

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

成瀬 加菜^{1,a)} 吉田 成朗¹ 世田 圭佑¹ 鳴海 拓志¹ 谷川 智洋¹ 廣瀬 通孝¹

概要: 口頭発表において緊張感が意思伝達の円滑性を損なう問題を解決するため、変換聴覚フィードバックを利用した緊張緩和システムを提案する。このシステムではマイクへの入力音声に変換処理を施すことで緊張状態での特性を打ち消した音声を生成し、発話者にフィードバックすることで自身が安静状態にあると知覚させることを図った。対面でのスピーチ実験による効果検証の結果、音声変換パラメータが緊張感に変化をもたらすことは示されなかったが、音声入出力間の遅延による影響をはじめ、フィードバック音声が発話者の心理に影響を与えた可能性が示唆された。

1. 研究の背景

人間が充実した生活を送るためには他者とのコミュニケーションが不可欠である。良好な人間関係の構築には会話が重要であるほか、面接や学会発表などの口頭発表場面では伝えたい情報を的確に伝えることが自身の活躍を促す。しかし、他者の前で話す場面は緊張感を喚起しやすいことが知られ [1]、それが意思伝達の円滑性を損なう要因になりうる。また、緊張感は発話者のスピーチパフォーマンスを低下させ、聞き手が発話者に対して抱く印象を悪化させることが知られている [2]。そこで本研究では発話活動における緊張感を緩和するためのシステムの構築を目指す。

心理学における緊張感とは、これから起きる物事に対して待ち受けている心の状態を指す。本研究で扱う口頭発表場面のほか、スポーツ競技や演奏場面なども緊張感を喚起する場面として知られている [3,4]。しかし、口頭発表では質疑などで即興的な発言を要求されることがあるため、万全な準備は不可能であり、緊張感の抑制が特に困難な状況であると考えられる。口頭発表で緊張感が喚起される原因は多岐に渡り、Bippus et al. が行ったアンケートでは、失敗への恐怖心や経験の浅さ、聴衆から自身がどのように見えているかを過度に意識することなどが緊張感を引き起こす代表的な要因として挙げられた [5]。この研究から、緊張感が生じる背景には成功イメージ形成の困難性や、聴衆の反応をはじめとした感覚情報があると考えられる。そこで本稿では、成功した感覚の擬似的な体験や即時的な感覚情報操作

を行うための手法を提案し、効果検証の結果を報告する。

2. 関連研究

2.1 変換聴覚フィードバックの概要と効果

本研究では発話活動と親和性の高い緊張緩和システムを構築するための手法として、人間の感情に作用する工学的手法の1つとして知られる変換聴覚フィードバック (Altered auditory feedback, AAF) を採用する。AAFとは、マイクへの入力音声に何らかの変換処理を施し、ヘッドホン等を通して発話者に変換音声をリアルタイムで聞かせる手法である。

AAFには人間の発声方法の調節や特定の感情の喚起などの用途がある。Burnett et al. の実験では、フィードバック音声のピッチを上下させると多くの発話者がそれと逆方向に実際の発話音声のピッチを変化させることが示された [6]。また、フィードバック音声の音量を上げると発話者が声を潜めることも明らかにされている [7]。これらの事例から、人間は普段より自身の声のフィードバックに基づいて発声方法を調節していると考えられている。

また、声のフィードバックは発声方法の調節の他に発話者の感情の推測にも用いられる。これを利用した AAF の用途として、発話者の感情操作がある。Aucouturier et al. は、入力音声を喜び・悲しみ・恐れ各感情が喚起された状態での発話音声のように変換するプラットフォームを開発した [8]。効果検証実験では、参加者が与えられた文章を朗読する間、発話音声を特定の感情が表現された変換音声へと徐々に変化させながらフィードバックした。その結果、ほとんどの参加者が音声変換処理に気がつかないまま、意図した感情が喚起されたことが示された。この結果

¹ 東京大学大学院情報理工学系研究科
〒113-8656 東京都文京区本郷 7-3-1

^{a)} naruse@cyber.t.u-tokyo.ac.jp

から、人間は自身の発話音声と感情とが合致するか否かを意識しているのではなく、発話音声から聞き取れた感情を自身の感情だと推測している可能性が示唆された。

本研究では AAF の心理的効果を利用し、緊張状態での発話音声に現れやすい特性を抑制した変換音声のフィードバックを行い、自身が緊張状態でないと発話者に知覚させることで緊張感の緩和を図る。

