Vol.2026-CN-128 No.1
2026/3/9

話者のみに聞こえる疑似フィラーを用いたフィラー抑制手法におけるフィラー提示タイミングの検討

大澤 怜二[1]　川口 一画[2]

概要：発話行動において，フィラーと呼ばれる命題的な価値を持たない慣習化された音声や単語が用いられることがしばしばある．特に独話におけるフィラーの使用は，聴衆が話者に対して抱く印象に悪影響を及ぼすと示されている．そのため，フィラー発生時に即時フィードバックする手法や，事後的にフィラー回数等の総括的情報を提示する手法などが提案されてきた．しかしこれらの手法では，フィラーの削減に注意を割くことでかえって発話の流暢性が低下し，不自然な間や言い淀みが増加するという懸念や，事前に訓練が必要になるという制約が存在する．そこで本研究では，フィラーの発生に先んじて話者にのみ話者自身の声音による疑似フィラーを聞かせる手法を提案する．これはフィラーが発生する理由の一つが沈黙を埋めるためであるという知見に基づき，疑似フィラーを以て沈黙を埋めることによってフィラーの発生を抑制することを目指す．フィラーの発生を抑制し，聴衆の話者に対する印象を向上させるという本研究の目的を達成するために，非発話区間が一定時間以上継続した場合に話者のみに疑似フィラーを提示するシステムを実装した．本稿では，本システムの適切な設定を明らかにするために予備実験を行った．非発話区間の継続時間として異なる閾値を設定した条件下において発話実験および主観評価を行い，閾値を決定した．

1. はじめに

発話行動においては，しばしばフィラーと呼ばれる発声音が生じることが知られている．フィラーとは，「エー」，「アノー」などに代表される命題的な価値を持たない慣習化された音声や単語である．先行研究において，フィラー発生に関わる要因として，発話内容を生成する際の処理の負荷 [1], [5], [12], [21], [27], [28], [36]，言い間違いをした際の修復過程 [21], [36]，注意の分散や不安・緊張などによる認知負荷の増加 [9], [10], [20]，沈黙回避や発話権保持などを目的とした発話継続の意思表明 [5], [26], [30], [36], [37], などが挙げられている．ここで，フィラーは複数人で相互に発話する対話（ダイアローグ）と，一人の話者が一定時間連続して発話する独話（モノローグ）で異なる役割を持つ．本研究では特に，独話におけるフィラーについて着目する．独話におけるフィラーの存在は，聴衆による話者への評価に影響し，自信に対する評価 [16], [19]，能力に対する評価 [6], [16]，誠実さに対する評価 [16] を低下させることが示された．また，発話内容に対する印象も低下する [19] と報告されている．そのため，特に発表場面などにおいては，フィラーの使用頻度を削減することが推奨されている [18], [30], [32]．そこで，発表場面などにおけるフィラーを削減するために，これまで様々な試みがなされてきた．具体的には，聴覚 [13], [35] や視覚 [3], [29]，触覚 [29] などによるリアルタイムのフィードバックや，フィラーの出現頻度や，使用量の事後的な提示といった手段が採られてきた [14], [22], [23], [24]．これらの研究の一部は，訓練を通じてフィラーの削減に寄与することが示されているものの，事前の訓練を前提とするため，即時的な状況や準備を伴わない発話場面への適用には制約があるという課題がある．また，注意がタスクから自己へと逸れることがタスクを行うにあたり必要なリソースを奪い，パフォーマンスに負の影響を及ぼしうることが報告されており [17]，さらに発表場面においては，そのような自己意識の高まりがフィラーの増加を招く可能性が示唆されている [4].

本研究の目的は，発話中に生じるフィラーの使用を抑制することにより，話者に対する印象を向上させることである．フィラーの削減を目的とした既存手法の多くは，話者に対してフィラーの使用を明示的に意識させる介入を含んでおり，その結果として話者の自己意識が高まり，発話生成過程に悪影響を及ぼす可能性が指摘されている．すなわ

1 筑波大学 情報理工学位プログラム
Master's Program in Computer Science, University of Tsukuba
2 筑波大学 システム情報系
Faculty of Engineering, Information and Systems, University of Tsukuba

ⓒ2026 Information Processing Society of Japan

ち,「フィラーを減らそうと意識すること」自体が，かえって発話の非流暢性を助長し得るという課題が存在する．そこで本研究ではこの課題に対し，フィラーの発生に先んじて話者にのみ話者自身の疑似フィラーを聞かせるというアプローチを採る．ここでいう疑似フィラーとは，事前に録音した話者自身のフィラー音声を，話者が普段聞いている音声に近くなるように変換したものである．沈黙を疑似フィラーによって埋めることで，沈黙に対する話者の主観的な負担を低減でき，その結果として実際のフィラー挿入が抑制されると考えられる．さらに，話者にフィラーを自己生成したと錯覚させることができれば，発話生成過程への影響を抑えながら，フィラー頻度のみを低減できる可能性がある．

本手法を実現するために，話者の発話をリアルタイムに取得し，発話状態に応じて疑似フィラーの提示を制御するシステムを設計した．本システムでは，話者の音声を常時取得し音量情報および音声区間検出によって非発話区間を判定，非発話区間が一定時間以上継続した場合に話者のみに疑似フィラーを提示する．本稿では，この提示を適切なタイミングで行うために必要となる無音区間が継続する時間の閾値を定めることを目的として，閾値決定のための予備実験を実施した．

