室蘭工業大学
学術資源アーカイブ
Muroran Institute of Technology Academic Resources Archive

音声情報からの知覚感性モデル化に関する研究

メタデータ	言語: Japanese
	出版者: 室蘭工業大学SVBL
	公開日: 2007-12-17
	キーワード (Ja):
	キーワード (En):
	作成者: 八戸, 駿, 李, 林甫, 鈴木, 育男, 久保, 洋
	メールアドレス:
	所属:
URL	http://hdl.handle.net/10258/311

音声情報からの知覚感性モデル化に関する研究

八戸駿（M2）† 李林甫‡ 鈴木育男‡ 久保洋†
†室蘭工業大学情報工学科，‡室蘭工業大学SVBL

1.はじめに

他人とのコミュニケーション（情報伝達）の有効な手段のひとつとして，音声情報が考えられる．音声は，話し手の意図や意思を聞き手に与えるのと同時に，「楽しさ」や「悲しさ」といった印象も与える．

従来の音声処理の研究では，音声の発話内容や意味の理解などが主であった．近年，人間とロボットとのコミュニケーションを円滑に行う手段として，音声情報に含む非言語的な情報が注目され始めており，その識別処理は非常に重要な課題となってきている．このような状況の中，話し手の音声情報から感性情報の認識や推定に関する研究が多くなっている．

これまで，我々は，ベイジアンネットワークを利用してユーザの音声から疲労度を含む心身状態を推定することを提案し，研究を行ってきた[1]．一方で，話し手の音声情報が，聞き手にどのような感情を与えるかについて研究する必要性もわかってきた．音声情報が，聞き手に与える感性の研究を発展させていくことによって，よりフレンドリーな man-machine コンミュニケーションの手段となることが期待される．

本研究では，音声を受ける際の聞き手側の感性に注目し，音声の特徴量と聞き手の心理推定との関連性をベイジアンネットワークとニューラネットワークのそれぞれの手法によってモデル化し，このモデルを利用した感性の推定結果について報告する．

2. 感性と音声

2.1. 感性

本研究には、人間の発する音声から感性を捉えようとした場合，「話し手の感性」と「聞き手の感性」に分類してモデル化する必要があると考えられる．この人間が発揮できる２つの感性の違いについて，加藤らによって「創出感性」と「知覚感性」とに分類されている[2]．

創出感性とは，自分がイメージした様々なメッセージ等を，他人の五感を通じて伝えようとするものである．一般には「表現力」等と呼ばれ，自分の意思を表現するための感性である．

一方，知覚感性とは，五感を通じて得た物理的な刺激を脳で認識して理解するための感性である．一般的に言われている感性は知覚感性のことであり，「感受性」等とも呼ばれている．

2.2. 音声

人が音声を利用して相手に伝達する情報としては，藤崎らによって「言語情報」，「パラ言語情報」，「非言語情報」の３つに分類している[3]．

この中で，「パラ言語情報」は感情的情報であり，話し手が意識的に制御して発声することができ，音声の韻律的特長（アクセント，声の大小やその変化）によって表現されると考えられている．

この音声情報は，音声波形に含まれるため，波形の特徴から様々な情報を得ることができると考えられる．代表的な特徴量としては「振幅」や「スペクトル」があげられるが，さらに音の高さを表すピッチ（基本周波数）やスペクトル上でのエネルギーの集中を示すフォルマント周波数等があげられる．

本研究では，感情を表現する上で音の高さとその変化が最も重要と考え，ピッチの情報を中心として特徴量を抽出した．

3. 推論技術

3.1. ベイジアンネット（Beyesian Network）

ベイジアンネットは，同時確率分布を用い，各確率変数間の依存関係を非循環有向グラフで表現した知識表現系ネットワークである[4]．このとき，各ノードに対応するデータがあれば，ネットワークを学習によって構築することが可能である．また学習時に情報量基準を用いることで情報理論的妥当性の検証を行うことができる．

本研究では，音声の特徴量，及び話しての印象をそれぞれにノードとして割り当て，特徴量と印象の確率的依存関係をネットワーク構造として表現する．学習については，音声サンプルの特徴量と，それに対して人間が評価した印象（感性情報サンプル）からネットワークを学習する．このとき，ネットワークはMDL基準を採用し，この値が最小となるようなモデルを全探索によって選択する．

3.2. ニューラルネット（Neural Network）

ニューラルネットは，人間の脳の働きをコンピュータで応用し，ニューロンとその結合をネットワークとして情報処理を行う技術である[5]．ニューラルネットにはパーセプトロンと呼ばれる基本的なネットワークから，それを層状につなぎ合わせた多層パーセプトロン等のネットワークがある．