2.2 緊張感と声に関する研究

緊張感の緩和を目的とした AAF システムを設計するためには、緊張状態での発話音声の特性について整理し、それを抑制する音声変換手法を考える必要がある。

人間の声質の違いは声道形状の違いに由来する。声道形状は感情によって変化することが知られており、声の音響パラメータと感情との間には対応関係が生じる [9]。特に感情の覚醒度の高さと声との間には強い相関があることが知られている。覚醒度と正の相関を持つ音響パラメータには、話速・基本周波数・周波数スペクトルのエネルギー分布における高周波数成分の比率などが挙げられる [10]。また、基本周波数の微小な振動も覚醒度の増大によって増加することが知られ [11]、AAF による感情喚起の研究においても、恐れを表す音声変換として入力音声のピッチを振動させる手法を採用している。

Russell は感情を 2 つの尺度 (快-不快・覚醒-鎮静) で説明する円環モデルを提案しており、緊張状態 (nervous) を覚醒度の高い不快感情として、また対極である安静状態 (calm) を覚醒度の低い快感情として位置付けている [12]。前述したように覚醒度と声の音響パラメータとの間には比較的強い関係性があり、覚醒度の高い状態での特性が緊張状態での声に現れる傾向がある。基本周波数の増加によるピッチの上昇、高周波数成分のエネルギー増大による声質の鋭利化、声の震えなどがその代表例である [13]。快感情の強度と発話音声との関係を調査した研究 [14-16] も存在するが、覚醒度に関する研究よりは少なく、報告されている結果も多岐にわたるため、覚醒度との関係ほどは体系化されていないと考える。

本研究では AAF による緊張緩和の手法として、緊張による発話音声の変化を音声変換により抑制し、発話者にフィードバックすることを考える。

3. 提案手法

感情は声の音響パラメータに影響を及ぼし、緊張状態にある人間の声にも幾分共通した変化が現れる傾向がある。また、感情的にニュートラルな発話者が AAF を使用することにより、フィードバック音声に表現された感情が実際に喚起された事例が存在する [8]。したがって、発話者が自身の声に上ずりや震えなどを確認した場合、自身が緊張状態にあると推論することで緊張感が喚起される可能性が

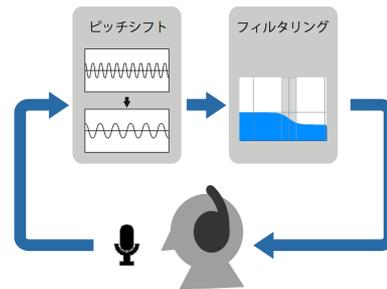


図 1 発話音声入力から変換音声出力までの流れ

ある。そこで、本研究では緊張感を抑制するため、緊張状態での声に現れる特性を音声変換により打ち消すことを考えた。

声から自身の感情の推論を行うためには、出力音声を自身の声だと知覚できることが前提条件となる。したがって、リアルタイムな音声変換処理を行い、発された言葉が発話と同時に聞こえるような AAF システムを設計する必要がある。本研究では音響パラメータのうち、リアルタイムな変換が比較的容易に実現できる基本周波数と周波数スペクトルのエネルギー分布に変換処理を行うものとした。

システム構築には音声信号処理に用いられることの多いビジュアルプログラミング言語である Max7 を用いた。発話音声の入力から変換音声の出力までの流れを図 1 に示す。マイクに入力された音声には初めにピッチシフトが行われる。その後、ピッチ変換された音声をフィルタリングし、周波数スペクトルのエネルギー分布を変換した後、ヘッドホンから出力する。

以下では、ピッチシフトとフィルタリングの 2 つの音声変換手法について詳細を説明する。

3.1 ピッチシフト

本システムにおけるピッチシフトとは、入力音声の基本周波数を恒常的に一定値移動させることで、聞こえる音の高さであるピッチを変化させることを言う。発話音声の基本周波数は感情の覚醒度の高まりによって増加する傾向があり、これは喉頭や声帯の筋緊張に由来すると言われている。本システムでは、入力音声のピッチを下げることで、緊張による発話音声の変化を抑制する。また、比較条件として、緊張状態での発話音声の特性を強調する音声変換では、入力音声のピッチを上げる変換を行った。AAF により感情を喚起するシステムである DAVID ?? が 30~50 セント程度のピッチシフトを行っていることから、本システムではフィードバック中に一貫して入力音声のピッチを 50 セント変化させるものとした。ピッチシフトにおける 1 セントは半音の 1/100 に相当し、ピッチが 1 セント上昇することは周波数が $2^{1/1200}$ 倍になることを意味する。ピッチ

シフトには Zynaptiq 社の ZTX *1を用いた。

3.2 フィルタリング

本システムではフィルタを用いて入力音声の周波数スペクトルのエネルギー分布を変換する。基本周波数と同様、高周波数成分のエネルギー比率も感情の覚醒度の高まりによって増大する傾向があることが知られている。そこで、ゲインを負の値に定めたハイシェルフフィルタを使用し、カットオフ周波数よりも高い周波数領域の信号を減衰させることで、緊張による声質の鋭利化の抑制を図った。緊張状態での発話音声の特性を強調する比較条件においては、ゲインを負の値に定めたローシェルフフィルタを用いることで、カットオフ周波数よりも低い周波数領域の信号を減衰させた。

