一般社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

HCGシンポジウム2016
HUMAN COMMUNICATION GROUP SYMPOSIUM 2016
HCG2016-I-2-1

話し手の説明戦略に影響する聞き手行動の分析

藍原　瞭† 　森　大毅†

† 宇都宮大学大学院工学研究科　〒 321–8585 栃木県宇都宮市陽東 7 丁目 1–2
E-mail: †{aihara,hiroki}@speech-lab.org

あらまし　現状の音声ガイダンスには, 被説明者の反応を一切考慮せず行われるため, 理解が難しいという問題がある. 本研究は, 対話システムが行う音声ガイダンスを, ユーザーの理解状態をモニタリングしながら行うことで, 説明を分かりやすく円滑なものに改善することを目的としている. 本論文では, 対話システムが音声ガイダンスを行う際, 被説明者の理解状態を推定するために有用であると考えられる言語行動について検討を行った. 一方の人間がもう一方の人間に道具の使い方を説明する様子を収録し, 話し手が聞き手のどのような反応を手掛かりにして理解状態を推定するのか, また, それを利用してどのように説明計画を漸次的に変えていくのかを観察し分析した.

キーワード　音声ガイダンス, タスク指向型音声対話, 聞き手反応, 反応潜時, ユーザ状態

Analysis of response tokens that affect speaker's explanation strategy

Ryo AIHARA† and Hiroki MORI†

† Utsunomiya University,　Yoto 7–1–2, Utsunomiya-shi, Tochigi, 321–8585 Japan
E-mail: †{aihara,hiroki}@speech-lab.org

Abstract　This study aims to improve the usability of spoken guidance by machine. Existing spoken guidance system do not take into account the response of listeners. In this paper, we have examined the verbal behavior that are considered to be useful for estimating user states. We have recorded dialog of a task where a participant is to explain how to use a tool to another person. Then we analyzed the way that speakers estimate the another person's state by means of the clues such as reaction of the lister, and that they dynamically modify the explanation plan with the clues.

Key words　speech guidance, task-oriented dialog, response tokens, response time, user state

1. はじめに

音声ガイダンスは, 券売機, 自動現金預払機, 運送業の再配達自動受付システムなどの社会インフラから HDD レコーダなどの家電製品に至るまで, 社会のあらゆる場面で使用されている. しかし, 現状の音声ガイダンスは時として十分に機能せず, 理解が困難だったり, 逆に説明が冗長でユーザをいらつかせたりすることがある.

現状の対話システムによる説明と人間による説明との大きな相異は, 前者が聞き手の反応を一切無視して一方的に話し続ける点にある. 人間同士のインタラクションでは, 例え説明者が聞き手にほとんど一方的に情報を伝える場面であっても, 説明と平行して聞き手の反応を常にオンラインで監視し, 状態をモニタリングしている.

聞き手の反応を利用して説明戦略を動的に変更するシステムを実現するためには, 聞き手の言語行動やノンバーバル行動から多岐にわたる特徴を抽出し, ユーザーの理解状態を総合的に推定する必要があると考えられる. 藤江ら [1] は, システムとの対話中にユーザが生成する否定的な相槌や聞き返しといった短い発話を認識した場合に, 既に説明した内容を補足する説明を挿入することを提案している. また, ユーザ側に対話システム側の状態をフィードバックするため, 相槌, 復唱や共同補完などの聞き手行動を生成させる試みもある [2]. しかしながら, ことばによる説明場面において, 説明者が聞き手発話に含まれるどのような特徴を聞き手の理解度の推定に利用し, またそれを説明戦略にどのように反映させているかについて直接研究した事例はほとんどなく, またそのような研究を可能にするような公開コーパスも存在しない.

本研究は, 対話システムの音声ガイダンスを, ユーザーの理解状態をモニタリングしながら行うことで, 説明を分かりやすく円滑なものに改善することを目的としている. そのためには対話システムが, ユーザーの聞き手としてのどのような行動に着目すればユーザーの理解状態を取得できるのか解明する必要がある.

