原著論文 Original Paper

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

Generating a Variety of Backchannel Forms Based on Linguistic and Prosodic Features for Attentive Listening Agents

山口 貴史 Takashi Yamaguchi
京都大学情報学研究科 Graduate School of Informatics, Kyoto University
takashi@sap.ist.i.kyoto-u.ac.jp

井上 昂治 Koji Inoue
（同　上）
inoue@sap.ist.i.kyoto-u.ac.jp

吉野 幸一郎 Koichiro Yoshino
奈良先端科学技術大学院大学情報科学研究科 Graduate School of Information Science, Nara Institute of Science and Technology
koichiro@is.naist.jp

高梨 克也 Katsuya Takanashi
京都大学情報学研究科 Graduate School of Informatics, Kyoto University
takanasi@sap.ist.i.kyoto-u.ac.jp

Nigel G. Ward Nigel G. Ward
テキサス大学エルパソ校計算機科学部／京都大学学術情報メディアセンター Department of Computer Science, University of Texas at El Paso / Academic Center for Computing and Media Studies, Kyoto University
nigel@utep.edu

河原 達也 Tatsuya Kawahara
京都大学情報学研究科／同学術情報メディアセンター Graduate School of Informatics / Academic Center for Computing and Media Studies, Kyoto University
kawahara@i.kyoto-u.ac.jp

keywords: spoken dialogue system, conversation agent, attentive listening, backchannel

Summary

There is a growing interest in conversation agents and robots which conduct attentive listening. However, the current systems always generate the same or limited forms of backchannels every time, giving a monotonous impression. This study investigates the generation of a variety of backchannel forms appropriate for the dialogue context, using the corpus of counseling dialogue. At first, we annotate all acceptable backchannel form categories considering the permissible variation in backchannels. Second, we analyze how the morphological form of backchannels relates to linguistic features of the preceding utterance such as the utterance boundary type and the linguistic complexity. Based on this analysis, we conduct machine learning to predict backchannel form from the linguistic and prosodic features of the preceding context. This model outperformed a baseline which always outputs the same form of backchannels and another baseline which randomly generates backchannels. Finally, subjective evaluations by human listeners show that the proposed method generates backchannels more naturally and gives a feeling of understanding and empathy.

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 [河原 13]. 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 [楡木 89]．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 [山本 09] 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 [目黒 12] 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」,「共感を示す発話をする」などが挙げられる．これらのうち，質問や共感発話を的確に生成するには，相手の発話の認識・理解が必要である．一方，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り，多くの研究が行われている．

相槌は会話を円滑に進める上で非常に重要な要素である．相槌は話し手の話を「聞いていること」,「理解していること」,「共感していること」などを表す役割がある [堀口 88]．また，相槌をうつことによって会話全体のリズムを生みだすこともできる．そのため，相槌をうつ対話システム [下岡 10, 横山 10, DeVault 14] や，相槌をうつタイミング [Koiso 98, 岡登 99, Ward 00, Kitaoka 05, 西村 09, Kamiya 10, Ozkan 11] について研究が行われてい

る．しかし，多くのシステムはあらかじめ決められた形態の相槌のみをうっており，その形態のバリエーションは乏しい．聞き手のうつ相槌が常に同じ形態の相槌（例えば「うん」）のみであると，話し手は相手が自身の話を聞いているのか，理解してくれているのかと不安や不満を感じる．さらに，会話のリズムに単調さや不自然さが生じる．一方，相槌の形態をランダムに決定すると，単調さの問題は改善されるかもしれないが，不自然さは残るであろう．これに対して，ユーザの話を傾聴するようなシステムにおいては，文脈に応じて適切に多様な形態の相槌をうつ必要がある．そこで，本研究ではそのような相槌の生成を目標として，相槌形態の分析・予測・生成に取り組む．

2. 相談対話コーパスと相槌の認定

2·1 相談対話コーパス

本研究では，上里ら [上里 14] の研究で収録された相談対話を用いる．対話のテーマは「日常の簡単な悩みや困りごと」であり，各セッションの対話は，話し役 1 名，聞き役 1 名で行っている．聞き役はスクールカウンセラ 2 名，話し役は大学生 8 名で，合計 8 対話が収録されている．対話時間は 20〜30 分である．この対話は大きく前半と後半に分かれており，前半はカウンセラが相談者の話を聞き，後半は相談者にアドバイスなどをしている．そこで，後述のアノテーションや予測・生成においては，前半部分を用いる（モデルの学習には全データを用いる）．

2·2 発話単位と相槌の認定

§1 発話単位の認定

本研究では発話単位として，間休止単位（IPU）と節単位の 2 つを用いる．

間休止単位（IPU）：間休止単位（IPU）は，笑いや咳などの「非言語発話」を除いた 200ms 以上の休止で挟まれた区間ごとに設定される発話単位である．

節単位：節単位は『日本語話し言葉コーパス』（CSJ）で定義された節境界 [高梨 04, 丸山 06] を区切りとした発話単位である．節境界には，その直後の構造的な切れ目の大きさという観点から切れ目の強さの順に絶対境界，強境界，弱境界の 3 種類があるが，発話単位の認定では，これら 3 種類を区別せずに扱う．

上記の通り，IPU は音響的な単位で，節単位は言語的な単位であるが，節末の大多数は IPU 末と一致している．また，本研究の分析対象において，節末の数は 1462 であり，節末に該当しない IPU 末の数は 1213 である．

§2 相槌の認定

相槌については様々な定義がなされているが，本研究ではメイナード [メイナード 93] の「話し手が発話権を持っている発話内で，話し手の発話に対して聞き手が発する，発話権の移動を伴わない発話」とした．相槌の形態ごとの分類には伝 [伝 15] による，「うん」や「ふんふん」といった促しや受容を表す応答系感動詞と，「あー」や「はー」といった興味や関心・共感を表す感情表出系感動詞の分類を参照した．