2. 関連研究

2.1 フィラー発生の原因

先行研究は，フィラーが単なる話し癖ではなく，発話生成過程の認知的負荷や話者の心理状態，および沈黙回避・発話権保持といった機能的要請と関連することを示している．発話内容の計画・修復が滞る場面では，思考時間を確保するためにフィラーが用いられやすい [12], [36]．また，語彙検索における負荷（低頻度語や予測しにくい語の産出など）が高い場合にも，言い淀みやフィラーが生じやすい [1], [21]．さらに，フィラーは遅延を聞き手に示して割り込みを抑制し，沈黙回避や発話権保持に寄与するほか [5], [26], [30]，不安や自己注目の高まりが認知的負荷を増大させ，フィラーの増加につながることも報告されている [9], [10].

2.2 フィラーの聴衆に対する影響

独話におけるフィラーは，話者自身の発話生成過程を反映する現象である一方，聴衆による話者評価に対して負の影響を及ぼすことが多く報告されている．

まず，話者の自信に対する評価への影響が指摘されている．Kirkland ら [16] は，フィラーや言い直しを含む発話が，流暢な発話と比較して聴衆によって知覚される話者の自信を有意に低下させることを示している．これは，非流暢性が話者を「自信がない」「ためらっている」人物として印象づける要因となることを示唆するものである．同様に，Laske ら [19] は，スピーチ中にフィラーが高頻度で出現する条件において，話者が自信に欠け，緊張していると評価されやすくなることを報告している．

次に，話者の能力に対する評価への影響も明らかにされている．Kirkland ら [16] の研究では，フィラーや言い直しの増加が，話者の能力評価，とりわけ新しい内容を分かりやすく伝える能力やタスク遂行能力の評価を低下させることが示された．また，Conrad ら [6] は電話調査における発話分析を通じて，適度なフィラー使用は必ずしも否定的に作用しない一方で，過度なフィラー使用は話者を「能力不足」「不慣れ」と印象づけ，結果として相手の協力意欲を低下させることを報告している．これらの知見は，フィラーの量が聴衆の能力評価に影響を及ぼす重要な要因であることを示している．

さらに，話者の誠実さや信頼性に対する評価についても影響が報告されている．Kirkland ら [16] は，フィラー単独では誠実さの評価に明確な影響は見られないものの，フィラーと言い直しが共起して頻出する場合には，誠実さの評価が有意に低下することを示している．これは，過度な非流暢性が，話者の信頼性や一貫性に対する疑念を生じさせる可能性を示唆している．

加えて，発話内容そのものの知覚にも影響が及ぶことが指摘されている．Laske ら [19] は，スピーチにおいてフィラー音が高頻度で挿入された場合，聴衆が「話の構成が不明瞭である」「非言語的な振る舞いが不十分である」といった評価を下す傾向を示している．これは，フィラーの多用が，実際の内容や構造とは無関係に，スピーチ全体の質に対する知覚を歪める可能性があることを示している．

このように，独話，特に発表場面のように話者が一方向的に情報を伝達する状況において，フィラーの頻出が聴衆の評価を総合的に低下させる要因となり得る．

2.3 話者への介入を通じてフィラー削減を試みた研究

話者に対してフィラー削減促すことを目的とした先行研究は，フィラー使用に対する介入のタイミングに基づき，リアルタイムフィードバックによる手法，および発話後の事後的フィードバックによる手法に大きく分類できる．

2.3.1 リアルタイムフィードバックによる介入

発話中にフィラーを検知し，その場で話者にフィードバックを与えることで，フィラー使用の抑制を試みる研究がこれまでに報告されている．

Zhang [35] らは，日常会話を対象として，フィラーをリアルタイムに検知し，聴覚的フィードバックを話者に提示するウェアラブルシステムを提案している．この研究では，発話中に即時的なフィードバックを与えることで，話者自身がフィラーの使用に気づくことを促す設計である．

Schneider[29] らは，発話中のフィラーをリアルタイムに検知し，視覚的および触覚的フィードバックを提示する

Presentation Trainer を提案している．また，発表練習を対象として，話者の行動に応じたリアルタイムの反応提示（例：仮想聴衆による非言語的反応）により，間接的に行動変容を促す枠組みも提案されている [3].

これらの研究は，リアルタイムにフィラー使用を検知し，話者に対してその使用を明示的に示すことで，行動変容を促そうとするアプローチである．一方で，先行研究では，話者に対して自身のパフォーマンスや評価対象となる行動を強く意識させることが，必ずしも発話の質の向上につながらない可能性が指摘されている．

Kluger と DeNisi [17] は，フィードバックが注意を自己評価へ向けさせる場合にパフォーマンスが低下し得ることを示している．また，Christenfeld と Creager [4] は，意識的な自己モニタリングがフィラー増加につながり得ることを報告している．したがって，発話中にフィラーを「減らすべき対象」として強く意識させる介入は，非流暢性を助長するおそれがある．

2.3.2 事後的フィードバックによる内省促進

フィラー削減に関するもう一つの主要なアプローチとして，発話後にフィラーの使用状況を可視化し，話者の内省を促す研究が報告されている．これらの研究では，発話中に介入を行うのではなく，録音・録画された発表を分析し，結果を可視化することで，話者自身による自己調整を支援する点に特徴がある．