そこで, まずは人間同士によるガイダンス形式の対話を収録し, 聞き手の理解状態がいかなる観察可能な聞き手行動として

現れるのかを調べる．これらの特徴は, 説明者の行動に無意識的に反映されている面も十分に考えられる．聞き手の理解状態が説明者の行動にどのような変化を与えているかを調べることは, 対話システムの説明をより円滑で自然なものとするための有用な知見をもたらすことが期待できる．

本論文では, 説明戦略の動的な変更の例として，次発話の生起タイミングに着目した．具体的には, 被説明者の相槌または復唱の生起タイミングが，説明者の次発話に影響を及ぼしているか否かを調べた．この目的のため，収録した対話音声中の被説明者の相槌および復唱の部分に着目し, その生起タイミングと説明者の次発話の生起タイミングとの関係を統計的に調べた．

2. 対話音声の収録

2.1 タスク設計

本研究では, 説明場面における話し手及び聞き手の行動を, 無意識下のものも含めて収録し, 観察及び統計的処理によってモデル化することを可能にすることが求められる．本研究では以下の要件を重視した．

（1） 基本的に話し手から聞き手への一方的な説明

（2） 高い自発性を持つ

（3） 説明談話を構成する談話単位がある程度の長さを持ち (数発話から十数発話程度), 一度にされても理解が困難だが, 順を追えば全体を理解することも十分に可能な程度の難易度を持つ

（4） 聞き手が受けた説明を正しく理解しているか否かを検証可能

(1) は問題を単純化して分析を容易にするため, (2) は対話参与者の無意識下の行動を誘発するため, (3) は種々の聞き手行動を自然に促し, また話し手に説明の仕方をリアルタイムで計画・更新することを促すため, (4) は聞き手に理解しようとする動機付けを与えるための要件である．

以上の要件を満たすものとして, 計算尺使い方説明タスクを考案した．説明者と被説明者は区切られた防音室内におり, お互いの手元は見えず, 相手からの音声はヘッドセットを通して聞こえる状態になっている．説明者と被説明者の手元には計算尺が置いてあり, 説明者は被説明者に計算尺の使い方を説明する．説明の内容は「C 尺と D 尺を使った掛け算」,「CI 尺と D 尺を使った掛け算」,「C 尺と CI 尺と D 尺を使った連続した掛け算」の 3 つについて行なった．説明者は事前に使用法を十分に修得し, 説明の際には手元に説明書やメモを一切置かなくて済むようにした．

計算尺使い方説明タスクでの対話収録の結果は以下の通りである．

- 説明者：1 名
- 被説明者：7 名
- セッション数：8 回
- 収録時間： 1 セッションあたり 10 分程度

収録した対話には, 当初期待した特徴を多く見ることができた．特に, 最小の談話単位がある程度大きくなり, 典型的には連続した数発話が説明-了解の単位となっていた点は当初の狙い通りであった．

表 1　聞き手反応 [3]

聞き手反応	タグ	例
応答系感動詞	B	はい, うん, ああ 等
感情表出系感動詞	E	あっ, えっ, へー, ふうん 等
語彙的応答	L	そうですね, なるほど 等
評価応答	A	すごい, 面白い 等
復唱	R	説明者の発話の全てまたは一部の繰り返し
共同補完	C	話し手の発話を聞き手が予測し補ったもの

表 2　聞き手反応と先行発話の関連付け

No	StartTime	tier	text	EndTime	refNo
1	3.600	L	はい	3.760	0
2	5.291	L	(F えと)	5.725	0
3	6.150	L	机の上に	6.938	0
4	7.403	R	{B はい }	7.570	3
5	8.311	L	道具が	8.701	0
6	9.054	L	あると思いますけど	10.116	0
7	10.195	R	{B はい }	10.389	5
8	11.237	L	これは計算尺	12.615	0
9	12.860	L	という名前の計算用具です	14.503	0
10	14.595	R	はい	14.867	0
11	15.620	L	計算尺	16.246	0
12	16.295	L	(F え)	16.734	0
13	16.359	R	{B はい }	16.551	11
⋮	⋮	⋮	⋮	⋮	⋮

3. アノテーション

200 ms 以上の無音区間の前後を境界として発話単位を設定した．付与した情報には各発話の開始時刻, 終了時刻, 文字情報及び表 1 に示す聞き手反応 [3] が含まれる．

3.1 聞き手反応と先行発話の関連付け

聞き手の相槌や復唱は必ずしも話し手の直前の発話によって誘発されたものとは限らない．そのため相槌や復唱が説明者 L のどの発話から誘発されたかを明記するために refNo (参照番号) の項目を追加した．この項目には誘発元になった説明者の text (発話内容) の No (番号) が記入されている．タグが付与されていない text の refNo には 0 が記入されている．

refNo の付与作業例を表 2 に示す．話し手の複数の発話を聞き手が一度に復唱する場合は参照される複数の発話のうち最後に現れる発話のみを refNo に記述した．

