

変換聴覚フィードバックを利用したオンラインスピーチの発話および印象改善

大沼 恵生¹ 大澤 恵二¹ 半田 篤史¹ 川口 一画²

概要：スピーチ中の緊張により、話者の発話の話速や声量が増加することがある。これらの発話変化はスピーチ聴衆の印象低下につながる。そこで本研究では、発話音声に音響的変換処理を加えたものを話者のみ聞かせる手法である変換聴覚フィードバック（Altered auditory feedback, AAF）に注目し、話者の話速の低下および声量の減少を誘発することによる、発話および印象の改善手法を提案する。話者の発話のピッチおよび音量を増加させた AAF をスピーチ中に実行し、AAF の有無による発話の話速、ラウドネスレベル、フィラー率、話者の主観評価、聴衆の印象評価を比較調査した。調査の結果、AAF を利用することで話者の話速が低下し、聴衆が話速の低下を好意的に評価することが確認された。一方で、ラウドネスレベル、フィラー率、聴衆視点のスピーチの総合評価に差は確認されなかった。また、スピーチ中の AAF 利用が、話者の孤独感に由来する緊張状態に影響する可能性が示唆された。

1. はじめに

人前でスピーチを行う際に、緊張する話者は多い [9], [21], [27]。話者の緊張時には話速や声量、声の震えといった話し方や声質に変化がみられることが多く [24], [33]、これらの変化は聴衆に対して負の印象を与える [12]。

緊張に伴う声質の変化に対するアプローチとして、専用デバイスの追加 [18] や環境の変更 [19] により緊張を緩和する手法が提案されている。しかし、デバイスの使用や環境の変更が困難な状況が存在する。別のアプローチとして、近年では機械学習により本人の声質を変化させるボイスチェンジャも開発されているが、事前にデータ取得と学習を行う等の準備が必要となる。また、話者が発話の変化を自覚して、声量や話速を調整することは重要である [11]。しかし、スピーチ中に話者が発話を修正することは困難である [1], [20]。本研究では、特殊なデバイスの使用や、事前学習等の準備を必要とせず、話者自身による無意識的な声質および話し方の変化を誘発する手法として、変換聴覚フィードバック（Altered auditory feedback, AAF）に着目する。

AAF とは、マイクを通じて取得した話者の発話音声に音響的な変換処理を施し、変換した音声を話者のみに対してリアルタイムで聞かせる手法である。AAF はその構造上、マイク、ヘッドフォン、およびコンピュータのみで実装可能である。また、AAF を用いることで、話者の感情および声質が変化する [2], [17]。特に、一人での発話時に音量を増加した AAF を利用することで、話者の声量が下がることが確認されている [14]。著者らの先行研究 [23]において、発話のピッチを上昇、または音量を増加させた音声を話者にフィードバックすることで、話者の声のかすれや声荒れといった発話の嗄声度が増加するほか、話者の話速が低下することが確認された。この知見から、ピッチ上昇および音量増加を行った AAF をスピーチ中に利用することで、聴衆に負の印象を与える話し方の改善が期待できると考えた。しかし、先行研究で得られた知見は、AAF を個人で利用した際の発話変化にとどまっており、発話の評価を行う聴衆は不在だった。そのため、スピーチの聴衆に与える影響については調査されていない。

本研究では、AAF をスピーチ中に利用することで、話者および聴衆に発生する効果に注目する。ピッチ上昇および音量増加を行った AAF をスピーチ中に利用することで、発話の話速および声量が変化するか、発話の変化が聴衆の評価を改善するかを検証する。ヘッドフォンおよびマイクを使用するオンラインスピーチ環境を構築し、AAF の有無によるスピーチ中の発話の差およびスピーチ全体の印象評価の差を比較することで、スピーチ中に AAF を利用す

¹ 筑波大学大学院 理工情報生命学術院 システム情報工学研究群 情報理工学位プログラム

Graduate School of Science and Technology, Degree Programs in Systems and Information Engineering, 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

ることの影響を調査する。

本論文ではまず、関連研究として緊張時の発話変化、既存の発話操作手法およびAAFが話者に与える影響に関する研究を紹介し、本研究の指針を述べる。次に、実装したシステムの詳細、およびシステムを用いた調査のための実験について述べる。最後に、実験の結果を基に議論および考察を行い、AAFを利用したスピーチ中の発話改善について本研究の結論を述べる。

2. 関連研究

本章では、まず緊張時の発話変化および既存の発話操作手法について説明し、本研究においてAAFに着目した理由を示す。その後、AAFに関する既存研究に対する本研究の位置づけを示すため、AAFが話者に与える影響について説明し、本研究との差分を示す。

2.1 緊張時の発話変化

緊張時の発話変化としてGallegoら[10]は、話者の緊張度が高いほどスピーチ時間が短くなる傾向にあることを示した。このほか、話者が不安や緊張を感じた際に話速が増加することが示されている[28], [33]。Jeffreyら[34]は、話者に作業負荷が発生すると声量が増加することを報告した。Demenkoら[8]は人間の発話を分析し、不安時にピッチが高くなることを報告した。また、岩田らの研究[12]により、話速の大きいスピーチを聞いた聴衆は、話者の不安が大きいと評価することが明らかになった。そのため本研究では、緊張時の発話変化および聴衆の印象に影響を与える発話変化として、話速および声量に注目する。