Ochoa らは，発表全体を対象とした事後フィードバックを提示するオープンソース基盤 OpenOPAF [24] を提案している．OpenOPAF では，フィラーについても発話中のどのタイミングで生起したかが可視化され，話者が自身の発話構造や癖を俯瞰的に理解できるよう設計されている．

また，Mateo ら [22] は，自動検出されたフィラー頻度をグラフとして提示する事後フィードバックを用い，学生の口頭発表を対象とした実践的な検討を行っている．この研究では，発表後に自身のフィラー使用状況を可視化して提示することで話者の自覚を促し，複数回のセッション間でフィラー使用頻度に有意な変化が見られたことが報告されている．ここでは，詳細な行動指示を与えるのではなく，客観的な使用状況を示すこと自体が内省を引き起こす手段として位置づけられている．さらに，トランスクリプト上の強調表示や割合指標など，比較的単純な可視化によって自己認識を促す事後フィードバックも報告されている [14]．このように，事後的フィードバックによるアプローチは，発話中の注意分断や自己意識の過度な喚起を避けつつ，話者自身の内省を通じた行動変容を促す手法として位置づけられてきた．ただし，これらの事後的フィードバック手法はいずれも，発話後の振り返りや反復的な発表を前提とした設計であり，即時的な発話場面や準備を伴わない状況への適用には制約がある．

2.4 事後的なフィラーの音声除去・置換に関する研究

発話に含まれるフィラーは，必ずしも理解や記憶を阻害するものではなく，聴取者の認知処理を支援する可能性があることが報告されている．先行研究では，フィラーや韻律的な非流暢性が，後続する発話内容の複雑さや話者の認知状態を示す手がかりとして機能し，結果として聴取者の理解や想起成績を向上させる場合があることが示されてきた [2], [8], [31].

こうした知見を背景として，録音後の音声からフィラーを除去・置換した音声を用い，聴取者の理解度や記憶への影響を検討する研究も行われている．この研究では，フィラーを保持した条件の方が，内容理解や記憶成績において有利となる結果が報告されることもあり [8]，フィラーを一律に排除すべき対象とみなさない立場が示されている．

一方で，これらの研究は主として聴取者側の理解や記憶に焦点を当てており，発表やスピーチといった独話場面における話者の発話体験や，聴衆による話者への印象評価を主たる対象とはしていない．したがって，本研究は，フィラーが理解を促進しうるという知見を踏まえつつも，事後的な音声加工による理解支援とは異なり，話者の発話行動およびその社会的・印象的影響に着目する点で，これらの研究とは異なる立場を取る．

3. 提案システム

本章では，本研究にて用いるシステムの設計について述べる．はじめに，疑似フィラー提示という介入を成立させるための設計指針を示し，続いて，それを実現するシステム構成および各処理の詳細について述べる．

3.1 設計指針

本研究では，一定以上の沈黙が生じたときに介入を行うことを基本方針とする．発表場面において，話者は沈黙を主観的に長く知覚しやすく，その結果として沈黙を回避するためにフィラーを使用することが指摘されている [30]．また，話者は沈黙が一定の長さを超えそうであることを予測した時点で，その沈黙を回避する目的でフィラーを発話する可能性が高いことが示唆されている [26]．以上を踏まえ，本研究では沈黙が顕在化する前段階において疑似フィラーを提示する設計を採用する．

3.1.1 疑似フィラーとして用いるフィラー

疑似フィラーとして用いる音声には，日本語において特に出現頻度が高いフィラーである「エー」を採用する．日本語の講義・発表場面におけるフィラー分布に関する先行研究において，「エー」は他のフィラーと比較して高頻度で使用されることが報告されている [33]．また，本研究では話者に違和感を与えない提示を重視し，事前に録音した話者自身のフィラー音声を用いることで，第三者の音声による介入と比較して自己生成的な感覚を保持できると考えた．

3.1.2 Filler Unit

本研究では，水上・山下による定義に基づき，フィラーを挟んだ前後の発話間の間隔を FU（Filler Unit），フィラーに前置する発話末からフィラー開始までの間隔を FFU（Front part of Filler Unit），フィラー末からフィラーに後置する発話頭までの間隔を BFU（Back part of Filler Unit）として扱う [37]．FFU の平均長は言語によって異なり，英語では約 350 ms，オランダ語では約 390 ms[7]，ハンガリー語では約 620 ms とされている [11]．日本語における FFU 長の体系的調査は確認されていないため，本研究ではパイロットスタディを行い，疑似フィラー再生の閾値を決定する．

3.2 システム構成

本システムは，話者が装着するヘッドセット型デバイス，音声処理を行うホスト端末（PC），およびホスト端末上で動作する音声解析・制御ソフトウェアから構成される．ヘッドセット型デバイスは，話者の音声入力および疑似フィラー音声の出力を担い，取得された音声信号はホスト端末へ送信される．ホスト端末では，入力された音声に対して逐次的な解析処理を行い，発話状態に応じて疑似フィラーの提示を制御する．システム構成の概要図を図 1 に示す．

