

多様な相槌をうつ傾聴対話システムのための相槌形態の予測

Prediction of Morphological Forms of Backchannels for Flexible Attentive Listening Agents

山口 貴史^{1*} 井上 昂治¹ 吉野 幸一郎² 高梨 克也¹

Nigel G. Ward^{3,4} 河原 達也^{1,4}

Takashi Yamaguchi¹ Koji Inoue¹ Koichiro Yoshino² Katsuya Takanashi¹

Nigel G. Ward^{3,4} Tatsuya Kawahara^{1,4}

¹ 京都大学 情報学研究科

¹ Graduate School of Informatics, Kyoto University

² 奈良先端科学技術大学院大学 情報科学研究科

² Graduate School of Information Science, Nara Institute of Science and Technology

³ テキサス大学エルパソ校 計算機科学部

³ Department of Computer Science, University of Texas at El Paso

⁴ 京都大学 学術情報メディアセンター

⁴ Academic Center for Computing and Media Studies, Kyoto University

Abstract: We design an attentive listening agent which can generate flexible backchannels. In our previous work, we analyzed the morphological forms (category) of backchannels by focusing on their relationship with the syntactic structure in the preceding utterances. In this paper, based on the analysis, we conduct machine learning to predict the category of backchannels using features of the preceding utterance. At first, we annotated all acceptable backchannel categories for every backchannel occurrence and regard them as a “correct” label for the reference in evaluating prediction of the backchannel category. This annotation also gives a good insight on the relationship between backchannel forms. The results of the prediction suggest that we can choose appropriate backchannels depending on the preceding utterance. The proposed model improved prediction accuracy in comparison with the baseline which always outputs the most frequent morphological form of backchannels. Furthermore, evaluations by human subjects show that our method obtained a significantly higher rating than the baseline method.

1 はじめに

近年、タスク指向型対話システムに加えて、雑談型対話システムも検討されるようになってきている [1]。雑談型対話システムの機能の一つにユーザの話を聞く傾聴がある。傾聴とは話し手の話に共感を示しつつ、話し手がより多く話せるように手助けをして話を聴くことである [2]。音声対話システムが傾聴を行うことにより、入院患者や高齢者の話し相手となること [3] や、ユーザの話したい、話を聞いてもらいたいといった欲求を満たすこと [4] が期待されている。傾聴を行う際に重要となる対話行為としては、話し手の発話に対して、「相槌をうつ」「質問をする」「共感を示す」などが挙げられる。これらのうち、質問や共感では相手の発話を十分に認識・理解する必要があるが、これは技術的に容易ではない。これに対して、相槌は先行発話の韻律や節末のパターンに基づいて生成できる可能性がある。

このような点を踏まえて、我々是对話の文脈に応じて多様な相槌を生成する傾聴対話システムの研究を進めている [5]。先行研究 [5] において、傾聴対話の典型例であるカウンセリング対話を分析し、カウンセラのうった相槌形態と話し手の先行発話との関係の分析を行い、相槌形態と先行発話の節境界、構文構造に関係があることを示した。そこで本稿では、この知見を用いて、先行発話の情報から次にうつべき相槌形態を機械学習により予測できるか実験を行う。また、生成した相槌を含む音声を用いた印象評定実験により予測された相槌形態の適切さを被験者の印象評定によって評価する。

2 関連研究

まず、傾聴を行う音声対話システムの研究としては、下岡ら [6]、横山ら [7, 8, 9]、Ozkan ら [10]、大竹ら [11, 12] の研究が挙げられる。これらの傾聴システムでは相槌をうつ基本的な行為が実装されているが、人間同士

