

エネルギー科学研究科

エネルギー社会・環境科学専攻修士論文

アフェクティブインタフェース

題目： のための表情・音声合成による
感情伝達手法に関する基礎研究

指導教官： 吉川 榮和 教授

氏名： 伊藤 京子

提出年月日： 平成13年2月7日(水)

論文要旨

題目：アフェクティブインタフェースのための表情・音声合成による感情伝達手法に関する基礎研究

吉川榮和研究室，伊藤 京子

要旨：

近年の情報通信技術の発展から、パソコンを利用し情報を得ることにより様々な利益を享受することができるようになってきている。その一方で、パソコンに「難しそう」、「よくわからない」などの心理的抵抗を持ち、情報を得たくても得られない人が多数いる。このように情報を得る人と得られない人との間に格差が生じることを「デジタルデバイド」というが、これは今後進展すると予想される情報社会で、社会全体に不利益をもたらす大きな問題とされ、その解決が求められている。

本研究では、私たち人間が生まれながらに持っている「感情」に注目し、コンピュータが人間の感情を推定し、また人間に感情を伝える、すなわち感情の要素をコンピュータに取り入れて、人間にとって親しみやすいインタフェースを実現し、「デジタルデバイド」の克服に活用することを研究の目標とする。具体的には、感情を取り扱うインタフェース(アフェクティブインタフェース)を作成し、コンピュータが表現する感情を人間が理解し、表現力を感じることができるかどうかの被験者実験を行って、コンピュータが表現したい感情を人間に効果的に伝えるためのあり方を考察する。

従来の受け取った命令に従って、ただ単に仕事をこなすだけのコンピュータに比べ、インタフェースに感情要素を取り入れることにより、コンピュータに心理的抵抗を持つ人々のコンピュータへの親しみやすさが増すことが期待される。また、これを人間と接する機会の多い生活場面に取り入れて、コンピュータエージェントとして実用化すると、様々な分野への応用が期待される。例として、次世代インターネットを用いた家庭や事務所のホームオートメーション、オフィスオートメーションの領域で空調、調理機器などのエネルギー機器での省エネ管理を人間の代わりにしてくれるコンピュータエージェントへの応用や、教育分野で生徒に学習への動機づけを与えるシステムへの応用などが挙げられる。

本研究では、まず、「感情」をどのようにコンピュータに取り入れるべきかを考えるために、感情についての学説を展望して、感情の社会的、文化的要因を考察したのち、これを背景に内外のアフェクティブインタフェース研究の状況を展望して、アフェクティブインタフェース技術の人間との親和性を評価する方法の開拓が重要なことを論ずる。次いで、この観点から行った本基礎実験では、コンピュータから人間への感情伝達の可能性とあり方を調べるために、まず、発話内容に合わせた感情を、音声と顔表情を用いて表現する手法を検討し、検討した手法に基づく物語を朗読するシステムを試作した。試作したシステムを用いて被験者実験を行い、コンピュータが意図した感情を人間に伝えることができるか、の観点からの評価を行うとともに、システムに対してなじみやすさや親近感、すなわち親しみやすさを感じるか、またコンピュータが朗読した物語が表現する感情に共感できるかを評価した。実験結果より、コンピュータから人間に意図した感情を伝達する可能性が示唆され、今後の検討課題を見出すことができた。

目次

第 1 章 序論	1
第 2 章 研究の背景と目的	3
2.1 研究の背景	3
2.1.1 ヒューマンインタフェースの観点から見た現状のパソコンの問題点	3
2.1.2 コンピュータを意識させないコミュニケーションの実現に向けて	5
2.1.3 従来研究の展望	8
2.2 研究の目的	12
第 3 章 物語朗読システム構成のための基礎的検討	14
3.1 システムの全体構成と感情推定・感情表出方法の概要	14
3.2 感情とその分類	16
3.3 ことばから人間の感情を推定する方法の検討	18
3.3.1 感情推定方法の検討	18
3.3.2 ことばと感情の対応付け	19
3.4 物語文朗読による感情伝達手法の検討	24
3.4.1 感情を表現する音声パラメータの検討	26
3.4.2 感情を表現する音声パラメータ値の決定	26
3.5 顔表情合成による感情伝達手法の検討	26
3.5.1 顔表情合成方法と感情を表現する顔表情パラメータの検討	27
3.5.2 感情を表現する顔表情パラメータ値の決定	30
3.6 検討結果に基づく物語朗読方法の提案	33
3.6.1 感情推定方法の提案	33
3.6.2 感情表出方法の提案	38
3.7 まとめ	44
第 4 章 物語朗読システムの作成	46

4.1	ハードウェア構成	46
4.2	ソフトウェア構成	46
4.2.1	感情推定サブシステム	47
4.2.2	感情表出サブシステム	48
第 5 章	物語朗読システムを用いた被験者実験	51
5.1	実験の目的	51
5.2	実験方法	51
5.2.1	実験手順	52
5.2.2	被験者	58
5.3	実験の結果	59
5.3.1	合成音声のみの物語文朗読による感情伝達の評価	59
5.3.2	物語朗読システムを用いた物語文朗読による感情伝達の評価	60
5.3.3	各表情の感情伝達の評価	64
5.3.4	コンピュータに対する態度(操作不安と接近願望)調査	69
5.3.5	実験全体の感想	73
5.3.6	多面的感情状態尺度(MMS)	73
5.4	結果の分析と考察	79
5.4.1	コンピュータに対して操作不安を持つ人と持たない人に分類した結果の分析	79
5.4.2	実験全体の結果と考察	81
5.5	まとめ	84
第 6 章	結論	86
	謝辞	89
	参考文献	90

目 次

2.1	SONY の AIBO	6
2.2	トミーのファービー	7
2.3	オムロンのたま	7
2.4	松下電器産業の高齢者支援ペットロボット	7
3.1	システムの全体構成	15
3.2	感情の 3 層モデル	17
3.3	感情推定のデータの流れ	20
3.4	感情語分類の実験風景	24
3.5	3 次元顔モデル	28
3.6	モデル化した表情筋とその配置	28
3.7	インタラクティブ進化計算のシステム構成	30
3.8	合成顔表情を評価する実験の画面構成	32
3.9	顔表情の結果 (1)	35
3.10	顔表情の結果 (2)	36
3.11	顔表情の結果 (3)	37
3.12	感情の埋め込み例	38
3.13	感情推定のアルゴリズム	39
3.14	感情表出方法の概要	40
3.15	感情識別子挿入物語テキストの 1 区切り	41
3.16	リップシンクデータの作成例	41
3.17	各母音を発音している顔表情	42
3.18	感情表出の流れとデータ処理例	43
4.1	ハードウェア構成	46
4.2	感情推定サブシステムの構成	48
4.3	感情表出サブシステムの構成	49
5.1	実験の手順	53

5.2	アニメ的な顔画像 (A) とリアルな顔画像 (B)	56
5.3	課題 3 の実験の風景	61
5.4	課題 4 の実験の風景	65
5.5	アンケート B-4 の結果 (全体)	68
5.6	アンケート B-4 の結果 (正答者のみの伝達度)	70
5.7	アンケート B-5 の結果	73
5.8	MMS の結果 (各課題前後の差の平均)	77
5.9	MMS の結果 (1 と 2 の差)	78
5.10	MMS の結果 (3 と 4 の差)	78
5.11	MMS の結果 (5 と 6 の差)	78
5.12	操作不安により分類したアンケート B-3 の結果	81

表 目 次

2.1	体の部位と対応する要素技術	11
3.1	感情による生理変化の分類	16
3.2	それぞれの文化で、あらかじめ示された感情を正しく識別したエクマンらの被験者のパーセント値 ^[20]	17
3.3	35種類の感情	19
3.4	感情語テーブルの例	21
3.5	選択した語彙分類	23
3.6	感情語分類実験結果(単語数)	25
3.7	感情表出用音声テーブル ^[26]	27
3.8	表情合成に用いる表情筋の種類	29
3.9	遺伝アルゴリズムのパラメータ	32
3.10	実験の終了世代数	34
4.1	ハードウェア要求仕様	47
5.1	アンケート実施時とアンケート名	54
5.2	MMSの項目一覧	58
5.3	アンケート B-2 の結果	59
5.4	課題2の感想	60
5.5	アンケート B-3 の結果	62
5.6	課題3の感想	63
5.7	アンケート B-4 の出題内容と結果(1)	66
5.8	アンケート B-4 の出題内容と結果(2)	67
5.9	「コンピュータに対する態度」の因子分析結果 ^[35]	71
5.10	アンケート B-5 の結果	72
5.11	アンケート B-6 の結果(1)	74
5.12	アンケート B-6 の結果(2)	75
5.13	MMSの結果(各回ごと)	76

5.14 MMSの結果(各課題前後の差の平均)	77
5.15 操作不安により分類したアンケート B-3の結果	80

第 1 章 序論

近年、爆発的に普及し「一家に一台」にせまる勢いのパソコンは、昔の大型コンピュータと比較すれば、人がコンピュータを使う際の入出力機能は非常に便利になっている。しかし、コンピュータに詳しい知識がなくても使いやすい、というには程遠いのが現状である。その操作方法は、非常に複雑であり、操作の手順をひとつでも間違えば目的が達成できないので、操作手順を詳しく知る必要があり、使う側の人間に大きな負担がかかる。現状のパソコン用インタフェースは、Windowsを始めとするグラフィカルユーザインタフェース (Graphical User Interface, GUI) であるが、マウスでアイコンをクリックする GUI が使いやすいとは決していえない。誰もが使いやすい、使ってみたいと考えるパソコンにするためには、パソコンと人間との接点であるヒューマンインタフェースを新たな観点から構成する取り組みが必要である。そこで、人間の「感情」に着目して、感情の要素を取り入れたヒューマンインタフェース (アフェクティブインタフェース) の研究が最近着目されている。このようなアフェクティブインタフェースの意義として、人間的感情を示すインタフェースを介することによりユーザの親しみやすさが向上することが期待される。最近、ワイドバンドの次世代インターネットを用いて家庭や事務所などのホームオートメーション、オフィスオートメーションの領域での空調、冷蔵庫、調理機器、その他のエネルギー利用機器の省エネ管理用の電子情報技術の研究開発が産学界で取り組まれているが、省エネ管理を人間の代わりに行う電子機器制御用のヒューマンインタフェースにアフェクティブインタフェースを取り入れたコンピュータエージェントへの応用が考えられている。

本研究では、感情を取り扱うアフェクティブインタフェースを構成するための基礎的な問題として、「どのようにすればコンピュータが人間に共感をもたらさうか」を調べるために、人間に様々な感情を伝える物語の朗読をコンピュータが行うシステムを構成し、それを用いた被験者実験を行って、コンピュータを用いた顔表情の合成や合成音声によって、どの程度コンピュータが伝達したい目的の感情を人間に効率的に伝えられるか (感情を伝達できるか) を多角的に検討した。

以下、本論文の構成を述べる。

まず、第 2 章では、研究の背景として、現状の人とパソコンの間のインタフェース (パソコンインタフェース) の問題点を示し、コンピュータを意識させない、人とコン

コンピュータとのコミュニケーションのあり方について考察した後、本研究に関連してアフェクティブインタフェースの従来研究を展望し、以上を背景に本研究の目的を示す。

次に、第3章では、まず、感情の社会的、文化的背景を展望し、本研究での感情分類方法を述べた後、合成音声で物語を朗読しながらそれに合わせた合成顔表情を提示するシステムを構成するために、物語を記述したテキスト(ことば)からテキスト中で表現される感情を推定する方法、物語文朗読による感情伝達手法、顔表情合成による感情伝達手法を順に検討し、検討結果に基づく方法を提案する。

そして、第4章では、提案した手法に基づいて試作した物語朗読システムのハードウェア構成、ソフトウェア構成について説明する。

さらに、第5章では、試作した物語朗読システムを用いて行った被験者実験について、その目的、方法、実験の結果と考察について述べる。

最後に、第6章で本論文のまとめと今後の課題を述べる。

第 2 章 研究の背景と目的

近年、情報技術の発展に伴い、情報機器が市場に氾濫するようになっている。しかし、携帯電話、パソコン、AV 機器などの技術開発では、小型化、高機能化に重点がおかれ、使う側に立った操作性に重点がおかれていない、という反省の声も聞かれるようになった。

本章では、研究の背景として、まず、人間とコンピュータの接点となるヒューマンインタフェースの現状での問題点を展望して、それらの問題点を解決するために、コンピュータの存在を意識せず、誰もが使えるインタフェースの実現に向けての、本研究での着眼点を述べる。次いで、アフェクティブインタフェースの実現例として人間のような対応をする「擬人化インタフェース」に関する従来研究を展望し、それらの背景を基に、本研究の目的を述べる。

2.1 研究の背景

情報通信技術の急速な発展に伴い、インターネット上の膨大な情報へのアクセス、電子メールなど、従来からのワープロ、表計算に加えてパーソナルコンピュータ (以下、PC) の利用はますます多様化している。

本節では、まず、その利用方法の多様化により、近年爆発的に普及してきた PC を対象に、現状の PC インタフェースの問題点を挙げた後、誰もが使えるインタフェースの実現に向けて、本研究で着眼する新しいヒューマンインタフェースであるアフェクティブインタフェースとはどのようなものをまず説明し、それらに関連する従来研究について述べる。

2.1.1 ヒューマンインタフェースの観点から見た現状のパソコンの問題点

21 世紀は、情報の世紀と言われている。従来の価値は実体の伴う「もの」、すなわちハードに付与されている場合が多かったが、近年、実体の伴わない「情報」、すなわちソフトに付与される価値が大きくなってきている。このような価値観の変遷には、イ

インターネットを代表とする情報技術の発展が大きな影響を与えている。2000年度の国内でのPC出荷台数は約1000万台、世帯普及率は約28%といわれ、2003年には50%を超えると予想されている。このような市場活性化の要因としては、主にインターネットへの接続や電子メールという使用目的が生まれたことが挙げられる。PCを利用し、様々な情報を得ることができれば、私たちの生活がより豊かで快適になると期待されている。しかし、情報化社会が発展するにつれて、PCを用いて情報を取得できる人だけがよりよい環境を得ることができ、情報を得られない人との格差が生じることが懸念され、社会全体としての不利益が予想される。情報を得る人と得られない人の格差をなくすためには、PCを利用したくても利用できない人たちにPCを利用できる環境を提供することが必要となる。ここで、PCを利用したくても利用できない主な理由として、PCに対する心理的障壁が大きいということが考えられる。PCに対する心理的障壁により、人間は、「難しそう」、「触ったら壊れてしまいそう」、「問題が起こったときにどうしたらいいのかわからない」などの印象をPCに対して持ち、PCの使用に抵抗を感じる。このような心理的障壁を与えるPC側の問題点をヒューマンインタフェースに関係する以下の3つの観点から考える。

1. PCの入力装置のなじみにくさ
2. PCからの出力のなじみにくさ
3. PC自体のなじみにくさ

1. の入力装置がなじみにくい理由として、現状のPCの入力装置であるマウスやキーボードの操作が難しいことが挙げられる。アルファベットは文字総数が少ないため、欧米ではタイプライタが普及しているが、日本の場合、欧米と異なりキーボードの利用経験をもたない人が多い。一方、マウスを使用する場合、画面を見ながらのマウスの移動や、短時間でマウスのボタンを2回押す(ダブルクリック)など、思ったように操作ができないことがある。このように、マウスやキーボードの習熟には時間がかかり、PCへの入力の困難さからPCへの心理的障壁が生じることが予想される。

2. PCからの出力がなじみにくい理由として、コンピュータ分野に特有の用語を用いたメッセージが提示されることが挙げられる。例えば、『システムのログファイルをハードディスクに書き込めません』というメッセージが画面上に提示された場合、PCになじみのない人は、対処の方法がわからず、PCが壊れてしまったかもしれないという不安に陥る。このように、日常生活では見たことのないようなメッセージが提示されることにより、PCに対する恐れが生じ、心理的障壁が生じることが予想される。

3. の PC 自体がなじみにくい理由として、PC が機械的なものであることへの心理的抵抗が考えられる。これまでの PC には人間性が感じられず、無機質な機械のように思え、これまで日常生活で得た経験が適用できない PC に対して、自分とはまったく相容れないものと認識し、関わりたくないという心理的障壁が生じることが予想される。また、電気機器の中では比較的その操作が簡単であると考えられるビデオでさえ、録画のための操作ができない人が数多く存在するが、そのビデオの操作と比べても、非常に操作が複雑であると感じられる PC に対して、恐れまでもを感じることを考えられる。

次項では、以上の3つの問題点を改善する方法について述べる。

2.1.2 コンピュータを意識させないコミュニケーションの実現に向けて

近年、人間とコミュニケーション可能なエンターテインメントロボットやぬいぐるみの開発が相次いでいる。このようなロボットやぬいぐるみは、人間とコミュニケーションすることにより、人間の心をなごます「癒し」の効果があると言われているが、その先駆けとなったのが AIBO である。AIBO(図 2.1) は SONY が開発したエンターテインメントロボットで、実際の子犬を真似た形状をもち、人が話しかけたりさわったりすると、本当の子犬のような反応をする。これに加えて、AIBO は自らの感情を持ち人間の行動に対して感情を含めたコミュニケーションが可能である。このような機能を持つため、このロボットは 1999 年 11 月に限定 1 万台販売されたが、一体 25 万円と高額にも関わらず 13 万 5000 件の申し込みが殺到し、発売と同時に話題をさらいヒット商品となった。同様の商品として、体に複数のセンサをつけ、人間とのコミュニケーションによって成長していくぬいぐるみの「ファービー」(図 2.2) や内部に感情モデルを持つねこ型ロボット「たま」(図 2.3)、高齢者用コミュニケーション支援器具としてのペットロボット(図 2.4) などが現れ、注目を浴びている。

前項で述べた PC に対する心理的障壁を取り除くために、AIBO などを参考に PC と人間との間で感情を含めた交流をもたらす技術を開発し、PC と人間との間を取り持つヒューマンインタフェースに適用しようというのが、本研究の着眼点である。PC と人間との間で感情を含めた交流が可能になれば、人間が PC に対して親しみを感じ、PC に対する心理的障壁を減らす、もしくはなくすことができると期待される。「親しみを抱く」とは親近感を持つことであり、自分と似たものや自分にとって好ましい反応を返してくれるものを感じられる。

以上の観点により、本研究では、前項で述べた問題点を解決するために、PC が人間

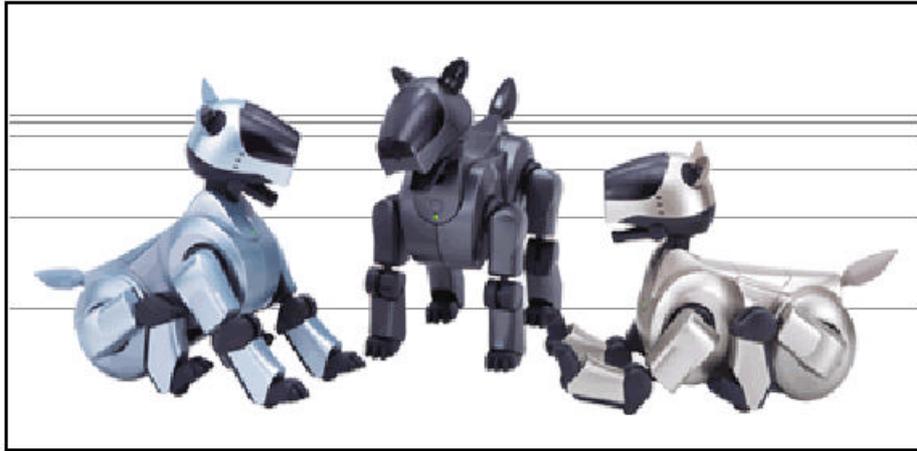


図 2.1: SONY の AIBO

の感情を推定し人間に感情を伝える、すなわち「感情を取り扱う」と音声対話を用いることに着目する。具体的には、まず、問題点3を解決するために、感情を取り扱うことにより人間に親しみやすさ、すなわちなじみがあり親近感を抱く気持ちを感じさせようとし、また、問題点の1、2を解決するために、人間の日常生活でもっともよく使われるコミュニケーション手段として、音声で会話するインタフェースを考えることとする。感情は、人間同士のコミュニケーションにおいても人間の行動の方向性を決める大きな要因になっている。この点に注目して、人間と機械との高次協調・高次機能発現をはかる「アフェクティブインタフェース」の研究が、最近この分野で試みられている。アフェクティブインタフェースは、MITメディアラボのR. Picardにより提唱された A[®]ffective Computing^[1] の概念をヒューマンインタフェースに応用したものである。アフェクティブインタフェースは、人間と機械の接点となるインタフェースに、人間らしい感情を組み込もうとする新しいヒューマンインタフェースの総称であり、人間の感情を推定したり、人間のような感情を表現させる仕組みを機械に組み込んで、より人間に適合するインタフェースを実現しようとするものである。

以上に述べたアフェクティブインタフェースの概念に基づき、感情を取り扱い、音声で会話するアフェクティブインタフェースを目指し、豊かな感情表現を実現するためには、その具体的な実現形として、擬人化インタフェースの利用が考えられる。擬人化インタフェースとは、まるで人間のように音声を用いて応答したり、また顔や姿を持つようなインタフェースのことである。その一例として、PCのモニタ上に3次元の顔や身体をもつ仮想の人間が現れ、ジェスチャや音声、表情を出力して人間とコミュ



図 2.2: トミーのファービー



図 2.3: オムロンのたま



図 2.4: 松下電器産業の高齢者支援ペットロボット

ニケーションし、あたかも本物の人間と対話しているようなインタフェースが挙げられる。このような擬人化インタフェースにより感情を表現するためには、擬人化インタフェースに顔表情や音声などを生成させることが考えられる。表情は、顔表面の形状を変化させることにより様々な感情を直接表現することができる。また、音声は発話内容により直接感情を表現できるだけでなく、発話の速さ、音量、抑揚などにより様々な感情を表現することが可能である。そのため、本研究では、感情を扱うことができる擬人化インタフェースを実現する要素として、顔表情と音声に着目する。

また、このような擬人化インタフェースを実現するにあたっては、人間が機械であるコンピュータに人間性を感じ、コンピュータから人間に感情を伝達することが可能であるか、という問題がある。人間とコンピュータとのコミュニケーションについて、1996年に発表された「The Media Equation」^[2]の中では、人間は、コンピュータを含む人工物のメディアに対しても、人に接するのと同様な対応を示す傾向をもつといわれている。これは人間の脳が周囲の人間との関係により社会性をもつインタラクションを学習して進化してきており、これが今日のコンピュータやメディアに対しても同様に働いていることに起因している。すなわち、人間はコンピュータに対して、もともと擬人性を感じることができ、さらにコンピュータとのインタフェースとして人間的な顔表情や音声を用いることで、感情を伝達できる可能性が高まると考えられる。

情報化が進む現代社会で、誰もがPCを利用することにより得られる利益を享受できるように、感情的交流がより強まるようなコンピュータとのコミュニケーション形態として、人間の顔を提示し、音声で会話する擬人化インタフェースは、難解な機械操作を誰もが利用できるようにするための仕掛けばかりではなく、感情を取り扱うことでコンピュータを身近に感じない人々に「コンピュータを意識させない」新しいコミュニケーション手段として大いに期待され、またアフェクティブインタフェースを実現するための基礎となり得る。