また，複雑化するネットワーク構造を学習するために誤差逆伝播法等の学習手法が確立している．

本研究では，3 層パーセプトロンを用い，音声サンプルからの特徴量を入力，感性情報サンプルを教師値として誤差逆伝播法によって学習し，音声特徴量から感性情報を推論するモデルを構築する．

4. 実験

4.1. 方法

サンプル音声と感性情報サンプルを採取し，音声サンプルからは特徴的な値を抽出する．その特徴量と感性情報サンプルから BN と NN のそれぞれにおいてモデルの構築を行い，最後に，新たな音声に対して人とモデルが評価した印象を照らし合わせて正答率を求めて評価する．

4.2. 手順

音声サンプル

第 1 印象となり易い「こんにちは」，話し手が様々な意味や感情を込め易い「ああ」，特に言語情報は特定しない「その他の声」を採取する．

感性情報サンプル

声の印象として「性別・高さ・太さ・速さ」，判断しやすい印象として「快さ・明るさ・感情的かどうか・自然さ・疲労度」，最後に感情の印象として「楽-悲・怒-恐・驚-期待・受容-嫌悪」を採取する．

音声サンプルの特徴量抽出

音声サンプルからは，音声分析ソフト Praat[6]を利用して，音声波形，ピッチ，フォルマントから特徴量を抽出する．

モデルの作成

収集した音声サンプルの特徴量と感性サンプルからベイジアンネット(BayoNet[7])とニューラルネットの各ネットワークの構造を学習することでモデルを作成する．

モデルの検証

作成したモデルにより，推論した感性情報を学習では利用したものとは別の感性サンプルと比較し，検証する．

5. おわりに

5.1. 結果・考察

音声サンプルは男女計10名から約1000サンプル採取し，感性情報サンプルを被験者（男 1，女 1）の計 2 名から採取した．音声情報サンプルからは 39 個の特徴量を抽出し，13 個の感性情報サンプルと合わせて学習サンプルを作成し，BN，NN においてモデルを構築した．

男性被験者の NN，BN モデルを利用した感性推定の検証結果の正答率は，図 1，図 2 のようであり，概ね同程度の正答率であることがわかる．しかし，この結果の場合，NN では「ああ」や「こんにちは」のような特定された音声の正答率が高く，BN では不特定音声の正答率が高い結果となっている．また，図では男性のみの正答率を掲載したが，女性においては男性ほど大きな正答率は得られなかった．

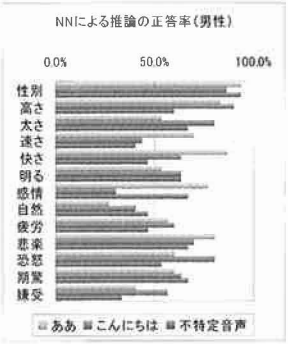

図 1 NN モデル推定結果

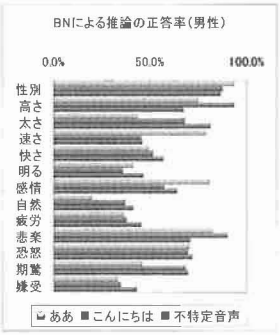

図 2 BN モデル推定結果

5.2. まとめ

本研究では，特定の聞き手の知覚感性についてモデル化を行ってきた．その中で，ある特定の状況ではモデル化は有効であり，ある程度の正答率を実現できることがわかった．しかし，人の感性は曖昧で変化があり，人によっても異なることも改めてわかった．

本研究では，知覚感性をモデル化する際に人間の心理に対しての知識を持っていない者を被験者としたが，人間の心理にしっかりとした概念を持つ被験者を採用することで優れた知覚感性モデルを構築することが可能であると考えられる．

今後の展望としては，まず特徴量の見直しによる正答率の向上が考えられる．また，今回は特定の聞き手についてモデル化したが，対象をグループ化することでそのグループ内での一般的な感性をモデル化できると思われる．さらに，言語情報・視覚情報との連携によるマルティメディア的な感性のモデル化が考えられる．しかし，これらの場合は曖昧な変化に対してそれぞれモデル化を行わなければならないため，聞き手の変化に対してモデルが変化させる等，新たな手法を提案する必要があると言える．

参考文献

[1] 原正一．声情報を利用した感性情報処理システムの構築に関する研究．2004
[2] 加藤俊一．多感覚情報のデータベース化と情報サービスへの応用．2003
[3] 藤崎博也．音声の音調的特徴のモデル化とその応用．2005
[4] 本村陽一．確率ネットワークと知識情報処理への応用．http://staff.aist.go.jp/y.motomura/DS/
[5] 中川聖一．パターン情報処理．1999
[6] Paul，David．Praat.
[7] http://www.fon.hum.uva.nl/praat/
[8] 数理情報システム．BayoNet. http://www.msi.co.jp/BAYONET/