

自律型アンドロイド ERICA による傾聴対話

Attentive Listening by Autonomous Android ERICA

山本 賢太* 井上 昂治 Divesh Lala 中村 静 高梨 克也 河原 達也
Kenta Yamamoto, Koji Inoue, Divesh Lala, Shizuka Nakamura,
Katsuya Takanashi, Tatsuya Kawahara

京都大学 大学院情報学研究科
Graduate School of Informatics, Kyoto University

Abstract: We demonstrate an attentive listening system by autonomous android ERICA. It is important for the attentive listening systems to generate the adequate type of responses at appropriate timing depending on the user utterance. The proposed system generates backchannels based on prosodic features. Moreover, using a focus word of the user utterance, the system generates responses such as repeating the focus word, elaborating questions on the focus word, assessment responses, and lexical responses. We select from the above responses according to the utterance timing and the order of priority in order to make the system engaging with real users.

1 はじめに

現在では対話システムは様々な場面で実用化されているが、新たな応用として傾聴対話が検討されている [1, 2, 3]. 傾聴対話とは、話し手の話を聞き手が耳を傾けながら聞く状況を指す。聞き手が適切な反応および共感を示すことで、相手の話をさらに引き出し、話を聞いてもらいたいという話し手の欲求を満たすことが目的である。さらに高齢者のユーザにおいては、話す過程において最近の出来事を思い出すことで、認知機能の維持が期待できる [4]. 今日の社会では、傾聴を行う聞き手役に対する需要が日々増加している。

我々の研究グループでは、自律型アンドロイド ERICA による音声対話に関する研究を進めている [5, 6]. ERICA は人間のような見かけを備えており、マルチモーダルなインタラクションを実現することで、他のロボットやエージェントにはない、人間らしい存在感を伴った対話の実現を目標としている。ERICA が人間らしい聞き手としてふるまうことで、傾聴対話システムの効果をより大きなものにすることが期待される。

本稿では、これまでに我々が研究開発を進めてきたアンドロイド ERICA における傾聴対話システムについて述べる。前回の報告 [7] では、聞き手応答の多様さという観点から、相槌、繰り返し、評価応答、掘り下げ質問、語彙的応答などについて、それぞれの生成方法について述べた。本稿では、実ユーザとの対話における実用性という観点から、上記の各応答をどのよう

なタイミングまた優先順位で出力するかについて述べる。さらに、各応答生成モジュールについても生成方法を見直し、頑健性の向上を図る。以上により、実利用の場面においても話が途切れることがない自然な傾聴対話を実現した。

2 傾聴対話の設定

本システムは、与えられたテーマに沿って 5 分程度の傾聴を行う。テーマは、「今までで一番印象に残っている旅行」、「最近気をつけている健康法」、「最近食べたもので一番美味しかったもの」などを設定している。ただし、ERICA が行う聞き手応答は、特定のドメインに依存しないため、ユーザ発話がどのような内容であっても対応することができる。

3 システム構成

システムの全体構成を図 1 に示す。入力装置はマイクロホンアレイと Kinect v2 である。これらのセンサはアンドロイド本体ではなく、ERICA の周囲に配置する。各処理は主に以下の 3 つに類別できる。

はじめに、音声処理について述べる。マイクロホンアレイに入力されるマルチチャンネル音声信号から、音源方向の定位、およびその方向の音声のみを強調する。また、Kinect v2 により三次元空間内でのユーザの位置を追跡して、音源定位結果と比較することで、ユーザの発話区間を検出する [5]. 音声強調および検出した発