本研究では，応答系の「うん」と「ふん」を同種のものとして扱い，その繰り返し回数によって，応答系 1 回の相槌，応答系 2 回の相槌，応答系 3 回以上の相槌とカテゴリ化することとした．ただし，引き伸ばし系の「うーん」と「ふーん」は，機能が異なるものも含まれている可能性があるため除いている [中村 16]．また，「あー」，「はー」，「へー」は応答系と振る舞いが異なる [常 09]，感情表出系の相槌として一つのカテゴリとした．つまり，対象とする相槌を以下の 4 つにカテゴリ化した．

- 応答系 1 回（「うん」など）
- 応答系 2 回（「うんうん」など）
- 応答系 3 回以上（「うんうんうん」など）
- 感情表出系（「はー」など）

本研究の分析対象において，節末 1462 個のうち，相槌があるのは 777，相槌がないのは 685 であった．また，節末に該当しない IPU 末 1213 個において，相槌があるのは 297，相槌がないのは 916 であった．

2·3 相槌の追加アノテーション

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ものだけに限られない．さらに，相槌をうつことができる箇所も個人によってさまざまである．そのような相槌の任意性に対処するために，相槌カテゴリを複数の作業者によってアノテーションし，うつことができる相槌形態を拡張することとした．具体的には，カウンセラが相談者の話を聞いて，頻繁に相槌をうっている前半部分に対して，次の 2 つの時点でこのアノテーションを行った．

(A) 相槌がうたれている節末：ここでは，どの相槌カテゴリがうてるかどうかのアノテーションを行った．

(B) 相槌がうたれていない節末と，節末に該当しない IPU 末：節末かつ IPU 末で相槌が生起している箇所は (A) に含まれているため，(A) と (B) は排他的である．ここでは，応答系 1 回の相槌か応答系 2 回の相槌が「うてる」か「うてない」かの 2 値でアノテーションした．これは，IPU 末では，うたれる相槌の大半が応答系であるためである．

ここでアノテーションされた相槌カテゴリは，4 章と 5 章での予測実験の際の正解ラベルとして用いる．

§1 アノテーション手順

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コーパス中の 4 対話である．ただし，提示音声は話し手役の相談者の発話のみで，カウンセラの発話や相槌は含まれていない．

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

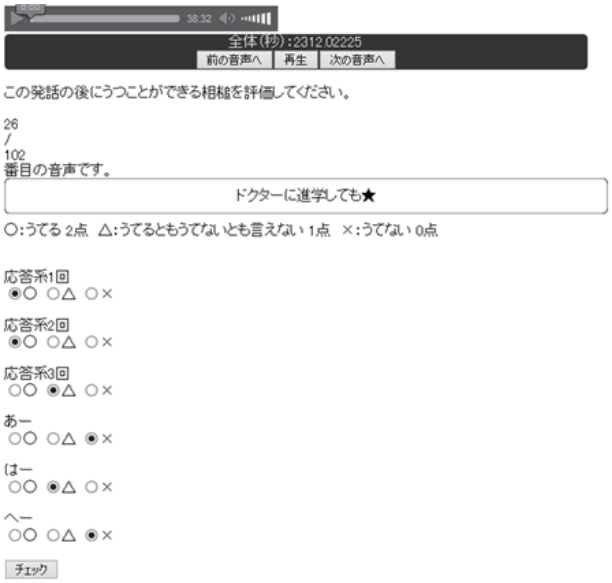

図 1 追加アノテーション用インターフェイス

音声の発話内容が文字で表示されており，アノテーションを行う該当箇所は★マークで示されている．該当箇所でカウンセラの相槌が生起していた場合にも，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区間へと移る．

(A) では，アノテータは節単位に区切られた提示音声を聞き，★の箇所で 4 つの相槌カテゴリのそれぞれについて,「○（うてる）」,「△（うてるともうてないとも言えない）」,「×（うてない）」の三段階で評価を行う．ただし，感情表出系のカテゴリでは,「あー」,「はー」,「へー」で振る舞いが異なる可能性があるため [常 08]，形態ごとに個別に評価する．

一方，(B) では，アノテータは IPU 単位もしくは節単位に区切られた提示音声を聞き，★の箇所で応答系 1 回もしくは応答系 2 回の相槌が,「○（うてる）」,「×（うてない）」の二段階の評価のみを行う．

§2 アノテーション結果

アノテータの一致率を Fleiss'κ 係数 [Fleiss 71] により求めた．(A) での全体の一致率は 0.099 (slight agreement) と非常に低い値であった．そのため，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，元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．認定された相槌カテゴリは 4 章と 5 章の予測実験の際に「正解」として扱う．

(B) の一致率は 0.298 (poor agreement) であった．ここでも，(A) と同様に，3 者一致したものを追加カテゴリとして認定した．認定されたカテゴリは 5 章の予測実験の際に「正解」として扱う．(A) と (B) の一致率の違いから，(A) の「どのような相槌形態がうてるか」という問題と，(B) の「相槌をうつことができるか」という問題では，(A) の方が任意性が高いことがわかる．

§3 アノテーション結果の分析

(A) の結果について，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）」とした割合をそれぞれ「許容度」と「非許容度」と定義し，許容できる追加形態の傾向を調べた．

図 2 に元形態 (a)〜(f) ごとの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 (a) 応答系 1 回と (b) 応答系 2 回では，それぞれ相互に許容度が高く，多くの発話に対して，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 回以上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も，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 (c) 応答系 3 回以上では，(a) と (b) に比べて，応答系 1 回への許容度が若干低下しているが，応答系 2 回と応答系 3 回以上では上昇している．また，(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 3 回以上の相槌は感情表出系に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 (d)「あー」と (e)「はー」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，感情表出系が追加形態として用いられる場合を分析すると,「へー」は許容されず,「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．

