原著論文 Original Paper

アンドロイドERICAの傾聴対話システム –人間による傾聴との比較評価–

An Attentive Listening System for Autonomous Android ERICA: Comparative Evaluation with Human Attentive Listeners

井上 昂治 Koji Inoue
京都大学 大学院情報学研究科 Graduate School of Informatics, Kyoto University
inoue.koji.3x@kyoto-u.ac.jp

ラーラー ディベッシュ Divesh Lala
（同　上）
lala@sap.ist.i.kyoto-u.ac.jp

山本 賢太 Kenta Yamamoto
（同　上）
yamamoto@sap.ist.i.kyoto-u.ac.jp

中村 静 Shizuka Nakamura
（同　上）
shizuka@sap.ist.i.kyoto-u.ac.jp

高梨 克也 Katsuya Takanashi
滋賀県立大学 人間文化学部 School of Human Cultures, The University of Shiga Prefecture
takanashi.k@shc.usp.ac.jp

河原 達也 Tatsuya Kawahara
京都大学 大学院情報学研究科 Graduate School of Informatics, Kyoto University
kawahara@i.kyoto-u.ac.jp

keywords: attentive listening, spoken dialogue system, autonomous android, backchannel, listener response

Summary

An attentive listening system for autonomous android ERICA is presented. Our goal is to realize a human-like natural attentive listener for elderly people. The proposed system generates listener responses: backchannels, repeats, elaborating questions, assessments, and generic responses. The system incorporates speech processing using a microphone array and real-time dialogue processing including continuous backchannel prediction and turn-taking prediction. In this study, we conducted a dialogue experiment with elderly people. The system was compared with a WOZ system where a human operator played the listener role behind the robot. As a result, the system showed comparable scores in basic skills of attentive listening, such as *easy to talk*, *seriously listening*, *focused on the talk*, and *actively listening*. It was also found that there is still a gap between the system and the human (WOZ) for high-level attentive listening skills such as *dialogue understanding*, *showing interest*, and *empathy towards the user*.

1. はじめに

スマートスピーカや会話ロボットの普及により音声対話システムの実用化が進んでいる．ただし，そこで実現されている対話は，機器操作，情報検索，簡単な案内などであり，人間どうしのような長い対話は想定されていない．そのため，ユーザの発話は必要最小限のコマンド程度になることが多い．一方，人間どうしの対話に目を向けると，１つのターンは短いポーズを挟む複数の発話で構成され，発話の合間で聞き手は相槌をうつことで長い対話を円滑に進めている．今後の音声対話システムが実社会でより活用されるようになるためには，このような人間どうしの長くて深い対話を実現する必要があると考えらえる．

本研究では，人間どうしのような長い対話タスクの１つとして傾聴対話に焦点をあてる．傾聴とは，相手の話に耳を傾けて聞く場面を指す．例えば，高齢者を主な対象とした傾聴ボランティアが存在する．そこでは，カウンセリングのような悩みごとの相談とは異なり，相手の話をしっかりと聴くことが重視されている．その際に重要となる聞き手のふるまいとして以下が挙げられている [三島 03].

- 語られたことをそのまま繰り返す
- 語られた内容を言い換えて繰り返す
- 語られた内容を要約する
- もっと語るように問い返す
- 話し手に共感し，気持ちを言葉にする
- 相槌などにより聴いていることを示す

図 1 自律型アンドロイド ERICA

このような姿勢で話を聴いてもらうことで，話し手の心が軽くなり，自己肯定感や安堵感が生まれることが期待されている [ホールファミリーケア 09]．ただし，傾聴ボランティアの確保は人的・時間的なコストが高く，高齢者にとって日常的に話を聴いてもらえる相手を探すことは依然として難しい．そこで，音声対話システムが傾聴対話を行うことで，これを代替することが望まれる．

本稿では，自律型アンドロイド ERICA [Glas 16, Inoue 16]（図 1）を用いた傾聴対話システムについて述べる [*1]．アンドロイドは人間に酷似した見た目を有しており，人間のような存在感を対話相手に示すことができる．したがって，まさに人間に話を聞いてもらっているという感覚を与え，人間どうしのように対話が長くて深くなることが期待される．また，このような人間らしい対話を実現することは，傾聴対話の主な対象である高齢者にとって話しやすさやなじみやすさといった点で重要といえる．

本研究では，上記のような人間らしい傾聴対話を指向した傾聴対話システムを構築した．聞き手応答として，相槌（「うん」など）や語彙的応答（「そうですか」などの定型表現）に加えて，ユーザ発話の焦点語に基づく繰り返しや掘り下げ質問，ユーザ発話の極性に基づく評価応答を生成し，話し手への理解や共感を醸し出すことを目指す．また，音声認識，音声合成，ターンテイキングシステム，マルチチャネル音響信号処理，人物位置追跡などの技術を統合することで，音声対話システムとしての頑健性を向上させ，傾聴対話が長く継続できるようにする．このシステムの評価として，高齢者との対話実験を実施し，提案システムと人間のオペレータによる傾聴（WOZ: Wizard-of-Oz）とを比較した．著者らの究極の目標は，人間と遜色のない傾聴対話を実現することである．この対話実験を通じて，これまでに構築してきた傾聴対話システムと人間の会話能力との相違（ギャップ）を明らかにすることで，この目標への道筋を探ることも本研究の目的の１つである．

本論文の構成は以下の通りである．まず，傾聴対話システムに関する関連研究を 2 章でまとめる．次に，アンドロイド ERICA のシステム全体を 3 章で，その中の傾聴対話システムを 4 章でそれぞれ説明する．続いて，上記の対話実験について 5 章で報告する．最後に，結論と今後の課題を 6 章でまとめる．

2. 関 連 研 究

はじめに，聞き手応答の中でも相槌に絞り，その生成方法に関する関連研究を紹介する．次に，その他の応答の生成方法も含めた傾聴対話システムに関する先行事例を紹介する．

2·1 相 槌 生 成

相槌とは「うん」,「はい」,「ふん」,「へー」といった短い応答（感動詞）であり，傾聴対話における基本的な聞き手応答である．上記のうち「うん」,「はい」は応答系,「ふん」,「へー」は感情表出系に分類される．また，相槌には「聞いている」,「理解している」,「共感している」という聞き手の状態を話し手に対して非言語的に伝達する機能・効果がある [堀口 88]．

傾聴において効果的な相槌をうつためには，適切なタイミングを予測する必要がある．相槌のタイミング予測の典型的な問題設定は，ポーズを検出した時点で先行する発話の韻律や言語情報をもとに相槌をうつか否かを予測するものである [Fujie 05, Koiso 98, Ward 00]．韻律情報で主に用いられてきたのは基本周波数（F0）であり，発話末での変動パターンにその特徴が表れるとされている．また，言語情報では，発話末の品詞情報や文・節境界といった統語情報が用いられている．これらの方法に対して本研究では，ポーズを待つのではなく，韻律情報に基づき常に予測を行うことで，ユーザの発話途中であっても相槌をうつことができるモデル [Lala 17] を実現する．また，相槌の形態（種類）の予測に関しては，先行研究は少数であり [Kawahara 16]，予測には言語情報が必要であるため，本研究では応答系のみに絞り，ランダムに選択する．