*連絡先：京都大学 大学院情報学研究科 知能情報学専攻
京都市左京区吉田本町
E-mail: yamamoto@sap.ist.i.kyoto-u.ac.jp

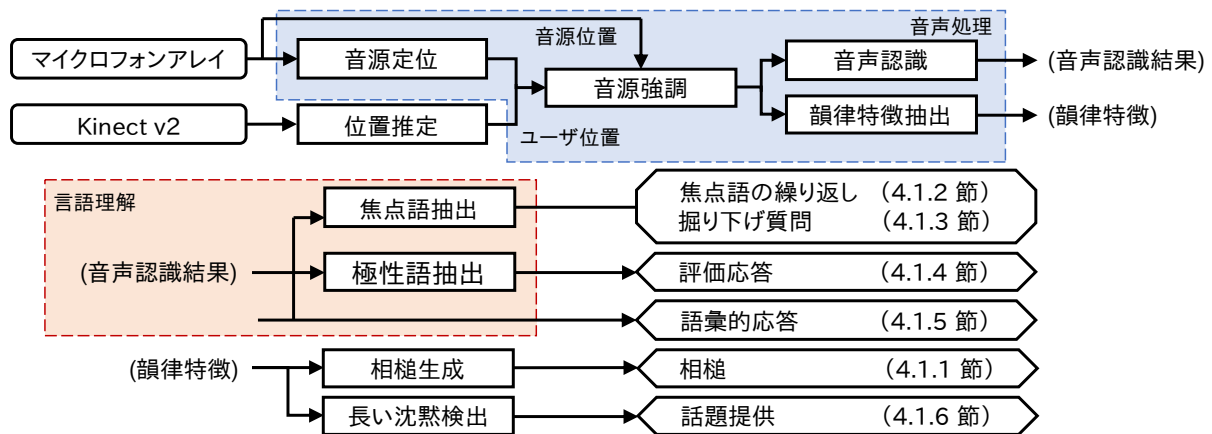


図 1: システム構成

話区間をもとに、音声認識 [8] および韻律特徴抽出 [9] を行う。音声認識のモデルは、単語単位の End-to-end (A2W) 型ニューラルネットワークによるものである。

次に、言語理解について述べる。音声認識結果を形態素解析し、ユーザの発話内で焦点となる単語（焦点語）を抽出する。ここでは、各単語の品詞情報に基づき、焦点語を決定する。具体的には、文末により近い名詞または形容詞が焦点語となる。また、単語の評価極性辞書を用いてポジティブがネガティブかを表す極性語も抽出する。音声認識結果では、単語ごとの信頼度スコアが得られるため、抽出した焦点語および極性語の信頼度スコアが閾値以下の場合には除外する。

最後に、応答生成について述べる。ここでは、ユーザ発話内の焦点語の有無、焦点語と疑問詞との関係、および極性語の有無によって複数の応答を生成する。また、ユーザ発話の途中では、韻律情報をもとに相槌のタイミングを予測して生成する。これらの生成方法については 4.1 節で述べる。また、生成された複数の応答から 1 つを選択し、実際に発話するタイミングについては 4.2 節で述べる。

4 聞き手応答生成

多様な聞き手応答の生成方法について述べる。さらに、各モジュールから生成された応答について、どの応答を実際に発話するかを決定する方法についても述べる。

4.1 応答生成モジュール

自然で効果的な傾聴を実現するためには、聞き手応答の多様性が重要である。ここでは、相槌、焦点語の繰り返し、焦点語に関する掘り下げ質問、評価応答、語

彙的応答、沈黙時の話題提供を聞き手応答として用いる。以下では、各応答の生成方法について述べる。

4.1.1 相槌

相槌は、相手の話の継続や受容を示す聞き手の反応であり、傾聴において重要である。相槌のタイミングは、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 [10]。このように、ユーザの発話途中であっても常に予測を行うことで、早いタイミングで相槌をうつことが可能になる。相槌の形態は、我々が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。

4.1.2 焦点語の繰り返し

抽出した焦点語を用いて、「(焦点語) + ですか」といった発話を生成する。ユーザ発話から焦点語が抽出できた場合にのみ行う。

4.1.3 焦点語の掘り下げ質問

「どんな」、「どの」、「なんの」、「どこの」、「いつの」、「だれの」などの 11 種の疑問詞と焦点語の組合せについて、その N-gram 確率を算出する。閾値以上かつ最大の N-gram 確率をもつ組合せを用いて、焦点語に関する掘り下げ質問を生成する。例えば、焦点語が「カレー」の場合に、「どんな」という疑問詞との組合せが選ばれると、「どんなカレーですか」という掘り下げ質問が生成される。N-gram 確率のモデルは、多様な話題に対応するために、現代日本語書き言葉均衡コーパス (BCCWJ) の全データから学習したものをを用いる。