4. 聞き手反応の統計的分析

4.1 聞き手反応の出現頻度

収録した 8 セッションにおける各聞き手反応の出現回数を表 3 に示す．相槌 (B) が最も多く, 次いで復唱 (R) が多く出現した．復唱の対象となったのは, 道具を構成する部品の名称に対してや道具の操作の指示に対するものが多かった．その一方で評価応答 (A) と共同補完 (C) は出現しなかった．また, 感情表出系感動詞 (E) と語彙的応答 (L) は相対的に少なかった．

4.2 反応潜時とポーズ長の関係

反応潜時とポーズ長の概要を図 1 に示す．ポーズ長とは話し

表 3　聞き手反応の出現回数

聞き手反応	タグ	出現回数
相槌	B	188
感情表出系感動詞	E	24
語彙的応答	L	23
評価応答	A	0
復唱	R	68
共同補完	C	0

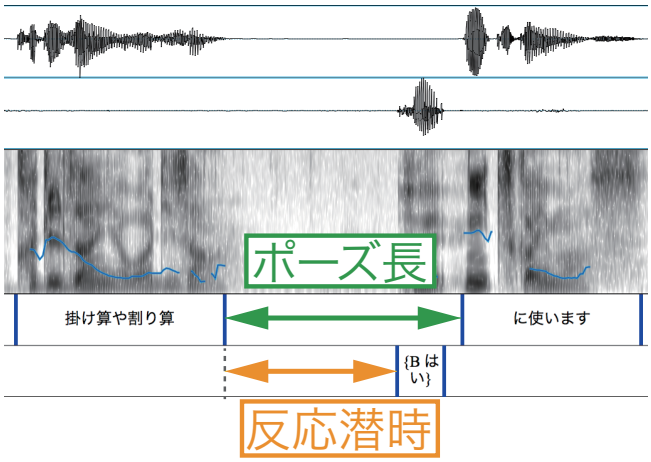

図 1　反応潜時とポーズ長

手の発話の終わりから次の発話までの時間のことを指す．反応潜時とは話し手の発話の終わりから聞き手の応答の始まりまでの時間を指す．図 1 の場合は，話し手の発話である「机の上に」と「道具が」の間の時間がポーズ長であり，「机の上に」から聞き手の「はい」までの時間が反応潜時となる．

図 2 に収録音声中に見られた特徴的なポーズ長の変化の例を示す．(a) は典型的な反応潜時の例である．一方，(b) は，聞き手の応答が遅い，すなわち反応潜時が長いため，話し手の次発話の開始が遅くなった，すなわちポーズ長が長くなった例を示している．これらの 2 つの例から，反応潜時に応じてポーズの長さも変化していると考えられる．(c) は一度は聞き手の応答を待ったものの聞き手が応答する前に説明を続けている例を示している．もし話し手が聞き手応答のタイミングをモニターしながら次発話のタイミングを調整しているならば，聞き手の応答が遅いと説明者の次発話の開始は遅くなるはずである．反対に，もし説明者のポーズ長が聞き手応答のタイミングと無関係ならば，説明者は聞き手の反応を省みずに自分のペースで発話タイミングを決めていると結論できる．

以上を検証するため，収録した対話音声中に現れた相槌と復唱に対して反応潜時とポーズ長の関係を調べた．もし話し手が聞き手応答のタイミングをモニターしながら次発話のタイミングを調整しているならば，これらは正の相関を示すはずである．

図 3 に相槌に関する散布図，図 4 に復唱に関する散布図を示す．横軸が反応潜時，縦軸がポーズ長である．相槌に関しては相関係数が 0.493 となっており，反応潜時とポーズ長の間には正の相関があることがわかる．復唱に関しては相槌と比較すると相関が弱いが，聞き手の反応潜時が長くなるのに伴い，短いポーズが存在しなくなっていることがわかる．これは被説明者の理解状態を説明にフィードバックするために説明者が反応を待っているためだと考えられる．