4. 実験 1 : AAF の緊張緩和効果の検証実験

AAF が発話者の緊張感に及ぼす効果を検証するため、緊張感を喚起する実験として知られる Trier Social Stress Test(TSST) [17] の手順に則り、希望職に就くための面接場面を設定し、AAF を使用しながら 5 分間のスピーチをする実験を行った。

4.1 実験条件と変換パラメータ

実験は 3 条件からなる参加者間計画で行った。実験参加者は実験開始前日までに後述の特性不安検査 [18] に回答し、性格特性として不安になりやすい程度を表す得点が各参加者について算出された。点数の平均値が条件間で可能な限り均等になるよう、各参加者に 1 つずつ条件を割り当てた。

用意した条件と使用した音声変換のパラメータは以下の通りであった。

- 安静声条件：緊張時の声に現れる特性を抑えた変換音声フィードバックした。
 - ピッチシフト：50 セント下降
 - フィルタ：カットオフ周波数 1 kHz、オクターブ毎のゲイン-9.5 dB のハイシェルフフィルタ
- 緊張声条件：緊張時の声に現れる特性を強調した変換音声フィードバックした。
 - ピッチシフト：50 セント上昇
 - フィルタ：カットオフ周波数 1 kHz、オクターブ毎のゲイン-9.5 dB のローシェルフフィルタ
- フィードバックなし条件：実験参加者はスピーチ中にマイク付きヘッドホンを装着したが、音声フィードバックは行わなかった。

参加者は 21 歳から 24 歳までの男性 8 名で、平均年齢は 22.0 歳だった。面接官役としてスピーチを観察する実験協

力者は男性 2 名で、それぞれ 22 歳と 23 歳だった。

4.2 実験手順

実験は大学内の静かで閉じた部屋で行った。実験の様子を再現を図 2 に示す。室内には机を 1 つと椅子を 3 つ配置し、机の上には PC を 2 台設置した。1 台の PC は音声変換プログラムの起動とアンケートの回答に使用し、マイク付きヘッドホンと接続した。もう 1 台の PC は皮膚電位の計測に使用し、計測器と接続した。

参加者は実験に関する説明文書を読み、参加への同意書に記入を行った。この時、緊張感の自己評価への余分な影響を排除するため、参加者には実験の真の目的(スピーチ中の AAF 使用による心理的効果の検証)については説明せず、スピーチ時の人間の振る舞いに関する研究であると伝えた。

初めに、スピーチ中に AAF を行う安静声条件と緊張声条件の参加者は、ヘッドホンからの出力音声の音量調整を行った。参加者は音声変換処理を行わないフィードバックの実施下で、事前に用意された文章を朗読しながら音量を調整した。音量は、実際の発話音声よりもヘッドホンからの出力音声の方がはっきりと聞こえる程度まで上げるように指示した。フィードバックなし条件の参加者はこの操作を行わなかった。

次に、皮膚電位計測のために参加者は腕に電極を装着し、3 分間目を閉じて安静状態をとった。その後、後述の状態不安検査 [18] に回答し、ベースラインにおける緊張感の自己評価を行った。

回答後、スピーチのテーマを参加者に提示し、5 分間のスピーチ準備時間を設けた。テーマは希望の職に就くための面接場面を想定した自己 PR とした。準備時間では A4 の用紙 1 枚をメモとして自由に使用することを認めたが、スピーチ中はメモを見てはいけないものとした。

準備時間終了後、協力者 1 名が入室し、机を挟んで参加者と向き合うように着席した。参加者はその場で起立し、5 分間のスピーチを行った。安静声条件と緊張声条件ではスピーチ開始の直前に AAF を開始した。スピーチ中、実験者は参加者の目につかない場所へ移動した。協力者はスピーチに対して頷き、相槌、表情の変化を含む一切の反応をせず、発話が途中で止まっても特に指示をしないものとした。

スピーチ終了後、実験者は元の位置に戻り、協力者は退室した。参加者は着席してヘッドホンを外し、再び状態不安検査に回答して緊張感の自己評価を行った。回答後、皮膚電位の計測を終了し、最後に参加者は実験に関する質問やヘッドホンからの出力音声に関する質問などを含む自由記述のアンケートに回答した。

*1 <http://www.zynaptiq.com/ztx/>



図 2 実験の様子の再現 (左は協力者, 右は参加者)

4.3 評価方法

スピーチにより喚起される緊張感は、状態-特性不安尺度 (State-Trait Anxiety Inventory, STAI) [18] と、皮膚電位反応 (Skin Potential Reflex, SPR) を用いて評価した。

