

宇都宮大学パラ言語情報研究向け音声対話データベースの構築*

○森 大毅 (宇都宮大), 粕谷 英樹 (国際医療福祉大/宇都宮大), △中村 真 (宇都宮大)

1 はじめに

音声言語に付随して伝達されるパラ言語情報は音声対話において核心的な役割を担っており、その運用、構造および効果を明らかにすることは、基礎研究のみならず音声合成・音声言語理解などの応用のためにも重要である。しかし、このような研究を主眼に置いて設計・構築された共通の音声コーパスはこれまで存在しなかった。

こうした背景から、宇都宮大学では自然で表情豊かな対話音声に含まれるパラ言語情報に関する研究への利用に適した音声対話データベース(以下、UUデータベース)を構築してきた。今回、UUデータベースを広く音声・言語研究者の利用に供すべく公開する。本稿では、UUデータベースの設計と収録、パラ言語情報の評定方法およびその統計情報について述べる。

2 設計と収録

UUデータベースは、大学生7ペア(女性12名、男性2名)による自然な対話における音声から成る。対話参加者は宇都宮大学に在籍する学部生(大半は文系学部)であり、音声収録にあたった第1著者の印象では概して快活である。対話参加者の募集は同学年の友人同士のペアを単位として行い、普段いわゆる「ため口」で会話できる間柄であることを条件とした。また、上記7ペアに含まれない男性2ペアについては、収録した対話が感情の起伏に乏しいと判断されたため除外した。このような構成としたのは、音声伝達する情報に占めるパラ言語情報の比重が、親近性の高さに関連すると予想されるからである。

対話参加者には、「4コマまんが並べ換え課題」と呼ぶ、ばらばらにされた4コマまんがの2コマ分をそれぞれ持ち、対話により元の順番を推定する課題を遂行させた。この課題では、納得の行くストーリーを想像するプロセスや、女子学生が好むキャラクターを取り入れるなどの課題に没入させるための工夫により、極めて自発性が高く、言語的にも音声学的にも多様な対話音声収録

できる。1編の4コマまんがに対応する一連の対話をセッションと呼んでいる。各ペアが遂行した対話数は3から7セッションで、データベース全体では27セッション、約130分である。各話者の発話は、スラッシュ単位と間休止単位に基づく発話単位[1]で切り出され、データベース全体では4737発話となっている。

3 パラ言語情報ラベルの付与作業

UUデータベース最大の特徴の1つは、音声から知覚される話者の感情状態を記述したパラ言語情報ラベルが付与されていることである。本節では、パラ言語情報ラベリングの方法および作業の実際について述べる。

自由な会話においては、「怒り」「悲しみ」のような狭い意味での感情を含む音声はむしろ稀であり、伝統的な感情カテゴリを用いたラベリングは困難である。これまでの研究で、著者らは「快-不快」「覚醒-睡眠」「支配-服従」「信頼-不信」「関心-無関心」「肯定的-否定的」の6抽象次元に基づき、発話から知覚される話者の感情状態を記述することを試みてきた。[2]では、UUデータベース中の2セッションについて、多数評定者によるラベリングの諸性質を詳細に検討している。その意味で、この2セッションは性質がよくわかっている「コア」サブセットである。

データベース全体に対するパラ言語情報ラベリングは、[2]の知見を基に2006年度に遂行された。[2]におけるラベラは、音声の研究室に在籍する学生が主体であったが、データベース全体に対する作業は長期に及ぶため、学生では負担が大きすぎると考えられた。

そこで、人材派遣会社を通じて6名のラベラ候補者を募るとともに、「コア」に対するラベリング作業を実施し、スクリーニングを行った。スクリーニングは以下の基準で行った。

基準1 一貫性

基準2 平均評定値との類似性

基準3 評定項目の独立性

*Utsunomiya University Spoken Dialogue Database for Paralinguistic Information Studies. by MORI, Hiroki (Utsunomiya University), KASUYA, Hideki (IUHW/Utsunomiya University), NAKAMURA, Makoto (Utsunomiya University)

