

活性-評価次元に基づくパラ言語情報ラベルの音響関連量*

森 大毅, 相田 千尋, 粕谷 英樹 (宇都宮大・工)

1 はじめに

対話音声には多様な情報が含まれている。特に、言語情報に付随して運ばれる、いわゆるパラ言語情報が対人コミュニケーションで演じる役割は重要である。近年、人間と機械のコミュニケーションにおいてもパラ言語情報を重視する立場が目立つようになり、話し方のニュアンスを認識する対話システム、あるいは表情豊かな合成音声を目指す研究が注目されている。これらを実現するためには、パラ言語情報の音響関連量を明らかにすることが必要不可欠である。

音響関連量を探究するためには音声コーパスが必要である。しかし、音声に含まれるパラ言語情報をラベリングする確立された方法はない。特に自発音声の場合は、感情を表現する語をリストから選択する方法が困難である点で、いわゆる感情音声のコーパスとは違った難しさがある。

一方、感情心理学の分野では、感情状態は少数の次元が作る空間上の点として表現できるという考え方があり、その代表例が、活性 (activation)-評価 (evaluation) 空間 [1] と呼ばれるものである。Schröder[2] は、テレビ番組中の対話に対して活性-評価空間に基づき付与されたラベルを用い、各々の次元の音響関連量を見出すことを試みた。その結果、活性次元は特に基本周波数 (f_0) に強く関連していることを示している。しかしながら、評価次元の音響関連量についてはこれまで明らかにされてこなかった。

著者らは活性および評価を含む次元に基づき、自発的な対話音声にパラ言語情報ラベルを付与したデータベースを構築する試みを行ってきた [3]。本研究は、このデータベースに基づき、特にこれまで明確でなかった評価次元の音響関連量を見出すことを目的とする。具体的には、多くのパラ言語 / 非言語情報に関連している f_0 に加え、過去の研究であまり考慮されてこなかった有声音源の気息性を反映する音響パラメータを導入し、このパラメータが特に「不快」な状態に強く関連していることを示す。

2 データベース

本論文が分析の対象とするのは、「4 コマまんが並べ換え課題」をタスクとする音声対話データベース [3] に含まれる 1 セッション分の対話である。話者は大学生 2 名 (FUE, FKC) で、合計 113 発話が含まれる。

各発話には、14 名のラベラにより評価された 6 軸の抽象次元に基づくパラ言語情報ラベルが付与されている [4]。これらの次元のうち、「覚醒-睡眠」は発話者の心理状態の活発さを表し、活性次元に相当する。また、「快-不快」は発話者にとっての良い / 悪い、好き / 嫌いなどに関連しており、評価次元に相当する。

ラベラに各発話の前後関係がわからないようランダムな順序で音声を呈示し、1 から 7 までの 7 段階で評価させた「快-不快」および「覚醒-睡眠」の値を、ラベラ全員で平均化し、当該発話のパラ言語情報ラベルとみなす。

3 音響パラメータ

本論文でパラ言語情報ラベルとの関連を分析する音響パラメータは以下のものである。

f_0 レンジ

1 発話中の対数 f_0 の最大値と最小値の差で定義する。 f_0 抽出エラー等は手作業で修正した。 f_0 レンジは覚醒レベルとの関連がたびたび指摘されてきたパラメータの 1 つである [5]。

平均 $f_{\text{aperiodic}}$

$f_{\text{aperiodic}}$ は有声音源スペクトルの周期的 / 非周期的成分境界周波数 [6] を表す。有声音の気息性を反映するパラメータと考えられ、非周期的成分の割合が増加すると $f_{\text{aperiodic}}$ は低下する。1 発話中の全有声音区間における $f_{\text{aperiodic}}$ の平均値で定義する。

4 分析結果

まず、「覚醒-睡眠」と f_0 レンジとの関連について述べる。相関係数は、FUE で 0.67、FKC で 0.54 と比較的大きく、抑揚のある音声は覚醒した状態と知覚されることが確かめられた。これは過去の研究を支持する結果になっている。

次に、「快-不快」と f_0 レンジの散布図を Fig. 1 および Fig. 2 に示す。図では、相槌など /uN/ だけから成る発話を区別して表示してある。FUE では f_0 レンジとの相関は比較的大きい (相関係数 0.64) が、評価値 3-4 (やや不快 ~ 中立) にかけては f_0 レンジがばらついていることがわかる。一方、FKC では相関は小さい (相関係数 0.34)。

* Acoustic correlates of paralinguistic information annotated with the activation and evaluation dimensions.
By MORI, Hiroki, AITA, Chihiro, KASUYA, Hideki (Utsunomiya University)

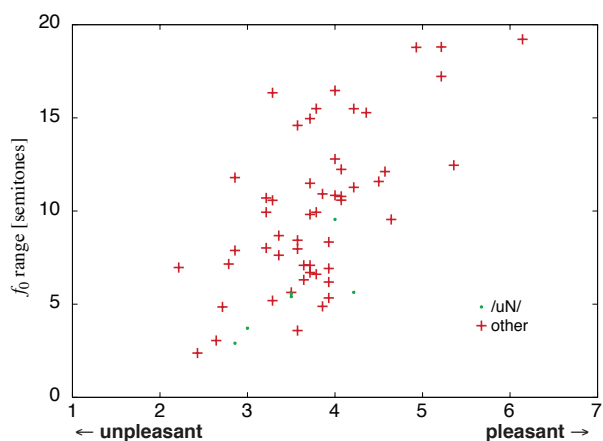


Fig. 1 「快-不快」と f_0 レンジの関係 (話者 FUE)

