2004－HI－110　(6)
2004／9／10

コミュニケーションロボットにおける ノンバーバル情報を用いた状況依存音声認識

岩瀬 佳代子[†‡]，神田 崇行[†]，石黒 浩[†*]，柳田 益造[‡]

あらまし　人間同士の対話において，音声表現や顔の表情などのノンバーバル情報は，相手に自分の感情を伝達する手段として重要であり，人間とロボットとの対話においても同様である．本研究では，ロボットがノンバーバル情報から相手の感情を認識することにより，現在の状況を限定して音声認識を行う手法を提案する．人間とロボットにおける簡単な対話実験から，文脈に依存しにくい感情「緊張」と，文脈に依存しやすい感情「喜び」などが表出しやすいことが示唆された．また，緊張の感情表出が，文脈に依存する感情表出を妨げ，状況の限定に影響を与える可能性があることを示唆した．さらに，音声と表情による感情認識システムを用いた評価実験を行い，これらシステムの有効性を示す．

Situated Speech Recognition based on Nonverbal Information for Communication Robots

Kayoko Iwase[†‡]，Takayuki Kanda[†]，Hiroshi Ishiguro[†*] and Masuzo Yanagida[‡]

Abstract　This paper describes speech recognition based on nonverbal information for communication robots. It will explain how a robot can extract human emotions from nonverbal information and limit the number of possible situations. First, the results of an experiment on interaction between a human and a robot, lead a conclusion that there are two types of emotions: one, including emotions (joy, anger, fear, etc.) which depend on the context, and the other, including emotions (strain, etc.) which do not depend on it. Also, discussed are the possibility that the presence of strain emotion prevents displaying context dependent emotions, and results obtained by using emotion recognition based on the prosodic features of voice and facial information.

† ATR 知能ロボティクス研究所，京都府
ATR Intelligent Robotics and Communication Laboratories, Kyoto Japan
‡ 同志社大学大学院　工学研究科，京都府
Graduate School of Engineering, Doshisha University, Kyoto Japan
* 大阪大学大学院　工学研究科，大阪府
Graduate School of Engineering, Osaka University, Osaka Japan

1. はじめに

近年, 人間とのコミュニケーションを可能としたロボットの研究開発が盛んになっている. そのようなロボットにおいて, 人間との自由な対話の実現を目的とする「音声対話」機能の充実が期待されている. また, このような機能を充実させるには, 人間同士の会話において用いられるようなノンバーバル情報を用いることが重要であると報告されている[1][2][3]. しかし, 現在のロボットは, バーバル情報を利用する音声対話は行っているが, ノンバーバル情報を利用した音声対話はあまり行われていない.

人間同士の音声対話では, 視覚・触覚などによる情報や音声に含まれる韻律的特徴などのノンバーバル情報を, 相手の感情を認識する重要な伝達情報として扱っている. これは, Mehrabian により, 相手にメッセージを伝える際の感情は, バーバル情報から 7%, 音声表現と顔の表情（ノンバーバル情報）から 93%が伝達されることとしても提唱されている[4].

また, 人間が表出する感情については, 心理学においてさまざまな研究がなされている[5][6]. 伊藤らの研究によると, 人間とロボットの対話において表出されやすい感情としては, 持続的感情としての「緊張」, 一時的感情としての「喜び」,「困惑」が報告されている[1].

本稿では, ATR 知能ロボティクス研究所で開発されたロボット Robovie [7]に, ノンバーバル情報を認識する機能を実装することを提案する. また, 本稿対話実験により, ロボットとの対話において表出されやすい感情は, 文脈に影響のある感情として「喜び」, あまり影響のない感情として「緊張」が表出されやすいこと, ならびに, 持続的感情である緊張を抑えることにより一時的感情が表出しやすくなることを示す.

この, ロボットにおける感情を認識する機能として, 音声の韻律情報から感情を認識するシステム[1]ならびに顔の表情から感情を認識するシステム[8]を用いる. まず緊張の有無を音声から, 次に喜びの感情の有無を表情から検出することによって現在の状況を認識し, 状況に応じた単語辞書を用いて音声認識を行うシステムを提案する.

