

# 初対面対話における場の和みのマルチモーダルな分析と検出

稲熊 寛文<sup>1</sup> 井上 昂治<sup>2</sup> 中村 静<sup>2</sup> 高梨 克也<sup>2</sup> 河原 達也<sup>2</sup>

<sup>1</sup>京都大学 工学部 情報学科

<sup>2</sup>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

人間同士の初対面対話では、本題に入る前に簡単な雑談を挟むことで緊張感の漂う場を和ませるアイスブレイキングがしばしば見受けられる。対話システムに関するこれまでの研究では、特定のドメインにおいて、課されたタスクを遂行することに主眼が置かれてきたが、対話開始時において人とシステムとの間での場の和みに着目したものはほとんどない。本研究では、対話中の振る舞いとして表出するマルチモーダルな特徴量と場の和みに関して分析を行い、特徴量からアイスブレイキングの予測を試みる。アイスブレイキングの定義に関して、対話中の笑いに着目し、二話者間で笑いを共有する Shared Laughter が初めて観測された時点を目印としてアイスブレイキングの表出点とする。分析データとして、人型アンドロイド Erica [1] を遠隔操作して収録した1対1の初対面対話を用いる。分析結果に基づき、回帰分析によりアイスブレイキングの予測精度を評価する。提案するモデルによって、雑談のトピックを切り替えたり、人間らしいタイミングで Shared Laughter を行ったりすることができると期待される。

## 2. 対話コーパス

人型アンドロイド Erica を用いて、1対1の初対面対話を収録した。被験者は15人で、1人ずつ部屋に入室し、別室の Erica 役1人によって遠隔操作されたアンドロイドと自由な対話を行ってもらった。対話は2段階から成り、はじめに生活や趣味等に関する雑談(フェーズ1)、その後、アンドロイド全般に関する本題(フェーズ2)が展開されるようにした。各対話とも10分程度であった。フェーズ1からフェーズ2へと移行するタイミングは、Erica 役が対話の流れに応じて判断した。ただし、フェーズ1は3分程度にするように指示した。音声は被験者が座る椅子の下に設置したワイヤレスマイクで、映像は被験者と Erica の両者が映るようにビデオカメラで、それぞれ収録した。対話の様子を図1に示す。

本研究で扱うマルチモーダル特徴量のアノテーションには、解析ツール ELAN(<https://tla.mpi.nl/tools/tla-tools/elan/>)を用いた。本稿では本題に入る前の対話中に表れる特徴量について考察するため、フェーズ1に対してのみアノテーションを行った。付与したラベルは、トピック、発話内容、笑いの3つである。トピックは、収録時に Erica 役の操作者にあらかじめ与えたトピックリストを参考に、挨拶・自己紹介(挨拶, 自己紹介), 大学生活(学業, サークル, 部活, アルバイト, 住まい, 自炊, 一人暮らし, 大学の志望理由), 経歴(出身地), 趣味(趣味), 特技(特技), 将来(将来)の6種類(サブトピックは14種類)に分けることができた。括弧中はサブトピックを示す。本稿での分析にはサブトピックを使用する。



図 1: 対話実験の様子 (左は被験者, 右は Erica)

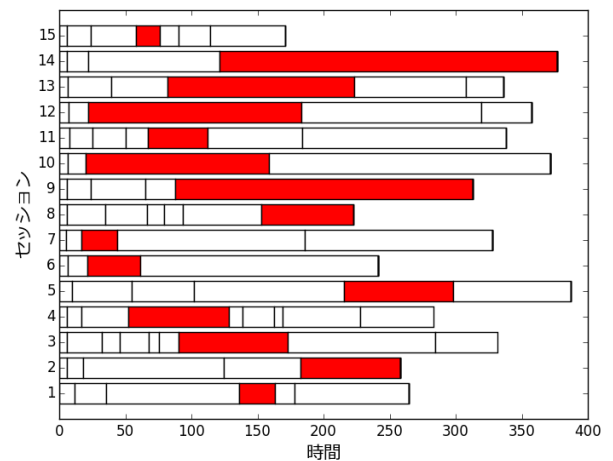


図 2: トピックの遷移 (赤はアイスブレイクトピック)