2·2 傾聴対話システム

傾聴対話システムの源流として ELIZA [Weizenbaum 66] が挙げられる．このシステムはカウンセラーを模したもので，入力であるユーザ発話と予め用意したテンプレートとのマッチングに基づいてシステム発話を生成する．ユーザ発話の内容に基づいてシステム発話を選択することで，あたかもユーザ発話を理解しているようにみせかけることに成功した．

音声言語処理技術の発展に伴い，その後種々の傾聴対話システムが構築され，様々な場面での実装が試みられてきた [*2]．まず，テキスト対話を対象としたシステムと

*1 本論文は，著者らの研究成果 [Inoue 20] に追加実験を実施した上で改訂したものである．

*2 ここで挙げているシステムが対象とする対話は，必ずしも傾

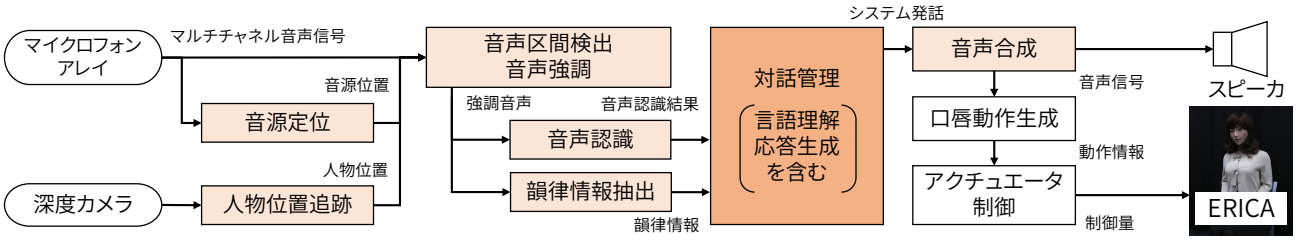

図 2 アンドロイド ERICA の音声対話システムの構成（傾聴対話システムは「対話管理」の内部に相当する）

して，大竹ら [大竹 14] や Han ら [Han 15] のシステムが挙げられる．次に，小林ら [小林 11, 小林 12]，Johansson ら [Johansson 16]，下岡ら [下岡 17] の研究では，卓上型，映像投影型，ペット型のロボットがそれぞれ用いられている．また，Schröder ら [Schröder 12] のシステムではバーチャルエージェントが用いられている．以上のシステムでは，繰り返し，掘り下げ質問，評価応答，語彙的応答といった，本研究で述べる我々のシステムと共通する聞き手応答の生成が試みられている [*3][*4]．このうち直近である下岡らのシステム [下岡 17] では，ユーザの発話の不足格を判定することで繰り返しや質問が生成されている．例えば,「昨日お花を貰いました」というユーザの発話に対して,「誰に」に相当する「ガ格」の情報が不足していると判断して「誰に貰ったのですか？」という質問を生成する．また，発話の感情極性を推定することで評価応答が生成されている．

本研究では，これまでの傾聴対話システムのコンセプトや聞き手応答の種類を踏襲しつつ，アンドロイドを用いて，人間らしい聞き手としてふるまう傾聴対話システムの実現を目指す．人間らしい傾聴対話の特徴として，対話の継続時間やユーザの話し方が考えられる．上記の関連研究のうち直近のもの [下岡 17] では，110 名の被験者との対話実験が実施されており，対話の継続時間は２～３分がほとんどであった．また，１分あたりの発話単語数はおおむね 10～20 個であった．したがって，そこでの対話は一問一答のようなもので，ユーザ発話の後にかなり間があいてからシステム応答がなされていたと推測される．これに対して本研究では対話時間をより継続させ，かつユーザがより多くの内容を話せるようになることを目指す．そのために本研究では，従来の聞き手応答の生成に加えて，相槌生成やターンテイキング予測といった機能にも焦点をあてる．このようなシステム統合の評価として，これまでの研究では行われていなかった人間による傾聴（WOZ）との比較を本研究では試みる．

聴という文脈ではないが，話を聞くことに焦点をあてているという点は一貫している．

*3 各聞き手応答の名称は，提案システムのものに合わせており，各文献でのものとは異なることに注意されたい．また，同じ種類の聞き手応答であっても，その生成方法は研究によって異なる．

*4 下岡らのシステム [下岡 17] では「相槌」が生成されるが，これは本研究での「語彙的応答」に対応する．

3. アンドロイド ERICA

本研究で用いるアンドロイド ERICA のシステム全体について紹介する．ERICA は，人間レベルの自然なインタラクションを実現するための研究開発プラットフォームである．その姿形はコンピュータグラフィックスにより人工的に設計されたものである．顔，頭部，肩，腰，腕，手の計 46 箇所に能動関節があり，空気圧アクチュエータで動作する．したがって，音声に加えて，視線，うなずき，表情，ジェスチャなどのマルチモーダルなふるまいの表現が可能である．

ERICA の音声対話システムの全体構成を図 2 に示す．入力となるセンサは，マイクロフォンアレイ（16 チャネル）と深度カメラ（Kinect v2）である．マイクロフォンアレイを用いることでユーザはマイクを意識することなく自然な形で対話を開始することができる．

システムの入力部分（図 2 左側）に相当する音声信号処理について述べる．マイクロフォンアレイから入力されるマルチチャネル音声信号から，MUSIC 法 [Schmidt 86] により音源方向を定位し，深度カメラで追跡する人物位置と比較する．これらが一致した場合に，その方向の音声のみを遅延和（Delay-and-Sum）ビームフォーミングにより強調し，音声認識へ入力する．音声認識のモデルは，サブワード単位の注意機構に基づく End-to-End 型ニューラルネットワークである．また，強調音声から基本周波数（F0）とパワーの韻律情報を抽出する [Ishi 16]．音声認識結果と韻律情報は，対話管理へと入力され，システム発話を生成する．この対話管理が，本研究の主題である傾聴対話システムに相当する．

次に，システムの出力部分（図 2 右側）について述べる．対話管理から出力されるシステム発話は，ERICA 用に設計・開発された音声合成 [*5] により再生する．ただし，聞き手応答のうち，相槌，語彙的応答，評価応答は予め録音したものを用いることで，自然な韻律パターンを実現している．それ以外の応答（繰り返しや掘り下げ質問）はその都度合成する．また，合成した音声に対してアンドロイドの口唇や頭部の付随動作も生成する [石井 13, 境 14]．

*5 https://voicetext.jp/news/product/151023/

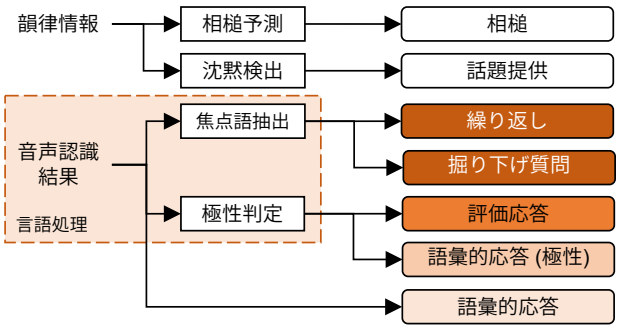

図 3 聞き手応答の一覧

4. 提案システム

アンドロイド ERICA を想定して実装した傾聴対話システムについて説明する．

4·1 傾聴対話の設定

提案システムでは，与えられたテーマに沿った 7 ～ 8 分程度の傾聴対話を想定する．テーマは,「今までで一番印象に残っている旅行」や「最近食べたもので一番美味しかったもの」などのように過去の出来事や心情を思い出しながら話すものに設定する．テーマ設定には，高齢者に対する心理療法である回想法 [Butler 63] のライフレビューの概念を参考にしている．話題提供を除き，特定のテーマやトピックに依存しないように設計・実装されている．