3.2.1 疑似フィラーの作成，および再生

疑似フィラーは，実験実施前に話者自身のフィラー音声を録音することで作成する．録音に際して発声の前に 0.5 秒の無音区間を設け，その区間にて録音した音声を用いて環境音を除去した．録音に際して話者が 3 秒間「エー」という発声を継続的に行った後，この音声から振幅のブレが小さい 1 秒区間を抽出した．抽出した音声に対して，話者が自己の声を内的に知覚している音質に近づけるための前処理を行う．具体的には，低音域あるいは中音域の強調や，高音域を抑制する変換を施す [25]．音声の個人差を加味し，複数の音質変換候補を生成した後，試聴により最適なものを選択できる手順を採用した．抽出した音声に対して 6 種類の変換（無変換，高域の抑制（弱／強），中域の強調（弱／強），低域の強調と高域の抑制）を作成し，話者が自己聴取に近いと感じる音質を選択した．

疑似フィラーの再生においては，提示開始時の遅延を最小限に抑えるため，疑似フィラー音声を常時ループ再生する方式を採用する．具体的には，事前に録音・前処理した疑似フィラー音声をシステム起動時から継続的に再生状態とし，通常時は音量を 0 に設定する．FFU の検知処理により，非発話区間の継続時間が設定した閾値を超えたと判定された場合にのみ，ループ再生中の疑似フィラー音声の音量を即時的に上昇させることで，話者に対して疑似フィラーを提示する．また，話者が再び発話状態に遷移した場合には，音量を再度 0 に設定する．このように，再生と停止を切り替えるのではなく，音量制御のみによって疑似フィラーの提示・非提示を切り替えることにより，再生開始に伴うバッファリングや処理遅延を回避する．

3.2.2 FFU の検知

FFU の検知処理は，Python を用いて実装した．音声の入出力には sounddevice ライブラリを用い，ヘッドセット型デバイスから話者の音声信号をリアルタイムに取得する．取得した音声信号は一定長のフレームに分割され，各フレームに対して逐次的な解析を行う．

非発話区間の検出には，音量情報と音声区間検出（Voice Activity Detection, VAD）を併用する．音量情報の算出には NumPy を用い無音区間を判定する．無音区間でなかった区間について，音声信号に音声活動が含まれるか否かを判別するために WebRTC VAD [34] を用いる．

本研究では，音量があらかじめ設定した閾値以下であり，かつ VAD により非発話と判定されたフレームが連続して出現した区間を非発話区間として扱う．この非発話区間の継続時間を逐次計測し，その長さが設定した閾値を超えた場合に，フィラーが発話される前段階，すなわち FFU に該当する状態であると判定する．

このように，音量情報と VAD を組み合わせることで，環境雑音や瞬間的な音声変動による誤検知を抑えつつ，発話状態の変化の検知を試みる．

4. 予備実験

本研究では，話者のみに聞こえる疑似フィラーを，話者が沈黙を過度に長く知覚する前段階で提示することで，フィラーの発生や不必要な沈黙回避行動を抑えることをねらう．このとき，疑似フィラーの挿入タイミングは「非発話区間の継続時間が閾値を超えた場合」に決定されるため，主実験に先立ち，適切な閾値の目安を得ることが必要である．そこで予備実験として，実験参加者自身が閾値を調整し，「言い淀みや詰まりが生じたときには挿入され，すらすら読んでいるときには挿入されない」状態となる閾値を探索した．併せて，疑似フィラーが読みやすさ・不快感・同一感に与える影響を調査した．

4.1 実験タスク

本実験では，疑似フィラーを挿入する閾値を，実験参加者が自身で調整しながらの音読をタスクとして設定した．音読における課題文として，正岡子規の『病牀六尺』を採用した [38]．採用の理由としては，現代とは異なる文体や表現を含む文章を用いることにより言い淀みや詰まりを誘発し，音読のように間が生じにくい状況でも非発話区間が生じやすくするためである．また，ルビにより間が生じる頻度が減ってしまわないように，青空文庫のルビを非表示にするツール [15] を用いて参加者に提示した．

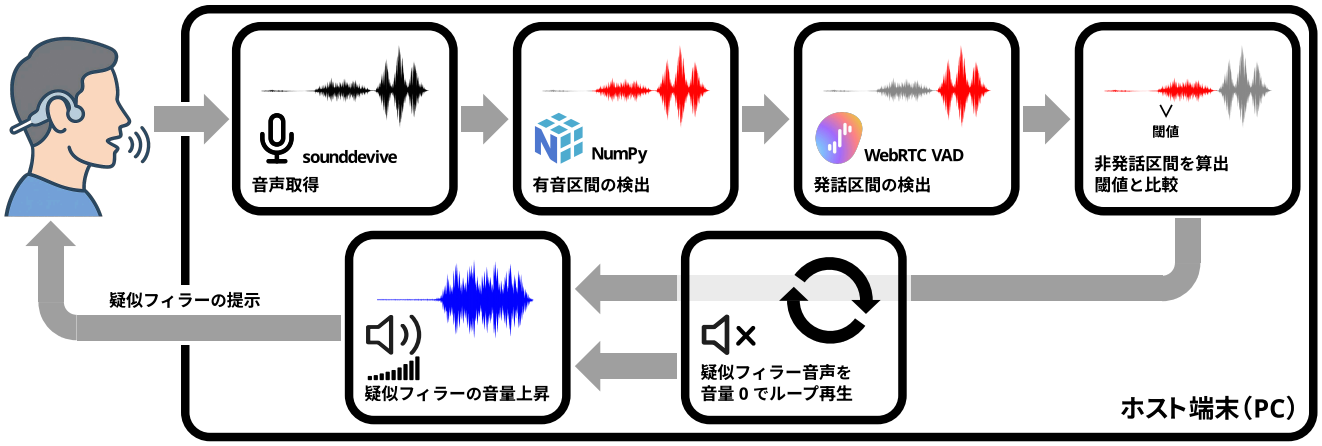

図 1: システム概要

4.2 実験参加者

実験参加者は研究室内で募集した計 7 名（男性 4 名，女性 3 名，平均 22.4 歳，標準偏差 1.13 歳）の大学生および大学院生である．実験全体の所要時間は 30 分程度であった．

