

話し手の修復行動を伴う聞き手の復唱の分析

Analysis of listener's repetitions that cause repair by speaker

藍原 瞭* 森 大毅

Ryo Aihara, and Hiroki Mori

宇都宮大学大学院 工学研究科

Graduate School of Engineering, Utsunomiya University

Abstract: Although spoken guidance has been spreading everywhere, it sometimes proceeds to the next explanation without regard to the listener. To solve this problem, it is necessary to make machines aware of listener's state of understanding.

In this study, we focused on the response tokens that may potentially be useful for estimating user's state of understanding. Among various kinds of response tokens, we have examined listener's repetitions that may cause speaker's repair. We have recorded dialogs of an explanatory scene where a participant is to explain how to use a tool that is called *slide rule* 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 listener, and that they dynamically modify the explanation plan with the clues. Among repair-initiation expressions, repetition was turned out to occupy a large proportion. To further investigate the influence of listener's repetitions to the speaker, the relationship between the listener's response time and speaker's reaction (continuation or repair) was also investigated.

1 はじめに

音声ガイダンスは社会のあらゆる所で利用されている。具体的には、駅の切符販売機、運送業の再配達自動受付システム、自動現金預払機 (ATM)、自動体外式除細動器 (AED) などの社会インフラや家庭用電気製品など多岐にわたる。これらは主に機器の使い方を説明する目的で利用されている。

しかし、まとまった量の情報を伝達するとき、被説明者の理解が追いついていないにも関わらず説明が次へ次へと移って行ってしまふ場合がある。かといって、説明の途中で逐一理解が追いついていないかどうかの確認を挟むというのも利用者からすれば煩わしい。このように現状の音声ガイダンスは被説明者の理解状態に合わせた説明をすることができない。

それに対し、人間による説明場面では円滑な対話が行われている。対話システムによる説明と人間による説明の違いは、システムが聞き手の反応を一切考慮せず一方的に説明を続ける点にある。人間同士のインタラクションにおいては、例えば説明場面のように話し手がほとんど一方的に情報を伝えるような場面でも、説明と平行して聞き手の反応をリアルタイムで監視し、状態をモニタリングしている。

このように人は説明する際に相手の状態をモニタリングして、そこから理解状態を推定し、その推定した結果を自身の説明にフィードバックしている。機械が人間の様に被説明者のどのような反応をどのように説明にフィードバックしているのかを解明することでシステムが行うガイダンスを改善できると思われる。

聞き手の反応を利用して説明戦略をリアルタイムで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様々な特徴を見出し、聞き手の理解状態を総合的に推定する必要がある。藤江ら [1] は、事前に生成した説明計画に従ってシステムが対話を行い、ユーザーによ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、システム発話をその反応に応じた計画に従ったものに変更することを提案している。一方で、システム側が主に聞き手になり、話し手であるユーザーの発話の間に相槌、復唱、共同補完などの聞き手行動を生成させることで円滑な対話を実現することを目指したシステムの提案もされている [2]。

しかしながら、人間同士の説明場面において、説明者が被説明者の反応に含まれるどのような特徴を聞き手の理解度として推定に利用し、また、それを自身の説明にどのようにフィードバックさせているのかを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するための対話コーパスも公開されていない。

*連絡先：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@speech-lab.org

そこで本研究では、被説明者の理解状態が観察可能などのような反応として現れるのかを解明するとともに、説明者がそれをどのように自身の説明にフィードバックしているのかを解明することを目的としている。

これまでに、対話音声コーパスの構築を行い、被説明者の理解状態が現れる反応として相槌と復唱に着目し、また、説明者が自身の発話の生起タイミングを被説明者の反応潜時によって調整していることを実データから示した [3]。しかし、説明者が被説明者の理解状態に問題があると判断し説明の修復を行う際、被説明者の復唱が理解状態に問題があることを伝える手段であった可能性が示唆されたものの十分な検討は行えていなかった。

復唱に関する研究としては、小磯ら [4] が、感動詞「あー」「へー」「ふーん」が、焦点要素が発話内に特定できるケースについては、発話の末尾よりも前に発話される傾向がある [5] のに対し、復唱の生起位置は焦点要素の位置に関わらず発話末を基点に一定時間内に反応する傾向があると指摘した。しかしながら、聞き手が自身の理解状態に問題があることを話し手に伝える役割を持ち、話し手の説明の修復行動にどのような影響を及ぼすのかについては明らかになっていない。