STAI とは質問紙により測定可能な不安尺度であり、性格特性としてどの程度不安かを表す特性不安 (A-Trait) と、測定時の心理状態としてどの程度不安かを表す状態不安 (A-State) を独立して測定することができる。本実験では参加者への条件の割り当てに特性不安検査を使用した。また、安静状態終了直後とスピーチ終了直後に状態不安検査を行うことで、スピーチによる不安度の増分を測定した。

SPR は短期的な皮膚電気活動の 1 つであり、ストレスや興奮による精神性発汗が皮膚表面の電気伝導性を変化させることに由来する反応である [19]。本実験では、3 分間の安静状態、5 分間のスピーチ準備時間、5 分間のスピーチ本番について SPR の総量 (基線からの振幅の積算量) をそれぞれ求め、各タイミングについて 1 分間あたりの平均を求めた。

また、実験の最後に行ったアンケートでは、参加者はフィードバック音声によるスピーチへの集中妨害度を 1 (全く感じなかった) から 5 (非常に感じた) の 5 段階で評価したほか、スピーチ中の思考やフィードバック音声への意見を自由記述で回答した。

4.4 実験結果

状態不安検査について、スピーチ課題直後での点数から安静状態終了直後での点数を引いた差分を図 3 に示す。安静声条件 ($M=-3.0$, $SD=8.89$)、緊張声条件 ($M=8.0$, $SD=16.37$)、フィードバックなし条件 ($M=19.5$, $SD=31.82$) の 3 条件を比較し、クラスカル・ウォリス検定を行った結果、有意差は得られなかった ($p = 0.4243$)。効果量として Morse の手法 [20] で求めた相関比は $\eta^2 = 0.2450$ だった。結果として有意差は得られなかったものの、安静声条件では、安静状態終了直後と比較してスピーチ課題直後の不安度が下がる傾向が見られた。

SPR の推移を図 4 に示す。安静声条件の参加者 1 名については、機器の不具合により正常な測定が不可能であったため、解析から除外した。フィードバックを使用した 2 条件においては、準備時間と比較してスピーチ中の反応量

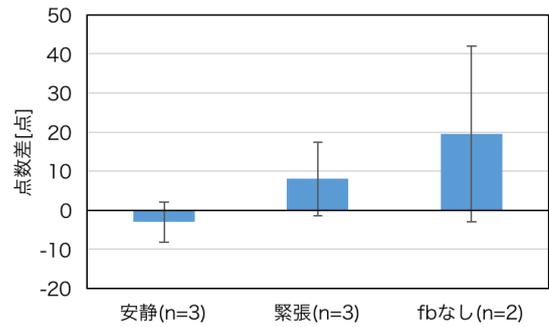


図 3 状態不安検査の点数差 (エラーバーは標準誤差)

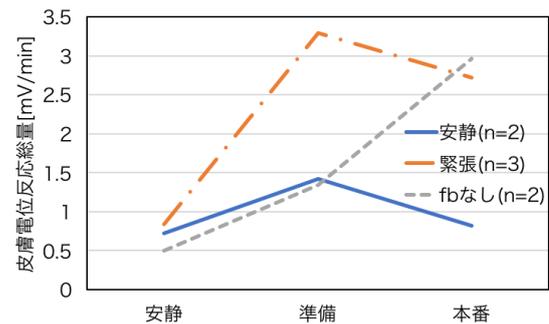


図 4 皮膚電位反応量の推移

質問：ヘッドホンから流れる音声が原因でスピーチ課題に集中できないと感じましたか？

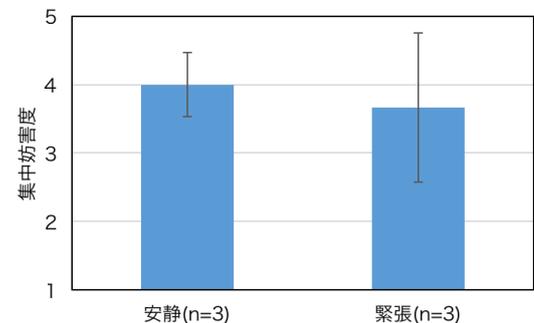


図 5 フィードバック音声によるスピーチの集中妨害度 (エラーバーは標準誤差)

は減少傾向にあった。一方、フィードバックなし条件ではスピーチ中に反応量が著しく増加した。

フィードバックを使用した 2 条件について、フィードバック音声によるスピーチへの集中妨害度についてのアンケート結果を図 5 に示す。安静声条件では $M=4.0$, $SD=0.816$ 、緊張声条件では $M=3.67$, $SD=1.89$ であった。

4.5 考察

状態不安検査では、安静声条件のみにおいてスピーチ終了直後の不安度が安静状態終了直後の不安度を下回る傾向が見られた。この結果から、安静声条件での AAF が緊張緩和効果を生じたことが示唆された。しかし、緊張声条件