*連絡先：京都大学情報学研究科知能情報学専攻 河原研究室
京都市左京区吉田本町
E-mail: takashi@ar.media.kyoto-u.ac.jp

の対話と比較して、その形態や韻律のバリエーションは乏しい。

一方、相槌の生成に関する研究としては、先行発話の韻律情報を用いた研究 [13, 14] や、言語情報を用いた研究 [15, 16, 17, 18, 19, 20, 21] など、相槌の生成タイミングの予測を行うものが数多くある。しかし、これらの研究でもあらかじめ決められた形態の相槌のみが対象とされて、多様な形態の相槌をうつ枠組みは検討されていない。これに対して、本研究では、対話の文脈に応じて適切な形態の相槌をうつことができる傾聴対話システムの実現を目標として、先行発話の特徴量から相槌形態の予測を行うことを目標とする。

著者らの先行研究 [5] では、相談対話 [22] を対象として、話し手の発話の節境界 [23, 24] にうたれる相槌に焦点を当てて分析を行い、節境界では感情表出系相槌と応答系相槌の繰り返しがあたれやすいこと、「うん」や「ふん」はどこでもうてることがわかった。この結果に基づき、応答系の「うん」「ふん」を同種として扱い、その繰り返し回数によって、応答系 1 回、応答系 2 回、応答系 3 回以上というカテゴリに分類した。また、「あー」、「はー」、「へー」は応答系と振る舞いが違うため [25]、感情表出系として一つのカテゴリとした。先行発話の特徴からうつべき相槌を予測するために、先行発話の節境界の種類や構文構造とこれらの 4 つの相槌との関係について分析し、節境界の種類は感情表出系と応答系の区別に有効であり、構文構造は応答系 1 回と 2 回の区別に有効であるという知見を得た。

そこで、本研究では、これらを先行発話から得られる特徴量とし、機械学習により、先行発話に沿った相槌形態を予測することにした。

3 相槌入れ替えアノテーション

3.1 相槌形態予測の予備実験とその問題点

2 節の最後で述べた先行研究の分析 [5] に基づいて、機械学習により先行発話の特徴量のみから相槌カテゴリを予測する実験を予備的に行った。相槌をうつタイミングが与えられているものとし、先行発話から得られる素性情報から山口ら [5] で定義した相槌カテゴリを予測する。傾聴対話 8 対話の交差検定を行ったところ、どの対話でも F 値が 0.3 程度と、低い分類精度となった。

この予測実験では、カウンセラがうったのと同じ一つのカテゴリのみを正解とし、分類精度を測った。しかしながら、一般に、ある発話に対してうつことのできる相槌形態は一つのカテゴリのものだけに限られないと考えられる。たとえば、強い同意を示すためには「あー」だけでなく「うんうんうん」なども可能かもしれない。そこで、この問題に対処するために、カウンセラが相槌をうっている箇所において、カウンセラのものとは異なるが、許容される相槌カテゴリを複数の作業者によってアノテーションする作業を行った。これにより、ある発話に対してうつことのできる相槌カ



図 1: アノテーション用インターフェイス

テゴリとして、元のカウンセラがうった相槌カテゴリだけでなく、これと入れ替え可能な相槌カテゴリも利用できるようにする。

3.2 アノテーション方法

節末に相槌がうたれている箇所のみアノテーションを行った。アノテータは 3 名 (男性 1 名、女性 2 名) である。使用する対話音声は上里ら [22] の実験で使用された 4 対話を用いるが、本研究では節末にうたれた相槌のみに焦点を当てているため、節単位に区切った提示音声を使用する。この提示音声には相談者の発話のみが収録されており、カウンセラの発話が入っていない。

アノテーションには、図 1 に示すインターフェイスを使用した。このインターフェイスでは、再生される提示音声の発話内容が表示されており、相槌がうたれている節末は★マークで示されるが、元のカウンセラがうった相槌形態は提示していない。「再生」を押すことで提示音声を聞くことができ、「次の音声へ」を押すことで表示が次の提示音声へ移る。

アノテータは提示音声を聞いた上で、★の箇所で各相槌カテゴリについて、○ (うてる)、△ (うてるともうてないとも言えない)、× (うてない) の三段階の評価を行う。ただし、感情表出系のカテゴリでは、「あー」「はー」「へー」で振る舞いが違うため [26]、これらの形態を個別に評価する。