## 3. 特徴量分析

### 3.1 アイスブレイキングの定義

対話中にアイスブレイキングの発生を予測したとき、本題に入るためには進行中のトピックを自然に終了させる必要がある。これまで、二話者間で共有する笑いである Shared Laughter とトピックの終結には相関があり、トピックの終結が近づくにつれて話者単体の笑いである Solo Laughter よりも Shared Laughter の方が発生しやすいということが示されている [2] [3]。そこで本稿では、二話者間の対話中におけるアイスブレイキングの指標として Shared Laughter に着目し、Shared Laughter が初めて観測された時点を目印としてアイスブレイキングの表出点と定義する。ここで、15セッションにおけるトピックの遷移を図2に示す。棒グラフ内部の境界はフェーズ1におけるトピック境界であり、各ブロックの長さはトピック長を表す。赤いブロックがアイスブレイキングが発生したトピックである。これより、アイスブレイキングが発生した後のトピック長は長くなり、またアイスブレイキングが早い段階で発生してもすぐには本題に入らないことがわかる。これは、アイスブレイキングが発生すると対話が盛り上がり、対話が持続する傾向があるからと考えられる。

Multi-modal Analysis and Detection of Ice-breaking in First Meeting Dialogue: Hirofumi Inaguma, Koji Inoue, Shizuka Nakamura, Katsuya Takanashi, Tatsuya Kawahara (Kyoto University)

表 1: 各特微量のトピック間での  $t$  検定 ( $p$  値)

	特微量	First & Break	First & Last
発話	発話回数/分 (Erica 役)	0.21390	0.18980
	発話回数/分 (被験者)	0.20090	0.04446*
	平均発話時間 (Erica 役)	0.37312	0.08324
	平均発話時間 (被験者)	0.99821	0.48838
	相対頻度	0.84810	0.49841
オーバーラップ	オーバーラップ回数/分	0.27496	0.01869*
	平均オーバーラップ時間	0.00628**	0.10272
	オーバーラップ頻度	0.41440	0.02472*
沈黙	沈黙回数/分	0.53603	0.74752
	平均沈黙時間	0.02956*	0.69428

(\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ )表 2: 各特微量の 30 秒区切りでの  $t$  検定 ( $p$  値)

	特微量	First & Break	First & Last
発話	発話回数 (Erica 役)	0.20765	0.25226
	発話回数 (被験者)	0.00452**	0.00676**
	平均発話時間 (Erica 役)	0.00008***	0.00583**
	平均発話時間 (被験者)	0.00887**	0.04177*
	相対頻度	0.00225**	0.00169**
オーバーラップ	オーバーラップ回数	0.30432	0.49685
	平均オーバーラップ時間	0.86638	0.10179
	オーバーラップ頻度	0.16577	0.58070
沈黙	沈黙回数	0.00408**	0.34101
	平均沈黙時間	0.88514	0.09098

(\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ )

### 3.2 分析方法

対話が盛り上がる状況では、二者間の発話量、オーバーラップ、沈黙時間、F0 などの韻律的特徴の同調傾向が観測されている [4]。また F0、パワー、話速の同調傾向が強い対話は盛り上がった印象を与えることも示されている [5]。以上より韻律的特徴と場の和みとの間にも相関があることが推測され、韻律的特徴がアイスブレイキングを予測する指標になることが期待される。したがって、本節では、発話 (回数・時間・相対頻度)、オーバーラップ (回数・時間・頻度)、沈黙 (回数・時間) のそれぞれについて分析を行う。発話に関しては、被験者と Erica 役の別々に分析する。

さらに笑い、沈黙、オーバーラップ、相槌などの特微量をまとめた Social Signals とトピックの変化との間にも関連性が示されている [6]。したがって、上記の各特微量について、トピック毎、一定時間毎に算出し、各対話の最初とアイスブレイキングが発生した区間 (First & Break)、最初と最後の区間 (First & Last) のそれぞれに関して両側  $t$  検定を行った。時間に関しては、各区間内の平均値を使用する。

