さらに朗読者を増やし、合成できる音声の種類を増やす予定である。

また、感情合成音声の応用例として、障害者や高齢者のための遠隔コミュニケーション支援システム(図7)を検討している。文章の入力はキーボード・エミュレータを介して行う。キーボード上のキーに、喜び、怒り、

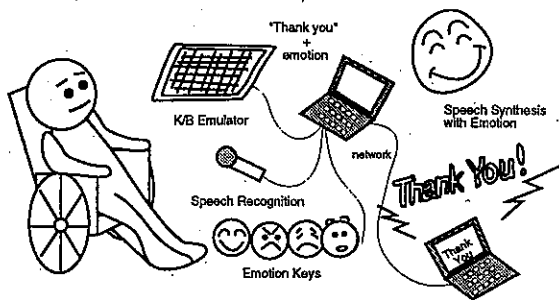


図7 システムイメージ
Fig. 7 System image.

悲しみが割り当てられ(感情キー), 送信者は入力文を送信するときに, 感情キーによって表現したい感情を選択して, 遠隔地の受信者に感情を含んだ文章を送信する。遠隔地の受信者は, 感情音声合成によって送信者の感情を反映した音声聞くことができる。これにより, 感情表現が困難な障害者や高齢者のコミュニケーションが円滑になることが期待される。

6. おわりに

本研究では, 原著者の感情を反映したテキストを収集・創作して喜び, 怒り, 悲しみの3感情の感情音声コーパスを作成し, それぞれを用いて, 自然音波形接続型音声合成システム CHATR により, 合成音声を作成し評価を行った。聴取実験による評価の結果, 作成した感情合成音声の正答率は有意水準 1% で有意に判別され, 本研究の有効性が認められた。今後は本手法による音声合成の実用化に向けて研究を行う。

謝辞 貴重なご意見をいただきました東京大学工学部の広瀬啓吉教授, ATR 音声翻訳通信研究所の芦村和幸研究技術員, 太田洋子氏, 慶應義塾大学大学院政策・メディア研究科の樋口文人氏, 伊賀聡一郎氏, および, 実験にご協力いただいた慶應義塾大学の皆さんに感謝いたします。

参考文献

- 1) <http://www.ibm.co.jp/kokoroweb/>
- 2) Murray, I.R. and Arnott, J.L.: Towards the simulation of emotion in synthetic speech: A review of the literature of human vocal emotion, *Journal of Acoustic Society of America*, Vol.93, No.2, pp.1097-1108 (1993).
- 3) 広瀬啓吉, 高橋 登, 藤崎博也, 大野澄雄: 音声の基本周波数パターンにおける話者の意図・感情の表現, 電気情報通信学会技術報告書, HC94-41, pp.33-40 (1994).

- 4) 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).
- 5) 小林 豊, 新美康永: 音声の感情を反映する韻律情報制御方式について, 日本音響学会秋季講演論文集, pp.233-234 (1993).
- 6) 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, 東京大学大学院工学系研究科 (1996).
- 7) Campbell, W.N. and Black, A.W.: CHATR: 自然音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).
- 8) Iida, A., Campbell, N. and Yasumura, M.: Emotional Speech as an Effective Interface for People with Special Needs, *Proc. Asia Pacific Computer Interaction 1998*, pp.266-271 (1988).
- 9) Shaver, P., Schwartz, J., Kirson, D. and O'Connor, C.: Emotion Knowledge: Further exploration of a prototype approach, *Journal of Personality and Social Psychology*, Vol.52, No.1, pp.1061-1086 (1987).
- 10) Campbell, W.N.: Processing a Speech Corpus for CHATR Synthesis, *Proc. The Int'l Conf. on Speech Processing*, pp.183-186 (1997).
- 11) 板橋秀一: 音声コーパス, 情報処理, Vol.38, No.11, pp.1012-1018 (1997).
- 12) Bance, R. and Scherer, K.: Acoustic Profiles in Vocal Emotion Expression, *Journal of Personality and Social Psychology*, pp.614-636 (1996).
- 13) 朝日新聞, いのち長き時代に, 1997年6月1日~12月20日。
- 14) 大平幸枝: 私らしく, 人間らしく, 神奈川中央出版社, 自費出版 (1995).
- 15) 轟木敏秀: 光彩—輝き続けるために—KB マウスで入力した筋ジストロフィー青年の記録, 自費出版 (1993).
- 16) 上村数洋: 明日を創る—頭脳損傷者の生活の記録, 三輪書店 (1990).
- 17) 国立国語研究所編: 国立国語研究所資料集, 分類語彙表, 大日本図書 (1964).
- 18) 国立国語研究所編: 国立国語研究報告 44, 形容詞の意味・用法の記述的研究, 秀英出版 (1972).
- 19) 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).
- 20) 藤澤 謙, 平井俊男, 樋口宣男: 波形接続型音声合成システム CHATR の基本周波数に関する音素単位選択基準の改良, 日本音響学会春季講演論文集, pp.219-220 (1997).

