

## 言語・表情など外観情報と心的状態遷移に基づく人間感情の認知について

任 福継

徳島大学大学院ソシオテクノサイエンス研究部

感情計算と知能技術は感性と理性の対立から高いレベルでの統一体になっていると認識しているが、従来の音声・言語などの外観情報だけのモデルでは、人間の感情をモデル化し伝えることは到底不可能である。我々は人間の心的な特性に着目し、話者の感情測定モデル及びコンピュータの感情シミュレーションモデル、即ち、人間感情の認知及び機械感情の創生ができる、汎用的なエージェントを開発している。この講演では言語・表情・音などの外観情報と我々の提案した心的状態遷移ネットワークに基づく人間感情の認知について述べる。さらに、発表では本稿で記述しきれなかった最新成果・アプリケーション事例も紹介する。

### Recognizing Human Emotion based on appearance information and Mental State Transition Network

Fuji Ren

Faculty of Engineering, The University of Tokushima

Many people still seem to have strong resistance toward interacting with machines in many business fields such as terminal devices and medical care systems. We focus on human psychological characteristics to develop general-purpose agents that can recognize human emotion and create machine emotion. We comprehensively analyze brain waves, voice sounds and picture images that represent information included in emotion elements of phonation, facial expressions, and speech usage. We analyze and estimate many statistical data based on the latest achievements of brain science and psychology in order to derive transition networks for human psychological states. We establish a speaker word model for researching computer simulation of psychological change and emotional presentation, developing emotion interface, and establishing theoretic structure and realization method of emotion communication. In the talk, a new approach for recognizing human emotion based on appearance information and Mental State Transition Network will be described and some new results for the project will be given.

#### 1. はじめに

携帯電話やカーナビなどの情報端末から介護システムまで、幅広いビジネス分野において、「機械との会話」に人間はまだ強い抵抗感を感じているように思われる。一方、自然言語の意味認識は、人間とのコミュニケーションを円滑に行うためになくはならない要素であるが、必ずしも、十分ではない。さらに、人間の感情（言動、表情、感性）に含まれる情報を認識し抽出することは、機械が「人間と違和感なく」コミュニケーションし、「感情通信」を実現するために、必要不可欠な要素である。しかしながら、従来の音声・言語などの外観情報だけのモデルでは、人間の感情をモデル化し伝えることは到底不可能である。

我々は、人間がある感情状態にある場合に、外部刺激および内的思考の推移により感情エネルギーがある一定の閾値を越えた場合に、他の感情状態に遷移する、感情エネルギーに基づく「心的状態遷移ネットワーク」モデルを考案した。さらに同モデルを中心に人に優しい情報処理技術を目指して、感情のメカニズムを解明するための様々な研究・開発を行っている。具体的に言えば、心理実験により遷移確率モデルの導出、大規模な感情データベースの構築、顔表情及び動的な人物表情コーパスの開発を通じ、人に優しい情報処理技術の実現方法の確立を目指している。

外観情報としては、発声の感情要素、話者の顔の表情、話者の言葉使いに含まれる情報、及び脳波変化状態などが挙げられる。このような外観情報を感情エネルギーに換算して、心的状態遷移ネットワークに作用させれば、人間の感情を認知できると考える。

この講演では、まず我々の提案している人間感情の認知及び機械感情の創生モデルを紹介し、心的状態遷移ネットワークの構成原理と構築手法を述べる。次に心理実験を通じ導出した静止状態心的状態遷移ネットワークを報告する。後半では、言語情報即ち言葉遣いによる感情認識手法、音声即ち発声による話者感情認識手法、顔表情と身振り表現による感情認識アルゴリズムを述べる。最後に最新成果の照会及び今後の課題を述べる。

## 2. 人間感情の認知及び機械感情の創生モデル

人間感情の認知は、従来行動心理学の分野において主に研究されてきたが、工学的な手段が欠如していたため、期待できる進展が得られなかった。一方、伝統的な人工知能では、人間の精神と文化を情報検索や推論手段で展開しているが、コンピュータでシミュレーションが出来る状況にはない。コンピュータのハードとソフト両面での飛躍的な進歩、特に自然言語理解領域での躍進によって、人間感情を認知する、という「研究者の長年の夢」が実現可能になり、近年になって復活されつつある。

