

Keep Generating and Nobody Explodes[🧨]: カスタマー操作型対話の自動化に向けた課題分析

佐藤志貴¹ 邊土名朝飛¹ 東佑樹² 岩田伸治¹

¹サイバーエージェント ²株式会社 AI Shift

{sato_shiki, hentona_asahi, azuma_yuki, iwata_shinji}@cyberagent.co.jp

概要

コールセンターのテクニカルサポートでは、オペレーターが音声のみでカスタマーの状況を把握し、手順書等に基づいて適切な環境操作を導く必要がある。本研究はこの設定を模したビデオゲーム『Keep Talking and Nobody Explodes』を題材に、リアルタイム音声対話モデルを分析担当者として用いた実験を行い、リアルタイム音声対話モデルを用いたオペレーター対応の自動化に向けた課題を検討する。実験の結果、27回の試行のうち成功回数は0回だった。エラー分析から、コンテキストの忘却、誤った操作の案内、前提を無視した発話が多く観察され、これらが実務での通話時間の増加や高リスク操作の誘発、対話停滞に直結しうることを議論した。

1 はじめに

近年のリアルタイム音声対話モデル [1] の飛躍的な性能向上に伴い、コールセンターのオペレーターの電話対応の自動化が注目されている。コールセンター対話は、タスク達成のための環境（状態を持つオブジェクト）を誰が操作するかによって、オペレーターが操作するオペレーター操作型対話と、カスタマーが操作するカスタマー操作型対話に大別できる。前者の例として、オペレーターがカスタマーの要望を聞き取りながら予約システムを操作し、レストランや診療の予約を行う場面が挙げられる。コールセンター対話におけるオペレーター対応の自動化に向けた先行研究は、この対話を主に扱ってきた [2]。

一方、実際のコールセンター業務ではカスタマー操作型対話も多く見られる。代表的な例として、カスタマーが所有する製品で発生している問題を解決するために、オペレーターが通話越しに製品の操作方法などを指示するテクニカルサポート対話が挙げ



図1 本研究で扱うビデオゲーム『完全爆弾解除マニュアル：Keep Talking and Nobody Explodes』プレイ中の処理担当者の画面のスクリーンショット。© Steel Crate Games, Inc. 後述する3種のモジュール（左下：ワイヤ、右下：ボタン、右上：キーパッド）が含まれる。モジュールの配置や各モジュールの設定（たとえばワイヤの本数・各ワイヤの色）は試行ごとに毎回異なる。

られる。カスタマー操作型対話では、音声のみを通じてカスタマー側の環境の状態を正確に把握したうえで適切な環境操作の実施を音声のみで導く必要があり、オペレーターが自ら環境を操作可能なオペレーター操作型対話とは異なる能力が求められる。しかし、カスタマー操作型対話におけるオペレーターの電話対応の自動化に関する研究は少なく、近年のリアルタイム音声対話モデルを同業務の自動化に利用する際の課題は明確ではない。

本研究では、カスタマー操作型対話と類似した設定を持つ Steel Crate Games 社のビデオゲーム『完全爆弾解除マニュアル：Keep Talking and Nobody Explodes』¹⁾ (図1) をプレイするリアルタイム音声対話モデルの振る舞いを分析することで、カスタマー操作型対話におけるオペレーター役の自動化を実現するうえでの課題を議論する。

2 関連研究

カスタマーサポート対応の自動化に向けた研究は長年取り組まれてきた [2, 3, 4]。近年では、深層学

1) <https://keeptalkinggame.com/>.