### 3.3 分析結果

トピック間で  $t$  検定を行った結果を表 1 に示す。回数に関しては、トピック長で正規化した。結果より、オーバーラップの特微量に有意差が見られた。また、発話の特微量では対話が進むにつれて  $p$  値が減少していくのに対し、沈黙の特微量では増加している。

次に、30 秒区切りで  $t$  検定を行った。結果を表 2 に示す。ただし、最後の区間長が 30 秒の 8 割 (24 秒) 未満の場合、直前の区間を最後の区間とした。また、最後の区間の回数は区間長で正規化した。結果より、発話の特微量に有意差が見られた。トピック間での結果より 30 秒区切りでの結果の方が有意差が見られる項目が多いため、以下の評価実験では 30 秒区切りのデータでアイスブレイキングの予測を試みる。

## 4. 評価実験

### 4.1 実験目的・条件

3.3 節の  $t$  検定による分析結果より、韻律的特徴がアイスブレイキングを予測する特微量として有用であること

表 3: 回帰分析による予測の評価

特微量	重回帰 (絶対値誤差)	ロジスティック回帰		
		適合率	再現率	F 値
発話	0.289	0.72	0.95	0.82
オーバーラップ	0.295	0.64	0.55	0.59
沈黙	0.302	0.74	0.29	0.42
発話、オーバーラップ	0.293	0.68	1.00	0.81
発話、沈黙	0.297	0.64	0.69	0.66
オーバーラップ、沈黙	0.301	0.74	0.32	0.45
発話、オーバーラップ、沈黙	0.301	0.68	0.86	0.76

が示唆された。そこで、重回帰分析とロジスティック回帰分析によりアイスブレイキングの予測精度を評価する。

重回帰分析では、入力は 3 節で分析した 11 個の特微量による、各対話の 30 秒毎の 11 次元の特微量ベクトル、出力は場の和み度である。ただし、場の和み度は最初の区間を 0、アイスブレイキングが発生した区間とそれ以降の区間を 1 とし、線形増加すると仮定する。評価尺度は、出力された場の和み度と正解値との絶対値誤差とする。

ロジスティック回帰分析では、アイスブレイキングが発生する前の区間と、後の区間とで 2 値分類し、適合率と再現率、F 値を算出した。

どちらも 1 セッション毎にテストデータとする交差検定によって精度を評価した。

### 4.2 実験結果

実験結果を表 3 に示す。発話の特微量のみを使用した場合の、重回帰分析による平均絶対値誤差は 0.289 で、ロジスティック回帰分析による適合率は 0.72、再現率は 0.95、F 値は 0.82 であった。どちらの回帰分析においても、発話の特微量のみを使用した場合に最も予測精度が良く、オーバーラップと沈黙の特微量を加えても精度の改善は見られなかった。以上の結果は、表 2 の結果と整合している。

## 5. おわりに

本稿では、初対面対話においてアイスブレイキングを予測するため、対話中に表出する韻律的特徴とアイスブレイキングに関して分析した。結果より韻律的特徴がアイスブレイキングの発生、対話の進行によって有意に変化し、予測実験では不均衡な分類タスクであるにも関わらず、ある程度のアイスブレイキングの予測ができることを確認した。今後は分析する特微量に F0、パワー、話速などの韻律的特徴や、視線、姿勢などを加え、データ数を増やす予定である。

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ションプロジェクトの支援を受けて実施された。

### 参考文献

- [1] 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. 人工知能学会研究会資料 SIG-SLUD, B502(02):21–24, 2015.
- [2] Elizabeth Holt. The last laugh: Shared laughter and topic termination. *Journal of Pragmatics*, 42(6):1513–1525, 2010.
- [3] Francesca Bonin *et al.* Laughter and topic changes: Temporal distribution and information flow. *Cognitive Infocommunications (CogInfoCom)*, 2012 IEEE 3rd International Conference on, 53–58. IEEE, 2012.
- [4] 綿貴啓子 *et al.* 対話における感情の変化の解析. 情報処理学会研究報告. SLP, 音声言語情報処理, 95(73):79–84, 1995.
- [5] 西村良太 *et al.* 音声対話における韻律変化をもたらず要因分析. 音声研究, 13(3):66–84, 2009.
- [6] Francesca Bonin *et al.* The discourse value of social signals at topic change moments. *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.