でも不安度の増大はフィードバックなし条件よりも小さく抑えられる傾向が見られ、緊張状態での声の特性を強調することによる緊張感の増大は確認されなかった。

SPRについても、安静声条件・緊張声条件ともにフィードバックなし条件よりもスピーチ中に緊張感が緩和された可能性が示唆された。AAFを使用した2条件間では緊張緩和効果に大きな差は見られなかった。

以上の結果から、音声変換のパラメータの違いは緊張感に大きく影響せず、AAFの使用の有無が緊張感に影響したと考える。実験の最後に行った自由記述式のアンケートにおいても、フィードバック音声の声質に関する意見は少なく、AAFのシステム自体に関する意見が多く見られた。

AAFに対する肯定的な意見には、自身の声をヘッドホンを通して聞くことにより、自身の発言や発声方法を客観的に確認できたという回答があった。

自分が言ったことがよく理解できた気がしたので喋りやすく感じた。恥ずかしかったがどうすればうまく喋れるかわかるような気がした。(緊張声条件)

反対に、AAFの欠点として多くの参加者が挙げたものは、音声入出力間の遅延の存在であった。スピーチ課題中にAAFを使用した6名の参加者を対象に、フィードバック音声に関する自由記述式のアンケートを行ったところ、6名中4名が遅延による発言や心理への影響について言及した。発言音声の入力からフィードバック音声の出力までに遅延が存在した場合、非吃音者の発言の流暢性が失われることが確認されている [21]。本実験でもAAFの使用により発言が困難化したという回答が多く、遅延がスピーチへの集中妨害の主要因となったと考える。

時間差で聴こえることでそちらに意識が持たかかれ、自分のスピーチ内容に集中できなかつた(緊張声条件)

しかし、遅延の存在によって、自分の声の変化を確認することが可能となり、スピーチに良い影響を与える場合もあった。AAF使用者は発言音声とフィードバック音声を異なるタイミングで聞くことになるため、しっかりと話せていることの自己確認が可能になったと考える。

自分の声が遅れて聞こえてくる(タイムラグ)が少し気になり、それに意識が向くと話しくく感じたが、後半その音に慣れてくると、自分がどういふ風に話しているのか、ちゃんと話せているというのがわかって安心する時もあつた。(安静声条件)

本実験では異なる2種類の変換パラメータを用いる条件を用意したが、2条件間で緊張緩和効果に大きな差が見られなかつた原因は、AAF使用者の意識が遅延に向けられ、フィードバック音声の声質が十分に意識されなかつたことだと考える。したがって、音声変換が緊張感に及ぼす影響

を検証するためには、遅延時間の短縮が必要となる。実験2では、遅延時間を短縮したシステムを新たに構築し、緊張緩和効果や集中妨害度に変化が見られるか検証する。

5. 実験2: システム遅延短縮と検証実験

5.1 遅延短縮システムの構築

実験1に使用したAAFシステムには音声入出力間に遅延が存在し、使用者の発言を阻害することがあつた。計測の結果、システムには常に50ms程度の遅延が存在し、機器の状態によっては数百msの遅延が生じる場合もあることが分かつた。各処理の所要時間を計測した結果、遅延の主な原因はピッチシフトであることが確認された。しかし、実験1ではピッチシフトに既成のシステムを使用していたため、処理時間の短縮は困難であつた。そこで実験2では、Python2を用いてAAFシステムを再構築することで、遅延時間が常に50ms以内の範囲となるまで短縮した。フィードバック音声の遅延と発言の流暢性との関係を調査した研究では、通常の話速では遅延時間が50ms以下であれば流暢性がほぼ損なわれないことが確認されている [21]。

5.2 実験条件

新たに構築したシステムを用いて実験1と同様のスピーチ実験を行い、スピーチへの集中妨害度や緊張感に変化が現れるか検証する。

実験は3条件からなる参加者間計画で行つた。用意した条件と使用した音声変換のパラメータは以下の通りであつた。

- 安静声条件: 緊張時の声に現れる特性を抑えた変換音声をフィードバックした。
 - ピッチシフト: 50セント下降
 - フィルタ: カットオフ周波数1kHz, オクターブ毎のゲイン-9.5dBのハイシェルフフィルタ
- 緊張声条件: 緊張時の声に現れる特性を強調した変換音声をフィードバックした。
 - ピッチシフト: 50セント上昇
 - フィルタ: カットオフ周波数1kHz, オクターブ毎のゲイン9.5dBのハイシェルフフィルタ
- フィードバックなし条件: 実験参加者はスピーチ中にマイク付きヘッドホンを装着したが、音声フィードバックは行わなかつた。

参加者は21歳から23歳までの14名(うち女性3名)で平均年齢は21.93歳だつた。参加者は謝礼として1名あたり1000円を受領した。また、面接官としてスピーチを観察する実験協力者は21歳から23歳までの男性3名で、平均年齢は22.0歳だつた。