2.1.3 従来研究の展望

本項では、本研究の主題であるアフェクティブインタフェースに関係する従来研究として、特に擬人化インタフェースに関する研究を概観する。次いで、その要素技術を展望し、要素技術の中で、特に顔表情を利用する研究についてまとめる。

擬人化インタフェース

擬人化インタフェースは、音声や画像を代表としたマルチメディア技術の進展を背景として、それをさらに進め、より自然なコミュニケーションに近い形態を実現しようとするものである。

最近の擬人化インタフェースは、顔や姿をもつものが主であるが、顔や姿をもたない擬人化インタフェースの考えは古くからあり、例えば、1968年のキューブリック監督のSF映画「2001年宇宙の旅」にはHAL9000というコンピュータが登場する。映画でのHAL9000は、人格を有し、人間との会話能力をもち、画像認識能力として個人特定、行動認識、読唇などができるとしている。

その他に、複雑・高機能化する情報空間での電子秘書の役割を果たす擬人化インタフェースについても古くからその考えが提唱されている^[3]。

一方、実際の擬人化インタフェースの研究開発は、最近、国内外で取り組まれ、既に実現化している。

東芝のTOSBURG II^[4]は、ハンバーガの注文を受け付けるマルチモーダル擬人化インタフェースであり、その特徴は音声処理部にある。この擬人化インタフェースは、発声中でもユーザからの発話を受け付けることが可能である。顔は漫画的な顔を用いている。この擬人化インタフェースはさらに発展し、社員の持つ知識を知識ベース化し、それを共有する秘書エージェントシステムとして運用されている。

ソニー CSL のTalkman^[5]は、ソニーのコンピュータ製品に関する質問応答システムであり、会話中の表情の生成に特徴がある。顔の3次元モデルは、実画像をテクスチャマッピングしている。加えて、2人の人間の会話を脇で聞いていて、会話の途切れや誤解等が生じているようなら割り込んで、会話が円滑に進むようなアドバイスの発言をするエージェント(自律的に動くソフトウェア)の試みも行われている。実用化のためには、さらなる意味、状況理解などの知識処理部分の充実が必要ではあるが、将来のパートナーとしての擬人化インタフェースの可能性を示している。

東京大学の石塚研究室では、3次元モデル上にテクスチャマッピングした自然感の高い顔を実時間動作させるメディア技術を開発した後、WWWブラウザとの結合を特徴とした擬人化インタフェースを研究している^[6]。

電総研が研究している擬人化インタフェースシステム^[7]は、細かな腕と指の動きをもつキャラクタエージェントに特徴があり、また登録しているユーザの顔を認識して名前を呼び、話し掛けてくる。このシステムと前述の東京大学のWWWブラウザとの結合技術を用いて、法律知識ベースに基づき法律に関する相談のマルチモーダルイン

タラクションを行う実験システムも作成されている。

他にも、NECの電子図書館での擬人化秘書インタフェース^[8]、オムロンの感情機能をもつマルチモーダルエージェント^[9]などがある。

一方、海外では、エージェントキャラクタのbelievability(直接的には、「信頼性を感じさせる性格」を意味するが、ここでは「人間味を感じさせる性格」を意味する)などを重要視した研究が多数ある。早くから Believable Agent に着目して研究を進めてきたのは、米国カーネギーメロン大学のOzプロジェクト^[10]である。キャラクタの外見は、ボールに目がついただけであるが、このCGキャラクタが感情をもっているように動き、ユーザとのインタラクションを行う。また、スタンフォード大学では、Virtual Theater プロジェクト^[11]が進んでいる。この実現例がSpencer's Barであり、バーのカウンタで個性的な姿をしたユーザの化身であるアバタと自律的なエージェントが、オンラインチャットを行う。

ビジネス面では、マイクロソフトのMicrosoft Agent ツールがある。マイクロソフトは、キャラクタエージェントによるマルチモーダルインタフェースに関心をもち、Persona プロジェクトを進めている^[12]。ワープロや表計算などの統合化環境であるMicrosoft Office97はユーザが操作に困ったと思われるときに姿をもつデジタルアシスタントが登場するが、現状では話し言葉を吹き出しの中に文字で出力する形式である。近いうちに人間が音声で質問し、その質問にシステムが音声で回答してくれるようになると予想される。

このようなエージェントシステムは、教育分野にも応用できる。教育分野で学習を指導するエージェントは教師エージェントと呼ばれ、その効用として、親しみのもてるインタラクションが可能であることに加えて、分かりやすく記憶に残るような教示効果をもたらすことが可能となる。例えば、ノースカロライナ州立大学のCOSMOシステムでは、奇妙な顔、姿のキャラクタエージェントがインターネット・プロトコルについて質問に答え、教示する^[13]。また、南カリフォルニア大学のシステムでは、Steveと呼ばれる教師エージェントが3次元環境中で動作し、艦船の機器の操作法や修理法について教示する^[14]。

要素技術

擬人化インタフェース実現のための要素技術を、人間の耳、目、口、顔、身体、頭脳の機能に対応させると表2.1のようになる。さらに、これらの機能を統合し調和すれば、より高次の機能を発現させることも可能である。これらの要素技術のうち、本研

究に直接関係する画像や音声の処理技術について展望する。

表 2.1: 体の部位と対応する要素技術

体の部位	対応する要素技術
耳	音声認識
目	画像認識 (表情、身振りの認識、読唇など)
口	音声合成
顔	顔表情の生成、唇の動きの生成
身体	動作、身振りの生成
頭脳	発話内容の意味の理解 表情、動作、身振りなどの意味理解 応答の生成 インタラクションの管理 ユーザからの情報の理解 知識

画像や音声などのメディアの入出力部については、近年利用できるソフトウェアなどが増えてきており、以前に比べると利用環境は向上しているが、必ずしも目的に適した機能がなかったり、インターフェースが適合しない、または性能が不足しているなどの問題がある。音声認識の性能は進歩してきてはいるが、制限のない自由対話を取り扱うことは、応答の生成も含めて困難であり、実際に利用するには人間に制限を感じさせないような対話の状況設定が必要となる。一方、音声合成に関しても、市販のソフトウェアを用いる場合、声の大きさ、読み上げ速度などを変化させると不自然な出力になり、現状では人間のように感情を含め、抑揚をつけた発話は困難である。このため、利用できる状況は機械的な応答が可能な場面、例えば自動発券機などに限られている。

顔画像の合成と認識に関しても、多くの研究が行われている。これらの研究では、顔画像における表情を分析するパラメータとして、FACS(Facial Action Coding System)と呼ばれる表情記述方式における顔面表情筋の運動単位 AU(Action Unit) が用いられることが多い。

顔画像認識では、FACS を直接利用する以外の手法としては、物理モデルに基づく手法、ニューラルネットを用いて AU パラメータを学習させる手法などが挙げられる。

本研究室では、一昨年度、動的顔表情認識による感情推定手法の提案を行った^[15, 16]。

一方、上記のような分析に基づき顔表情を合成するには、主にワイヤースケルトンに代表される3次元モデルを変形させることによって実現される。本研究室では、昨年度、顔面の主要な筋肉の収縮をシミュレーションして、顔表面の変形を算出することにより表情変化を合成する、筋肉モデル法を用いて顔表情を動的に合成する研究を行った^[17, 18]。表情合成の研究は、これ以外にも、AUに基づき3次元モデルの特徴点を移動させる方法、顔面を解剖学的なばね構造と考え、顔面筋の移動をシミュレートすることで自然な表情を合成する方法などがある。

2.2 研究の目的

以上の背景より、これまでも様々な擬人化インタフェースの研究が行われてきたが、音声・画像の認識・生成、頭脳の実現など、技術面の充実が中心であった。これに対して、本研究では、システムを使用する「人間」を中心に捉え、感情の社会的、文化的要因を考察し、利用者層が共有する感情生起特徴を明らかにすることに注目する。

これより、本研究では、顔を持ち人と会話できる擬人化インタフェースを、アフェクティブインタフェースの実例として着目し、作成する擬人化インタフェースからの感情の出力を人間が認識できるか、またどのように感じるかを中心とした基礎研究を行う。

そこで、本論文の具体的研究として、「コンピュータによる音声合成と動的顔表情合成により、どの程度効果的に感情的共感をもたらさうか？」をアフェクティブインタフェース実現上の検討課題と考え、その基礎研究として、音声合成と動的顔表情合成によって、物語を朗読するシステムを構成することとする。具体的には、物語文を入力すると、自動的にその内容を感情を込めて朗読し、それに合わせた顔表情を提示するシステムを試作する。

そのための要素研究として、以下の1、2を検討する。

1. 物語文を感情を込めて合成音声で朗読させるために、素材の物語文をどのように加工し、どのように朗読する合成音声を用いると効果的か。
2. 合成顔画像を物語文の朗読に合わせて、どのように表情変化させると効果的か。

次いで、その2つを検討した結果に基づいて以下を行う。

1. 「物語朗読システム」を試作する。

2. 上記システムを用いて、被験者実験を行う。
3. 被験者実験の結果を分析して、今後の研究展開のための課題を導出する。

以上を、本研究の具体的な目的とする。

第 3 章 物語朗読システム構成のための基礎的 検討

本章では、2.2 節で述べた「物語朗読システム」を試作するために、合成音声で物語を朗読しながら物語の内容に合わせた顔表情をディスプレイ上に提示することにより、感情を表現するシステムの構成を検討する。まず、感情の社会的、文化的要因を考察したのち、言語と感情は密接な関係があるという観点から、物語の感情を表現するために、物語文に含まれる単語に基づいて表現する感情を推定する方法を検討する。次に、感情表現において大きな役割を担っている音声と顔表情に着目し、推定した感情を合成音声と合成顔表情を用いて表現する方法をそれぞれ検討する。そして、検討結果に基づいたシステムを提案する。

以下では、システムの全体構成と感情推定・感情表出方法の概要を述べた後、感情について考察し、本研究で用いる感情の分類方法を説明し、システムを試作するために、ことばから人間の感情を推定する方法、物語文朗読による感情伝達手法、および顔表情合成による感情伝達手法を順に検討した後、検討結果に基づき物語朗読システムを提案し、最後に本章のまとめを述べる。

3.1 システムの全体構成と感情推定・感情表出方法の概要

本研究では、2.2 節で述べたように、「物語を朗読するシステム」を作成するために、具体的に、(1) 物語文をどのように加工するか、(2) 物語文をどのように朗読するか、また、(3) 朗読に合わせてどのように表情を変化させるか、の 3 点を検討し、物語の朗読内容に適している感情を表現する必要がある。ここで、システムから人間に何らかの感情を伝達するためには、伝達される側が『何の感情が表現されているか』を理解し、その感情が『よく表現されている』と感ずることが必要である。本研究では、感情表現の認識は個人的属性と深く関係すると考え、物語朗読システムの試作では、システム作成者の主観ではなく、感情を伝達される側の意見に基づく表現手法を検討する。

以下、本研究では、物語テキストに基づいてその文が表現する感情を推定することを「感情推定」、感情推定で推定された感情を、合成音声と合成顔表情を用いて表現す

ることを「感情表出」と呼ぶこととする。

まず、物語文が表す感情をどのように推定するかが問題となる。ここで、人間は物語文に含まれる「ことば」から何らかの感情が表現されていることを認知するが、本研究では人間が物語文から認知する感情をシステム側が推定可能にするための感情推定方法を検討する。そして、物語文の朗読、すなわち推定した感情を音声を用いて表現する方法を検討する。また、音声で朗読しながら物語の内容に合わせた表情を表現する、すなわち顔表情合成による感情伝達手法を検討するが、ここで感情推定方法と同様に、朗読を視聴する人が感情を表現していると感じる表情を作成することを中心とした伝達手法を検討する。

このような着眼点に基づき、まず、素材の物語文(以下、物語テキスト)を加工する方法として、物語テキストを入力すると自動的にその文が表現している感情を推定する方法(感情推定)を検討する。次いで、推定した感情を合成音声と合成顔表情を用いてそれぞれ表現する方法(両者を合わせて、感情表出)を検討する。

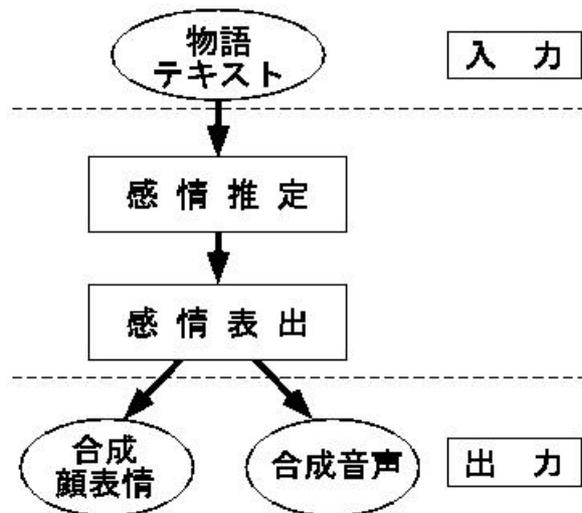


図 3.1: システムの全体構成

図 3.1 に、このような2つの方法に基づいたシステムの全体構成を示す。入力には物語テキストであり、感情推定、感情表出の2つの処理を行い、出力は合成音声で物語を朗読しながらそれに応じて表情を変化させる。ここで感情推定は、用意された対応を参照することにより、物語テキストから内容が表現する感情を推定することにより実現する。

3.2 感情とその分類

人は様々な生活体験の場面で感情を抱き、それによって身体に様々な生理的变化が生じているが、それらは表 3.1 に示すように、変化が他人に見えるもの(見えるシグナル)と見えないもの(見えないシグナル)がある。

表 3.1: 感情による生理変化の分類

他人に見えるもの (見えるシグナル)	顔の表情、声の抑揚、身振り、 姿勢、瞳孔拡大
他人に見えないもの (見えないシグナル)	呼吸、心拍、体温、 皮膚電位、筋電位、血圧

また、感情には生得的なものと後天的なものがあり、前者の感情を第一次感情、後者を第二次感情という。例えば、大きな物音を耳にすると最初に驚き、この後一体何が起こったのかと判断する認知活動が始まるが、このような外界刺激により最初に感情が生じるような反射的な感情生起が、第一次感情である。一方、人が成長するに伴い、社会の様々な場面の見聞体験で、特定の対象・状況を感情的経験としてカタログ化した結果として、認知的に生じる感情が第二次感情である。

次に感情の種類と強さを取り扱う場合を考える。これは、アナログ量として定量的に扱う方法として、arousal(覚醒)の高低と、valence(好意の正負感情)の2軸空間にマッピングする考え方がある。また、離散的、デジタル的な識別に、fear(恐れ)、anger(怒り)、sadness(悲しみ)、happiness(幸福)、disgust(嫌悪)、surprise(驚き)の6つの基本感情(Ekmanによる^[19])を基にしてそれらの同時的、継時的組み合わせで、複雑な感情、複合した感情を表す方法がある。

人間は心身の発達に応じて、高度な感情の表現を行うようになる。また、高度な感情表現をできる能力は、個々人の社会的適合の上で知能指数(IQ)以上に重要であるとして、情緒指数EQ(Emotional Quolity)という表題の本もベストセラーになったことがある。このようなヒトの感情の発達を説明するモデルとして、図 3.2 のように感情の進化を反応層、熟考層、自己観察層の順に3層化するモデルがある。

このように、人間の感情には非常に社会的・文化的要因が大きいことが指摘される。ここで、Ekmanらが6つの基本感情を表す顔表情の絵を見ることで、どの程度正しい感情を推定するかを、国民、人種の違いの間で比較した結果を表 3.2 に示す。

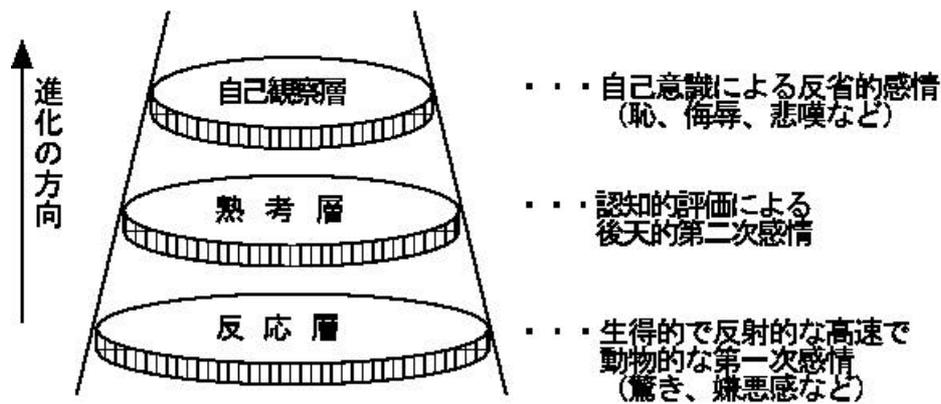


図 3.2: 感情の3層モデル

表 3.2: それぞれの文化で、あらかじめ示された感情を正しく識別したエクマンらの被験者のパーセント値^[20]

国/地域	幸福	驚き	悲しみ	恐れ	嫌悪	怒り
エストニア	90	94	86	91	71	67
ドイツ	93	87	83	86	61	71
ギリシア	93	91	80	74	77	77
香港	92	91	91	84	65	73
イタリア	97	92	81	82	89	72
日本	90	94	87	65	60	67
スコットランド	98	88	86	86	79	84
スマトラ	69	78	91	70	70	70
トルコ	87	90	76	76	74	79
アメリカ	95	92	92	84	86	81

このように、人種によらず共通しているといわれる基本感情でさえ、顔表情を読む上でそれぞれの国の文化的背景により異なっている。このことは、一国内でも年代や学歴、仕事の職種、その他の社会的背景から社会の各層で共通する感情要因を調べて、その感情要因に調和して感情表出を考えないと、アフェクティブインタフェースも社会に浸透しないという問題がある。すなわち、感情は社会的・文化的要因による違いが大きいため、感情の要素を取り込んだアフェクティブインタフェースは、以下の事項をその設計に取り入れることが重要と思われる。

『アフェクティブインタフェースの想定する利用者層の感情の社会的・文化的バックグラウンドを調査して、利用者層が共有する感情生起特徴を明らかにすること』

本研究の場合、物語を朗読するので、

1. 「ことば」に関してどのような感情イメージを共有しているか
2. 「ことば」を読み上げる声や朗読の仕方にどのような感情イメージを共有しているか
3. 顔表情の変化から想起する感情にどのような共通性があるか

といったことを把握していないと、コンピュータが生成する合成音声や顔表情で伝えたい感情要素も相手に伝わらないことになる。

以上により、本研究では、研究の第一段階として、年齢・学歴の個人的属性がほぼ共通な20才前後の大学生・大学院生を対象とする。

また、本研究では、感情とそれを表現する言語とは密接なつながりがあるという観点から、言語に基づいて分類された感情を用いる。具体的には、人間の知覚や認識の構造を視覚的に理解しやすくするために、幾何学的モデルで表現する多次元尺度構成法を用いて、日本語の感情を表す語彙に対応し、感情を35種類に分類した心理学の実験結果^[21]を選択する。本研究で用いる35種類の感情を表3.3に示す。

次節では、ことばからそれが表現する感情を推定する方法を検討する。

3.3 ことばから人間の感情を推定する方法の検討

本研究では、感情はそれを表現することば、すなわち個々の単語と密接な関係があることから、単語に基づく感情の分類方法を選択した。そして、3.1節で述べたように、物語を聞く人がことばからどのような感情を推定するかという点に着目する。

以下では、まず感情推定方法の概要を検討し、ことばに着目して感情を推定するために、感情とことばの対応付けを行う方法、および結果について述べる。

3.3.1 感情推定方法の検討

物語文が表現する感情を推定するために、ことばと感情との対応、および聞く人の属性の2点に着目し、まず事前に聞く人と同じ属性を持つ人、すなわち本研究の場合

表 3.3: 35 種類の感情

	感情		感情		感情
1	喜び	13	嫉妬	25	寂しさ
2	楽しい	14	後悔	26	苦痛
3	幸福	15	忍耐	27	嫌い
4	暖かい	16	耐えられない	28	失望
5	愛情・好き	17	哀れみ	29	憎悪
6	あこがれ	18	退屈	30	屈辱
7	尊敬	19	ぼんやり	31	理不尽
8	満足	20	不愉快	32	驚き
9	優しい	21	ためらい	33	怒り
10	好奇心	22	恥ずかしい	34	侮辱
11	疑問	23	不安	35	恐れ
12	プライド	24	悲しみ		

は 20 才前後の大学生・大学院生の意見を反映して、単語と感情を対応付ける必要がある。そして、単語と感情を 1 対 1 に対応付けた結果を「感情語テーブル」とし、そのテーブルを参照して、物語文からその内容が表現している感情を推定する。作成する感情語テーブルの例を表 3.4 に示す。

このようなテーブルを用いて感情を推定するには、図 3.3 に示すように、感情語テーブルを参照して物語テキストから感情を表現する単語 (感情語) を抽出し、対応している感情を物語文に埋め込むことが必要である。

次項では、感情推定に必要なことばと感情の対応付け、すなわち感情語テーブルの作成方法と作成結果について述べる。

3.3.2 ことばと感情の対応付け

3.2 節で述べたように、本研究では朗読する物語を聞く人を 20 才前後の大学生・大学院生とする。朗読を聞く人が共有する感情生起特徴に合うように、単語とその単語が表す感情を対応付けた感情語テーブルを作成するため、文献 22 を参考に、以下の 2 つの実験を行う。

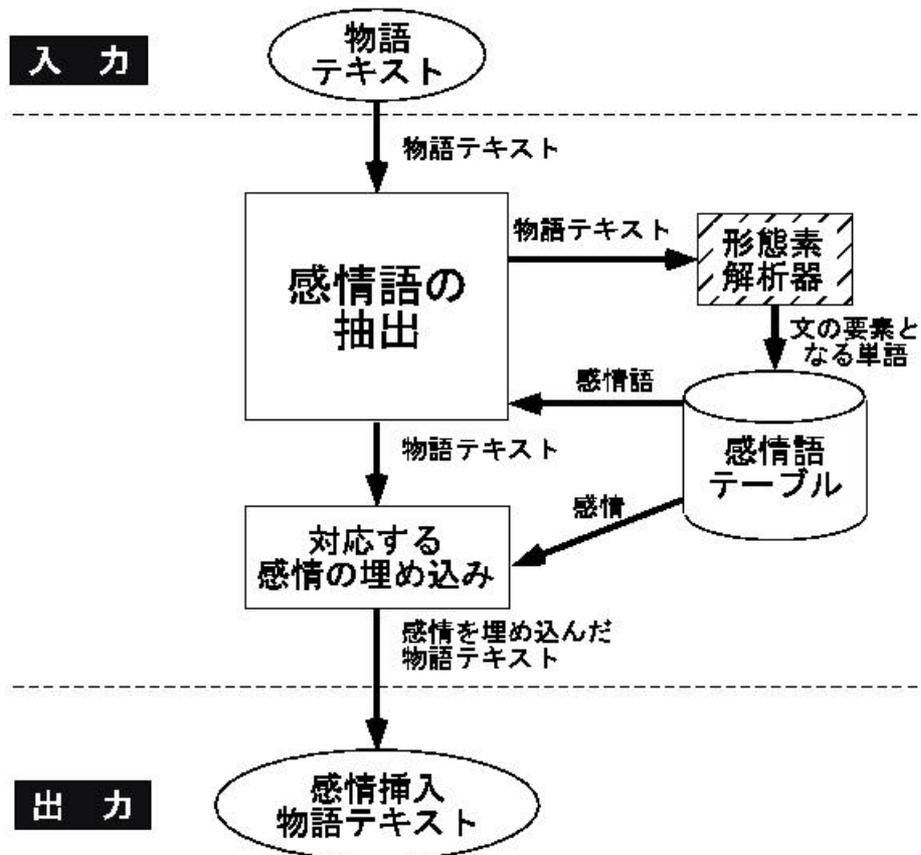


図 3.3: 感情推定のデータの流れ

表 3.4: 感情語テーブルの例

単語	感情
あきあき	退屈
あきれる	失望
あこがれ	あこがれ
あつけらかん	ぼんやり
あっさり	満足
あやまる	後悔
ありがたい	喜び
あわれみ	哀れみ
いじいじ	ためらい
いじける	苦痛