4.3 実験手順

実験では，最初に事前アンケートを実施した．その後，システムの説明，タスクの説明，注意事項の説明を行った．注意事項として，話す速度，間のとり方，抑揚などを音読用に工夫せず読むよう指示した．

実験参加者は，練習として課題文における最初の 1 段落をシステムを用いず音読した．その後，同課題文の 2 節まで，音読を進めながら閾値を調整した．閾値は 100 ms から開始し，調整の判断基準として下記を両方満たすよう指示した．

- 言い淀んだ時には疑似フィラーが挿入されること
- すらすら読めている時には挿入されないこと

閾値の調整は音読中に逐次行い，概ね 1 行程度読むごとに，必要に応じてボタン操作により閾値を変更するよう求めた．調整用 GUI には「-10 ms」「-50 ms」「+10 ms」「+50 ms」の 4 種類のボタンを用意し，参加者は上記の判断基準を満たす値に近づけるため，2 節末に到達するまで閾値を増減させてもよいものとした．システム GUI を図 2 に示す．

また，2 節末までの調整で不十分であると感じた場合には，調整を継続するために 3 節以降も音読するように指示した．音読終了後にはタスクに対するアンケートを実施した．また，アンケート回答後にインタビューを行った．

4.4 実験環境

実験環境を図 3 に示す．実験は静かな室内環境で実施した．実験端末は MacBook Air（Apple M4, メモリ: 16 GB, OS: macOS Sequoia）である．音声の入出力には骨伝導ヘッドセット（Shokz OpenRun Pro，Bluetooth 5.1）を用いた．表示環境として，PC のメイン画面に課題文を表示し，1 行あたり 51 文字が表示される程度に拡大率を調整した．また，iPad Air（11 インチ）を Sidecar によりサブモニタとして接続し PC の右側に配置，システムの GUI を表示した．参加者は当該 GUI ウィンドウから疑似フィラー挿入の閾値を調節した．

図 2: システム GUI

4.5 収集データ

実験では，事前アンケートにおいて年齢，性別，所属，これまでのスピーチや朗読，発話表現に関する専門的なトレーニングそれに類する経験の有無を収集した．事後アンケートでは，最終的な閾値，および以下の 4 項目を 7 段階のリッカート尺度で評価させた．

- システムの介入に寄らない，原稿そのものの読みやすさ（1: 読みにくい–7: 読みやすい）
- 疑似フィラー挿入の読みやすさへの影響（1: 影響しない–7: 読みやすい）
- 疑似フィラー挿入の不快感（1: 不快でない–7: 不快である）
- 疑似フィラーの自分の音声との同一感（1: 同一であると感じない–7: 同一であると感じる）

図 3: 実験環境

5. 予備実験の結果および考察

本章において，4 章にて述べた実験における結果を示し，閾値の決定，および考察を行う．

5.1 最終閾値の分布およびアンケート結果

実験によって得られた最終的な閾値と事後アンケートの結果を表 1 に示す．各参加者が最終的に選択した閾値は，の平均は 260 ms ，標準偏差は 94.34 ms であった．次に，4 つのアンケート結果を示す．原稿そのものの読みやすさ（1：読みにくい–7：読みやすい）は平均 1.43（SD=0.53）であった．疑似フィラー挿入の読みやすさへの影響（1：影響しない–7：影響する）は平均 4.43（SD=1.72）であった．疑似フィラー挿入の不快感（1：不快でない–7：不快である）は平均 3.57（SD=1.51）であった．疑似フィラーの自分の音声との同一感（1：同一であると感じない–7：同一であると感じる）は平均 4.86（SD=1.35）であった．

5.2 考察

本予備実験では，参加者自身が「言い淀み時には挿入され，流暢時には挿入されない」と感じる境界として閾値を探索した．以下では，表 1 の結果とインタビューの内容を踏まえ，疑似フィラーが参加者の体験に与える影響と，閾値設計に関する示唆を述べる．

5.2.1 疑似フィラーが沈黙の主観的負担へ与える効果

自由記述では，疑似フィラーについて「口が止まっていることが明示されて良かった」（P1），「テンポを掴めた／心地よい」（P2）といった反応が得られた．また，「詰まっているときに何も発していないように感じず，詰まっているように感じなかった」（P6）という報告もあり，疑似フィラーが沈黙を「無音」ではなく「何らかの発声が継続している状態」として知覚させ，主観的な停滞感を弱める可能性が示唆される．これにより，本研究が想定する発表場面において，沈黙に伴う心理的負担を軽減することが期待される．一方で，「急かされる用に感じた」（P3），「漢字の読み方を考える時邪魔に感じた」（P5）といった意見も見られた．これらの意見は非発話区間中に疑似フィラーが鳴り続けていることによる影響と考えられるため，主実験においては，疑似フィラー長を適切に調整することが望ましい．