従来の研究開発は感性という表層的なものだけに力を入れ、根本的に人間の感情を認知することにはまだ達していない。例えば、フランスの研究機関の情報では、最近いくつかのベンチャー企業によって感情認識が出来たという報告があったが、どれも文章や音声認識からの感情語彙辞書による認識であり、リズム・音律からの認識までできていない。日本では、音声特性からリズムを読み取り、怒り・喜び・悲しみの感情認識を利用して、音声認識の認識辞書を分割し、感情に合わせて切り替える手法を取ってきたが、現在の技術では人間の感情を認識することができない。

それで、我々は言語・音声・表情などの外観情報のみならず、人間の心的状態遷移メカニズムにも着目して人間の感情認知を研究している。なお、数年以内で、人間の心的状態遷移メカニズムを完全に究明することは困難なため、現在は心理学的な実験により導出した心的状態遷移ネットワークを利用して人間の感情を推定するという方法論を採っている。

図1に提案する人間感情認知と機械感情創生ができる感情インターフェースの概念を示す。

外観情報として、言葉遣いに含まれる情報（本稿では感情エネルギーと呼ぶ）、発声に含まれる情報、身振りや顔表情に含まれる情報が考えられるが、その他に脳波と感情との関係を調査するための実験も行った。しかしながら、当面、脳波情報を工学的に実用可能な手段で獲得するのは困難なので、本稿では脳波情報には言及しない。それぞれの外観情報を感情エネルギーに転換し、感情エネルギー合成により外部刺激を生じる。外部刺激はさらに心的状態遷移ネットワーク（次の節で定義する）に作用して、次の感情状態を推定する。これによりその時点の人間の感情を認知する。

## 3. 心的状態遷移ネットワーク

現在、我々は言語・音声・顔情報を融合して人間の感情を測定するためのプロジェクトを立ち上げて、いろいろな実験を行っている。さらに、人間の生理学的な反応や精神的な反応のある特定の遷移ネットワークにより近似する考えを検討した。即ち、人間の脳のメカニズム自体を解明するのではなく、ブラックボックスにより推論しようというものである。

ここに人間の感情は幾つかの状態に位置し、情報処理過程に対応するいくつかの離散的状态の間を遷移すると仮定するが、ここでは、これを「心的状態」と呼ぶ。人間の心的状態は、ある条件で、一つの状態から別の状態へ遷移することができる。状態間の遷移は同じ確率ではないが、外界の要因を無視すると一定の期待値が存在する。大量のデータを分析し、人間の個性情報に基づいて、図2に示した心的状態遷移ネットワークモジュールを構築することができる。