4·2 聞き手応答の生成

聞き手応答の一覧と処理の流れを図 3 に示す．繰り返しと掘り下げ質問には話に対する理解や興味，評価応答と語彙的応答（極性）にはユーザに対する共感をそれぞれ示しているようにみせることが期待される．各聞き手応答の生成方法を以下で述べる．

§ 1 相 槌

ユーザの発話に合わせて,「うん」や「うんうん」などの相槌を生成する．相槌を生成するためには，そのタイミングと形態を決定する必要がある．タイミングは，100 ミリ秒毎に，その時点から 500 ミリ秒以内に相槌をうつか否かを，韻律情報を用いてロジスティック回帰により予測する [Lala 17]．特徴量は，基本周波数（F0）およびパワーについての，平均，最大，最小，レンジなどの統計値である．このように，ユーザ発話の途中であっても連続的に予測を行うことで，ユーザ発話の終了を待たずして相槌をうつことが可能になる．形態は，著者らが過去に収録した傾聴対話データにおいて，実際に使用された相槌の形態の分布に基づいて選択する．

§ 2 繰り返し

音声認識結果からユーザの発話内の焦点である単語（焦点語）を抽出し，オウム返しのように焦点語を繰り返す応答を生成する．これにより話を理解しているようにみせる．焦点語は文末により近い名詞または形容詞とする．名詞が連続する場合は複合語とみなして，その複合語を焦点語の候補とする．応答のフレームは「(焦点語）ですか」とする．例えば,「去年，シンガポールへ行きました」という音声認識結果が得られた場合,「シンガポールですか」という応答が生成される．ただし，焦点語の候補となる単語の音声認識の信頼度が閾値以下の場合には，焦点語はなしとする．

§ 3 掘り下げ質問

「どんな」,「どの」,「なんの」,「なにの」,「どこの」,「いつの」,「だれの」といった 7 種の疑問詞と焦点語を組み合わせることで，焦点語に関する掘り下げ質問を生成する．これにより話を理解しているようにみせつつ，さらに話題（焦点語）に関して対話を掘り下げていく．はじめに，上記のすべての組合せについて，その N-gram 確率を算出する．そして，N-gram 確率が閾値以上かつ最大のものを用いて応答を生成する．例えば，焦点語が「家族」の場合に,「どんな」という疑問詞との組合せが選ばれると,「どんな家族ですか」という質問が生成される．N-gram 確率のモデルは，現代日本語書き言葉均衡コーパス（BCCWJ）から学習したものを用いる．

§ 4 評価応答

ユーザ発話の極性（ポジティブまたはネガティブ）に応じた応答を生成することで，話し手に共感しているようにみせる．極性の判定には，日本語アプレイザル辞書 [*6] の「情動」と「心状」に該当する単語群，ならびに SNOW D18:日本語感情表現辞書 [*7] を用いる．前者の辞書には各単語にポジティブまたはネガティブの極性が付与されている．後者の辞書には，各単語が 48 感情へ分類されているため，48 感情をポジティブまたはネガティブへと再分類した．これらの辞書の極性に基づきユーザ発話の極性を判定する．ただし，ユーザ発話中の複数の単語が異なる極性を持つ場合には，全体で多い方の極性を採用する．また，各単語の極性を判定する際に，その単語の音声認識の信頼度が閾値以下の場合には，その単語の極性は考慮しない．極性の判定結果がポジティブの場合には「いいですね」または「素敵ですね」，ネガティブの場合には「大変ですね」または「残念でしたね」という応答を出力する．

§ 5 語彙的応答

上記のような動的に生成される応答は，常に生成できるとは限らない．そのため,「そうですか」,「そうなんですね」,「なるほど」といった定形表現の応答も用意しておく．ただし，ユーザ発話の単語数が少ない場合には「はい」という短い語彙的応答を出力する．

*6 https://www.gsk.or.jp/catalog/gsk2011-c/
*7 http://www.jnlp.org/SNOW/D18

§ 6 語彙的応答（極性）

評価応答と語彙的応答の中間として，ユーザ発話の極性に応じて語彙的応答の韻律パターンを使い分ける．例えば，極性がポジティブな場合には，ポジティブな印象を与える韻律で「そうですか」と応答する．極性を判定する辞書には，評価応答の生成で用いたものよりも幅広いもの [小林 05] を用いる．したがって，言語的に破綻するリスクを抑えつつ，話し手への共感を示すことが期待される．

§ 7 話題提供

ユーザが話す内容について思いつかずに一定時間以上沈黙している場合には，バックアップとして話題提供を行う．あらかじめ設定される傾聴のテーマに応じて，話題提供発話を用意しておく．例えば,「今までに行った旅行」が傾聴のテーマの場合には,「そのあとはどこに行きましたか」などを用いる．

4·3 応 答 選 択

提案システムは複数の応答生成モジュールで構成されており，ユーザ発話が入力されると並列に応答生成が試みられる．したがって，これらの中から１つを選択して，実際の発話を行う．ただし，相槌と話題提供はそれぞれ独自のタイミングで発話がなされる．それ以外の発話に関しては,「評価応答」,「掘り下げ質問」,「繰り返し」,「語彙的応答（極性）」,「語彙的応答」の優先順位とした．提案システムの聞き手応答のねらいには「理解」と「共感」があるが，最終的には感情のレベルでユーザとの関係性を構築することが重要と考え，ここでは「共感」を優先し評価応答を最優先とした．続いて,「理解」の中でも生成される頻度が少ない順に「掘り下げ質問」,「繰り返し」とした．また，評価応答を生成するための辞書は極性が明確に表れているものに絞られているため，評価応答が生成される頻度は少ない．その観点からも最も高い優先度に設定している．

4·4 ターンテイキング

ターンテイキングは，音声認識結果の単語系列を用いて，LSTM（Long Short-Term Memory）ベースのニューラルネットワークによりユーザのターン終了の事後確率を予測する．加えて，状態遷移モデルを用いてターンを獲得するまでのポーズ長を決定する [Lala 18, Raux 09]. ターン終了の事後確率が大きいほどポーズ長は短く，小さいほどポーズ長は長くなる．

4·5 対 話 例

対話例を図 4 に示す．ただし，相槌はユーザのターン中にもなされるため，ここでは対話のわかり易さのために表示していない．

話者	発話内容
S	これまでで一番印象に残っている旅行についてお話を聞かせてください
U	去年の春に京都にある公園に行きました
S	どの公園ですか（**掘り下げ質問**）
U	八坂というところにある円山公園です
S	円山公園ですか（**繰り返し**）
U	そうです
S	はい（**語彙的応答**）
U	天気がよかったので賑わっていました
S	そうなんですね（**語彙的応答**）
U	そして私は有名な枝垂れ桜の周りを散歩しました
S	散歩ですか（**繰り返し**）
U	はい．とても楽しかったです
S	素敵ですね（**評価応答**）
	（長い沈黙）
S	そこでは他に何をしましたか（**話題提供**）

