

高齢者対話における音響的および韻律的エントレインメントの分析

杉本 浩之[†] 田中 宏季[†] 中村 哲[†]

[†] 奈良先端科学技術大学院大学先端科学技術研究科 〒630-0192 奈良県生駒市高山町 8916 番地の 5
E-mail: †{sugimoto.hiroyuki.se8,hiroki-tan,s-nakamura}@is.naist.jp

あらまし 会話中のエントレインメント (会話相手に対する話し方の適応) は, 言語的および非言語的特徴において生じる. 加えて, エントレインメントはコミュニケーションの良し悪しと関連する. 本研究では, エントレインメントの詳細な分析を目的とし, 音響的特徴量 (基本周波数とラウドネス) および韻律的特徴量 (発話速度) に焦点を絞り, 高齢者対話における臨床心理士, 介護士, 大学院生の発話におけるエントレインメントについて, 種々の条件 (対話行為タグ, 割り込みの有無, 会話中の発話回数) において, 探索的に分析を行った. 結果, 割り込みの有無について, ラウドネスの最大値におけるエントレインメントの違いを統計検定 ($p < 0.05$) により, 臨床心理士, 介護士, 大学院生の全ての場合で確認した. 加えて, その効果量の大きさについて, 介護士が最も大きく, 続いて臨床心理士, 大学院生の順となっていた. 一方, 対話行為タグや会話中の発話回数をを用いた分析では, 主に介護士の発話においてのみ, エントレインメントの違いを統計検定 ($p < 0.05$) により確認した.

キーワード エントレインメント, 発話音声, 音響的特徴量, 韻律的特徴量

Acoustic and Prosodic Entrainment Analysis in Conversation With Elderly People

Hiroyuki SUGIMOTO[†], Hiroki TANAKA[†], and Satoshi NAKAMURA[†]

[†] Graduate School of Science and Technology, Nara Institute of Science and Technology Takayama 8916-5,
Ikoma-shi, 630-0192 Japan

E-mail: †{sugimoto.hiroyuki.se8,hiroki-tan,s-nakamura}@is.naist.jp

1. ま え が き

会話におけるエントレインメントとは, 会話相手に対する話し方の「適応」である. 会話という行為は, 何を言うのかだけでなく, どのように言うのか, どのような顔の表情や身振り手振りを加えるのかといった, さまざまな要素から成り立っている. それと同様に, エントレインメントにおいても, 話す文章の構文構造 [1] や文章の中で用いられる語彙 [2] といった言語的要素や, 会話中の顔や頭の動き [3] といった非言語的要素, そして, 声の大きさや基本周波数, 話す速度といった音響的もしくは韻律的要素 [4] [5] においても生じることが報告されている. また, 人間同士の会話中において生じるエントレインメントとその会話の良し悪しについて, その関連性が報告されている. Lubold ら [6] は, 音響的および韻律的エントレインメントとその会話により達成されたラポール (協力関係) スコアに関連があることを報告しており, Borrie ら [7] は, 2 人で協力しながらあるタスクに取り組む際に行われた会話における音響的および韻律的エン

トレインメントが強いほど, 成し遂げたタスクの成功度も高くなることを報告している. 加えて, エントレインメントはロボットとそのユーザーにおけるインタラクションにおいても重要であるとされている. Westlund ら [8] は, 会話相手の声に合わせて自身の声を調整するロボットは, そのユーザーに対して, 笑顔や喜びといった顔の表情を表す機会を増加させ, 恐れや失望といった顔の表情を表す機会を減少させることができたことを報告している. また, Metcalf ら [9] は, 会話相手の話す分量の度合いに応じて話す分量の度合いを変化させるロボットを用いることで, ロボットに対するユーザーの好ましさ, および信頼度を向上させることができたことを報告している.

