

車内状況アウェアネスのための感情音データベースの設計と評価*

○ 滝口 哲也, 有木 康雄 (神戸大・工)

1 はじめに

これまでの音声対話によるカーナビゲーションシステムでは、ユーザーがシステムに目的地などを音声で問い合わせる事により、ユーザーの要求に応答するものが多い。つまりシステムへの「問い合わせ」が必要である。一方、人間同士の対話では、例えば楽しそうに海の話をしていれば、「海に行きたいのだな」と理解可能であり、また、会話がつまらなそうであれば、話題を変えるなど、その場を読んだり、状況を察する事により、互いの要求に応答する事が出来る。そこで我々は、明らかな問い合わせがなくても、「状況や雰囲気を感じ」、「場を読む」事の出来る（状況アウェアネス）対話ナビゲーションシステムを目指し、研究を進めている。本稿では、感情による状況アウェアネスに注目し、実際に60人の被験者が発話した音声データから感情の抽出を行った。収録データの主な状況としては、「気分が良い」、「疲れている」、「怒っている」、「ノーマル（通常）」の4種類とした。本稿では、収録内容、及び予備実験の結果について報告する。

2 感情データベース

2.1 車内状況の設定

被験者は、以下に示す12個の状況に従い、発話（と動作）を行った。また、具体的な発話データ（スクリプト）は、被験者が自分の体験等をもとに即興で発話したものであり、予め用意された原稿を読み上げたものではない。言葉以外の発声として、ため息、笑い、鼻歌等も含まれる。

- 1 好きな音楽が流れて喜んでいる。
- 2 テレビに流れている情報を見て、喜んでいる。または、ラジオ等を聞いて喜んでいる。
- 3 外の景色を見て、気分が良い。
- 4 渋滞等で、いらいらしている。
- 5 仕事や遊びで疲れている。

- 6 眠気がしている。

- 7 道を間違えた（怒っている。例えば自分ではカーナビの指示通りに運転したつもりなのに、間違えている。その他、助手席の人が間違えて道を教えた等。）

- 8 道を間違えた（明らかに、自分のミスで間違えた場合。）

以下の状況9から11に関しては、状況1から8以外で、指定した感情に対応する各自が体験した状況を設定する。

- 9 気分が良い。うれしい。楽しい。

- 10 疲れている。

- 11 怒っている。

- 12 ノーマル（通常）な状況での会話等。

2.2 収録方法

本収録作業では、感情データに注目するため、状況（発話内容）のみ車内を想定し、実際の収録作業は車内ではなく屋内にて行った。従って、音声データには、エンジン音や様々な背景雑音は含まれていない。

被験者は、一人ずつ椅子に座り、データ収録を行った。収録は複数人による対話形式ではなく、各被験者は、車内に相手がいる場合を想定するか、もしくは独り言を想定し、各状況毎に数回発話を行った。以下に、収録データの文例を示す。

状況4 モー何分たってると思ってんの、はよーな、はよー、もー、もーいくで、ほら信号変わるやん。

状況5 あーーーー、もー、今日泳ぎ過ぎたわ、ほんま、めっちゃけだるい。あーなんか、ごめん、飲み物とってくれへん、ありがとう、あけて。

* A design and an evaluation of emotional speech database for in-car situation awareness. by Tetsuya Takiguchi and Yasuo Ariki (Kobe University)



Fig. 1 収録背景

Table 1 被験者の年齢層，男女別人数

年代	男性	女性
18～30	20 人	20 人
31～50	10 人	10 人
合計	30 人	30 人

次に，収録背景を図 1 に示す．被験者は図の右側の椅子に座り収録を行う．マイクロフォン（SONY ECM-66B）は 2 個使用した．一つは被験者の胸元に取り付け，もう一つはマイクスタンドを使用して，被験者の頭部前方数 10cm に取り付けた．音声データは，サンプリング周波数 44.1kHz，量子化ビット数 24bit で録音されている．また，本実験では使用していないが，Digital Hi-Vision カメラ（Victor GR-HD1）を話者の前方に設置して，被験者の表情等の収録も行った．

表 1 に，被験者の年齢層と男女別人数を示す．被験者の男女比率は等しくなるようにした．

3 認識実験

3.1 評価方法

感情音認識における音響特徴量としては，ピッチ，パワー等に関連した統計量がしばしば使われる [1, 2, 3]．またプロソディ情報だけでなく，MFCC も考慮した特徴量が使われる事もある [4]．本実験では，各フレームで MFCC+ Δ MFCC とピッチ，パワーを連結し，各感情毎に GMM を作成し識別を行った．GMM の混合数は 4 とした．

3.2 実験条件

本実験では，「怒り」，「疲労」，「ノーマル（通常）」の 3 種類の感情識別を行う．各々，状況 10，11，12 に対応している．ただし，収録したデータは，必ずしも同一状況下の全ての時間において，指定した感情とはなっていなかったため，収録したデータに対して，おおまかなラベル付け

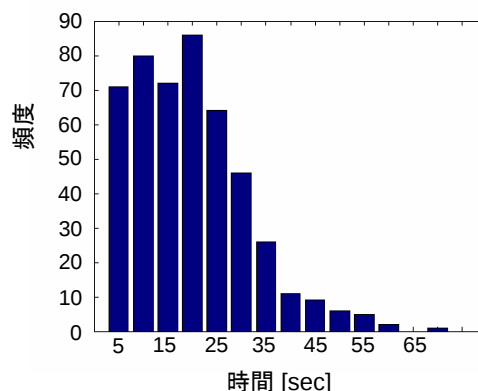


Fig. 2 発話時間の分布

を行い，感情が変化する毎にファイルの分割を行った．3 つの状況下で収録した音声ファイルの発話時間の分布を図 2 に示す．平均発話時間は，20.9 秒となった．

GMM の学習用に 50 人（男性 25 人，女性 25 人）を使用し，評価用データとして 10 人分（男性 5 人，女性 5 人）を使用した．

3.3 実験結果

感情識別実験を行った結果，MFCC+ Δ MFCC（32 次元）で 71.4%，更にピッチ+パワー（2 次元）を加えることにより，72.7%となった．

4 おわりに

本稿では，車内状況アウェアネスのための感情音データベースの内容について述べた．今後は感情識別に適した音響特徴量について検討していく．

参考文献

- [1] V. Hozjan, Z. Kačič, “Improved emotion recognition with large set of statistical features,” Eurospeech 2003, pp. 133-136.
- [2] B. Schuller, G. Rigoll, and M. Lang, “Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine - belief network architecture,” ICASSP 2004, pp. 577-580.
- [3] S. Hock, F. Althoff, G. McGlaun, G. Rigoll, “Bi-modal fusion of emotional data in an automotive environment,” ICASSP 2005, pp. 1085-1088.
- [4] Y. Wang and L. Guan, “Recognizing human emotion from audiovisual information,” ICASSP 2005, pp. 1125-1128.