5.3 実験手順・評価方法

実験1と同様の手順で、TSST [17]の手順に則つたス

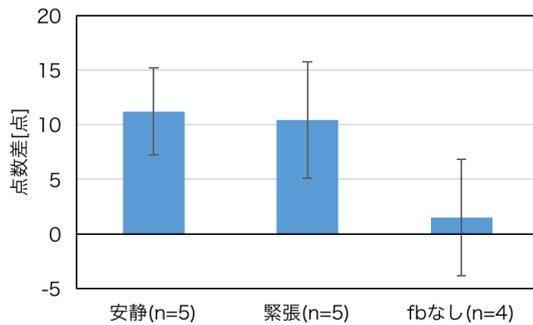


図 6 状態不安検査の点数差 (エラーバーは標準誤差)

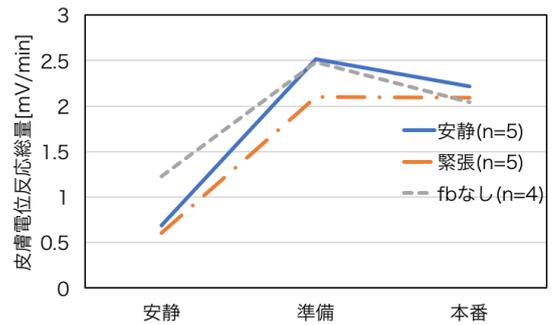


図 7 皮膚電位反応量の推移

ピーチ実験を実施した。

評価方法は実験 1 と同様に、主観指標として STAI [18] の状態不安検査を、客観指標として SPR 計測を採用した。SPR については、3 分間の安静状態、5 分間のスピーチ準備時間、5 分間のスピーチ本番における SPR の総量をそれぞれ計測し、各タイミングについて反応量 (基線からの振幅の積算量) の 1 分間あたりの平均を求めた。

また、実験終了後のアンケートには、実験 1 と同様に AAF による集中妨害度を 1(全く感じなかった) から 5(非常に感じた) の 5 段階で評価する質問のほか、フィードバック音声をどの程度自分自身の声だと思えたかを 1(全くできなかった) から 5(非常にできた) の 5 段階で評価する質問を追加した。これは、実験 1 において、外部機器によって自分の声を認識しているという意識が緊張感を喚起したという意見があったためである。

5.4 実験結果

状態不安検査について、スピーチ課題直後での点数から安静状態終了直後での点数を引いた差分を図 6 に示す。安静声条件 ($M=11.2$, $SD=8.90$), 緊張声条件 ($M=10.4$, $SD=11.93$), フィードバックなし条件 ($M=1.5$, $SD=10.63$) の 3 条件を比較し、クラスカル・ウォリス検定を行った結果、有意差は得られなかった ($p = 0.2903$, 効果量 $\eta^2 = 0.1908$)。しかし、実験 1 とは異なり、フィードバックなし条件で最も緊張感の増大が抑えられる結果となった。

SPR の推移を図 7 に示す。本実験では 3 条件に差は見られなかった。

フィードバックを使用した 2 条件について、フィードバック音声によるスピーチへの集中妨害度についてのアンケート結果を、実験 1 の結果と比較して図 8 に示す。実験 2 の結果は、安静声条件では $M=3.4$, $SD=1.2$, 緊張声条件では $M=3.6$, $SD=0.8$ だった。

参加者がフィードバック音声を自分自身の声だと思えたことのできた程度を図 9 に示す。安静声条件 ($M=3.4$, $SD=0.8$), 緊張声条件 ($M=3.2$, $SD=1.17$) の 2 条件を比較し、ウィルコクソンの順位和検定を行った結果、有意差は得られなかった ($p = 0.9048$, $r = 0.0378$)。

質問：ヘッドホンから流れる音声が原因でスピーチ課題に集中できないと感じましたか？

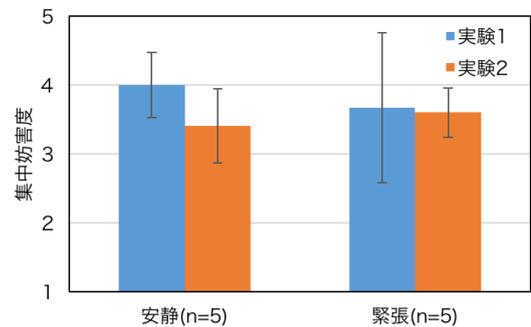


図 8 フィードバック音声によるスピーチの集中妨害度 (エラーバーは標準誤差)

質問：ヘッドホンから流れる音声を自分自身の声だと思えたことができましたか？

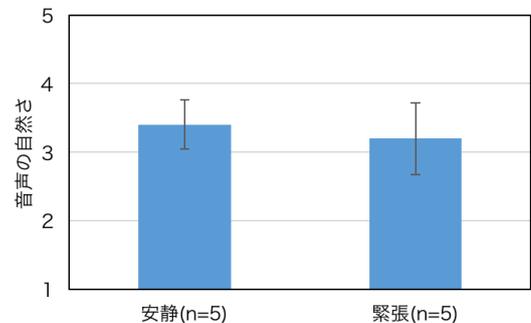


図 9 出力音声の自然さ (エラーバーは標準誤差)