(1) 感情語選択実験

(2) 感情語分類実験

(1) 感情語選択実験では、与えられた単語が感情を表現するかしないかを被験者に評定してもらい、被験者が「感情を表現する」と評定した単語を用いて(2)感情語分類実験を行い、表 3.3 に示す 35 種類の感情に分類してもらう。

以下、各実験の詳細を示す。

(1) 感情語選択実験

感情語選択実験では、用意した単語が感情を表現するものであるかどうかを被験者に評定させ、感情を表現するとされた単語を感情語として選択する。実験では、被験者に与える単語として分類語彙表^[23]を用いる。分類語彙表は、国立国語研究所が作成した日本語の基本語彙設定のための基礎データであり、各単語がどのような意味で用いられているかを一覧できるように単語を分類したものである。現在、約 87,000 語が収録されており、大項目として、1. 名詞、2. 動詞、3. 形容詞に分類されている。

以下に、実験の目的、方法、および結果を述べる。

目的 日本語における感情語は多数あり、どの単語が感情を表すと感じるかは個人の属

性によって異なると考えられるが、本実験は、20才前後の大学生・大学院生を対象にして感情を表す単語を選択することを目的とする。

方法 被験者 大学生、大学院生計20名(男性16名、女性4名)。

用意した単語 本実験では、文献21を参考に、分類語彙表のうち感情に関する項に分類されている「人間の活動における精神と行為に関する単語」を選択する。選択した分類番号とその見出し、およびそれぞれの単語数を表3.5に示す。実験では、これら4,705語のうち、単語の重複等を取り除いた4,303語を用意する。

手続 上記の4,303語を200語ずつ22群(第22群のみ103語)に分け、各群を評価欄とともに2ページ(第22群のみ1ページ)に印刷した小冊子を用意する。この22群に対して1群あたり5名の被験者を配し、各被験者は上記の小冊子を配布された後、「各単語を情緒、感情、情動、気分などの表現に用いるか」という問いについて評価するように教示される。被験者に配布した手順書と感情語を選択するための小冊子の例を付録Aに示す。

結果 各単語ごとに評価の結果を集計し、5名中、過半数である3名以上の被験者が、「感情などの表現のために用いる」と評価した単語684語を選び出した。

(2) 感情語分類実験

感情語分類実験では、(1)の実験で感情語として選択された単語の意味を被験者に評価させ、単語と感情を対応付ける。この実験の方法は、イメージスケールの作成方法^[24]を参考にした。イメージスケールとは、イメージのものさしであり、対象を意味付けしようとして開発されたものである。イメージスケールを作成するためには、3~5人の人間が合意したイメージを対象のイメージとして決定する。ここで3~5人とは、合意が得られる最大の人数とされている。この方法を真似て、(2)の実験では、各単語に対して4名の被験者を配し、全員の合意が得られる感情を割り当てるよう指示した。

以下に、実験の目的、方法、および結果を述べる。

目的 「感情などの表現のために用いる」と評価された上記684語が、どの感情を表しているかを調べることを目的とする。

方法 被験者 大学生・大学院生、計20人(男性5名、女性15名)。

表 3.5: 選択した語彙分類

分類番号	見出し	単語数
1.3004	感情・気分	80
1.3010	快・驚き・喜び	83
1.3011	苦悩・悲哀・恐れ・怒りなど	164
1.3012	安心・焦燥・満足	101
1.3020	対人感情（好悪など）	300
1.3021	対人感情（敬意・信頼など）	113
1.3030	表情（顔つき・泣き・笑いなど）	186
1.3040	自我・信念・努力・忍耐など	114
1.3041	自信・名誉・勇気	110
1.3042	欲望・期待など	107
1.3044	態度・反省など	43
1.3045	意志	121
1.3046	誠・徳・義・信・恩など	165
1.3047	信仰・宗教	111
2.3000	感覚・疲労・睡眠など	364
2.3010	気分・情緒	476
2.3020	対他感情	285
2.3030	表情・態度	263
2.3040	努力・忍耐	199
2.3041	志望・反省	198
2.3042	誇り・恥じ・気どり・ひがみなど	114
3.3000	意識・感覚	209
3.3010	驚き・楽しい・快い	160
3.3011	苦しい・悲しい・こわい	118
3.3012	はずかしい・ほしい・くやしい・ありがたい	130
3.3020	好き・きれい・かわいい・にくらしい	170
3.3030	しくしく・にこにこ・ぷりぷり	63
3.3040	かしこい・おろか	158
合計		4705

用意した単語 感情語選択実験により得られた 684 語。

手続き 上記の 684 語を 150 語ずつ 5 群 (第 5 群のみ 84 語) に分け、評価欄とともに 1 ページ 25 語ずつ、印刷した小冊子を用意する。被験者は 4 名ずつ 5 つのグループに分けられ、各語に対応する感情を表 3.3 の 35 種類の中から選択し、単語を分類する。分類は、4 人のグループで、全員の見解が一致した感情を選択するように教示される。各グループは、各 1 群の分類を行う。被験者に配布した手順書と語彙分類用紙を付録 B に、実験風景を図 3.4 に示す。



図 3.4: 感情語分類の実験風景

結果 表 3.6 に分類結果 (単語数) を、付録 C に分類結果 (全単語) を示す。

この実験の結果に基づき、単語と感情を対応付け、感情語テーブルとして用いる。

3.4 物語文朗読による感情伝達手法の検討

人間が音声を用いて表現できる感情は、言語的なものと非言語的なものがあるが、本節では、感情推定方法で推定した感情を非言語的な音声情報を用いて表現する方法を検討する。

以下では、まず、感情を表現する音声パラメータとして、本研究では何を選択するかを検討した後、感情を表現するための各パラメータ値の決定について述べる。

表 3.6: 感情語分類実験結果 (単語数)

番号	感情	単語数	番号	感情	単語数
1	喜び	24	19	ぼんやり	24
2	楽しい	21	20	不愉快	15
3	幸福	19	21	ためらい	15
4	暖かい	18	22	恥ずかしい	15
5	愛情好き	28	23	不安	37
6	あこがれ	19	24	悲しみ	25
7	尊敬	10	25	寂しさ	17
8	満足	38	26	苦痛	29
9	優しい	7	27	嫌い	8
10	好奇心	24	28	失望	25
11	疑問	3	29	憎悪	15
12	プライド	22	30	屈辱	13
13	嫉妬	15	31	理不尽	9
14	後悔	23	32	驚き	21
15	忍耐	11	33	怒り	32
16	耐えられない	43	34	侮辱	11
17	哀れみ	17	35	恐れ	17
18	退屈	14		合計	684

3.4.1 感情を表現する音声パラメータの検討

Mehrabian によると、コミュニケーション時のメッセージ伝達において、音声の果たす役割は 38%もある^[25]。

音声の特徴は、声の大きさ、音の高低の幅、音色、発話の速さ、リズム、発音などであり、これらの特徴を組み合わせることで感情を表現できる。しかし、音声によって感情を推定する能力は個人差が大きく、Davitz によると、聞き手によりその的中率に 20～50%の幅がある^[26]。彼によると、音声から感情をかなり確実に推定できる人には、自分の感情表現がしっかりしており、音感や言語の能力が優れているという特徴がある。音声による感情表現で、一番正解率の高い感情は、怒りといわれている。また、喜びや憎しみもわかりやすいが、愛や恥などは音声では判別しにくい。音声は個人差や状況次第で変化も著しいため、一定の基準を示すことは困難である。一方、本研究では合成音声を用いて物語文を朗読するが、現状の音声合成技術では音声の特徴を多数含め自然に音声を合成することは困難である。

以上より、本研究では、感情の種類による音声表現の特徴のなかで、表現のしやすい「声の大きさ」と「話す速さ」を合成音声のパラメータとする。

3.4.2 感情を表現する音声パラメータ値の決定

合成音声を用いて感情を表現するためには、前項で選択したパラメータに対して、感情とそれに対応する音声の特徴を対応付けたテーブル(感情表出用音声テーブル)を作成する必要がある。しかし、すべての感情について感情を表現する音声の特徴があるわけではなく、上述したように音声を用いて表現しやすい感情とにくい感情があるため、合成音声を用いて表現可能な感情を選択する必要がある。このため、本研究では、既存の知見^[26]を参考にして顕著に感情を表現する音声パラメータを扱う。表 3.7 に本研究で扱う感情とそれを表現する音声の特徴を記した感情表出用音声テーブルを示す。

3.5 顔表情合成による感情伝達手法の検討

本節では、感情を表現する顔表情を合成するために、どのような合成方法とパラメータを用いるべきかを検討した後、そのパラメータを用いて各感情を表現する顔表情のパラメータ値を決定する方法、および結果について述べる。

表 3.7: 感情表出用音声テーブル^[26]

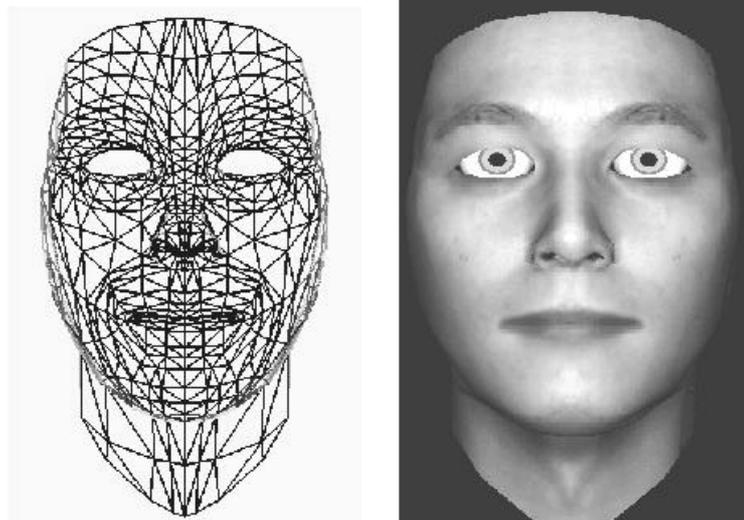
感情	大きさ	速さ
愛情	小さい	ゆっくり
怒り	大きい	速い
退屈	普通	ゆっくり
喜び	大きい	速い
悲しみ	静か	ゆっくり
満足	普通	普通

3.5.1 顔表情合成方法と感情を表現する顔表情パラメータの検討

顔表情を合成する方法は、鼻の長さや眉毛の位置など顔の構造に関する情報をパラメータとして扱うパラメータ法や、人間の顔表情をコンピュータに取り込んで合成する Performance Driven 法などいくつかある。本研究では、(1)感情と関連付けやすい、(2)自然な表情が合成できる、(3)表情表出過程を提示できる、という観点から、筋肉モデル法を用いて表情を合成する。

この方法は、人間の表情は表情筋と呼ばれる顔の筋肉の収縮による皮下組織や皮膚表面の変形として現れることに注目し、筋肉の収縮をシミュレーションして顔表面の変形を算出することによって表情変化を合成するものである。具体的には、各表情筋の収縮率をパラメータとして顔表情を合成する。筋肉モデル法の長所は、あらかじめ個々の表情に関する顔の形状データを用意する必要がないこと、および人間の顔の解剖学的な構造をシミュレーションしているため、リアルな表情が比較的容易に表現できることである。

本研究では、図 3.5(A) に示すような 882 ポリゴンからなる 3 次元顔形状モデルに顔写真のテクスチャを貼り付け (図 3.5(B))、表情の表出に合わせて筋肉モデルによりポリゴン頂点を移動させることにより表情を合成する。ここで、モデル化する表情筋は、図 3.6 に示す主要な筋肉 29 種類であり、左右対称にある筋肉のパラメータは同じ値をとるものとして、表情合成には表 3.8 に示す 16 種類の表情筋のパラメータを用い、これらを本研究における顔表情パラメータとする。



(A) 3次元形状モデル (B) テクスチャの貼り付け

図 3.5: 3次元顔モデル

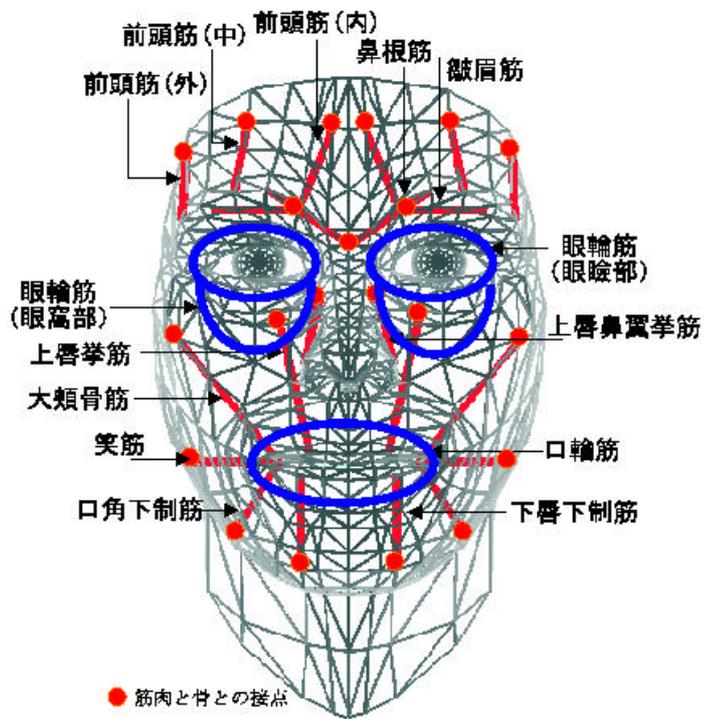


図 3.6: モデル化した表情筋とその配置

表 3.8: 表情合成に用いる表情筋の種類

番号	表情筋
1	前頭筋 (中)
2	前頭筋 (外)
3	前頭筋 (内)
4	皺眉筋
5	鼻根筋
6	上唇鼻翼挙筋
7	上唇挙筋
8	大頬骨筋
9	口角下制御
10	下唇下制筋
11	笑筋
12	眼輪筋 (上)
13	眼輪筋 (眼窩部)
14	眼輪筋 (下)
15	口輪筋
16	あごの回転

3.5.2 感情を表現する顔表情パラメータ値の決定

本研究では、物語の朗読を聞く人の感情想起特徴に合わせて、表現している感情を聞く人が認識できる顔表情を合成するために、人間の感情想起特徴に基づき、「感情を表現する表情」に関する解を求める計算方法が必要となる。

顔表情のように、人間が主観的に評価するものの場合、人間の心の中にある印象、好み、わかりやすさなどが求められる出力であり、主観的価値観を扱う感性レベルで評価をすることが望ましい。このような出力を最適化するためには、人間の主観を最適化系に組み込み、その評価に基づいてコンピュータに最適化させるという方法が考えられる。このように人間とコンピュータとの相互作用によって人間の主観的評価に基づく最適化を行う手法として、進化計算を用いるインタラクティブ進化計算がある^[27, 28]。インタラクティブ進化計算は、人間の感性に関する解を求めるのに適した方法であり、遺伝的アルゴリズム、すなわち進化計算を人間とコンピュータのインタラクションに応用した技術である。

本研究では、感情を表現する顔表情を決定するためにインタラクティブ進化計算を用いる。インタラクティブ進化計算のシステム構成を図3.7に示す。

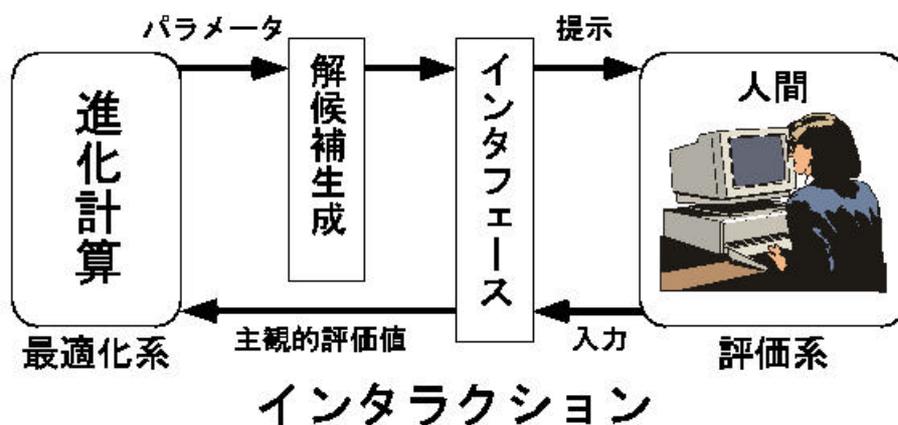


図 3.7: インタラクティブ進化計算のシステム構成

インタラクティブ進化計算では、図3.7に示すように、人間が評価系、進化計算が最適化系として全体システムを構成する。進化計算は、解候補を構成するパラメータを生成し、解候補を生成するモジュールにわたす。ここで、パラメータとは解候補の特徴を表す変数のことである。人間は、インタフェースを通して、提示された解候補に主観的な評価を与える。進化計算は、人間から与えられた評価値に基づいて、新たな解候補のパラメータを生成する。このインタラクションを繰り返して、評価している

人間が求める方向にパラメータが最適化される。インタラクティブ進化計算で用いられる遺伝的アルゴリズム [29, 30, 31, 32] については、付録 D で説明する。

以下では、感情を表現する表情を求めるためにインタラクティブ進化計算を用いて行った実験について、その目的、方法および結果を述べる。

感情を表現する顔表情のための顔表情パラメータを決定する実験

目的 インタラクティブ進化計算を用いて、感情を表現する顔表情パラメータを決定する。

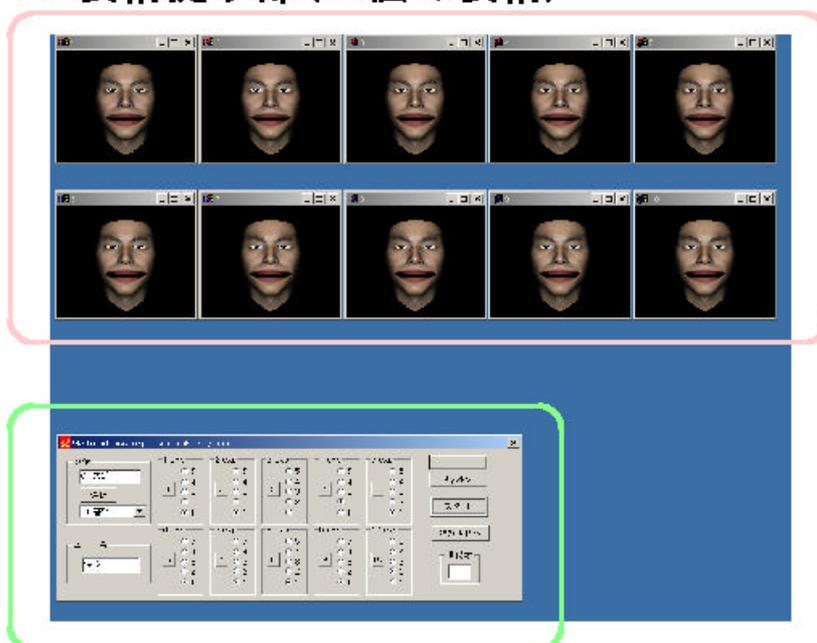
実験システム 実験に先立ち、顔表情パラメータを最適化するインタラクティブ進化計算システムを作成した。システムの画面構成を図 3.8 に示す。画面は、1：表情提示部と 2：表情評価部より構成される。1：表情表示部には、進化計算により生成された解候補として 10 個の合成顔表情が提示され、被験者はこれら 10 個の顔表情に対して、事前に与えられた感情をどの程度表現しているかを 5 段階で評価する。被験者が 10 個の評価を終えた後、「次の世代へ」ボタンをマウスでクリックすると、システムが遺伝的アルゴリズムを用いて次世代の解候補を生成し、新たに 10 個の表情が提示される。実験システムの遺伝的アルゴリズムで用いたパラメータを表 3.9 に示す。各世代について解候補として 10 個ずつの表情を提示するので、1 世代あたりの個体数は 10 である。前項で述べたように、表 3.8 に示す 16 種類の表情筋の収縮率を本研究の顔表情パラメータとしたが、この 16 種類の顔表情パラメータについて、およそ人間の顔で動かすことが可能だと考えられる収縮範囲を設定し、それらを 4 段階に分割した。実験で用いた顔表情パラメータの値を付録 E に示す。また、各顔表情パラメータの値は 4 段階に分割しているため、各顔表情パラメータの遺伝子長は 2 となり、遺伝子長の合計は 32 となる。遺伝的アルゴリズムにおける交叉率は、経験的に 0.6 ~ 0.95 といわれており [29]、本研究では 0.8 とした。突然変異率は、一般的に各世代各個体が 1 遺伝子変異する値をとることが多い [29] ため、本研究では 0.03 とした。

方法 本実験では、システムで提示される表情の変化に慣れ、事前に、被験者自身に終了条件を想定してもらうために各表情で同一被験者にそれぞれ 2 回ずつ試行してもらう。各試行の終了条件は、1 回目は世代数が 21 になること、2 回目は被験者自身が提示された感情を表現していると感じられる表情が提示されること、である。以下、実験の手順を述べる。

表 3.9: 遺伝アルゴリズムのパラメータ

1世代あたりの個体数	10
遺伝子長	32
交叉率	0.8
突然変異率	0.03

1: 表情提示部(10個の表情)



2: 表情評価部

図 3.8: 合成顔表情を評価する実験の画面構成

- 手順 1. まず、被験者に実験手順書(付録F)を読んでもらった後、感情を表現する顔表情を決定するために画面に提示される合成顔表情を5段階で評価することを説明する。
2. 次に、被験者は、本システムにより表情がどのように変化していくかを体験するために、1回目の試行として、世代数が21になるまで評価を繰り返す。
3. 2回目の試行では、被験者は、世代数に関わらず、与えられた感情を表現していると感じられる表情が現れるまで評価を繰り返し、最終的に、感情を

最もよく表現していると感じられる顔表情を1つだけ選択する。

被験者 被験者は、20代前後の大学生・大学院生、計16名(男性13名、女性3名)であり、各感情に関して1名の被験者が表情を決定した。ここでは、各表情の決定に対して被験者の割り当ては無作為に行い、被験者と共通の学歴・世代の感情想起を示す一例とした。