そこで本論文では、人間同士の説明場面において、修復を伴った復唱と修復を伴わなかった復唱が説明者にどのような影響を及ぼしたのかについて分析を行った。

2 分析対象

本研究で分析対象とする音声対話コーパスとして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。

既存の音声対話コーパスには、日本語地図課題コーパス [6] などがあるが、これは双方向の説明を必要とするため要件 1. を満たさず、また、一つ一つの説明が短

いことから要件 3. も満たさない。そのため、新しくコーパスを設計する必要がある。

上述した要件を満たすものとして、計算尺使い方説明タスク [3] を考案し、収録を行った。計算尺は電子計算機の登場以前に使われていた計算道具であり、被験者として募りやすい工学部の学生にとっては仕組みがわかれば容易に扱うことができる。しかし、使ったことのない者に使い方を説明するのは容易ではなく、したがって、ひとつひとつの説明が長くなることが期待できる。

計算尺使い方説明タスクは、説明者と被説明者の手元に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音声のみで説明する。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 つについて行った。

収録は防音室内で行った。説明者と被説明者は区切られた防音室内にあり、お互いの手元は見え、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。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

3 アノテーション

間休止単位 [6] を用いて、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。アノテーション作業は著者 1 名によって行った。以下に対話音声に付与した情報を示す。

- 発話開始時刻、発話終了時刻
- 発話内容の書き起こし
- 聞き手反応
- フィラー
- 修復行動
- どの発話に誘発されたか
- 誰に向けられたか

「どの発話に誘発されたか」については、聞き手反応と修復行動のみに記述した。

表 1: 聞き手反応 [7]

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	話し手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

3.1 聞き手反応

対話中に話し手が話している最中にも聞き手は相槌やうなずきといった反応を送ることで対話に参加していることを話し手に示している。このような発話権を行使しない反応を聞き手反応 [7] という。表 1 に聞き手反応を示す。発話内容の書き起こしに対して表 1 の例に当てはまる聞き手反応のタグを付与した。ただし、応答系感動詞については、説明者の発話末より後に生じたものは一度受け取った発話権を再度説明者に渡す意図で発話されたと考え、応答感動詞からは取り除いた。また、書き起こしでは形式上復唱の例に当てはまるが、明示的な情報要求としての意図で発話されたものは復唱から取り除いた。また、表 1 のどの例にも当てはまらない応答詞「はあ」の取り扱いについては、付録 A で説明する。

3.2 修復行動

修復連鎖 [8] についてのアノテーションを行った。修復連鎖を構成する 3 つの要素として問題源 (source), 修復開始要素 (initiation), 修復 (solution) がある。対話音声に記述を行う際は、まず説明者が修復を行った発話に「solution」タグを付与し、それを誘発した発話に「initiation」タグを付与した。さらに「initiation」の発話を誘発した「source」のタグを付与した。

修復を伴った復唱の例を表 2 に、修復を伴わなかった復唱の例を表 3 に示す。表 2 の対話では被説明者の No. 83 の「シー尺」という発話が復唱であり、これにより説明者の後続の発話 No. 84 と No. 85 により「シー尺がどこにあるのか」を補足するために修復が行

表 2: 修復を伴った復唱の例 (S003:103.728-122.915)(A:説明者 B:被説明者)

No	StartTime	who	text	EndTime	refNo	Repairtype	Addressee
78	103.728	A	掛け算するときに	104.877		No-repair	B
79	105.477	A	何を使うかという	106.716		No-repair	B
80	107.26	A	滑尺の	108.142		No-repair	B
81	108.476	B	{B はい}	108.65	80	No-repair	A
82	108.981	A	{F え} シー尺	109.821		source	B
83	110.551	B	{R シー尺}	110.996	82	initiation	A
84	111.245	A	シーエフシーアイエフケーシーアイシー	114.298	83	solution	B
85	114.908	A	ってあると思いますけど	115.917		No-repair	B
86	116.052	B	{B はい}	116.181	85	No-repair	A
87	116.693	A	そのシー尺	117.363		No-repair	B
88	118.524	A	と	118.706		No-repair	B
89	119.453	A	{F え} 固定尺のディー尺を使います	122.915		No-repair	B