エントレインメントの分析の際に取り扱う会話の場面設定においては, あるタスクを協力して取り組む際の会話を分析対象としている研究 [2] もあれば, 日常的な会話 [10], 薬物中毒者のためのカウンセリングにおける会話 [11], 裁判所における口頭弁論 [12] といった場面に着目している研究もある. 一方で, 会話の 1 側面に焦点を当てたエントレインメントの分析も存在し,

Ludusan ら [13] は、会話において生じる笑い (Laughter) に着目し、分析を行っている。加えて、Soto ら [14] は、発話する 1 つの文章の途中で用いる言語が変化するコードスイッチングとエントレインメントとの関連性を明らかにしている。

このように、エントレインメントは人間同士の会話およびロボットと人間とのインタラクションに重要であり、また、様々な場面におけるエントレインメントの分析が行われている。このエントレインメントを人間の会話の評価指標に利用することや、エントレインメントを対話システムに組み込むことを考えた際、どのような場面において、どのような特徴量において、どの程度のエントレインメントを生じさせる必要があるのかを分析する必要がある。一方で、人間の会話における要素の多次元性に加え、既存の研究において、各々の会話の場面設定やデータセット、および分析手法を利用しているため、分析結果が一致しない場合も存在する。加えて、Pérez ら [15] は、エントレインメントが生じない場合のほうが、会話相手に対して良い影響を与える場合があると報告している。そこで、より詳細な分析を行う必要があり、会話を話者同士の関係 [2] や会話の目的 [16]、もしくは会話中の各発話を発話行為タグ [17] や割り込みの観点 [18] で分類してからエントレインメントを分析することが有効であると考えられる。

そこで、本研究では、高齢者対話において、各発話を簡単な規則による割り込みの有無の分類および発話行為タグでの分類を行った上で音響的および韻律的エントレインメントの観点から違いが生じるのかを分析する。また、音響的および韻律的エントレインメントが発言する度にどのように変化するかを分析する。さらに、一般的に高いコミュニケーション能力が要求される臨床心理士や、高齢者に対して会話を行う機会が多い介護士、そして、比較対象としての大学院生における発話において、結果の一貫性の有無を調べることを目的とする。

2. 分析手法

2.1 データセット

本研究では、Yoshino ら [19] が収録したデータセットを用いる。このデータセットは互いに初対面である 2 者間の会話が収録されたものである。それぞれの会話において、一方の話者が高齢者であり、もう一方の話者が臨床心理士、介護士、大学院生のいずれかである。高齢者側は、あるニュースについて会話相手に話し、臨床心理士、介護士、大学院生側は対話相手である高齢者に対して積極的傾聴を行うように指示されている。参加者は高齢者 24 人、臨床心理士、介護士、大学院生はそれぞれ 5 人ずつである。高齢者と臨床心理士、介護士、大学院生との会話が各 20 セット、合計 60 セットの会話データが収録されており、各会話の継続時間が約 20 分から 30 分となっている。会話の様子は話者ごとに wav データで収録されており、また、会話中の各発話に対して、その発話の書き起こし、開始時刻と終了時刻、および ISO24617-2 [20] [21] 基準の対話行為ラベルが付与されている。本研究では、発話中に生じる相槌 (Back channel) については、相槌だけで成り立つ発話については除外し、1 つの発話の途中で生じる相槌についてはそのままにする。加えて、高齢者と臨床

心理士の会話 20 セットのうち、アノテーションされた各発話の開始時間と終了時間を用いて、会話全体の wav ファイルから発話ごとに適切に切り出すことができなかった 2 セットについては、今回の分析からは除外する。

2.2 音響的および韻律的特徴量

本研究では、音響的特徴量として基本周波数 (F0) とラウドネスを、韻律的特徴量として発話速度を分析対象の特徴量とする。

F0 およびラウドネスについては、会話内の各発話からそれぞれ抽出する。抽出には、オープンソースツールキットである openSMILE [22] を利用する。ある 1 つの会話における各発話に対して、フレームサイズ 0.02 秒、フレームシフト 0.01 秒でフレームごとに特徴量を抽出する。これを全発話分行い、その後、各話者ごとに、その発話全体からの平均値と標準偏差値を計算し、発話ごとにフレーム単位で算出した特徴量に対し標準化を行う。標準化を行った後、各発話ごとに、F0 の平均値と最大値、ラウドネスの平均値と最大値を計算し、これらを各発話に対する音響的特徴量とする。

一方、発話速度については、日本語の特性上、音素単位ではなく、モーラ単位で数えることとする [23]。1 つの発話内において、1 秒あたりいくつのモーラが発せられているのかを考える。各発話の書き起こしテキストからオープンソースライブラリである Mecab^(注1)を用いてカナ読みに変換したあと、そのモーラ数を数える。その後、モーラ数を発話時間で割り、その値を発話速度とする。