2.2 発話操作に関する既存手法

発話を変化させる手法として、音声変換（Voice Conversion, VC）が存在する。VCとは発話の言語情報を維持しつつ非言語情報のみを変換する技術である[32]。しかし、VCは音声変換の設定を機械学習等で決定するため、VCの利用には話者の音声のデータセット作成が必要である。したがって十分な準備時間が確保できない場合において、VCは有効ではない。一方、AAFは1章にある通り話者の発話をフィードバックする手法であるため音声のデータセットは不要である。

また、別の発話操作手法として、遅延聴覚フィードバックが挙げられる。発話から50 msから105 ms程度の時間をおいた発話音声のフィードバックにより、吃音の改善が報告されている[7], [29]。しかし、この発話の明瞭化は吃音者にのみ有効であり、非吃音者に対しては発話の流暢性の低下や不快感の増加が発生すると報告されている[23], [30]。これに関連して、栗原ら[13]は、1000 msを超える遅延を設定した聴覚フィードバックを利用することで話者の発話が阻害されることを報告した。遅延による発話阻害効果は、

遅延時間が30 msを超える場合に確認されている[3], [15]。一方、AAFについては、2.3節に示すように話者の属性に関わらず、発話を阻害しない範囲で共通の効果を発揮する。また、本研究で用いるAAFの遅延時間は3.1節に記載する通り23 ms程度であるため、遅延時間の観点からAAFおよび遅延聴覚フィードバックは別手法であるといえる。そのため本研究では、事前準備を必要とせず、かつ話者に共通して発話変化を誘発するAAFを発話の操作手法として採用する。

2.3 AAFが話者に与える影響

AAFは話者の発話を変化させる効果を持つ。Burnettら[4]は、AAFにおいてフィードバック音声のピッチが下がると発話のピッチが上がり、フィードバック音声のピッチが上がると発話のピッチが下がることを示した。Laneら[14]は、AAFにおいてフィードバック音声の音量が上がると発話の声量が下がることを示した。Coughlerら[6]はAAFを使用した際の基本周波数、フォルマント、および摩擦音の周波数重心への影響を調査した。その結果、4歳以上の子供は、フィードバック音声のパラメータ変換と逆向きの声質の変化が見られること、および大人よりもAAFの効果が小さいことを報告した。このほか、著者らの先行研究[23]において、フィードバック音声のピッチ上昇および音量増加を行うことにより、話者の発話が明瞭になるほか、話速が低下することが示された。しかしこれらの研究では話者の発話変化にのみ注目しており、対話相手への影響については調査されていない。

また、AAFの別の効果として、話者の不安軽減が存在する。Jeanら[5]はピッチ下降、フォルマント下降、およびローシェルフフィルタを利用したAAFにより、対立中の話者の不安が軽減されることを発見した。この研究では不安に関する生理学的分析および主観的評価を行っている一方で、発話の変化や不安以外の印象評価については触れられていない。成瀬ら[22]は、ピッチ下降およびハイシェルフフィルタを利用したAAFにより、対面でのスピーチにおいて話者の緊張感が低下、あるいは増大することを報告した。この研究においても話者の不安感について生理学的分析および主観的評価を行っているが、聴衆による話者およびスピーチの評価は行われていない。

したがって本研究では、スピーチ中のAAF利用が話者および聴衆に与える影響について調査する。具体的には、ピッチ上昇および音量増加を行った音声をスピーチ中の話者にフィードバックして話速の低下および声量の減少を誘発することにより、聴衆の話者に対する印象評価への影響を調査する。また、評価指標を発話音声の分析のみ、話者の主観評価のみ、聴衆の印象評価のみに絞らず、3つすべてを用いることで、スピーチ中のAAF利用がスピーチに与える影響を探索的に調査する。

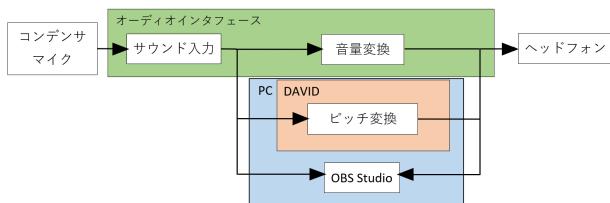


図 1 システム構成図

3. 実験

本研究では、AAF を利用したオンラインスピーチが話者および聴衆に与える影響を調査するための実験を行った。実験は 2 段階に分かれ、話者側の影響を確認するスピーチタスクおよび聴衆側の影響を確認する映像視聴タスクを順に実施した。本章では、実験に用いたシステムの説明、実験条件、実験参加者、実験タスク、実験環境、および評価指標を述べる。

3.1 実験用システム

本実験では、著者らの先行研究 [23] にて実装した AAF 提示システムを用いた。コンデンサマイク (audio-technica AT2050) から入力した音声を、オーディオインターフェース (NATIVE INSTRUMENTS KOMPLETE Audio 6) を経由してオープンソースソフトウェアプラットフォームである DAVID[26] に与える。オーディオインターフェースおよび DAVID にて音声変換操作を行った後に、ヘッドフォン (SONY WH-1000XM4) を用いて参加者に変換音声を提示する。システムの構成を図 1 に示す。なお、マイクに音声を入力してから参加者に変換音声が提示されるまでの時間は、著者の環境 (3.5 節に記載) にて 23 ms 程度であった。

3.2 実験条件

実験条件は以下の 2 条件で、すべて参加者内配置で設計した。

- AAF あり条件
- AAF なし条件

AAF あり条件ではピッチおよび音量を増加させた音声を話者にフィードバックする。ピッチは、DAVID の PITCH パラメータを操作して、話者の肉声から 100 cent 上昇するように設定した。音量は、オーディオインターフェースのボリュームを操作して、参加者ごとに設定した音量の基準から音量が 50% 程度増加するように設定した。これらの変換設定は、ピッチの影響を調査した先行研究 [4] および著者らの先行研究に基づき、話速の低下および声量の増加を狙って設定された。なお、AAF なし条件では参加者にフィードバック音声を提示しない。