3.3 アノテーションの結果の分析

アノテータの一致率を Fleiss' κ 係数 [27] により求めた。全体の一致率は 0.099 (slight agreement) であったため、アノテータ 3 名とも「○: うてる」と評価した相槌カテゴリのみを、もとのカウンセラがうった相槌形態と「入れ替え可能」な相槌カテゴリと認定した。認

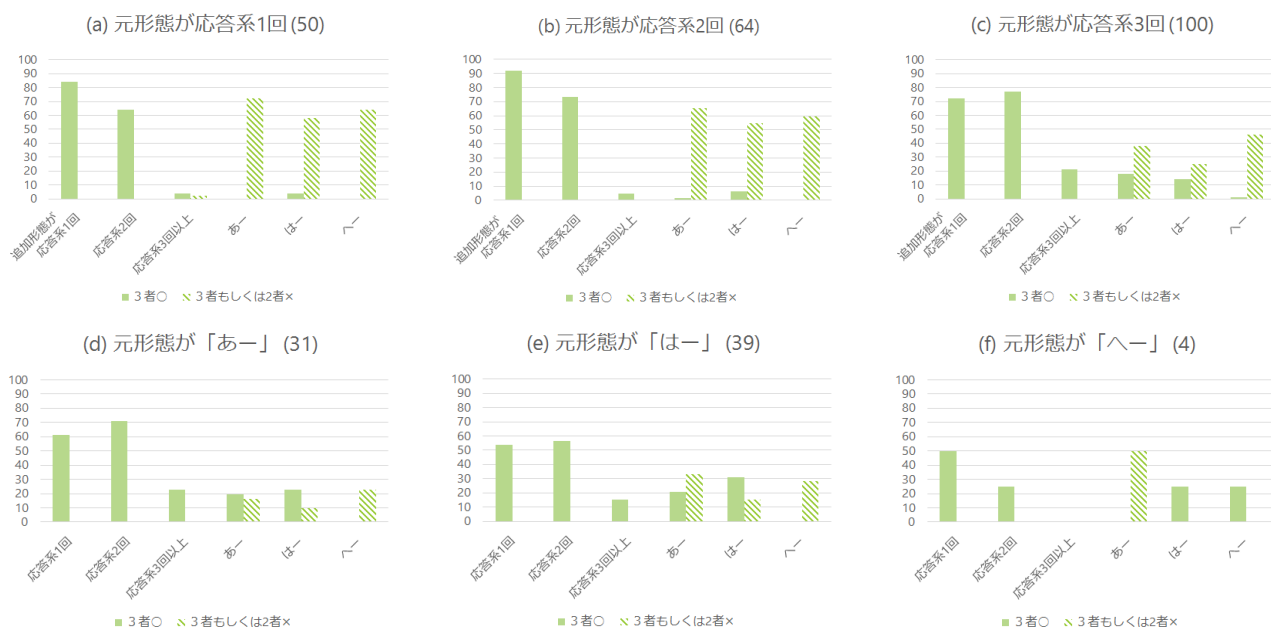


図 2: それぞれの元形態における各追加形態の許容傾向: 各形態ごとにおけるグラフの左は各追加形態の許容度 (3名ともが○)、右は非許容度 (3名もしくは2名が×) を表す。

定された相槌カテゴリは4節の予測実験の際に「正解」として扱う。

次に、元のカウンセラがうった相槌形態 (以下「元形態」とアノテーションされた相槌カテゴリ (「追加形態」) の関係について調べた。アノテータ3名とも「○: うてる」とした割合とアノテータ3名もしくは2名が「×: うてない」とした割合を各追加形態の間で比較し (それぞれ「許容度」と「非許容度」と呼ぶ)、許容できる追加形態の傾向を調べた。