結果 各感情に対して、被験者と、その被験者により表情が決定した世代数を表3.10に示す。また、実験の結果の表情を図3.9~図3.11に示す。この結果の顔表情パラメータを各感情を表現する顔表情として、感情表出用表情テーブルを作成する。

3.6 検討結果に基づく物語朗読方法の提案

本節では、3.3節~3.5節の検討結果に基づき、具体的に物語朗読方法を提案する。まず、3.3節のことばから人間の感情を推定する方法の検討結果に基づき、感情推定方法を提案し、次いで、3.4節の物語文朗読による感情伝達手法、および3.5節の顔表情合成による感情伝達手法の検討結果に基づき、感情表出方法を提案する。

3.6.1 感情推定方法の提案

ことばから感情を推定する方法の検討より、感情語テーブルを作成し、作成した感情語テーブルを用いた感情推定方法を検討した。以下、検討結果に基づく感情推定方法を提案する。

感情推定での入力には物語テキストとし、感情語テーブルを参照することで物語テキストに含まれる感情語を抽出し、抽出した感情語に対応する感情識別子を物語テキストに埋め込む。ここで、感情識別子とは、物語テキストの該当箇所が表現しているとされる感情を識別できるように、物語テキスト中に埋め込む識別子のことである。この感情識別子は、感情を表現するための始まりの記号として< EMOTION = \感情の種類" >という文字列(以下、タグ)、感情の保持終了記号として< = EMOTION >タグを用いて記される。感情識別子を埋め込む場所は、感情語の直前であり、各感情識別子は1文が終わるまで保持される。もし、1文が終了するまでに新たな感情語が含まれていたら、その直前まで感情識別子を保持する。

図3.12に感情推定方法における感情識別子の埋め込み例を示す。ここでは、感情語テーブルに「泣く」という単語があり、その単語に「悲しみ」の感情が対応している

表 3.10: 実験の終了世代数

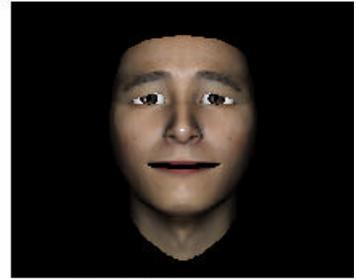
	感情	被験者	終了世代数
1	喜び	KD	19
2	楽しい	NK	25
3	幸福	OF	15
4	暖かい	EK	26
5	愛情好き	SK	18
6	あこがれ	KT	15
7	尊敬	EK	25
8	満足	NK	24
9	優しい	SK	15
10	好奇心	OF	8
11	疑問	KT	7
12	プライド	NK	12
13	嫉妬	EK	35
14	後悔	OF	10
15	忍耐	SK	14
16	耐えられない	KT	3
17	哀れみ	SR	21
18	退屈	TT	18
19	ぼんやり	MT	13
20	不愉快	FM	8
21	ためらい	SR	11
22	恥ずかしい	IR	30
23	不安	MT	4
24	悲しみ	OY	15
25	寂しさ	YM	31
26	苦痛	AU	37
27	嫌い	MT	4
28	失望	MT	13
29	憎悪	YM	22
30	屈辱	NS	20
31	理不尽	YM	10
32	驚き	NS	42
33	怒り	NS	39
34	侮辱	SM	14
35	恐れ	SM	21



1 喜び



2 楽しい



3 幸福



4 暖かい



5 愛情・好き



6 あこがれ



7 尊敬



8 満足



9 優しい



10 好奇心



11 疑問



12 プライド

図 3.9: 顔表情の結果 (1)



13 嫉妬



14 後悔



15 忍耐



16 耐えられない



17 哀れみ



18 退屈



19 ぼんやり



20 不愉快



21 ためらい



22 恥ずかしい



23 不安



24 悲しみ

図 3.10: 顔表情の結果 (2)



25 寂しさ



26 苦痛



27 嫌い



28 失望



29 憎悪



30 屈辱



31 理不尽



32 驚き



33 怒り



34 侮辱



35 恐れ

図 3.11: 顔表情の結果 (3)

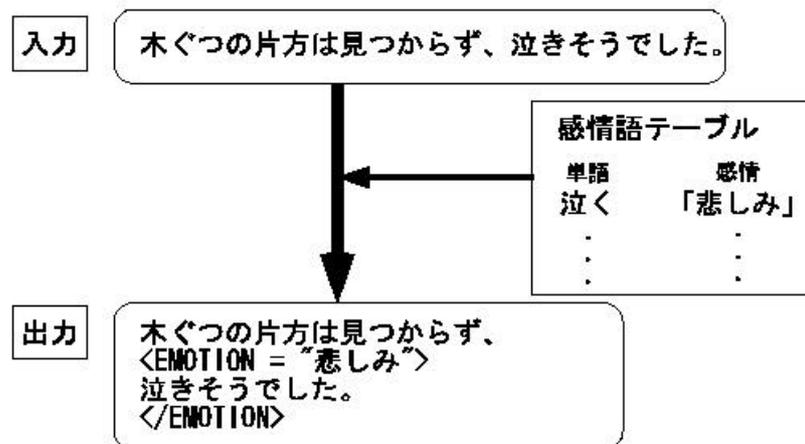


図 3.12: 感情の埋め込み例

とする。入力されたテキストの中で「泣く」という単語が抽出され、抽出された単語の直前に < EMOTION = \悲しみ" > タグが追加される。この文にはこの他に感情語がないため、この感情識別子の終了は文末となり、文末に < = EMOTION > が追加される。

図 3.13 に感情推定のアルゴリズムを示す。感情推定では、物語テキストの 1 文ごとに感情を推定する。まず、物語テキストの 1 文が入力されると、1 文ごとに形態素解析を行い、単語ごとに分解した後、各単語の基本語が求められる。ここで、基本語とは活用語の原形のことである。求められた基本語から、感情語テーブルを参照して感情語を抽出する。抽出した感情語に対応する感情を、感情語テーブルを参照して求め、対応する感情語の直前に感情開始のタグを追加し、それと対応する感情終了のタグを文末に追加する。もし、1 文中に他の感情語がある場合は、次に現れる感情語の直前に感情終了のタグを追加する。出力は、入力した物語テキストに感情を埋め込んだ感情識別子挿入物語テキストとする。

3.6.2 感情表出方法の提案

ここでは、物語文朗読と顔表情合成による感情伝達手法の検討結果に基づき、物語文より推定された感情を表現する感情表出方法について説明する。感情表出方法では、感情推定方法と同様に、聞く人側の視点で感情を表現していると納得できる表現ができれば、感情の伝達が可能であるという点に着目し、聞く人の感情想起特徴に応じた音声や顔表情を合成し、それらを朗読に合わせて提示する方法を検討した。一方、物

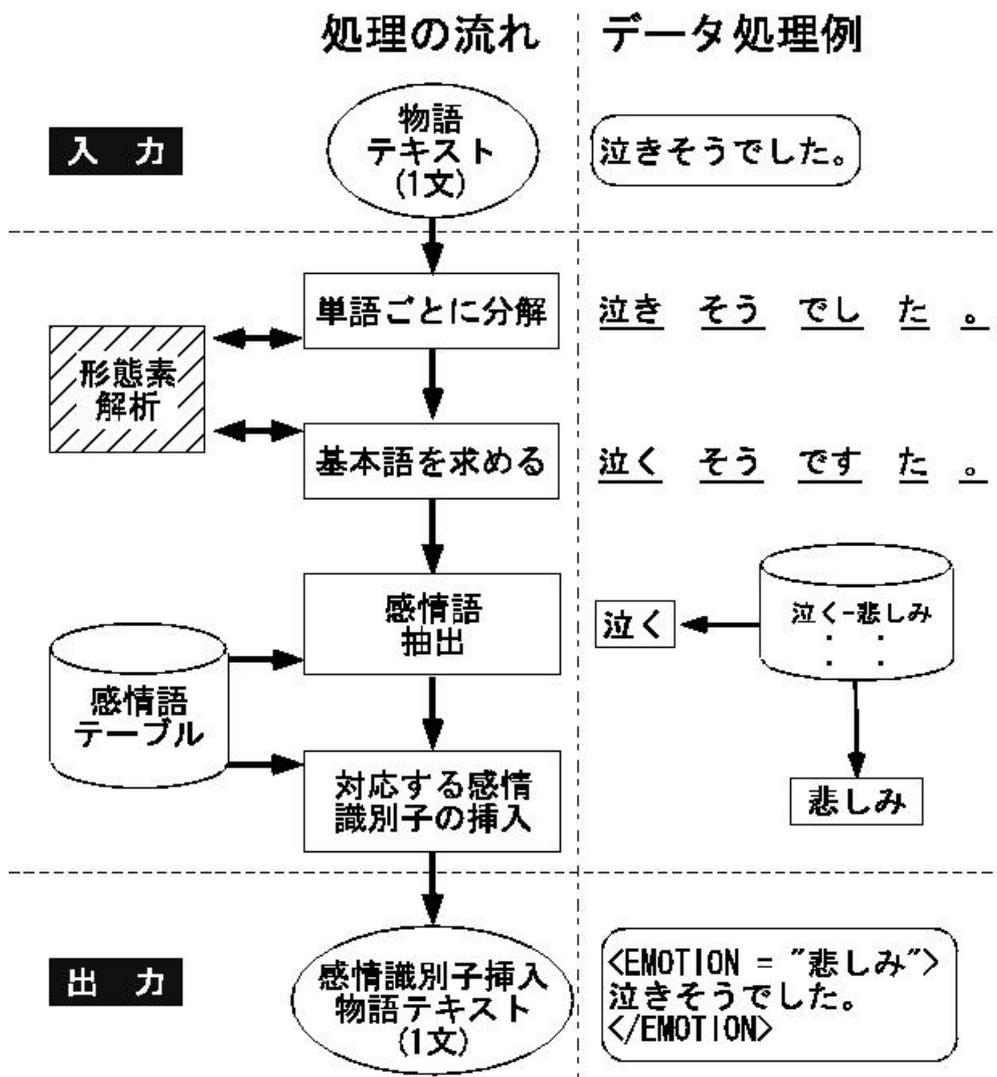


図 3.13: 感情推定のアルゴリズム

語を朗読しているときに、朗読内容に合わせて唇が動いていないと、朗読自体が不自然に感じられ、感情伝達を阻害する原因になる可能性が考えられるため、発話に合わせた唇の動き (lip synchronization, 以下リップシンク) の生成を提案する。

以上の点に基づき図 3.14 に感情表出方法の概要を提案し、以下で説明する。まず、

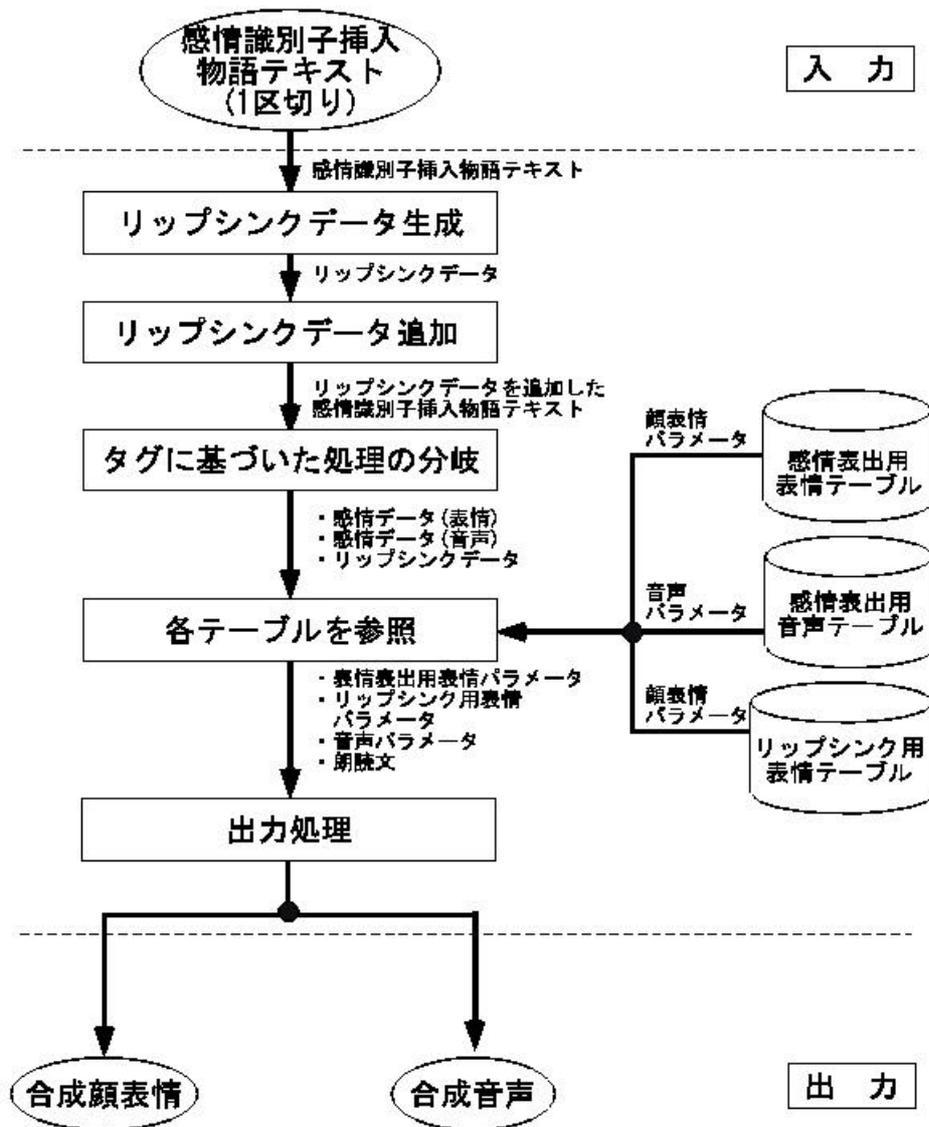


図 3.14: 感情表出方法の概要

感情表出の入力は、感情推定の出力である感情をタグとして埋め込んだ感情識別子挿入物語テキストの1区切りであり、出力はそれに対応する合成顔表情と合成音声である。ここで、感情識別子挿入物語テキストの1区切りとは、以下のものである。

² 感情識別子が挿入されていない場合は、1文を1区切りとする。

- 2 感情識別子が挿入されている場合は、タグの前後を区切りとし、その間を1区切りとする。

図 3.15 に感情識別子挿入物語テキストの例を示す。

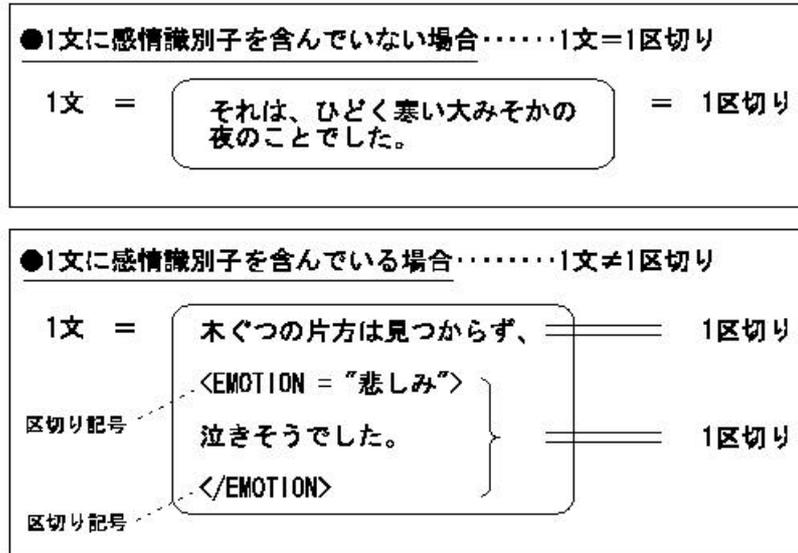


図 3.15: 感情識別子挿入物語テキストの1区切り

この1区切りの発話内容に応じて、リップシンクためのリップシンクデータを加える。リップシンクデータとしては、発話内容の読みから母音を抽出し< LIPSYNC = \母音の並び" >というリップシンクタグを作成する。図 3.16 にリップシンクデータの作成例を示す。次いで、感情識別子挿入物語テキストに加えられた感情タグで示され

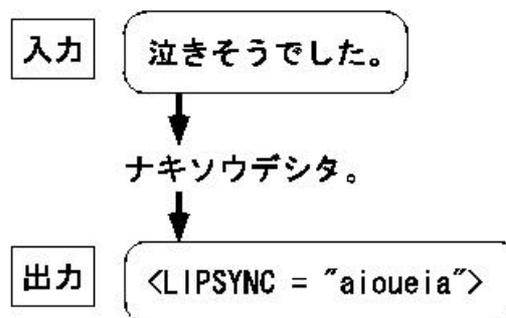


図 3.16: リップシンクデータの作成例

る感情とリップシンクタグで示されるリップシンクデータに基づいて、感情を表出す

る表情、感情を表出する音声、リップシンクのための口の動きを合成する。ここでは、それぞれ3.5節で検討した感情表出用表情テーブル、3.4節で検討した感情表出用音声テーブル、およびリップシンク用表情テーブルを参照する。

ここで、リップシンク用表情テーブルは、日本語の母音である「あ」、「い」、「う」、「え」、「お」、および「ん」の唇の形を顔表情パラメータを用いて表現したテーブルである。このテーブルは、人間の発話の唇の形を参考にして、顔表情パラメータを設定する。図3.17にリップシンク用表情テーブルの顔表情パラメータにより各母音を発話している際の唇の形を示す。

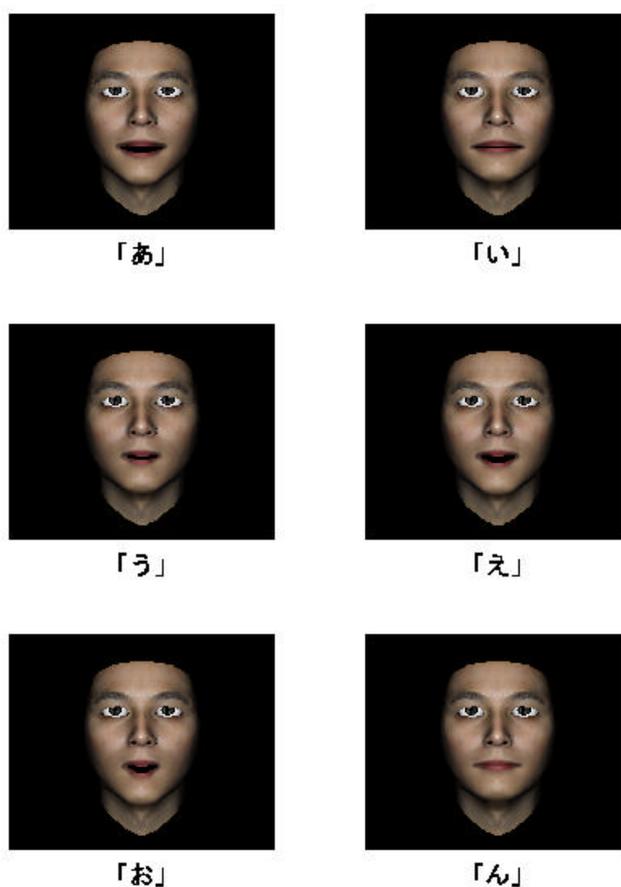


図 3.17: 各母音を発音している顔表情

そして、表情表出用表情パラメータとリップシンク用表情パラメータを顔表情合成の出力処理へ、音声パラメータと朗読文を合成音声の出力処理へそれぞれ送り、合成顔表情と合成音声を出力する。

以上で述べた各パラメータおよび各テーブルを用いて感情を表出する流れを例とともに図3.18に示す。感情表出のためには、感情識別子挿入物語テキストが一区切りず

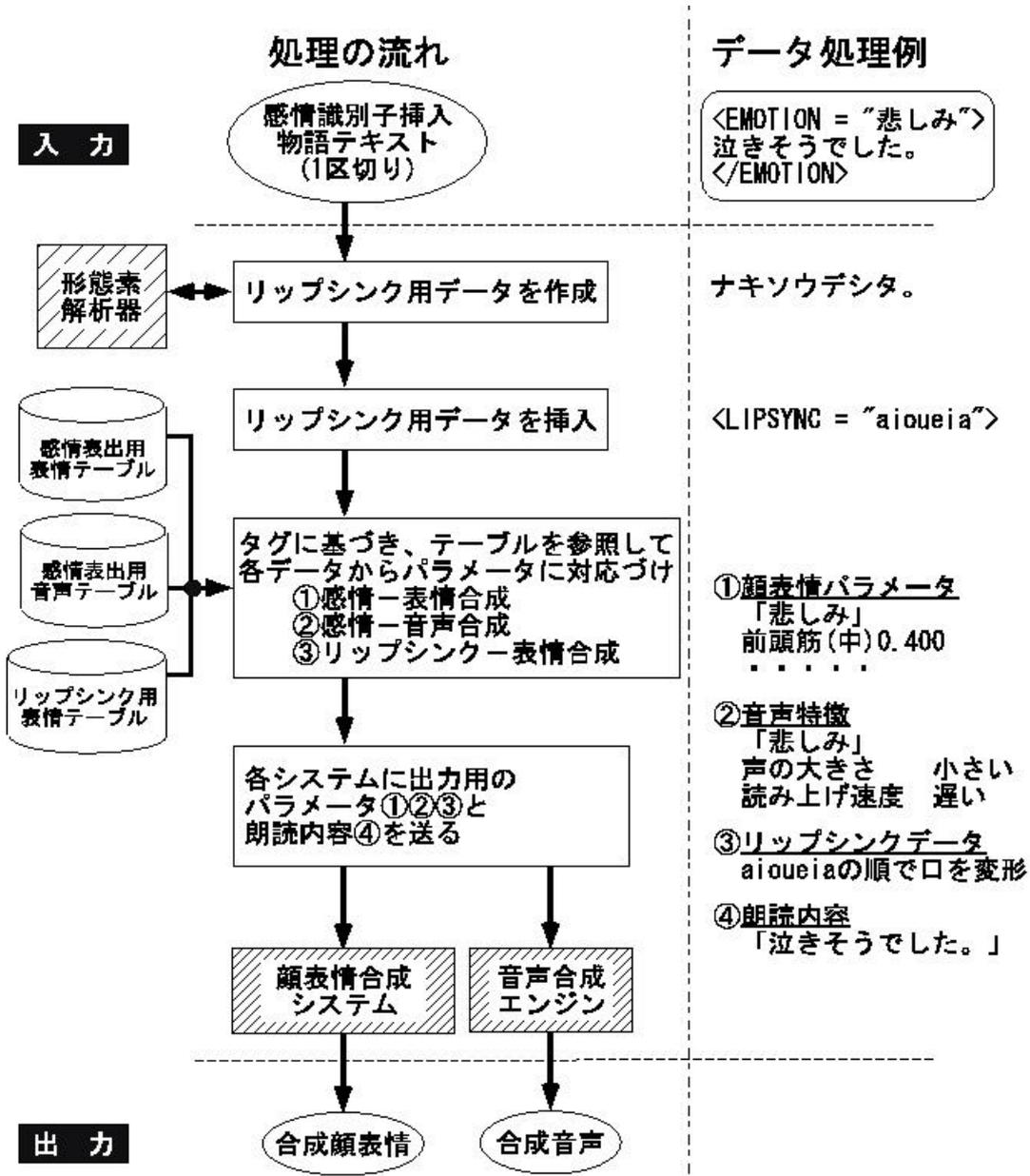


図 3.18: 感情表出の流れとデータ処理例

つ入力される。まず、入力されたテキストの中の読み上げ部分に基づき、リップシンクのための母音を抽出する。次いで、対応する部分に抽出した母音をリップシンクタグとして挿入する。ここで、1区切りの朗読部分について、読み上げテキスト、感情推定方法により埋め込まれた表現すべき感情、およびリップシンク用データが揃う。以上のデータに基づき、3つのテーブル(感情表出用表情テーブル、感情表出用音声テーブル、リップシンク用表情テーブル)を参照して、各パラメータに対応付けた後、顔表情合成システムに顔表情パラメータ、音声合成エンジンに音声パラメータと朗読内容を送る。これにより、合成音声で感情を表現する朗読とそれに合わせた合成顔表情が出力される

