2—2—7

フェイスマークを用いた 自然発話音声における感性情報の分析*

◎醫野和博 (奈良先端大 情報)　△柏岡秀紀 (奈良先端大/ATR)
ニック キャンベル (奈良先端大/ATR/CREST)

1. はじめに

近年，人とコンピュータのインターフェイスにおいて音声を活用する動きが現実味のあるものとなってきている．より人間に近い機能を持ち合わせたインターフェイスを考える時，発話の意図や態度といった音声に含まれる感性情報の付与が不可欠である．これを実現するためには，普段われわれが話している日常会話について，感性情報を示すと思われる韻律情報の分析が必要である．

日常会話には基本感情以外に意図や態度をはじめさまざまな発話様式が入り組んでいる．本研究ではこうした音声の持つ感性情報，特に発話の意図や態度などをフェイスマークで表現することを提案する．フェイスマークには，1)汗などの微妙なニュアンスの表現が可能である，2)電子メール・携帯電話で盛んに用いられており一目で直感的に分かる，といった利点がある．また，認知心理学の研究において，絵文字・顔文字情報は音声言語コミュニケーションにおけるプロソディと同等の効果をもたらす[1]という報告もある．

本稿では，音声信号の中に含まれる韻律情報と，人間の知覚によって分類化された発話者の意図や態度など感性情報を表すフェイスマークとの対応関係を分析した結果を報告する．

2. 音声試料

音声試料として，JST/CREST 発話様式プロジェクトで作成中の自然発話音声データ[2,3,4]を用いた．このデータベースは，以下のような特徴を持つ．

－自由対話（電話を含む）
－関西弁の特定の女性話者 1名
－2年間に渡って継続的に収録
－1収録あたり 6分～30分程度
－多様な対話者

今回の分析では，発話の単位として，明らかなポーズ，または明らかなピッチの立ち上がりが知覚された場合に半自動で区切った句を扱うことにした．その中から，人手で付与されているラベルをもとに，文内容に加えて韻律や声質による情動が含むとされる発話302個を選択し分析対象とした．

3. フェイスマークによるラベル付与

3.1. 使用したフェイスマーク

本研究で使用したフェイスマーク（以下，マーク）は表 1に示す 19種類である．

表 1:使用フェイスマーク

(1)	(^-^)	(2)	(*^_^*)	(3)	(^O^)	(4)	(^.^)	(5)	(≧∀≦)
(6)	(^_^)	(7)	(-_-#)	(8)	(￣□￣)	(9)	Σ(°□°;)!!	(10)	(v_v)
(11)	(--)	(12)	(´Д`)=3	(13)	(´Д`)	(14)	(>_<)	(15)	(?_?)
(16)	(°o°)	(17)	(-´-)	(18)	(￣￣)	(19)	(‥?)		

3.2. 知覚評価実験

2節で既述した 302個の各発話が，表 1の 19種類のマークのどれにあたるかを決定するために知覚評価実験を行った．被験者は日本人学生 3人である．評価実験は無音室でヘッドホンを使用し，302 個の各発話の聴取回数は任意とした．

3.3. 評価結果

302 個の発話のうち，3 人（S1，S2，S3）が一致したマークを付与した発話は 184 個（61%）であった．また，2 人が一致したマークを付与した発話は 101 個（33%），付与したマークが 1 人も一致しなかった発話は 17 個（6%）であった．また，各マークにおける 3 人の評価頻度，および 3 人が一致した評価頻度を図 1に示す．

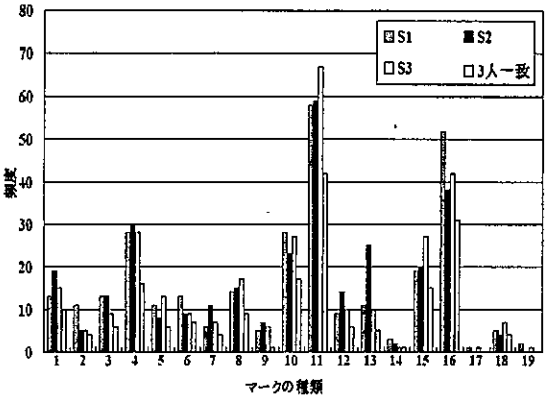

図 1: 各マークにおける評価頻度

4. 決定木学習による分析

4.1. 学習に用いる韻律パラメータ

分析に用いる韻律情報として，基本周波数(F0)，パワー(Power)，発話時間長(Duration)，発話速度の以下に示す各パラメータ（図 2）を 14個用いた．

－F0：レンジ，平均，傾き 5つ(図 2中の A,B,C,D,E)
－Power：最大，最小，レンジ，平均，分散
－Duration
－発話速度（mora/sec）

但し，F0 の 5つの傾き A,B,C,D,E はそれぞれ，発話開始位置・最大値位置間の傾き，最大値位置・発話終了位置間の傾き，発話開始位置・最小値位置間の傾き，最小値位置・発話終了位置間の傾き，最大値位置・最

* "Analysis of affect information in spontaneous speech with face mark" by Kazuhiro SHONO (Nara Institute of Science and Technology (NAIST)), Hideki KASHIOKA (NAIST/ATR) and Nick CAMPBELL (NAIST/ATR/CREST)