図2に元形態 (a)~(f) 各追加形態の許容傾向を示す。各元形態の見出しの横の () 内の数字は生起度数を表す。各追加形態の左 (塗りつぶし) のグラフは許容度を表し、右 (斜線) のグラフは非許容度を表している。

まず、元形態が応答系1回 (a) と応答系2回 (b) では、それぞれ相互に許容度が高いため、互いに追加形態になりやすいことがわかる。また、応答系3回や感情表出系へは入れ替えが難しいこともわかる。(a) と (b) はグラフ全体の傾向が似ていることから、応答系1回と応答系2回は似た性質を持っていると考えられる。

一方、元形態が応答系3回 (c) では、(a) と (b) に比べて、応答系1回への許容度が若干下がっているが、応答系2回と応答系3回では上がっている。(c) は (a) や (b) と比べて、感情表出系への許容度が高くなっており、逆に非許容度は下がっている。そのため、応答系1回や応答系2回と比べて、応答系3回以上の相槌は感情表出系により近い性質を持っていると考えられる。

次に、元形態が感情表出系の「あー」(d)、「はー」(e)、「へー」(f) について述べる。まず、(d) と (e) では、応答系1回や応答系2回が他の追加形態と比べて許容されやすいことがわかる。また、追加形態が感情表出系の場合では、非許容度の大きさから、(d) と (e) どちらにおいても「あー」よりも「はー」の方が追加形態と

して好ましいことがわかる。最後に (f) では、サンプルが少ないものの、特に、追加形態に「あー」が許容されないことがわかる。これらのことから、感情表出系の「あー」「はー」「へー」の中では、「はー」が他の感情表出系の相槌との入れ替えが最も許容されやすい形態と考えられる。

本節では、ある発話に対してうつことのできる相槌形態を拡張することを目的とし、カウンセラが節末にうった相槌を他のカテゴリの相槌と入れ替えるアノテーションを行い、次の知見を得た。

- 応答系1回と応答系2回は似た性質をもち、相互に入れ替え可能な相槌である
- 応答系3回は感情表出系に近い性質を持つ
- 感情表出系では「はー」が他と感情表出系のもの比べてより柔軟に使える

次節では、本節のアノテーションによって入れ替え可能 (3名ともが○) と認定された相槌カテゴリを「正解」ラベルとみなして予測実験を行う。

4 先行発話からの相槌形態の予測

4.1 実験方法

先行発話の特徴量から機械学習により相槌カテゴリを予測する実験を行った。使用する特徴量を表1に示す。先行研究 [5] で相槌形態との関係について知見が得られている、先行発話の節境界の種類と文節数、構文木の深さ、係り受けの数に加え、節境界ラベルや末尾語とその品詞を用いた。また、一つ前の節境界の種類とそこでの相槌形態も特徴量に加えた。

表 1: 予測に用いる特徴量の種類

節境界の種類
末尾語
末尾語の品詞
節境界ラベル
文節数
構文木の深さ
係り受けの数
一つ前の節境界の種類
一つ前の相槌形態

表 2: 学習データ・テストデータごとの各相槌カテゴリの生起度数

		応答系 1 回	応答系 2 回	応答系 3 回以上	感情表出系	合計
対話 1	学習	135	89	234	183	641
	テスト	14	31	30	18	93
対話 2	学習	127	121	252	186	686
	テスト	19	3	24	21	67
対話 3	学習	150	97	226	176	649
	テスト	2	27	36	20	85
対話 4	学習	122	125	253	188	688
	テスト	15	3	10	15	43

相談対話 [22] 8 対話のうち、1 対話を評価用、残りの 7 対話を学習用とした交差検定を行った。ただし、評価データは第 3 節で追加形態をアノテーション付与した 4 対話とする。表 2 に学習データの内訳を示す。分類器には線形 SVM (Liblinear [28]) を使用した。

