



音情報を利用した感情・疲労度推定システムの構築： 感性情報処理システムへの拡張に向けて

メタデータ	言語: jpn 出版者: 室蘭工業大学SVBL 公開日: 2008-02-20 キーワード (Ja): キーワード (En): 作成者: 北越, 大輔, 原, 正一, 魚住, 超, 久保, 洋 メールアドレス: 所属:
URL	http://hdl.handle.net/10258/377

音情報を利用した感情・疲労度推定システムの構築 ー感性情報処理システムへの拡張に向けてー

北越大輔[†], 原正一^{††}, 魚住超^{††}, 久保洋^{††}

[†] 室蘭工業大学 SVBL, ^{††} 室蘭工業大学情報工学科

1. はじめに

音声は、人間対人間の意思疎通に重要な役割を担っているのは周知の通りである。このため、音声をコンピュータに認識させ、人間とのやりとりを行うインタフェースとして利用することを目的とした研究も盛んに行われている。一般的に、音声による会話の主要要素として、その意味内容自体を特定する言語情報に注目することが多いが、近年、会話音声の大きさや高さ、会話スピードなど、言語情報以外の要素に注目することによって、人間の感情を推定することを目的とする研究も増加しつつある[1]。ユーザがコンピュータを用いて何らかの処理を行う際、ユーザの音声から疲労度を含む現在の心身状態を認識できれば、より快適な作業の遂行が期待できる。現在まで、心身状態の推定には様々な手法が提案されているが、我々は確率的モデルの一つであるベイジアンネットワーク(Bayesian Network: BN)を利用した音声特徴量・心身状態のモデル構築を試みる。

本研究では、実際に会話する話者の音声から得られる特徴量を入力として、現在の話者の疲労度、および抱いている感情を推定する、感情・疲労度推定システムを提案する。具体的には、話者の会話音声から抽出する各種特徴量、およびアンケートの結果をサンプルデータとして、音声特徴量・心身状態モデルの構造を決定する。出力となる心身状態は、モデルの構造をもとに、入力となる音声特徴量を条件とした確率推論によって決定される。提案システムの導入によって、一般的な音声対話システムを、現在のユーザの心身状態を踏まえた上での出力を可能とする感性情報処理システムへと拡張することが可能となる。

以降、準備として、提案システムを構成する各要素について説明する。続いて、システムの概要について述べた後、適用例とその結果、考察を示す。最後にまとめと今後の展望について紹介する。

2. 準備

本節では、提案する感情・疲労度推定システムを構築するための理論的準備として、システムへの入力となる音声情報、出力となる感情・疲労度のモデル、およびシステム全体のモデル化に利用する BN について、順を追って紹介する。

2.1 音声情報

音声には、その意味内容を表す言語情報以外の非言語的な情報が含まれている。中でも、パラ言語情報と呼ばれる音声の韻律的特徴(例: アクセント、声の大小やその変化等)によって表現される情報は、発話者の感情や意図の表出に寄与していると考えられている。本研究では、上述のパラ言語情報をはじめとする音声の特徴量を抽出し、発話者の感情、疲労度の推定を図る。提案システムで利用する音声特徴量のうち、主要なものについて以下に簡単に示す。

- ピッチ … 声の高さを特徴づける。音声のアクセント、イントネーションや話者の感情等を伝える主要な物理的要因。
- 振幅 … 声の大きさを特徴付ける。ピッチと同様、感情伝達における主要な要因。
- パワースペクトル … 音声信号中に一つ以上含まれる、周波数成分それぞれにおける強さの分布を表したもの。人間の可聴域は 20Hz ~ 20KHz といわれているが、可聴域外の周波数成分が人間の感情に影響を与えているという報告もある[2]。
- ゆらぎ … 一般的には、パワースペクトルの回帰直線の傾きを表す(ただし横軸は対数軸をとる)。この値が-1の時が俗に言う“1/f ゆらぎ”である。同様の特徴を有するものとして、人体の心拍周期や α 波があり、自然界の多くの現象でも観測することができる。1/f ゆらぎを有した刺激を感じた時、人間