2.3 分析のためのフレームワーク

本研究は Levitan ら [4] の提案する手法を参考にし、ある話者の現在の発話が直前の会話相手の発話の影響を受ける、言い換えると、異なる話者からの隣接する 2 つの発話の音響的または韻律的な特徴量が似るか否かを分析する。仮に、自身と会話相手との発話間における特徴量が似れば、その発話間にエントレインメントが生じていると考えられる。また、エントレインメントの分析において、時間情報は重要であるため、時間情報を考慮しない発話単位の分析と、時間情報を考慮する会話セッション単位の分析と 2 つに分け、ノンパラメトリックの統計検定手法を用いて分析する。加えて、有意性が認められる結果が得られた場合、その効果量 (r) についても算出する。

2.3.1 音響的もしくは韻律的特徴量における距離

2 つの発話間の特徴量の類似性を差の絶対値により計算し、その値をエントレインメントの分析のための特徴量とする。本研究では、臨床心理士、介護士、大学院生の発話が分析対象となるが、1 つの発話ターンにおいて、複数の発話がなされる場合も存在する。今回はその中でさらに、発話ターンが会話相手である高齢者から臨床心理士、介護士、大学院生へと移った際の最初の発話を本研究の分析対象の発話とする。対象となった各発話と高齢者の発話の中で最も近接している発話とのペアを考える。本研究で対象となった発話は、臨床心理士について 913 発話、介護士について 1367 発話、大学院生について 1038 発話となって

(注1) : Mecab: Yet Another Part-of-Speech and Morphological Analyzer, <https://taku910.github.io/mecab/>

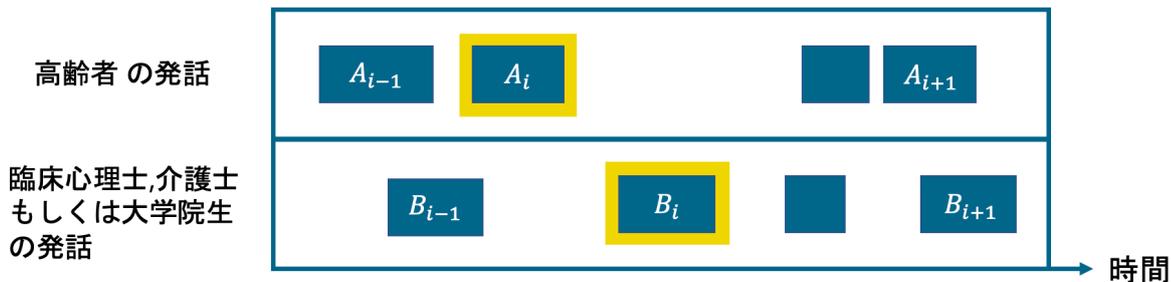


図1 図中の青色の長方形は発話を表しており、その中で B_j は会話中の発話の中で対象となる、臨床心理士、介護士もしくは大学院生の j 番目の発話を示す。一方で、 A_j は B_j に対してペアとなる発話を表している。例えば、特徴量距離を考える際、黄色で色付けされた2つの発話 A_i と B_i が i 番目ペアとなる。また、図において A や B が記されていない発話は今回の研究では考慮しない発話である。

いる。図1は例として、1つの会話における i 番目のペアを表した概略図である。そして、お互いにペアとなった発話から抽出した音響的および韻律的特徴量に対し、その差の絶対値を考える。この音響的もしくは韻律的特徴量における距離（以下、特徴量距離）を用いて分析を行う。

2.3.2 発話単位の分析

発話単位の分析においては、割り込みの有無を用いる分析、および ISO24617-2 基準による発話行為タグを用いた分析を行う。

割り込みの有無について、特徴量距離を考えた発話のペアそれぞれにおいて、臨床心理士、介護士もしくは大学院生の発話の開始時間とそれに対応する高齢者の発話の終了時間が被っているか否かにより分類する。そして、割り込みの有る発話と割り込みの無い発話との間で特徴量距離に差が生じているのかを調べるため、マン・ホイットニーの U 検定を行う。

ISO24617-2 基準の発話行為タグを用いた分析においては、General purpose function に分類されている、Information seeking (IS) と Information providing (IP) を利用する。IS および IP タグが付けられた発話数が十分であることに加え、情報を求める行為と情報を与える行為は相対する行為であると考えられることにより上記タグを採用する。そして、IS と IP がつけられた発話との間にエンタテインメントの違いが生じるのかをマン・ホイットニーの U 検定を用いて調べる。

