

相槌が伝達するパラ言語情報の文脈を考慮した生成*

長塚健二, 森大毅 (宇都宮大院・工学研)

1 はじめに

近年、音声対話システムへの応用を目的とした音声合成についての研究が盛んに行われている。西村ら [1] は人間同士の対話の印象と韻律変化との間にもどのような関係性があるかを分析し、塚原 [2] は韻律情報と文脈情報を用いて機械が打つ相槌の使い分けを行った。これらは共に、音声対話システムにおける人間と機械の対話の円滑化を目指したものである。

人間と機械の対話を更に円滑なものにするためには、音声合成において対話の文脈を考慮した適切なパラ言語情報の生成、即ち、システムが応答発話の話し方の違いを決定し、それらを音声合成に適用することが必要である。そこで、本研究では、対話の文脈と応答発話のパラ言語情報との関係をモデル化することを目的とした検証を行った。

その第一歩として、相槌を対象とした実験を行った。相槌は、対話を円滑に進めるという役割を持っているとされており、この機能を音声対話システムに持たせる事は有用である。

本稿では、対話の文脈を考慮した相槌のパラ言語情報の推定について報告する。具体的には、相槌が打たれた発話、つまり、先行する発話の音響特徴量およびパラ言語情報から、機械学習により、システムが打つ相槌のパラ言語情報の推定を行う。

2 対話コーパス

音声データとして、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 (UUDB) [3] を用いた。このコーパスには、大学生 7 組 14 名が「4 コマ漫画並べかえ課題」というタスクを行った対話音声収録されている。そして、それぞれの音声ごとに 6 次元のパラ言語情報 (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) が 3 名のラベラにより 1~7 の 7 段階の値でラベリングされている。

本研究では、UUDB 中の全発話 4840 発話の中から「相槌」とラベリングされている 847 発話を対象として実験を行った。

3 相槌が伝達するパラ言語情報の多様性

通常の発話に比べ、相槌のパラ言語情報は画一的である可能性もある。従って、まずパラ言語情報を生

Table 1 各パラ言語情報の平均評価値の平均値および標準偏差

	全発話		相槌	
	平均値	標準偏差	平均値	標準偏差
快-不快	4.11	0.80	3.87	0.64
覚醒-睡眠	4.31	1.09	3.20	0.81
支配-服従	4.18	1.21	2.85	0.38
信頼-不信	4.49	0.80	4.31	0.65
関心-無関心	4.83	0.88	4.05	0.93
肯定的-否定的	4.53	0.81	4.36	0.72

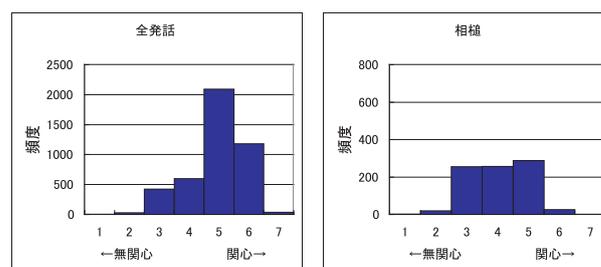


Fig. 1 関心-無関心の平均評価値の分布

成する意義の検証を行う必要がある。

パラ言語情報の各次元に対する平均評価値の平均値および標準偏差を Table 1 に示す。Table 1 より、快-不快、覚醒-睡眠、信頼-不信、肯定的-否定的では標準偏差に大きな違いはなかった。一方、支配-服従は標準偏差が著しく減少した。これより、支配-服従については、相槌のパラ言語情報が画一的であることが分かる。逆に、関心-無関心 (Fig. 1) は最も標準偏差が大きく、全発話の場合よりも多様性が大きくなっている。関心-無関心は評価値 3~5 に多様な分布を示していることが分かる。

以上の結果より、支配-服従を除けば相槌のパラ言語情報は多様であり、機械学習により推定を行う意義があると言える。

4 パラ言語情報の生成モデル

パラ言語情報の生成モデルの構築を、回帰木および重回帰分析により行った。これらを実行するツールには WEKA [4] に搭載されている REPTree および LinearRegression を使用した。説明変数としては、先行発話のみから得られるパラメータ 14 個、および、

* Paralinguistic information generation based on contextual attributes in dialogue for emotional back-channel speech synthesis. by NAGATSUKA, Kenji, MORI, Hiroki (Utsunomiya University)