図 4 対話例（S：システム，U：ユーザ）

5. 対話実験による評価

高齢者との対話実験により提案システムを評価した．著者らの最終的な目標は，人間と同じレベルの傾聴対話を実現することである．提案システムによる傾聴と人間のオペレータによる傾聴（WOZ: Wizard-of-Oz）とを比較することで，現状のシステムと人間とのギャップ（または同レベルの点）を調査した．

5·1 条 件

被験者は高齢者 20 名で，年代は 70 代と 80 代でそれぞれ 14 名と 6 名ずつであった．対話時間は 7 分に設定した．７分を経過してシステムがターンを取得した時点で対話が終了する旨をシステムが発話し，対話を終了した．また，WOZ の場合には，オペレータに 7 分を経過したら対話を終了するように事前に伝えた．ただし，提案システムの場合には，用意した話題提供（４種類）を全て発話した場合には，７分に満たない時点でも対話を終了した．各被験者には，提案システムと WOZ のそれぞれと対話をしてもらい，表 2 に示す 19 項目（７段階評価）について個別に評価をしてもらった．これらは相槌生成システムの評価項目 [Kawahara 16] を参考にして設計されている．また，被験者には提案システムの対話能力やもう一方が WOZ であることは伝えなかった．傾聴のテーマは,「印象に残っている旅行」と「最近食べて美味しかったもの」の２つとし，対話相手の各条件（システムまたは WOZ）に対して１つのテーマを割り当てた．対話相手の条件とテーマの組合せおよび対話の順序は被験者毎にランダムに設定した．本実験での音声認識

図 5 対話実験における WOZ のオペレータ

図 6 ERICA の非言語行動に関する WOZ のオペレータ用の操作マニュアル

の単語誤り率の平均は 33.8%であった．

WOZ の設定について説明する．この条件では，別室のオペレータ（操作者）が ERICA 役としてマイクに向かって発話し，その音声を ERICA のスピーカからそのまま再生した．音声合成とオペレータの音声との違いをなくすために，オペレータに文を作成（または選択）してもらいそれを音声合成で再生することも考えられるが，本実験のように被験者が話し続ける設定では，できるだけ早く聞き手応答を発話することが要求される．相槌，評価応答，語彙的応答などの固定的な表現では選択による発話も可能であるが，繰り返しや掘り下げ質問に関しては焦点語や疑問詞を決定して文を構成する必要があるため，今回のようにオペレータが直接発話する方式を採用した．図 5 に対話実験におけるオペレータの様子を示す．また，うなずき，視線，ジェスチャの非言語行動はオペレータの手元のコントローラ（Xbox 360 コントローラ）で操作してもらった．オペレータへ提示したコントローラの操作マニュアルを図 6 に示す．傾聴の方法として，提案システムが生成する聞き手応答の範囲内で応答を行ってもらった．ただし，応答の種類，タイミング，内容（焦点語や極性）はオペレータ自身で適切なものを考えてもらうようにした．つまり，この条件は，提案システムの枠組みにおいて，人間による適切な応答生成がなされた場合に相当する．オペレータは 3 名の女性が交替で務めた．参加した対話数はそれぞれ 7，7，6 だった．彼女らは ERICA の設定の年齢に近い女性劇団員であり，発話する際の韻律を ERICA の音声合成のものにできるだけ近づけてもらうようにした．また，これまでにも同様の対話実験にオペレータ役として参加した経験を有していた．対話実験に先立ち，オペレータには，上記の傾聴の方法について説明を受けてもらうとともに，対話の練習を十分に行ってもらった．また，練習に際してオペレータには被験者役として提案システムと対話をしてもらい，提案システムのおおよその動作を把握してもらった．

5·2 提案システムの各発話の評価

WOZ との比較を行う前に，提案システムがどの程度妥当な聞き手応答を発話できていたかを把握するために，5·1 節で説明した対話実験におけるシステムの各発話について，著者らによる評価を実施した．ここでは，システム応答の内容によっては対話が破綻するリスクがある「繰り返し」,「掘り下げ質問」,「評価応答」,「語彙的応答（極性）」を評価の対象とした．一方で，ほとんどの文脈に対して発話することができる「相槌」,「話題提供」,「語彙的応答」は対象には含めなかった．

評価基準は以下の 3 種類である．1 つ目は，その応答が文脈的に破綻しなかったかどうかを焦点語から客観的に評価したものである．したがって，評価対象は焦点語を用いる応答の「繰り返し」と「掘り下げ質問」である．評価基準は，応答生成に用いた焦点語が，被験者の先行発話に含まれていなければ破綻とした．つまり，被験者が述べていないことを繰り返したり，掘り下げたりした場合は破綻とした．評価応答と語彙的応答（極性）は直前の発話を直接参照しないため，この評価からは除外した．

2 つ目は，その応答に対する被験者の反応に基づいて客観的に評価したものである．具体的には，そのシステムの発話の直後に，被験者が「はい」や「そうなんですよ」などの肯定的な反応を示したかどうかを調べた．「肯定的な反応」の基準を客観的にするために反例として「いえ」や「違います」などの否定,「えっ」などの聞き返し，無視を定めた．また,「掘り下げ質問」に関しては，質問に対して被験者が回答していればそれを肯定的な反応とした．「語彙的応答（極性）」に関しては，話し手によっては何らかの反応を示さずに話を継続しても不自然ではないため，ここでは評価の対象に含めなかった．システムの発話が多少不自然なものであったとしても，実際には話し手が対話に満足しているかが重要であるので，このような肯定的な反応に基づく評価方法も有用であるといえる．

3 つ目は，聞き手応答として適切か否かというものである．具体的には,「繰り返し」は焦点語,「掘り下げ質問」は焦点語と疑問詞の組合せ,「評価応答」と「語彙的応答（極性）」は極性が，それぞれ被験者の先行発話の内容に対して適切であるかが判断基準である．ただし，そもそも焦点語がない，あるいは極性がニュートラルの場合もあり，そのときに上記の対象となる聞き手応答が発話されていれば適切ではないと判断するようにした．

以上の 3 項目を著者らによって評価した．1 つ目と 2 つ目は客観的に評価することができるが，3 つ目に関しては評定者によって焦点語と極性の判断にばらつきが生

表 1 提案システムの各発話の評価

聞き手応答	(1) 破綻なし		(2) 反応あり		(3) 適切さ	
	○	×	○	×	○	×
繰り返し	83	7	79	11	57	33
掘り下げ質問	16	0	13	3	11	5
評価応答	-	-	32	13	31	14
語彙的応答（極性）	-	-	-	-	25	37
計	99 (93.4%)	7 (6.6%)	124 (82.1%)	27 (17.9%)	124 (58.2%)	89 (41.8%)

話者	発話内容	(1)	(2)	(3)
S	ワインは 250 円でもいいかなと思うんですが，日本人としたらお水 250 円は相当当時高かった感じがしますけれど，私はお酒飲めないんで			
E	お酒ですか（繰り返し）	○	×	×
S	ワイン飲めないんで，もう，しょうことなしに水を 250 円で買いました			
E	そうなんですね（語彙的応答）			
S	それとアメリカのニューヨークに行ったときも食事の値段がすごく安いんです			
E	どんな食事ですか？(掘り下げ質問)	○	○	○
S	朝はコーヒーとアップルデニッシュ，今よう思い出したわ（続く）			