5.5 考察

状態不安検査では、AAF を使用した安静声条件と緊張声条件での点数の増加分が、共にフィードバックなし条件を上回る結果となった。また、SPR では 3 条件に目立った差異はなかった。これらの結果から、システムの遅延時間を短縮した本実験では、短縮前の実験 1 とは異なり、AAF の使用がスピーチ中の緊張感を高める傾向にあったと考える。また、実験 1 と同様に、本実験でも変換パラメータが緊張感に影響した可能性は低いと考える。

以上の結果を得た原因として、2 つの可能性を考える。1

つは、遅延が存在しない分、AAF 使用者がフィードバック音声の違和感に意識を向けるようになった可能性である。実験の最後に行った自由記述式アンケートでは、多くの参加者が遅延に言及した実験 1 とは異なり、本実験では遅延が存在したという回答は見られなかった。しかし、遅延が短縮されたにもかかわらず、AAF の使用によるスピーチへの集中妨害度に大きな減少は確認されなかった。アンケートでは、聞きなれない変換音声が入力されることへの抵抗感や、自身の声のみが増幅されて聞こえることによる周囲からの疎外感などに関する指摘が多く見受けられた。遅延が短縮されたにもかかわらず、AAF の使用によるスピーチへの集中妨害度にあまり大きな減少が見られなかった。以上から、実験 1 で遅延に向けられていた使用者の意識が、実験 2 ではフィードバック音声の違和感へと向けられるようになったことで、AAF による緊張緩和効果が得られなくなったと考える。

普段聞き慣れていない自分の声が入ることで、違和感がずっとあった。(安静声条件)

始めの数十秒だけ、音声が戸惑って口ごもることがあった。(安静声条件)

2 つ目は、遅延そのものが緊張緩和に寄与していた可能性である。実験 1 では、遅延の存在によって AAF 使用者が声の変化を知覚することが可能となり、緊張緩和へ繋がった可能性があると考えた。実験 1・2 ともに変換パラメータの違いにより結果に大きな差は見られなかったため、発話音声とフィードバック音声の相違点が意識された可能性は低いと考える。しかし、発話音声とフィードバック音声を異なるタイミングで聞くことで、使用者が声の変化を無意識に知覚し、しっかりと話すことが出来ている感覚を知覚したことが緊張緩和に寄与した可能性がある。

遅延には発話阻害を生じるという欠点があり、通常ではスピーチ中の使用に適さないと考える。そのため、以後は発話阻害を生じない方法で声の変化を知覚させるようシステムを改善する必要がある。

6. まとめと今後の展望

本研究では、口頭発表場面での緊張感が原因となる意思伝達の円滑性低下の問題を解決するため、緊張緩和を目的とした AAF システムを構築し、効果検証を行った。

人間の感情と発話音声との間には対応関係があることが知られている。先行研究には、入力音声を特定の感情が表現された声に変換しフィードバックする AAF システムを用いて、発話者に感情を喚起させた事例がある。本研究では AAF の心理的効果を利用し、緊張状態での発話音声に現れる特性を抑制した変換音声をフィードバックすることで、自身が安静状態であると発話者に知覚させることを図った。

提案手法の緊張緩和効果を検証するため、面接場面を想

定したスピーチ実験を AAF 使用のもとで行った。その結果、AAF の使用によりスピーチ中の緊張感が緩和される傾向が見られたが、変換パラメータの違いによる緊張緩和効果の差は確認されなかった。この結果に関し、システムに存在していた音声入出力間の遅延が発話阻害を起こしたことで、AAF 使用者がフィードバック音声の音質を十分に意識できなくなったことが原因であると考えた。

変換パラメータが緊張感に及ぼす影響を詳細に検証するため、遅延時間を短縮した AAF システムを新たに構築し、同様の実験を行った。その結果、1 回目の実験の結果に反し、AAF の使用により緊張感が増大したことが示唆された。また、2 回目の実験においても変換パラメータの違いが緊張感に変化をもたらしたことは確認されなかった。この結果を得た原因として、遅延が解消されたことで AAF 使用者がフィードバック音声の音質の違和感へと意識を向けるようになった可能性と、実験 1 では遅延の存在により使用者が声の変化を無意識に知覚し、しっかりと話していることの自己確認が可能となったことが緊張緩和に寄与していた可能性を考えた。