2. ノンバーバル情報の認識と状況

2.1 対話における状況

対話において, 状況を考慮することは非常に重要である. 語用論において, 人間同士の会話の局部支配機構のひとつ, “ 隣接ペア ” という概念が存在する. これは, たとえば, 問い―返答, 挨拶―挨拶, 申し出―受容などのことを示し, 発話者 A（第一部分）の発話によって発話者 B（第二部分）の発話が, ある一定の範囲に限定されることを意味する. これを状況の適切性という[9].

このように, 対話において, ある問いかけに対する相手の返答は両者ともにある程度予測できる範囲の内容であると言える. すなわち, その対話において, 両者間にある状況が形成されており, 人間はその状況を認識し適切な内容の返答をしているのである.

2.2 ノンバーバル情報の重要性

状況を認識できる要素として, ノンバーバル情報が挙げられる. ノンバーバル情報とは, 人々が意思を伝える際の言語（バーバル情報）などの明示的な情報ではなく, 意味が表に現れない暗示的な情報である. 具体的には, 顔の表情, 言語情報以外の音声情報（以下, 言語情報以外の音声情報を「音声表情と」呼ぶ.）, 手や腕のジェスチャー, 視線などである.

このノンバーバルな情報は, バーバル情報とは異なり, 人間の行為や行動は感情から切り離せないという点で特に重要である. なぜなら, 人間は, 意識する／しないに関係なく, ノンバーバルな情報で物事を判断してしまうことが多いからである. また, それらは, 言語内容以上の意味を持つことがしばしばある.

Mehrabian は, メッセージ全体の印象を次のような公式で示し, ノンバーバル情報の与える影響が大きいことを示している[4].

メッセージ全体の印象＝
0.07（言語内容）＋0.38（音声）＋0.55(表情)

つまり, 顔の表情の印象が最大で, 次が音声の表情, 最後が言語内容ということである. これは, 顔表情と言語内容が矛盾する場合は, 顔表情によって伝えられる印象の方が優先し, 音声表情と言語内容が矛盾することがあれば, 音声表情が優先して全体の印象を決定するということになる.

またこの公式は, 言語内容より行動という伝達手段が, 他人に感情や態度を伝えることと密接な関連を持っている. すなわち, 前述した公式は, 左辺「メッセージ全体の印象」を「感情の統計」と置き換えることにより, すべての感情に当てはめることができる[4]. したがって, ノンバーバル情報, 特に音声表情と顔表情は, 対話の相手の感情を認識することができ, さらに現在の状況を限定することができると言える.

2.3 感情モデル

人間の感情は, 古くから心理学の分野において研究されている. また, 感情の表出に関しては, 顔の表情や姿勢, 音声, ジェスチャーなどの行動および自律反応として, ノンバーバルな情報を対象とした研究がされている. その中でも, 感情と表情に関する研究は非常に多くの知見を得ている. 例えば, Ekman らによる基本感情説, Russell らによる次元説が挙げられる[7].

Ekman によると, 感情には, だれもが普遍的に持っている 6 つの基本感情（喜び, 驚き, 怒り, 嫌悪, 恐れ, 悲しみ）が存在すると示している. また, 次元説は, 次元上の 1 つのベクトルによって感情を表現することが可能であるという考え方である. たとえば, Russell による「快－不快」,「覚醒－眠気」の 2 つを軸とした円環モデルとして, 図 1 に一部示すように感情をあるベクトルの向きとして表すことが提唱されている.