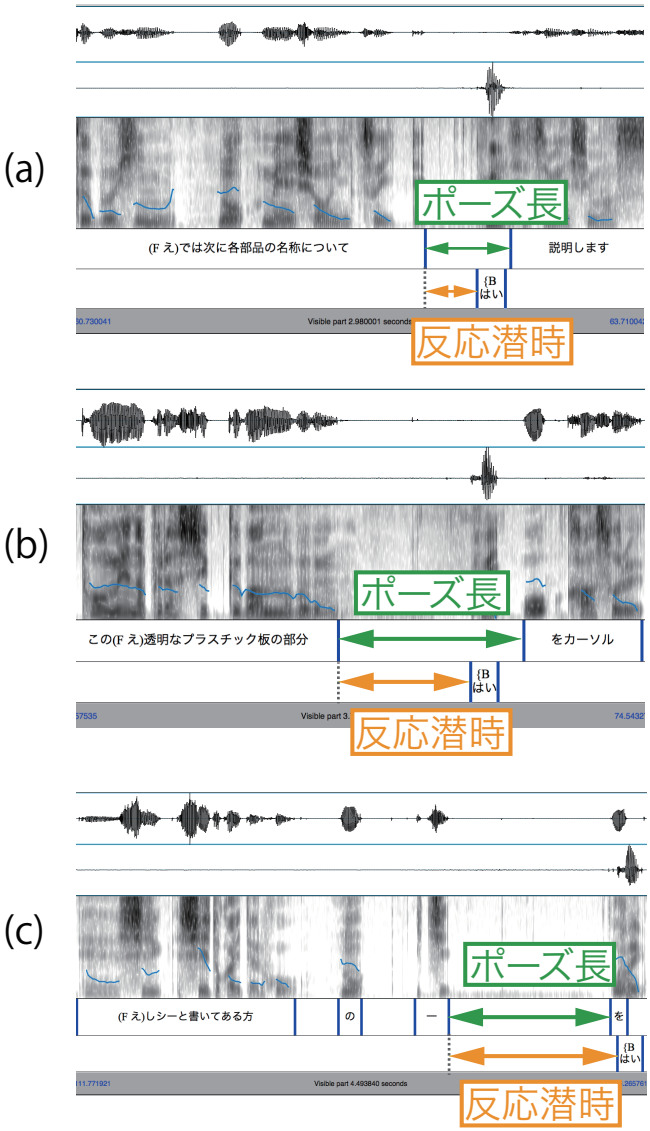

図 2　反応潜時とポーズ長の例

表 4 に各被験者の相槌と復唱に対する反応潜時とポーズ長の相関係数を示す．S005 と S007 の被験者の 2 名に関しては他の被験者に比べて相関が弱かった．S005 の被験者は相槌の出現回数が 8 回と非常に少なく分析は難しい．一方，S007 の被験者は相槌の出現回数が 38 回と全被験者中最多であるにも関わらず最も相関が弱い事がわかる．このことから，S007 の被験者は他の被験者とは異なる振る舞いをしていると考えられる．

図 5 に各被験者の相槌の反応潜時の分布を示す．S007 の話者は他の被験者と比較して相槌の生起タイミングが早く，分布も狭いことがわかる．S007 の被験者は説明者の発話が途切れたら，自身が理解しているかどうかに関わらず相槌を行う傾向がある．これにより説明者は被説明者の理解状態に問題が無いと推定し，説明計画を変えることなく説明を続けるため短いポーズ長が出やすくなる．これらの結果，反応潜時とポーズ長の分布が共に狭くなり，相関が弱くなったと考えられる．以上をまと