3.7 まとめ

本章では、物語朗読システム構成のための基礎的検討として、まずシステムの全体構成と方法の概要を説明した後、感情の社会的、文化的要因を考察し、本研究で用いる感情の分類方法を決定した。そして、ことばによる感情推定手法、物語文朗読による感情伝達手法、および顔表情合成による感情伝達手法を順に検討し、検討結果に基づく物語朗読システムを提案した。

まず、ことばから感情を推定するために、単語と感情を1対1に対応付け、その結果を用いて感情を推定する方法を検討し、単語と感情を対応付けた感情語テーブルを被験者実験の結果に基づき作成した。次いで、物語文を朗読することにより感情を伝達するために、「声の大きさ」と「読む速さ」を音声パラメータとし、感情表出用音声テーブルを作成した。一方、顔表情合成を用いて感情を伝達するために、人間の主観を反映させる計算方法としてインタラクティブ進化計算を用い、各感情を表現する顔表情を被験者実験の結果に基づき作成した。

感情の感じ方や表現は、前述のように年齢、性別、文化、生活環境などの個人的属性によってばらつきがある。提案した手法は、この点を考慮して、事前に物語を聞く人を想定し、その人に近い年齢・学歴を持った人々を対象とした被験者実験により作成した感情語テーブル、表情表出用表情テーブルを用いることを特徴とする。特に、表情表出用表情テーブルの作成にあたっては、「表情」という具体的イメージのつかみにくいものを被験者自ら直接作成するのではなく、次々と提示される表情の評価を繰り返すことで被験者がイメージする表情を作り上げていくインタラクティブ進化計算を用いている。

本研究では、アフェクティブインタフェースの基礎研究として、合成顔表情と合成音声により感情を込めた物語文を朗読するシステムを試作し、被験者実験によってコンピュータから人間への感情伝達要因を検討することを主眼としている。そのため、本章で提案した手法では、実際にシステムを試作して被験者実験を行うことを優先しており、以下のように簡略化している部分がある。

- 2 提案した感情推定方法では、物語文から単語を抜き出し、その感情語テーブルにより単に単語から感情に変換している。そのため、物語文全体が表す意味や文脈を考慮していない。
- 2 感情表出で表出する合成顔表情は、各感情に応じてそれぞれ一人の被験者が個人的主観で決定したものである。そのため、試作するシステムにより朗読される物語は視聴する人の感性に合うように感情を表現できているとは限らない。

第5章で述べる被験者実験では、これら手法の簡略化による限界を考慮する必要がある。このため、以下の点が課題として残されている。

- 2 否定語や助動詞に対応し、表現する感情の打ち消しや強度の変化を推定するとともに、物語文全体が表す意味を推定する感情推定方法を検討する。
- 2 合成顔表情の作成において、35種類の顔表情を統一させるために、1人の被験者にすべての顔表情を決定してもらい、または複数の被験者に顔表情を決定してもらい平均するなど、合成顔表情の作成方法を検討する。

第 4 章 物語朗読システムの作成

本章では、第 3 章で提案した感情推定方法、および感情表出方法に基づいて試作した物語を朗読するシステムについて、そのハードウェア構成およびソフトウェア構成を述べる。

4.1 ハードウェア構成

図 4.1 に物語を朗読するシステムのハードウェア構成を示す。作成したシステム(以下、本システム)のハードウェアは、表 4.1 の性能をもつ汎用のノートパソコン (IBM 社製 ThinkPad A20p) である。本システム実行時には、液晶パネル上に合成顔表情の動画像が提示され、内臓スピーカから合成音声が出力される。



図 4.1: ハードウェア構成

4.2 ソフトウェア構成

本システムは、第 3 章で検討した感情推定方法、感情表出方法に基づき、それぞれの方法を実現するための以下の 2 つのサブシステムから構成されている。

2 感情推定サブシステム

プロセッサ	モバイル Pentium III 750MHz
メモリ	256MB
ビデオ・サブシステム	ATI Ragei Mobility 128
ビデオ RAM	16MB
解像度	1024 × 768
オーディオ	ステレオスピーカ、Sound Blaster Pro 互換

表 4.1: ハードウェア要求仕様

2 感情表出サブシステム

以下、各サブシステムについて詳細を述べる。

4.2.1 感情推定サブシステム

感情推定サブシステムの構成を図 4.2 に示す。感情推定サブシステムは、以下の3つの部分より構成される。

2 テキスト処理部

2 感情語テーブル

2 形態素解析器

感情推定サブシステムでは、1文ごとの物語テキストを入力とし、対応する感情識別子を埋め込んだ物語テキスト(感情識別子挿入物語テキスト)を1文ごとに出力する。

まず、物語テキストの1文がテキスト処理部と形態素解析器に入力される。形態素解析器では、入力された文を単語ごとに分解した後、各単語の基本語が求められる。形態素解析器には、日本語形態素解析システム「茶筌」^[33]を用いた。これらの基本語から、感情語テーブルを参照して感情語を抽出する。抽出した感情語に対応する感情を、感情語テーブルを参照して求め、感情タグとして対応する感情語の直前に挿入する。ここでの出力は、テキスト処理部で感情タグを挿入した物語テキスト(感情識別子挿入物語テキスト)の1文となる。

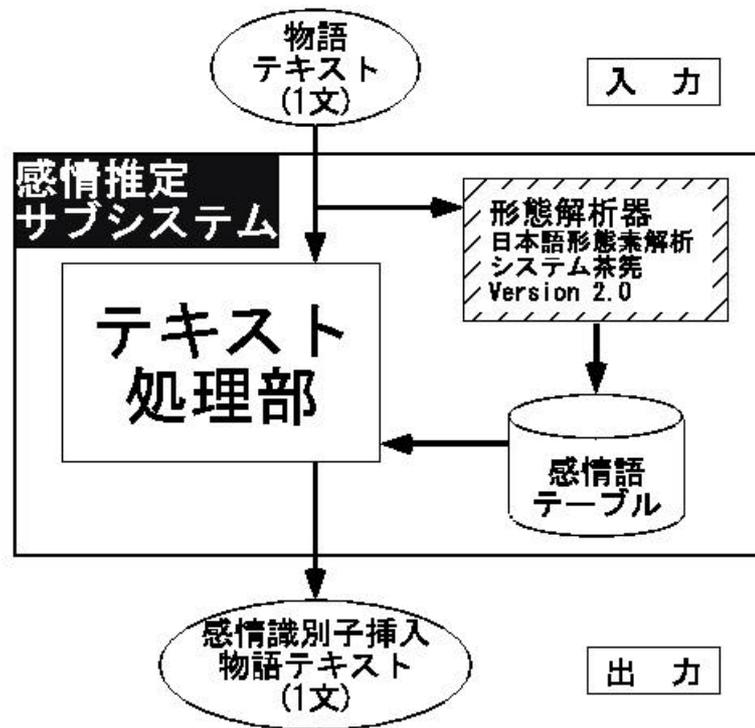


図 4.2: 感情推定サブシステムの構成

4.2.2 感情表出サブシステム

図 4.3 に感情表出サブシステムの構成を示す。感情表出サブシステムは、以下より構成される。

- 2 リップシンクデータ生成部
- 2 表情・音声パラメータ生成部
- 2 顔表情合成システム
- 2 音声合成エンジン
- 2 感情表出用表情テーブル
- 2 感情表出用音声テーブル
- 2 リップシンク用表情テーブル
- 2 形態素解析器

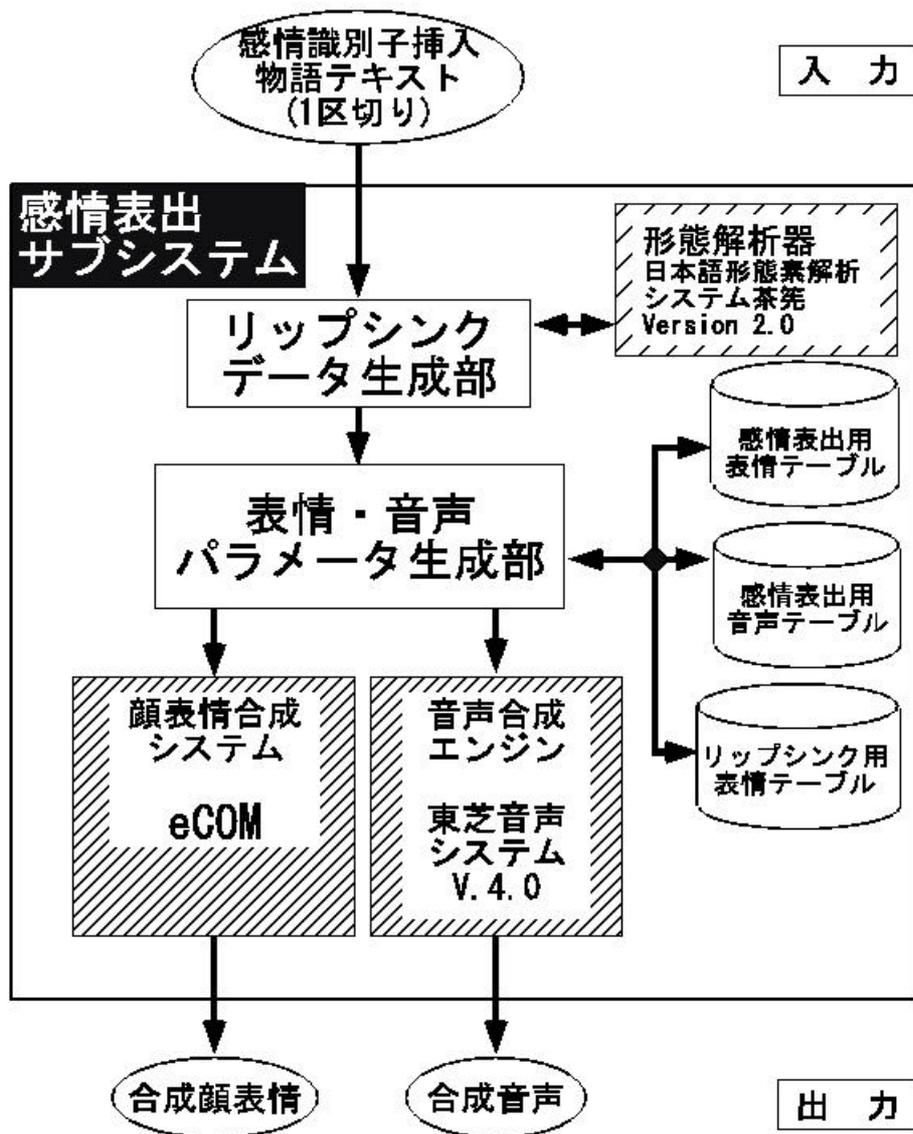


図 4.3: 感情表出サブシステムの構成

感情表出サブシステムは、1区切りずつの感情識別子挿入物語テキストを入力とし、合成顔表情と合成音声を出力する。まず、入力されたテキストの中の読み上げ部分を形態素解析器により読み方に変換し、リップシンクのための母音を抽出する。次いで、対応する部分に抽出したリップシンクタグを挿入する。これにより、1区切りの朗読部分について、物語文である読み上げ用のテキストと、感情推定サブシステムで挿入された感情タグに加えて、リップシンクタグが挿入される。以上のデータに基づき、3つのテーブル(感情表出用表情テーブル、感情表出用音声テーブル、リップシンク用表情テーブル)を参照して、音声パラメータと顔表情パラメータを生成した後、顔表情合成システムに顔表情パラメータ、音声合成エンジンに合成音声パラメータと朗読内容を送る。顔表情合成サブシステムは、昨年度本研究室で開発したリアルタイム動的顔表情合成システム eCOM^[17, 18] を用いた。音声合成エンジンは、東芝音声システム V.4.0^[34] を用いた。以上の処理により、感情を込めた合成音声と、それに合わせた顔表情が出力される。

第 5 章 物語朗読システムを用いた被験者実験

本章では、第 4 章で述べた物語朗読システムを利用して、合成顔表情と合成音声を用いてコンピュータから人間にどの程度感情を伝達できるかを評価するための実験を行った。

以下では、実験の目的、方法、結果を述べる。

5.1 実験の目的

本実験では、合成顔表情と合成音声を使用した物語朗読システムを用いた被験者実験を行うことにより、以下の項目について評価することを目的とする。

項目 1:感情伝達 システムが伝達したい目的の感情をどの程度効果的に被験者に伝えられるか、すなわち被験者に感情を伝達することができるか

項目 2:共感の誘起 システムが朗読する物語の内容や提示する表情をもとに、被験者にその感情を追体験してもらえるか、すなわち被験者に共感させることができるか

項目 3:親しみやすさ システムの朗読方法や顔、表情をもとに、被験者がシステムに親近感を抱き、親しみを感じるか、すなわち被験者に親しみやすさを感じさせることができるか

5.2 実験方法

本実験では、作成した物語朗読システムを利用して、システムが表現する感情を被験者に理解してもらい、適切に表現されているかどうか、すなわちシステムから被験者に感情を伝えられるかどうかを調べるために、(1) 文章を読む、(2) 合成音声の朗読を聞く、(3) 合成音声の朗読とともにそれに合わせた顔表情を視聴する、の 3 つの課題を同じ物語の内容で被験者に行ってもらおう。ここで、(1) と (2) を比較することにより合成音声による感情伝達を評価してもらい、(2) と (3) を比較することにより顔表情を追加した朗読を評価してもらうために、(1)、(2)、(3) の順序で実験を行う。以下、具体的な方法について述べる。

被験者に与える課題は以下の4つである。

課題1 物語の内容を文書にて黙読

課題2 合成音声で読み上げた物語のテキストを静聴

課題3 作成した物語朗読システムを用いた合成顔表情付きの合成音声による読み上げを視聴

課題4 各感情を表現する顔表情の感情表現に関する評価

課題1~3で用いた物語は、アンデルセンの「マッチ売りの少女」である。実験で使用する物語テキストを付録Gに示す。

また、実験目的の各項目に対する評価方法について述べる。

項目1の評価方法 感情伝達の評価については、課題3の後でアンケートによる評価を行い、課題4で各表情に対して個別に評価する。

項目2の評価方法 物語の朗読内容が表現する感情を被験者に伝えることができたか、すなわち共感の誘起に関しては、課題3の後でアンケートによる評価を行うとともに、各課題の前後で多面的感情状態尺度(MMS)を用いて、各課題前後の感情状態の変化を測定する。MMSについては、後述する。

項目3の評価方法 親しみやすさの評価については、課題3の後でアンケートによる評価を行う。

なお、各課題の開始前の感情状態を平静に保つために、課題1~3の前にはリラックス用のビデオを視聴してもらおう。リラックス用ビデオの内容は、特別な情動を喚起しないよう、落ち着いた音楽が流れる風景の画像である。

5.2.1 実験手順

実験手順を図5.1に示す。実験は、大きく(A)~(D)の4つのパートに分かれ、各パートは前項で述べた4つの課題に対応している。実験開始後、実験について簡単に説明した後、氏名、年齢などを記入するアンケートB-1に回答してもらおう。次いで、課題1を行うパート(A)の部分に進む。まず、リラックスビデオを見てもらい、感情状態を一定にしてもらおう。次いで、MMS用紙に回答してもらい、課題1を行った後、再度MMS用紙に回答してもらおう。以下、課題2、課題3についても同様に、課題の前後でMMS

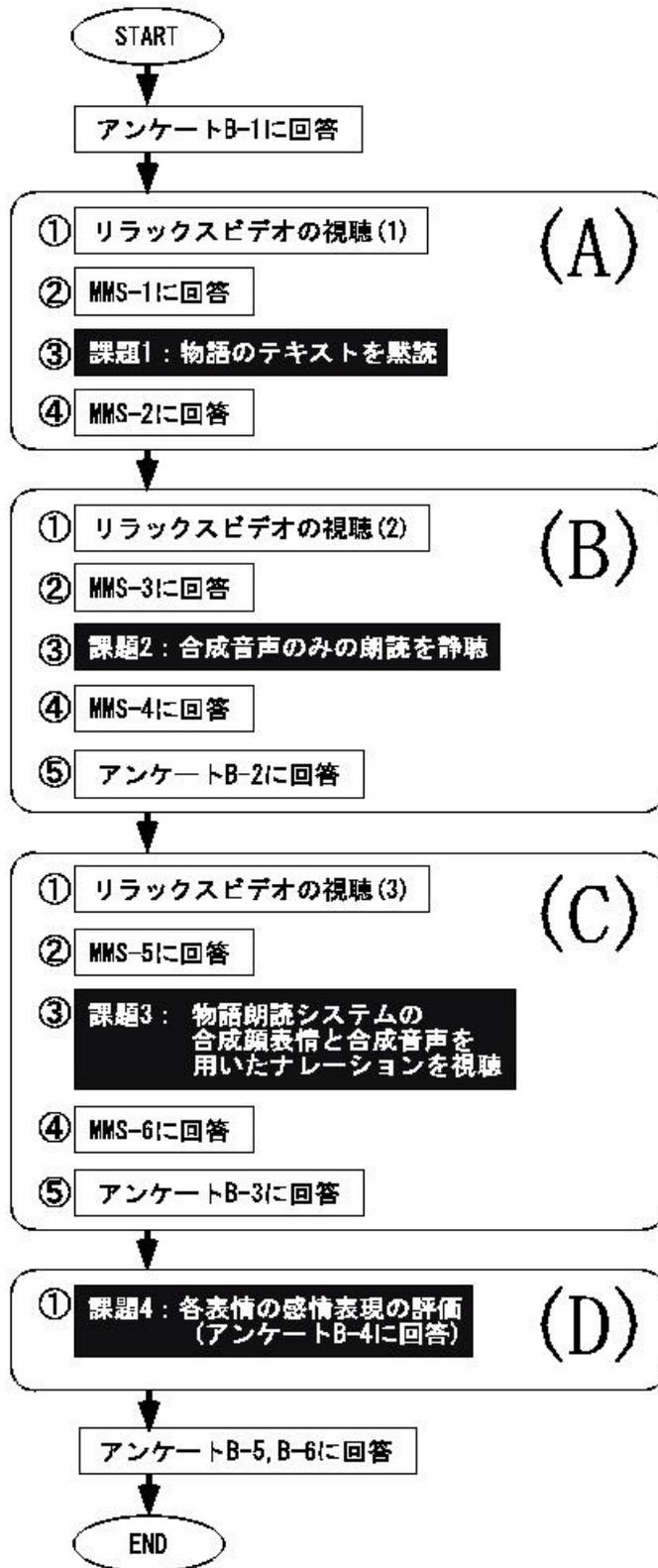


図 5.1: 実験の手順

用紙に回答してもらおう。パート(B)では、課題2、すなわち合成音声のみの朗読を聞いてもらい、その後、課題1と課題2を比較するアンケートB-2に回答してもらおう。パート(C)では、課題3、すなわち物語朗読システムを視聴してもらった後、システムを評価するアンケートB-3に回答してもらおう。

パート(D)では、課題3の朗読に含まれる感情を表現している部分を視聴してもらい、それらの朗読が感情を伝達しているか、またそれはどのような感情かをアンケートB-4に記入してもらおう。最後に、「コンピュータに対する態度」を調べるために、アンケートB-5に回答してもらおう。

表5.1にアンケート実施時とそのアンケート名をまとめる。また、アンケート用紙

表 5.1: アンケート実施時とアンケート名

アンケート実施時	アンケート名
課題1開始前 (MMS-1 回答前)	B-1
課題2終了後 (MMS-4 回答前)	B-2
課題3終了後 (MMS-6 回答前)	B-3
課題4実施中	B-4
課題4終了後	B-5, B-6
課題1直前	MMS-1
課題1直後	MMS-2
課題2直前	MMS-3
課題2直後	MMS-4
課題3直前	MMS-5
課題3直後	MMS-6

B-1～B-5を付録Hに、MMS用紙を付録I示す。以下では、各アンケートの目的と方法について説明する。

アンケート B-1

目的 被験者の属性を知ること。

方法 用紙に氏名、年齢、性別、専攻分野などを記入してもらおう。

アンケート B-2

目的 課題1の黙読と課題2の合成音声での朗読とを比較して、朗読する物語文が表現している感情をどちらの方が伝達できたか、すなわちどちらが感情伝達できたか、また、合成音声での朗読に対して感情伝達の程度や朗読方法の評価を知ること。

方法 「黙読と朗読について、どちらが感情を伝達できたか」、および「合成音声での朗読の音質、朗読方法、間(ま)はどうだったか」という質問に対して、7段階(-3 ~ +3)で評価してもらう。

アンケート B-3

目的 アンケート B-3は、以下の項目について知ることを目的とした。

1. 物語朗読システムを用いた朗読を、合成音声のみの朗読と比較して、物語文の表現する感情をどちらの方が伝達できたか
2. 物語朗読システムの朗読は、内容に応じた感情を表現していたか
3. 物語朗読システムの表情に関して、変化のタイミングは適切であったか
4. 物語朗読システムの表情に関して、速さは適切であったか
5. 物語朗読システムの表情に関して、保持時間は適切であったか
6. 物語朗読システムの表情に関して、リップシンクは適切であったか
7. 物語朗読システムを用いた朗読の内容、すなわち物語文の表現する感情に共感できたか
8. 物語朗読システムの朗読は自然であったか
9. 物語朗読システムの朗読は人間らしかったか
10. 物語朗読システムの朗読が好きか嫌いか
11. 将来のコンピュータが、物語朗読システムのように発話しながらそれに応じた表情を提示してくれることを望むか
12. アニメ的な顔画像と物語朗読システムでの顔画像とを比較して、どちらが親しみやすいか
13. アニメ的な顔画像と物語朗読システムでの顔画像とを比較して、どちらが人間らしいか