話者	発話内容	(1)	(2)	(3)
S	そんなことでいつもお付き合いがあったものですから，高級なお料理屋さんに一度家族みんな連れて行ってくれまして			
E	どんな家族ですか（掘り下げ質問）	○	○	×
S	うちは息子二人と主人とあたしとの四人なんですけれど，向こうは姉たちは夫婦だけで行きました			
E	そうなんですね（語彙的応答）			
S	はい．楽しかったです			
E	素敵ですね（評価応答）	-	○	○
S	はい．兄弟がやっぱり仲良くね，同じ婿同士いうのが仲良くしているって（続く）			

図 7 対話例とシステム発話の評価（S：被験者，E：ERICA，(1) 破綻なし，(2) 反応あり，(3) 適切さ）

じる可能性がある．そこで，この項目については，著者らのうち２名が話し合いながら３対話分を評価し，すべての評価が一致するようにした．その後，残りの対話について，この２名のうち１名が評価した．

結果を表 1 に示す．破綻に関しては，９割以上の応答で破綻がなかった．また，８割以上の応答で被験者の肯定的な反応が観測された．このことから，表層的には自然な傾聴対話が実現できているといえる．一方，適切さに関しては，適切と判定された応答は 6 割弱にとどまった．つまり，４割ほどの応答に関しては改善が必要であり，これらが次節以降で述べる WOZ との差につながっていると考えられる．図 7 に対話例とその評価結果を２つ示す．上の例の最初の発話では，既に「お酒」の話をしているため，焦点語は「お酒」ではなく「飲めない」（あるいは「焦点語なし」）とするのが妥当である．例えば，過去に提案されている焦点アノテーションの基準 [春日 19] では，ここでの焦点は「飲めない」となる．ただし，現状のシステムでは焦点語は名詞または形容詞に限定しているため「飲めない」を焦点語とすることができないので，焦点語候補の幅を広げていくことが今後必要となる．また，そのあとの掘り下げ質問については，被験者が思い出しながら話す場面を引き出せており，内容的にも適切な聞き手応答であるといえる．下の例の最初の発話では,「家族」について掘り下げるよりも「家族みんな」を焦点語として繰り返し応答を行う方が適切といえる．

さらに，音声認識誤り（平均単語誤り率 33.8%）が後段の処理，特に焦点語抽出へどの程度影響したのかを分析した．上記の「適切さ」の評価において，適切ではないと判断された焦点語を用いた応答 38 個（繰り返し 33 個，掘り下げ質問 5 個）について，ユーザが発話していない単語を誤認識して焦点語とした場合（誤受理），本来焦点語であるべき単語を正しく認識できなかった場合（誤棄却）に該当するかを確認した．その結果，繰り返しでは 33 個のうち 15 個（誤受理が 7，誤棄却が 8），掘り下げ質問では 5 個のうち 1 個（誤棄却），両者を合わせると 38 個のうち 16 個（42.1%）が音声認識誤りによる

表 2 提案システムと WOZ による傾聴の主観評価結果（平均と標準偏差，7 段階評価）および t 検定の結果 ($n = 20$)

	評価項目	システム	WOZ	p 値
(ロボットのふるまいへの印象)				
Q1	ロボットが話した言葉は自然だった	5.0 (1.6)	5.9 (0.9)	.003 **
Q2	ロボットはタイミングよく反応していた	4.8 (1.4)	5.6 (1.3)	.022 *
Q3	ロボットはこまめに反応していた	5.5 (0.7)	5.8 (1.0)	.005 *
Q4	ロボットの反応は人間らしかった	4.4 (1.3)	5.2 (1.3)	.008 *
Q5	ロボットの反応はあなたの話を適切に促していた	5.0 (1.4)	5.2 (0.9)	.359
Q6	ロボットの反応の頻度は適切だった	5.1 (1.1)	5.4 (1.1)	.232
(ロボット自体への印象)				
Q7	このロボットとまた話したい	4.6 (1.3)	5.4 (1.5)	.005 *
Q8	このロボットは話しやすい	4.9 (1.3)	5.4 (1.2)	.116
Q9	ロボットは親身だと感じた	4.7 (1.4)	5.6 (1.2)	.004 **
Q10	ロボットは真面目に話を聞いていた	5.6 (1.1)	6.0 (1.1)	.072
Q11	ロボットは集中して話を聞いていた	5.6 (1.1)	5.7 (1.1)	.681
Q12	ロボットは積極的に話を聞いていた	5.4 (1.3)	5.6 (1.1)	.385
Q13	ロボットは話を理解していた	5.0 (1.1)	5.9 (1.4)	.002 **
Q14	ロボットは話に対する関心を示していた	5.2 (1.3)	5.8 (1.2)	.028 +
Q15	ロボットはあなたに対して共感を示していた	5.1 (1.4)	5.8 (1.0)	.015 *
Q16	ロボットは裏で人間が操作していたと思う	3.3 (1.3)	2.9 (1.1)	.286
Q17	ロボットは会話の間の取り方がうまい	4.5 (1.1)	4.8 (1.3)	.209
(会話全体への印象)				
Q18	会話について満足した	4.6 (1.5)	5.3 (1.5)	.012 *
Q19	会話でのやりとりはスムーズだった	4.6 (1.4)	5.3 (1.4)	.002 **

(+ $p < .05$, * $p < .025$, ** $p < .005$)

ものであることがわかった．

5·3 WOZ との比較：主観評価

次に，提案システムと WOZ との比較を行う．主観評価の結果を表 2 に示す．ただし，5·5 節において別の比較を行うため，Bonferroni 法により有意水準を 0.025（=0.05/2）に補正している．まず，提案システムの評定値の平均は，ほとんどの項目で 7 段階（1 から 7）の 4 から 6 に位置しており，高い評価を得られていることがわかる.「ロボットのふるまいへの印象」については，Q1〜Q4 でシステムと WOZ に有意な差がみられたが，Q5 と Q6 ではそれがみられなかった．したがって，被験者はシステムと WOZ の違い（Q1〜Q4）を知覚したが，聞き手応答の適切さ（Q5）や頻度（Q6）については大きな違いを感じなかったといえる．次に,「ロボット自体への印象」については,「また話したい（Q7）」や「親身だと感じた（Q9）」では有意な差がみられたが,「話しやすい（Q8）」ではみられなかった．したがって，話しやすさについては人間と同程度といえるが，人間相手の場合ほど親身さを感じずまた話したいとはならなかったと推察される．傾聴対話システムが継続的に利用されるためには，また話したいとユーザが感じることが重要となる.「ロボット自体への印象」に関する詳細な項目（Q10 以降）を分析したところ,「真面目に話を聞いていた（Q10）」,「集中して話を聞いていた（Q11）」,「積極的に話を聞いていた（Q12）」という項目では有意な差がみられなかった．一方,「話を理解していた（Q13）」と「あなたに対して共感を示していた（Q15）」という項目では有意な差がみられた．また,「話に対する関心を示していた（Q14）」では有意傾向がみられた．したがって，話を聞くための姿勢や基本的な傾聴スキル（Q10〜Q12）については，人間と同程度のものを達成しているが，深い理解を要するスキル（Q13〜Q15）については現在のシステムと人間とではその能力にギャップがあるといえる．また,「人間が操作していたと思う（Q16）」という項目については，WOZ の条件で高い数値が得られることを期待していたが，実際には低くなった．これは，WOZ の条件が存在することを被験者には特に伝えていなかったため，どちらも自律のシステムと解釈された可能性がある．今後，チューリングテストのような直接的な比較（「どちらが人間だったと思いますか」など）を導入していくことも視野に入れる必要がある．最後に,「会話全体への印象（Q18，Q19）」についても有意な差がみられた．これらの項目における WOZ との差を詰めるためには，上記で挙げた高度な傾聴スキルを向上させていく必要がある．