は快適さを感じる傾向があるといわれている。

- ホルマント周波数 … 音声波形のスペクトル上において、特定の周波数領域にエネルギーが集中して現れる領域のうち、最も振幅の大きい部分の周波数をいう。その周波数の低い順から第一ホルマント周波数、第二ホルマント周波数、…と呼び、母音の特定に利用される。

2.2 感情・疲労度のモデル

提案システムによって感情・疲労度の推定が可能となれば、これをもとに一般的な音声対話システムを感性情報処理システムへと拡張可能であることは 1. 節で述べた。ここで、本研究における感性情報とは、感情に代表されるような主観性、多義性、状況依存性等を有する情報を指す。例えば、同じ相手に同一の会話内容を話す場合においても、その際の自分および相手の肉体的、精神的状態や、会話が行われる環境によって、自分の話し方も相手の受け取り方も異なる。したがって、相手の音声から、その意味のみでなく、現在の心身状態も推定可能であれば、状況に応じた適切な出力(返答)を可能とするシステムを構築できると考えられる。我々は提案システムの出力の一つとして、感性情報の中でも基本的な要素である感情を取り扱う。心理学の分野では、感情を扱うために様々な感情モデルが提案されており、本研究ではそれらの一つである基本感情モデルを採用する。これは、本質的な感情である基本感情が複数結合、混合することによって異なる感情が生起するという考えをもとに構築されるものであり、提案システムにおけるモデルとしては Plutchik の基本感情モデル(図 1)を採用する。Plutchik のモデルでは、図に示される通り、対角線上にある感情が対を成しており、中央部は conflict と呼ばれる対となる 2 感情が拮抗している状態を表している。本研究では conflict を、対となる感情のうちのどちらでもないという意味で、“平静”の状態として考える。

また、提案システムにおけるもう一つの推定対象となる疲労度のモデルについては、ごく単純に、快調・通常・不調の 3 状態を有するモデルを想定する。

3. ベイジアンネット

ベイジアンネット(Bayesian Network :

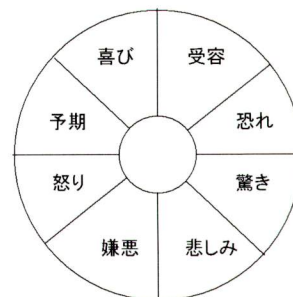


図 1 : Plutchik の 8 基本感情

BN)は、同時確率分布 $P(X_1, \dots, X_n)$ を用い、各確率変数間の依存関係を非循環性有向グラフとして表現した知識表現系ネットワークである。確率変数をノードとし、変数間に確率的依存関係が強いと判断される場合に方向付けられたリンクを設ける。依存関係を確率的相関と同一視した場合、同時分布 P は、以下のような条件付確率分布の積で表現される。

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i | \pi(X_i))$$

$\pi(X_i)$ は、確率変数 X_i と相関を持つ確率変数のうち、 X_i へのリンクを有するもの(親ノード)からなる同時確率変数 $(X_{e_1}, \dots, X_{e_h})$ とする。

本研究では、情報理論的妥当性がある MDL 基準を用いたモデル選択を採用する。MDL 基準は、

$$MDL(\hat{\theta}, d) = -\log P_{\hat{\theta}}^N(D) + \frac{d \log N}{2}$$

と定義され、この情報量が最小となるモデルを全探索によって選択する。ここで、パラメータ $\hat{\theta}$ は最尤法により得られたものである。

4. システム概要・適用例

本節では、提案する感情・疲労度推定システムの概要、システム動作、およびその性能評価のための適用例と結果の考察について述べる。

4.1 システムの概要と動作

まず、提案システムの枠組を示す(図 2)。感情・疲労度の推定に利用する BN の構造は、会話音声データと会話の際に得られたアンケートデータをもとに、MDL 基準を