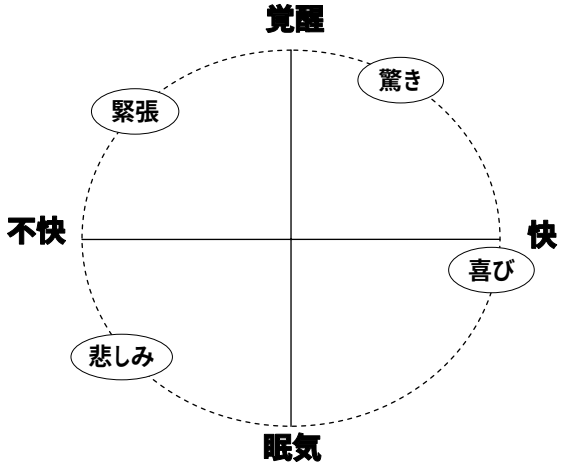

図 1 Russell による感情の円環モデル

伊藤らによると, 感情はその遷移速度によって一時的感情と持続的感情に分けられる [1]. 一時的感情は, 怒り, 喜びなどのように, 発話単位での変化が見られる感情であり, 持続的感情は, 緊張など個人の性格にも依存し, 発話単位では変化しにくいものである. したがって, 発話内容など文脈に依存しやすい感情は一時的感情であり, 反対に持続的感情は文脈には依存しにくいことが推測される. さらに, ロボットとの対話において多く見られる「緊張」の感情が強く表出している場合, 文脈に影響を与えやすい感情への, ベクトルの向きが変化は起こりにくくなることが推測される.

3. 対話実験

本章では, 人間とロボットとの対話実験を行い, 表出されやすい感情, さらに, 状況の限定に用いることができる感情についての調査を行う.

3.1 実験条件

人間とロボットの対話において, 人間が表出する感情を調査するため, 以下のような条件で対話実験を行った.

＜実験条件＞
ロボットは被験者に, いくつかの問いかけを同じ内容で繰り返し行う. それに対し被験者は,
（1）自由に回答する
（2）肯定的に回答する
（3）否定的に回答する
という条件を与えられる.

＜被験者＞
大学・大学院生　6 名　（男女　各 3 名）

＜対話例＞
R（ロボット）:「おはよう.」
P（被験者）:「お, おはようございます.」
R:「僕はロボビーだよ.」
P:「えっと, 私は, ○○です.」
R:「一緒に遊ぼうよ.」
P:「良いですよ.」
R:「じゃんけんしようよ.」
P:「よし, じゃんけんしましょう.」
R:「ロボビーかわいいでしょ？」
P:「うん, かわいいですね.」
R:「バイバイ.」
P:「はい, またね.」

また, ロボットの発話終了時から被験者の発話終了時までの画像・音声を，DV（Digital Video）に記録した．以降，この記録したデータを用い, ロボットとの対話における人間の感情について調査する．

3.2 評価尺度法による感情ラベリング

3.2.1 評価に用いる感情の種類

本研究では, 第 2 章で挙げた心理学研究における感情モデルのうち, Ekman の基本 6 感情に注目する．また，その他の感情として，伊藤ら[1]も注目しているように，人間がロボットと対面した場合などに見られる持続的感情, 緊張の感情にも注目する．また，緊張の感情表出が強く持続している場合，図 1 に示す Russell の円環モデルにおいて，不快，覚醒の間にベクトルが向いたままになってしまい, 他の感情の表出を妨げてしまうことが考えられる. そのため, ロボットが状況を認識するための文脈に影響を与える一時的感情を検出しにくくなると考えている．

次節において, 評価尺度法による感情のラベリングを行い, 持続的感情としての緊張の感情が表出している場合に, 一時的感情の表出が抑えられてしまう可能性があることを示す．

3.2.2 感情ラベリング

対話実験において記録した画像データについて，ロボットの問いかけ終了後 200～400msec の被験者の表情を静止画として切り出した．画像データ数は，各被験者 12 表情ずつ切り出し，全体で 72 フレームである．さらに，前述した 7 つの感情（喜び，驚き，怒り，嫌悪，恐れ，悲しみ，緊張）について，評価尺度法によるラベリングを行った. このラベリングは，対話実験の被験者ではない第三者 4 名（男 3 名，女 1 名）に対して行った．また，評価尺度は，それぞれの感情に対して，図 2 に示すような「とてもある」から「全くない」までの 6 段階の尺度を用いた．