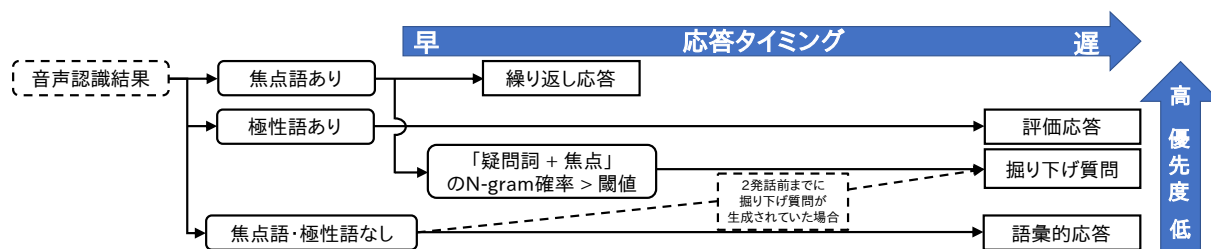


図 2: 聞き手応答生成の流れ

4.1.4 評価応答

ユーザ発話に、感情に関する極性語が含まれる場合に、その極性に対応した評価応答を行う。極性語の抽出には、日本語アプレイザル辞書 [11] を用いた。この辞書の「情動」,「心状」に該当する単語を用いた。各単語にはポジティブまたはネガティブの極性が付与されており、これをそのまま判定結果として用いる。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結果とする。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」、逆にネガティブの場合には、「大変ですね」または「残念でしたね」といった応答を出力する。

4.1.5 語彙的応答

語彙的応答とは、「そうですね」や「なるほど」といった定形表現である。ほとんどのユーザ発話に対して用いることができるため、バックアップの応答としての側面を持つ。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。これにより、ユーザがフィラーを発した場合などに、誤出力の影響を低減することが期待される。

4.1.6 話題提供

ユーザが話す内容について思いつかずに困っている場面では話題提供が必要となる。ここでは、ユーザの長い沈黙に対して、話題提供発話を行う。あらかじめ設定される傾聴のテーマに応じて、用意された話題提供発話が生成される。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？」などと出力される。これについてはテーマに依存して指定する。

4.2 応答選択と発話タイミング

前述の複数の聞き手応答について、どのタイミングでどの応答を実際に発話するのかについて述べる。こ

こでは、図 2 に示すように、発話のタイミングと各応答の優先度という観点で整理する。同じタイミングの場合には優先度が高い応答を、優先度が同じ場合にはタイミングが早い応答を先に発話する。まず、焦点語の繰り返し発話が最もタイミングが早く、焦点語が取得できた段階で早めに出力する。このタイミングは約 1 秒である。また、優先度も最も高い。焦点語の繰り返しは、他の応答に比べて安定的に生成できる傾向から、このような設定にした。他の応答は、焦点語の繰り返しよりも遅いが、すべて同じタイミングで発話する。つまり、焦点語が抽出できた場合には、焦点語の繰り返しの後、ユーザがターンを取らなければ、次のタイミングで他の応答を発話する。他の応答の中で最も優先度が高いのは評価応答である。評価応答はユーザに対して共感を示す役割があり、生成できた場合には発話をするべきである。また、使用する極性辞書は明らかに極性を表しているものであるため、再現率は低いが精度は高いといえる。したがって、優先度を高めに設定した。次に優先度が高いのは掘り下げ質問である。掘り下げ質問も精度が高くなるように N-gram 確率の閾値を高めに設定している。また、一度生成された掘り下げ質問は、しばらく経ってから発話しても概ね問題ない。少し前にユーザが発話した焦点語について、ユーザの話がひと段落した段階で、「さきほどの(焦点語)ですが、どんな(焦点語)ですか？」と発話することができる。したがって、掘り下げ質問に関しては、一度生成されると、ユーザの発話が後に 2 発話以内であれば、その掘り下げ質問は内部に保持しておく。そして、掘り下げ質問よりも優先度が低い語彙的応答が発話されるタイミングで、掘り下げ質問が保持されていれば発話する。優先度が最も低いのは語彙的応答である。語彙的応答はほとんどのユーザ発話に対して対応できるため、このように設定している。また、上記の応答とは別に、ERICA の発話のあとに、ユーザが数秒間沈黙している場合には、話題提供を行う。

本システムでは、対話の進行度合いに応じて応答の種類に制約を課す。対話開始後 1 分間の序盤に対応する場面では、ユーザは話す話題についての概要を説明すると予想される。そのため、ある程度は話す内容は

固まっているはずなので、無理に掘り下げ質問などを生成する必要はない。また、話の導入部分であるため、評価応答の必要性も低い。したがって、相槌、繰り返し応答、語彙的応答のみを用いる。対話中盤では、全ての応答を用いる。対話終盤の残り1分になると、終了時間が迫っていることを伝える発話を行い、まとめとなる発話を促す。また、対話を終了させるため、話題提示は行わない。これにより、ERICAが対話を無理に打ち切ることなく、ユーザの方から自然に終了することが期待できる。