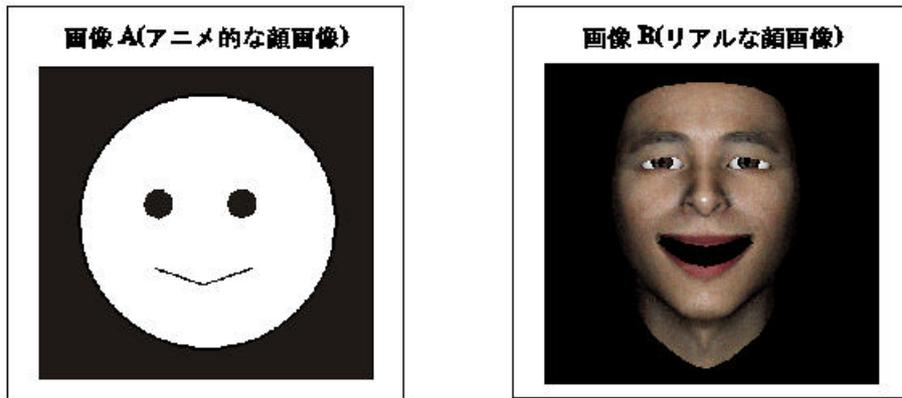


図 5.2: アニメ的な顔画像 (A) とリアルな顔画像 (B)

14. アニメ的な顔画像と物語朗読システムでの顔画像とを比較して、どちらに信頼感を感じるか
15. アニメ的な顔画像と物語朗読システムでの顔画像とを比較して、どちらが好きか

方法 上記の目的に対応した質問をアンケートにより行い、7段階 (-3 ~ +3) で評価してもらう。なお、目的 12 ~ 15 で提示したアニメ的な顔画像とリアルな顔画像 (物語朗読システムで用いる顔画像) を図 5.2 に示す。

アンケート B-4

目的 物語朗読システムを用いて感情を表現している朗読の一部分を被験者に視聴してもらうことにより、その部分がどの程度伝達したい感情を効果的に伝えているかを知ること。

方法 課題 3 での朗読中、感情を表現していた部分 (47 部分) をすべて取り出し、それぞれを提示した後、「感情を表現しているかどうか」を 7 段階 (-3 ~ +3) で評価してもらい、「どの感情を表現しているか」を 35 種類の感情の中から選択してもらう。提示する順序は、物語に沿った順序ではなく、あらかじめランダムに順序を入れ替えている。

アンケート B-5

目的 被験者のコンピュータに対する態度、すなわち「コンピュータ操作不安」および「コンピュータ接近願望」を知ること。

方法 コンピュータに関する15項目の質問に回答してもらい、川嶋らがコンピュータに対する態度を調べるために行った因子分析の結果^[35]を利用して、「コンピュータ操作不安」、および「コンピュータ接近願望」を求める。

アンケート B-6

目的 本実験全体に関する感想を聞くこと。

方法 実験に対する感想を自由記述してもらう。

多面的感情状態尺度 (MMS)

目的 各課題(課題1~3)の前後で感情状態の変化があるかどうかを調べることにより、物語文の表現する感情を伝えることにより物語に感情移入してもらうことができたか、すなわち共感をよびおこしたかどうかを調べること。具体的に、本実験では、物語の内容が「悲しい」内容(「マッチ売りの少女」)であるため、「悲しい」もしくは「悲しい」に関係する感情の変化が起こるかどうかを調べること。

方法 本実験で心理評価手段として用いた多面的感情状態尺度について述べる。本実験では、感情状態の評定尺度として、MACLを日本語化し、さらに因子分析により質問項目を40項目に削減した短縮版の多面的感情状態評価尺度(Multiple Mood Scale; MMS)^[36]を用いる。多面的感情状態評価尺度では、感情状態は8つに分類され、それら8つの感情は正の感情、負の感情および中性的な感情の3つの範疇のいずれかに属する。正の感情状態として、「活動的快」、「非活動的快」、「親和」の3つ、また負の感情状態として、「抑うつ・不安」、「敵意」、「倦怠」の3つ、さらに中性的感情状態として「集中」、「驚愕」の2つ、合計8つの感情状態に分類しておき、感情状態がそのいずれにあるかを質問するため、各々5項目の設問で構成する。具体的な設問内容を表5.2に示す。

設問に対する回答方法として、

1. 「全く感じていない」(-2点)

2. 「あまり感じていない」(-1点)
3. 「少し感じている」(1点)
4. 「はっきり感じている」(2点)

の尺度で評定を行い、各感情状態に関する設問に対する回答の合計点を算出する。

表 5.2: MMS の項目一覧

抑うつ・不安	不安な	敵意	敵意のある
	悩んでいる		憎らしい
	気がかりな		うらんだ
	自信がない		攻撃的な
	くよくよした		むっとした
倦怠	だるい	活動的快	活気のある
	疲れた		気力に満ちた
	つまらない		元気いっぱい
	退屈な		はつらつとした
	無気力な		陽気な
非活動的快	のんびりした	親和	恋しい
	おっとりした		いとおいしい
	ゆっくりした		好きな
	のどかな		愛らしい
	のんきな		すてきな
集中	丁重な	驚愕	びっくりした
	慎重な		驚いた
	ていねいな		びっくりとした
	注意深い		動揺した
	思慮深い		はっとした

5.2.2 被験者

被験者は、大学生及び大学院生の男性 6 名、女性 14 名の計 20 名である。

5.3 実験の結果

本実験では、一室に被験者全員を集め、一斉に実施した。そのため、課題2、3、4で提示する音声や顔表情は事前に試作システムを用いてビデオテープに記録したものを、実験室の60インチモニタとスピーカにより被験者に提示した。以下では、まず、アンケートB-2、B-3、B-4、B-5、B-6について、それぞれの結果を述べる。なお、アンケートB-1は個人的属性を問うものであり、その詳細な回答は割愛する。

5.3.1 合成音声のみの物語文朗読による感情伝達の評価

結果

表5.3にアンケートB-2の質問文と、その結果を示す。ここで、質問番号1の「感情を伝達」は、システムから被験者に物語文の表現している感情を伝達できたかどうかを尋ねている。表中には、20名の被験者の回答の平均と標準偏差を示している。また、課題2について回答欄の欄外に任意に書いてもらった感想を表5.4に示す。

表 5.3: アンケート B-2 の結果

質問番号	質問文	評価点	
		平均	標準偏差
1	音声で読み上げた朗読 (課題2: 音声朗読) を物語のテキストを黙読した場合 (課題1: テキスト黙読) と比較して、感情をよく伝達したのはどちらだと感じられましたか? (-3: テキスト ~ +3: 音声朗読)	-2.1	1.4
2	感情を伝達するという観点からみて、音声で読み上げた朗読に関して、各項目を評価してください。		
	・音質 (-3: 不十分 ~ +3: 十分)	-1.6	1.5
	・朗読方法 (-3: 不十分 ~ +3: 十分)	-2.4	0.7
	・文章と文章の間 (ま) (-3: 不適切 ~ +3: 適切)	-2.1	1.4

結果のまとめ

上記の結果より、合成音声を用いた朗読は、物語文の黙読と比較して感情を伝達できないことがわかる (平均-2.1)。アンケート結果より、この原因として、合成音声の音

表 5.4: 課題 2 の感想

被験者番号	感想
3	音がとんでいて、聞きにくい場所がいくつかあった。棒読みであまり感情が伝わらなかった。
4	はじめは、人が朗読している感じがしていたけれど、しばらくすると文章のとぎれるところが変わったりして聞きづらかった。
6	音質が機械的 (人工的) な声だったことと、文章の間のとぎれ方がおかしかったので、感情が伝わらなかった。冷酷な印象を受けた。
9	感情がなさすぎて、聞いていて何もつたわってこなかった。イントネーションが無い。
14	発音・アクセントがわかりにくく、聞き取りにくい部分があった。また、文章が途中変な部分で切れたり、逆にさらっと流れたりして、聞き取りにくいと感じた。
18	「。」で終わっていないと思われる所で、妙に長い間があった。読みの区切れが変なところがあった。

質 (平均-1.6)、特に朗読方法 (平均-2.4)、次いで文章と文章の間 (平均-2.1) が悪いということが考えられ、それらが大きく影響していることと考えられる。

5.3.2 物語朗読システムを用いた物語文朗読による感情伝達の評価

結果

図 5.3 に課題 3 の実験風景を示し、表 5.5 にアンケート B-3 の質問文と、その結果を示す。表中には、20 名の被験者の回答の平均と標準偏差を示している。また、回答用紙の欄外に任意で書いてもらった課題 3 の感想を表 5.6 に示す。

結果のまとめ

アンケート結果より、それぞれの質問番号に対応して以下のようにまとめられる。

- 1, 2 物語朗読システムを用いた朗読は、音声のみの朗読に比べて感情が表現できている (平均 1.3) と考えられる。これより、合成音声と合成顔表情を用いて、感情を伝達できる可能性が示されている。しかし、質問番号 2 の回答 (平均-0.7) より、内容に応じた感情伝達は難しいことがわかり、感情推定方法が適切ではなかったと考えられる。



図 5.3: 課題 3 の実験の風景

表 5.5: アンケート B-3 の結果

質問番号	質問文	評価点	
		平均	標準偏差
1	音声のみの朗読 (課題 2 : 音声朗読) と顔表情付きナレーション (課題 1 : 顔付き) を比較して、どちらが感情を表現していましたか? (-3:音声朗読 ~ +3:顔付き)	1.3	1.2
2	物語の内容に応じた感情を表現していると感じましたか? (-3:表現していない ~ +3:表現している)	-0.7	1.7
3	表情の変化のタイミングは適切でしたか? (-3:不適切 ~ +3:適切)	-0.6	1.7
4	表情の変化する速さは、適切でしたか? (-3:不適切 ~ +3:適切)	-0.7	1.6
5	表情の保持時間は、適切でしたか? (-3:不適切 ~ +3:適切)	-0.4	1.6
6	発話に対して、唇の動きは、自然でしたか? (-3:不適切 ~ +3:適切)	-2.5	0.8
7	実験 3 のナレーションに共感できましたか? (-3:共感できない ~ +3:共感できた)	-1.5	1.3
8	実験 3 のナレーションは自然だと感じましたか? (-3:不自然 ~ +3:自然)	-2.2	0.8
9	実験 3 のナレーションに人間らしさを感じましたか? (-3:人間らしくない ~ +3:人間らしい)	-2.1	1.2
10	実験 3 のようなナレーションは、好きですか、嫌いですか? (-3:嫌い ~ +3:好き)	-1.9	1.3
11	実験 3 のようなナレーションがコンピュータのディスプレイ上に現れて対応してくれるようになることを期待しますか? (-3:期待しない ~ +3:期待する)	-0.3	2.4
12	下記の画像 A(アニメ的な顔画像) と画像 B(実験 3 で用いたリアルな顔画像) とを比較して、どちらに親しみやすさを感じますか? (-3:画像 A ~ +3:画像 B)	-1.8	1.6
13	下記の画像 A と画像 B を比較して、どちらに人間らしさを感じますか? (-3:画像 A ~ +3:画像 B)	1.6	1.8
14	下記の画像 A と画像 B を比較して、どちらに信頼感を感じますか? (-3:画像 A ~ +3:画像 B)	0.3	2.0
15	下記の画像 A と画像 B を比較して、どちらが好きですか? (-3:画像 A ~ +3:画像 B)	-1.2	1.9

表 5.6: 課題 3 の感想

被験者番号	感想
1	顔付き朗読は、あまりにも表情全体が不自然で不気味だった。顔の筋肉はゆるやかに変化するものだが極端すぎて親しみやすさは感じられず、恐いとばかり感じる。ナレーションも文の区切りが不自然で聞きづらかった。ナレーションだけで表現できるようにするともっと良いと思う。
2	リアルな顔が気持ち悪いと思う反面、妙な顔の変化がおもしろかった。
3	話し終わっても、まだ口が動いているので、不自然な感じがした。目がずっと一定を見ているだけなので、恐かった。目が死んでいる感じがした。
4	アニメ的な顔とリアルな顔とを比べると、リアルな顔の方が人間的だとはおもうけれど、見なれていないこともあって不自然な感じがする。それに比べて、アニメ的な顔はテレビのアニメなどでよく見るような感じだから人間的でなくても身近さを感じる。
5	ナレーションの顔が、その時の話の状況と合っていなかったと思う。また、文章の区切れが不自然だった。
6	リアクションがオーバーすぎた。
9	顔の表情と音声が全く合っていなかったのを見ていてきもちわるかった。表情は豊かだが、不適切であるし、不自然である。唇の形が不自然すぎる。機械、映像的である。
13	うれしい表情などがあまりそのようには見えなかった。とても不自然に思った。
14	「うれしい」などの (+) の感情は顔の表情からある程度うまく伝わってきた。逆に「悲しい」などの (-) の感情は、顔の表情とマッチしていない部分もあったように思われる。また、話の内容上、大体は素の顔である部分だけ表情がかわると全体としての雰囲気損なわれるように感じた。話の流れや内容に合わせて表情を変え、その中でさらに表情を変えるともっと効果的かもしれない。
16	リアルな顔画像について、唇が厚い、歯がない、口自体の大きさがコロコロ変わる、三白眼である、そのため、へんにリアルで人間らしくないと思った。
19	もっと目元の変化も表情には必要だと思う。口の動きが不自然すぎる。

- 3~6 顔表情合成に関して、変化のタイミング(平均-0.6)や変化の速さ(平均-0.7)、保持時間(平均-0.4)が被験者には適切に感じられなかったと考えられる。加えて、リップシンクに関しては、発話内容とタイミングが合わなかったため、非常に不適切に感じられた(平均-2.5)ということが示されている。
- 7 物語朗読システムを用いた朗読は、共感をよぶことが難しかった(平均-1.5)と考えられる。
- 8~10 物語朗読システムを用いた朗読は、あまり自然さ(平均-2.2)や人間らしさ(平均-2.1)が感じられず、また、好まれない(平均-1.9)ということが示されている。
- 11 平均値が0に近く(平均-0.3)、標準偏差が大きいことより、このような朗読の実現を望む人も望まない人もいるということが考えられ、個人の好みに依存することが示されている。
- 12~15 アニメ的な顔画像と比較すると、本システムで用いた顔画像は、あまり親しみやすさが感じられず(平均-1.8)、好まれない(平均-1.2)と考えられる。しかし、信頼感(平均0.3)に関しては、アニメ的なものと同じ程度であり、人間らしさ(平均1.6)に関しては本システムで用いた顔画像のほうが感じられると考えられる。これらの回答は、本システムで用いた顔画像は、非常にリアルであるのにも関わらず、白目部分が大きいなど、一部分がデフォルメされているため、アニメ的な単純な顔画像に比べて、どこが不自然なのか被験者自身にもわからないが不自然さが感じられるためだと考えられる。また、同様に本システムの顔画像がリアルなために、人間らしさが感じられ、信頼感が生じると考えられる。

5.3.3 各表情の感情伝達の評価

結果

図5.4に課題4の実験風景を示し、表5.7、表5.8、および図5.5にアンケートB-4の結果を示す。表中の「セリフ」とは表情の提示とともにシステムが発話した内容であり、「感情」とは伝えたい感情を示している。また、「正答者数」とは、20人中正しい感情を選択した人数であり、「感情伝達度」とは、各表情に対して「選択した感情がどの程度表現できていたか」という項目に7段階(-3~+3)で回答してもらった結果である。図5.5は、横軸が「感情伝達度」(7段階;-3~+3)、縦軸が正答者数であり、1~47までの各番号の結果をプロットしたものである。



図 5.4: 課題 4 の実験の風景

表 5.7: アンケート B-4 の出題内容と結果 (1)

番号	出題内容		結果		
	セリフ	感情	感情選択	感情伝達度	
			正答者数	平均	標準偏差
1	ほっとしました。	満足	5	-0.8	1.9
2	ずきずきしていました。	苦痛	12	1.4	1.6
3	強がりました。	プライド	7	-1.8	1.6
4	がっかりしていました。	失望	10	1.0	1.6
5	しんみりとしていました。	寂しさ	13	0.7	1.7
6	耐えました。	忍耐	9	-0.6	1.4
7	やるせなくなりました。	耐えられない	3	1.2	1.5
8	にこりしました。	喜び	11	-1.2	1.9
9	かわいそうな一人の少女が歩いて いました。	哀れみ	14	0.8	1.6
10	憂えました。	ぼんやり	0	-0.9	1.5
11	うっとりしました。	あこがれ	9	1.3	1.4
12	びっくりすることに、今度はあっ というまもありませんでした。	驚き	9	0.2	2.0
13	がっかりしました。	失望	10	0.7	1.8
14	いやだとおもいました。	嫌い	11	1.3	1.5
15	哀れでした。	哀れみ	14	0.8	1.8
16	びっくりしました。	驚き	13	-0.4	2.4
17	笑っていました。	喜び	3	-0.2	1.6
18	心残りでした。	後悔	9	-2.3	1.4
19	案じていました。	不安	13	0.3	1.9
20	にっこりしました。	幸福	6	0.6	2.0
21	ふさいであるだけ。	失望	2	0.2	1.7
22	冷笑します。	侮辱	9	-0.2	1.9
23	なつかしいおばあさんはもういま せん。	暖かい	1	-1.1	1.9
24	泣きそうでした。	悲しみ	12	2.0	1.5

表 5.8: アンケート B-4 の出題内容と結果 (2)

番号	出題内容		結果		
	セリフ	感情	感情選択	感情伝達度	
			正答者数	平均	標準偏差
25	ぬくぬくするかもしれません。	暖かい	12	1.6	1.1
26	あこがれていました。	尊敬	1	-1.2	1.5
27	いとおいしいと思ってくれたたった一人の人でした。	愛情・好き	9	1.8	0.8
28	つまらなく思いました。	退屈	15	1.7	1.4
29	悔しく感じました。	屈辱	8	-0.5	1.8
30	笑みが浮かびました。	幸福	2	0.6	1.8
31	笑っていました。	喜び	7	-0.7	1.3
32	かつとし、ぜったいホッペをぶつにちがいありません。	怒り	11	1.8	1.1
33	いきりたち、	憎悪	5	-0.1	1.5
34	じいんとしました。	暖かい	5	1.6	1.3
35	いたわってくれました。	優しい	8	1.1	1.5
36	むくれています。	理不尽	0	0.9	1.6
37	うらやましくなりました。	嫉妬	11	-0.7	2.4
38	くたびれていました。	ぼんやり	3	-0.2	1.9
39	むずがゆく感じました。	不愉快	4	1.4	1.7
40	ドキドキしました。	好奇心	10	-0.4	1.9
41	びくびくしました。	恐れ	11	2.0	0.9
42	はにかみました。	恥ずかしい	5	-1.8	1.6
43	迷いました。	ためらい	7	0.2	1.9
44	うれしいかもしれません。	喜び	10	-0.2	2.0
45	ねえ、わたしをいっしょに連れてってくれるの？	疑問	5	-1.1	2.1
46	うきうきしました。	楽しい	7	1.3	1.3
47	驚きました。	幸福	1	-1.0	1.7
平均			7.7	0.0049	1.63

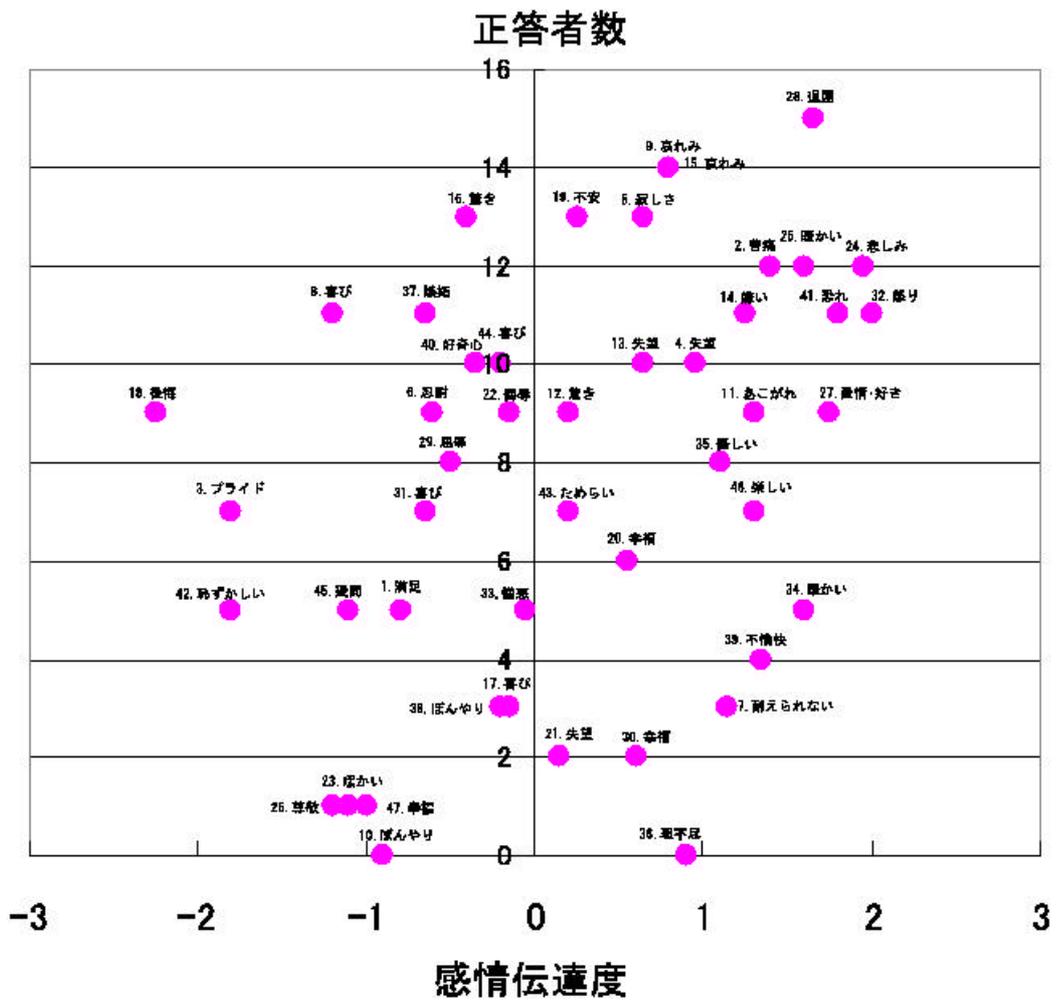


図 5.5: アンケート B-4 の結果 (全体)

結果のまとめ

アンケート結果より、発話内容とその対応する感情が一致していない、すなわち感情推定方法がうまく機能していない回答(回答番号 10、23、26、47)では、正答者数が少なく、感情伝達度が小さいことがわかる。また、発話内容とその対応する感情がうまく一致している回答(回答番号 2、25、24、28)は、正答者数が多く、感情伝達度が大きいことがわかる。

これより、被験者はまず、発話内容を聞いて表現すべき感情を選択し、提示された表情が選択した感情をどの程度表現しているかを評価したと考えられる。

以上の結果より、まず、正答者数が少ない回答は、発話内容と感情の対応がうまくとれていないと考えられ、これは感情推定方法に問題がある。一例を挙げると、提案した感情推定手法では、文中に否定語が含まれている場合に対応していないため、回答番号 23 の場合には、「…もういません」となっているにもかかわらず、「暖かい」感情を表現し、推定がうまく機能していない。このことから、物語文から自動的に感情を推定するためには、否定語の存在を考慮する必要があると考えられる。

一方、伝えたい感情をどの程度効果的に伝えることができたか、すなわち感情伝達度を見るためには、表情から正しい感情を選択できた人の感情伝達度を見る必要がある。図 5.6 は、横軸が「正答者のみの感情伝達度の平均」(7 段階;-3 ~ +3)、縦軸が正答者数であり、1~47 までの各番号の結果をプロットしたものである。