習モデルを用いた自動対話技術やその達成に向けたデータ作成が盛んに取り組まれている。具体的にはカスタマーサポート時に大規模言語モデルを用いるための知識表現・ガイドライン考慮・検索拡張生成とのハイブリッドシステム [5, 6, 7], 音声やテキスト対話データセット作成 [8, 9, 10] が存在する。しかし、従来の研究はオペレーター操作型対話を主に扱ってきた。一方、カスタマー操作型対話に関する研究は少なく、既存研究においてもテキストと画像のみで対話する研究 [11] や、大規模言語モデル同士の対話を扱う研究 [12] 等、人とリアルタイム音声対話モデルで音声対話を行った際の課題は明確にはされていない。

3 タスク

本節では、Steel Crate Games のビデオゲーム『完全爆弾解除マニュアル：Keep Talking and Nobody Explodes』の概要を説明したうえで、本研究の実験におけるタスク設定を述べる。

3.1 完全爆弾解除マニュアル

本ゲームは、2人のプレイヤーが会話をしながら制限時間内にゲーム画面上の爆弾の爆発を阻止するビデオゲームである。2人のプレイヤーは、ゲーム画面上で爆弾を操作して解除する『処理担当者』と、『爆弾解除マニュアル』を参照しながら解除方法を指示する『分析担当者』に分かれる。爆弾の解除には、爆弾自体の情報（たとえば、何色のワイヤが何本あるか）と爆弾解除マニュアル中の情報（たとえば、何色のワイヤが何本ある場合にどのワイヤを切断すべきか）を照らし合わせる必要がある。ただし、処理担当者はマニュアルを、分析担当者は爆弾自体を直接確認できないため、分析担当者は爆弾の情報を正確に処理担当者から聞き出し、マニュアルから特定される適切な操作を導く必要がある。

本研究では、分析担当者に求められるこれらの能力が1節で述べたカスタマー操作型対話においてオペレーターに求められる能力と多くの共通点を持つことを踏まえ、同ゲームを題材としたタスクを用いた実験を実施する。

3.2 タスク設定

本研究では、リアルタイム音声対話モデルを分析担当者に割り当てたうえで、モデルが対話を通じて必要な情報を収集し、マニュアルに基づく推論によ

り適切な操作手順を特定し、処理担当者による正しい操作を導くことができるかを評価する。

入出力。 分析担当者は、処理担当者の適切な操作を特定するためのルール集合である爆弾解除マニュアルと、処理担当者からの音声入力を受け取り、処理担当者に対する音声を出力とする。両参加者に対してターンテイキングに関する制約は設けないため、分析担当者は処理担当者が発話中かにかかわらず音声を出力できる。

評価指標。 タスクの成功率を評価指標として用いる。ここで、分析担当者の指示を通じて処理担当者が制限時間内に爆弾を解除し、爆弾の爆発を回避することをタスクの成功と定義する。また、時間切れあるいは誤操作の累積により、爆発に至ることを失敗と定義する。

4 実験設定

本研究では、人間の被験者を処理担当者としたうえで、上述のタスクにおける近年のリアルタイム音声対話モデルの性能を評価した。

リアルタイム音声対話モデル。 Gemini Multimodal Live API²⁾を介して対話を行った。モデルは gemini-2.5-flash-native-audio-preview-09-2025 を用いた。API 利用時のパラメータは全て初期値を用いた。システムプロンプトを付録 A に示す。

解除対象となる爆弾。 ゲーム内のチュートリアルである“Section 1: Introduction”の“1.3 The First Bomb”における爆弾を解除対象とした。この爆弾の解除の制限時間は5分で固定されている。また、誤操作によりストライクが蓄積し、3回に達すると爆発する。本ゲームにおける爆弾は複数の解除対象（モジュール）から構成され、爆発前に爆弾の全てのモジュールについて適切な操作が行われた場合に解除成功となる。当該爆弾は常にワイヤ、ボタン、キーパッドの3つのモジュールからなる。各モジュールの概要を以下に示す：

- **ワイヤ**：複数本のワイヤが提示され、ワイヤの本数・色などの情報に基づいて、切断すべき1本のワイヤがルールとして定まる。
- **ボタン**：ボタンの色・ラベル（文字）・長押しした際の点灯色などに基づき、ボタンの押し方や離し方がルールとして定まる。
- **キーパッド**：4つのキーそれぞれに描かれている合計4種類の記号の組み合わせに基づいて、

2) <https://ai.google.dev/gemini-api/docs/live>.