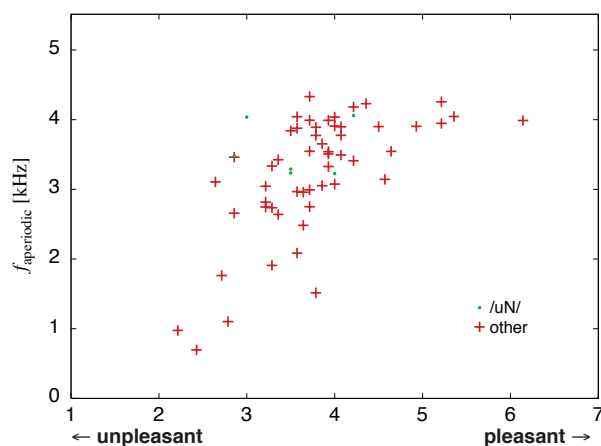


Fig. 3 「快-不快」と $f_{\text{aperiodic}}$ の関係 (話者 FUE)

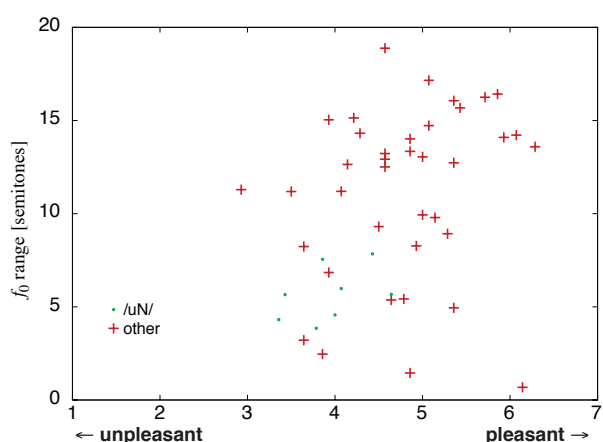


Fig. 2 「快-不快」と f_0 レンジの関係 (話者 FKC)

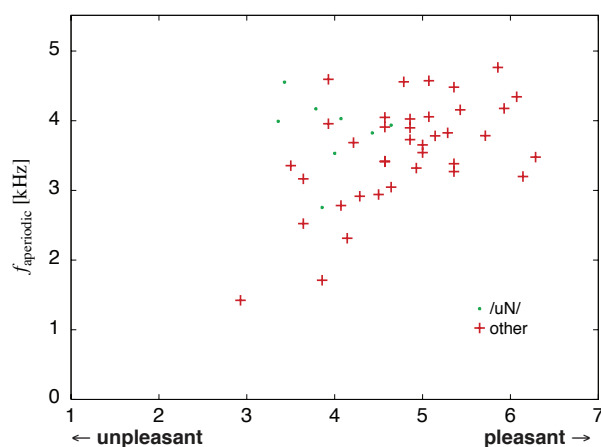


Fig. 4 「快-不快」と $f_{\text{aperiodic}}$ の関係 (話者 FKC)

Fig. 3 および Fig. 4 に、「快-不快」と平均 $f_{\text{aperiodic}}$ の散布図を示す。FUE では、平均 $f_{\text{aperiodic}}$ との相関が比較的大きいことがわかる (相関係数 0.60)。特に、 $f_{\text{aperiodic}} < 3$ [kHz] ではほとんどの場合で評価値が 4 未満 (不快) であり、このことから有声音に氣息性成分が含まれていると不快な状態であると知覚されていることがわかる。FKC でも同様の傾向が見られる (相関係数 0.39)。

f_0 レンジと平均 $f_{\text{aperiodic}}$ を説明変量として重回帰分析を行った場合の重相関係数は、FUE で 0.74、FKC で 0.48 となった。これより、これらのパラメータはいずれも快-不快に強く関連していることがわかった。

5 おわりに

活性-評価次元により記述されたパラ言語情報と音響パラメータとの関連を分析した。その結果、 f_0 だけでなく氣息性を表すパラメータが評価次元と強く関連していることがわかった。今後は分節的特徴との関連についても調査したい。

参考文献

- [1] R. Cowie *et al.*, "Emotion recognition in human-computer interaction," *IEEE Signal Processing Magazine*, **18**, 32–80 (2001).
- [2] M. Schröder, "Speech and emotion research: An overview of research frameworks and a dimensional approach to emotional speech synthesis," PhD thesis, Saarland Univ. (2004).
- [3] H. Mori *et al.*, "Some considerations for designing spoken dialogue database from the viewpoint of paralinguistic information," *Acoust. Sci. & Tech.*, **24**, 376–378 (2003).
- [4] 森大毅他, "対話音声のパラ言語情報ラベリングの安定性," 日本音響学会誌 (掲載予定).
- [5] K. Scherer, "Vocal correlates of emotional arousal and affective disturbance," in *Handbook of Social Psychophysiology* (John Wiley & Sons, New York, 1989), pp. 165–197.
- [6] T. Ohtsuka and H. Kasuya, "Aperiodicity control in ARX-based speech analysis-synthesis method," *Proc. Eurospeech 2001*, pp. 2267–2270 (2001).