以上をまとめると，次の知見が得られた．

- 応答系 1 回と応答系 2 回は似た性質を持ち，相互に入れ替え可能となる場合が多い
- 応答系 3 回以上は感情表出系に近い性質を持つ
- 感情表出系では「はー」が他の感情表出系のものと比べてより柔軟に使える

感情表出系の相槌はそれぞれ性質が異なるが，応答系とは明らかに役割が異なる上に，個別にモデル化・評価を行うのはデータ量の点から困難であるので，1 つのカテゴリにまとめる．

3. 先行発話の統語構造と相槌の形態の関係の分析

話し手の発話の特徴から，相槌形態を予測できるか調べるために，先行発話の持つ特徴と後続する相槌との関係を分析した．ここでは，形態素列の表層的な特徴によっ

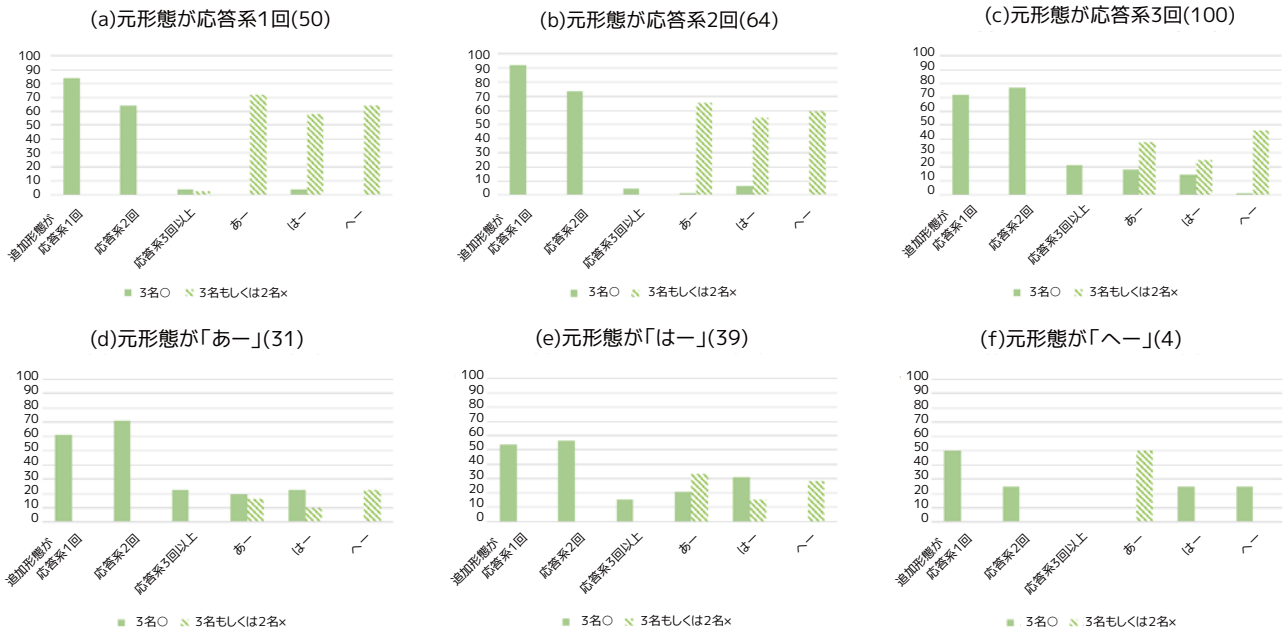

図 2 それぞれの元形態における各追加形態の許容傾向：各形態ごとにおけるグラフの左は各追加形態の許容度（3 名ともが○），右は非許容度（3 名もしくは 2 名が×）を表す．

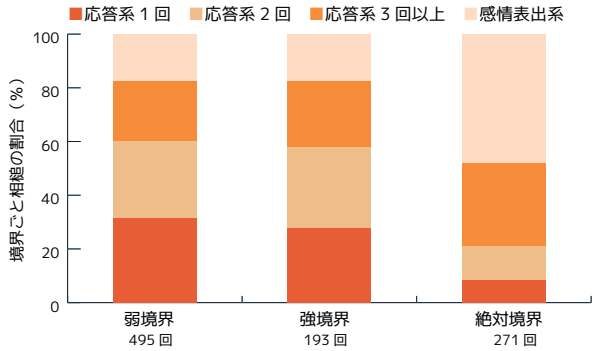

図 3 節境界の種類ごとの相槌形態のカテゴリの割合

て得られる節境界（3･1 節）と係り受け解析によって得られる構文構造（3･2 節）の 2 種類について，それぞれの先行発話から得られる情報と後続する相槌との関係を分析した．

3･1 先行発話の節境界の種類と相槌の形態の関係

先行発話の節境界の種類によって，うたれる相槌の形態が異なると考えられる．図 3 に各節境界ごとにうたれた相槌のカテゴリの割合を示す．下段の回数は各節境界の総数である．絶対境界では感情表出系の相槌が最も多くうたれ，次に応答系 3 回以上の相槌が多くうたれている．一方，弱境界と強境界では応答系 1 回の相槌と応答系 2 回の相槌が同様の割合で多いが，先行発話の節境界の種類だけでは弱境界と強境界の違いは区別することができない．

表 1 先行発話の構文構造と相槌の形態との関係（平均値）

	応答系 1 回	応答系 2 回	応答系 3 回以上	感情表出系
文節数	**4.73**	**5.52**	5.42	5.15
構文木の深さ	**2.18**	**2.57**	2.56	2.54
係り受けの数	1.88	2.00	1.89	1.75

3･2 先行発話の構文構造と相槌の形態の関係

先行発話の複雑度によって，うたれる相槌の形態が異なることも考えられる．そこで，先行発話の構文構造の観点から分析を行う．先行発話の構文構造の複雑さの指標として，節単位中の文節の数と，節末を根とする構文木の深さ，節末文節にかかる係り受けの数を用いる．構文解析には KNP [Kawahara 06] を使用した．