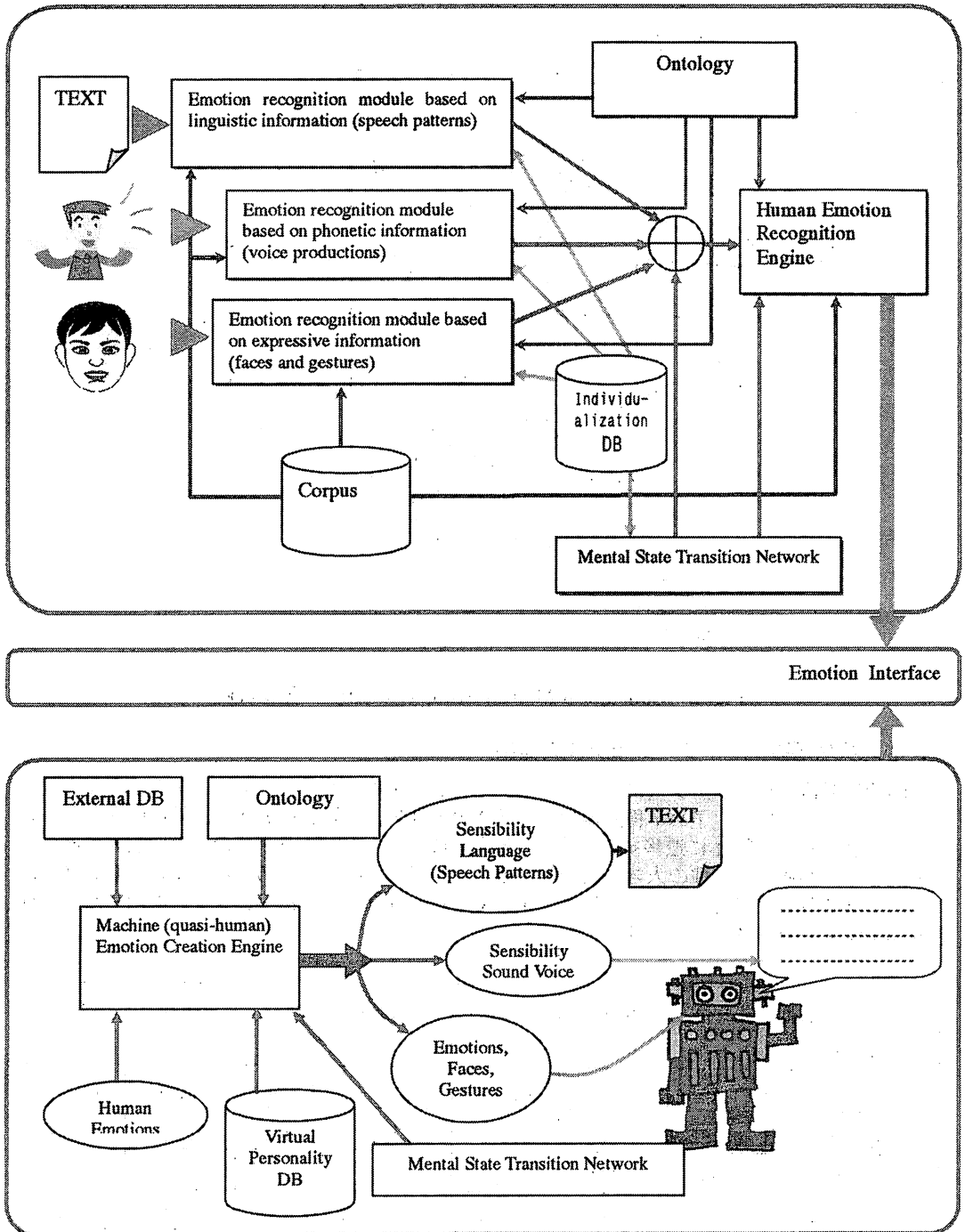


Fig. 1 Interface structure which can recognize human emotions and create artificial emotions.

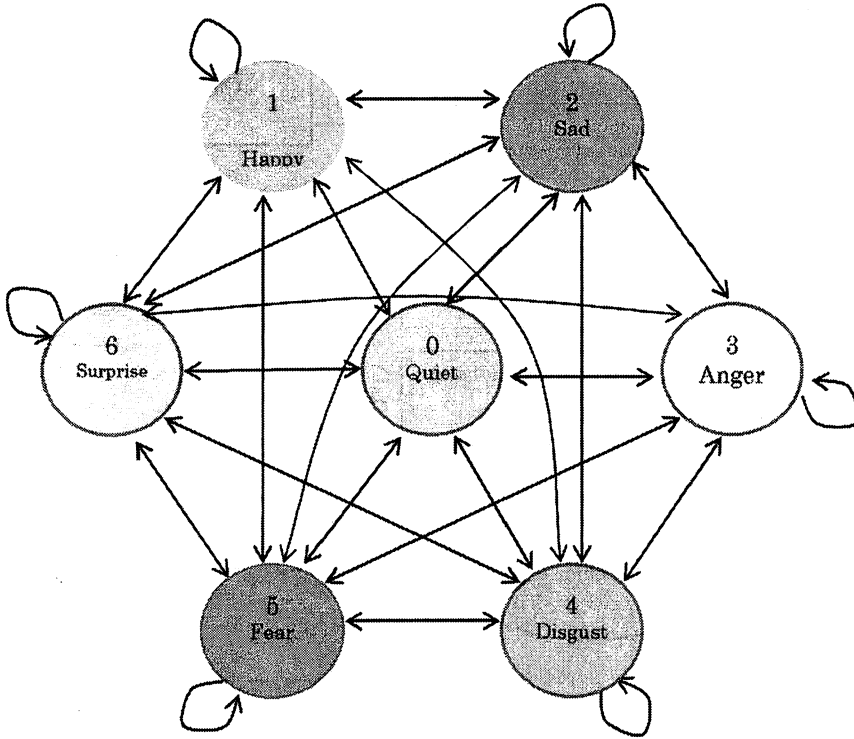


Fig. 2 Concept of Mental State Transition Network

#### 4. 心的状態遷移ネットワークモデルの導出

心的状態遷移ネットワーク (MSTN) モデルを導出するために、我々は以下のような心理的な実験を行った。まず、予め7種感情状態の遷移表  $S(7, 7)$  を作成する。外部刺激がない条件で、状態  $i$  から状態  $j$  に遷移する可能性を被験者により1から10までの数字で記入する。数字の大きさは遷移の可能性の大きさと一致する。そして、外部刺激がある場合に、同じ実験を行った。200名の被験者を通じて原始データを得たが、統計的な手法で、表1を導出した。これを外部刺激がない場合のMSTNとする。