表 3 被験者の発話の分析（平均値と標準偏差）および t 検定結果（$n = 20$）

分析項目	システム		WOZ		p 値
発話時間 [秒] / 分	38.3	(5.5)	37.5	(5.9)	.287
単語数 / 分	107.5	(19.1)	112.0	(23.1)	.177
単語の種類数 / 分	29.0	(4.4)	32.6	(5.1)	.003 **
内容語数 / 分	53.2	(9.8)	55.6	(12.3)	.220
内容語の種類数 / 分	23.3	(4.1)	26.3	(4.4)	.008 **

(** $p < .01$)

表 4 各聞き手応答の頻度（括弧内は 1 対話あたりの平均）

聞き手応答	システム		WOZ	
相槌	1,601	(80.1)	1,573	(78.7)
繰り返し	90	(4.5)	48	(2.4)
掘り下げ質問	16	(0.8)	25	(1.3)
評価応答	45	(2.3)	126	(6.3)
語彙的応答（極性）	62	(3.1)	-	
語彙的応答	325	(16.3)	259	(13.0)
話題提供	12	(0.6)	3	(0.2)
その他（笑い）	-		7	(0.4)

表 5 オペレータ毎の各聞き手応答の頻度（1 対話あたりの平均）

聞き手応答	オペレータ		
	A	B	C
相槌	102.7	42.7	94.0
繰り返し	3.3	3.0	1.0
掘り下げ質問	0.3	2.0	1.3
評価応答	8.2	5.4	5.6
語彙的応答	16.2	14.0	9.1
話題提供	0.0	0.0	0.4
その他（笑い）	0.5	0.0	0.6

5·4 WOZ との比較：ふるまい分析

5·3 節の主観評価結果を補完するものとして被験者の発話についての客観的な分析を行った．ここでは，発話時間，単語数，単語の種類数，内容語数，内容語の種類数について，単位時間あたりの数値をシステムと WOZ との間で比較した．これらの数値が高いほど被験者の発話が促進されたと解釈することができる．単語分割は MeCab[*8] によって行い，内容語は名詞，動詞，形容詞，副詞，接続詞とした．分析結果を表 3 に示す．発話時間，単語数，内容語数には有意な差がみられなかったが，単語と内容語の種類数には有意な差がみられた．つまり，単語と内容語の種類数が話の豊富さの近似であると解釈すれば，WOZ の方がより豊富な内容を引き出せていたと推定できる．また，関連研究で挙げた下岡らの対話実験 [下岡 17] でも同様の分析が行われており，1 分あたりの発話単語数はおおむね 10〜20 個であった．これに対して本実験での提案システムの場合は 1 分あたり 100 個以上である．ただし，本実験と下岡らの実験の条件を比較すると，「過去の旅行」というドメインは共通しているが，本研究は高齢者のみを対象としている．また，使用したロボットはペット型とアンドロイドの違いがある．

システムの発話についても比較を行った．システムと WOZ のオペレータのそれぞれが発話した各聞き手応答の頻度を表 4 に示す．ただし，オペレータの発話について，「語彙的応答」と「語彙的応答（極性）」の区別が難しかったため，どちらも「語彙的応答」にまとめた．また，オペレータによる笑いがいくつか観察され [*9]，それらは「その他」とした．両者の分布を比較すると，システムは「繰り返し」が多かったのに対して，WOZ では「評価応答」が多かった．この違いが，聞き手（ERICA）に対する印象に影響を及ぼしていたと考えられる．つまり，話し手が共感的で強めの応答である「評価応答」を期待していた場面で，システムは比較的弱めの応答である「繰り返し」を発話する傾向にあった．提案システムによる繰り返し 90 件のうち 13 件（20 対話のうち 10 対話）でそのような例がみられた．例えば，「招待してもらったということで余計うれしかったです」という被験者の発話に対して，システムは「招待ですか」と応答していた．別の例では，「ほんとに黒い猫がいたおかげで私たちはすごく癒されたんです」という発話に対して，システムは「黒い猫ですか」と応答していた．これらの応答は 5·2 節の「適切さ」の評価では「適切」とされていたが，「素敵ですね」という評価応答の方がより効果的であったと考えられる．他の例の中で，「適切でない」と判定されたものでは，「結婚式は出られるしそれと旅行もできるということで非常に楽しかったんです」という発話に対して，「結婚式ですか」という応答があった．ここでも「素敵ですね」という評価応答の方が効果的であり，「適切さ」の評価でも「適切でない」から「適切」に改善すると考えられる．

続いて，WOZ による傾聴の例を図 8 に示す．上の例は，上記で述べた評価応答の適切な使用例である．下の例では，掘り下げ質問と評価応答を巧みに使い分けている．この例では，オペレータは掘り下げ質問を用いて話を展開させて，それに対する被験者の感想が発話された時点，あるいは話の一定の区切りの時点で評価応答を発

*8 https://taku910.github.io/mecab

*9 事前のインストラクションにおいて，オペレータには笑わないように伝えたが，わずかでも笑い声が聞こえれば「笑い」が生じたとした．

話者	発話内容
S	パンをオーブンで焼いているときのにおいがね，もうたまらないんですね
E	いいですね（評価応答）
S	バターのにおいとそのパンの焼きあがるこう焼けていくにおい．もう家じゅうににおいが溢れて幸せでした
E	素敵ですね（評価応答）
S	はい．なんか幸せな時間でしたね

話者	発話内容
S	アイヌの人の生活も踊りも見せてもらいました．
E	どんな踊りですか（掘り下げ質問）
S	なんかね，櫂みたいな，船を漕ぐ櫂みたいなのを持って踊ってましたけどね
E	なるほど（語彙的応答）
S	そうでした．それも楽しかったです
E	いいですね（評価応答）

図 8 WOZ での対話例（S：被験者，E：ERICA）

話している．対話の流れを考慮した応答選択は現在の提案システムでは実現できておらず，今後採り入れていきたい点である．このように提案システムとオペレータのふるまいを比較することで，提案システムの改善の方向性を見出すことができる．今後は，言語的なふるまいだけでなく，視線やうなずきといった非言語的なふるまいの影響も分析対象に含めていく予定である．

WOZ の 3 名のオペレータにより使用される聞き手応答の傾向の違いについても分析を行った．表 5 に各オペレータの聞き手応答の 1 対話あたりの平均頻度を示す．分類方法は表 4 の WOZ の場合と同様で，「語彙的応答」と「語彙的応答（極性）」の区別はせず，両者を「語彙的応答」にまとめた．分析の結果，おおよその傾向は 3 名で共通していることがわかった．

5·5 オペレータの制約の有無の比較

ここまでの実験では，WOZ のオペレータの発話の種類に関して，システムが生成できる範囲に留めるように制約を課していた（以降「制約あり」と呼ぶ）．これは提案システムが生成できる聞き手応答が 1 章で紹介した三島ら [三島 03] により挙げられているものをある程度包含できていること，さらには提案システムと WOZ を比較しやすくするためという意図に基づいていた．この妥当性を検証するために，何も制約を課していない WOZ での対話実験も実施した（以降「制約なし」と呼ぶ）．実験の参加者（被験者とオペレータ）および実験条件はこれまでと同じである．傾聴のテーマは「印象に残っている旅行」と「最近食べて美味しかったもの」からランダムに割り当てた．