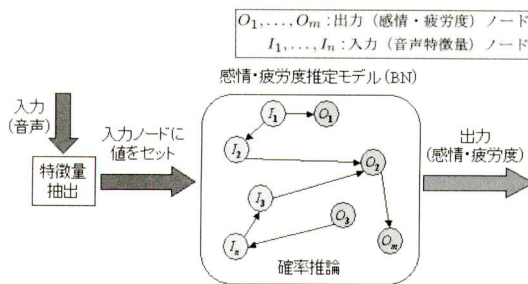


図 2：感情・疲労度推定システムの枠組

利用して予め選択されたものを用いる。また、提案システムにおける BN は、図 2 に示される通り、出力(感情・疲労度)および入力(音声特徴量)を表す 2 種類のノード群から構成される。出力ノードのそれぞれは、Plutchik の基本感情における対面する二つの感情、および疲労の度合いに対応する。一例を挙げると、“喜び-平静-悲しみ”といったようになり、確率変数の値としては、それぞれの感情に対応する整数値として $\{0,1,2\}$ を割り当てる。また、疲労度に対応する出力ノードについても、“快調-通常-不調”の 3 状態に対応する整数値を確率変数値として割り当てる。入力ノードはそれぞれ、音声から得られるピッチの最大値、平均値、レンジ、時間変化、振幅の最大値、平均値、音声の持続時間、第 1、第 2、第 3 ホルマント周波数、スペクトル中における最大音圧の周波数、およびゆらぎの値に対応する。確率変数の値としては、最大値や平均値、レンジ、周波数、およびゆらぎについては“低い(小さい)-標準-高い(大きい)”の 3 状態に対応する整数値を、ピッチの時間変化については、その回帰直線の傾きから“下がっている-一定である-上がっている”の 3 状態に対応する整数値を割り当てる。値を割り当てる際の閾値は、BN の構造選択の際に得られるサンプルデータをもとに決定する。

続いて、システムにおける感情・疲労度の推定動作の流れについて以下に述べる。システムへの入力音声となる。入力された音声の各種特徴量が抽出され、それらを BN の各種ノードに対応する確率変数値へと変換した後、BN に値をセットし、出力となる心身状態について確率推論を行う。具体的なシステムの出力決定法を以下に示す。

1. 要素数 $m-1$ の感情に関する出力候補集合 $OUTPUT = \{o_j^* | j = 1, \dots, m-1\}$ の要素 o_j^* を以下の式で選択する。

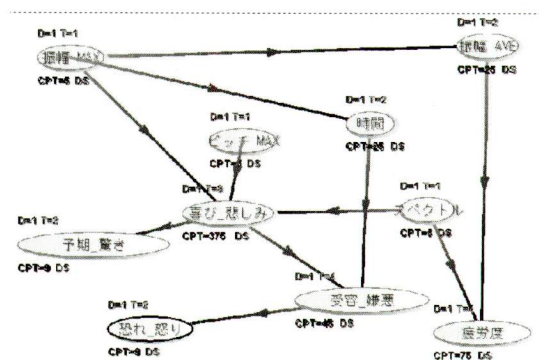


図 3：感情・疲労度推定モデル

$$o_j^* = \arg \max_{x \in M} P(O_j = x | I_1 = i_1, \dots, I_n = i_n) \\ (j = 1, \dots, m-1)$$

ただし M は各確率変数の変数値で $M = \{0,1,2\}$ 。

2. 以下の式を満たす一つの感情を選択する。

$$output = \arg \max_{o_k^* \in OUTPUT'} P(O_k = o_k^* | i_1, \dots, i_n)$$

ここで、

$$OUTPUT' = OUTPUT \setminus \{o_p^* | o_p^* = 1\}.$$

すなわち、出力候補の集合 $OUTPUT$ から、その値が平静(確率変数値=1)となっている要素を取り除き、平静以外の感情の中から条件付確率が最大のものを一つ出力する。ただし、 $OUTPUT'$ の要素が全て平静となった場合は平静を出力する。また、疲労度については上記 1. と同様の式によって最も条件付確率の値が大きくなる状態を出力とする。