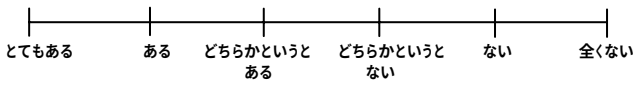

図 2 感情のラベリングに用いた評価尺度

さらに，それぞれの段階について，（とてもある）3 点，2 点，1 点，−1 点，−2 点，−3 点（全くない）の点数をつけ，ラベリングを行った 4 名の大学・大学院生 (a,b,c,d) の点数 $\left(f_{a,emo}, f_{b,emo}, f_{c,emo}, f_{d,emo}\right)$ の平均値 $\bar{f}_{emo}$ を，次式によって求めた．

$$\bar{f}_{emo} = \left(f_{a,emo} + f_{b,emo} + f_{c,emo} + f_{d,emo}\right)/4$$

$$\left(\begin{array}{l} emo \text{ は，怒り, 嫌悪, 恐れ, 喜び,} \\ \quad \text{悲しみ, 驚き, 緊張のいずれか} \end{array}\right)$$

また, その表情に対する感情は, 平均値 $\bar{f}_{emo}$ の最大値 $Max\left(\bar{f}_{emo}\right)$ をとるものとした．

3.3 ラベリング結果

表 1 に, 対話実験における感情ラベリングの結果を示す．左の項目には感情の種類を，右の項目は，3.2 節で行ったラベリングによって求めた $Max\left(\bar{f}_{emo}\right)$ について，各感情に対する全体の割合を示したものである．これによると，人間とロボットの対話において，「喜び」と「緊張」の感情が表出しやすいことが示されている．

表 1 感情ラベリング
（最大値 $Max\left(\bar{f}_{emo}\right)$ を示した感情）

感情	各感情が $Max\left(\bar{f}_{emo}\right)$ を示す割合
怒り	7.6%
嫌悪	18.9%
恐れ	0.8%
喜び	41.7%
悲しみ	6.1%
驚き	0.0%
緊張	25.0%

表 2 感情ラベリング
（返答内容への影響）

感情	肯定的返答	否定的返答
怒り	4.3%	15.0%
嫌悪	16.3%	25.0%
恐れ	0.0%	2.5%
喜び	52.3%	17.5%
悲しみ	8.7%	0.0%
驚き	0.0%	0.0%
緊張	18.5%	40.0%

次に, 返答の内容が感情の表出に与える影響を調べるため, 肯定的な返答をした場合と否定的な返答をした場合に分け, 表 2 にその結果を示す．ここで，3.1 節の実験条件において，肯定的な返答すると条件においた場合は肯定的返答, 否定的な返答すると条件においた場合は否定的返答，さらに，自由に返答すると条件においた場合は実際に返答の内容を聞き，肯定・否定に分類した. 自由に返答する条件の場合は, 返答内容を聞き，その分類を行った.

表 2 からは, 肯定的な返答をする場合は喜びの感情, 否定的な返答をする場合は嫌悪など喜び以外の感情が表出していることが分かる. したがって, 喜びの感情を示す場合は肯定的な返答をしている場合が多く, それ以外の否定的な感情を示す場合は否定的な返答をしている場合が多いと示唆される.

表 1，表 2 より，持続的感情である緊張の感情が, 比較的大きな割合を持つ感情として存在している．これは，前述したように，機械などに慣れていない人間がロボットと初めて対面した場合などに表出しやすい感情であると考えられる.

表 3 は，評価尺度法によって求めた緊張の平均値 $\bar{f}_{緊張}$ について，$\bar{f}_{緊張} > 0$ の場合と $\bar{f}_{緊張} \leq 0$ の場合に分け, さらに返答内容の肯定と否定によって場合分けした場合の集計である.