表 1: 最終閾値およびアンケート結果（$n = 7$）

参加者	最終閾値 [ms]	原稿の読みやすさ	影響度	不快感	同一感
P1	320	2	5	5	6
P2	340	2	6	2	5
P3	240	1	6	4	6
P4	400	1	2	5	3
P5	180	2	5	5	6
P6	150	1	5	2	3
P7	190	1	2	2	5
平均	260.00	1.43	4.43	3.57	4.86
SD	94.34	0.53	1.72	1.51	1.35

5.2.2 閾値の目安と個人差

最終閾値は 150–400 ms に分布し，変動係数も 0.36 と小さくないことから，適切な挿入タイミングは参加者間で一様ではない．一方で中央値は 240 ms であり，200–300 ms 付近に集中する傾向も見られた．したがって主実験では，260 ms 前後を初期値（デフォルト）として提示し，参加者ごとに微調整できる設計とする．

6. 予備実験における制約および今後の課題

本予備実験には，手続き・課題設定・評価方法に関する制約がある．本章ではそれらを整理し，今後の課題を述べる．

6.1 参加者数と募集範囲

本予備実験の参加者数は $n = 7$ と少なく，研究室内での実施により参加者の属性も限定されている．そのため，結果の一般化には限界がある．主実験では，より広く参加者を募集し，参加者数を増やした上で検証を行う．

6.2 閾値調整手続きの影響

本予備実験では，参加者が音読しながら閾値を調整する手続きを採用した．この手続きでは「調整すること」自体が参加者の注意や発話行動に影響し，介入の受け止め方が主実験（閾値を固定して使用する状況）と異なる可能性がある．主実験では，キャリブレーション（調整）区間と評価区間を分離し，評価は閾値確定後に実施するなど，手続き由来の影響を低減する．

6.3 課題と想定場面の差

本予備実験の課題は難読文の音読であり，主実験で想定する発表場面とは発話の性質が異なる．例えば「スピーチ

は普段よりゆっくりに話すのではないか」（P7）という意見が得られており，話速や沈黙の現れ方が変化する可能性がある．その結果，予備実験で得られた閾値や主観評価が，発表場面では異なる形で現れる可能性がある．主実験では，発表場面における話速や沈黙の特徴もあわせて計測し，音読課題で得られた所見との比較を行う．

6.4 音声における自分の声との同一性に関する課題

疑似フィラーの自分の音声との同一感は高めに評価した参加者がいる一方で，低めに評価した参加者も存在した．自由記述では「自分の声とは違う，機械的な音声に感じた」（P5），「自分の声を録音しているからそういった刷り込みはあるが，自分の声かなとは思わなかった」（P7）といった意見が得られている．このことから，音声変換（あるいは提示音声生成）の品質が受容性に影響する可能性がある．また，本研究では「自分の声に近い提示音声が有効である」という前提に立って設計しているが，そもそも提示音声として「自分の声」を用いること自体が最適であるかは仮説であり，その妥当性は今後の検証が必要である．例えば，sin 波や他人の声と比較し，同一感や違和感だけでなく，フィラーの発生頻度や発話負荷への影響を定量的に評価することで，自分の声を採用するかことが適切かの判断を行う．主実験に向けて，音色や抑揚の自然さを高める調整，および個人差に応じた生成・補正手段の検討など，音声変換機能の改善を進める．

7. おわりに

本研究では，発表を始めとした独話場面におけるフィラーの使用を抑制し，聴衆が抱く話者への印象を向上させることを目的として，フィラーの発生に先んじて話者にのみ話者自身の声音による疑似フィラーを聞かせる手法を提案した．システムの有効性を検証する主実験に先立ち，疑似フィラーの提示タイミング（非発話区間の継続時間の閾値）の目安を得るため，参加者が音読しながら閾値を自己調整する予備実験を行った．閾値が「言い淀んだ時には挿入され，すらすら読めている時には挿入されない」境界となるよう探索させた結果，最終閾値は 150–400 ms に分布し平均 260 ms であり，個人差がある一方で 200–300 ms 付近に集中する傾向が見られた．主観評価と自由記述からは，疑似フィラーが沈黙の主観的負担や停滞感を弱めることに寄与しうる可能性が示唆された一方，提示が「急かされる」「思考を妨げる」と感じられる場合もあり，提示長や介入強度の調整が重要であることが示唆された．これらより，主実験では 260 ms 前後を初期値として提示しつつ，参加者ごとの微調整を可能にする設計とする．

一方で，予備実験には参加者数・募集範囲の限定，音読中の閾値調整という手続きそのものの影響，音読課題と想定場面の差，ならびに疑似フィラーの同一感（自己の声らしさ）に関する課題がある．今後は，これらの課題を補完するための改善を行った後，発表場面に近い課題設定で検証を行う．また，音声変換の品質向上や個人差に応じた補正，提示長・音量・提示条件の最適化，および話速や沈黙特性を踏まえた適応的な閾値設計を進めることで，話者に違和感を与えずに沈黙負担を軽減し，結果としてフィラーを自然に減らす介入としての実用可能性を高めていく．

参考文献

[1] Beattie, G. and Butterworth, B.: Contextual Probability and Word Frequency as Determinants of Pauses and Errors in Spontaneous Speech, *Language and Speech*, Vol. 22 (online), DOI: 10.1177/002383097902200301 (1979).