表 1 に相槌カテゴリとこれらの指標の関係の結果を示す．各カテゴリ間で t 検定（MS-Excel の TTEST 関数（2 標本が非等分散の場合のウェルチの t 検定）を使用）を行った結果，文節数と構文木の深さについて，応答系 1 回の相槌と 2 回の相槌の間で有意差が見られた（表の太字，有意水準 5%）．前節の節境界の種類に関する分析では，これらの区別はできなかったが，直前の発話の構文構造を用いれば区別できる可能性があることがわかった．

4. 相槌形態の予測

前章の分析に基づいて，先行発話から得られる特徴を用いて相槌形態の機械学習による予測を行う．本章では，相槌をうつタイミングは所与とし，相槌がうたれている節末のみを予測対象位置とする．

表 2 相槌形態の予測結果（相槌がうたれた節単位末）

		応答系 1 回	応答系 2 回	応答系 3 回以上	感情表出系	平均
提案モデル	適合率	0.982 (54/55)	0.826 (19/23)	0.488 (63/129)	0.654 (53/81)	0.656 (189/288)
	再現率	0.460 (23/50)	0.484 (31/64)	0.830 (83/100)	0.703 (52/74)	0.656 (189/288)
	F 値	0.626	0.611	0.615	0.678	0.656
言語的特徴のみ	適合率	0.980 (48/49)	0.714 (15/21)	0.468 (65/139)	0.646 (51/79)	0.622 (179/288)
	再現率	0.400 (20/50)	0.391 (25/64)	0.840 (84/100)	0.676 (50/74)	0.622 (179/288)
	F 値	0.568	0.505	0.601	0.660	0.622
韻律的特徴のみ	適合率	0.829 (29/35)	0.800 (8/10)	0.429 (75/175)	0.601 (41/68)	0.531 (153/288)
	再現率	0.160 (8/50)	0.250 (16/64)	0.860 (86/100)	0.581 (43/74)	0.531 (153/288)
	F 値	0.268	0.381	0.572	0.592	0.531
チャンスレート	適合率	0	0	0.417 (120/288)	0	0.417 (120/288)
	再現率	0.040 (2/50)	0.047 (3/64)	1 (100/100)	0.203 (15/74)	0.417 (120/288)
	F 値	0	0	0.417	0	0.417
ランダム生成	適合率	0.777	0.750	0.414	0.389	0.535
	再現率	0.334	0.394	0.707	0.521	0.535
	F 値	0.464	0.513	0.521	0.458	0.535

4·1 使用する特徴量

予測に使用する言語的特徴として，3 章の分析に用いた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て，節境界ラベル（CBAP-csj [丸山 04] で定義されたもの）や末尾語の表層形とその品詞を用いた．さらに，相槌の履歴として一つ前と二つ前の節境界の種類とそこで生起した相槌カテゴリも特徴に加えた．また，相槌のタイミングに関する多くの研究 [Kitaoka 05, 西村 09, Ozkan 11, Ozkan 13] で先行発話末から得られる F0 やパワーなどの韻律情報が用いられている．そこで，このような韻律情報から得られる特徴についても検討した．具体的には，西村ら [西村 09] にならい，先行発話末 150ms から得られる F0 とパワーについて 1 次回帰係数とレンジを特徴に含めた．ただし，F0 は 10 を底とする対数 F0 とした．さらに，発話長と先行発話末 150ms の話速と軋みさ（creakiness）も特徴として検討した．発話長は節の先頭から節末までの発話の時間である．軋みさ（creakiness）は末尾 150ms の間で，F0 におけるジッタが起こった回数をカウントしたものである．特徴を以下にまとめる．

言語的特徴：節境界の種類，節境界ラベル，末尾語（表層形），末尾語の品詞，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，一つ前の相槌カテゴリ，二つ前の節境界の種類，二つ前の相槌カテゴリ

韻律的特徴：発話長，対数 F0 の 1 次回帰係数，パワーの 1 次回帰係数，対数 F0 のレンジ，パワーのレンジ，話速，軋みさ（creakiness）

4·2 実 験 条 件

相談対話コーパスの 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 2·3 節で追加形態のアノテーションを行った 4 対話とした．分類器には ロジスティクス回帰（Liblinear [Fan 08]）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，各カテゴリごとで算出した．評価の際には，元形態だけでなく，相槌が生起した節末のみで行った (A) のアノテーション（2·3 節）で追加形態と認定された相槌カテゴリも正解としている．再現率を計算する際の母数はカウンセラがうった元相槌であるが，ある予測位置において，元形態ではなく，追加アノテーションによる形態を予測した場合も正解とする．この場合，追加形態ではなく，元形態に対する正解としてカウントする．

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（カウンセラにより）うたれていた回数が最も多い応答系 3 回以上をすべての予測箇所で出力する「チャンスレート」と，学習データにおける相槌カテゴリの度数分布にしたがって相槌形態をランダムに決める「ランダム」の二つを採用した．ただし，ランダム生成の結果 (適合率・再現率・F 値) は 1000 回試行の平均とした．

4·3 実 験 結 果

各相槌カテゴリごとの結果を表 2 に示す．
提案モデルについては，言語的特徴のみ及び韻律的特徴のみの場合の結果もあわせて示している．言語的特徴のみの方が韻律的特徴のみよりも予測精度が高いが，両者を組み合わせることの効果も確認できる．
ランダム生成の結果（適合率・再現率・F 値）は 1000 回

試行の平均のため，表中に個数は載せていない．チャンスレートにおける応答系 1 回や 2 回などの再現率は，応答系 3 回以上が追加形態となる箇所があるため，算出されている．各相槌カテゴリにおいて，提案手法は二つのベースラインを上回る結果となった．この結果から，分析に基づいた予測モデルは有効であるといえる．