表 4: 修復開始要素となった発話の内訳

反応	修復開始要素	発話数
聞き手反応	相槌	0
	感情表出系感動詞	1
	語彙的応答	1
	復唱	17
	共同補完	0
その他	明示的な情報要求	11

われている。一方で、表 3 の対話では被説明者の No. 88 の「カーソル」が復唱であるが、それに対して説明者は修復を行わず説明を続けている。

3.3 発話の宛先

収録した対話音声の中の被説明者の発話には独り言のように聞こえる発話がいくつか含まれているが、独り言への反応義務はないため、そこからの説明者の反応潜時を求めることは無意味であり、データから除外する必要がある。そこで、発話が誰に対して向けられたのか (Addressee) を記述した。一対一の対話であることから基本的には A (説明者) の発話の Addressee には B (被説明者) と記述し、B の発話の Addressee には A と記述した。独り言のような発話だった場合には Addressee に nobody と記述した。

4 結果と考察

表 4 に被説明者の修復開始要素となった発話の内訳を示す。修復開始要素となった発話は 30 発話あり、そのうち 19 発話が聞き手反応であった。さらにその中でも復唱が 17 発話存在した。この結果から、被説明者が説明を理解できていないことを説明者に伝える場合、およそ半分は聞き手反応のような短い応答によってなされるということが分かる。加えて、その聞き手反応の中でも復唱がその役割のほとんどを担っていることも分かる。

表 3: 修復を伴わなかった復唱の例 (S002:105.69-113.419)(A:説明者 B:被説明者)

No	StartTime	who	text	EndTime	refNo	Repairtype	Addressee
82	105.69	A	で	105.868		No-repair	B
83	106.346	A	それ以外の三つ目の	107.678		No-repair	B
84	108.301	A	(F え) 透明なプラスチック板	110.006		No-repair	B
85	110.377	A	がついた	110.843		No-repair	B
86	111.154	A	部分	111.433		No-repair	B
87	111.697	A	これをカーソル	112.438		No-repair	B
88	112.758	B	{R カーソル }	113.032	87	No-repair	A
89	113.229	A	といいます	113.774		No-repair	B
90	113.251	B	はい	113.419		No-repair	A

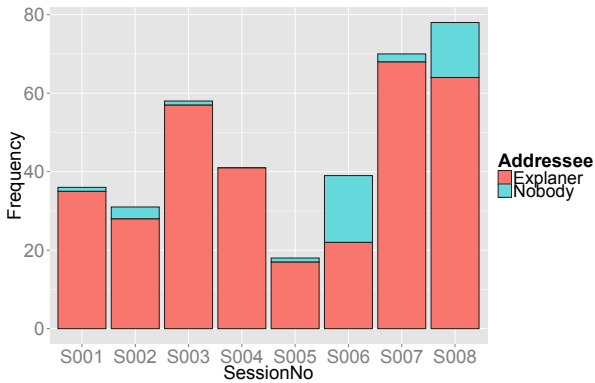


図 1: 各被験者の聞き手反応における誰にも向けていない発話の割合

表 5 に被説明者の復唱 96 発話のうち修復開始要素となったものの数を示す。誰にも向けられていないにも関わらず修復を伴った発話は 3 発話存在した。このことから、説明者は必ずしも自身に向けられた発話でなくとも被説明者の理解状態に問題があると判断すれば修復を行うことがあることが確認できた。

図 1 に各被験者の聞き手反応における独り言の割合を示す。S006 と S008 の被験者が他の被験者に比べて独り言の回数が多い。しかし、割合を見ると S008 は全体の聞き手反応の出現回数が多いため他の被験者と変わらず、S006 の被験者が聞き手反応における独り言の割合が大きいことが分かる。

説明者が、相槌のように理解状態に問題がない場合に使われる復唱と理解状態に問題がある場合に使われる復唱をどのように区別していたのかを調べる。聞き手が話し手に対して懐疑的な姿勢をとっているとき、相槌の反応潜時が長くなることが指摘されている [9]。このことから、復唱においても理解状態に問題があると