4 個のキーを押す順番がルールとして定まる。

マニュアル. Steel Crate Games が公開している爆弾解除マニュアル (BombManual)³⁾のうち、当該チュートリアルで必要となる範囲 (先頭 7 ページ) を画像としてリアルタイム音声対話モデルに提示した。

処理担当者. 処理担当者は、著者の所属組織に所属する計 9 名とした。実験開始前に、各被験者には著者のうち 1 名が分析担当者を務める設定で同チュートリアル (1.3) をクリアするまでプレイしてもらい、少なくとも人間が分析担当者であれば当該タスクを達成可能な水準に到達させた。

実験手順. 各被験者は、リアルタイム音声対話モデルを分析担当者とする条件で合計 3 回、本チュートリアル (1.3) の解除を行った。各回において、3 種類のモジュールすべての操作を実施する前に爆弾が爆発しうるため、被験者が操作に着手した回数のモジュール間での偏りを可能な限り抑制する必要がある。そこで、各回でモジュールを解く順序を事前に指定し、被験者には指定された順序で解くよう対話を主導させた。

5 実験結果

本実験では 9 名の被験者が 3 回ずつ爆弾解除に取り組み、計 27 試行を実施した。

5.1 成功率

表 1 にモジュール別の解除結果と爆弾解除の成功率を示す。ワイヤは実施 10 回のうち 7 回成功、ボタンは実施 17 回のうち 9 回成功であり、失敗も見られるものの、一定の割合で適切な操作の実施に成功した。一方、キーパッドは実施 21 回のうち成功 0 回であり、結果として爆弾解除は全 27 試行で成功 0 回であった (表 1)。

本実験の全被験者が分析担当者を人間とした試行において爆弾解除に成功した経験を有するため、今回評価したモデルは人間と比べ分析担当者としての支援能力が不十分であった可能性が高い。

5.2 エラー分析

実験実施時の記録をもとに、タスク実施中に発生したリアルタイム音声対話モデルのエラーを特定したうえで、各エラーの出現頻度を集計した。表 2 に、モジュールごとの試行回数に対して、あるエ

3) <https://www.bombmanual.com/ja/index.html>.

表 1 全 27 試行でのモジュール別の解除結果と爆弾解除の成功率。モジュールごとの失敗回数のうち、処理担当者が途中でそのモジュールの操作を諦め別のモジュールの操作に移行した回数を括弧内に示す。成功率は成功回数をそのモジュールや爆弾に対する操作実施回数で除算することで算出した。

対象	実施	未実施	成功	失敗	成功率
ワイヤ	10	17	7	3 (1)	.70
ボタン	17	10	9	8 (3)	.53
キーパッド	21	6	0	21 (1)	.00
爆弾解除	27	0	0	27	.00

表 2 モジュールごとの実施試行に対して、各エラーが 1 回以上発生した対話 (試行) の数と割合 (実施試行数で除した値)。

エラー種別	ワイヤ (n = 10)	ボタン (n = 17)	キーパッド (n = 21)
コンテキストの忘却	3 (.30)	3 (.18)	14 (.67)
誤った操作の案内	3 (.30)	9 (.53)	14 (.67)
前提を無視した発話	0 (.00)	0 (.00)	11 (.52)
相手意図の不理解	1 (.10)	4 (.24)	4 (.19)
その他	2 (.20)	0 (.00)	6 (.29)

ラーが 1 回以上発生した対話 (試行) の数と割合を示す。以下に、各エラーの概要と考えられる発生要因を述べる。

コンテキストの忘却. 既に処理担当者から得た情報 (たとえば、キーパッドの 4 個の記号やワイヤの本数・色) を保持できず、同じ情報のヒアリングの要求や、直前の前提と矛盾する指示が発生したエラーを指す。爆弾解除の失敗に直接繋がらないが、タイムロスにより制限時間内での操作の完遂を困難にする可能性がある。対話自体は制限時間により 5 分以内となるため、対話長が今回評価対象としたモデルの許容する文脈長を超過したことは本エラーの発生原因として考えにくい。そのため、入力に含まれる音声対話履歴を適切に参照することが現状のリアルタイム音声対話モデルにとって難しい場合があることが本エラーの発生原因である可能性がある。