3.3 実験参加者

実験参加者は計 10 名（男性 9 名、女性 1 名、平均 22.5 歳、標準偏差 1.05 歳）の大学生および大学院生である。実験は参加者内配置で行われ、各参加者が 2 条件すべてを実施した。

3.4 実験タスク

実験は 2 段階に分けて実施された。1 段階目は話者への影響を調査するため、オンライン環境でのスピーチを行った。2 段階目は聴衆への影響を調査するため、1 段階目のスピーチ映像視聴を行った。本節では、話者側および聴衆側に分けてタスク内容を説明する。

3.4.1 話者側スピーチタスク

スピーチ中の AAF 利用が話者に与える影響を調査するため、オンライン環境でのスピーチタスクを実施した。最初に、参加者に対して実験で使用するシステムの音量調整を行った。音量調整は、ヘッドフォンから再生される音声の大きさおよび参加者の感じる骨導音の大きさが、参加者の主観で同じ大きさになるよう指示した。

その後、すべての条件に先駆けて音声の変換およびデータ計測を行わないトレーニングを実施した。トレーニングは参加者の「自分の声を聞きながら発話する」経験の有無による実験結果への影響を排除する目的で行われ、実験者側で用意した原稿を、参加者の主観で読み慣れたと感じるまで繰り返し音読するよう指示した。

各条件にて参加者は実験用 PC を用いて zoom に接続した後に、参加者自身が事前に用意した原稿を用いて 2 分間のスピーチを行った。スピーチ原稿は「好きな本・曲・アーティスト」「ストレス解消法」の 2 種類であり、実験実施前日に原稿作成期間を設けた。原稿のトピックは、内容が平易かつ等質で、スピーチに感情が乗りやすいものを選定した。参加者は 1 条件につき 1 つの原稿を用いてスピーチを行った。スピーチ中の原稿持ち込みは許可したが、聴衆とのアイコンタクトを可能な限り維持するよう参加者に指示した。また、後日実施する聴衆側映像視聴タスクのため、参加者スピーチ中の zoom 画面を録画した。なお、条件の実施順序および条件とトピックの対応は参加者毎にランダム化した。

話者側スピーチタスクにおいて、参加者のスケジュール調整が困難であったため、仮想的な聴衆として実験協力者 2 名を zoom に配置した。実験協力者にはスピーチ中に適度な視線移動および相槌を取ること、スピーチの内容や参加者によって反応に差を付けないよう指示した。なお、実験協力者は全ての参加者を通じて同一である。

各条件にて原稿の音読およびタスクに対するアンケートを行い、すべての条件が終了した後に実験全体に関する半構造化インタビューを行った。アンケートは Tanveer らの研究 [31] を参考に、「まったく当てはまらない」から「と

ても当てはまる」の5段階リッカート尺度を用いた。アンケートの内容は以下の通りである。

- スピーチの出来に満足している
- 適切に声の大きさを調整できた
- 話速は適切だった
- 適切な身振り手振りを示せた
- 聴衆を意識して視線を配った
- リラックスしてスピーチに取り組めた
- フィードバック音声はスピーチの役に立った（AAFあり条件のみ）
- フィードバック音声の存在でスピーチに集中できなかった（AAFあり条件のみ）
- フィードバック音声の有用性は集中阻害の影響を上回った（AAFあり条件のみ）
- 実際のスピーチでもフィードバック音声を利用したい（AAFあり条件のみ）

また、スピーチ音声に含まれる以下の3つの特徴量を抽出して、タスクの評価指標とした。

- 話速
- ラウドネスレベル
- フィラー率

このうち話速は1秒間の平均発音文字数である。また、ラウドネスレベルは発話の音量に関する指標である。フィラー率はスピーチ全体の発音文字数に占めるフィラーの割合であり、割合が低いと聴衆が好意的にスピーチを評価すると考えられる。これらの指標は、特徴量抽出に用いたPythonのライブラリであるSurfboard[16]において抽出可能な特徴量、および著者らの先行研究から選出された。

3.4.2 聴衆側映像視聴タスク

すべての参加者が話者側スピーチタスクを終えた後に、スピーチ中のAAF利用が聴衆に与える影響を調査するため、録画したスピーチ映像を視聴してアンケートに回答するタスクを行った。参加者のスケジュール調整を柔軟に行うため、タスクはオンライン上で非同期的に行われた。参加者は自分以外のすべてのスピーチ映像18本（=9人×2条件）を視聴して、それぞれの映像に対してアンケートによる評価を行った。すべての評価について、原稿の内容は評価対象に含めないことを参加者に指示した。なお、映像の視聴順については、同一参加者の映像を連続して視聴するように指示した。

アンケートは総合評価および印象評価に分かれ、総合評価は同一参加者のスピーチ2種類のどちらが好ましいかを選択させた。印象評価はTanveerらの研究[31]を参考に、「まったく当てはまらない」から「とても当てはまる」の5段階リッカート尺度を用いた。印象評価アンケートの内容は以下の通りである。

- 話者はスピーチ中に間を置かなかった
- 話者は聞き取りやすい話速でスピーチをした



図2 スピーチ中の様子。個人情報保護のため顔を修正。役割は著者により追記。

- 話者は聞き取りやすい声量でスピーチをした
- 話者はフィラー（えー、あの、など）を多用しなかった
- 話者は聴衆とアイコンタクトを保っていた
- 話者は自然な身振り手振りを行った
- 話者はスピーチに集中していた（気が散っているようには見えなかった）