なお、対象となった全発話において、割り込みの有無による発話数、および IS、IP とタグ付された発話数は表の通りである（表1、表2）。

表1 臨床心理士、介護士、大学院生の発話における割り込みの有無

	臨床心理士	介護士	大学院生
割り込み有りの発話数	544	787	588
割り込み無しの発話数	369	580	450

表2 臨床心理士、介護士、大学院生の発話における IS および IP タグ

	臨床心理士	介護士	大学院生
Information seeking (IS)	213	324	253
Information providing (IP)	291	508	447

2.3.3 会話セッション単位の分析

会話セッション単位の分析では、特徴量距離が会話が進むにつれてどのように変化が生じるのかを、2つの手法を用いて分析する。

今回用いるデータセットに収録されている会話はその継続時間が約20分から30分となっており均一ではない。そこで、本研究では会話の進みの程度を、会話の継続時間ではなく、特徴量距離を考える際に対象となった臨床心理士、介護士、大学院生の発話数により考える。

1つめの手法では、各会話において対象となった発話とその発話番号をもとに前半の発話と後半の発話に分ける。そして、前半の発話と後半の発話それぞれで特徴量距離の平均値を計算する。そして、前半の発話における特徴量距離の平均値と後半の発話における特徴量距離の平均値における違いをウィルコクソンの順位和検定により分析する。

2つめの手法では、特徴量距離の変化の度合いを傾きとして算出することを目的に、回帰分析を利用する。各発話において対象となった発話とその発話番号をもとに前から順に10分割し、それぞれの分割部を1, 2, ..., 10と番号付ける。その後、1から10それぞれの分割部において、特徴量距離の平均値を算出する。そして、分割部の番号である1から10を説明変数に、各分割部それぞれに対応する特徴量距離の平均値を目的変数として回帰分析を行う。なお、回帰分析を行う前に、説明変数と目的変数それぞれに対し標準化を行う。

3. 結果

3.1 発話単位におけるエンタテインメント

まず、割り込みの有無を利用した分析結果（ p 値および効果量）を示す（表3）。有意性が確認できた特徴量の中で、ラウドネスの最大値において、臨床心理士、介護士、大学院生全ての場合で一貫した結果が得られた。このラウドネスの最大値において、特徴量距離の値の違いを確認するための箱ひげ図を示す（図2）。ここでも、全ての場合において、割り込みがある場合の特徴量距離が割り込みが無い場合と比較して中央値が大きくなっているという、一貫した結果を得た。このことにより、会話相手の発話

に被せて発話をする際には、相手の発話のラウドネスの最大値と自分の発話のラウドネスの最大値の差が会話相手の発話に被さない場合と比較して大きくなることを示している。そしてこれは、会話相手の発話に被せて発話をする際には、エントレインメントの程度が弱くなることを意味する。また、F0の最大値について、介護士の発話においてその有意差が認められた。

次に、発話行為タグ IS および IP を利用した分析結果を示す(表 4)。有意性が見られたのは、介護士の発話におけるラウドネスの平均値、およびラウドネスの最大値のみであった。

表 3 割り込みの有無における検定結果 (p 値) と効果量

	臨床心理士	介護士	大学院生
ラウドネスの平均値	0.166	0.495	0.945
ラウドネスの最大値	0.010*** (0.086)	<0.001*** (0.155)	0.025*** (0.069)
F0 の平均値	0.153	0.934	0.299
F0 の最大値	0.687	<0.001*** (0.174)	0.469
発話速度	0.089† (0.056)	0.194	0.792

各セルにおいて、左の値が p 値、括弧内の値が効果量
有意水準は、***p<0.05, †p<0.10

表 4 IS と IP における検定結果 (p 値) と効果量

	臨床心理士	介護士	大学院生
ラウドネスの平均値	0.116	0.043*** (0.070)	0.321
ラウドネスの最大値	0.808	<0.001*** (0.154)	0.191
F0 の平均値	0.946	0.691	0.309
F0 の最大値	0.407	0.952	0.420
発話速度	0.398	0.376	0.149