評価には、適合率と再現率、これらの調和平均である F 値を使用し、各対話ごと、各カテゴリごとで算出した。第 3 節で述べたように、今回の実験では評価の際、元形態だけでなく追加形態と認定された相槌カテゴリも正解としている。なお、比較のためのベースラインとしては、コーパスにおいて節末でうたれていた回数が最も多かった応答系 3 回をすべての予測箇所ですべての出力するチャンスレートを採用した。

4.2 結果と考察

各対話ごとの予測結果を表 3 に示す。提案モデルは全対話においてベースラインを上回る結果となった。次に、各カテゴリごとの結果を表 4 に示す。各相槌カテゴリにおいても、ベースラインを上回る結果となった。これらの結果から、先行研究の知見に基づく本予測モデルは有効であるといえる。

表 4 より、応答系 1 回と応答系 2 回は適合率が高く、予測のほとんどが正解になっている。この結果は、3.3 節で考察した追加形態の傾向で、多くの先行発話に対し応答系 1 回や応答系 2 回が許容されていた点と整合する。また、応答系 3 回以上は、生起頻度が最も多いため再現率が高いものの、適合率は低い。これは、応答系 3 回以上が追加形態となる箇所があまり多くなかったことによると考えられる。

次に感情表出系では、適合率が低いことから、不適

表 3: 対話ごとの予測結果 (適合率=再現率=F 値)

	提案モデル	ベースライン
対話 1	0.591 (55/93)	0.376 (35/93)
対話 2	0.582 (39/67)	0.448 (30/67)
対話 3	0.647 (55/85)	0.529 (45/85)
対話 4	0.651 (28/43)	0.233 (10/43)
平均	0.618	0.397

表 4: 相槌カテゴリごとの予測結果

相槌カテゴリ	適合率	再現率	F 値
応答系 1 回	0.917 (44/48)	0.420 (21/50)	0.576
応答系 2 回	0.879 (29/33)	0.375 (24/64)	0.526
応答系 3 回以上	0.444 (48/108)	0.780 (78/100)	0.566
感情表出系	0.566 (56/99)	0.730 (54/74)	0.637
合計	0.615 (177/288)	0.615 (177/288)	0.615
ベースライン	0.417 (120/288)	0.417 (120/288)	0.417

切な箇所感情表出系の相槌を予測していることがわかる。この点に関しては 5 節で考察する。

5 予測された相槌形態の印象評価

5.1 実験条件

予測された相槌形態を含む音声を用いた印象評定実験を行った。被験者は 20 代の男女 9 名 (男性 5 名、女性 4 名) である。各被験者は相槌が挿入された対話音声を聞き、相槌の印象を評価する。ここでは比較のため、相談対話の音声に対して、元の対話で相槌がうたれていた節末に、以下の三条件の相槌音声を挿入した提示刺激を作成した。ベースライン条件の相槌形態は前節の実験と同様、生起度数が最も多かった応答系 3 回とした。

1. ベースライン条件:すべての相槌形態が応答系 3 回
2. 予測条件:提案モデルにより出力された相槌形態
3. カウンセラ条件:もとの対話収録時にカウンセラがうったのと同じ相槌形態

相槌音声は、応答系として「うん」「うんうん」「うんうんうん」、感情表出系では「あー」「はー」を使用する。ただし、予測条件において、感情表出系をうつ場合には、3.3 節のアノテーションにおいて「あー」よりも許容性があるとされた「はー」を使用した。相槌音声は三条件とも同じ音声 (元のカウンセラの音声ではなく別の話者により発せられた) を使用した。使用した対話音声は前節で用いた 4 対話から 1 分半~2 分程度連続した箇所を 2 箇所ずつ抜粋した合計 8 サンプルである。抜粋箇所は、「話の内容がまとまっている」、「カウンセラが比較的話しかけたりせず相槌をうっている」、「カウンセラの相槌をうつ際の癖が出にくい」などを基準とした。実験に際し、抜粋した音声は対話の