提案モデルの結果について，相槌形態ごとに詳しく見ると，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，2·3 節で考察した追加形態のアノテーションにおいて，多くの先行発話に対して応答系 1 回と応答系 2 回が許容可能であった点と整合する．次に，応答系 3 回以上は，生起頻度が最も多いため再現率が高くなるものの，適合率は応答系 1 回と応答系 2 回と比べて低い．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．また，感情表出系では，適合率がランダム生成と比べて高い．このことから，不適切な箇所で感情表出系を予測する頻度が低く，感情表出系を文脈に応じて予測できていることがわかる．

5. 多様な形態の相槌の生成

前章では，多様な形態の相槌の中から適しているものを予測できるかを確かめるため，対象となる生起位置をカウンセラが相槌をうった節末のみに限定し，そこで 4 クラス分類を行った．しかし，システムによる相槌の生成のためには，相槌を「うつ」か「うたない」かも含めた予測を行う必要がある．そこで本章では，節末であるかどうかや，相槌の有無にかかわらず，相槌が生起可能なすべての候補位置で「応答系 1 回の相槌」,「応答系 2 回の相槌」,「応答系 3 回以上の相槌」,「感情表出系の相槌」,「うたない」の五つのカテゴリのいずれかの予測を行う．予測対象位置としては，カウンセラによる相槌の生起にかかわらず，音響的区切りである IPU 末を用いる．ただし前述の通り，節末の大多数は IPU 末である．

5·1 予 測 モ デ ル

相槌を「うつ」,「うたない」を含めた予測を行うために，以下の二通りのモデルを検討する．

5 クラス分類モデル：4 章の 4 クラスの予測モデルに，「うたない」を追加して 5 クラスの予測に拡張したものである．

2 クラス分類モデル：四つの相槌カテゴリそれぞれについて，そのカテゴリを「うつ」か「うたない」かの 2 クラスを予測するモデルを構成する．すなわち，2 値分類のモデルを 4 種類構築し，これらを組み合わせて最終的に出力する相槌カテゴリを決定する．まず，それぞれの相槌カテゴリのモデルにより，そのカテゴリを「うつ」確率値と「うたない」確率値を求める．次に，各カテゴリの「うつ」確率値が閾値以上なら，それを「うてる相槌カテゴリ」とみなし，これがなければ「うたない」を予測結果として出力する．また,「うてる相槌カテゴリ」が二つ以上ならば，その中で最も確率値が高いものを予測結果として出力する．閾値は，0.5 から順に下げていき，「相槌をうつ」と「うたない」の頻度分布が，評価データの頻度分布に最も近くなったときの値を採用する．その結果，閾値は 0.275 とした．このとき，各発話に対して「うてる」と判断された相槌カテゴリの個数は平均 2.3 個であった．

5·2 実 験 条 件

実験条件は前章の 4 クラス予測と同様である．ただし，本章では予測対象位置をすべての節末と IPU 末に拡大しているため，2·3 節の (A) だけでなく (B) のアノテーションも用いる．使用する特徴は 4·1 節と同様であるが，本実験では，予測対象位置を節単位より細かい IPU に拡大したため，一つ前の IPU の末尾語の品詞を加えている（多くの場合これは直前の節単位に含まれている）．比較のためのベースラインとしては，前章と同じく，ランダムに生成する方法を用いる．

5·3 実 験 結 果

予測結果を 表 3 に示す．提案手法である 5 クラス分類モデルと 2 クラス分類モデルともに，ランダム生成と比べて有効に予測できていることがわかる．この結果は，提案手法の有効性を示すものである．

提案手法同士の比較では，すべてのカテゴリで 2 クラス分類モデルの方がよい．5 クラス分類モデルは，再現率が「うたない」以外の 4 カテゴリで 2 クラス分類モデルよりも大幅に低く，そもそも相槌をうつと予測できている箇所自体が少ない．これに対して，2 クラス分類モデルはこの点が大幅に改善されている．

さらに，相槌形態ごとに詳しく見ると，2 クラス分類モデルは 5 クラス分類モデルに比べて，応答系 1 回と応答系 2 回の適合率が高いことから，これらを適切な箇所で予測できていることがわかる．しかし，応答系 3 回以上や感情表出系では，いずれの手法も適合率が低い．これは，節末でない IPU 末に対して,（B）のアノテーションでは応答系 1 回と応答系 2 回が「うてる」かどうかだけをアノテーションしたためと考えられる．なお，ランダム生成では，感情表出系の適合率が非常に低い．感情表出系は出現の頻度がそれほど多くなく，そのパターンも限定されているためと考えられる．

6. 音声を用いた印象評価実験

6·1 音 声 サ ン プ ル

前章の予測に基づいて生成される相槌について，音声データを用いた印象評定実験による評価を行う．

5 章で最も結果がよかった 2 クラス分類モデルの予測

表 3 相槌のタイミング及び形態の予測結果（すべての IPU 末）

		応答系 1 回	応答系 2 回	応答系 3 回以上	感情表出系	うたない	平均
5 クラス分類	適合率	0.676 (25/37)	0.750 (15/20)	0.293 (27/92)	0.357 (5/14)	0.648 (468/722)	0.610 (540/885)
	再現率	0.189 (20/106)	0.225 (20/89)	0.311 (38/122)	0.173 (13/75)	0.911 (449/499)	0.610 (540/885)
	F 値	0.295	0.346	0.302	0.223	0.757	0.610
2 クラス分類	適合率	0.657 (67/102)	0.820 (41/50)	0.333 (64/187)	0.342 (20/56)	0.769 (377/490)	0.643 (569/885)
	再現率	0.311 (33/106)	0.382 (34/89)	0.672 (82/122)	0.467 (35/75)	0.775 (385/499)	0.643 (569/885)
	F 値	0.422	0.521	0.454	0.405	0.772	0.643
ランダム生成	適合率	0.510	0.464	0.156	0.125	0.584	0.431
	再現率	0.170	0.195	0.281	0.232	0.597	0.431
	F 値	0.253	0.272	0.156	0.160	0.591	0.431