3.5 実験環境

話者側スピーチタスクに関して、オーディオインターフェース（NATIVE INSTRUMENTS KOMPLETE Audio 6）を接続したラップトップコンピュータ（CPU:Intel Core i7-9750H, GPU:GeForce GTX 1650, メモリ：16GB）を用いて実験を行った。オーディオインターフェースにはコンデンサマイク（audio-technica AT2050）およびヘッドフォン（SONY WH-1000XM4）を接続した。参加者はヘッドフォンを装着し、コンデンサマイクに向かって声を出した。また、ラップトップコンピュータにはzoomの画面を映し、ラップトップコンピュータ内蔵のカメラを用いて参加者の映像をzoomに投影した。スピーチ中の様子を図2に示す。聴衆側映像視聴タスクについては、3.4.2項にある通り、オンライン上で実験を行った。

3.6 評価指標

評価指標として、3.4.1項および3.4.2項に記載した、アンケート結果およびスピーチ音声から抽出した特徴量を用いる。

4. 結果

本章では、実験の各評価指標についての結果を示す。3.6節に示した特徴量について、AAFあり条件のスピーチおよびAAFなし条件のスピーチから特徴量を抽出し、条件間に比較および分析を行った。特徴量抽出後AAFあり条件のスピーチおよびAAFなし条件のスピーチのそれに対して、F検定による等分散性およびシャピロウィルク検定による正規性を確認した。その後、正規性が認められなかったためウィルコクソンの符号順位検定を行った。

話者側スピーチタスクのアンケートにて、2条件に共通して回答のある項目については、アンケート項目ごとに条件間でウィルコクソンの符号順位検定を行った。聴衆側映像視聴タスクのアンケートにて、印象評価については、アンケート項目ごとに条件を固定因子、話者および評価者を変量因子とした混合効果モデルを作成して分析を行った。また、総合評価については、二項検定を行った。なお、話者1名分の録画映像ファイルが破損していたため、該当する話者への聴衆側評価の結果（印象評価18件、総合評価9件）を分析から除外した。

すべての検定において p 値が0.05未満であることを統計的に有意とみなした。また、統計分析にはR[25]を用いた。

4.1 特徴量分析

AAFありのスピーチおよびAAFなしのスピーチの話速、ラウドネスレベル、フィラー率をそれぞれ条件毎に箱ひげ図としてまとめた結果を図3に示す。ウィルコクソンの符号順位検定を行った結果、AAFあり条件の話速が有意に減少した（ $Z = 2.06, p = 0.039, d = 0.65$ ）。ラウドネスレベル（ $Z = 1.52, p = 0.13$ ）、フィラー率（ $Z = 0.11, p = 0.91$ ）に有意差はみられなかった。

4.2 話者側アンケート

3.4.1項に示したアンケートについて、2条件に共通する項目の結果を条件毎に箱ひげ図としてまとめたものを図4に示す。2条件共通項目についてウィルコクソンの符号順位検定を行った結果、AAFあり条件の身振り手振りが有意に增加了（ $Z = 1.98, p = 0.048, d = 0.63$ ）。スピーチの満足度（ $Z = 1.17, p = 0.24$ ）、リラックス度（ $Z = 1.90, p = 0.058$ ）、視線配分（ $Z = 1.72, p = 0.086$ ）、話速（ $Z = 1.43, p = 0.15$ ）、声量（ $Z = 1.90, p = 0.058$ ）に有意な主観評価の差はみられなかった。また、AAFあり条件のみの項目の結果を箱ひげ図としてまとめたものを図5に示す。「フィードバック音声はスピーチの役に立った」「フィードバック音声の存在でスピーチに集中できなかった」「フィードバック音声の有用性は集中阻害の影響を上回った」「実際のスピーチでもフィードバック音声を利用したい」の平均スコアはそれぞれ4.2、1.6、3.8、4.2であった。

4.3 聴衆側アンケート

3.4.2項に示したアンケートについて、印象評価の結果を条件毎に箱ひげ図としてまとめたものを図6に示す。混合効果モデルを用いて分析した結果、AAFあり条件の話速の聞き取りやすさが有意に向上した（ $Z = 2.40, p = 0.017, d = 0.75$ ）。間の置き方（ $Z = 0.45, p = 0.66$ ）、声量の聞き取りやすさ（ $Z = 0.15, p = 0.88$ ）、フィラー量（ $Z = 0.78, p = 0.44$ ）、アイコンタクト（ $Z = 1.02, p = 0.31$ ）、身振り手振り

（ $Z = 1.81, p = 0.071$ ）、話者の集中度（ $Z = 0.98, p = 0.33$ ）に有意な印象差はみられなかった。また、総合評価の結果を棒グラフとしてまとめたものを図7に示す。二項検定を行った結果、有意差はみられなかった（ $Z = 0.89, p = 0.37$ ）。

5. 議論および考察

本研究ではスピーチ中にピッチ上昇および音量増加を行ったAAFを利用してすることで、話速の低下および声量の減少を誘発すること、および聴衆の話者に対する印象評価を改善することを期待した。話速に注目すると、AAFあり条件にて有意に話速が減少しているほか、話速の聞き取りやすさについての評価スコアが有意に大きい。これはAAFを利用することにより話者の話速が低下し、かつ低下した話速が聴衆にとって好意的に捉えられたことを意味する。一方で話者側の主観評価に注目すると、話速に関して有意差はみられていない。そのため話速の観点からは、AAFを利用することにより話者の話速を低下させ、聞き取りやすい話速のスピーチを実現できると考えられる。また、実験を通して発話阻害は確認されておらず、話速調整によるスピーチ支援システムとして有用であるといえる。