小値位置間の傾きである．また，F0，Power のレンジは最大値・最小値幅である．

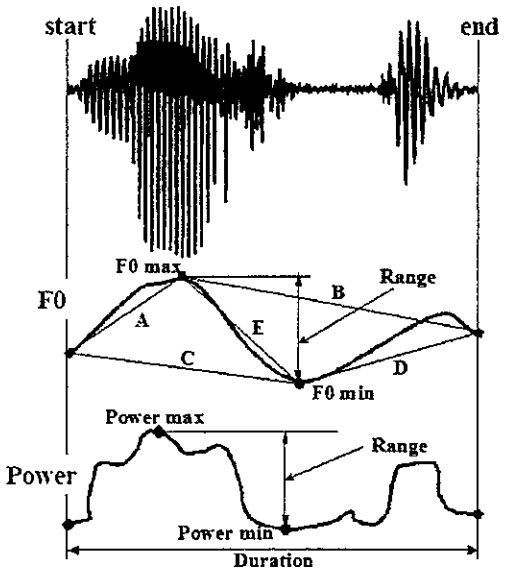

図 2: 韻律パラメータ

4.2. 決定木を用いた学習

各発話に付与されたマークと音声信号から抽出した韻律パラメータとの間にある相関を調べるために，決定木を用いた学習を行った．学習データとして，3 節において知覚評価を行った 302 発話およびその評価結果を使用し，3 人の評価結果をもとに延べ 906 発話を用意した．決定木の学習には C4.5 アルゴリズム[5]を利用した．

4.3. closed なデータによる決定木学習

用意した 906 発話の内，表 1 中の(17)，(19)のマークが評価実験で付与された 5 発話を除いた 901 発話を訓練データとして，closed な決定木の学習を行った．これは，図 1 より明らかなにデータ量が不足していると考えられるためである．

決定木の精度評価は，決定木の判別率と判別の信頼度を示す κ 値，および分類精度を示す再現率(Recall)・適合率(Precision)・F-measure を用いる．表 2 上段に，closed な決定木の判別率および κ 値を示す．また，表 3 に各マークにおける再現率，適合率，F-measure を示す．

この判別率と κ 値から，作成された決定木による判別は精度が高く信頼できるものであると考えられる．一方で分類の精度を見ると，発話データの頻度が少ないマークにおいて，表 3 中の(9)，(13)のように分類精度がよいとは言い難いものがある．今後，各マークにおける発話データ数を揃えた分析を行う必要があると言える．

4.4. 交差検定による判別の評価

次に，4.3 と同様の 901 発話を用い，これに対して 20 分割の交差検定を行った．

交差検定による精度評価についても 4.3 同様，判別率と判別の信頼度を示す κ 値，および分類精度を示す再現率・適合率・F-measure を用いる．交差検定による判別率および κ 値を表 2 下段に示す．また，表 3 に各マークにおける再現率，適合率，F-measure を示す．

判別率，κ 値共に，closed なデータによる学習に比べ精度が低くなっている．また，分類の精度においても同様のことが言える．改善には，聴覚特性に基づいた韻律パラメータの検討と共に，4.3 節同様に偏りのない学習データによる分析が望ましいと考えられる．

表 2:決定木学習の判別精度

	判別率	κ 値
closed な木の評価	84.68%	0.829
交差検定による評価	71.92%	0.687

表 3:分類精度（closed / open set）

		Recall	Precision	F-measure
(1)	(￣-￣)	0.923 / 0.68	0.766 / 0.723	0.837 / 0.701
(2)	(*^_^*)	0.833 / 0.667	0.714 / 0.571	0.769 / 0.615
(3)	(^0^)	0.867 / 0.657	0.743 / 0.657	0.8 / 0.657
(4)	(^_^)	0.757 / 0.67	0.907 / 0.686	0.825 / 0.678
(5)	(≧∀≦)	0.795 / 0.727	0.969 / 0.75	0.873 / 0.738
(6)	(^_^)	0.867 / 0.697	0.839 / 0.75	0.852 / 0.719
(7)	(-_-#)	0.833 / 0.636	0.833 / 0.583	0.833 / 0.609
(8)	(￣□￣)	0.824 / 0.75	0.913 / 0.783	0.866 / 0.766
(9)	Σ(°□°;)!!	0.786 / 0.667	0.611 / 0.333	0.688 / 0.444
(10)	(v,v)	0.845 / 0.778	0.91 / 0.808	0.877 / 0.792
(11)	(-,-)	0.873 / 0.764	0.897 / 0.81	0.885 / 0.7786
(12)	(´Д)=3	0.833 / 0.76	0.758 / 0.576	0.794 / 0.655
(13)	(´Д`)	0.714 / 0.489	0.652 / 0.478	0.682 / 0.484
(14)	(>_<)	0.833 / 0.75	0.833 / 0.5	0.833 / 0.6
(15)	(?_?)	0.932 / 0.73	0.833 / 0.697	0.88 / 0.713
(16)	(°o°)	0.878 / 0.757	0.871 / 0.78	0.875 / 0.769
(18)	(￣~￣)	1 / 0.75	0.75 / 0.75	0.857 / 0.75

5. まとめ

本稿では，自然発話音声を用いて，音声信号の中に含まれる韻律情報と人間の知覚によって分類化された発話者の意図や態度など感性情報を表すフェイスマークとの対応関係を分析した．

今後は，各マークにおける発話数に偏りがない学習データを用いた分析を行う必要があると同時に，人間の聴覚特性に基づいた韻律パラメータの検討をする必要があると考えられる．

謝辞 本研究の一部は科学技術振興機構戦略的基礎研究推進事業(JST/CREST)の援助により行われた．

参考文献

[1] 八田,インターネット・パソコン通信における文字情報の伝達効率改善に関する認知心理学的研究,電気通信普及財団研究調査報告書,2003
[2] Campbell, Mokhtari, "Voice Quality, the 4th prosodic dimension", Proc ICPhS 2003,pp.2414-2420,2003
[3] Campbell, "Labelling natural conversational speech data",音講論 1-10-22,pp.273-274,2002-10.
[4] http://feast.his.atr.jp/
[5] J,Ross Quinlan, C4.5:Programs for Machine Learning, Morgan Kaufmann Publishers