

聞き手の反応をモニターしながら説明する音声ガイドシステム

An automated speech guidance that is aware of listener's response

森本洋介* 森大毅

Yosuke Morimoto Hiroki Mori

宇都宮大学

Utsunomiya University

Abstract: We aim to realize an automated speech guidance that monitors listener's response such as backchannels and fillers, and controls the timing of explanation. As long as regular backchannels are detected, the system continues to explain. Contrastively, if backchannels are not detected for a while, the system makes an utterance to check whether the user understands. In addition, when a filler is detected, the system stops talking immediately and waits for user's utterance. We built a prototype system, and verified its operation.

1 はじめに

人間同士の説明場面では、たとえ話し手が聞き手に一方的に情報を伝える場面であっても、話し手は聞き手の反応をたえず監視し、発話計画を動的に調整している。一方、音声ガイドは一方的に説明するだけで、聞き手の反応をモニターすることは一切ない。

本研究では、相槌およびフィラーといった聞き手の反応をリアルタイムで検出することで聞き手の理解状態をモニターし、それに応じて説明のタイミングを調整する音声ガイドシステムを提案し、試作と動作検証を行なった。提案システムにより、効率が高く、心的負荷の小さい音声ガイドの実現を目指す。

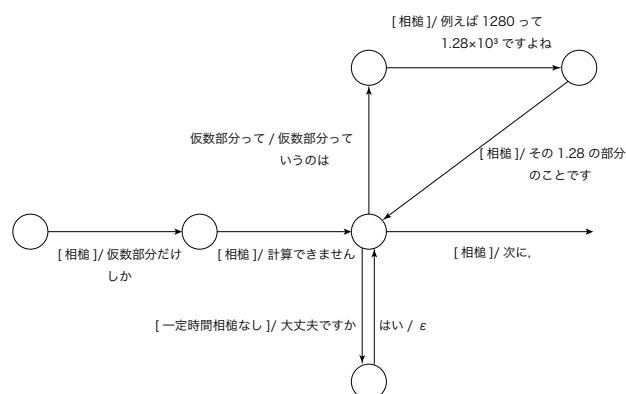


図 1: 状態遷移図 (一部)

2 システムの概要

2.1 聞き手の反応に対するシステムのふるまい

聞き手反応に関する過去の研究 [2][3] に基づき、システムのふるまいを以下のように定めた。

- 相槌が検出された場合、説明をそのまま続ける
- 相槌が検出されなかった場合、ユーザの相槌を少し待つ
- 長い間相槌が検出されない場合、「大丈夫ですか？」と発話し応答を促す

- フィラーが検出された場合、システムは即座に発話を中断してユーザの発話を待つ

また、ユーザが直前の説明に現れた用語を繰り返した場合 (例: 仮数部分って) や、「何それ」などの明示的な質問をした場合には、システムはその部分に対する追加説明を行う。

図 1 にシステムの状態遷移図の一部を示す。各アークは有限状態トランスデューサ (FST) の入力記号/出力記号を表す。ユーザが相槌を打っている間はシステムはメインの説明を続ける。一定時間入力がなかった場合も、相槌があった場合と同じ状態遷移を行う。ただし、しばらく相槌がなかった場合は「大丈夫ですか」と発話しユーザの応答を待つ。また、フィラーが入力された場合には、次のユーザの発話まで状態遷移しない。