今後は 2 回の実験で明らかとなった問題点を解消し、システム改善に努める。フィードバック音声への違和感を軽減するほか、遅延による発話阻害を生じることのない方法で声の変化を知覚させ、しっかりと話すことが出来ている感覚を提示する必要があると考える。また、フィードバック音声の質を向上させるため、本研究で行った音声変換が安静状態での声を生成する手法として十分な妥当性を有していたか検証し、変換手法の改善を行う必要があると考える。

また、効果検証の精度を向上させるため、検証方法に関しても検討すべきであると考えられる。今後はサンプルの偏りの解消や拡大に努めるほか、AAF 使用者の印象やスピーチパフォーマンスの客観的な評価により、提案手法が実際に使用者へ利益をもたらすかを検証していく。

AAF の即時的な心理的効果を活用することで、口頭発表のみならず、面接や日常的なコミュニケーションなど万全な対策を行うことが困難な場面にも適用可能な緊張緩和システムが実現できると考える。今後は更に緊張緩和効果やパフォーマンス変化の検証を進めながらシステムの改善や使用性の追求を行い、意思伝達の円滑性向上を可能とするインターフェースの開発を目指す。

参考文献

- [1] 有光興記：「あがり」のしろうと理論：「あがり」喚起状況と原因帰属の関係，社会心理学研究，Vol. 17, No. 1, pp. 1-11 (2001).
- [2] 岩田彩香，川井智理，齋藤順一，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).

- [3] Jarvis, M.: *Sport psychology: A student's handbook*, Routledge (2006).
- [4] Craske, M. G. and Craig, K. D.: Musical performance anxiety: The three-systems model and self-efficacy theory, *Behaviour Research and Therapy*, Vol. 22, No. 3, pp. 267–280 (1984).
- [5] Amy M. Bippus & John A. Daly: What do people think causes stage fright Naive attributions about the reasons for public speaking anxiety, *Communication Education*, Vol. 48, No. 1, pp. 63–72 (1999).
- [6] 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 (1997).
- [7] 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 (1961).
- [8] Aucouturier, J.-J., Johansson, P., Hall, L., Segnini, R., Mercadié, 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* (2016).
- [9] 北村達也：情動による声道形状変化のMRI観測，第23回日本音声学会全国大会予稿集，pp. 45–50 (2009).
- [10] Briefer, E.: Vocal expression of emotions in mammals: mechanisms of production and evidence, *Journal of Zoology*, Vol. 288, No. 1, pp. 1–20 (2012).
- [11] Juslin, P. N. and Scherer, K. R.: Vocal expression of affect, *The new handbook of methods in nonverbal behavior research*, pp. 65–135 (2005).
- [12] Russell, J. A.: A circumplex model of affect., *Journal of personality and social psychology*, Vol. 39, No. 6, p. 1161 (1980).
- [13] Laukka, P., Linnman, C., Åhs, F., Pissioti, A., Frans, Ö., Faria, V., Michelgård, Å., Appel, L., Fredrikson, M. and Furmark, T.: In a nervous voice: Acoustic analysis and perception of anxiety in social phobics' speech, *Journal of Nonverbal Behavior*, Vol. 32, No. 4, p. 195 (2008).
- [14] Taylor, A. M., Reby, D. and McComb, K.: Context-related variation in the vocal growling behaviour of the domestic dog (*Canis familiaris*), *Ethology*, Vol. 115, No. 10, pp. 905–915 (2009).
- [15] Fichtel, C., Hammerschmidt, K. and Jürgens, U.: On the vocal expression of emotion. A multi-parametric analysis of different states of aversion in the squirrel monkey, *Behaviour*, Vol. 138, No. 1, pp. 97–116 (2001).
- [16] Soltis, J., Blowers, T. E. and Savage, A.: Measuring positive and negative affect in the voiced sounds of African elephants (*Loxodonta africana*), *The Journal of the Acoustical Society of America*, Vol. 129, No. 2, pp. 1059–1066 (2011).
- [17] Kirschbaum, C., Pirke, K.-M. and Hellhammer, D. H.: The 'Trier Social Stress Test' – a tool for investigating psychobiological stress responses in a laboratory setting, *Neuropsychobiology*, Vol. 28, No. 1-2, pp. 76–81 (1993).
- [18] Spielberger, C. D., Gorsuch, R. L. and Lushene, R. E.: Manual for the state-trait anxiety inventory (1970).
- [19] 牛山美和，中揮朗，小林正義，千島亮，佐藤陽子，牛山喜久：皮膚電位反応を用いた情動反応評価に関する基礎的検討，紀要，Vol. 22, pp. 105–112 (1996).
- [20] Morse, D. T.: MINSIZE2: A computer program for determining effect size and minimum sample size for statistical significance for univariate, multivariate, and non-parametric tests, *Educational and psychological measurement*, Vol. 59, No. 3, pp. 518–531 (1999).
- [21] Stuart, A., Kalinowski, J., Rastatter, M. P. and Lynch, K.: 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 (2002).