声量やフィラー率については、すべての評価指標において有意差がみられなかった。聴衆側の総合評価結果についてもAAFの有無で有意差はみられておらず、AAFを利用することで聴衆の評価を改善できると断定できない。Laneらの先行研究[14]では、フィードバック音声の音量を増加すると話者の声量が減少することが確認されていた。本実験のAAFでは音量と同時にピッチを操作したため、複数の音声変換を同時に使うと、単体の音声変換を行ったAAFとは別の効果が表れる可能性がある。一方で、話者側スピーチタスクの半構造化インタビューから「AAFありの方がフィラーや言い淀みが少なくなった気がする」(P3)、「AAFありの方が噛む回数が少なかった気がする」(P4)といった意見が得られている。物理量や聴衆の印象に変化を与えるしないものの、AAFを利用することにより話者の「主観的な話しやすさ」に何らかの変化が発生した可能性がある。

話者の主観に注目すると、話者側スピーチタスクの半構造化インタビューから「AAFありだとヘッドフォンをしていないかのような自然さで話せた」(P5)、「AAFがあった方が自信を持って話せた」(P7)といった意見が得られている。また、AAFあり条件のみのアンケート項目に注目しても、AAFが話者にとって概ね好意的に捉えられると考えられる。このほか、話者側アンケートの「リラックスしてスピーチに取り組めた」がAAFを利用することで増加傾向にあることを踏まえると、AAFを利用することで話者の緊張状態に影響を与える可能性が考えられる。Jeanらの研究[5]ではフィードバック音声のピッチを下降させ、高音域を抑制することにより話者の不安が減少する