[2] Brennan, S. E. and Williams, M.: The Feeling of Another's Knowing: Prosody and Filled Pauses as Cues to Listeners about the Metacognitive States of Speakers, *Journal of Memory and Language*, Vol. 34, No. 3, pp. 383–398 (online), DOI: 10.1006/jmla.1995.1017 (1995).

[3] Chollet, M., Wörtwein, T., Morency, L.-P., Shapiro, A. and Scherer, S.: Exploring Feedback Strategies to Improve Public Speaking: An Interactive Virtual Audience Framework, *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 1143–1154 (online), DOI: 10.1145/2750858.2806060 (2015).

[4] Christenfeld, N. and Creager, B.: Anxiety, alcohol, aphasia, and ums, *Journal of Personality and Social Psychology*, Vol. 70, No. 3, pp. 451–460 (online), DOI: 10.1037/0022-3514.70.3.451 (1996).

[5] Clark, H. H. and Fox Tree, J. E.: Using Uh and Um in Spontaneous Speaking, *Cognition*, Vol. 84, No. 1, pp. 73–111 (online), DOI: 10.1016/S0010-0277(02)00017-3 (2002).

[6] Conrad, F. G., Broome, J. S., Benkí, J. R., Kreuter, F., Groves, R. M., Vannette, D. and McClain, C.: Interviewer Speech and the Success of Survey Invitations, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, Vol. 176, No. 1, pp. 191–210 (online), DOI: 10.1111/j.1467-985X.2012.01064.x (2012).

[7] Fox Tree, J. E.: Listeners' Uses of Um and Uh in Speech Comprehension, *Memory & Cognition*, Vol. 29, No. 2, pp. 320–326 (online), DOI: 10.3758/bf03194926 (2001).

[8] Fraundorf, S. H. and Watson, D. G.: The Disfluent Discourse: Effects of Filled Pauses on Recall, *Journal of Memory and Language*, Vol. 65, No. 2, pp. 161–175 (online), DOI: 10.1016/j.jml.2011.03.004 (2011).

[9] Garcia-Lopez, L. J., Díez-Bedmar, M. B. and Almansa Moreno, J. M.: From Being a Trainee to Being a Trainer: Helping Peers Improve their Public Speaking Skills, *Revista de Psicodidactica / Journal of Psychodidactics*, Vol. 18, pp. 331–342 (online), DOI: 10.1387/RevPsicodidact.6419 (2013).

[10] Goberman, A. M., Hughes, S. and Haydock, T.: Acoustic Characteristics of Public Speaking: Anxiety and Practice Effects, *Speech Communication*, Vol. 53, No. 6, pp. 867–876 (online), DOI: 10.1016/j.specom.2011.02.005 (2011).

[11] Gósy, M.: Occurrences and Durations of Filled Pauses in Relation to Words and Silent Pauses in Spontaneous Speech, *Languages*, Vol. 8, No. 1, p. 79 (online), DOI: 10.3390/languages8010079 (2023).

[12] Goto, M., Itou, K. and Hayamizu, S.: Speech Comple-

tion: On-Demand Completion Assistance Using Filled Pauses for Speech Input Interfaces, *Proceedings of the 7th International Conference on Spoken Language Processing (ICSLP 2002)*, pp. 1489–1492 (online), DOI: 10.21437/ICSLP.2002-449 (2002).

[13] Hazel, M., McMahon, C. and Schmidt, N.: Immediate Feedback: A Means of Reducing Distracting Filler Words during Public Speeches, *Basic Communication Course Annual*, Vol. 23, p. 6 (2011).

[14] Hoque, M. E., Courgeon, M., Martin, J.-C., Mutlu, B. and Picard, R. W.: MACH: My Automated Conversation Coach, *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 697–706 (online), DOI: 10.1145/2493432.2493502 (2013).

[15] Kepe: 青空文庫 ルビ削除ツール α 版, `https://www.kepe.net/ruby/`. accessed: 2026-02-08.

[16] Kirkland, A., Gustafson, J. and Székely, É.: Pardon My Disfluency: The Impact of Disfluency Effects on the Perception of Speaker Competence and Confidence, *Proceedings of Interspeech 2023*, pp. 5217–5221 (online), DOI: 10.21437/Interspeech.2023-887 (2023).

[17] Kluger, A. N. and DeNisi, A.: The Effects of Feedback Interventions on Performance: A Historical Review, a Meta-Analysis, and a Preliminary Feedback Intervention Theory, *Psychological Bulletin*, Vol. 119, No. 2, pp. 254–284 (online), DOI: 10.1037/0033-2909.119.2.254 (1996).

[18] Kowal, S., O'Connell, D. C., Forbush, K., Higgins, E. and Clarke, D.: Interplay of Literacy and Orality in Inaugural Rhetoric, *Journal of Psycholinguistic Research*, Vol. 26, No. 1, pp. 1–31 (online), DOI: 10.1023/A:1025043620499 (1997).