表 4 被験者による相槌の印象評定結果

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	**-0.42**	**1.04****	0.79
Q2：全体を通してテンポよく進んでいましたか	**0.25**	**1.29***	1.00
Q3：全体を通して真面目に聞いてくれていると感じましたか	0.25	1.04	1.08
Q4：全体を通して集中して聞いてくれていると感じましたか	0.50	1.29	0.96
Q5： 全体を通して積極的に聞いてくれていると感じましたか	0.63	1.21	1.08
Q6：全体を通して親身に聞いてくれていると感じましたか	0.33	1.25	0.96
Q7：全体を通して理解してくれていると感じましたか	**-0.13**	**1.17****	0.79
Q8：全体を通して関心を持ってくれていると感じましたか	**0.21**	**1.21***	1.04
Q9：全体を通して共感してくれていると感じましたか	**0.13**	**1.04***	0.46
Q10：このカウンセラと話したいと思いましたか	**-0.33**	**0.96****	0.29
		* $p < 0.05$	** $p < 0.01$

結果に基づいて相槌を生成し，音声データを作成した．相槌を挿入する位置は 5 章と同様，節末もしくは IPU 末である．比較のため，相談対話の音声に対して，以下の 3 条件の相槌音声を挿入した提示サンプルを作成した．相槌音声はアンドロイド ERICA の TTS（HOYA 音声合成ソフトウェア VoiceText ERICA） [井上 15] 用に収録したものを使用した．

(1) ベースライン条件：ランダムに生成されたカテゴリ（5 章と同じ）
(2) 予測条件：提案手法の 2 クラス分類モデルにより生成された相槌カテゴリ
(3) カウンセラ条件：相談対話収録時にカウンセラがうったのと同じ相槌カテゴリ（ただし，相槌音声は元のカウンセラのものではなく，他の条件と同様に，TTS 用のものを使用）

相槌音声の形態としては，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，予測条件とベースライン条件において感情表出系をうつ場合には，2·3 節のアノテーションにおいて「あー」よりも許容度があるとされた「はー」を使用した．相槌音声は 3 条件とも同じ音声を使用した．各形態の音声は，相談対話に合うように，なるべく控えめなものを選んだ．

使用した対話音声は，4 章と 5 章で用いた 4 対話から 1 分半〜2 分程度連続した区間を 2 か所ずつ抜粋した合計 8 対話セグメントである．

6·2 印 象 評 定

聴取する被験者は 20 代の男女 9 名（男性 5 名，女性 4 名）である．各被験者は音声データを聴取して，相槌の印象を評価する．抜粋した音声は対話の途中から始まるため，被験者が話の内容を理解しやすいように，そこまでの対話内容の概略を口頭で説明した．

8 対話セグメント × 3 条件で合計 24 のサンプルがあるが，各サンプルについて，3 名の異なる被験者から回答が得られるようにした．また，対話セグメン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．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 [堀口 88] の相槌の機能を参考にして作成した．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 7 件法（-3：全くそう思わない 〜 3：非常にそう思う）で評定する．

6·3 評 価 結 果

各被験者の評価値の平均を 表 4 に示す．どの項目においてもベースライン条件が最も低い評価値であった．

ランダムに生成されるベースライン条件と提案法による予測条件との間に有意差があるか t 検定（MS-Excel の TTEST 関数（２標本が非等分散の場合のウェルチの t 検定）を使用）を行ったところ，有意水準 1%で項目 Q1，Q7，Q10，有意水準 5%で項目 Q2，Q8，Q9 で有意差が認められた（表の太字）．このように提案法では，半数以上の評価項目でベースライン条件よりも高い評価を得た．一方，予測条件とカウンセラ条件の間について検定したところ，すべての項目で有意差がなかった．この結果から，提案法はカウンセラと同程度の評価値を得ていると考えられる．

以下，各項目ごとに考察を行う．まず，項目 Q1 の「自然さ」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，相槌形態やタイミングがランダムに生成されたものでは，不自然な印象を与えることがわかる．また，項目 Q10 の「カウンセラと話したいか」に関しても，ランダム生成では，話したいという印象を与えることができないことがわかる．

項目 Q2 の「テンポの良さ」では，予測条件やカウンセラ条件の方がベースライン条件と比べて評価値が非常に高い．ベースライン条件では対象位置で相槌をうつかどうか自体がランダムであるのに対して，予測条件では相槌をうつタイミングがより適切であるため，テンポよく会話が進んでいるという印象を与えていると考えられる．ここでは，相槌が生起可能なすべての位置で予測しているため，相槌を「うつ」か「うたない」かの判断が被験者の印象に大きく影響を与えると考えられる．実際に，5 章の予測実験で，ランダム生成での「うたない」の適合率や再現率が低いことから，うつべきでない位置でうっていると考えられる．一方，予測条件では,「うたない」の適合率や再現率は十分に高いとまではいえないが，ベースライン条件と比べれば高く，テンポに関する項目 Q2 の評価もよい．このことから，タイミングに関しては，ある程度以上の精度があれば，必ずしも悪い印象を与えることはないと考えられる．

次に，相槌の機能に関する項目 Q3〜Q6 の「聞いてくれていると感じるか」では，ベースライン条件と予測条件で有意な差が見られなかった．このことから，単に聞いてくれているという印象を与えるだけなら，どのような形態の相槌でも十分であるという可能性もうかがえる．しかし,「理解」や「共感」に関する項目 Q7，Q8，Q9 では，いずれも予測条件はベースライン条件を有意に上回るとともに，カウンセラ条件とも差がない．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ように，聞き手の理解には深さの度合いがあり，聞き手はこの理解の深さを反応の強さによって表すと考えられる [Allwood 92]．5 章の予測実験でランダム生成での感情表出系の適合率が非常に低かったことから，特に「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出現していたことの影響が大きいと考えられる．