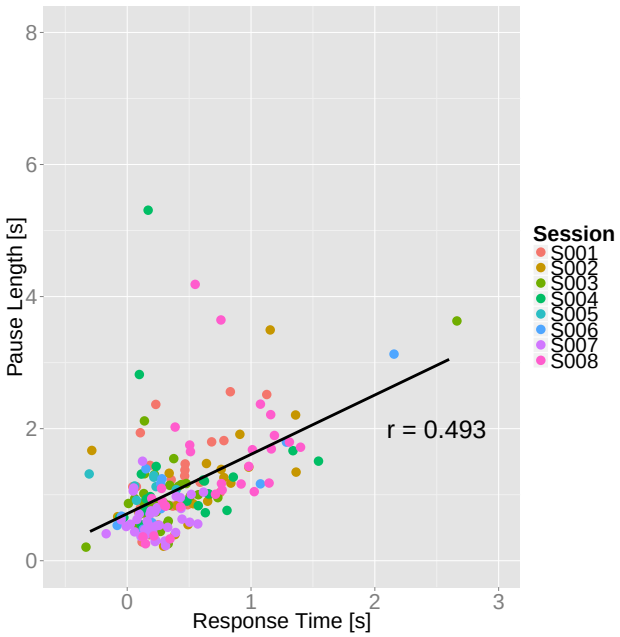

図 3　相槌 (B) の反応潜時とポーズ長の関係

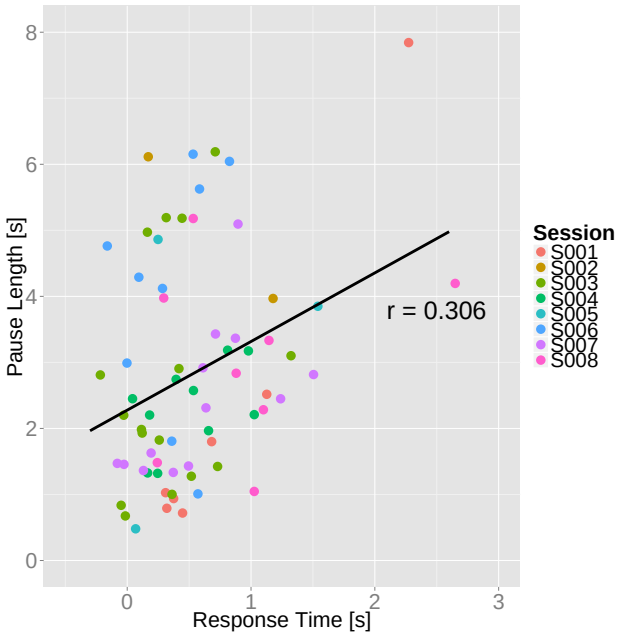

図 4　復唱 (R) の反応潜時とポーズ長の関係

めると, S007 の他の被験者と異なる振る舞いについては, S007 は自身が理解できているかどうかよりも対話を円滑に進めることを優先する行動傾向があると考えられる.

図 6 に各被験者の復唱の反応潜時を示す. 他の被験者よりも早く復唱が生起された被験者はおらず相槌とは異なる結果となった. これらより, 復唱の生起タイミングは相槌の生起タイミングに比べて被験者による差が小さいことがわかる.

図 7 に各聞き手反応に対するポーズの分布を示す. 相槌に対する説明者のポーズはその他の聞き手反応に対して最も短い. これは被説明者の理解状態に問題がなく, 説明者が説明内容を考えなおす必要がなかったためだと考えられる. 一方で, 復唱

表 4　各被験者の反応潜時とポーズ長の相関係数

被験者 ID	相槌 (B)		復唱 (R)	
	出現回数	相関係数	出現回数	相関係数
S001	25	0.494	7	0.679
S002	22	0.613	2	—
S003	35	0.390	16	0.315
S004	20	0.325	10	0.456
S005	8	-0.238	3	—
S006	11	0.763	9	0.283
S007	38	0.117	13	0.709
S008	29	0.666	8	0.143

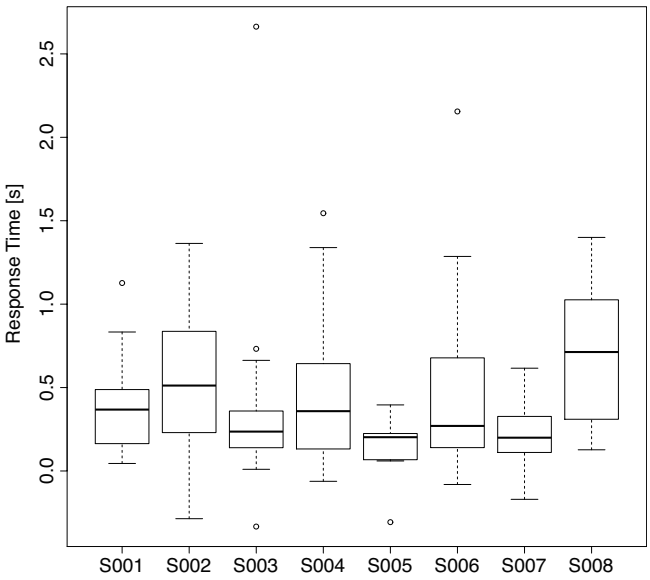

図 5　各被験者の相槌の反応潜時

に対するポーズはその他の聞き手反応に対して長く, また, 広く分布している. ポーズが長くなったのは, 復唱を聞いた説明者が被説明者の理解状態に問題があると推定し説明計画を変えて説明内容を考えなおしたためだと考えられる. 以上より復唱は理解状態に問題があることを伝える機能と相槌に近い機能を持っており, そのため分布が広くなったと考えられる.