各セルにおいて、左の値が p 値、括弧内の値が効果量
有意水準は、***p<0.05

3.2 会話セッション単位におけるエントレインメント

会話における、前半と後半の発話での特徴量距離の違いにおける分析結果を示す(表 5)。基本周波数の平均値と最大値や発話速度といった特徴量においては有意差が見られなかった。一方、ラウドネスの平均値においては、介護士において有意差が見られ、臨床心理士および大学院生についても、p<0.10 であり、有意傾向が見られた。続いて、ラウドネスの平均値について、前半の発話と後半の発話で特徴量距離の値がどうなっているのかを箱ひげ図を用いて示す(図 3)。箱ひげ図より、臨床心理士、介護士、大学院生のいずれの場合も、前半の発話よりも後半の発話における特徴量距離の中央値が小さくなっていることが分かる。このことより、ラウドネスの平均値において、そのエントレインメントの程度は会話中の全発話数における後半の発話のほうが大きくなるということを示す。一方で、ラウドネスの最大値については、介護士においてのみ有意性が認められた。

表 5 前半の発話と後半の発話の違いにおける検定結果 (p 値) と効果量

	臨床心理士	介護士	大学院生
ラウドネスの平均値	0.081† (0.294)	0.002*** (0.466)	0.083† (0.277)
ラウドネスの最大値	0.369	0.002*** (0.466)	0.231
F0 の平均値	1.000	0.596	0.189
F0 の最大値	0.702	0.143	0.927
発話速度	0.523	0.498	0.143

各セルにおいて、左の値が p 値、括弧内の値が効果量
有意水準は、***p<0.05, †p<0.10

続いて、ラウドネスの平均値および最大値に着目した回帰直線の傾きを示す(表 6)。ラウドネス関連の特徴量全ての場合において、傾きの値が負となったが、特に、発話数を前半と後半

で分けた際の分析において有意性もしくは有意傾向が認められた臨床心理士におけるラウドネス平均値、介護士におけるラウドネスの平均値および最大値、大学院生におけるラウドネスの平均値における傾きの絶対値の大きさは大きくなっている。さらに、介護士における傾きの絶対値の大きさが臨床心理士や大学院生と比較して大きい。これは、ラウドネスの平均値および最大値について、発話が増加していくにつれて変化するエントレインメントの度合いが臨床心理士や大学院生と比較して大きいことを示す。

表 6 回帰分析をした際の直線の傾き

	臨床心理士	介護士	大学院生
ラウドネスの平均値	-0.168	-0.319	-0.107
ラウドネスの最大値	-0.036	-0.210	-0.083

4. 考 察

本研究では、高齢者対話における臨床心理士、介護士および大学院生の発話における音響的および韻律的エントレインメントに着目し、エントレインメントがどのように生じているか、臨床心理士、介護士、大学院生の発話において一貫した結果を得ることができるのか、それとも結果に違いが見られるのかについて分析を行った。

発話単位の分析において、割り込みの有無および、IS と IP の発話行為タグを用いた。割り込みの有無を用いた分析においては、割り込みが有る発話において、ラウドネスの最大値におけるエントレインメントの程度が、割り込みが無い発話と比較して小さくなるという結果が得られた。会話において、会話相手に自身の声を届けることが不可欠である。その際、相手が話している最中に発話をする場合、声を相手に届けるためには声の大きさを大きくする必要がある。今回の分析結果はその現象について示していると考えられる。これをエントレインメントの観点から考えると、相手の発話と被せて発話をする場合、エントレインメントの程度が弱くなる、もしくはエントレインメントが生じないということになる。

IS と IP タグが付けられたそれぞれの発話間のエントレインメントの違いにおいては、介護士の発話におけるラウドネスの平均値および最大値においてのみ有意差が認められた。ここで、IS と IP における特徴量距離の中央値を見ると、ラウドネスの平均値については、IS の場合は 0.289、IP の場合は 0.338、ラウドネスの最大値について、IS の場合は 1.362、IP の場合は 1.829 となっていた。つまり、情報を求めるための発話よりも情報を与えるための発話のほうがラウドネスの平均値および最大値の距離が大きくなっている。このことは、情報を与えるための発話の際には、ラウドネスにおいてエントレインメントが弱くなることを意味する。しかしながら、この現象が一般的であるのか、今回の状況設定に限ったものなのかについては更なる分析が必要である。

