

自律型アンドロイドによる対話における「同調的笑い」の生成

勝見 久央[†] 井上 昂治[‡] 中村 静[‡] 高梨 克也[‡] 河原 達也[‡]

[†] 京都大学 工学部情報学科

[‡] 京都大学 大学院情報学研究科 知能情報学専攻

1. はじめに

音声対話システムの応用として、人間と自然な日常会話をを行うことを目指した自律型アンドロイドの研究が行われている。また、「うん」、「はー」といった相槌を自然なタイミングで打ったり、ユーザの発話に対して適切な質問を返すといったシステムの研究も行われている [1,2]。実際の間どうしの対話では笑いが頻繁に見られ、適切なタイミングでの笑いは重要な要素であり、さらに一方が笑うともう一方がうつられて笑うといった同調的笑いは対話の場を和ませたり、盛り上げたりするなどの効果があると考えられる。したがって、自律型アンドロイドによる対話においても笑いを実現することは重要である。そこで本研究ではまず、人間どうしの対話でみられるような同調的笑いを雑談対話システムにおいて実現するため、女性の外見をした自律型アンドロイド ERICA [3] と人間との対話データを収録し、アンドロイドが人間につられて笑うべきタイミングを様々な特徴量を用いて分析を行う。さらに、分析結果に基づき、同調的笑いを行うか否かの予測を行う (図 1)。

2. 対話コーパス

本研究では、オペレータが遠隔操作したアンドロイド ERICA と被験者 1 名の 10 分程度の初対面対話 42 セッションを録音、録画したデータを用いた。被験者には、大学の研究室の教授に会いに来た研究室見学の学生や、営業に来た会社員といった、教授との面会のために来訪した人物を演じてもらった。実際に応接室を模した部屋に入室し、教授の秘書である ERICA と被験者の間で来訪の目的などを問いながら、不在の教授が戻るのを一緒に待つ、という流れの中で対話を行ってもらった。なお、オペレータは女性 6 名で交代して行った。これらの対話中に発生した被験者の笑いについて以下の分析を行った。

3. 分析

最初に、ERICA と被験者それぞれの笑いの発生箇所を人手によりアノテーションした。ただし発話単位は IPU (間休止単位 200msec) である。また、2 秒以上の間隔を開けずに同一話者から発された笑いは連続した 1 箇所の笑いであるとみなした。

3.1 同調的笑いの発生の有無による分類

被験者が発した笑いを、笑いと同調的笑いとの位置関係に着目した Gupta らの分類方法 [4] を参考にし、次の 3 つに分類した。

- 被験者の笑いのうち、その直後 4 秒以内に ERICA が同調して笑わなかったもの (stand-alone subject's laughter: St-S)
- 被験者の笑いのうち、その直後 4 秒以内に ERICA も同調して笑ったもの (subject-lead shared laughter: Sh-S)

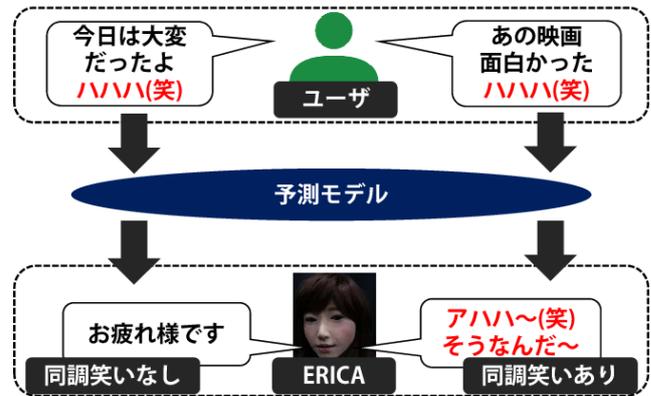


図 1: システム概要図

表 1: 3 種類の被験者笑いの回数

種類	St-S	Sh-S	Sh-E	計
	197	79	22	298

- 被験者の笑いのうち、その直前 4 秒以内に ERICA の笑いが発生しており、被験者がそれに同調して笑ったもの (ERICA-lead shared laughter: Sh-E)

3 種類の笑いの発生回数を表 1 に示す。42 セッションを通して被験者の笑いは 298 回発生した。被験者の関与する笑いのうち、被験者が主導して笑う St-S と Sh-S が全体の 92.6% (= (197+79)/298) を占めていた。以降ではこの被験者主導の笑いの St-S と Sh-S の 276 回について分析を行う。

3.2 直前の両者の発話の極性による分類

人間どうしの対話であれば、ポジティブな内容の会話では同調的笑いが発生しやすく、逆にネガティブな内容の会話中には同調的笑いが発生しづらいと考えられる。そこで、発生した被験者主導の笑いを、直前の発話の極性に着目した Gupta らの分類方法 [5] を参考に、以下の手順で、positive laughter、neutral laughter、negative laughter のいずれかに分類した。

1. 対象の笑いの直前の発話を両者合わせて 5 発話まで遡り、それぞれに positive、neutral、negative いずれかの極性をつける。
2. その 5 発話の極性から笑いの直前の会話の極性を次のようにして決定する。
 - 5 発話の極性が neutral の 1 種類のみからなる場合、その極性を neutral とする。
 - 上記にあてはまらない場合、neutral な発話を取り除き、時間的に最も笑いの発生箇所に近い発話の極性を、その極性とする。
3. 分類する笑いを、前の手順で求めた極性が positive であれば、positive laughter、neutral であれば neutral laughter、negative であれば negative laughter とする。

Generating shared laughter by autonomous android : Hisao Katsumi, Koji Inoue, Shizuka Nakamura, Katsuya Takanashi, Tatsuya Kawahara (Kyoto Univ.)