誤った操作の案内. 処理担当者から提示された情報の理解の失敗や、提示された情報とマニュアルに基づく正しい操作の特定の失敗により、誤った操作手順を提示するエラーである。誤操作によるストライクの蓄積を考慮すると解除の失敗に直結する深刻なエラーである。同エラーは、特にキーパッド

の操作で多発した。ワイヤ・ボタンとキーパッドの相違点として、後者では厳密な言語化が困難な視覚情報を、音声対話を介して共有する必要がある点が挙げられる。分析担当者が適切な操作を案内するには、マニュアル中の記号一覧から、キーパッドに描かれた記号と一致するものを特定しなければならない。しかし、図1に示すキーパッドの例のように、キーに記載された記号の形状は既存の単一の文字では十分に表現できない。このため分析担当者は、記号一覧に掲載された各記号の特徴を事前に把握したうえで、処理担当者が主観に基づいて言語化した特徴から候補を絞り込む必要がある。処理担当者の説明のみに基づいて記号一覧から該当記号を同定することが現状のリアルタイム音声対話モデルにとって難しく、そのことがキーパッドでの誤った操作の案内の多発に繋がった可能性がある。

前提を無視した発話。 処理担当者がマニュアルを参照できない状況にもかかわらず、「特定の列・行」「該当ページ」など、処理担当者側でのマニュアルの参照を前提とした情報提示・指示を行うエラーである。本エラーは、キーパッドの操作のみで多発した。多発の原因として、キーパッドが操作の成功率の低さで示されているように難易度の高いモジュールであるために、同モジュールに関する対話履歴が長く複雑になり、タスクにおける前提の考慮が応答生成時に疎かになった可能性が考えられる。

相手意図の不理解。 処理担当者の要望（例：別解の提示、前提の確認、失敗後のリカバリ）を取り違え、意図に合致しない説明や指示を継続するエラーである。各モジュールの操作で発生したものの、発生件数は比較的少なかった。

6 議論：実用における課題

前節では、タスク遂行中に観察されたエラーを整理した。本節では、これらのうち多発した3種類のエラーがコールセンター業務（特にテクニカルサポートのようなカスタマー操作型対話）におけるオペレーター役の自動化において、どのような実務上の課題となるかを議論する。

コンテキストの忘却。 コールセンターでは、通話中に獲得した情報（契約者情報、機器型式、エラーコード、直前に実施した手順など）がその後の対応を決定するうえで複数回必要になる場合がある。コンテキストの忘却が生じると、一度取得したこれらの情報が再び必要になった際に再確認するこ

ととなり、通話時間の増加に直結する。また、過去の対応と矛盾する指示を提示した場合には、カスタマー側の操作を無駄にするだけでなく、信頼の毀損やクレーム発生に繋がらう。とくに長時間の通話において、対話履歴を保持及び参照する能力を向上させる必要がある。

誤った操作の案内。 テクニカルサポートでは、案内の誤りがそのまま誤操作・誤設定に繋がらう。機器の状態悪化といった実害を生む。本研究の設定ではストライク蓄積が失敗に直結したが、実務においても「やり直しが効かない操作」（初期化、課金操作など）が存在し、誤案内は特に高リスクである。文書および対話履歴に基づく適切な操作の同定のための推論能力やカスタマー情報の正確な把握のための対話能力の向上に加え、指示する操作のリスクに応じた段階的手順（たとえば、破壊的操作に先行する安全な検証）を組み込むことも重要となる。

前提を無視した発話。 実務のコールセンターでは、カスタマーが参照できる情報・操作可能な範囲は、端末種別やリテラシーなどによって大きく異なる。にもかかわらず、カスタマーが持たないリソース（手順書、管理画面、専門用語の知識等）を前提とした指示を出すと、対話が停滞し、状況確認からやり直す必要が生じる。したがって、対話が長期化・複雑化しても前提の遵守に対する注意を継続できるリアルタイム音声対話モデルの実現が、頑健な自動化にとって重要となる。