current \ next	happy	quiet	sad	surprise	angry	fear	disgust
happy	0.421	0.213	0.084	0.190	0.056	0.050	0.047
quiet	0.362	0.509	0.296	0.264	0.262	0.244	0.252
sad	0.061	0.090	0.320	0.091	0.123	0.137	0.092
surprise	0.060	0.055	0.058	0.243	0.075	0.101	0.056
angry	0.027	0.039	0.108	0.086	0.293	0.096	0.164
fear	0.034	0.051	0.064	0.076	0.069	0.279	0.075
disgust	0.032	0.042	0.068	0.048	0.121	0.092	0.313

Table 1 Mental State Transition Network

## 5. 外観情報の獲得

次に、外観情報とする発声、言葉遣い、顔表情からの感情エネルギーを獲得する手法を紹介する。なお、詳細には関連文献に譲る。

### 5.1 発声情報に基づく感情推定

人間の感情、例えば怒りや悲しみ、喜び、笑い、興奮などは音声の抑揚やリズムなどに込められていると考えられる。コンピュータにより人間の感情を認知するには、音声情報の利用が重要である。従来、音声情報からその感情を知ろうとした研究としては、通常の音声認識での感情認識が盛んに行われてきた。一方でコンピュータ上での仮想人格をユーザーに合わせて自然に反応させることを目的に、最近では継続時間、発話速度、発声の休止の有無、振幅、基本周波数の分析での感情認識の手法の研究が主流になっている。また、ニューラルネットを利用し、感情分析や和音による感情分析の例もある。通常の音声認識での感情分析では、感情の抑揚による表現が認識できないため抑揚による感情の起伏や変化の認識は困難である。そこで、音韻からその感情特性を求めようとしたアプローチもある。しかし、ニューラルネットを利用した感情認識では、判断の軌跡が読めなく、再現性と正確さにかけてくるため、実際に感情を認識することは困難である。

話者の発声の感情要素の抽出について、ニューラルネットワークや隠れマルコフでのスペクトル学習ではなく、抑揚やその変化量から人間の感情を認識する。そして、任意音声試験及び連続会話からの感情遷移情報検出、音楽の感情分析試験と感情の伝播試験と感情を含んだ連続会話からの感情認識試験を行い、実際の人間の感情変化との比較において推定手法の適応性を検証し修正する。

### 5.2 言葉遣いによる感情推定

会話文の内容をスーパー関数によって理解し、言葉遣いにより、話者の感情を認識する。例えば、「先生が私に重傷と言っていました。」について、「恐れ+不安」という感情、「徳島大学工学部教員徳島太郎 氏の御母堂様におかれましては、平成16年X月Y日（水）に御逝去されましたので、謹んでお知らせいたします。告別式等（仏式）は下記により執り行われます。」について、「悲しみ」という感情、「さて、貴殿よりご応募頂きました戦略的情報通信研究開発推進制度の課題につきましては、審査結果により、採択されましたので、ご連絡致します。」について、「喜び」という感情を認識する。

言葉遣いによる感情認識の詳細をここに記述できないが、本研究で構築したシステムの概要について述べる。システムは大きく分けて3つの処理を行う。最初の処理は、入力された会話文の構文解析結果を解析し、格フレーム化する感情計算の前処理である。次の処理では、格フレーム化された会話文の格要素ごとに好感度を設定し感情計算を行う。最後の処理では、感情計算で得られた感情値と、文末の様相をもとにそれぞれの感情を計測する。図3は、プロトタイプシステムのインターフェース画面である。

### 5.3 顔表情と身振り表現から感情認識

顔は、目、鼻、口、耳などの諸器官を共通の構成要素として、ほぼ一定の構造をしているが、これら顔を構成している各器官の形状、大きさ、色、全体的な配置、といった特徴には個人差があり微妙に異なっている。人間が顔によって人物を特定したり、性別、年齢などの人物の属性、人物に対する印象、人種や民族を判定したりすることができるのは、このような顔が持つ静的な特徴の多様性によるものである。そして同時に顔は、動的な多様性も併せ持っている。時々刻々と変化する喜びや悲しみといった感情表現として無意識のうちに、あるいは相手に対して何らかの意図を伝達するために人間は表情を作る。これによって、顔を構成する諸器官の形状やその全体的な配置が変化し、真顔からの変位として様々な表情が表出する。このような内的な情動や意図的な感情を表出させた顔の表情を、人間は言葉によらないコミュニケーションの手段として有効に活用している。我々の研究では、顔表情から感情エネルギーを自動認識するために、顔画像を六つの基本感情に無表情を加えた七つのカテゴリーを用いて分類するという方法論を用いる。