付録 感情音声コーパス中のテキスト修正

(1) 感情を移入しやすくする工夫

- 句点(.)を疑問文では疑問符(?)に、感嘆文では感嘆符(!)に変換:

例:(喜)「感謝感激だった!」

- 動作主が3人称の場合は1人称に修正:

例:(悲)「山本さんは母親を励ましたくて、来る日も来る日も絵はがきを故郷に送り続けた。」「私は母を励ましたくて……」

(2) 全体的な感情の統一

原著者の感情が一時的に他の感情に移行したことが見られる文章を元の感情を表す文章に差し替える。

例:(悲)「私のために懸命に明るくしてくれる主人や子供たちの前で泣くこともできず、いい母親を演じている自分が許せない。この仮面をぬぐことができない自分が腹立たしい。」「…いい母親を演じている。でも、1人になると切なくて涙が止まらない。」

(3) 感情語の補完による感情強化^{17),18)}

例:(悲)「その夜自責の念が押し寄せた。」「その夜初めて深い自責の念が押し寄せた。」

- 形容詞:美しい, 悲しい, わびしい, せつない, 嬉しい, 照れくさい, 腹立たしい, やるせない
- 程度を表す形容詞:はなはだしい, 著しい, 恐ろしい, えらい, 素晴らしい, ものすごい, ひどい, 深い
- 副詞:すごく, とても, 大変, 非常に, 甚だ, ひどく, ついに, ただただ
- 副助詞:さえ, こそ
- 感嘆詞:ああ, なんと, おお
- 形容動詞:きれいな, ばかな, 愚かな

(4) 朗読者の感情を高揚させるリズム・テンポ作り

- 語りかけ: 例(喜)「あなたを通して教えられました」→「あなたを通して教えてもらいました。ね、桂木さん。」
- その感情特有の表現を段落最後尾に追加: まったくだらない! ふざけるな!
例(怒)10年後には3,500円にもなるというではないか! けしからん!
- 段落末への感嘆表現追加: 例(怒)「自民党の『地元の強い要望』という根拠も、世論調査で慎重派が多数を占めたことで崩れた。」「……崩れたじゃないか。」

- 段落末へ感情文追加: 例(悲)「むくみが出た父親の顔は見る見るうちに別人のようになった。手や足も驚くほど腫れ上がった。」に、「私たち家族にとってそんな父の姿は痛ましく、いたたまれなかった。」を追加。

(平成10年6月3日受付)

(平成11年1月8日採録)



飯田 朱美 (学生会員)

1985年上智大学外国語学部英語学科卒業。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)を経て、現在、慶應義塾大学大学院政策・メディア研究科博士後期課程。日本音響学会会員。



ニック キャンベル

1985年ランカスター大学大学院自然言語学修士課程修了。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科学研究所センター・フェロー, 1989年エジンバラ大学主幹研究員, 1990ATR自動翻訳電話研究所客員研究員。現在, ATR音声翻訳通信研究所第二研究室室長。多言語音声合成用の大規模データベース作成, 韻律処理, 単位選択の研究に従事。ESCA Speech Synthesis SIG Vice Chairman, COCODA Coordinating Committee Member。



安村 通晃 (正会員)

1947年生。1971年東京大学理学部物理学科卒業。1975~1977年UCLA留学。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、1990年4月より慶應義塾大学環境情報学部助教授。現在、同教授。理学博士。ベクトルコンパイラ, Lisp処理系, マルチモーダルインタラクションなどの研究に従事。プログラミングシステム, 並列ソフトウェア, 実世界指向インタフェース等に関心を持つ。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, ヒューマンインタフェース学会, ACM, IEEE Computer Society 各会員。