5 対話例

対話例を以下に示す。また、対話動画¹も公開している。以下では、**U**はユーザ、**E**はERICAの発話をそれぞれ表す。

- E1** 今日は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。
- U1** はい。今日は、去年の春に京都の祇園にある公園に行った話をします。
- E2** 公園ですか。(焦点語の繰り返し)
- U2** はい。
- U3** 桜の季節だったので、とても賑わっていたんですよ。
- E3** そうですか。(語彙的応答)
- U4** そうなんですよ。
- E4** どの公園ですか?(焦点語の掘り下げ質問)
- U5** 八坂というところにある円山公園という所です。
- E5** 円山公園ですか。(焦点語の繰り返し)
- U6** そうです。
- E6** はい。(短い語彙的応答)
- U7** そこは有名な枝垂れ桜の木があったりするので、
- E7** はい。(相槌)
- U8** その周りを散歩しました。
- E8** 散歩ですか。(焦点語の繰り返し)
- U9** はい。とても楽しかったです。
- E9** 素敵ですね。(評価応答)
(長い沈黙)
- E10** そこでは他にどんなことをしましたか?(話題提供)

6 おわりに

本稿では、自律型アンドロイドERICAにおける傾聴対話システムについて述べた。聞き手応答生成では、ユーザ発話に応じて、相槌、焦点語の繰り返し、焦点語に関する掘り下げ質問、評価応答、語彙的応答を生成

する。これらの応答を選択する基準として、タイミングと優先度という観点から基準を整理した。これにより、実ユーザとの対話において一定の頑健性を実現した。今後は、実ユーザによる評価実験を予定している。

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 JPMJER1401 の支援を受けて実施した。

参考文献

- [1] Marc Schröder, Elisabetta Bevacqua, Roddy Cowie, Florian Eyben, Hatice Gunes, Dirk Heylen, Mark ter Maat, Gary McKeown, Sathish Pammi, Maja Pantic, Catherine Pelachaud, Björn Schuller, Etienne de Sevin, Michel Valstar, and Martin Wöllmer. Building autonomous sensitive artificial listeners. *IEEE Trans. on Affective Computing*, Vol. 3, No. 2, pp. 165–183, 2012.
- [2] Martin Johansson, Tatsuro Hori, Gabriel Skantze, Anja Höthker, and Joakim Gustafson. Making turn-taking decisions for an active listening robot for memory training. In *ICSR*, pp. 940–949, 2016.
- [3] 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.
- [4] 大武美保子. 認知症の予防と支援に役立つ人工知能と高齢者とともにつくる認知症予防支援サービスの開発. 人工知能, Vol. 31, No. 3, 2016.
- [5] Koji Inoue, Pierrick Milhorat, Divesh Lala, Tianyu Zhao, and Tatsuya Kawahara. Talking with ERICA, an autonomous android. In *SIGDIAL*, pp. 212–215, 2016.
- [6] Tatsuya Kawahara. Spoken dialogue system for a human-like conversational robot ERICA. In *IWSDS*, 2016.
- [7] 井上昂治, Divesh Lala, Pierrick Milhorat, 石田真也, 趙天雨, 高梨克也, 河原達也. 自律型アンドロイドERICAにおける多様な聞き手応答を用いた傾聴対話. 人工知能学会研究会資料, SLUD-B508-11, 2017.
- [8] Sei Ueno, Hirohumi Inaguma, Masato Mimura, and Tatsuya Kawahara. Acoustic-to-word attention-based model complemented with character-level CTC-based model. In *ICASSP*, 2018.
- [9] Carlos T. Ishi, Chaoran Liu, Jani Even, and Norihiro Hagita. Hearing support system using environment sensor network. In *IROS*, 2016.
- [10] Divesh Lala, Pierrick Milhorat, Koji Inoue, Masanari Ishida, Katsuya Takanashi, and Tatsuya Kawahara. Attentive listening system with backchanneling, response generation and flexible turn-taking. In *SIGDIAL*, pp. 127–136, 2017.
- [11] 佐野大樹. 『日本語アプレイザル評価表現辞書—態度評価編— jappraisal 辞書 ver1.0』. 言語資源協会発行, 2011.

¹デモ動画 <https://youtu.be/qnYS8JcqANI>