変換聴覚フィードバックが話者の発話特性に与える影響の調査

大沼 怜生[1]　市川 あゆみ[2]　川口 一画[3]

概要：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にのみ聞かせる手法である変換聴覚フィードバックにおいて，話者の発話の音響特性を個別に変換した際の発話特性への影響について調査した．調査では，原稿を音読中の話者の発話のピッチ，抑揚，およびラウドネスをそれぞれ正負方向に変化させた変換聴覚フィードバックを行い，その際の話者の発話特性の変化を調査した．調査の結果，フィードバック音声の抑揚増加およびラウドネス減少において共に発話のシマの増加が確認された．また，フィードバック音声のラウドネス増加において発話のジッタおよびシマの減少，そしてラウドネスレベルの増加が確認された．調査結果および半構造化インタビューで得られた意見を基に変化の原因を考察し，調査に用いたシステムの課題を検証した．

Investigation of the Effect of Altered Auditory Feedback on Speakers' Speech Characteristics

1. はじめに

近年，オンライン通話ツール [9], [15], [25] の普及に伴いオンラインでの通話機会が増加している．これらのツールはユーザのカメラ映像を有効にするか無効にするか選択可能である物が多い．カメラ映像が無効の場合，表情等の視覚情報による非言語情報が伝達されないため，参加者は相手の声の高さや抑揚といった発話特性から気分や感情等を判断する必要がある．しかし，十分なボイストレーニングを受けていない話者が自分の声の発話特性を意図通りに制御することは困難である．

このような問題を解決する手法の 1 つとして，ボイスチェンジャの利用が挙げられる．近年の機械学習の発展に伴い，モデル生成型ボイスチェンジャにおける音声の変換精度は向上している [10], [21]．しかし，ボイスチェンジャを通した音声は発話の音響特性を変換して生成された音声であるため，声質がボイスチェンジャを通す前の音声から大きく変化する可能性がある．

その一方，話者の発話特性を変化させる手法の 1 つとして，変換聴覚フィードバック（Altered auditory feedback, AAF）が存在する．AAF とは，マイク等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 を用いることで話者の感情や発話特性をある程度制御することが可能である [2], [14]．本研究では，このような AAF の特性を活用することで，ボイスチェンジャを用いることなく，話者の発話特性を変化させるシステムを実現することができると考えた．発話特性の変化がインタラクションに与える影響として，発話音声のピッチ操作による話者の身体的および社会的優位性の向上 [18] や，発話音声のピッチおよびラウドネス操作による話者の発話の説得力の変化 [24] が示されている．このように，人間の発話において音響特性は非言語情報として意味を持つ．AAF によってボイスチェンジャを用いずにこれらの特性を変化させることができれば，話者の意図通りの気分や感情等を，話者に負担をかけることなく聞き手に伝えることが可能で

[1] 筑波大学 情報学群 情報メディア創成学類
College of Media Arts, Science and Technology, School of Informatics, University of Tsukuba
[2] 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
[3] 筑波大学 システム情報系
Faculty of Engineering, Information and Systems, University of Tsukuba

あると考えた．

しかし，AAF において発話の音響特性に与える影響および変換する音響特性のパラメータの違いによる発話特性への影響については，基本周波数，フォルマント，摩擦音の周波数重心，およびラウドネスに対する操作および分析が行われているのみである [6], [8], [11]．変換する音響特性および発話特性への影響について調査することは，対話支援システムを設計する上で発話特性の自在な変換を実現するために必要である．よって本研究では，これまでに調査が行われていない音響特性として抑揚にも着目し，AAF において発話の音響特性を個別に変換した際の話者の発話特性への影響を調査する．

ピッチ，抑揚，およびラウドネスについての調査に際し，特に抑揚に関しては前述した発話特性である基本周波数，フォルマント，摩擦音の周波数重心，およびラウドネスの分析では影響が判断できないと考えられる．そのため本研究では，抑揚の分析項目として新たに音声波形の周期方向の乱れであるジッタおよび振幅方向の乱れであるシマを追加する．

2. 関連研究

本章では，発話特性に影響を与える要因の 1 つとして，AAF が話者の心理状態に与える影響を調査した研究について述べる．その後，話者の発話特性に与える影響を直接調査した AAF 研究について述べる．

2.1 話者の心理状態に与える影響

AAF を用いることによって，話者の気分や感情といった心理状態に対して影響を与えることが知られている．Arakawa ら [1] は，機械学習を用いた AAF を使用することで話者の自己表象に影響を与える可能性を示している．Jean ら [2] は，スピーチ中の話者に気づかれることなく，話者の感情を幸せ，悲しみ，および恐怖の 3 種類に変化させることに成功した．この研究では，ピッチ，フォルマント，声の震え，および帯域通過フィルタを利用している．Taguchi ら [23] は，フィードバック音声のピッチ上昇により，映像会議中に会議参加者の気分を向上させ会話が活発になることを明らかにした．Jean ら [7] は，フィードバック音声のピッチ下降，フォルマント下降，およびローシェルフフィルタの利用により，対立中の話者の不安が軽減されることを発見した．また，フィードバック音声のピッチを下降させることにより，話者が自身をより力強く感じるようになると報告した．また，成瀬ら [17] は，フィードバック音声のピッチ下降およびハイシェルフフィルタの利用により，対面でのスピーチにおいて AAF の使用が話者の心理状態に影響を与える可能性を示唆している．