表 3　感情ラベリング
（緊張の有無への影響）

感情	$\bar{f}_{緊張} > 0$ (緊張あり)		$\bar{f}_{緊張} \leq 0$ (緊張なし)	
	肯定的返答	否定的返答	肯定的返答	否定的返答
怒り	0.0%	0.0%	6.3%	27.3%
嫌悪	2.9%	0.0%	21.9%	45.4%
恐れ	0.0%	7.1%	0.0%	0.0%
喜び	29.4%	7.1%	59.4%	27.3%
悲しみ	5.9%	0.0%	12.4%	0.0%
驚き	0.0%	0.0%	0.0%	0.0%
緊張	61.8%	85.8%	0.0%	0.0%

これより，$\bar{f}_{緊張} > 0$ の場合, ほとんどの場合で緊張が 7 感情の中での最大値 $Max\left(\bar{f}_{緊張}\right)$ となってしまう，すなわち，以下の式に示すように, その他の感情の平均値を上回らない結果となっている.

$$Max\left(\bar{f}_{緊張}\right) > Max\left(\bar{f}_{emo}\right),\ emo \neq 緊張$$

これは，3.2.1 で示したように，文脈に依存しにくい緊張の感情がある場合, それに依存する他の感情が表出しにくい, 状況を限定しにくいことを示している. また，$\bar{f}_{緊張} \leq 0$ の場合は, 表 2 において示唆されたように, 肯定的返答の場合は喜び，否定的返答の場合はむしろ，嫌悪のような否定的な感情が表出する確率が高くなっている.

4.　提案システム

4.1　状況依存音声認識システム

ここでは，第 2 章に挙げた概念に基づき，状況依存音声認識システムを提案する. 図 3 にその処理の流れを示す.

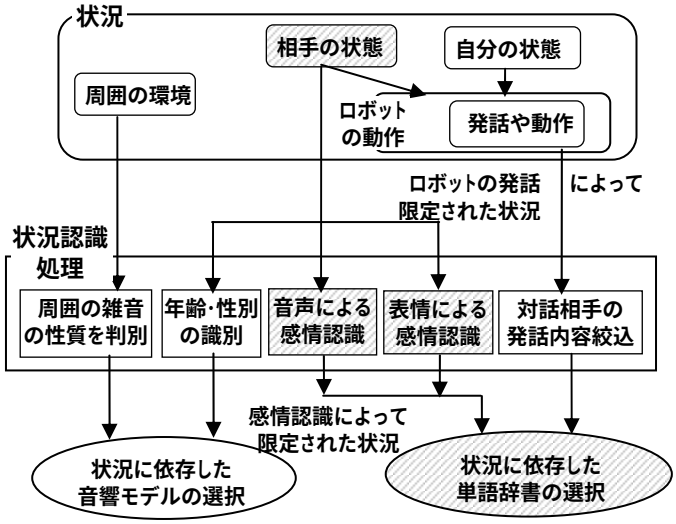

図 3　状況依存音声認識システム

状況依存音声認識システムは, ロボットの周りの状況を雑音の性質や発話内容, 対話の相手の感情として認識し, それに最適な音響モデルや単語辞書を用いて音声認識をするシステムである．すなわち，現在の状況において，相手が発話すると考えられる単語を絞り込み, 最も適当な単語辞書などを用いて音声認識を行うのである．既にある状況としては，相手・自分の状態，周囲の環境などが挙げられ，ロボットはさまざまなセンサーにより, これらの情報を取り入れることが可能である．さらに，ロボッ

トは, 認識した状況からさらに新しい状況生成行動（発話や動作）を行って自らが状況を作り出すことも可能である.

[10]において，図 3 のうち，自分自身（ロボット）の発話や動作を認識することにより，状況に適切な単語辞書を選択すると音声認識の性能が向上するという状況依存の可能性を示した．本稿では，図 3 の斜線部分の処理，対話の相手の状態から感情を認識することにより，音声認識で用いる単語辞書を限定するという，感情認識部の処理を提案する.

4.2 感情認識部

4.1 節で述べたように，相手の状態から状況を認識することによって，感情を認識する．ここで，第 3 章の感情のラベリング結果より，人間がロボットと対話する場合, 緊張の感情が他の感情の表出を妨げているということが示唆されている．また，発話内容の否定，肯定は，喜びの感情を検出することが重要であることも示唆された.