一方、用いる発話行為タグの選択の際には、そのサンプル数が十分であることが必要であった。ISO24617-2 における発話行為

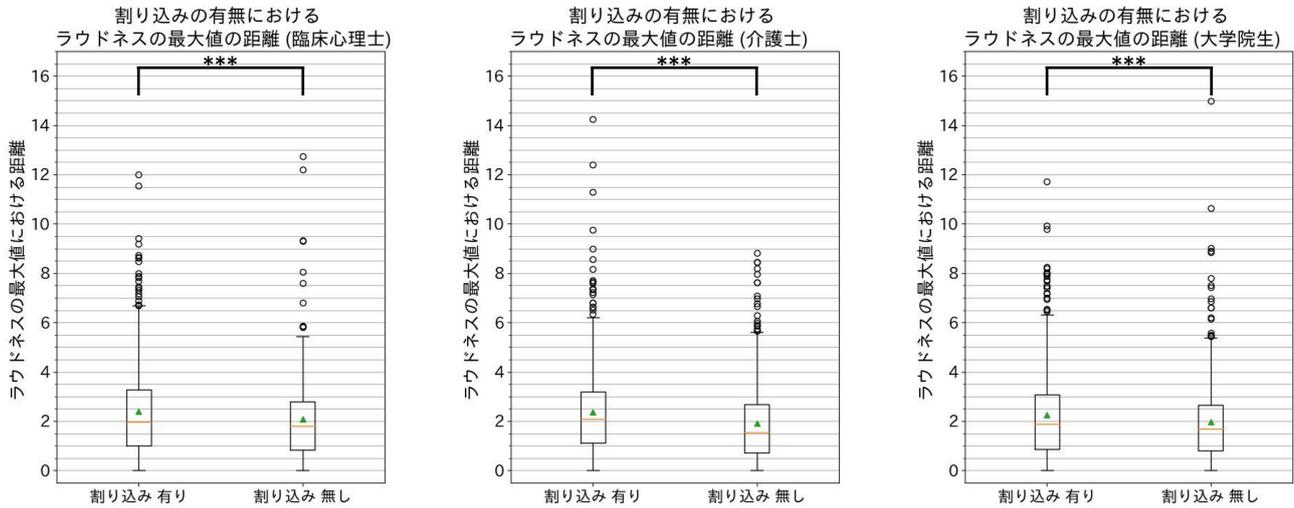


図 2 割り込みの有無で分けた、各発話のラウドネスの最大値における距離
有意水準は、*** $p < 0.05$

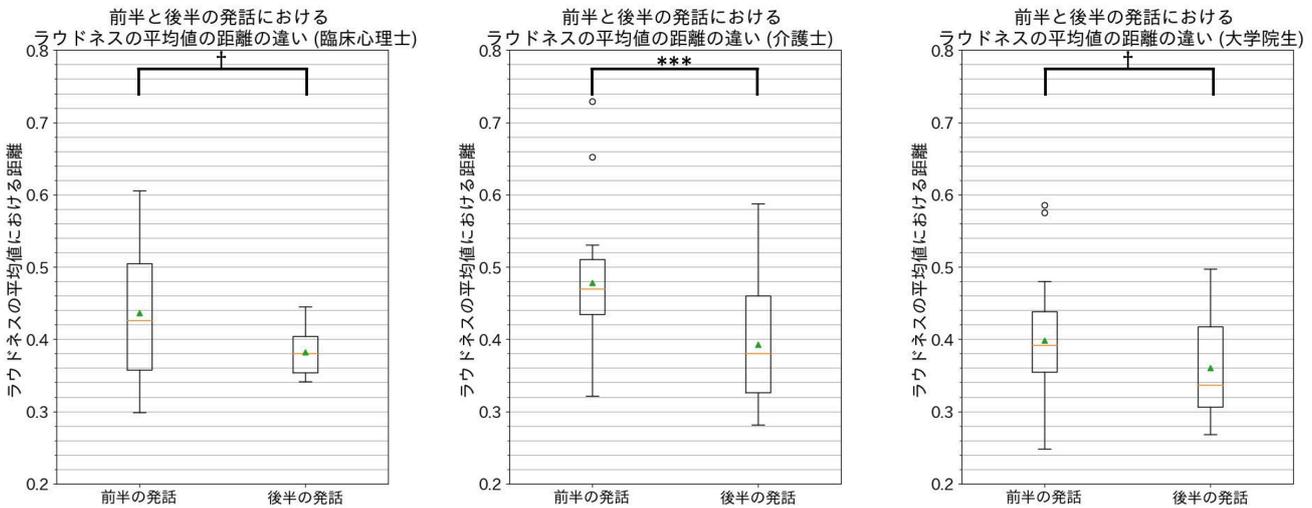


図 3 前半と後半の発話に対する、ラウドネスの平均値における距離
有意水準は、*** $p < 0.05$, † $p < 0.10$

タグには Directive という、会話相手に指示を与える発話に付けられるタグもあり、音響的および韻律的エンタテインメントの分析に役立つものであると期待されるが、本研究では、サンプル数が十分ではないと判断し、利用しなかった。