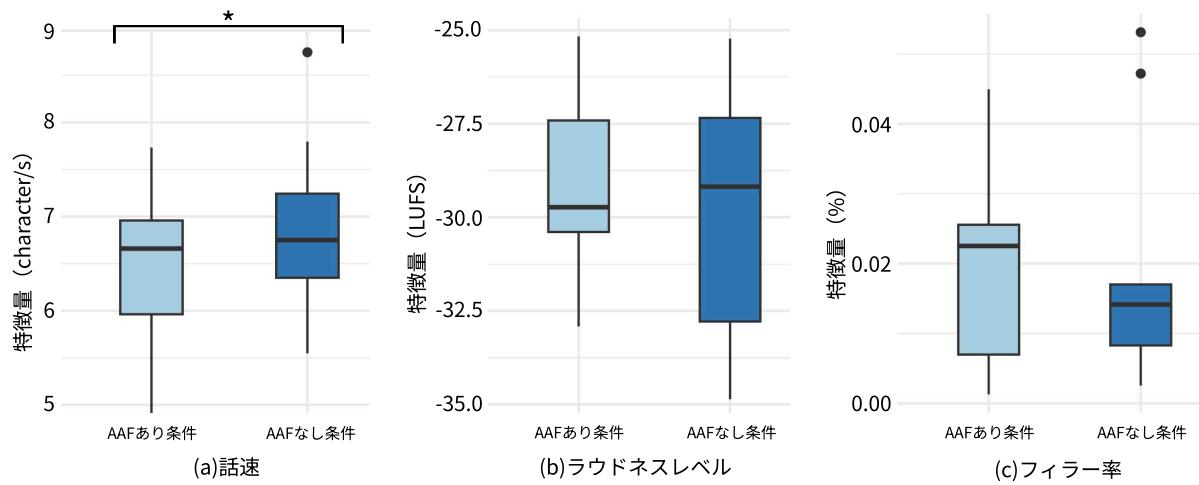


図 3 各条件でのスピーチに含まれる特徴量 (a) 話速 (b) ラウドネスレベル (c) フィラー率。
有意差の出た項目に*を示す。

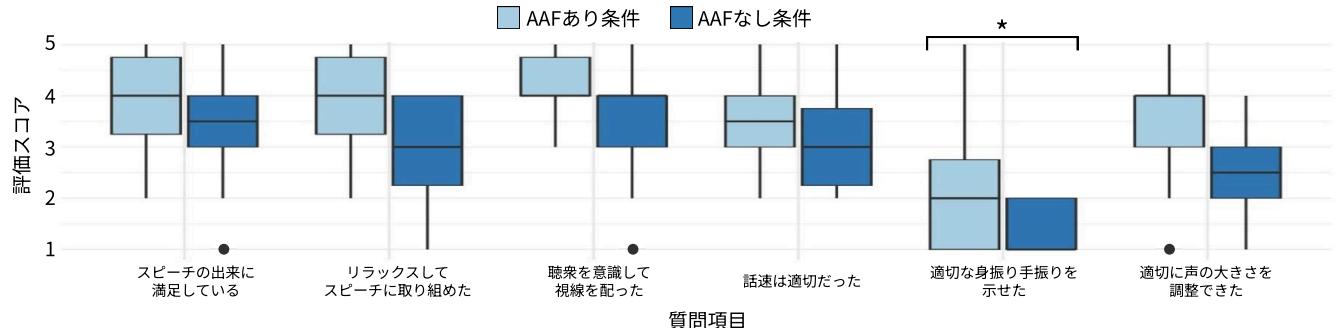


図 4 話者側アンケート (2 条件共通) の結果。有意差の出た項目に*を示す。

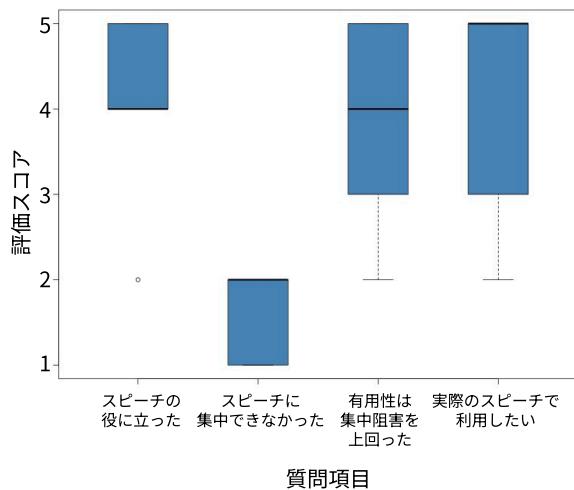


図 5 話者側アンケート (AAF あり条件のみ) の結果。

ことを報告している。本実験にて行った音声変換はフィードバック音声のピッチおよび音量を増加させたものであるが、上述した主観的な話しやすさへの影響および半構造化インタビューから得られた「AAF ありは周りに人がいるような感覚がする」(P4) という意見から、何らかの変換操作を行った AAF を利用することで、話者の孤独感による緊張状態に影響を与える可能性があると考えられる。し

かし、検証のためには別の調査が求められる。

6. 制約および今後の課題

本章では、本研究の制約および今後の課題を述べる。

6.1 原稿の内容および印象評価

3.4.2 項にて説明した通り、本研究にて行った実験ではスピーチの印象評価に原稿の内容を含めないよう指示した。しかし、聴衆側アンケートの自由記述項目にて「どうしても原稿の内容を意識してしまう部分はある」(P10) という意見を得た。本実験での原稿トピックは、聴者の知識や専門性への影響が小さいものを著者の主觀で選定したが、今後も原稿の内容が話し方の印象評価に影響しないようなトピックの選定および分析が求められる。

6.2 スピーチ聴講環境の再現

本研究にて行った実験では、聴衆のいるスピーチ環境を再現するため実験協力者を配置し、参加者は後日スピーチの録画を視聴する形式を採用した。そのため、参加者は他の参加者のスピーチをリアルタイムで聴講していない。リアルタイムなスピーチの聴講およびその場での印象評価を行うことで、より実際のスピーチ環境を再現可能である。

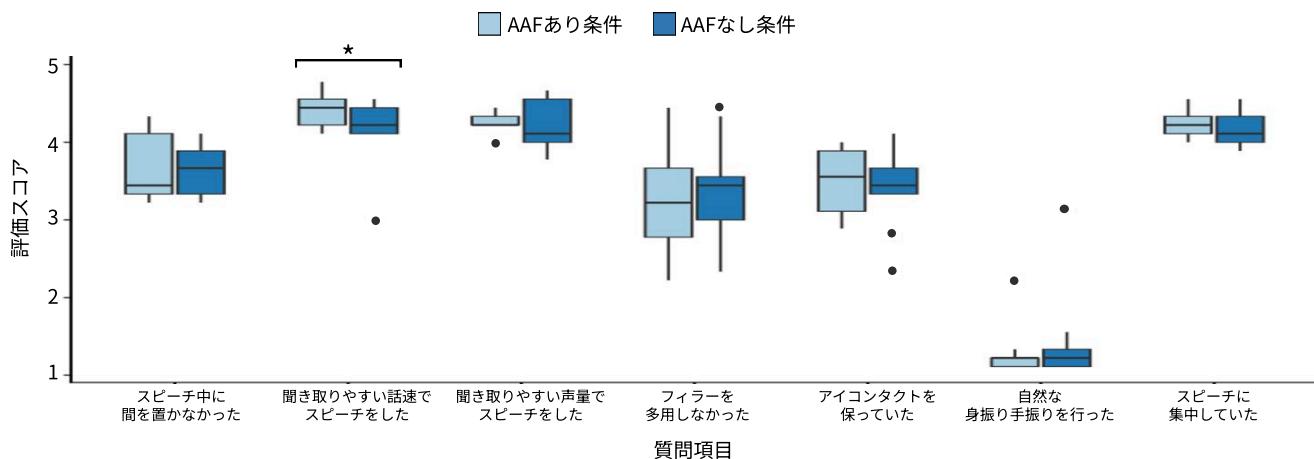


図 6 聴衆側アンケート（印象評価）の結果。有意差の出た項目に*を示す。

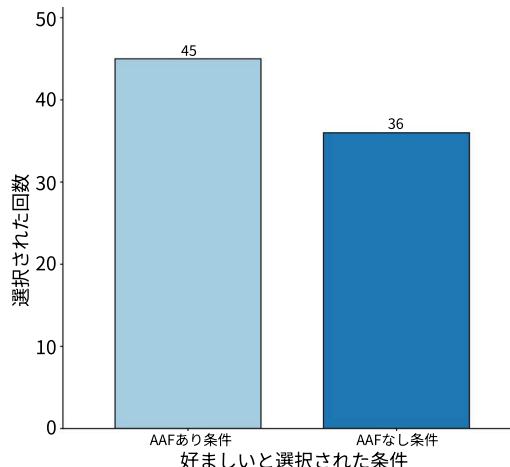


図 7 聴衆側アンケート（総合評価）の結果。

したがって、聴衆側の実験設計を再構成する必要がある。

6.3 発話変化の個人差

本研究にて行った実験では、参加者のスピーチ経験や参加者個人の緊張時の発話変化に注目していない。そのため、緊張時に話速や声量が変化しない参加者が含まれていた可能性がある。参加者個人の緊張時の発話変化を考慮することで、より正確な分析が可能になると考えられる。

