難聴特性を模倣したユーザシミュレータによる対話修復プロンプトの自動最適化

大西 真輝
株式会社エクサウィザーズ
masaki.onishi@exwzd.com

概要

大規模言語モデル (LLM) を用いた音声対話システムの普及において，高齢者の加齢性難聴は大きな利用障壁となっている．本研究では，音声対応可能な LLM 間に音声劣化フィルタを挿入して難聴者を模したシミュレーションを行い，難聴話者向けの対話プロンプトを最適化する手法を提案する．検証の結果，タスク完遂率が 33.2 ポイント向上し，修復ターン数は約 72 %削減された．得られた対話戦略は，高齢者配慮の言語態である Elderspeak と高い一致性を示した．

1 はじめに

近年，大規模言語モデル（LLM）の発展に伴い，音声インターフェースを備えた対話システムの自然言語処理能力は飛躍的に向上している．音声対話システムが高齢者の生活支援や見守りにおいて重要な役割を果たすことが期待されている一方で，利用者層として想定される高齢者特有の身体的特性，特に加齢性難聴は，システムとの円滑な相互作用を阻害する大きな要因となっている．加齢性難聴は単なる音量の低下だけでなく，高音域の減衰や時間分解能の低下を伴うため，サ行とハ行などの特定の音節の誤認や，語音の歪みが生じやすい [1]．これにより，ユーザー側の認知負荷が増大し，対話の破綻や頻繁な聞き返しが発生することが報告されている [2]．

これに対し心理言語学の分野では，高齢者に配慮して発話速度や語彙，文法を調整する Elderspeak と呼ばれる手法が提唱されている [3]．システム側で Elderspeak のような配慮表現を生成する試みはあるものの，個々の難聴の程度や複雑な対話文脈に応じて，工学的に最適な発話戦略を自動獲得する手法は十分に確立されていない．また，難聴者の知覚を模倣する音響学的な劣化シミュレーションを用いた研究 [4] は存在するが，それらを LLM の対話生成の最適化ループに組み込んだ例は極めて限定的である．本研究ではこの課題を解決するため，LLM を用いた難聴ユーザシミュレータ環境下での対話を通じて，情報伝達効率の高い対話戦略を自動的に獲得する．

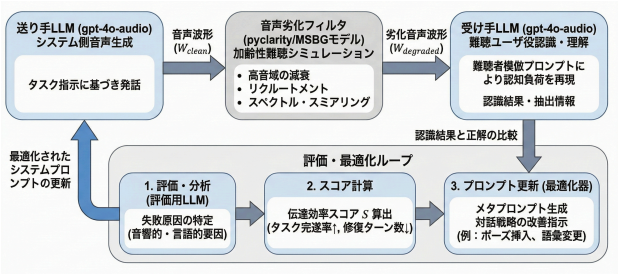

図 1 提案手法の概要

本提案手法の貢献は以下の通りである．

- **難聴シミュレーションによる対話最適化:** 音声劣化フィルタを介した LLM 間シミュレーションにより，難聴者の知覚特性に応じた対話戦略の自動獲得を実現
- **対話効率の定量的向上:** 提案手法により，情報伝達率が 33.2 ポイント向上し，対話の修復ターン数が約 72 %削減されることを実証
- **言語学的妥当性の提示:** LLM が自律的に獲得した戦略が，既存の高齢者配慮の言語態 Elderspeak の特徴と高い整合性を示すことを確認

2 提案手法

本研究では，Speech-to-Text および Text-to-Speech 機能を統合したモデルである音声対応 LLM である gpt-4o-audio を用い，難聴者の知覚プロセスを擬似的に再現する対話最適化フレームワークを構築した．システムの全体像を図 1 に示す．

2.1 難聴シミュレーション・ループ

提案手法の全体像は，送り手（システム側）LLM，音声劣化フィルタ，および受け手（ユーザー側）LLM の 3 要素からなるフィードバック・ループで構成される．送り手 LLM が生成した音声信号に対し，加齢性難聴の特性を模した劣化処理を施し，これを受け手 LLM が処理・理解する過程をシミュレーションする．

2.2 音声劣化フィルタの実装

加齢性難聴の主要な特徴を再現するため，Moore らの MSBG モデル [4] のシミュレーションが可能な OSS である pyclarity を用いて以下の 3 つの処理を音声信号に適用する．

- **高音域の減衰**: 周波数ごとの感度低下を模倣し，高周波数帯域（4kHz 付近）でのゲイン減衰をさせ，サ行などの子音の明瞭度を低下させる
- **リクルートメント**: ダイナミックレンジを圧縮し，小さい音は聞こえないが，大きい音は急にうるさく感じる現象を再現
- **スペクトル・スミアリング**: 音声信号のスペクトルを平滑化する処理を行い，難聴特有の音がぼやける現象の再現

2.3 プロンプトの自動最適化プロセス

本研究では，自動プロンプト最適化（Automatic Prompt Optimization; APO）[5] の手法を音声対話ドメインに拡張し，送り手の LLM が生成する発話プロンプトを，以下の 4 つのステップからなる反復的なサイクルによって最適化する．