7 おわりに

本研究では、カスタマー操作型対話（テクニカルサポート等）におけるオペレーター役の自動化に向けた課題を明らかにするため、ビデオゲーム『完全爆弾解除マニュアル：Keep Talking and Nobody Explodes』を題材に、音声入出力が可能な大規模言語モデルを分析担当者として用いた実験を行った。9名の処理担当者による計27試行の結果、爆弾解除の成功回数は0回であり、特にキーパッドにおいて成功が得られなかった。エラー分析からは、コンテキストの忘却、誤った操作の案内、前提を無視した発話が多く観察された。また、これらのエラーは実務において、通話時間の増加、誤案内に伴う高リスク操作の誘発、対話の停滞といった問題に直結することを議論した。今後の取り組みとして、多様なモデルを評価対象とした実験の拡充や、実際の製品等を題材としたタスク設計が挙げられる。

謝辞

本研究の許可をいただいた Steel Crate Games 社に感謝いたします。また、被験者として実験にご協力いただいた株式会社サイバーエージェントの同僚の皆様にも感謝いたします。

参考文献

- [1] Wenqian Cui, Dianzhi Yu, Xiaoqi Jiao, Ziqiao Meng, Guangyan Zhang, Qichao Wang, Steven Y. Guo, and Irwin King. Recent advances in speech language models: A survey. In **Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 13943–13970, 2025.
- [2] Sinarwati Mohamad Suhaili, Naomie Salim, and Mohamad Nazim Jambli. Service chatbots: A systematic review. **Expert Systems with Applications**, Vol. 184, p. 115461, 2021.
- [3] Allen L. Gorin, H. Hanek, R. C. Rose, and L. Miller. Spoken language acquisition for automated call routing. In **3rd International Conference on Spoken Language Processing (ICSLP 1994)**, pp. 1483–1486, 1994.
- [4] Marilyn Walker, Irene Langkilde, Jerry Wright, Allen L Gorin, and Diane Litman. Learning to predict problematic situations in a spoken dialogue system: experiments with how may i help you? In **1st Meeting of the North American Chapter of the Association for Computational Linguistics**, 2000.
- [5] Hanchen Su, Wei Luo, Yashar Mehdad, Wei Han, Elaine Liu, Wayne Zhang, Mia Zhao, and Joy Zhang. LLM-friendly knowledge representation for customer support. In **Proceedings of the 31st International Conference on Computational Linguistics: Industry Track**, pp. 496–504, 2025.
- [6] Xiangyu Wen, Jianyuan Zhong, Zhijian Xu, and Qiang Xu. Guideline compliance in task-oriented dialogue: The chained prior approach. In **Findings of the Association for Computational Linguistics: NAACL 2025**, pp. 6750–6776, 2025.
- [7] Priyaranjan Pattnayak, Amit Agarwal, Hansa Meghwani, Hitesh Laxmichand Patel, and Srikant Panda. Hybrid AI for responsive multi-turn online conversations with novel dynamic routing and feedback adaptation. In **Proceedings of the 4th International Workshop on Knowledge-Augmented Methods for Natural Language Processing**, pp. 215–229, 2025.
- [8] Yuki Saito, Eiji Iimori, Shinnosuke Takamichi, Kentaro Tachibana, and Hiroshi Saruwatari. Calls: Japanese empathetic dialogue speech corpus of complaint handling and attentive listening in customer center. In **Interspeech 2023**, pp. 5561–5565, 2023.
- [9] James Gung, Emily Moeng, Wesley Rose, Arshit Gupta, Yi Zhang, and Saab Mansour. NatCS: Eliciting natural customer support dialogues. In **Findings of the Association for Computational Linguistics: ACL 2023**, pp. 9652–9677, 2023.
- [10] Mengze Hong, Wailing Ng, Chen Jason Zhang, Yuanfeng Song, and Di Jiang. Dial-in LLM: Human-aligned LLM-in-the-loop intent clustering for customer service dialogues. In **Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing**, pp. 5896–5911, 2025.
- [11] Nan Zhao, Haoran Li, Youzheng Wu, and Xiaodong He. JDDC 2.1: A multimodal Chinese dialogue dataset with joint tasks of query rewriting, response generation, discourse parsing, and summarization. In **Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing**, pp. 12037–12051, 2022.
- [12] Timothy Ossowski, Jixuan Chen, Danyal Maqbool, Zefan Cai, Tyler Bradshaw, and Junjie Hu. Comma: A communicative multimodal multi-agent benchmark. **arXiv preprint arXiv:2410.07553**, 2024.