[19] Laske, M. M. and DiGennaro Reed, F. D.: Um, so, like, do Speech Disfluencies Matter? A Parametric Evaluation of Filler Sounds and Words, *Journal of Applied Behavior Analysis*, Vol. 57, No. 3, pp. 574–583 (online), DOI: 10.1002/jaba.1093 (2024).

[20] Lay, C. H. and Burron, B. F.: Perception of the Personality of the Hesitant Speaker, *Perceptual and Motor Skills*, Vol. 26, No. 3, pp. 951–956 (online), DOI: 10.2466/pms.1968.26.3.951 (1968).

[21] Levelt, W. J. M.: Monitoring and Self-Repair in Speech, *Cognition*, Vol. 14, No. 1, pp. 41–104 (online), DOI: 10.1016/0010-0277(83)90026-4 (1983).

[22] Mateo, W., Eras, L., Carvajal, G. and Domínguez, F.: Detecting Speech Disfluencies Using Open-Source Tools in Automatic Feedback Systems for Oral Presentation Training, *Proceedings of the 16th International Conference on Computer Supported Education (CSEDU)*, pp. 213–220 (online), DOI: 10.5220/0012622100003693 (2024).

[23] Ochoa, X., Domínguez, F., Guamán, B., Maya, R., Falcones, G. and Castells, J.: The RAP System: Automatic Feedback of Oral Presentation Skills Using Multimodal Analysis and Low-Cost Sensors, *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, pp. 360–364 (online), DOI: 10.1145/3170358.3170406 (2018).

[24] Ochoa, X. and Zhao, H.: OpenOPAF: An Open Source Multimodal System for Automated Feedback for Oral Presentations, *Journal of Learning Analytics*, Vol. 11, No. 3, pp. 224–248 (online), DOI: 10.18608/jla.2024.8411 (2024).

[25] r/audioengineering: Is there a way to edit our voice, in such a way that it sounds like how we hear our voice in heads? : r/audioengineering, `https://www.reddit.com/r/audioengineering/comments/12choib/is_there_a_way_to_edit_our_voice_in_such_a_way/` (2023). reddit; accessed 2026-02-08.

[26] Rose, R. L. and Watanabe, M.: A Crosslinguistic Corpus Study of Silent and Filled Pauses: When Do Speakers Use Filled Pauses to Fill Pauses?, *Proceedings of the 19th International Congress of Phonetic Sciences (ICPhS 2019)*, pp. 2615–2619 (2019).

[27] Schachter, S., Christenfeld, N., Ravina, B. M. and Bilous, F.: Speech Disfluency and the Structure of Knowledge, *Journal of Personality and Social Psychology*, Vol. 60, pp. 362–367 (1991).

[28] Schnadt, M. J. and Corley, M.: The Influence of Lexical, Conceptual and Planning Based Factors on Disfluency Production, *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 28 (2006).

[29] Schneider, J., Börner, D., van Rosmalen, P. and Specht, M.: Presentation Trainer, Your Public Speaking Multimodal Coach, *Proceedings of the 2015 ACM International Conference on Multimodal Interaction*, pp. 539–546 (online), DOI: 10.1145/2818346.2830603 (2015).

[30] Seals, D. R. and Coppock, M. E.: We, um, have, like, a problem: Excessive Use of Fillers in Scientific Speech, *Advances in Physiology Education*, Vol. 46, No. 4, pp. 615–620 (online), DOI: 10.1152/advan.00110.2022 (2022).

[31] Watanabe, M., Hirose, K., Den, Y. and Minematsu, N.: Filled Pauses as Cues to the Complexity of Following Phrases, *Proceedings of Interspeech 2005*, pp. 37–40 (online), DOI: 10.21437/Interspeech.2005-33 (2005).

[32] Watanabe, M. and Toyama, S.: 『日本語話し言葉コーパス』と対照可能にデザインされた英語話し言葉コーパスにおけるフィラーの分布の特徴, *NINJAL Research Papers*, Vol. 12, pp. 141–203 (2017).

[33] Watanabe, M. and Ishi, C. T.: The Distribution of Fillers in Lectures in the Japanese Language, *Proceedings of the Sixth International Conference on Spoken Language Processing (ICSLP 2000)*, Vol. 3, pp. 167–170 (2000).

[34] Wiseman, J.: py-webrtcvad, `https://github.com/wiseman/py-webrtcvad`. GitHub repository; accessed 2026-02-08.

[35] Zhang, Y., Janaka, N., Ram, A., Yin, P., Tian, Y., Zhao, S. and Dragicevic, P.: WSCoach: Wearable Real-time Auditory Feedback for Reducing Unwanted Words in Daily Communication, *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, Vol. 9, No. 3, p. 152 (online), DOI: 10.1145/3749531 (2025).

[36] Zulhemindra, Z., Munir, S., Yulnetri, Y. and Putra, K. P.: Investigating Filled Pauses Found in English Students' Conversation, *Ahmad Dahlan Journal of English Studies*, Vol. 9, No. 1 (online), DOI: 10.26555/adjes.v9i1.35 (2022).

[37] 悦雄水上, 耕二山下: 対話におけるフィラーの発話権保持機能の検証, 認知科学, Vol. 14, No. 4, pp. 588–603（オンライン）, DOI: 10.11225/jcss.14.588 (2007).

[38] 正岡子規: 病牀六尺, `https://www.aozora.gr.jp/cards/000305/files/43537_41508.html`. 青空文庫, accessed: 2026-02-08.