表 5: 被験者による相槌の印象評価結果

評価項目	ベースライン条件	予測条件	カウンセラ条件
Q1: 全体を通して相槌は自然でしたか	-0.58	0.21	0.46
Q2: 全体を通してテンポよく進んでいましたか	0.00	0.42	0.79
Q3: 全体を通して聞いてくれていると感じましたか	0.04	0.42	0.54
Q4: 全体を通して理解してくれていると感じましたか	-0.63	0.08	0.21
Q5: 全体を通して共感してくれていると感じましたか	-0.75	-0.08	0.38
Q6: このカウンセラと話したいと思いましたか	-0.67	-0.08	0.46

途中から始まるため、被験者が話の内容を理解しやすくなるよう、音声聞いてもらう前に対話の概略を口頭で説明した。

8 サンプル× 3 条件で合計 24 のスロットがあるため、各スロットについて、3 名の異なる被験者から回答が得られるようにする。そのため、各被験者に 8 サンプルを提示する際には、スロット間で同じ被験者の組み合わせの重複が最小限になるようにした。また、サンプルを聞く順番も被験者ごとに異なっている。

評価項目は堀口 [29] の相槌の機能を参考に作成した。被験者はそれぞれのサンプルに対して、これらの評価項目について 5 段階 (-2: 全くそう思わない -1: そう思わない 0: どちらともいえない 1: そう思う 2: 非常にそう思う) で評定した。

5.2 評価結果

各被験者の評価値の平均を表 5 に示す。どの項目においてもカウンセラ条件、予測条件、ベースライン条件の順で評価値が高くなる傾向であった。応答系 3 回という決められた形態の相槌のみが挿入されているベースライン条件と提案法である予測条件との間に有意差があるかを t 検定 ($p < 0.05$) で検定したところ、項目 Q1、Q4~Q6 で有意差が認められた (表の太字)。

まず、項目 Q1 では、カウンセラ条件や予測条件と比べて、ベースライン条件の評価が大きく低いことから、同じ形態が続くと不自然な印象を与えることがわかる。また、項目 Q2 においても、形態が異なっている方がテンポよく会話が進んでいるという印象を与えることがわかる。

次に、相槌の機能に関する項目について、まず項目 Q3 では、ベースライン条件と予測条件で有意な差がない。このことから、聞いてくれているという印象を与えるだけなら、同じ形態の相槌の繰り返しでも十分であるといえる。しかし、項目 Q3、Q4、Q5 については、どの条件でも評価値が項目 Q3 → Q4 → Q5 と下がっていくという傾向が見られ、その下がり幅はベースライン条件で最も大きく、予測条件ではそれほどではない。特に、項目 Q4 と Q5 においては、ベースライン条件では予測条件やカウンセラ条件と比べて評価値が非常に低い。

一般に、聞いていないと理解できず、理解できていないと共感できないということからもわかるように、聞き手の理解には深さの度合いがあり、これに応じて聞き手はこうした理解の深さを反応の強さによって表す

必要があると考えられる。この観点から、ベースライン条件のように同じ形態の相槌の連続では、聞き手の深い理解や共感といった強い反応を表すのに不十分であることが推測される。

しかしその一方で、予測条件はカウンセラ条件と比べると低い評価であった。これは、予測条件では理解や共感を表す機能を持つ感情表出系が不適切な箇所でも出現していたことによるのではないかと考えられる。この点は前節の分類実験でも感情表出系の適合率が低かったことと一致する。逆に、カウンセラ条件が項目 Q4 (理解) よりも項目 Q5 (共感) で高い評価であるのは、カウンセラ条件ではより強い印象を与える感情表出系が適切な箇所でもうたれていたからではないかと考えられる。

最後に、項目 Q6 は傾聴システムの実装を想定した総合評価的な質問項目であるが、ベースライン条件の評価が特に低く、また予測条件もカウンセラ条件と比べて低い。この結果は理解 (Q4) や共感 (Q5) での評価値の傾向と類似していることから、これらの項目での評価が全体に影響しているのではないかと考えられる。