*連絡先: 宇都宮大学
栃木県宇都宮市陽東 7 丁目 1-2
E-mail: morimoto@speech-lab.org

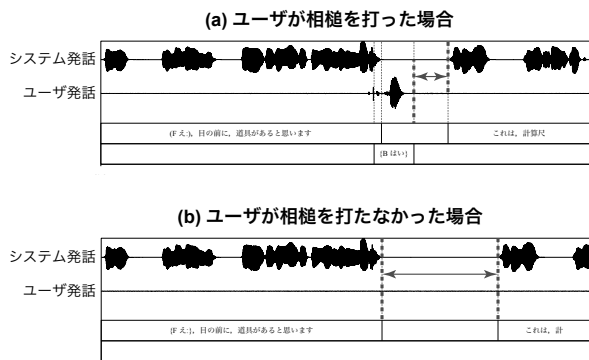


図 2: ユーザの相槌の有無によるシステム発話のタイミングの違い

MMDAgent [1] を用いて音声ガイドシステムを構築した。MMDAgent には、聞き手反応検出モジュールを追加し、音声認識モジュールと並列に動作させた。聞き手反応検出モジュールは、音声活動を検出すると端末フリーの聞き手反応分類を行い、相槌またはフィラーに分類されたならば、音声区間終了の検出を待たずに聞き手反応があったことを知らせるメッセージをイベントキューに送る。音響特徴量には、パワー、MFCC などのセグメント単位の統計量を用い、分類には計算尺使い方説明タスクコーパス [4] で学習した 16 混合 GMM を用いた [5]。

2.2 対話スクリプトの作成

MMDAgent の対話管理は FST により状態遷移を記述したファイルに従って行われる。2.1 で述べた動作を実現する FST ファイルは複雑で、手作業で記述するには労力が要る。そこで、説明内容を書いたテキストファイルを入力すれば、聞き手反応を監視しながら説明するための FST ファイルを出力できるスクリプトを作成した。

また、聞き手に自然な聞き手反応を出させるため、なるべく人間同士の説明場面に近づけるようにした。具体的には、「えーと」などのフィラーを発する。それに合わせて 3D エージェントが軽く視線をそらす、などの工夫をした。

3 動作例

図 2 にユーザの相槌の有無によってシステム発話のタイミングが変わる例を示す。(a) では、ユーザの相槌が検出されたため、システムはすぐに次の発話を開始している。一方 (b) では、システムはユーザの相槌を

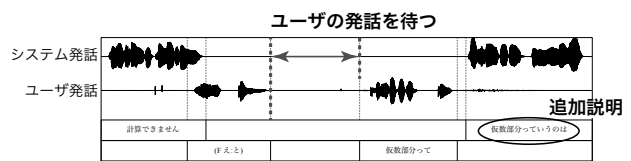


図 3: フィラーによりシステムがユーザの発話を待つ様子

少しの間待っている。このように、ユーザの理解状況に応じたテンポの良い説明が実現できている。

図 3 にフィラーによりシステムがユーザの発話を待つ例を示す。フィラーが検出されるとシステムは即座に発話を中断して、ユーザの発話を待っている。その後ユーザの質問に応じて追加説明を行っている。このように、相槌以外の聞き手反応や質問に応じて柔軟に説明戦略を変えることができています。

4 まとめ

本稿では、聞き手の反応によって説明のタイミングを調整する音声ガイドシステムを提案した。聞き手の反応として相槌とフィラーに着目し、相槌は発話タイミングの調整と話を聞いているかの確認、フィラーはシステム発話に対する割り込みとして機能する。今後は、本システムが有用であるかを調べるための実験と評価を行う予定である。

参考文献

- [1] 李晃伸, 大浦圭一郎, 徳田恵一: 魅力ある音声インタラクションシステムを構築するためのオープンソースツールキット MMDAgent, 信学技報, Vol. 111, No. 364, pp. 159–164 (2011).
- [2] Den Y., Koiso H., Takanashi K., and Yoshida N.: Annotation of response tokens and their triggering expressions in Japanese multi-party conversations, *In Proc. LREC 2012*, pp. 1332–1337 (2012).
- [3] 山根智恵: 日本語の談話におけるフィラー, くろしお出版 (2002).
- [4] 藍原瞭, 他: 話し手の説明戦略に影響する聞き手行動の分析, *HCG シンポジウム*, pp. 436–440 (2016).
- [5] 森本洋介, 森大毅: イベント継続時間モデルを用いた聞き手反応の検出, 音講論, pp. 905–906 (2019).