以上より，まず，人間の緊張の感情の有無を検出する必要がある．もし緊張があれば，それを軽減させる状況を新しく形成して緊張を軽減し, 喜びの感情を検出する. このようにして，人間の感情を認識することによって, 現在の状況の限定を行った後に, 限定された単語辞書を用いて音声認識を行う手法を提案する.

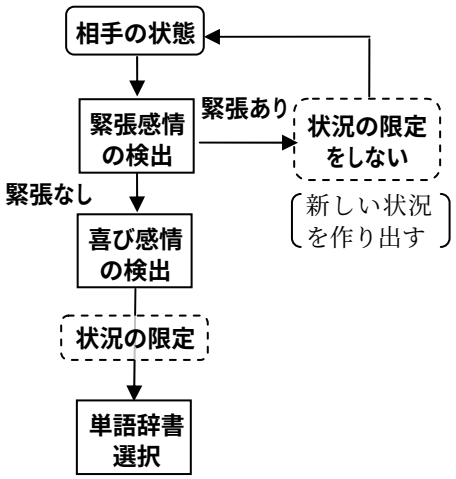

図 4 感情認識モデル

図 4 は，図 3 において，感情認識の部分の処理を表したものである．緊張の感情がない，または小さいと認識した場合, 喜びの感情を認識し, 状況を限定して音声認識を行うというモデルである．特に，肯定的な返答であるか否定的な返答であるかを認識し, 状況を限定するのである．また，緊張があると判断した場合は，現時点では状況の限定を行わず, ロボットから緊張をほぐすような発話などを行い, 人間の緊張の感情を小さくし, その他の感情が表出しやすいような状況を作り出す.

4.3 人間型ロボット Robovie への実装

本研究では, これらの感情認識システムを実際のロボットに実装し, 人間との音声対話の性能の向上を目指す．実装するロボットは，ATR 知能ロボティクス研究所で開発された人間型ロボット，Robovie である（図 5）.

図 5 Robovie

Robovie は聴覚，視覚，触覚，超音波センサーを持ち, 外界からのさまざまな情報を取り入れ，認識することができるロボットである．これらのセンサーを用い, ノンバーバル情報から感情を認識して状況の限定を行うことにより，人間との自由なコミュニケーションを可能とするロボットの実現を目指す.

特に，感情認識において，視覚センサーは人間の顔を見つけてその表情から，また，聴覚センサーからは話し方やニュアンスなどの韻律情報から，人間の感情（緊張と喜び）を判別する．判別に用いる感情認識システムは，次節に述べる 2 つのシステムを使用する.

4.4 感情認識システムの評価実験

4.4.1 音声による感情認識システム

音声による感情認識システムは, 伊藤らによって作成された．特徴量として，基本周波数，パワー，発話間間隔など 29 の特徴量を用い，

C5.0 または SVM を用いて,「緊張」と「喜び」の感情の有無を検出するシステムである. このシステムにおける感情判別率は，SVM を用いた場合の喜び感情の判別率が 74.1%，C5.0 を用いた場合の緊張の感情の判別率が 87.0%を示している.

第 3 章で示した対話実験で収録した音声データに対して, このシステムで感情判別を行った. 音声データは, 第 3 章で切り出した画像データにおける人間の発話を用いた. このシステムを用いた喜びの感情判別は，成功率 43.8%, 失敗率 24.5%であった．ここで，第 3 章で評価尺度法を用いてラベリングした感情とシステムの出力が, 一致した場合を成功, 一致しなかった場合を失敗としている. また, 緊張の感情判別に関しては，認識に用いる C5.0 のパラメータを調節したところ，成功率 81.0%，失敗率 19.0%を示した.

問題点として, 収録した音声データに大きな雑音が含まれているため, 感情の判別結果に影響が出ると考えられる. よって, ロボットに実装する際，再学習をする必要がある.

4.4.2 表情による感情認識システム

G. Littlewort らによって開発された表情認識システムは，FACS（Facial Action Coding System）を用い, Ekman の基本 6 感情と普通の 7 表情（怒り，嫌悪，恐れ，喜び，悲しみ，驚き，普通）を判別するシステムである.「喜び」の感情判別に関しては, 成功率 87.0%であるとされている.