6 まとめ

本研究では、多様な形態の相槌をうつことのできる傾聴対話システムの実現を目標として、先行研究の分析結果に基づいて相槌形態の予測を行った。ある発話に対してうつことのできる相槌形態を拡張することを目的とした相槌形態のアノテーション結果の傾向から相槌形態の性質を考察することができた。また、アノテーション結果を踏まえた先行発話の特徴量に基づく相槌形態の予測では、あらかじめ決められた形態をうつベースラインと比較して、有効に相槌形態を予測することができた。さらに、音声による相槌の印象評価実験では、ベースラインと比較して、相槌の自然さや、理解や共感などの表現において有意に高い評価を得た。しかし、感情表出系の相槌をうつのに適切な箇所については、さらなる検討が必要である。

謝辞

アノテーションと印象評価実験にご協力いただいた皆様に感謝いたします。

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による。

参考文献

- [1] 河原達也. 音声対話システムの進化と淘汰: 歴史と最近の技術動向. 人工知能学会誌, Vol. 28, No. 1, pp. 45–51, 2013.
- [2] 楡木満生. 積極的傾聴法. 医学教育, Vol. 20, No. 5, pp. 341–346, 1989.
- [3] 山本大介, 小林優佳, 横山祥恵, 土井美和子. 高齢者対話インタフェース: 『話し相手』となって, お年寄りの生活を豊かに. 電子情報通信学会技術研究報告. HCS, Vol. 109, No. 224, pp. 47–51, 2009.
- [4] 目黒豊美, 東中竜一郎, 堂坂浩二, 南泰浩. 聞き役対話の分析および分析に基づいた対話制御部の構築. 情報処理学会論文誌, Vol. 53, No. 12, pp. 2787–2801, 2012.
- [5] 山口貴史, 井上昂治, 吉野幸一郎, 高梨克也, 河原達也. 傾聴対話における相槌形態と先行発話の統語構造の関係の分析. 人工知能学会研究会資料. SLUD, 言語・音声理解と対話処理研究会, Vol. 73, pp. 21–26, 2015.
- [6] 下岡和也, 徳久良子, 吉村貴克. 音声対話ロボットのための傾聴システムの開発. 人工知能学会研究会資料. SLUD, 言語・音声理解と対話処理研究会, Vol. 58, pp. 61–66, 2010.
- [7] 横山祥恵, 山本大介, 小林優佳, 土井美和子. 高齢者向け対話インタフェース: 雑談継続を目的とした話題提示・傾聴の切替式対話法. 情報処理学会研究報告. SLP, 音声言語情報処理, Vol. 2010, No. 4, pp. 1–6, 2010.
- [8] 小林優佳, 山本大介, 土井美和子. 高齢者対話インタフェース-ユーザの聴き手になる音声対話インタフェース. 人工知能学会研究会資料. SLUD, 言語・音声理解と対話処理研究会, Vol. 59, pp. 1–6, 2010.
- [9] 小林優佳, 山本大介, 土井美和子. 高齢者対話インタフェース: 発話間の共起性を利用した傾聴対話の基礎検討. 情報科学技術フォーラム講演論文集, Vol. 10, No. 2, pp. 253–256, 2011.
- [10] D. Ozkan and L.-P. Morency. Modeling wisdom of crowds using latent mixture of discriminative experts. In *Proc. ACL/HLT*, 2011.
- [11] 大竹裕也, 萩原将文. 高齢者のための発話意図を考慮した対話システム. 日本感性工学会論文誌, Vol. 11, No. 2, pp. 207–214, 2012.
- [12] 大竹裕也, 萩原将文. 評価表現による印象推定と傾聴型対話システムへの応用. 日本知能情報ファジィ学会論文誌, Vol. 26, No. 2, pp. 617–626, 2014.
- [13] 岡登洋平, 加藤佳司, 山本幹雄, 板橋秀一. 韻律情報を用いた相槌の挿入. 情報処理学会論文誌, Vol. 40, No. 2, pp. 469–478, 1999.
- [14] N. Ward and W. Tsukahara. Prosodic features which cue backchannel responses in English and Japanese. *J. Pragmatics*, Vol. 32, No. 8, pp. 1177–1207, 2000.