Table 2 説明変数一覧

内容	P1	S1	P2	S2	内容	P1	S1	P2	S2
快-不快 (Pleasantness)					後自発話とのポーズ長				
覚醒-睡眠 (Arousal)					後自発話との Pleasantness の傾き				
支配-服従 (Dominance)					後自発話との Arousal の傾き				
信頼-不信 (Credibility)					後自発話との Dominance の傾き				
関心-無関心 (Interest)					後自発話との Credibility の傾き				
肯定的-否定的 (Positivity)					後自発話との Interest の傾き				
発話のモーラ数					後自発話との Positivity の傾き				
発話継続時間					後自発話との SR の傾き				
発話速度 (SR)					後自発話との f_0 max の傾き				
f_0 の発話内最大値 (f_0 max)					後自発話との f_0 min の傾き				
f_0 の発話内最小値 (f_0 min)					後自発話との f_0 ave の傾き				
f_0 の発話内平均値 (f_0 ave)					後自発話との f_0 ran の傾き				
f_0 の発話内レンジ (f_0 ran)					後自発話との intmax の傾き				
パワーの発話内最大値 (intmax)									

Table 3 推定値と正解ラベルとの RMS 誤差

	OR	回帰木		重回帰分析	
		14 個	83 個	14 個	83 個
快-不快	0.606	0.604	0.594	0.596	0.603
覚醒-睡眠	0.778	0.778	0.770	0.764	0.841
支配-服従	0.299	0.302	0.301	0.293	0.403
信頼-不信	0.585	0.606	0.557	0.574	0.603
関心-無関心	0.801	0.806	0.796	0.784	1.072
肯定的-否定的	0.603	0.603	0.597	0.589	0.630

それ以前の発話も含めたパラメータ 83 個を用いた。説明変数を Table 2 に示す。Table 2 において、P1 は相槌が打たれた相手発話についての情報、S1 は相槌の直前に発した自発話についての情報、P2 は P1 の直前に発した相手発話についての情報、S2 は S1 の直前に発した自発話についての情報であり、それぞれの情報に含まれているパラメータを「」で示している。目的変数としては、当該相槌のパラ言語情報 6 項目の平均評価値を用いた。

推定値と正解ラベルとの RMS 誤差を Table 3 に示す。ここで、OR とは、全相槌に対する平均値を推定結果とする方法であり、これをベースラインとする。

Table 3 より、快-不快、信頼-不信においては、説明変数 83 個を用いた回帰木の RMS 誤差が最も小さくなった。覚醒-睡眠、支配-服従、関心-無関心、肯定的-否定的においては、説明変数 14 個を用いた重回帰分析の RMS 誤差が最も小さくなった。

最適な機械学習アルゴリズムを用いた場合の快-不快、支配-服従、信頼-不信、肯定的-否定的に対する RMS 誤差は 0.6 未満となった。これは、森ら [5] で示されている「同一の発話に対する評定値の標準偏差は 0.5 付近に分布する」という知見を考えると、精度の良い推測が行われていると言える。

次に、快-不快、信頼-不信に対する回帰木 (説明変数 83 個) を分析した。ルートノードにおける質問は、

快-不快では S1 の快-不快であり、信頼-不信では S1 の信頼-不信であった。また、快-不快、信頼-不信の両方に共通して、S1 および S2 に関する質問が多く見られた。これより、S1 および S2 の情報は、相槌の快-不快、信頼-不信を決定する上で重要であると言える。

最後に、覚醒-睡眠、支配-服従、関心-無関心、肯定的-否定的に対する回帰式 (説明変数 14 個) を分析した。全体を通して P1 の肯定的-否定的の回帰係数が高く、P1 の肯定的-否定的の変化が、相槌の覚醒-睡眠、支配-服従、関心-無関心、肯定的-否定的の変化に対して比較的大きな影響があると言える。

5 まとめ

本稿では、対話の文脈を考慮した相槌のパラ言語情報の生成モデルの構築について報告した。

その結果、機械学習を用いることによって、各パラ言語情報の推定精度が良くなることが確認された。また、最適な機械学習の方法は、パラ言語情報ごとに違いがあることも確認できた。

今後は、今回の実験で推定精度が最も低かった関心-無関心を、精度良く推定できるようなモデルの構築をしていく予定である。

参考文献

- [1] 西村良太, 北岡教英, 中川聖一, 音講論 (春), 2-10-15, 2008.
- [2] 塚原涉, 東京大学工学研究科博士論文, 2000.
- [3] 森大毅, 粕谷英樹, 中村真, 音講論 (秋), 1-4-10, 2007.
- [4] WEKA, <http://www.cs.waikato.ac.nz/ml/weka/>
- [5] 森大毅, 相澤宏, 粕谷英樹, 日本音響学会誌, 61 巻 12 号, pp. 690-697, 2005.