表 5: 被説明者の全復唱に占める修復開始要素

Addressee	修復開始要素	全復唱
A(説明者)	14 発話	60 発話
nobody	3 発話	36 発話
計	17 発話	96 発話

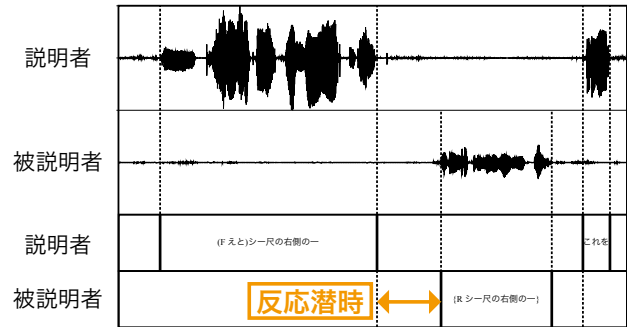


図 2: 被説明者の反応潜時

きは被説明者の反応潜時が長くなると予想した。以上より、説明者が復唱を受けて説明を修復するかどうかを判断するとき復唱の反応潜時を判断の基準としているかどうかを調べるため、修復を伴った復唱と修復を伴わなかった復唱の反応潜時を比較する。反応潜時の定義を図 2 で説明する。図 3 に誰にも向けられていない発話を除く、修復を伴った復唱と修復を伴わなかった復唱の被説明者の反応潜時の分布を示す。これについて t 検定を行ったが有意な差は見られなかった。すなわち、説明者が復唱を受けて修復を行う必要があるかを判断するとき、復唱の反応潜時を判断基準にしているという仮説を支持する証拠は得られなかった。

5 おわりに

本論文では、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に、計算尺使用方説明タスクコーパスの収録を行い、被説明者の復唱が説明者の修復行動に及ぼした影響に着目し分析を行った。

収録した対話音声の中では、修復を伴った被説明者の発話は復唱の割合が多かった。このことから、復唱が被説明者の理解状態に問題があることを伝える手段であることが示唆された。加えて、説明者は被説明者の発話が自分に向けられていなくても理解状態に問題があると判断すれば修復を行うことがあることが確認で

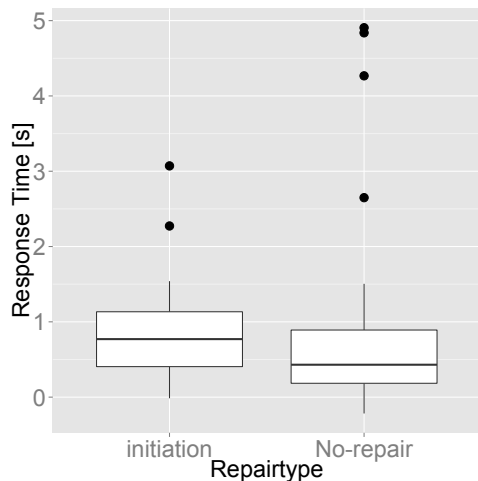


図 3: 修復開始要素と修復を伴わない復唱の被説明者の反応潜時の分布

きた。また、修復を伴った復唱と修復を伴わなかった復唱の生起タイミングに有意差は認められず、説明者が復唱の反応潜時を修復が必要かどうかの判断の手がかりにしているという仮説は裏付けられなかった。

今後は、説明者を追加して収録を行った同タスクの対話音声についても同様の分析を行い、一般的な結論を導く。また、説明者の説明のどのような違いによって被説明者による反応が変わったのかを調べる必要がある。

参考文献

- [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, Vol. 2009-SLP-77, No. 22, pp. 1–6, 2009.
- [3] 藍原瞭, 森大毅, “話し手の説明計画に影響する聞き手行動の分析,” HCG シンポジウム, pp. 436–440, 2016.
- [4] 小磯花絵, 伝康晴, “会話における (部分) 復唱発話の分析,” 日本認知科学会第 28 回大会発表論文集, pp. 250–255, 2011.
- [5] 高梨克也, 常志強, 河原達也, “聞き手の興味・関心を示すあいづちの生起する会話文脈の分析,” 人工知能学会研究会資料, Vol. SIG-SLUD-A903, No.58, pp. 25–30, 2010.
- [6] 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.3-4, pp. 295–321, 1998.