4.2 適用例および結果・考察

提案システムの適用例と、その結果・考察について示す。準備として、19～26 歳の被験者のべ 51 名(男性 49 名、女性 2 名)から会話音声を収録した後、アンケートに記入してもらい、音声特徴量とアンケートデータの組からなる合計 1250 個のサンプルデータをもとに、感情・疲労度推定モデルとなる BN の構造を決定した。会話時間は一定の目安にしたがって実験の実施者側が設定する。また、話者の会話音声は、会話後にとったアンケートの結果をもとに切断し、切断されたそれぞれの音声データから特徴量を抽出した。上記の操作により構造選択された感情・疲労度推定モデルを図 3 に示

表 1 : 推定結果

心身状態	テスト数	結果: 推定率	
		結果1	結果2
喜び-悲しみ	20	40%	50%
恐れ-怒り	10	0	0
予期-驚き	23	0	0
受容-嫌悪	46	17%	37%
平静	10	40%	30%
感情全体	109	18%	28%
疲労度	95	41%	41%

す。構造選択の結果、他のノードと独立（矢印なし）となった音声特微量ノードについては表示していない。

構築されたモデルを用いたシステムにおいて、新たにユーザ 2 名の会話音声を収録し、それらから抽出した音声特微量を入力とした場合のシステムの出力（感情、疲労度）が、ユーザの回答したアンケート結果とどの程度合致しているかを調査することによって、提案システムの性能について評価、考察する。結果を表 1 に示す。表中のテスト数とは、会話全体においてユーザが各感情対中の“平静”以外のいずれかの感情を何度抱いたかをカウントしたものであり、音声データ数と等しい。また、結果の値は、音声データをシステムの入力とした場合、その出力がアンケート結果と合致した割合を示している。結果 1 は、先述したシステム動作と同様の方法で心身状態を出力した結果との比較である。実際のシステム出力としては、感情として平静が出力される割合が非常に高い。このような結果となった主な原因の一つとして、感情・疲労度推定モデルを構築するためのサンプルデータの収集方法に不適切な部分があったと考えられ、様々な感情が含まれた十分な量のデータを収集することによって、モデル自体がより適切なものとなり、推定率が向上する可能性が高い。結果 2 は、結果 1 を受けて、出力が平静であった場合、BN の推論結果における条件付確率の値が 2 番目に大きかった感情に注目し、「出力が平静となる確率」と「出力が第二候補の感情となる確率」の差が定数 α （本研究では 0.2）以内であれば、第二候補を出力として推定率を計算したものである。表に示されるとおり、結果 1 と比較して平静の推定率が減少している反面、喜び-悲しみ、および受容-嫌悪の感情についての推定率は向上している。人間同士が会話をするような状況においては、自分の感

表 2 : 該当データ数

心身状態	該当データ数
喜び-悲しみ	431
恐れ-怒り	137
予期-驚き	179
受容-嫌悪	378
平静	328
合計	1453

情を表出するための手段となる言語、音声、表情のうち、感情の 55%は表情によって表現され、ついで音声で 38%、さらにその音声の意味内容である言語が残り 7%の割合を占めるという研究報告がなされている [3][4]。このことから、他の研究成果との更なる比較検討の必要が残されているとはいえ、結果 2 において推定可能であった二つの感情対に関しては、感情認識の手がかり全体のうちおよそ 4 割程度である音声のみによって推定した結果としては妥当な成果が得られているといえる。加えて、従来研究と異なり、本研究における感情・疲労度推定モデルの構造決定には複数話者の一般的な会話音声についての特微量データを利用しているが、喜び-悲しみの感情対については、他の感情対と比較して高い推定率を得られている。このため、喜び-悲しみの感情が表出した際の音声には、話者の個人差に影響されない、共通の特徴があるものと推測される。なお、疲労度については結果 1、2 とともに同じ推定法を用いているため率に変化はなく、恐れ-怒り、および予期-驚きの感情についてはいずれの推定法においても正しく推定を行うことができなかった。結果 1 における考察と同様、収集したサンプルデータの数、およびその内容が推定性能に依存する可能性は非常に高い。また、特に恐れ-怒り、予期-驚きの 2 感情対における推定率が 0 であったことについては、これらが一般的な友人や知人同士の会話において表出しにくい感情であったり、今回のデータ収集の際に十分な量のサンプル数が得られなかったことが原因として考えられる。ここで、BN 構築に利用したサンプルデータにおいて、それぞれの感情対に対応するデータがいくつ含まれていたのかを表に示す（表 2）。表中のデータ総数がサンプルデータ数と一致していないのは、音声データに対応する感情をアンケート結果として回答してもらう際、複数の感情にチェック