第 3 章にて収録した画像データについて表情判別を行ったところ，成功率 77.8%，失敗率 23.3%であった. また, 成功率, 失敗率は, 4.4.1 に示したものと同様である.

5. まとめ

本研究では, ロボットに搭載する音声認識において, 視覚や聴覚のセンサーを用いて取り入れたノンバーバル情報を用いることの重要性を示した. 特に, 表情と音声表情について述べ, 実際に人間とロボットの対話実験を行い, 状況とそれらの表出の関係について調査を行った.

これによると, ロボットの問いかけに対して, 肯定的な返答をする場合は喜びの感情, 否定的な返答をする場合は喜び以外の感情が多く表出することが示された. したがって，喜びの感情が判別された場合の多くは, 人間が肯定的な返答をする状況であると限定することが可能である. また, 持続的感情である緊張の感情について調査したところ, ロボットとの対話における多くの場合でその感情の表出が見られ, 他の一時的感情が表出しにくいことが示唆された. したがって, ノンバーバル情報を音声認識に有効に利用するには, この緊張の感情を取り除く必要があると考えられる.

さらに，音声，および，表情による感情認識システムについて紹介し, 緊張の感情を音声による感情認識システム, 喜びの感情を主に表情による感情認識システムから検出できる可能性があることを示した.

今後, これらのシステムをロボットに実装し, ノンバーバル情報を用いた状況依存音声認識システムとして, 音声認識の性能について検討する.

謝辞

本研究は情報通信研究機構の研究委託「超高速知能ネットワーク社会に向けた新しいインタラクション・メディアの研究開発」により実施したものである.

また, 感情認識システムの利用に関してご支援頂いた，京都大学 河原達也教授，伊藤亮介氏，および，San Diego 大学 J. R. Movellan 氏に深く感謝いたします.

参考文献

[1] 伊藤亮介, 駒谷和範, 河原達也, 奥乃博：“ロボットとの音声対話におけるユーザの心的状況の分析”, 情報処理学会研究会資料, SLP-45-18, （2003.2）

[2] 佐藤賢太郎, 広瀬啓吉, 峰松信明：“生成過程モデルに基づくコーパスベース感情音声合成とその評価”, 情報処理学会研究会資料, SLP-50-8, （2004.2）

[3] 森山剛，斎藤英雄，小沢慎治：“音声表現における感情表現語と感情表現パラメータの対応付け”, 電子情報通信学会論文誌, D-II, Vol.J82-D-II, No.4, pp.703-711, （1999.4）

[4] A. Mehrabian 著，西田司 津田幸男 岡村輝人 山口常夫 訳：“非言語コミュニケーション”, 聖文社，1986 年

[5] 濱治世，鈴木直人，濱保久：“感情心理学への招待 ―感情･情緒へのアプローチ―”，サイエンス社，2001 年

[6] 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, 2000 年
[7] 神田崇行, 石黒浩, 小野哲雄, 今井倫太, 前田武志, 中津良平：“研究用プラットホームとしての日常活動型ロボット”Robovie”の開発”, 電子情報通信学会論文誌, D-I, Vol.J85-D-I, No.4, pp.380-389, （2002.4）
[8] G. Littlewort, M. S. Bartlett, I. Fasel, J. Chenu, T. Kanda, H. Ishiguro and J. R. Movellan： “Towards social robots: Automatical evaluation of human- robot interaction by face detection and expression classification”, International Conference on Advan- ces in Neural Information Processing Systems, Vol.16, MIT Press, （2003.12）
[9] S. C. Lebinson 著, 安井稔 奥田夏子 訳： “英語用語論”, 研究社出版, 1990 年
[10] 岩瀬佳代子, 伊藤亮介, 神田崇行, 河原達也, 石黒浩, 柳田益造：“日常活動型ロボットの状況依存音声認識”, 情報処理学会関西支部支部大会 講演論文集, B-04,（2003.10）