表 6 に「制約あり」と「制約なし」の WOZ の比較結果を示す．ただし，5·3 節において提案システムと WOZ（制約あり）の比較を行ったため，Bonferroni 法により有意水準を 0.025（=0.05/2）に補正している．多くの項目で「制約なし」の方がやや高い評価を得ているが，有意傾向がみられたのは「話を適切に促していた（Q5）」と「集中して話を聞いていた（Q11）」のみで，その他の項目では有意な差がみられなかった．したがって，被験者による主観評価においては，「制約なし」と「制約あり」では大きな差はなかったといえる．ただし「制約なし」では，以下のようなオペレータのふるまいがみられた．被験者が「私らはもう七十歳越えてますから，そんな量は食べられないんです．そういうことも考えて，大変満足して帰ってきました」と発話したときに，オペレータは「ちょうどいい量だったんですね」と発話していた．また別の例では，「国際免許証を取って私もハンドルを握ってドライブしました」と被験者が発話したときに，オペレータは「どんな気分でしたか」と質問していた．前者はユーザの発話をより理解している，さらに後者は理解するだけでなくユーザの心理状態に焦点をあてて質問をしている．このような機能は現在の提案システムでは実現できておらず，今後の研究開発において採り入れていきたい点である．

6. 結論と今後の課題

本稿では，自律型アンドロイド ERICA のための傾聴対話システムについて述べた．提案システムは，相槌，繰り返し，掘り下げ質問，評価応答，語彙的応答といった聞き手応答を生成する．また，音声認識，音声合成，ターンテイキングシステム，マルチチャネル音響信号処理，人物位置追跡などの技術を，アンドロイドのための音声対話システムとして統合した．

提案システムの性能を評価するために，高齢者との対話実験を実施し，人間（WOZ）と提案システムとを比較した．その際に，提案システムの個々の発話を客観的に評価したところ，ほとんどの発話で破綻がなく，被験者の肯定的な反応を引き出せていたことがわかった．また，聞き手応答としての適切さという観点では，約 6 割の応答が適切と判定された．WOZ との比較では，「話しやすい」，「真面目に話を聞いていた」，「集中して話を聞いていた」，「積極的に話を聞いていた」といった基本的な傾聴スキルに関する評価では，人間のオペレータの場合と遜色がないことがわかった．ただし，「話を理解していた」，「関心を示していた」，「共感を示していた」といったより高度な傾聴スキルに関しては，人間のオペレータとはギャップがあることがわかった．また，「反応は人間らしかった」，「また話したい」，「親身だと感じた」という総合的な項目でも人間のオペレータと比べて有意な差がみ

表 6 WOZ による傾聴（制約あり／制約なし）の主観評価結果（平均と標準偏差，7 段階評価）および t 検定の結果 ($n = 20$)

	評価項目	制約あり	制約なし	p 値
(ロボットのふるまいへの印象)				
Q1	ロボットが話した言葉は自然だった	5.9 (0.9)	6.0 (0.8)	.716
Q2	ロボットはタイミングよく反応していた	5.6 (1.3)	5.9 (0.8)	.349
Q3	ロボットはこまめに反応していた	5.8 (1.0)	5.9 (0.7)	.541
Q4	ロボットの反応は人間らしかった	5.2 (1.3)	5.5 (1.2)	.301
Q5	ロボットの反応はあなたの話を適切に促していた	5.2 (0.9)	5.6 (0.8)	.035 +
Q6	ロボットの反応の頻度は適切だった	5.4 (1.1)	5.6 (0.9)	.591
(ロボット自体への印象)				
Q7	このロボットとまた話したい	5.4 (1.5)	5.1 (1.4)	.204
Q8	このロボットは話しやすい	5.4 (1.2)	5.6 (1.0)	.562
Q9	ロボットは親身だと感じた	5.6 (1.2)	5.5 (0.8)	.629
Q10	ロボットは真面目に話を聞いていた	6.0 (1.1)	6.1 (0.7)	.330
Q11	ロボットは集中して話を聞いていた	5.7 (1.1)	6.1 (0.7)	.025 +
Q12	ロボットは積極的に話を聞いていた	5.6 (1.1)	5.8 (0.8)	.453
Q13	ロボットは話を理解していた	5.9 (1.4)	6.0 (0.6)	.330
Q14	ロボットは話に対する関心を示していた	5.8 (1.2)	5.8 (0.8)	.804
Q15	ロボットはあなたに対して共感を示していた	5.8 (1.0)	5.7 (0.8)	.716
Q16	ロボットは裏で人間が操作していたと思う	2.9 (1.1)	3.2 (1.3)	.287
Q17	ロボットは会話の間の取り方がうまい	4.8 (1.3)	5.0 (1.2)	.545
(会話全体への印象)				
Q18	会話について満足した	5.3 (1.5)	5.4 (1.1)	.267
Q19	会話でのやりとりはスムーズだった	5.3 (1.4)	5.5 (0.8)	.330

(+ $p < .05$)

られた．今回は聞き手（ロボット）の見かけや音声言語以外のふるまいを条件間で統制するために WOZ との比較を実施したが，今後はマルチモダリティの観点でも人間とシステムとを比較していく必要があり，この意味で人間どうしが対面した状態での傾聴対話との比較も実施していくことが求められる．

今後は，上述のギャップを埋めるために，深い理解に基づく聞き手応答の生成に取り組む必要がある．例えば，1 章で挙げた三島ら [三島 03] による聞き手応答のうち「語られた内容を要約する」については現在のシステムでは実現できておらず，発話内容に対するより深い理解が求められる．加えて，近年注目を集めている BERT [Devlin 19] などの事前学習モデルを聞き手応答生成に利用することを検討している．また，提案システムの改善を図りつつ，高齢者施設などの実際の現場で傾聴対話システムの実証実験を実施していきたい．さらに，提案システムは他のロボットや CG エージェントにも実装可能であるため，傾聴対話の効果が聞き手の実体の違いによりどのように変わるのかについても調査したい．

謝　辞

本研究は，JST ERATO（JPMJER1401）ならびに JSPS 科研費（JP19H05691，JP20K19821）の支援を受けた．

◇ 参 考 文 献 ◇

[Butler 63] Butler, R. N.: The life review: An interpretation of reminiscence in the aged, *Psychiatry*, Vol. 26, No. 1, pp. 65–76 (1963)

[Devlin 19] Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K.: BERT: Pre-training of deep bidirectional transformers for language understanding, in *Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pp. 4171–4186 (2019)

[Fujie 05] Fujie, S., Fukushima, K., and Kobayashi, T.: Back-channel feedback generation using linguistic and nonlinguistic information and its application to spoken dialogue system, in *INTERSPEECH*, pp. 889–892 (2005)

[Glas 16] Glas, D. F., Minaot, T., Ishi, C. T., Kawahara, T., and Ishiguro, H.: ERICA: The ERATO intelligent conversational android, in *International Conference on Robot and Human Interactive Communication (ROMAN)*, pp. 22–29 (2016)