をつけることを許したためである。表に示されているとおり、推定率が 0 となっている二つの感情対についての該当データ数は他の感情対に対応するデータ数のほぼ半分しかなく、推定モデルとしての BN の構造が適切に選択できなかったため、推定が不能になったと考えられる。したがって、今後推定率の更なる向上を目指すためには、適切な量と内容のサンプルデータを収集可能な環境の整備に加え、会話に特化した新たな感情モデルの提案も必要だと考えられる。

5. まとめ・今後の展望

本研究では、話者の会話音声から得られる特徴量を入力とし、BN による確率推論によって話者の感情や疲労度を推定するシステムを構築した。実際に得られた話者のサンプルデータ、およびアンケート結果をもとに構築した BN を感情・疲労度推定モデルとして利用した。システムの適用例として、話者の音声から話者自身の抱えている感情や発話時の疲労度を推定する実験を行い、その性能評価を行った。

実験の結果、提案システムでは話者を限定せず、複数の感情候補の中からの推定を可能とした。また、喜び・悲しみの感情については、話者の個人差に影響されない特定の音声特徴量が存在することを示唆する結果を得た。その反面、受容・嫌悪の感情における受容の推定率等、従来の研究においても十分な議論が行われていない部分も多く残されており、今後の課題として考慮すべき問題となる。また、本研究における BN 構築に用いたサンプルデータは、知り合い同士の一般的な会話を採取したため、怒りや恐れ、嫌悪といった感情と対応付けられる音声データの数が十分でなかったと推測される。加えて、本研究では過去に多くの研究で利用されてきた Plutchik の 8 基本感情をモデルとして利用したが、一般的な会話において表出する感情を表現するモデルとして適切であったとは言い難い。したがって、より会話に適した感情モデルの導入、または新たな感情モデルの提案も必要となる。

本研究と並行して、我々は会話音声における“間”、および話者の会話における話し始め、話し終わりに関する音声特徴量と感情の関係に注目した研究[5]や、現在までの研究例が少ない、効果音が人間の感情に与える影響についての研究[6]を行ってきた。これら全ての研究に共通する課題として、

BN の構造選択に用いるサンプルデータの収集法が挙げられる。会話テーマを予め決める、多様な特徴量を有する効果音をピックアップして聞かせる等の方法を取ることによって、十分な量の多様なサンプルデータを収集し、感情や疲労度に関するより詳細なモデルを構築することが求められる。最終的には、これらの研究成果を統合し、さらに顔画像データや脈拍、血圧等の生体情報も考慮することによって、推定率を向上させ、広範な音声情報に関する人間の感情推定が可能なシステムの構築を目指していく。

参考文献

- [1] 金城良治, 中村かおり, 長山格, 音声情報における感情表現の分析, 電気学会・電子情報通信学会合同講演会論文集, 2000.
- [2] 人体と音のコミュニケーションに関する調査研究報告書, 財団法人サウンド技術振興財団, 1993
(<http://www.sound-zaidan.com/04rs-mokuji.html>).
- [3] A. マレービアン, 西田司 (ほか) 共訳, 非言語コミュニケーション, 聖文社, 1986.
- [4] 斎藤康之, appearance ベースと model ベースによる眼鏡顔画像からの眼鏡なし顔画像の推定, 北陸先端科学技術大学院大学博士論文, 2002.
- [5] 室野敬太, 会話音声を用いた感性情報システムに関する基礎研究, 平成 15 年度室蘭工業大学卒業論文, 2004.
- [6] 八戸駿, 効果音が感情に与える影響に関する基礎研究, 平成 15 年度室蘭工業大学卒業論文, 2004.