A システムプロンプト

以下に、分析担当者を務めるリアルタイム音声対話モデルに与えたシステムプロンプトを示す：

あなたは『Keep Talking and Nobody Explodes (完全爆弾解除マニュアル)』の熟練した分析担当者（エキスパート）です。

あなたの役割は、添付のマニュアル（PDF）に基づき、現場の処理担当者（ユーザー）へ音声で指示を出し、爆弾を解除することです。

1. セッション開始時の最初の発話

セッションが接続されたら、ユーザーの入力を待たずに、直ちに以下の主旨で話しかけてください。

> 「こちら分析担当。聞こえるか？ まずは爆弾の側面にある情報を確認したい。シリアルナンバーとバッテリーの本数を読み上げてくれ。」

2. 音声対話のためのガイドライン

* **一問一答を心がける:** 音声でのやり取りなので、一度に多くの質問をせず、一つずつ情報を聞いてください。

* **簡潔に話す:** 長い説明は避け、短く切って話してください。

* **復唱する:** 聞き間違いを防ぐため、重要な情報（ワイヤの色や数字）は必ず復唱して確認してください（例：「赤、青、赤だな？」）。

3. 爆弾の基本情報（エッジワーク）の確認

多くのモジュール（ワイヤ、サイモン、複雑ワイヤ等）で、以下の情報が必須となります。これらを会話の冒頭、または必要になった瞬間に聞き出してください。

* **シリアルナンバー:** 特に「最後の数字が奇数か偶数か」「母音が含まれているか」が重要です。

* **バッテリー:** 本数を確認してください（単1か単3かの区別が必要な場合もあります）。

* **インジケーター:** 「FRK」「CAR」などのラベルと、点灯しているかどうか重要です。

* **ポート:** パラレルポートの有無などが必要になることがあります。

4. 各モジュールへの対応指針

ユーザーから見えているものを聞き、マニュアルに従って解除手順を指示します。

* **ワイヤ (Simple Wires):** 本数と色を上から順に聞き、切るべきワイヤを特定する。

* **ボタン:** 色、ラベル（「中止」「起爆」など）を聞く。長押しが必要な場合は、リリースするタイミング（ストリップの色と数字）を教える。

* **キーパッド:** ユーザーが口頭で伝える記号の特徴（例：「中に点があるA」「Ωのような形」）を聞き取り、マニュアルの図形と照合して押す順番を教える。

* **サイモン (Simon Says):** シリアルナンバーの母音の有無によって表を使い分け、点滅した色に対応する「押すべき色」を伝える。

* **「誰だ」(表比較 / Who's on First):** まずディスプレイの言葉を聞いて「読むべきボタンの位置」を教え、次にそのボタンの単語を聞いて「押すべきボタン」を指示する。

* **記憶 (Memory):** 過去のステージで押したボタンの位置やラベルを記憶しておく必要があるため、ステージごとに情報を整理して指示する。

* **モールス信号:** 短点（・）と長点（-）のリズムを聞き取り、単語を特定して対応する周波数で「送信」を押させる。

* **複雑ワイヤ:** ワイヤごとに「赤/青/白」「★の有無」「LEDの点灯」を聞き、ベン図の条件（S, C, D, P, B）に従って切断指示を出す。

* **順番ワイヤ:** 赤・青・黒のワイヤが「これまでに何本出現したか」をカウントし続ける必要があります。

5. エラー処理

* もし爆発までの残り時間が少ない場合や、ユーザーがパニックになっている場合は、声をかけて落ち着かせてください。

* ユーザーがミスをした場合は、サイモンゲームなど条件が変わるモジュールがあるため、現在のミス回数を確認してください。