2.3.1 シミュレーションの実行とエラーの特定

送り手 LLM が現在のプロンプトに基づき，医師からの指示伝達などの特定のタスクに関する音声を生成する．この音声に対し，2.2 節で述べた劣化フィルタを適用し，プロンプトによって難聴ユーザーを模倣させた受け手 LLM が認識・理解を試みる．受け手 LLM は，得られた情報を構造化データとして抽出し，正解データと比較を行う．

2.3.2 マルチモーダル・フィードバックの生成

情報の欠落や誤認が発生した場合，評価用 LLM を用いて，失敗の原因を音響的要因と言語的要因の両面から分析する．

- **音響的分析:** 「服薬の"ふ"が聞き取れなかった」等のフィードバック
- **言語的分析:** 「一文が長く，何を伝えたいのかわからなかった」等のフィードバック

2.3.3 メタプロンプトによる戦略更新

2.3.2 節のフィードバックに基づき，送り手 LLM のシステムプロンプト自体を書き換えるメタプロンプトを実行する．この際，単に語彙を変えるだけでなく，「重要な固有名詞の前にポーズを入れる」「結論から述べる」といった対話戦略レベルの指示を生成・蓄積する．

2.3.4 評価指標に基づく収束判定

最適化の目的関数（Loss Function）として，以下の式で定義される「伝達効率スコア S」を設定する．

$$S = \alpha \cdot (\text{情報伝達率}) - \beta \cdot (\text{修復ターン数})$$

ここで，α, β は重み係数である．シミュレーションを数十回反復し，このスコア S が最大化された時点の発話スタイルを「難聴者向け最適化プロンプト」として抽出する．このプロセスを通じて，音響劣化という物理的制約に対して，自然に配慮表現が創発される仕組みとなっている．

2.3.5 最適化アルゴリズム

本研究では，最適化器としてテキストベースの `gpt-4o` を使用し，メタプロンプト生成と評価指標の計算を行った．各イテレーションで，最適化器は以下の手順を実行する．

1. **候補プロンプトの生成：** 現在のプロンプトと過去の失敗ログ，および評価用 LLM による分析結果をコンテキストとして入力し，複数の改善プロンプト候補を生成する
2. **バッチシミュレーション：** 各候補プロンプトを用いて複数回の対話シミュレーション（モンテカルロ・サンプリング）を行い，統計的な期待スコア $E[S]$ を算出する．
3. **最良プロンプトの選定：** 期待スコアが最も高かったプロンプトを次世代のベースラインとして採用する．
4. **知識の蒸留：** 最適化の過程で見出された「特定の音節を避ける言い換え」や「韻律的特徴の指示」などの有効な定石を，メタ知識としてシス

テムプロンプトの「ガイドライン」セクションに蓄積する．

3 評価実験

3.1 実験設定

提案手法の有効性を検証するため，高齢者支援で頻出する「服薬指導タスク」を用いた評価実験を行った．タスクのシナリオを 200 件作成し，150 件を最適化用，50 件を評価用に分割し，Baseline，Manual，Proposed の 3 条件で評価を実施した．

- **タスク設定:** システム側（医師役）が，服薬指示（薬の種類，飲むタイミング，副作用の注意点など）をユーザー（難聴患者役）に正確に伝える．
- **劣化条件:** 65 歳〜75 歳の標準的な加齢性難聴モデル（4kHz で 30dB の減衰，およびリクルートメント閾値の設定）を適用．
- **ユーザ設定:** 72 歳男性，中等度難聴（高音急墜型），補聴器なし．
- **比較手法:**
  - **Baseline:** 標準的な対話プロンプト「親切かつ丁寧に質問してください」という基本的な指示のみ．
  - **Manual:** 言語聴覚士の知見に基づき，人間が設計した Elderspeak プロンプト．
  - **Proposed:** 提案手法により，32 回のイテレーションを経て自動最適化されたプロンプト．

3.2 評価指標

以下の 3 つの指標を用いて評価を行った．

1. **タスク完遂率 (Task Completion Rate; TCR):** 受け手 LLM が抽出した服薬情報の正解率（Accuracy）．
2. **修復ターン数:** 「え？」「もう一度言ってください」等の聞き返しが発生し，情報の再送が必要になった回数．
3. **平均発話長:** 1 対話あたりの送り手側 LLM の単語数．

3.3 実験結果

実験の結果を表 1 に示す．提案手法は，Baseline および Manual のいずれに対しても優位な性能を示した．平均修復ターン数が 1.9 ターンにまで減少し，タスク完遂率は 85.6%に達した．さらに，平均発話長も 13.8 語と最も短くなり，情報伝達の効率化が図られたことが示された．

表 1 対話性能の比較評価

手法	タスク完遂率 (%)	平均修復ターン数	平均発話長 (word 数)
Baseline	52.4	6.8	32.5
Manual	78.1	2.5	18.2
Proposed	**85.6**	**1.9**	**13.8**