続いて、会話セッション単位の分析結果を考察する。会話の始めから「聞く」という役割について認識している臨床心理士、介護士、大学院生の発話におけるエンタテインメントを分析すると、発話するにつれて、ラウドネスの平均値において、そのエンタテインメントの度合いが大きくなっていく、もしくはその傾向があることが分かった。聞くことを意識する以上、エンタテインメントは発話するにつれても変化はしない、言い換えると会話の最初から望ましい程度のエンタテインメントが達成されていると仮定していたが結果は異なったものとなった。これが意味することは、望ましいエンタテインメントを達成するためにはある程度の時間がかかる可能性があること、もしくは、エンタ

テインメントには無意識的に生じる側面がある可能性があるということである。今回用いたデータセットにおいては、会話相手同士がお互いに初対面であるため、1つめの可能性を示すためには、初対面同士の会話と知り合い同士の会話とを比較する必要がある。

最後に、今回の発話単位の分析および会話セッション単位の分析における統計的検定において、有意性が見られた特徴量における効果量、または、回帰係数の絶対値から、ラウドネス関連の特徴量において、介護士の発話におけるエンタテインメントの程度の大きさが臨床心理士や大学院生と比較して大きいことが分かった。エンタテインメントが生じるか否かやその程度は、会話の場面設定や会話相手による影響を受けると考えられるが、今回取り扱った会話は高齢者との会話であり、普段から高齢者と会話する機会が多い介護士は、高齢者との会話に慣れているため、エンタテインメントの程度も大きいということが推測で

きる。また、エンタテインメントを分析する上で、それぞれの話者が普段どのような相手と会話することが多いかといった個人的な情報も考慮に入れる必要があることも示唆している。

5. ま と め

本研究では、高齢者対話における臨床心理士、介護士、大学院生の発話における音響的および韻律的エンタテインメントを種々の条件を用いて詳細に分析した。発話単位の分析では、会話相手に対する割り込み発言について、割り込みをしない発話と比較したところ、ラウドネスの最大値におけるエンタテインメントの程度が小さくなることを臨床心理士、介護士、大学院生それぞれの場合で確認した。一方、IS および IP の対話行為タグがつけられた発話におけるエンタテインメントの違いに関して、ラウドネスの平均値および最大値について、介護士の発話にのみ有意性が認められた。加えて、会話セッション単位の分析では、ラウドネスの平均値において、発話を重ねるごとにエンタテインメントの程度が大きくなることを、介護士の発話においては有意に、臨床心理士と大学院生の発話においても有意傾向を示すことを確認した。また、回帰分析では、ラウドネスの平均値及び最大値いずれの場合においても、介護士の発話における傾きの絶対値の値は大きいことが分かった。本研究により、エンタテインメントの分析について、割り込みの有無や発話行為タグといった発話の分類条件は詳細な分析を可能にすることに加え、臨床心理士や介護士、大学院生といった違いがエンタテインメントの分析の結果に影響することが分かった。エンタテインメントを詳細に分析するためには、従来の言語的及び非言語的特徴量に加え、発話ごとの分類、会話相手を含めた会話場面の設定や職業といった話者個人の情報などの利用が有効であると考えられる。