7. おわりに

本研究では、スピーチ中にAAFを利用することによる発話改善、聴衆の印象改善、およびスピーチに与える影響について調査を行った。AAFの変換パラメータとして発話のピッチ上昇および音量増加を設定し、AAFの有無によるスピーチ中の話速、ラウドネスレベル、フィラー率、スピーチに対する話者の主観評価、スピーチに対する聴衆の印象評価を比較した。

調査の結果、AAFを利用することで話者の話速が低下、かつ聴衆が話速の低下を認知して好意的に評価した。そのため、ピッチ上昇および音量増加AAFをスピーチ中に利

用することで、話者の発話を阻害せずに聞き取りやすい話速でのスピーチを実現できると判断した。一方でラウドネスレベル、フィラー率、聴衆視点のスピーチの総合的な好ましさに差はみられなかった。また、話者の主観評価および聴衆の印象評価結果を総合することで、AAFの利用が話者の孤独感に由来する緊張状態に変化を与える可能性が示唆された。

今後は参加者の緊張時の発話変化の個人差や、より実際の状況に近いスピーチ聴講環境を考慮した実験設計の修正を行う。また、本研究により得られた、話者の孤独感による緊張状態への影響について調査を行う。

参考文献

- [1] Arnfield, S., Roach, P., Setter, J., Greasley, P. and Horton, D.: Emotional stress and speech tempo variation, *ESCA/NATO Workshop on Speech under Stress*, pp. 13–15 (1995).
- [2] Aucouturier, J.-J., Johansson, P., Hall, L. and Watanabe, K.: Covert digital manipulation of vocal emotion alter speakers' emotional states in a congruent direction, *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 113, No. 4, pp. 948–953 (online), DOI: 10.1073/pnas.1506552113 (2016).
- [3] Black, J. W.: The Effect Of Delayed Side-Tone Upon Vocal Rate And Intensity, *Journal of Speech and Hearing Disorders*, Vol. 16, No. 1, pp. 56–60 (online), DOI: 10.1044/jshd.1601.56 (1951).
- [4] Burnett, T. A., Senner, J. E. and Larson, C. R.: Voice F0 responses to pitch-shifted auditory feedback: a preliminary study, *Journal of Voice*, Vol. 11, No. 2, pp. 202–211 (online), DOI: 10.1016/S0892-1997(97)80079-3 (1997).
- [5] Costa, J., Jung, M. F., Czerwinski, M., Guimbretière, F., Le, T. and Choudhury, T.: Regulating Feelings During Interpersonal Conflicts by Changing Voice Self-Perception, *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, New York, NY, USA, Association for Computing Machinery, pp. 1–13 (online), DOI: 10.1145/3173574.3174205 (2018).
- [6] Coughler, C., de Launay, K. Q., Purcell, D. W., Cardy, J. O. and Beal, D. S.: Pediatric Responses