図 5.6 より、基本 6 感情 (Ekman による^[19]) のうち、「悲しみ」、「恐れ」、「怒り」、「嫌悪 (嫌い)」は、感情伝達度が 1 以上となっている。これより、基本 6 感情の中でもネガティブな感情を示す表情は、表情を作成しやすく、また作成した表情は理解されやすい、と考えられる。一方、「幸福」については、感情伝達度にばらつきがあり、「驚き」の感情伝達度は 0 に近い。正答者数が 1 である「47. 幸福」以外は、「幸福」の感情伝達度も 1 以上となっている。「驚き」については、Ekman の定義による「驚き」の表情の大きな特徴である、『顎を下げ唇を開く』^[19] が「驚き」の表情に表れていない(図 3.11 参照) ため、感情伝達度が低くなったと考えられる。

5.3.4 コンピュータに対する態度 (操作不安と接近願望) 調査

結果 表 5.9 の因子分析の結果^[35]に基づいて、表 5.10 と図 5.7 にアンケート B-5 の結果を示す。

結果のまとめ 図 5.10 より、操作不安が正となった被験者は 5 名 (被験者 4、7、10、12、

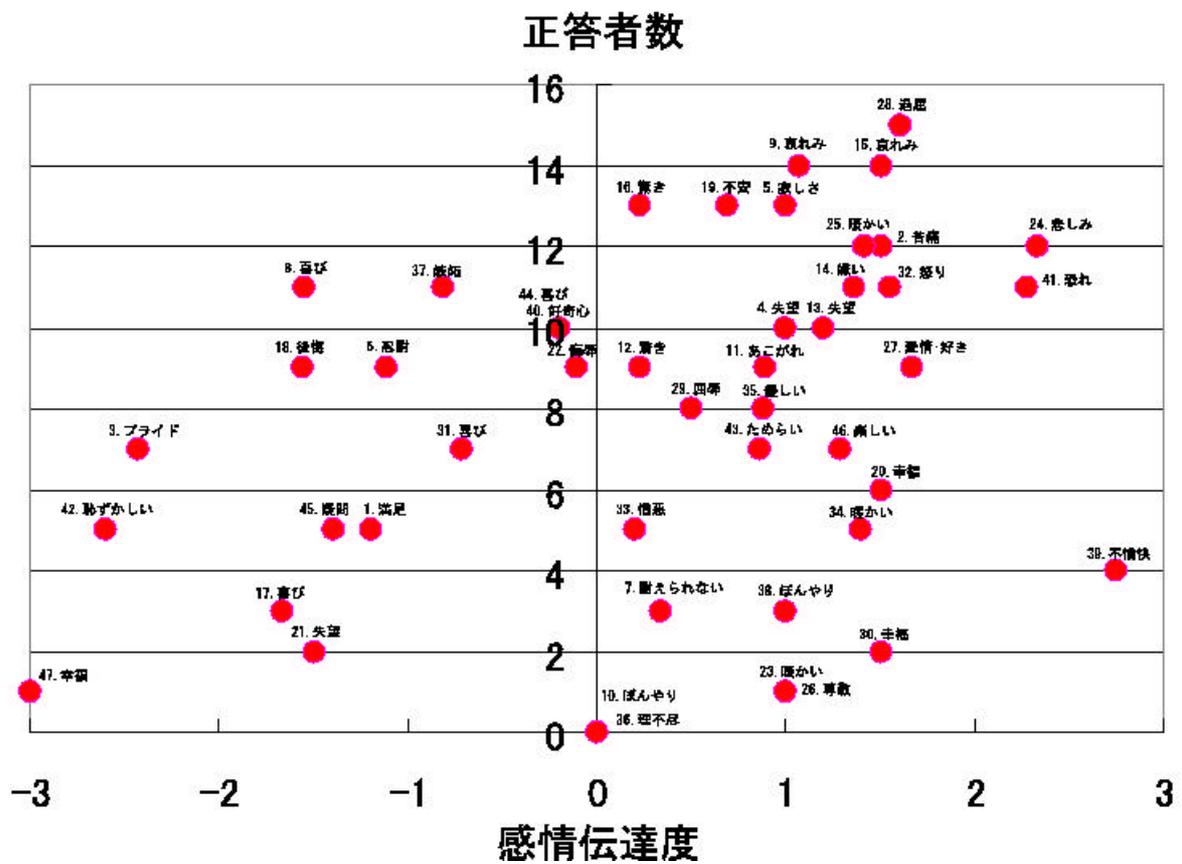


図 5.6: アンケート B-4 の結果 (正答者のみの伝達度)

表 5.9: 「コンピュータに対する態度」の因子分析結果^[35]

番号	項目	操作不安	接近願望
1	人が見ている前でコンピュータを操作すると恥をかきそうだ	0.84	-0.28
2	コンピュータの前に座っただけでとても緊張してしまう	0.80	0.17
3	コンピュータを使うのは不安だ	0.77	-0.18
4	コンピュータと聞いただけでももうお手上げの状態である	0.77	0.28
5	コンピュータに慣れていないので使うのをためらう	0.73	-0.11
6	まちがいをするのが怖くてコンピュータを使うのをためらう	0.71	-0.10
7	コンピュータを使用するとき操作を誤って何かを壊しそうな気がする	0.71	0.05
8	コンピュータのキーボードを見るとうんざりする	0.66	-0.26
9	違うキーを押してたいせつな情報を消してしまうのが恐ろしい	0.65	-0.09
10	コンピュータについてもっと知りたいと思う	-0.29	0.77
11	コンピュータを操作している人を見ると自分も早くそうなりたいと思う	-0.04	0.69
12	これからの社会ではコンピュータについて何も知らないことは恥ずべきことだ	0.16	0.66
13	将来的にコンピュータを使う可能性がある	-0.31	0.41
14	コンピュータについてもっと知りたいと思う何も知らないと思われ ても平気だ	0.30	-0.63
15	コンピュータの利用は得意な人に任せておけばよい	0.21	-0.69

表 5.10: アンケート B-5 の結果

被験者	操作不安	接近願望
1	-5.35	10.34
2	-19.20	12.33
3	-13.63	8.33
4	3.21	9.47
5	-4.52	8.23
6	-0.50	4.97
7	3.78	8.60
8	-13.41	3.38
9	-16.41	13.94
10	0.39	4.07
11	-20.09	14.54
12	2.61	5.86
13	-5.58	2.65
14	-1.28	10.31
15	-19.02	6.23
16	4.33	7.83
17	-7.47	2.79
18	-17.15	2.39
19	-15.44	7.16
20	-14.44	13.86
平均	-7.96	7.86
標準偏差	8.68	3.88

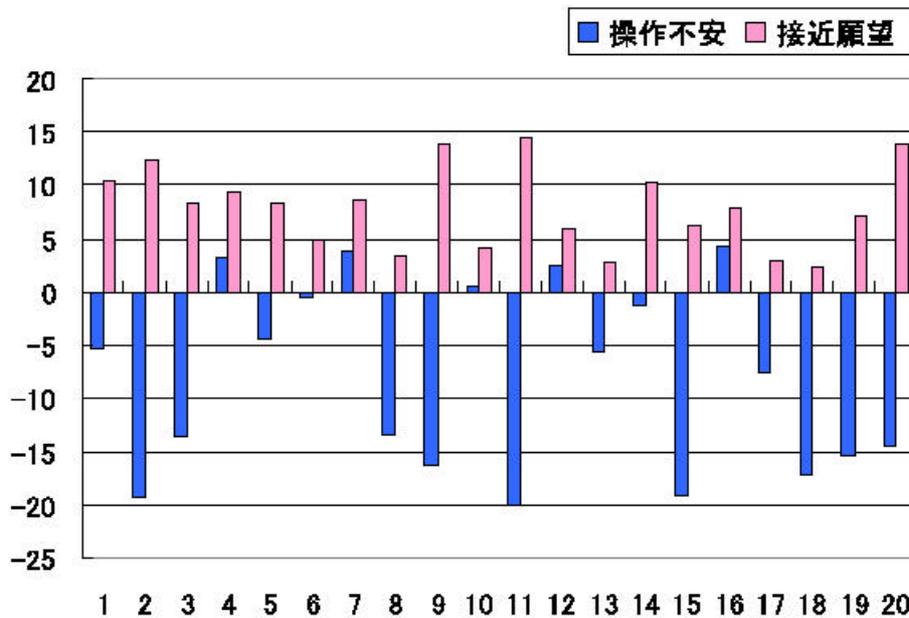


図 5.7: アンケート B-5 の結果

16) のみであり、接近願望が負になった被験者はいなかった。これより、本実験に参加した被験者は、コンピュータに対する操作不安が比較的小さく、接近願望が大きい人が多い。近年の大学生はコンピュータに接する機会が多いと予想され、コンピュータに対して何らかの関心を持っていると考えられる。

5.3.5 実験全体の感想

結果 アンケート B-6 で尋ねた実験の感想を表 5.11 ~ 表 5.12 に示す。

結果のまとめ アンケートより、物語朗読システムを用いたナレーションが今後出現することへの期待や不安、また、表情の不自然さに対して、口の動き、目の動き、表情の左右対称性などの意見が挙げられている。また、提示された顔そのものが不自然であり、親しみにくいという意見が多くみられる。

5.3.6 多面的感情状態尺度 (MMS)

結果 課題 1 ~ 3 の前後で、各課題により感情状態がどのように変化したかを知るために行った MMS の各回ごとの結果を表 5.13 に、各課題前後の差の平均を表 5.14 に、

表 5.11: アンケート B-6 の結果 (1)

被験者番号	感想
1	今後こういうもの(物語の人)ができるといいなあと思った。
2	興味深かった。表情の変化がおもしろかった。
3	目を変化させた方がいい。
4	このような画像を見慣れていないから、リアルすぎて笑ってしまった。これからの時代はこういう風な画像が当たり前になっていくのかなあと思うと複雑な気持ちになる。
5	かなり人間とコンピュータが近づいてきたと思うが、まだ人間には人間特有の暖かさや表情があり、コンピュータでは伝わってこない部分があると思う。これからコンピュータがあらゆるところで活用されると思うが、すべてコンピュータに委ねてしまっているのか、不安が残る。
6	合間にリラックスできるビデオを見れてよかった。
7	顔の映像がもっと豊富になってほしい。黒目を動かしてほしい。まだ人工的で見ている人に不快感を与えてしまいがちになる。
8	顔の表情を表現ごとに変えてあっておもしろかった。顔の各器官をいろんな組み合わせを変えていくことで様々な感情表現ができることに改めておどろいた。
9	人間の表情は表現しにくいものだった。人とコミュニケーションを図ったり、何かを相手に伝える時に言語(言葉)と同じくらい表情も大切、むしろ表情がないと駄目。
10	リラクゼーションのVTRも短時間ではあまり気分は変わらないのではないかと考えた。黙読だけと音声によるものでは同じ内容でも全く違った感じに受けとめるものなのだなあと考えた。
11	もっとほほが動いたら表情に幅がでてくると思う。1回ごとに真顔に戻るのも、その時に感情が途絶えてしまっている気がする。人が話しているように表現するのは同時にいくつものことをしなくてはいけないのでとても難しいと思う。実社会で使われるようになったら素敵だと思う。
12	とても感動した。近い未来にコンピュータが今日のようになせるようになり、コンピュータだけでなくロボットも登場したら素晴らしいと思った。

表 5.12: アンケート B-6 の結果 (2)

被験者番号	感想
13	コンピュータを使ってこういうナレーションができることに驚いた。ただ、とても不自然でやはり人の表情は豊かであるということに改めて気付いた。人にしかできないと思う。コンピュータがこれからもっと発達しても人の上をいくものはない気がする。
14	1つ1つの表情と朗読の流れの中での同じ表情とは、多少感じ方が違うものであると思う。
15	合成表情の口の動きとせりふとがきちんと同期していないため、違和感を強く感じた。また、B-4の問題では口がセリフに合わせて動いてしまうため、表情を目や眉といった顔の上半分ではしか判定することができなかった。
16	リアルな顔は恐く、じっと見ていられない。かえって不快に感じる。もう少し自然な表情になればと思う。話している間と口の動いている間を同じにした方がいいと思う。
17	目の動きで受ける印象がかなりちがうと思う。ナレーションのときに、感情の顔で表現したあとに常に同じ表情にもどるのが不自然に感じられた。
18	TVで朗読というのはあまり見たことがないので、顔の表情の変化を作るシステム作りには朗読ではなく、他の顔の表情を見せるもので作った方がよいのではと思う。例えば落語のように演者が複数の役を行うものであれば、表情の変化というものを受け入れやすいと思う。また、左右非対称にはできないものか？
19	人間の感情や表情といったものをコンピューターに出すのはまだまだ難しいようだと思った。表情やしぐさは地域性があると思うがどうだろうか。
20	途中の心を落ち着かせるための映像をある程度省略できないか？

グラフを図 5.8 示す。また、課題 1～3 の前後の比較を、図 5.9～図 5.11 に示す。

表 5.13: MMS の結果 (各回ごと)

		抑うつ・不安	倦怠	非活動的快	集中	敵意	活動的快	親和	驚愕
MMS1	平均	-2.85	0.30	3.70	-0.70	-8.35	-2.15	1.45	-6.50
	標準偏差	5.24	5.11	5.16	5.03	2.70	6.32	5.75	3.95
MMS2	平均	-1.95	-1.40	-1.65	-4.95	-6.65	-6.00	1.15	-4.15
	標準偏差	4.17	4.73	5.44	4.17	4.09	5.70	5.04	4.61
MMS3	平均	-5.70	0.30	4.60	-4.75	-8.90	-3.00	1.25	-7.90
	標準偏差	3.74	5.58	5.83	4.20	2.17	4.47	6.00	2.53
MMS4	平均	-3.50	2.75	-1.60	-4.35	-5.45	-7.75	-5.55	-6.35
	標準偏差	4.39	4.84	3.97	5.16	4.47	2.57	2.89	3.79
MMS5	平均	-6.90	-1.80	5.10	-4.35	-8.70	-3.40	0.50	-7.90
	標準偏差	3.92	5.15	4.81	4.57	2.70	4.57	5.81	2.38
MMS6	平均	-5.05	-1.10	-3.65	-4.00	-5.60	-4.60	-4.50	-3.1
	標準偏差	4.30	5.52	4.25	4.78	4.90	5.00	4.71	6.11

結果のまとめ 課題 1～3 の前後で主な差が見られる非活動的快・活動的快・親和・驚愕について、それぞれ述べる。

非活動的快 課題 1、2、3 と進むにつれて減少が大きくなっている。これは、課題 2 については合成音声、課題 3 についてはさらに顔表情が加えられ、刺激が大きくなっていったためと考えられる。

活動的快 課題 1 の前後に比べて、課題 2 の前後で大きく下がっている。まず、課題 1 の前後では実験の終了による活動性の低下と見られる。一方、課題 2 の音声のみでの朗読の評価が低く、朗読時間が長かった (15 分) ことから、被験者の意欲が薄れたことが原因と見られる。ここで、課題 3 の前後ではあまり変化がないのは、顔表情が提示されることによって比較的刺激があったためだと考えられる。

親和 課題 2 の前後で大きく減少している。これは、合成音声での朗読が不自然で、親しみを感じられなかったためだと考えられる。また、課題 3 の前後でも減

表 5.14: MMS の結果 (各課題前後の差の平均)

		抑うつ・不安	倦怠	非活動的快	集中	敵意	活動的快	親和	驚愕
A	平均	0.90	-1.70	-5.35	-4.25	1.79	-3.85	-0.30	2.35
	標準偏差	6.10	4.33	6.66	5.20	3.90	8.18	4.51	4.04
B	平均	2.20	2.45	-6.20	0.40	3.45	-4.75	-6.80	1.55
	標準偏差	4.89	3.63	6.31	3.87	3.78	3.45	6.86	3.36
C	平均	1.85	0.70	-8.75	0.35	3.10	-1.20	-5.00	4.80
	標準偏差	4.13	5.00	5.67	5.04	4.36	5.68	5.43	4.60

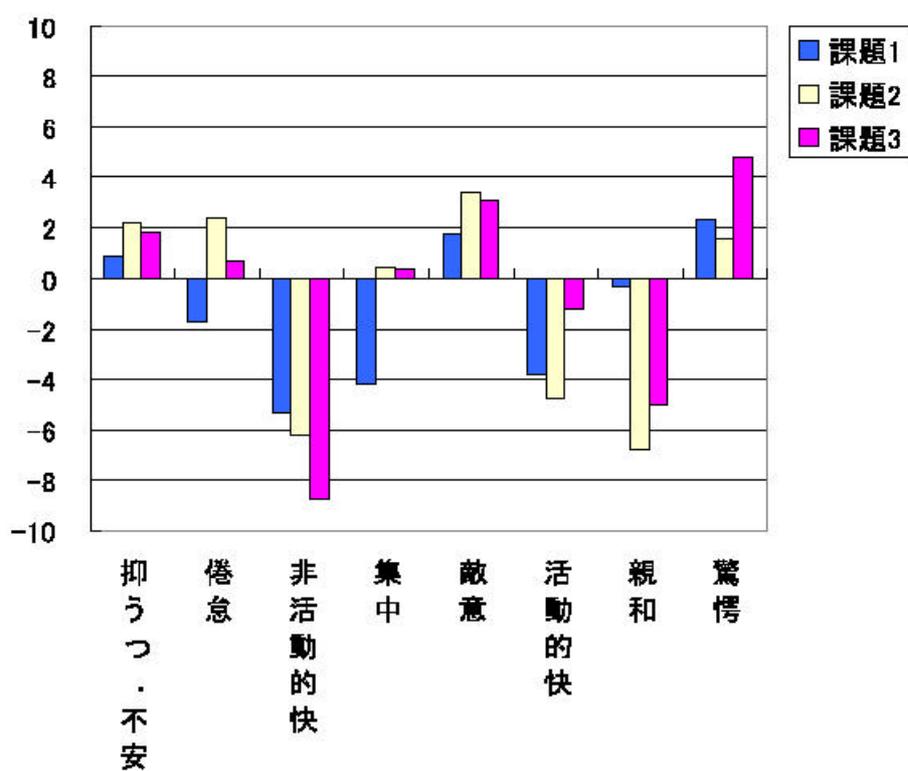


図 5.8: MMS の結果 (各課題前後の差の平均)

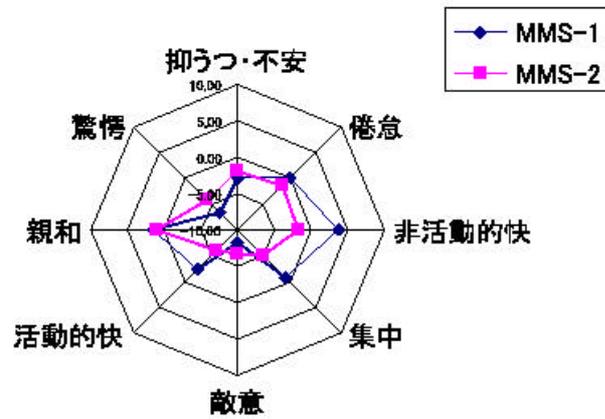


図 5.9: MMS の結果 (1 と 2 の差)

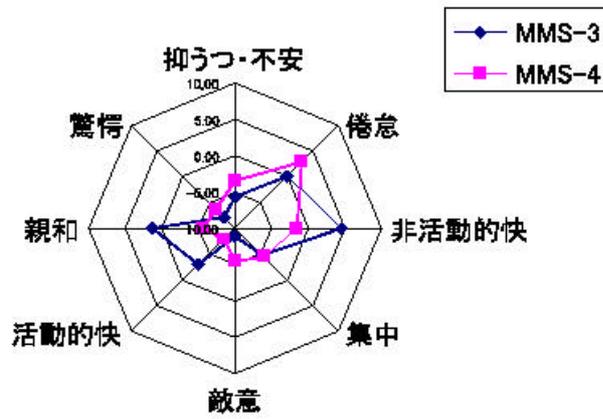


図 5.10: MMS の結果 (3 と 4 の差)

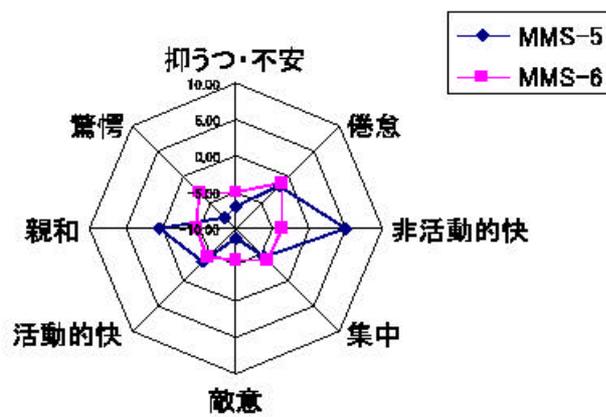


図 5.11: MMS の結果 (5 と 6 の差)

少しているが、これも顔表情に対する親しみが低かったことが原因だと考えられる。

驚愕 課題3の前後で大きくなっている。これは、見慣れない合成顔表情が提示され、表情を変化させたことによる驚きではないかと考えられる。

以上をまとめると、課題2の前後では、合成音声の朗読により違和感が感じられたと考えられる。また、課題3の前後では、見慣れない顔表情からの刺激と違和感が感じられたと考えられる。

5.4 結果の分析と考察

本節では、前節での個々のアンケート結果を踏まえて、実験全体としてどのような結論を導くことができるか、また、どのような検討課題が挙げられるかを述べる。

まず、コンピュータに対する態度の違いからアンケート結果を分析した後、実験全体の結果と考察を述べる。

5.4.1 コンピュータに対して操作不安を持つ人と持たない人に分類した結果の分析

2章で述べたように、本研究ではコンピュータに心理的障壁を持っている人にも親しみやすく、使いやすいと感じさせるインタフェースを目指すため、アンケートB-5のコンピュータに対する態度の結果より、「操作不安」が正の値をとる被験者(4、7、10、12、16)に着目し、これらの人がどのように感じているかを分析する。表5.15に被験者全体を「コンピュータ操作不安」が正の値をとる人と負の値をとる人、すなわち「操作不安」を感じている人と感じていない人の2群に分けた場合の、アンケートB-3の回答の平均、および2群間の差を示す。

この結果より、操作不安の正負でもっとも差が顕著な項目は、質問番号14の「アニメ的な顔画像(A)と本実験で用いたリアル顔画像(B)のどちらに信頼感を感じるか?(-3:アニメ ~ +3:リアル)」という質問に対する回答であり、操作不安を感じている被験者は、操作不安を感じていない被験者と比べて、本実験で用いたリアルな顔画像に信頼感を感じない傾向がある。また、同様に質問番号12の「(A)と(B)のどちらに親しみやすさを感じるか」、質問番号15の「(A)と(B)のどちらが好きか」についても、操作不安を感じていない被験者と比較して、リアルな顔画像には親しみやすさが感じられ