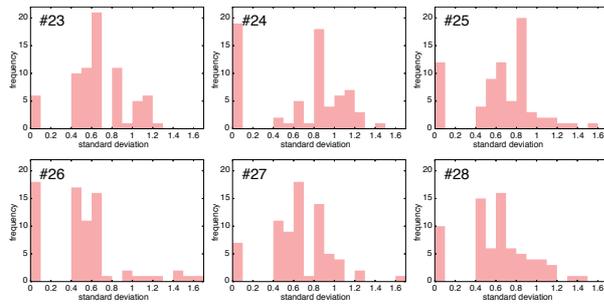


Fig. 1 同一発話に対する評定値の標準偏差分布

Table 1 平均評定値との相関係数

	ラベラ					
	#23	#24	#25	#26	#27	#28
快-不快	0.76	0.74	0.69	0.48	0.69	0.68
覚醒-睡眠	0.75	0.74	0.73	0.65	0.81	0.66
支配-服従	0.63	0.55	0.60	0.64	0.56	0.66
信頼-不信	0.63	0.64	0.65	0.56	0.57	0.60
関心-無関心	0.66	0.60	0.57	0.68	0.55	0.41
肯定的-否定的	0.61	0.60	0.62	0.49	0.63	0.50

基準1は、同一発話に対して同一のラベリングができているかを見るためのもので、順序をランダム化した「コア」中に混入させた8個の同一発話(12種類)の評定値のばらつきで判定する。Fig. 1に6名のラベラ候補者(#23から#28)の標準偏差の分布を示す。#24, #25は標準偏差が0.8を超えるばらつきの大きい発話が多い。それ以外のラベラの標準偏差の分布は、[2]における大部分のラベラの標準偏差と比べて大きいものではなかった。なお、#26は一見すると非常に安定したラベラのように見えるが、これは#26の評定が保守的(レンジが狭い)であるためである。

基準2は、大多数のラベラと類似した評定ができているかを見るためのもので、過去[2]に収集した22名による評定値の平均との相関係数(Table 1)で判定する。6名のラベラ候補者の中には、特に問題は見出されなかった。

基準3は、ラベリング結果を観察して事後的に設けた基準である。#24は、全発話のうち30%以上について、6次元の評定項目全てに対して同じラベルを付与していた。ほかのラベラの場合には数%程度であるから、#24は集中力あるいは評定項目の理解が不足していたと考えられる。

以上の結果、基準1および基準3で問題があった#24および#25、および作業継続が不可能となった#28を除外した残りの3名をラベラとして選出し、全セッションに対するラベリングを遂行させた。

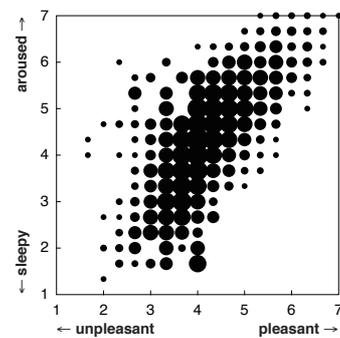


Fig. 2 UU データベースの平均評定値分布

4 評定値分布と評定の一致率

パラ言語情報を記述する抽象次元の中でも、「快-不快」「覚醒-睡眠」の2次元は、話者自身の感情状態の最も一般的な指標であるとされている。Fig. 2は、UUデータベースに含まれる4737発話に対して、選出した3名のラベラが評定した「快-不快」および「覚醒-睡眠」の平均値の分布である。図中、円の面積は発話数に比例している。「快-不快」「覚醒-睡眠」の間には正の相関があり($r = 0.53$)、Fig. 2でも対角線付近のデータが多い。しかしながら、不快かつ覚醒(左上の領域)と知覚される発話も相当数含んでおり、全体として見れば、多様な感情状態での発話を含んだデータベースとなっていると言える。

3名のラベラによる評定値の一致率は、完全一致(3名とも同じ)が22.09%(偶然水準2.04%)、部分一致(2名が同じ)が83.92%(偶然水準38.78%)となっている。

5 おわりに

UUデータベースは、国立情報学研究所・音声資源コンソーシアムを通じて配布する予定である。UUデータベースの最新情報は、<http://uudb.speech-lab.org/>より入手できる。

謝辞

実験協力者の佐竹智幸さんに感謝します。この研究は科研費(17700172)の支援を受けている。

参考文献

- [1] 森大毅: “対話音声のパラ言語情報研究用データベースの分析単位,” 人工知能学会研究会資料, SIG-SLUD-A603, pp. 9-14, 2007.
- [2] 森大毅, 相澤宏, 粕谷英樹: “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, 61(12), 690-697, 2005.