- to Fundamental and Formant Frequency Altered Auditory Feedback: A Scoping Review, *Frontiers in Human Neuroscience*, Vol. 16 (online), DOI: 10.3389/fnhum.2022.858863 (2022).
- [7] Daliri, A. and Max, L.: Stuttering adults' lack of pre-speech auditory modulation normalizes when speaking with delayed auditory feedback, *Cortex*, Vol. 99, pp. 55–68 (online), DOI: 10.1016/j.cortex.2017.10.019 (2018).
- [8] Demenko, G. and Jastrzebska, M.: Analysis of natural speech under stress, *Acta Physica Polonica A*, Vol. 121, No. 1A (2012).
- [9] Dwyer, K. K. and Davidson, M. M.: Is Public Speaking Really More Feared Than Death?, *Communication Research Reports*, Vol. 29, No. 2, pp. 99–107 (online), DOI: 10.1080/08824096.2012.667772 (2012).
- [10] Gallego, A., McHugh, L., Penttonen, M. and Lapalainen, R.: Measuring public speaking anxiety: self-report, behavioral, and physiological, *Behavior Modification*, Vol. 46, No. 4, pp. 782–798 (2022).
- [11] Hoogterp, B.: *Your Perfect Presentation: Speak in Front of Any Audience Anytime Anywhere and Never Be Nervous Again*, McGraw-Hill Education (2014).
- [12] 岩田彩香, 川井智理, 斎藤順一, 嶋 大樹, 熊野宏昭: 社交不安傾向によるスピーチ場面でのパフォーマンス低下に関する検討, 早稲田大学臨床心理学研究, Vol. 15, No. 1, pp. 53–63 (2015).
- [13] Kurihara, K. and Tsukada, K.: SpeechJammer: A System Utilizing Artificial Speech Disturbance with Delayed Auditory Feedback (2012).
- [14] Lane, H. L., Catania, A. C. and Stevens, S. S.: Voice Level: Autophonic Scale, Perceived Loudness, and Effects of Sidetone, *The Journal of the Acoustical Society of America*, Vol. 33, No. 2, pp. 160–167 (online), DOI: 10.1121/1.1908608 (1961).
- [15] Lee, B. S.: Artificial Stutter, *Journal of Speech and Hearing Disorders*, Vol. 16, No. 1, pp. 53–55 (online), DOI: 10.1044/jshd.1601.53 (1951).
- [16] Lenain, R., Weston, J., Shivkumar, A. and Fristed, E.: Surfboard: Audio Feature Extraction for Modern Machine Learning (2020).
- [17] Lincoln, M., Packman, A. and Onslow, M.: Altered auditory feedback and the treatment of stuttering: a review, *J Fluency Disord*, Vol. 31, No. 2, pp. 71–89 (online), DOI: 10.1016/j.jfludis.2006.04.001 (2006).
- [18] MacLean, D., Roseway, A. and Czerwinski, M.: Mood-Wings: a wearable biofeedback device for real-time stress intervention, In *Proceedings of the 6th International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '13)*, Vol. 66, pp. 1–8 (online), DOI: 10.1145/2504335.2504406 (2013).
- [19] Matthews, M., Snyder, J., Reynolds, L., Chien, J. T., Shih, A., Lee, J. W. and Gay, G.: Real-Time Representation Versus Response Elicitation in Biosensor Data, *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, New York, NY, USA, Association for Computing Machinery, p. 605–608 (online), DOI: 10.1145/2702123.2702485 (2015).
- [20] Mehrabian, A. and Williams, M.: Nonverbal concomitants of perceived and intended persuasiveness., *Journal of personality and social psychology*, Vol. 13(1), pp. 37–58 (online), DOI: 10.1037/h0027993 (1969).
- [21] Merz, C. J., Hagedorn, B. and Wolf, O. T.: An oral presentation causes stress and memory impairments, *Psychoneuroendocrinology*, Vol. 104, pp. 1–6 (online), DOI: 10.1016/j.psyneuen.2019.02.010 (2019).
- [22] 成瀬加菜, 吉田成朗, 世田圭佑, 鳴海拓志, 谷川智洋, 廣瀬通孝: リアルタイムな変換聴覚フィードバックによる緊張緩和効果の基礎的検討, ヒューマンインタフェース学会研究報告集, Vol. 20, pp. 105–112 (オンライン), 入手先 <https://cir.nii.ac.jp/crid/1520572358880607488> (2018).
- [23] 大沼怜生, 市川あゆみ, 川口一画: ピッチとラウドネスを増減させた変換聴覚フィードバックおよび遅延聴覚フィードバックの発話明瞭度への影響調査, 情報処理学会研究報告, Vol. 2025-HCI-211, pp. 1–8 (2025).
- [24] Puyvelde, M. V., Neyt, X., McGlone, F. and Pattyn, N.: Voice Stress Analysis: A New Framework for Voice and Effort in Human Performance, *Front Psychol*, Vol. 9 (online), DOI: 10.3389/fpsyg.2018.01994 (2018).
- [25] R Core Team: *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria (2018).
- [26] Rachman, L., Liuni, M., Arias, P., Lind, A., Johansson, P., Hall, L., Richardson, D., Watanabe, K., Dubal, S. and Aucourturier, J.-J.: DAVID: An open-source platform for real-time transformation of infra-segmental emotional cues in running speech, *Behavior Research Methods*, Vol. 50, pp. 323–343 (online), DOI: 10.3758/s13428-017-0873-y (2018).
- [27] Sawyer, C. R.: *16. Communication Apprehension and Public Speaking Instruction*, pp. 397–426 (online), DOI: doi:10.1515/9781501502446-017, De Gruyter Mouton (2016).
- [28] Scherer, K. R., Wranik, T., Sangsue, J., Tran, V. and Scherer, U.: Emotions in everyday life: Probability of occurrence, risk factors, appraisal and reaction patterns, *Social Science Information*, Vol. 43, No. 4, pp. 499–570 (2004).
- [29] Sparks, G., Grant, D. E., Millay, K., Walker-Batson, D. and Hynan, L. S.: The effect of fast speech rate on stuttering frequency during delayed auditory feedback, *Journal of Fluency Disorders*, Vol. 27, No. 3, pp. 187–201 (online), DOI: 10.1016/S0094-730X(02)00128-6 (2002).
- [30] Stuart, A., Kalinowski, J. and abd Kerry Lynch, M. P. R.: Effect of delayed auditory feedback on normal speakers at two speech rates, *The Journal of the Acoustical Society of America*, Vol. 111, No. 5, pp. 2237–2241 (online), DOI: 10.1121/1.1466868 (2002).
- [31] Tanveer, M. I., Lin, E. and Hoque, M. E.: Rhema: A Real-Time In-Situ Intelligent Interface to Help People with Public Speaking, *Proceedings of the 20th International Conference on Intelligent User Interfaces*, IUI '15, New York, NY, USA, Association for Computing Machinery, p. 286–295 (online), DOI: 10.1145/2678025.2701386 (2015).
- [32] Toda, T., Black, A. W. and Tokuda, K.: Voice Conversion Based on Maximum-Likelihood Estimation of Spectral Parameter Trajectory, *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 15, No. 8, pp. 2222–2235 (online), DOI: 10.1109/TASL.2007.907344 (2007).
- [33] Wang, Q., Xu, F., Wang, X., Wu, S., Ren, L. and Liu, X.: How Anxiety State Influences Speech Parameters: A Network Analysis Study from a Real Stressed Scenario, *Brain Sciences*, Vol. 15, No. 3, p. 262 (2025).
- [34] Whitmore, J. and Fisher, S.: Speech during sustained operations, *Speech Communication*, Vol. 20, No. 1, pp. 55–70 (online), DOI: [https://doi.org/10.1016/S0167-6393\(96\)00044-1](https://doi.org/10.1016/S0167-6393(96)00044-1) (1996). Speech under Stress.