表 5.15: 操作不安により分類したアンケート B-3 の結果

質問番号	操作不安		差 (あり-なし)
	あり	なし	
1	0.8	1.4	-0.6
2	-1.6	-0.4	-1.2
3	-0.8	-0.5	-0.3
4	-1.6	-0.4	-1.2
5	-0.4	-0.3	-0.1
6	-2.8	-2.4	-0.4
7	-2.6	-1.1	-1.5
8	-2.6	-2.0	-0.6
9	-2.8	-1.9	-0.9
10	-2.8	-1.5	-1.3
11	-1.0	0.0	-1.0
12	-2.6	-1.5	-1.1
13	1.2	1.7	-0.5
14	-1.6	0.9	-2.5
15	-2.0	-0.9	-1.1

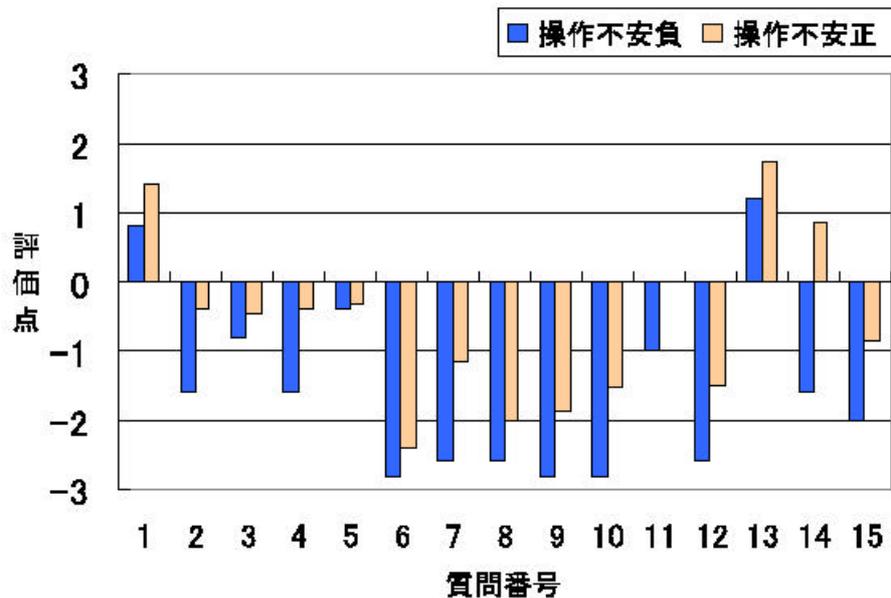


図 5.12: 操作不安により分類したアンケート B-3 の結果

ず、好まれていない。一方、ナレーションに対して、質問番号7の「ナレーションに対する共感」、および質問番号10の「ナレーションの好き、嫌い」も、操作不安が負の被験者は、ナレーションに共感せず、嫌いであることがわかる。また、質問番号4の「表情の変化の速さ」についても、不適切さを強く感じている。これより、コンピュータに対して操作不安を感じている被験者は、人工的な画像や音声にネガティブな印象を持っているようである。

5.4.2 実験全体の結果と考察

まず、合成音声の朗読、本システムの朗読、個々の表情による感情伝達について以下にまとめる。

合成音声の朗読 アンケートB-2の結果より、合成音声に対する評価は非常に低かった。

これは、現状での合成音声は、人間の朗読に比べて抑揚がなく、発音も不明瞭な部分があり、また、間が適切でなかったためと考えられる。しかし、音声合成の技術は開発途上であり、今後、人間の朗読に近い発音のできる合成音声エンジンの開発が期待される。

本システムの朗読 アンケート B-3 の結果より、ディスプレイ上に提示された顔表情に対する感想は、人により異なった。本システムのような朗読を近未来的だと捉え、今後の発展を期待する人がいる一方で、コンピュータが人間のような振る舞うことを懸念する人もいた。全体を見ると、「顔」そのものの不自然さが大きく現れているが、結果には被験者がこのようなシステムの開発を期待するか、しないかも関わっていると考えられる。

個々の表情による感情伝達 個々の表情伝達については、選択肢が 35 種類と多いにも関わらず、正答者数の平均値は 7.7(20 人中) と低くなかった。これより、顔や表情に対する人間の感受性の強さが予想される。また、表情は豊かであるが、不自然さやリアルさが、感情伝達を阻害した部分もあったと考えられる。

以上を踏まえて、5.1 節で述べた実験目的の各項目について、以下に考察する。

項目 1:感情伝達 アンケート B-3 質問番号 1 の結果 (A) から、合成顔表情と合成音声をを用いて感情を伝達できる可能性は高いと考えられる。しかし、この結果は「音声のみの朗読」と比較したものであることから、アンケート B-4 の結果も考慮する必要がある。アンケート B-4 の結果である正答者数 (B) から、人間が合成顔表情を見て、その表情が表現している感情を推定できる可能性は高いと考えられる。以上より、両者の結果から本システムで感情伝達できる可能性が示唆される。

項目 2:共感の誘起 アンケート B-3 質問番号 7 の結果 (C) からは、本研究で提案した合成音声と合成顔表情を用いた感情伝達方法では、被験者に共感してもらえなかったと考えられる。しかし、ここで (C) の質問文が「実験 3 のナレーションに共感できましたか?」となっており、「朗読内容に共感できたか」ということを適切に質問文中で表現できておらず、「このようなナレーション方法に対して、共感できるか」という意味にもとれるため、この結果のみからではどちらかわからない。ここで、「マッチ売りの少女」は「悲しい」話であるため、この話の朗読から誘起される感情は、MMS の項目 (表 5.2) の中には「抑うつ・不安」の上昇、および「活動的快」の下降であると推定される。表 5.14、図 5.11 を見ると、課題 3 後での「抑うつ・不安」は、わずかに増加している。ここで、課題 1 の前後でも「抑うつ・不安」の変化はほとんどなく、また、課題 3 では同じ物語を読んだり聞いたりするのが 3 回目となっていることより、被験者が内容を熟知していることから、特に「抑うつ・不安」の感情は誘起されなかったとも考えられる。一方、「活動的快」

は、課題1、2の後では下降しているが、課題3の後では下降の度合いが小さい。課題1、2の後の「活動的快」の下降は、被験者に「悲しい」感情を誘起させたと見ることもできるが、それよりも実験の疲れが大きいと考えられる。課題3後については、前述したように、物語を朗読しながら顔表情が提示されることで刺激が生じたと考えられる。

以上より、MMSの結果からは、被験者に「悲しい」感情を誘起させたとはいえない、すなわち共感の誘起は実現されなかったと考えられる。

項目3:親しみやすさ アンケートB-3 質問番号12の結果(D)から、本システムで用いた顔画像はアニメ的な顔画像と比較して親しみやすさを感じさせることが難しい、ということがわかった。前節でも述べたように、これは本システムで用いた顔画像が人間に非常に似ているリアルなものである反面、目が大きいなど一部分がデフォルメされているため、顔に対する感受性の強い人間にとっては中途半端で、違和感を感じさせる原因となっていると考えられる。ここで、(D)とアンケートB-3 質問番号15の「アニメ的な顔画像とリアルな顔画像のどちらが好きか(-3:アニメ ~ +3:リアル)」という質問の結果(E)との相関係数は0.6となっており、リアルな顔画像の好き嫌いとの相関がある。また、アンケートB-3 質問番号8の「ナレーションの人間らしさ」の結果(F)と質問番号9の「ナレーションの自然さ」の結果(G)の相関係数は0.73、(G)と質問番号10の「ナレーションが好きか、嫌いか」の結果(H)の相関係数は0.76となり、これらの評価点の平均が低いことから、人間らしさ、自然さ、好き嫌いが、親しみやすさに関連があると考えられる。

以上をまとめると、以下のような結果となる。

1. 本システムの合成音声と合成顔表情を用いて、伝えたい感情を被験者が正しく想像してくれる、すなわち感情を伝達できる可能性が示唆された。
 2. 本システムの合成音声と合成顔表情を用いて、共感の誘起は実現できなかった。
 3. 本システムの合成音声と合成顔表情を用いて、親しみやすさは実現できなかった。
- 2、3が実現できなかった理由として、以下のような点が考えられる。
1. 顔表情の合成、とくに唇の動きが発話と合っておらず不自然だった。
 2. 合成顔表情の顔自体が、被験者に親しみやすさを感じさせなかった。

3. 合成音声による朗読、とくに朗読方法や文章と文章の間が適切ではなかった。
4. 顔や表情、音声など、総合的に不自然さが残ったため、共感を誘起させるというより、むしろ不自然さの方が意識され、物語文の内容で表現される感情を被験者に共有してもらい、すなわち共感を与えることができなかった。

1. については、今後システムを改良し、より自然にみえる顔表情や唇の動きを合成することで改良が可能である。2. については、人間が親しみを感じるような顔画像がどのようなものであるかを調査し、その結果を反映した顔画像を作成することが必要となる。3. については、市販の音声合成システムの性能は日々向上しているため、人間の自然な発話を生成できる高性能な音声合成エンジンの出現が待たれる。4. については、1. ~ 3. の問題点が総合的に絡み合っており不自然さが感じられると予想されるので、問題点を1つずつ解決した後に、それらをうまく組み合わせることが必要となる。

5.5 まとめ

本章では、3章で述べた物語朗読システムを用いて、合成音声と合成顔表情を用いてコンピュータから人間に伝えたい感情を伝達できるかどうかを調べるために行った被験者実験について、その目的、結果、および考察を述べた。

実験の目的は、試作した物語朗読システムを使い、物語文の内容が表現する感情について(1)伝えたい感情を被験者に伝達できるか、すなわち感情伝達できるか、(2)伝えたい感情と同じ感情を被験者に感じてもらえるか、すなわち共感を誘起できるか、および(3)システムが提示する顔表情や音声に対して被験者が親近感を抱き、親しみを感じるか、すなわち親しみやすさを感じてもらえるかを調べることであり、(A)文書による黙読、(B)合成音声のみの朗読、および(C)システムを用いた朗読、の3つを比較することにより、本システムの有効性を調べた。評価方法として、アンケート、および被験者の感情状態の推移を測定するための多面的感情状態尺度(MMS)を用いた。実験結果より、本システムを用いて、伝えたい感情を人間に伝達できる可能性は示唆された。しかし、共感の誘起や親しみやすさを実現することはできなかった。この理由として、以下のことが考えられる。

2 発話とともに行われる合成顔表情の唇の動きが不自然だった

2 顔画像が適切でなかった

2 合成音声による朗読が聞きづらいものだった

以上の問題点を解決するためには、以下の点を検討する必要がある。

2 唇の動きを自然にする

2 顔画像を変更する

2 朗読を自然に見せるための全体の調整する

第 6 章 結論

本研究では、人間と機械との高次協調・高次機能発現をはかるために、人間と機械の接点となるインタフェースに人間らしい感情を組み込んだ新しいヒューマンインタフェース「アフェクティブインタフェース」の実現を目指し、その基礎研究としてコンピュータによる感情表現・伝達の可能性の検討を行うことを目標とした。これにあたって、人間のコンピュータへの親和性を高めるために、より親しみやすいインタフェースの実現も目指すものとした。

本研究では、そのようなインタフェースの具現化として擬人化インタフェースを対象にとりあげ、その感情表現・伝達のあり方を、使用者である「人間」を中心に捉え、感情の社会的、文化的要因を考察し、利用者層が共有する感情生起特徴を明らかにするという観点から検討を進めた。具体的には、音声合成と動的顔表情合成技術を用いた擬人化インタフェースによる物語の朗読・表現システムを構築し、これを用いてコンピュータによる感情表現・伝達手法のあり方を実験検討した。

本論文では、まず研究の背景として、現状の PC インタフェースの問題点を述べた後、それらの問題点を解決するために人間が親しみやすさを感じる要素を「感情」と捉えた。そのような「感情」の表現を用いたシステムとして、表情により感情を表現し音声対話が可能な擬人化インタフェースの実現を、アフェクティブインタフェースの基礎研究と位置付けた。そうした擬人化インタフェースが人間へ感情を伝達することの可能性を検討課題とし、物語文を入力すると自動的に感情を推定し、それを表現する物語朗読システムを作成することとした。

次に、物語朗読システムの全体構成とシステム作成に必要な手法の概要を述べた後、感情の社会的、文化的要因を展望した。また、本研究では言葉と感情との関係に着目した感情の分類方法を用いた。そして、感情が主観的なものであり各個人の属性により感じ方が異なることを考慮し、物語の朗読を聞く人の属性に着目した感情伝達方法を検討するために、(1) ことばから人間の感情を推定する方法、(2) 物語文朗読により感情を伝達する方法、および (3) 顔表情合成により感情を伝達する方法を順に検討した。まず、(1) では、感情を表現する単語と表現している感情を対応付けるテーブル(感情語テーブル)を作成し、そのテーブルを参照することにより、物語文から感情を表現する単語を抽出し、抽出した単語に感情を割り当てる方法を検討した。さらに、2つの被

験者実験(感情語選択実験、感情語分類実験)を行い、その結果に基づく感情語テーブルを作成した。

次いで、(2)では、物語文朗読により感情を伝達するために、感情を表現する音声パラメータとして「声の大きさ」と「読む速さ」を選択し、感情を表現する音声パラメータ値を決定した。そして、(3)では顔表情合成により感情を伝達するために、筋肉モデル法に基づいて16種類の表情筋を本研究における顔表情パラメータとして選択した。そして、各感情を表現する顔表情パラメータを決定するために、人間の感覚を反映できる方法を検討した。具体的には、遺伝的アルゴリズムに基づいたインタラクティブ進化計算を用いた実験システムを構築し、提示された顔表情を実験の被験者に評価してもらうことにより、各感情を表現する顔表情を決定した。以上の検討結果に基づいて、物語朗読方法として、(i)物語文から物語の内容が表現する感情を推定する「感情推定方法」、および(ii)合成音声で物語文を朗読しながらそれに合わせた顔表情を提示する「感情表出方法」を提案した。

そして、提案した感情推定方法、感情表出方法に基づいて、物語を朗読するシステムを試作した。このシステムは擬人化インタフェースが合成音声により物語を朗読するとともに、合成音声と合成顔表情により物語中の感情を表現するものである。

最後に、物語を朗読するシステムを用いた被験者実験を行った。実験では、(A)物語文を黙読する、(B)合成音声のみでの朗読を聞く、(C)システムを用いた朗読を聞く、の3通りを被験者に比較してもらい、(1)システムから人間へ伝えたい感情を伝達できるか、(2)システムが朗読する物語文の内容に対して人間に感情移入させられるか、すなわち共感を誘起させられるか、(3)システムが人間に親近感を持たず、すなわち親しみやすさを感じさせることができるか、を調べた。被験者は、20才前後の大学生・大学院生計20名であった。評価方法として、アンケートと感情状態の変化を調べる多面的感情状態尺度(MMS)を用いた。これらの結果を分析することにより、本システムを用いて、被験者に感情を伝達することが可能であるということが示唆された。一方、共感をもたらす、および親しみやすさを感じさせる、という点に関しては、実現が困難であった。その原因としては、顔表情がリアル過ぎること、表情変化が不自然であること、合成音声が不自然であることなどが挙げられた。この結果より、物語朗読システムの改良、表情表出方法の検討などが課題として残された。

以上、本研究では、アフェクティブインタフェースを実現するために表情・音声合成による感情推定・伝達手法を検討し、人間の感情が持つ社会的、文化的要因を考慮した感情推定・伝達手法を提案した。提案した手法に基づき物語を朗読するシステム

を作成し、作成したシステムを用いて被験者実験を行った。実験結果より、以下のことがわかった。

- 2 合成音声と合成顔表情を用いて、コンピュータから人間に伝えたい感情を伝達する可能性が示唆された。
- 2 合成音声と合成顔表情を用いて、コンピュータが人間に共感をもたらし、親しみやすさを感じさせるには、感情表出方法の再検討などが必要である。

また、今後の発展としては、以下の課題が考えられる。

- 2 人間が親しみやすさを感じる顔の作成方法の検討
- 2 人間が親しみやすさを感じる感情表現方法の検討
- 2 人間に共感を誘起させるための感情表現方法の検討

謝 辞

本研究を進めるにあたり、研究テーマの設定から、研究会での的確な助言、論文の執筆や内容の提案にいたるまで、柔軟な発想と細かなご指導を頂きました吉川榮和教授に深く感謝いたします。

種々の助言や論文の執筆と、具体的なご指導を頂きました下田宏助教授に深く感謝いたします。

技術的なご指導から種々の助言、論文執筆にいたるまで、いつも暖かなご指導を頂きました石井裕剛助手に深く感謝いたします。

システム作成中様々な技術的助言を頂き、ともに論文執筆に立ち向かい、かつ数々の心情の吐露を受け止めていただきました修士2回生の小牧大輔氏、大坂融弘氏に深く感謝いたします。

論文作成時の細かなご指導や日常の有益な助言をいただきました博士3回生の大林史明氏に深く感謝いたします。

被験者実験の準備やデータ解析を手伝っていただきました修士1回生の神月匡規氏、松崎剛士氏、学部4回生の新田和弘氏、小林隆氏に深く感謝いたします。

常に有益なご指導や助言をいただきました博士3回生の小澤尚久氏、博士2回生の山本倫也氏に深く感謝いたします。

被験者実験に参加していただきました修士2回生の笹井寿郎氏、米田賀一氏、修士1回生の社領一将氏、岡田芳信氏、学部4回生の遠藤啓介氏、鮫島良太氏、谷友美秘書、藤岡美紀秘書、吉川万里子秘書、石井令乃奈秘書、宮沢研究室の沖陽三氏、高橋とも氏、近藤寛子氏、早瀬賢一氏に感謝いたします。

研究室が異なるにもかかわらず、遠路かつ多忙の中、長時間の被験者実験に参加していただきました二宮啓氏、佐藤正晃氏に深く感謝いたします。

最後に、6年前の私の選択を許し、常に健康を気遣いながら応援してくれた家族に、心から感謝します。

参考文献

- [1] Rosalind W Picard : Affective Computing, The MIT Press(1997).
- [2] Reeves, Nass : The Media Equation, Cambridge University Press(1996).
- [3] A. Key : Computer software, Scientific American, Vol.251, No.3, pp.41-47(1984).
- [4] 竹林 洋一 : 音声自由対話システム TOSBURG II-ユーザ中心のマルチモーダルインタフェースの実現に向けて-, 電子情報通信学会論文誌, Vol.J77-D-II, No.8, pp.1417-1428(1994).
- [5] 長尾確 : マルチモーダルインタフェースとエージェント, 人工知能学会誌, Vol.11, No.1, pp.32-40(1996).
- [6] 土肥浩, 石塚満 : Face-to-Face 型擬人化エージェント・インタフェースの構築, 情報処理学会論文誌, Vol.40, No.2, pp.547-555(1999).
- [7] 長谷川修, 坂上勝彦, 速水悟 : 実世界視覚情報を対話的に学習・管理する人間型ソフトウェアロボット, 電子情報通信学会論文誌, Vol.J82-D-II, No.10, pp.1666-1674(1999).
- [8] 神谷俊之 : 3次元ウォークスルーとCG秘書を用いた電子図書館インタフェースの開発, 情報処理学会 IM 研資料, IM19-5(1995).
- [9] 牛田博英, 平山祐司, 中嶋宏 : 自律的行動決定モデルに基づくインタフェースエージェント, 電子情報通信学会論文誌, Vol.J82-D-II, No. 10, pp.1655-1555(1999).
- [10] J. Bates : The role of emotion in believable agents, Communications of the ACM, Vol.37, No.7, pp.122-125(1994).
- [11] B. Hayes-Roth and P.Doyle : Animate characters, Autonomous Agent and Multi-Agent Systems, Vol.1, No.2, pp.195-230(1998).

- [12] G. Bell, D.Ling et al. : Lifelike computer characters, Thepersona project at microsoft research, Software Agent(J. M. Bradshaw, ed.), Chapter 10, pp.191-222, AAAI Press/The MIT Press(1997).
- [13] J. Lester, et al. : Deictic believability: Coordinating gesture, locomotion and speech in lifelike pedagogical agents, Applied Artificial Intelligence, Vol.13, No.4-5, pp.383-414(1999).
- [14] J. Rickel and W. L. Johnson : Animated agents for procedural training in virtual reality: Perception, cognition, and motor control, Applied Artificial Intelligence, Vol.13, No.4-5, pp.343-382(1999).
- [15] 國弘 威, 動的表情認識による感情推定手法に関する基礎研究, 京都大学大学院エネルギー科学研究科修士論文 (1999).
- [16] 下田宏, 國弘威, 吉川榮和 : 動的顔画像からのリアルタイム表情認識システムの試作, ヒューマンインタフェース学会論文誌, Vol.1, No.2, pp.25-32, (1999).
- [17] 楊大昭 : バーチャル空間における顔表情の動的合成に関する研究, 京都大学大学院エネルギー科学研究科修士論文 (1999).
- [18] 下田宏, 楊大昭他 : 筋肉モデルによる表情合成システムと筋肉パラメータの決定, ヒューマンインタフェース学会研究報告集, Vol.2, No.3, pp.67-72(1999).
- [19] Ekman, P., Friesen, W. V. 著, 工藤力訳編 : 表情分析入門, 誠信書房 (1987).
- [20] Cornelius, R. 著, 斎藤勇監訳 : 感情の科学, 誠信書房 (1999).
- [21] Masaaki Yoshida, Reiko Kinase, Junko Kurokawa : Multi-Dimensional Scaling of Emotion, Japanese Psychological Research, Vol. 12, No. 2, pp.45-61(1970).
- [22] 松山義則, 浜治世, 川村安子, 三根浩 : 情動語の分析, 心理学研究, Vol. 49, No. 4, pp.229-232(1978).
- [23] 国立国語研究所 : 分類語彙表 (増補版) 電子化データ, 国立国語研究所 (1996).
- [24] 道江義禎, 室田理子 : 改訂 インテリアとカラーコーディネート, 山海堂, 第3章 (1996).

- [25] 松尾太加志：コミュニケーションの心理学, ナカニシヤ出版 (1999).
- [26] 福井康之：感情の心理学, 川島書店 (1990).
- [27] 北野宏明編：遺伝的アルゴリズム 4, 産業図書, 第 13 章 (2000).
- [28] 高木英行, 畝見達夫, 寺野隆雄：対話的進化計算の研究動向, 人工知能学会誌, Vol. 13, No. 5, pp.692-703(1998).
- [29] 伊庭斉志：遺伝的アルゴリズムの基礎-GA の謎を解く -, オーム社 (1994).
- [30] 石田良平, 村瀬治比古, 小山修平：パソコンで学ぶ遺伝的アルゴリズムの基礎と応用, 森北出版株式会社 (1997).
- [31] 萩原将文：ニューロ・ファジィ・遺伝的アルゴリズム, 産業図書 (1994).
- [32] Holland, J.H.: Adaptation in natural and artificial systems, University of Michigan Press(1975).
- [33] 松本裕治他：日本語形態素解析システム『茶筌』version 2.0 使用説明書 第二版 (1999), <http://chasen.aist-nara.ac.jp/chasen/distribution.html.ja>.
- [34] 株式会社 東芝：東芝音声システム V.4.0.
- [35] 鎌原雅彦, 宮下一博, 大野木裕明, 中澤潤 編著：心理学マニュアル 質問紙法, (株) 北大路書房, pp.152-153(1998).
- [36] 寺崎正治, 岸本陽一, 古賀愛人：多面的感情状態尺度の作成, 心理学研究, Vol.62, No.6, pp.350-356(1992).