表 2: 6 種類の被験者主導笑いの回数

	St-S	Sh-S	計
positive laughter	80	21	101
neutral laughter	84	32	116
negative laughter	41	18	59
計	205	71	276

表 3: 4 種類の被験者主導笑いの回数

	St-S	Sh-S	計
speech laughter	112	56	168
stand-alone laughter	93	15	108
計	205	71	276

表 4: 4 種類の被験者主導笑いの回数

	St-S	Sh-S	計
obvious laughter	72	32	104
breath laughter	133	39	172
計	205	71	276

各発話の極性は、MeCab と CaboCha を用いて日本語極性辞書 [6, 7] にエンタリーが存在する名詞、用言を抽出し、それらの極性のスコア平均とした。そして、各被験者主導の笑いをその極性 3 種と直後の同調的笑いの発生の有無により 6 種類に分類したところ、それぞれの発生回数は表 2 のようになった。

3.3 発話と笑いの関係性による分類

発生した被験者主導の笑い (St-S、Sh-S) を、発話と笑いの位置関係に着目した Tian らの分類方法 [8] を参考に、次の 2 つに分類した。

- 発話しながら発された笑い (speech laughter)
- 発話を伴わず単独で発された笑い (stand-alone laughter)

そして、被験者主導笑いを上述の 2 種と直後の同調的笑いの発生の有無により 4 種類に分類したところ、それぞれの発生回数は表 3 のようになった。

3.4 笑い声の聞こえ方による分類

発生した被験者主導の笑い (St-S、Sh-S) を笑い声の聞こえ方に着目した大原らの分類方法 [9] を参考にし、次の 2 つに分類した。

- 「アハハハ」、「フフフ」というように声の繰り返し構造、強い震えが聞き取れる笑い
- 上記に分類されない笑いであり、息を吸う音、息を吐く音、鼻笑音で笑い声の大部分が構成されているような笑い

特に後者は大原らの研究で「引き笑い声」、「吸気音」、「鼻笑呼気音」とされているものであり、ここでは前者を obvious laughter、後者を breath laughter と呼ぶこととする。発生した被験者主導笑いを上述の 2 種と直後の同調的笑いの発生の有無により 4 種類に分類したところ、それぞれの発生回数は表 4 のようになった。

4. 予測実験

4.1 条件

表 5 にまとめた前述の分類に基づく特徴量を入力ベクトルとし、出力を St-S もしくは、Sh-S すなわち、直後

表 5: 使用した特徴量

極性	positive/neutral/negative
発話と笑いの関係性	speech/stand-alone
聞こえ方	obvious/breath

表 6: 実験結果

適合率	再現率	F 値	AUC
0.3176	0.6692	0.4250	0.4252

に同調的笑いが発生するか、発生しないかの 2 値とするロジスティック回帰による予測モデルを作成した。

4.2 結果

全 42 セッションの対話データ内の合計 276 回の被験者主導の笑い (St-S、Sh-S) を、訓練データが 8 割、残りの 2 割がテストデータとなるように無作為に分割し、ロジスティック回帰モデルでの学習、予測を行った。同調的笑いを発生させる Sh-S の適合率、再現率、F 値、Precision-Recall curve の Area under the curve (AUC) を計算した。1000 回の試行による平均値を表 6 に示す。結果として F 値がチャンスレート 0.286 (= 79/276) を上回っており、表 5 に挙げた特徴量が同調的笑いの発生の予測に有用であることがわかった。

5. おわりに

本稿では、自律型アンドロイドによる対話において同調的笑いを生成するため、対話中に発生するどのような人間の笑いにアンドロイドは同調して笑うべきか分析し、これに基づく予測モデルの構築を試みた結果、チャンスレートよりも高い精度を得ることができた。今後は、直前発話のパワーや基本周波数などの韻律的特徴、発話行為の種類、笑いの連鎖位置などの言語的特徴を検討し、精度の向上を目指す。

謝辞 本研究は、JST ERATO 石黒共生ヒューマンロボットインタラクティブプロジェクトの支援を受けて実施された。

参考文献

- [1] 山口貴史 et al. 傾聴対話システムのための言語情報と韻律情報に基づく多様な形態の相槌の生成. 人工知能学会論文誌, 31(4):C-G31-1, 2016.
- [2] 石田真也 et al. 傾聴対話システムのための発話を促す聞き手応答の生成. SIG-SLUD, 5(01):1-6, 2016.
- [3] 井上昂治 et al. 自律型アンドロイド Erica のための音声対話システム. SIG-SLUD, 75:21-24, 2015.
- [4] Gupta et al. Analysis and modeling of the role of laughter in motivational interviewing based psychotherapy conversations. INTERSPEECH, 2015.
- [5] Gupta et al. Laughter valence prediction in motivational interviewing based on lexical and acoustic cues. INTERSPEECH, 2016.
- [6] 小林のぞみ et al. 意見抽出のための評価表現の収集. 自然言語処理, 12(3):203-222, 2005.
- [7] 東山昌彦 et al. 述語の選択選好性に着目した名詞評価極性の獲得. 言語処理学会第 14 回年次大会論文集, 584-587, 2008.
- [8] Tian et al. When do we laugh? SIG-SLUD, 360, 2016.
- [9] 大原察. 対話音声の笑い声や笑い方についての分析. 奈良先端科学技術大学院大学修士論文, 2005.