文 献

- [1] D. Reitter, Johanna D. Moore, and Frank Keller. Priming of syntactic rules in task-oriented dialogue and spontaneous conversation. 2006.
- [2] Alla Menshikova, Daniil Kocharov, and Tatiana V. Kachkovskaia. Lexical entrainment and intra-speaker variability in cooperative dialogues. In *Interspeech*, 2021.
- [3] Max M. Louwerse, Rick Dale, Ellen Gurman Bard, and Patrick Jeuniaux. Behavior matching in multimodal communication is synchronized. *Cognitive science*, Vol. 36 8, pp. 1404–26, 2012.
- [4] Rivka Levitan and Julia Hirschberg. Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. In *Interspeech*, 2011.
- [5] Alla Menshikova, Daniil Kocharov, and Tatiana V. Kachkovskaia. Phonetic entrainment in cooperative dialogues: A case of russian. In *Interspeech*, 2020.
- [6] Nichola Lubold and Heather Pon-Barry. Acoustic-prosodic entrainment and rapport in collaborative learning dialogues. *Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge*, 2014.
- [7] Stephanie A. Borrie, Nichola Lubold, and Heather Pon-Barry. Disordered speech disrupts conversational entrainment: a study of acoustic-prosodic entrainment and communicative success in populations with communication challenges. *Frontiers in Psychology*, Vol. 6, , 2015.
- [8] Jacqueline Kory Westlund and Cynthia Lynn Breazeal. Exploring the effects of a social robot’s speech entrainment and backstory on young children’s emotion, rapport, relationship, and learning. *Frontiers in Robotics and AI*, Vol. 6, , 2019.
- [9] Katherine Metcalf, Barry-John Theobald, Garrett Weinberg, Robert Lee, I-M. Jonsson, Russell Webb, and Nicholas Apostoloff. Mirroring to build trust in digital assistants. In *Interspeech*, 2019.
- [10] Antje Schweitzer and Natalie Lewandowski. Convergence of articulation rate in spontaneous speech. In *Interspeech*, 2013.
- [11] Bo Xiao, Zac E. Imel, David C. Atkins, Panayiotis G. Georgiou, and Shrikanth S. Narayanan. Analyzing speech rate entrainment and its relation to therapist empathy in drug addiction counseling. In *Interspeech*, 2015.
- [12] Stefan Benus, Agustín Gravano, Rivka Levitan, Sarah Ita Levitan, Laura Willson, and Julia Hirschberg. Entrainment, dominance and alliance in supreme court hearings. *Knowl. Based Syst.*, Vol. 71, pp. 3–14, 2014.
- [13] Bogdan Ludusan, Marin Schröer, and Petra Wagner. Investigating phonetic convergence of laughter in conversation. In *Interspeech*, 2022.
- [14] Víctor Soto, Nishmar Cestero, and Julia Hirschberg. The role of cognate words, pos tags and entrainment in code-switching. In *Interspeech*, 2018.
- [15] Juan Manuel Pérez, Ramiro H. Gálvez, and Agustín Gravano. Disentrainment may be a positive thing: A novel measure of unsigned acoustic-prosodic synchrony, and its relation to speaker engagement. In *Interspeech*, 2016.
- [16] Andreas Weise and Rivka Levitan. Investigating the influence of personality on acoustic-prosodic entrainment. In *Interspeech*, 2022.
- [17] Masahiro Mizukami, Koichiro Yoshino, Graham Neubig, David R. Traum, and Satoshi Nakamura. Analyzing the effect of entrainment on dialogue acts. In *SIGDIAL Conference*, 2016.
- [18] Rivka Levitan, Stefan Benus, Agustín Gravano, and Julia Hirschberg. Entrainment and turn-taking in human-human dialogue. In *AAAI Spring Symposia*, 2015.
- [19] Koichiro Yoshino, Hiroki Tanaka, Kyoshiro Sugiyama, Makoto Kondo, and Satoshi Nakamura. Japanese dialogue corpus of information navigation and attentive listening annotated with extended iso-24617-2 dialogue act tags. In *International Conference on Language Resources and Evaluation*, 2018.
- [20] Harry Bunt, Jan Alexandersson, Jean Carletta, Jae-Woong Choe, Alex Chengyu Fang, Kôiti Hasida, Kiyong Lee, Volha Petukhova, Andrei Popescu-Belis, Laurent Romary, Claudia Soria, and David R. Traum. Towards an iso standard for dialogue act annotation. In *International Conference on Language Resources and Evaluation*, 2010.
- [21] Harry Bunt, Jan Alexandersson, Jae-Woong Choe, Alex Chengyu Fang, Kôiti Hasida, Volha Petukhova, Andrei Popescu-Belis, and David R. Traum. Iso 24617-2: A semantically-based standard for dialogue annotation. In *International Conference on Language Resources and Evaluation*, 2012.
- [22] Florian Eyben, Martin Wöllmer, and Björn Schuller. Opensmile: the munich versatile and fast open-source audio feature extractor. *Proceedings of the 18th ACM international conference on Multimedia*, 2010.
- [23] Keiko Ochi, Nobutaka Ono, Keiho Owada, Masaki Kojima, Miho Kuroda, Shigeki Sagayama, and Hidenori Yamasue. Quantification of speech and synchrony in the conversation of adults with autism spectrum disorder. *PLoS ONE*, Vol. 14, , 2019.