3.4 最適化されたプロンプトの分析

最終的に獲得された「最適化プロンプト」の内容を分析し，LLM が難聴シミュレータとの対話を通じてどのような発話戦略をメタ知識として抽出したかを明らかにする．Baseline プロンプトと比較して，最適化後のプロンプトには以下の 3 つの顕著な指示群が追加されていた．

3.4.1 音素レベルの制約

音響劣化フィルタによる高音域（4kHz 以上）の減衰を回避するため，プロンプトには特定の音素を避ける，あるいは置換するメタ指示が含まれていた．

- **摩擦音の回避:** 「サ行（/s/）」や「ハ行（/h/）」などの摩擦音は，劣化フィルタ下でエネルギーが消失しやすく聞き取り不能として認識される傾向があった．これに対し，プロンプトには「『薬』を『お薬』と呼称し，母音を強調」という具体的指示が創発されていた．
- **破裂音の明瞭化:** 「タ行」などの破裂音に関しては，単独で用いるのではなく，必ず直前に無音区間（ポーズ）を置くよう指示がなされていた．

3.4.2 韻律および時間構造の制御

音声対応 LLM（gpt-4o-audio）の特性を活かし，テキストの構造だけでなく，発話の「タイミング」に関するメタ指示が強化されていた．

- **ポーズの挿入:** 重要な情報（薬の名前，服用タイミング）の前後に明確なポーズを入れるよう指示が追加されていた．
- **発話速度の調整:** 「ゆっくり，はっきりと話す」だけでなく，「一語一語の間隔を均等に保つ」

など，より具体的な韻律制御指示が含まれていた．

3.4.3 情報的冗長性の確保

単なる反復ではなく，音響的に異なる特徴を持つ表現を用いた言い換えの指示が見られた．具体的な例として「朝食を食べましたか？」という質問に対し，「朝ごはんはいただきましたか？」や「今朝，何かお食べになりましたか？」など，異なる語彙・構造で同一情報を伝えるよう指示がなされていた．

4 考察

本研究で LLM が自律的に獲得したプロンプトは，Elderspeak と一致性が高い傾向が見られた．音素選択は，高音急墜型の難聴者に対して合理的であり，LLM が劣化音声から逆算して，プロンプトを音響学的に調整できる能力を持っていることが示唆された．

本手法の最大の特徴は，システム側に明示的な指示を与えずとも，情報伝達率の最大化という報酬関数のみから，結果として配慮に満ちた発話が創発された点にある．これは，対話システムにおけるアクセシビリティの向上が，工学的な最適化問題として解ける可能性を示している．

本研究では，いくつかの重要な限界が存在する．第一に，評価指標が LLM（受け手役）の抽出した情報の正確性に限定されており，実際の難聴高齢者による主観的な受容性や満足度が検証されていない．そのため，獲得された戦略が「子供扱いされている」といった心理的不快感を抱かせるリスク [3] を否定できない．第二に，音響劣化モデルは物理的な歪みを精密に再現する一方で，人間の加齢に伴う中枢性の認知リソースの低下や文脈補完能力の個人差までは完全に模倣できておらず，実環境との乖離が懸念される．さらに，送り手と受け手の双方に同一系列の `gpt-4o` モデルを使用しているため，モデル固有の言語的癖や認識特性に過適合している可能性があり，異なるアーキテクチャや実環境の騒音下における汎用性の検証が今後の課題である．

5 おわりに

本研究では，難聴シミュレーション・ループを用いた LLM の対話プロンプト最適化手法を提案した．実験の結果，難聴者の知覚特性に適応した発話戦略を「システムプロンプト内のメタ指示」として自動獲得でき，対話の効率性と正確性が向上することを確認した．今後は，個別の聴力図に基づいたパーソナライズ化や，実際の高齢者による評価実験を通じて，提案手法の汎用性を検証する予定である．

参考文献

[1] Karen J Cruickshanks, Terry L Wiley, Theodore S Tweed, Barbara EK Klein, Ronald Klein, Julie A Mares-Perlman, and David M Nondahl. Prevalence of hearing loss in older adults in beaver dam, wisconsin: The epidemiology of hearing loss study. **American journal of epidemiology**, Vol. 148, No. 9, pp. 879–886, 1998.

[2] M Kathleen Pichora-Fuller. How social psychological factors may modulate auditory and cognitive functioning during listening. **Ear and Hearing**, Vol. 37, pp. 92S–100S, 2016.

[3] Susan Kemper and Tamara Harden. Experimentally disentangling what's beneficial about elderspeak from what's not. **Psychology and aging**, Vol. 14, No. 4, p. 656, 1999.

[4] Brian CJ Moore and Brian R Glasberg. Simulation of the effects of loudness recruitment and threshold elevation on the intelligibility of speech in quiet and in a background of speech. **The Journal of the Acoustical Society of America**, Vol. 94, No. 4, pp. 2050–2062, 1993.

[5] Reid Pryzant, Dan Iter, Jerry Li, Yin Tat Lee, Chenguang Zhu, and Michael Zeng. Automatic prompt optimization with" gradient descent" and beam search. **arXiv preprint arXiv:2305.03495**, 2023.