- [15] H. Koiso, Y. Horiuchi, S. Tutiya, A. Ichikawa, and Y. Den. An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese map task dialogs. *Language and Speech*, Vol. 41, No. 3, pp. 295–321, 1998.
- [16] 西村良太, 中川聖一. 応答タイミングを考慮した音声対話システムとその評価. 情報処理学会研究報告. SLP, 音声言語情報処理, Vol. 2009, No. 22, pp. 1–6, 2009.
- [17] 竹内真士, 北岡教英, 中川聖一. 韻律・表層的言語情報を発話タイミング制御に用いた雑談対話システム. 情報処理学会研究報告. SLP, 音声言語情報処理, Vol. 2004, No. 15, pp. 87–92, 2004.
- [18] N. Kitaoka, M. Takeuchi, R. Nishimura, and S. Nakagawa. Response timing detection using prosodic and linguistic information for human-friendly spoken dialog systems. *J. Japanese Society for Artificial Intelligence*, Vol. 20, pp. 220–228, 2005.
- [19] 小林一樹, 船越孝太郎, 小松孝徳, 山田誠二, 中野幹生. ASEに基づく相槌によるロボットとの対話体験の向上. 人工知能学会論文誌, Vol. 30, No. 4, pp. 604–612, 2015.
- [20] 大野誠寛, 神谷優貴, 松原茂樹. タグ付けの安定性を備えた音声対話コーパスに基づくあいづち生成タイミングの検出. 情報処理学会研究報告, 2010-SLP-84-4, 2010.
- [21] Y. Kamiya, T. Ohno, and S. Matsubara. Coherent backchannel feedback tagging of in-car spoken dialogue corpus. In *Proc. SIGdial*, pp. 205–208, 2010.
- [22] 上里美樹, 吉野幸一郎, 高梨克也, 河原達也. 傾聴対話における相槌の韻律的特徴の同調傾向の分析. 人工知能学会研究会資料. SLUD, 言語・音声理解と対話処理研究会, Vol. 70, pp. 7–13, 2014.
- [23] 日本語話し言葉コーパスの構築法. 独立行政法人国立国語研究所, 2006.
- [24] 高梨克也, 内元清貴, 丸山岳彦. 『日本語話し言葉コーパス』における節単位認定. 『日本語話し言葉コーパス』同梱マニュアル, 2004.
- [25] 常志强, 高梨克也, 河原達也. ポスター会話におけるあいづちの韻律的特徴に関する印象評定. 人工知能学会研究会資料. SLUD, 言語・音声理解と対話処理研究会, Vol. 56, pp. 31–36, 2009.
- [26] 常志强, 高梨克也, 河原達也. ポスター会話におけるあいづちの形態的・韻律的な特徴分析と会話モード間との相関の分析. 人工知能学会研究会資料. SLUD, 言語・音声理解と対話処理研究会, Vol. 54, pp. 7–13, 2008.
- [27] L. J. Fleiss. Measuring nominal scale agreement among many raters. *Psychological bulletin*, Vol. 76, No. 5, pp. 378–382, 1971.
- [28] R. E. Fan, K. E. Chang, C. H. Hsieh, X. R. Wang, and C. J. Lin. Liblinear: A library for large linear classification. *The Journal of Machine Learning Research*, Vol. 9, pp. 1871–1874, 2008.
- [29] 堀口純子. コミュニケーションにおける聞き手の言語行動. 日本語教育, No. 64, pp. p13–26, 1988.