

## 原稿の有無が説明者の発話と被説明者の反応に与える影響の分析\*

☆高松屋友翼, 森大毅 (宇都宮大)

## 1 はじめに

人間同士の対話では、相槌や復唱が頻繁に見られる。話し手は説明を繰り返す、補足する、言い換える、話すスピードを調整する方法で、聞き手からのフィードバックを説明に反映させる。このような聞き手反応 [1] は、人間同士の対話を円滑にする重要な役割を担っている。それに対して対話システムとの対話では、聞き手がシステムに対して相槌や復唱などで反応する場面はほとんど見られない。

本研究の目的は、対話システムとの対話において聞き手反応が見られない原因を明らかにするとともに、その原因を取り除いた対話システムとの対話で聞き手反応が出現するようになることを確認することである。

本論文では、人間同士の対話にだけ存在する対話の要素の有無を操作することを目的とした対話実験のデザインについて述べ、対話の要素と聞き手反応の出現頻度との関係を調べた結果を報告する。

## 2 対話実験のデザイン

本研究では、話し手が人間の場合と対話システムの場合との相異点を、以下に例示する要素の有無で特徴づける。

**H** 人間の声 (肉声) で話しているように聞こえること

**N<sub>1</sub>** フィラー、言い誤りなどの非流暢な発話

**N<sub>2</sub>** 笑い、叫び、感情表出系感動詞などのノンバーバル行動

**V** 自発音声の言語的特徴 (断片的、倒置などの破格)

**P** 自発音声の韻律的特徴

実験では、システムがユーザにニュース記事の内容を説明する場面を想定した人間同士の対話を収録・分析した。その際、上述の要素の有無を操作するための道具立てとして、説明用原稿の有無および種類の違いを利用し、以下に述べる3通りの条件を設定した。

条件① HNV では、ニュース記事の説明を原稿なしで行う。システム役の説明者は記事の内容を暗記しておくが、手元に用意したメモを確認することは許す。条件① HNV では、日常会話でエピソードを紹介する場面に類似した対話になると予想される。また、

入力 (聞き手反応)	出力 / 状態遷移
相槌・無反応	⇒ φ / 主計画継続
上がり口調の復唱	⇒ 復唱単語の繰り返し / 主計画継続
想定内の内容に関する WH 質問	⇒ φ / 副計画に遷移
想定外の内容に関する WH 質問	⇒ 直前の文・単語の繰り返し / 主計画継続
想定内の内容に関する Y/N 質問	⇒ 「そうだね」「違うよ」 / 主計画継続
想定外の内容に関する Y/N 質問	⇒ 「わからない」 / 主計画継続

Fig. 1 条件② H, 条件③ HV におけるシステム役の説明者の発話/行動規則

Table 1 システム役の説明者に課す条件

	人間らしさ	非流暢性 / ノンバーバル行動	自発音声の 言語的特徴	自発音声の 韻律的特徴
	H	N	V	P
① HNV	○	○	○	○
② H	○	×	×	?
③ HV	○	×	○	?

人間同士であるため要素 H を有し、自発音声であるため要素 N<sub>1</sub>, N<sub>2</sub>, V を有すると想定される。

条件② H では、ニュース記事の説明を原稿ありで行う。原稿はニュース記事を要約し口語化したものである。システム役の説明者は、ユーザからの質問などの反応がない限り原稿に忠実に発話し、反応があった場合には Fig. 1 に示す対話制御規則に従って次発話の内容と行動を決める。特に、想定された WH 質問があった場合には、用意しておいた回答や補足情報を説明してから元の流れに戻る。これは、ユーザの能動的情報要求に応じて情報の補完を行う音声対話システム [2] の動作の人間による模倣である。条件② H では、このような対話システムに説明を受ける場面に類似した対話になると予想される。また、条件① HNV と異なり、システムの発話に非流暢性は無く、ノンバーバル行動は無く、自発音声の言語的特徴は持たない。

条件③ HV では、条件② H と同じく原稿ありで説明するが、原稿の作り方が異なる。条件③ HV の原稿は、条件① HNV で収録した対話の書き起こしを基に、フィラー、笑い、感情表出系感動詞を削除して作成した。条件③ HV では、自発音声と同じように発話が断片的・破格であるが、原稿に頼った発話であるため自発音声と異なる韻律的特徴を持つ可能性がある。以上の条件をまとめて Tab. 1 に示す。

26 名のユーザ役の被説明者を対象に対話実験を行った。システム役は第 1 著者である。各被説明者には 3 条件のうち 1 つを無作為に割り当てた。各条件の平

\*Effects of scripted/unscripted tutorial on tutor utterance and tutee response. by TAKAMATSUYA, Yusuke, MORI, Hiroki (Utsunomiya University)