顔表情の推定には、顔領域検出、特徴点及び特徴量の算出と顔表情の推定プロセスがある。表情の判定には FACS と最小距離識別法を用いる。まず FACS を用いて各表情に特有な AU の動作状況

を確認し、最も動作一致率の高い表情を第1候補とする。このとき、動作一致率があらかじめ設定しておいた閾値を超えていない場合や、第1候補と第2候補との差が少ない場合は最小距離識別法を用いて辞書との照合を行う。それでも判定がきわどく際どくなる場合、第1候補がFACS、最小距離識別法共に一致していればその表情、一致していない場合、第2候補との差が大きい方の表情と判定する。

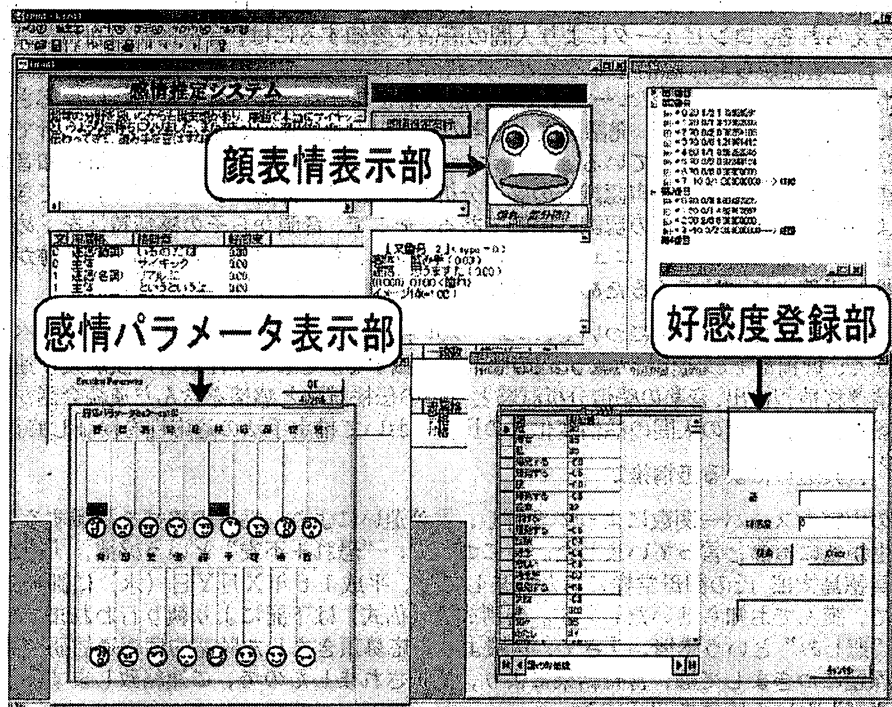


Fig. 4 User Interface of the prototype system

## 5. おわりに

人間の感情メカニズムのモデルとして心的状態遷移ネットワークを提案した。そして心理的な実験を実施し、外部刺激がない場合のモデルを導出した。

この招待講演で上記の心的状態遷移ネットワークを紹介して、外観情報とする発声・言葉遣い・顔表情などからの感情エネルギーを獲得するアルゴリズムも述べる。さらに、感情認識に関連するアプリケーション及び今後の研究報告にも言及する。

## 謝辞と注

本講演で言及したプロジェクトに参加している黒岩真吾助教授を始め徳島大学工学部知能情報工学科任研究室の各位に感謝します。本稿の一部は既に発表された文献から引用したものであり、参考文献の多くも省略された点について特記する。

## 参考文献

- [1] Fuji Ren, Dianbing Jian, Hua Xiang, Shingo Kuroiwa, Tetsuya Tanioka, Zhong Zhang, Chengqing Zong, Mental State Transition Network and Psychological Experiments, the International Conference on Artificial Intelligence and Soft Computing, pp.439-444, 2005
- [2] Peilin Jiang, Hua Xiang, Fuji Ren, Shingo Kuroiwa, An Advanced Mental State Transition Network and Psychological Experiments, IFIP International Conference on Embedded And Ubiquitous Computing, pp.1026-1035, 2005
- [3] Fuji Ren, Recognizing Human Emotion and Creating Machine Emotion, Invited Paper, Information, Vol.8, No.1, pp.7-20, 2005