- [7] Den, Y., Koiso, H., Takanashi, K., and Yoshida, N., “Annotation of response tokens and their triggering expressions in Japanese multi-party conversations,” In Proceedings of the 8th Language Resources and Evaluation Conference, 2012.
- [8] Schegloff, E. A., Jefferson, Sacks, H., “The preference for self-correction in the organization of repair in conversation,” Language, Vol. 53, pp.361-382, 1977.
- [9] Mori, H., “An analysis of switching pause duration as a paralinguistic feature in expressive dialogues,” Acoust. Sci. & Tech, Vol. 30, 2009.

A 応答詞「はあ」の扱いについて

収録した対話音声の中に「はあ」という応答詞が存在した。この応答詞「はあ」は聞き手反応同様、説明者の発話中に挿入される被説明者による短い応答としての性質を持っており、応答系感動詞や感情表出系感動詞に近いものと思われる。しかし、収録した対話音声の中では応答系感動詞と感情表出系感動詞のどちらにも当てはまる使われ方をしており分類の基準を定めるのは容易ではない。

収録音声における実際に応答詞「はあ」が含まれた対話の例を表6と表7に示す。表6の例は応答詞「はあ」が修復開始要素となった対話の例である。この対話は $2 \times 3 \times 5 \times 2$ の計算をしている途中の対話である。説明者が 2×3 で6ができたという説明をしたがNo 355で応答詞「はあ」が生起し、それを聞いた説明者が被説明者に対し正しく計算できたか確認する旨の発話をしている。このNo 355には「はあ」ではなく「はい/いいえ」が入っていたとしても不自然ではない。実際にその後のNo 362までのやり取りで正しく計算できていなかったことをお互いに認識しているため、被説明者はNo 355で「いいえ」と発話していた可能性も考えられる。また、もし「いいえ」と発話されていた場合、その後のNo 356からの正しく計算できているか確認を行う修復の発話も別の発話になっていた可能性も考えられる。このことから、この「はあ」は「はい/いいえ」と同じ箇所に生起するものでありながら、「はい/いいえ」のどちらでもない曖昧な返事を返しているものと考えられる。

表7に応答詞「はあ」が修復開始要素にならなかった

場合の対話の例を示す。この例は6の場合と違い「はい/いいえ」で置き換えられないものである。No 222の「はあはあはあ」の直前に修復連鎖の修復が起きており、それに対する反応としてこのような反応が表れた。収録音声の中でこの箇所に置き換えられる別の表現としては「あ、なるほど」といった反応が考えられる。

以上のように応答詞「はあ」は役割が複数あり、既存の聞き手反応のいずれかに分類するのが容易ではない。しかし、表6の例のように被説明者が理解状態に問題があることを説明者に伝えるシグナルとして使われることがあるのは明白なので、既存の聞き手反応の中では最も近いと思われる感情表出系感動詞としてアノテーションを行った。

表 6: 応答詞「はあ」が修復開始要素となった例 (S008:610.283-625.899)

No	StartTime	who	text	EndTime	RepairType
352	610.283	L	次にカーソルを動かしてシー尺の三に合わせてください	613.899	source initiation solution
353	614.194	R	はい	614.395	
354	616.883	L	これでひとまず六が出来ました	618.733	
355	620.994	R	{E はあ }	621.195	
356	622.807	L	ディー尺	623.346	
357	623.884	L	をカーソルが六を指してるよね	625.899	
358	628.323	R	{E え }	628.359	
428	628.71	R	ごめ	628.954	
360	628.731	L	{E え }	628.877	
361	629.24	R	{E え }	629.467	
362	629.734	R	指してない {E え }	630.583	

表 7: 応答詞「はあ」が修復開始要素とならなかった例 (S004:391.113-413.751)

No	StartTime	who	text	EndTime	RepairType
215	391.113	L	なんでこの二種類あるのか	392.8	source initiation solution
216	394.288	L	計算の仕方が二種類あるのかというと	396.155	
217	396.393	L	これを交互に行うことで連続した	399.179	
218	399.598	L	掛け算をすることができる	401.069	
219	401.585	L	からです	402.183	
220	403.162	R	{R 連続した掛け算ができる }	404.665	
221	405.357	L	(F ま) 二掛ける三掛ける五掛ける二	407.654	
222	408.993	R	{E はあはあはあ }	409.586	
223	409.32	L	みたいな	409.665	
224	412.1	L	(F ま) 今からそれをやっっていくんですが	413.751	