5 章の 2 クラス分類モデルの結果では,「うたない」以外の 4 カテゴリの精度はあまり高くないように見える．しかし，提案法の予測条件は，カウンセラ条件と同程度の評価を得ており，予測精度が十分でなくても，ユーザに対して一定の印象を与えられると考えられる．ただし，カウンセラ条件も含めて評価値は必ずしも高くない．その理由として，形態以外の要因が考えられる．具体的には，上里ら [上里 14] のように韻律の調整を適応的に行うことや，タイミングについても先行する発話へのオーバーラップを含めて精密に調整することが考えられる．

7. ま と め

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，相槌の形態と先行発話の統語構造や韻律的特徴との関係について分析し，これらの先行発話の特徴から相槌の形態の予測と生成を行った．まず，本研究で用いる相談対話コーパスに対して，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アノテーションを行った．この結果から相槌形態の性質を考察することができた．次に，話し手の発話の区切りに出現する相槌を対象として，先行する発話の統語構造と相槌の形態の関係を分析した．その結果，節境界や構文構造が相槌形態を区別するのに有用であることがわかった．この知見に基づいて，先行発話の特徴から相槌形態を機械学習により予測できるか実験を行った．その結果，決められた形態のみやランダムに生成する方法と比べて効果的に予測できることが示された．そして，相槌が生起可能なすべての位置を対象とした予測実験を行った結果，ランダムに生成する方法と比べて有効に予測できることが示された．さらに，このシステムで生成した相槌の音声データを用いた印象評価実験を行った結果，ランダムに生成する方法と比較して，相槌の自然さや「理解」,「共感」などの表現において有意に高い評価を得ることができた．

本論文では，2 名のカウンセラのデータを用いてモデル学習を行ったが，構築されたモデルは，これらのカウンセラに依存している可能性がある．また，相槌には男女による差があることも指摘されている [辻本 07]．実際に本研究で用いた 2 名のカウンセラは男女 1 名ずつであったが，相槌の生起パターンに違いが見られた．そもそも，全体の学習データ量も多いとはいえない．より一般的なモデル学習のためには，より大規模なデータを収集する必要があり，性別や状況別のモデル化も検討する必要があると考えられる．

今後の展望として，上里ら [上里 14] の知見を用いて，相槌を生成する際に韻律的特徴を調整することが挙げられる．これにより，相槌生成モジュールを自律型アンドロイド ERICA の音声対話システム [井上 15] に実装することが可能になる．

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．本研究の一部は，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 ◇

[Allwood 92] Allwood, J., Nivre, J., and Ahlsén, E.: On the semantics and pragmatics of linguistic feedback, *The Journal of semantics*, Vol. 9, No. 1, pp. 1–26 (1992)

[伝 15] 伝 康晴：対話への情報付与, 小磯花絵 (編)『講座 日本語コーパス 3: 話し言葉コーパス—設計と構築—』, pp. 101–130, 朝倉書店 (2015)

[DeVault 14] DeVault, D., Artstein, R., Benn, G., Dey, T., Fast, G. E., Gainer, A., Georgila, K., Gratch, J., Hartholt, A., Lhommet, M., Lucas, G., Marsella, S., Morbini, F., Nazarian, A., Scherer, S., Stratou, G., Suri, A., Traum, D., Wood, R., Xu, Y., Rizzo, A., and Morency, L.-P.: SimSensei Kiosk: A virtual human interviewer for healthcare decision support, in *Proceedings of International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pp. 1061–1068 (2014)

[Fan 08] Fan, R. E., Chang, K. E., Hsieh, C. H., Wang, X. R., and Lin, C. J.: LIBLINEAR: A library for large linear classification, *The Journal of Machine Learning Research*, Vol. 9, pp. 1871–1874 (2008)

[Fleiss 71] Fleiss, L. J.: Measuring nominal scale agreement among many raters, *The Journal of Psychological bulletin*, Vol. 76, No. 5, pp. 378–382 (1971)

[堀口 88] 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. 13–26 (1988)

[井上 15] 井上 昂治, 河原 達也：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 言語・音声理解と対話処理研究会 (SLUD), Vol. SLUD-B502-5 (2015)

[Kamiya 10] Kamiya, Y., Ohno, T., and Matsubara, S.: Coherent back-channel feedback tagging of in-car spoken dialogue corpus, in *Proceedings of Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGdial)*, pp. 205–208 (2010)

[Kawahara 06] Kawahara, D. and Kurohashi, S.: A fully-lexicalized probabilistic model for Japanese syntactic and case structure analysis, in *Proceedings of Human Language Technology and the North American Chapter of the Association of Computational Linguistics (HLT/NAACL)*, pp. 176–183 (2006)

[河原 13] 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)

[Kitaoka 05] Kitaoka, N., Takeuchi, M., Nishimura, R., and Nakagawa, S.: Response timing detection using prosodic and linguistic information for human-friendly spoken dialog systems, *The Journal of Japanese Society for Artificial Intelligence*, Vol. 20, No. 11, pp. 220–228 (2005)

[Koiso 98] Koiso, H., Horiuchi, Y., Tutiya, S., Ichikawa, A., and Den, Y.: An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese map task dialogs, *The Journal of Language and Speech*, Vol. 41, No. 3, pp. 295–321 (1998)

[丸山 04] 丸山 岳彦, 柏岡 秀紀, 熊野 正, 田中 英輝：日本語節境界検出プログラム CBAP の開発と評価, 自然言語処理, Vol. 11, No. 3, pp. 39–68 (2004)

[丸山 06] 丸山 岳彦, 高梨 克也, 内元 清貴：節単位情報,『日本語話し言葉コーパスの構築法』, pp. 255–322, 国立国語研究所 (2006)