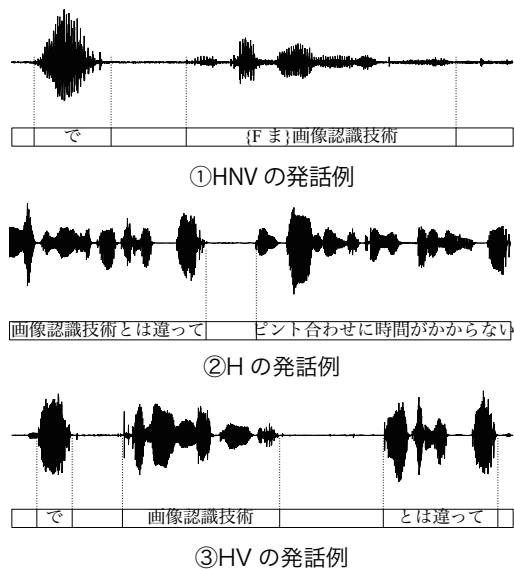


Fig. 2 各方法の発話例

Table 2 説明者の発話の特徴

	①HNV	②H	③HV
1分間あたりのフィラー数	3.80	0.120	0
発話間ポーズ	0.716	0.662	0.931
1発話あたりの文節数	1.86	3.85	1.71
1対話あたりの発話数	324	107	200

均対話時間は、条件① HNV:11.2分、条件② H:6.6分、条件③ HV:7.7分である。Figure 2に説明者の発話の例を示す。① HNVと③ HVの発話が断片的であること、① HNVの発話にフィラーが含まれていることがわかる。

### 3 実験結果

#### 3.1 説明者の発話の特徴

説明者の発話の特徴(1分あたりのフィラー数、発話間ポーズの平均、1発話あたりの文節数、1対話あたりの発話数)をTab. 2に示す。

条件② H、条件③ HVではフィラーが少なく、要素Nを狙い通りに排除できていることがわかる。また条件② Hでは、1発話あたりの文節数が多く、書き言葉を読み上げた時のように情報量が多いことがわかる。

#### 3.2 被説明者の聞き手反応

1分あたりの聞き手反応の回数の平均は、条件① HNV:10.5回、条件② H:5.96回、条件③ HV:22.0回であった。条件① HNVを基準とした聞き手反応の回数は、条件② Hでは有意に少なく( $p = 0.002$ )、条件③ HVでは有意差がなかった( $p = 0.52$ )。Figure 3に被説明者の聞き手反応(相槌、感情表出系感動詞、

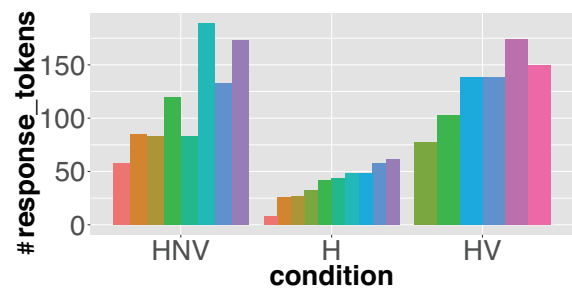


Fig. 3 各被説明者ごとの聞き手反応の回数

復唱)の回数を示す。

#### 3.3 考察

条件② Hでは、途切れることのない流暢な説明が、ユーザにとって機械的な発話に感じられ、相槌などの反応をしても無意味な非社会的存在と認識されたことが聞き手反応減少の原因となっている可能性がある。条件② Hでは、発話数が他の条件に比べ少なくなっているため、単に反応できる場所が少なかっただけの可能性も考えられる。しかし、条件① HNVの発話数が条件③ HVに比べ多いのにも関わらず、これらの条件間で聞き手反応の数に違いは見られない。このことから、聞き手反応減少の原因を発話数の違いだけによって説明することはできないと考えられる。

また、条件② Hと、条件③ HVの間で聞き手反応の回数に違いが見られることから、自発音声の言語的特徴である断片的な発話や、話し言葉独特の表現が、相手が機械ではなく反応をしなければならない社会的存在であるとユーザに認識させる助けになった可能性がある。よって、自発音声の言語的特徴を有する合成音声システムに実装することで聞き手反応をするようになる可能性があると考えられる。

### 4 おわりに

話し手が人間の場合にのみ存在する要素のうち非流暢性/ノンバーバル行動および自発音声の言語的特徴有無に着目して、その要素と聞き手反応の頻度との関係を調べた。条件① HNVを基準とした聞き手反応の回数は、条件② Hでは有意に少なく、条件③では有意差がなく、自発音声の言語的特徴が聞き手反応の頻度に関連していることを示唆する結果を得た。

#### 参考文献

- [1] Den et al., Proc. LREC 2012, pp. 1332-1337, 2012.
- [2] 高津 他, 人工知能学会論文誌, pp. DSH-C 1-24, 2018.