[Han 15] Han, S., Bang, J., Ryu, S., and Lee, G. G.: Exploiting knowledge base to generate responses for natural language dialog listening agents, in *Annual SIGdial Meeting on Discourse and Dialogue (SIGDIAL)*, pp. 129–133 (2015)

[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)

[Inoue 16] Inoue, K., Milhorat, P., Lala, D., Zhao, T., and Kawahara, T.: Talking with ERICA, an autonomous android, in *Annual SIGdial Meeting on Discourse and Dialogue (SIGDIAL)*, pp. 212–215 (2016)

[Inoue 20] Inoue, K., Lala, D., Yamamoto, K., Nakamura, S., Takanashi, K., and Kawahara, T.: An attentive listening system with android ERICA: Comparison of autonomous and WOZ interactions, in *SIGdial Meeting on Discourse and Dialogue (SIGDIAL)*, pp. 118–127 (2020)

[石井 13] 石井 カルロス寿憲, 劉 超然, 石黒 浩, 萩田 紀博：遠隔存在感ロボットのためのフォルマントによる口唇動作生成手法, 日本ロボット学会誌, Vol. 31, No. 4, pp. 401–408 (2013)
[Ishi 16] Ishi, C. T., Liu, C., Even, J., and Hagita, N.: Hearing support system using environment sensor network, in *International Conference on Intelligent Robots and Systems (IROS)*, pp. 1275–1280 (2016)
[Johansson 16] Johansson, M., Hori, T., Skantze, G., Höthker, A., and Gustafson, J.: Making turn-taking decisions for an active listening robot for memory training, in *International Conference on Social Robotics (ICSR)*, pp. 940–949 (2016)
[春日 19] 春日 悠生, 井上 昂治, 山本 賢太, 高梨 克也, 河原 達也：ヒューマンロボットインタラクションコーパスへの焦点アノテーションの基準と予備的分析, 人工知能学会研究会資料, SIG-SLUD-B901-03 (2019)
[Kawahara 16] Kawahara, T., Yamaguchi, T., Inoue, K., Takanashi, K., and Ward, N.: Prediction and generation of backchannel form for attentive listening systems, in *INTERSPEECH*, pp. 2890–2894 (2016)
[小林 05] 小林 のぞみ, 乾 健太郎, 松本 裕治, 立石 健二, 福島 俊一：意見抽出のための評価表現の収集, 自然言語処理, Vol. 12, No. 3, pp. 203–222 (2005)
[小林 11] 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, pp. 253–256 (2011)
[小林 12] 小林 優佳, 山本 大介, 土井 美和子：高齢者向け対話インタフェース: 病院スタッフ・患者間の対話モデルを使用したコミュニケーションロボット, 人工知能学会研究会資料, SIG-SLUD-64, pp. 75–80 (2012)
[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, *Language and Speech*, Vol. 41, No. 3, pp. 295–321 (1998)
[Lala 17] Lala, D., Milhorat, P., Inoue, K., Ishida, M., Takanashi, K., and Kawahara, T.: Attentive listening system with backchanneling, response generation and flexible turn-taking, in *Annual SIGdial Meeting on Discourse and Dialogue (SIGDIAL)*, pp. 127–136 (2017)
[Lala 18] Lala, D., Inoue, K., and Kawahara, T.: Evaluation of real-time deep learning turn-taking models for multiple dialogue scenarios, in *International Conference on Multimodal Interaction (ICMI)*, pp. 78–86 (2018)
[三島 03] 三島 徳雄, 久保田 進也：積極傾聴を学ぶ–発見的体験学習法の実際 (2003)
[大竹 14] 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, 知能と情報, Vol. 26, No. 2, pp. 617–626 (2014)
[Raux 09] Raux, A. and Eskenazi, M.: A finite-state turn-taking model for spoken dialog systems, in *Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pp. 629–637 (2009)
[境 14] 境 くりま, 石井 カルロス寿憲, 港 隆史, 石黒 浩：発話者の音声に対応する動作生成と遠隔操作ロボットへの動作の付加効果, 人工知能学会研究会資料, SIG-Challenge-B303, pp. 7–13 (2014)
[Schmidt 86] Schmidt, R.: Multiple emitter location and signal parameter estimation, *IEEE Transactions on Antennas and Propagation*, Vol. 34, No. 3, pp. 276–280 (1986)
[Schröder 12] Schröder, M., Bevacqua, E., Cowie, R., Eyben, F., Gunes, H., Heylen, D., Maat, ter M., McKeown, G., Pammi, S., Pantic, M., Pelachaud, C., Schuller, B., Sevin, de E., Valstar, M., and Wöllmer, M.: Building autonomous sensitive artificial listeners, *IEEE Transaction on Affective Computing*, Vol. 3, No. 2, pp. 165–183 (2012)
[下岡 17] 下岡 和也, 徳久 良子, 吉村 貴克, 星野 博之, 渡部 生聖：音声対話ロボットのための傾聴システムの開発, 自然言語処理, Vol. 24, No. 1, pp. 3–47 (2017)
[Ward 00] Ward, N. and Tsukahara, W.: Prosodic features which cue back-channel responses in English and Japanese, *Journal of Pragmatics*, Vol. 32, pp. 1177–1207 (2000)
[Weizenbaum 66] Weizenbaum, J.: ELIZA–a computer program for the study of natural language communication between man and machine, *Communications of the ACM*, Vol. 9, No. 1, pp. 36–45 (1966)
[ホールファミリーケア 09] ホールファミリーケア 協会（編）：新傾聴ボランティアのすすめ–聴くことでできる社会貢献 (2009)

〔担当委員：岡田 将吾〕

2021 年 5 月 2 日 受理

著 者 紹 介

井上 昂治（正会員）

2015 年 京都大学大学院情報学研究科修士課程修了．2018 年 同研究科博士後期課程研究指導認定退学．博士（情報学）．現在，京都大学大学院情報学研究科助教．音声対話システム，音声言語処理，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．

ラーラー ディベッシュ

2015 年 京都大学大学院情報学研究科博士後期課程修了．博士（情報学）．同年，日本学術振興会外国人特別研究員．現在，京都大学大学院情報学研究科特定研究員．ヒューマンエージェントインタラクション，マルチモーダルインタラクションに関する研究に従事．

山本 賢太（学生会員）

2017 年 京都大学工学部情報学科卒業．2020 年 同大学院情報学研究科修士課程修了．現在，同研究科博士後期課程在学中．日本学術振興会特別研究員（DC1）．音声対話システムに関する研究に従事．情報処理学会 学生会員．

中村 静（正会員）

2012 年，早稲田大学大学院国際情報通信研究科博士後期課程修了．博士（国際情報通信学）．日本学術振興会特別研究員（DC1），大阪大学大学院言語文化研究科助教などを経て，京都大学大学院情報学研究科研究員．音声科学的観点から音声コミュニケーションに関する研究に従事．日本音響学会，ISCA，IPA 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，京都大学大学院情報学研究科研究員などを経て，滋賀県立大学人間文化学部教授．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，質的心理学会 各会員．

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 同大学院工学研究科修士課程修了．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．2020 年 同大学情報学研究科長．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．2006～20 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，特に音声認識およ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，情報処理学会理事，APSIPA 理事，ISCA 理事，APSIPA Transactions on Signal and Information Processing 編集委員長を歴任．IEEE Fellow. 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，ISCA，APSIPA 各会員．日本学術会議連携会員．