5. おわりに

口頭での説明場面において, 話し手が聞き手の理解状態の手がかりをどのように利用し, 説明行動に反映させているかを調べるため計算尺使い方説明タスクの対話を収録した. 分析の結果, 一部の被説明者において反応潜時が長い場合に説明者が次の発話の開始を遅らせている傾向が見出された. また, 復唱に関しては被説明者が理解状態に問題があることを伝える機能と相槌に近い機能の両方があることが分かった.

今後の課題は説明者と被説明者を増やしてデータを蓄積することで, より一般的な結果を導きたい. また, 同じ復唱応答でも, 相槌のように理解状態に問題がない場合と, 明示的な聞き返しのように理解状態に問題がある場合の両方があることがわかったが, 説明者が復唱発話のどのような特徴からそれらを区別しているのかを, 復唱発話の音響分析等により明らかにしていく必要もある.

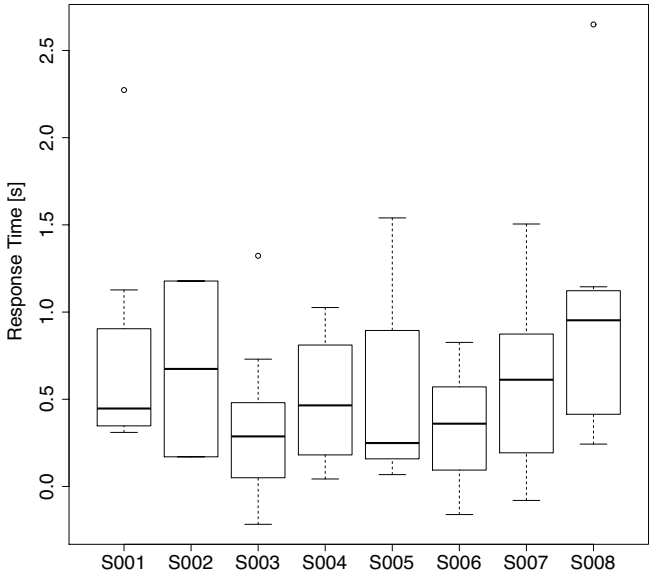

図 6　各被験者の復唱の反応潜時

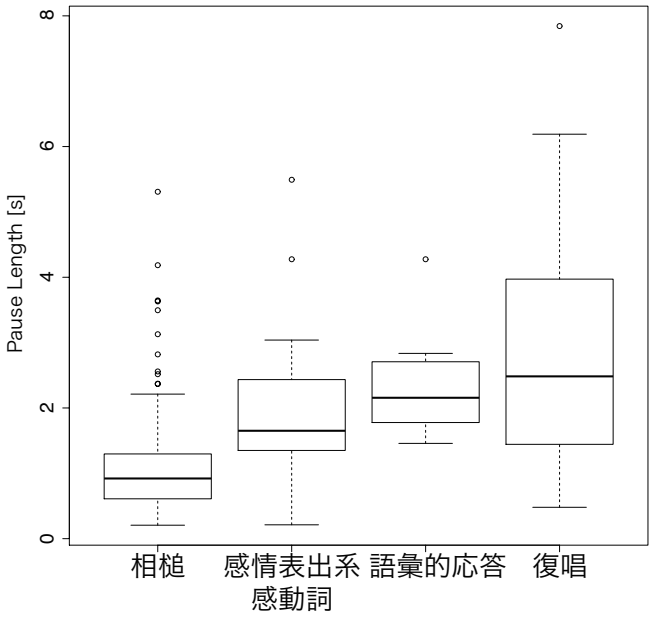

図 7　各聞き手反応に対するポーズの分布

文　　献

[1] 藤江 真也, 福岡 維新, 麥田 愛純, 高津 弘明, 林 良彦, 小林 哲則, “聞き手の反応を考慮した発話計画を用いた音声対話システム,” 日本音響学会 2015 年秋季研究発表会講演論文集, pp. 39–42 (2015).

[2] 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, **2009-SLP-77**, 22, pp. 1–6 (2009).

[3] Y. Den, H. Koise, K. Takanashi, and N. Yoshida, “Annotation of response tokens and their triggering expressions in Japanese multi-party conversations,” Proc. LREC 2012, pp. 1332–1337 (2012).