[メイナード 93] 泉子・K・メイナード：会話分析, くろしお出版 (1993)

[目黒 12] 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)

[中村 16] 中村 静, 高梨 克也, 山口 貴史, Ward, N., 河原 達也：相槌「うん」と「うーん」の表記の問題と韻律的な特徴, 人工知能学会研究会資料 言語・音声理解と対話処理研究会 (SLUD), Vol. SLUD-B503-10 (2016)

[楡木 89] 楡木 満生：積極的傾聴法, 医学教育, Vol. 20, No. 5, pp. 341–346 (1989)

[西村 09] 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 音声言語情報処理研究会 (SLP), Vol. 2009-SLP-77-22, (2009)

[岡登 99] 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)

[Ozkan 11] Ozkan, D. and Morency, L.-P.: Modeling wisdom of crowds using latent mixture of discriminative experts, in *Proceedings of the Annual Meeting of the Association for Computational Linguisticsa and Human Language Technologies (ACL/HLT)* (2011)

[Ozkan 13] Ozkan, D. and Morency, L.-P.: Prediction of visual backchannels in the absence of visual context using mutual influence, in *Proceedings of Intelligent Virtual Agents (IVA)*, pp. 451–454 (2013)

[下岡 10] 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 言語・音声理解と対話処理研究会 (SLUD), Vol. SIG-SLUD-A903-11 (2010)

[高梨 04] 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)

[辻本 07] 辻本 桜子：あいづちの男女差に関する一考察: トーク番組における司会者のあいづちを通して, 日本言語文化研究, Vol. 11, pp. A33–A45 (2007)

[上里 14] 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 言語・音声理解と対話処理研究会 (SLUD), Vol. SLUD-B303-02 (2014)

[Ward 00] Ward, N. and Tsukahara, W.: Prosodic features which cue backchannel responses in English and Japanese, *The Journal of Pragmatics*, Vol. 32, No. 8, pp. 1177–1207 (2000)

[山本 09] 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ヒューマンコミュニケーション基礎研究会 (HCS), Vol. HCS2009-56 (2009)

[横山 10] 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音声言語情報処理研究会 (SLP), Vol. 2010-SLP-80-4 (2010)

[常 08] 常 志強, 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 言語・音声理解と対話処理研究会 (SLUD), Vol. SLUD-A802-02 (2008)

[常 09] 常 志強, 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 言語・音声理解と対話処理研究会 (SLUD), Vol. SLUD-A901-06 (2009)

〔担当委員：坂本 真樹〕

2016 年 3 月 8 日　受理

著 者 紹 介

山口　貴史

2014 年 和歌山大学システム工学部デザイン情報学科卒業. 2016 年 京都大学大学院情報学研究科知能情報学専攻修士課程修了. 同年，株式会社日立製作所入社. 在学中，音声対話システムに関する研究に従事.

井上　昂治

2013 年 久留米工業高等専門学校専攻科修了．2015 年京都大学大学院情報学研究科修士課程修了．現在，同博士後期課程在学中，および日本学術振興会特別研究員（DC1）．音声言語処理，画像処理，マルチモーダルインタラクションに関する研究に従事．情報処理学会，日本音響学会，電子情報通信学会，IEEE，ISCA 各会員．

吉野　幸一郎

2009 年慶應義塾大学環境情報学部卒業．2011 年京都大学大学院情報学研究科修士課程修了．2014 年同博士後期課程修了．同年日本学術振興会特別研究員（PD）．2015 年より奈良先端科学技術大学院大学情報科学研究科特任助教．博士（情報学）．音声言語処理および自然言語処理，特に音声対話システムに関する研究に従事．2013 年度本学会研究会優秀賞受賞．IEEE，ACL，情報処理学会，言語処理学会 各会員．

高梨　克也（正会員）

2000 年京都大学大学院人間・環境学研究科博士課程単位取得退学．博士（情報学）．独立行政法人情報通信研究機構専攻研究員，京都大学学術情報メディアセンター特定助教，科学技術振興機構さきがけ専従研究者などを経て，現在京都大学情報学研究科研究員．コミュニケーションの組織化を支える認知的・社会的プロセスの解明に従事．言語処理学会，日本認知科学会，社会言語科学会，組織学会 各会員．一般社団法人社会対話技術研究所理事．

Nigel G. Ward

received the Ph.D. in Computer Science from the University of California at Berkeley in 1991. He was on the faculty of the University of Tokyo for ten years before joining The University of Texas at El Paso in 2002. In 2015-2016 he was a Fulbright scholar at Kyoto University.

河原　達也（正会員）

1987 年 京都大学工学部情報工学科卒業．1989 年 同大学院修士課程修了．1990 年 京都大学工学部助手．1995 年 同助教授．2003 年 同大学学術情報メディアセンター／情報学研究科教授．現在に至る．この間，1995～96 年 米国・ベル研究所客員研究員．1998～2006 年 ＡＴＲ客員研究員．2006 年～ 情報通信研究機構短時間研究員・招へい専門員．音声情報処理，特に音声認識及び対話システムに関する研究に従事．博士（工学）．科学技術分野の文部科学大臣表彰 (2012 年度)，日本音響学会から粟屋潔学術奨励賞 (1997 年度)，情報処理学会から坂井記念特別賞 (2000 年度)，喜安記念業績賞 (2011 年度)，論文賞 (2012 年度) を受賞．IEEE ASRU 2007 General Chair，INTERSPEECH 2010 Tutorial Chair，IEEE ICASSP 2012 Local Arrangement Chair，言語処理学会理事，情報処理学会音声言語情報処理研究会主査，APSIPA 理事，情報処理学会理事を歴任．情報処理学会，日本音響学会，電子情報通信学会，言語処理学会，IEEE，ISCA，APSIPA 各会員．