しかし，多くの AAF 研究は話者の心理状態を対象としており，音響特性のパラメータを対象とした AAF 研究は少ない．よって本研究では AAF における音声の音響特性に注目し，音響特性を示すそれぞれのパラメータが，話者の発話特性に与える影響について調査する．

2.2 話者の発話特性に与える影響

Burnett ら [6] は，AAF においてピッチを変更した際の話者の発話音声の変化を調査している．その結果，フィードバック音声のピッチが下がると発話音声のピッチが上がり，フィードバック音声のピッチが上がると発話音声のピッチが下がることを示した．Lane ら [11] は，AAF において音量を変更した際の話者の発話音声の変化を調査している．その結果，フィードバック音声の音量が上がると発話音声の音量が下がることを示した．Coughler ら [8] は，小児期の子供に対して AAF を用いた際の，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論文をまとめている．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発話特性の変化が見られることを発見した．また，いくつかの例において，大人よりも 4 歳以上の子供の方が AAF の効果が小さいことを報告した．

これらの研究により，ピッチ，ラウドネスに関する成人の発話特性への影響および子供の基本周波数，フォルマント，および摩擦音の周波数重心への効果は示されている．しかし，実際の会話への応用を考慮すると，成人に関してはピッチおよびラウドネスのみでは調査が不十分である．そのため本研究では，ピッチおよびラウドネスに加えて，これまでに調査が行われていない音響特性として，声の抑揚についての調査を行う．

また，抑揚は瞬間的なピッチの変動であることから，音声全体のピッチおよびラウドネスのみの分析では抑揚への影響が判断できないと思われる．そのため分析項目として新たにジッタおよびシマを追加し，AAF がジッタおよびシマに与える影響の分析を行う．

3. システムの設計

本研究では，AAF において変換する音響特性が発話特性へ与える影響を調査するため，ピッチ，抑揚，およびラウドネスを個別に変換可能な実験用システムを実装した．本章では，本研究で用いたシステムの設計について述べる．最初にシステムの構成を述べた後，変換する音響特性について述べる．

3.1 システムの構成

本システムは，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[20] を用いて実装を行った．DAVID を実行可能なコンピュータのサウンド入力を DAVID の入力とし，仮想オーディオケーブルを介して変換後の音声を再生する．また，サウンド入力および仮想オーディオケー

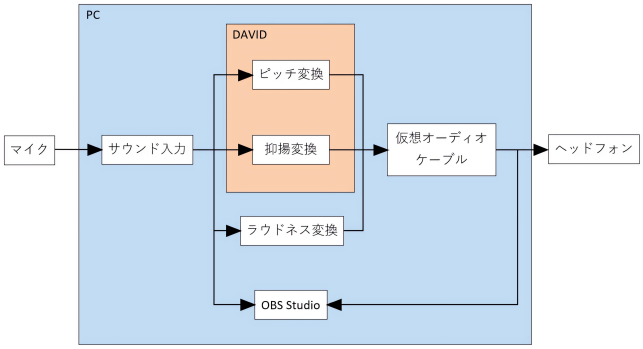

図 1 システム構成図
Fig. 1 System configuration diagram

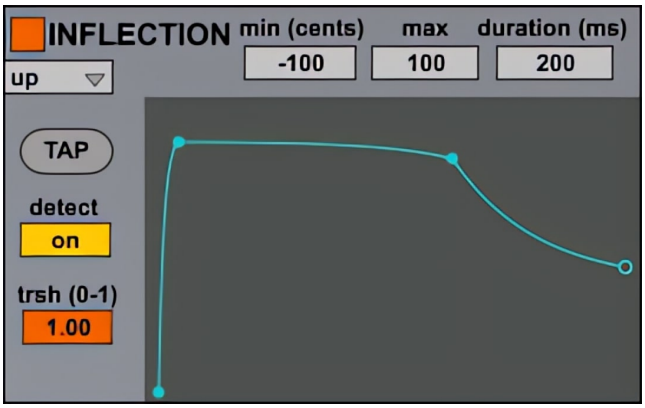

図 2 抑揚を増加させる設定
Fig. 2 Settings that increase inflection

ブルからそれぞれ発話音声およびフィードバック音声を取得し，Open Broadcaster Software（OBS Studio）[3] を用いて録音した．なお，著者の環境において，システムに入力された音声が変換され再生されるまでの時間（以降，遅延と記す）は 120 ms 程度であった．システムの構成を**図 1** に示す．

3.2 音響特性の変換方法

本研究で変換する音響特性のパラメータは，音声のピッチ，音声の抑揚，および音声のラウドネスの 3 種である．これらのパラメータは，音の三要素（音の高さ，音色，および音の大きさ）を基に，DAVID で変換可能なものを考慮して選出した．

ピッチは，DAVID の PITCH パラメータの操作により変換する．ピッチの影響を調査した先行研究 [6] を参考に，100 cent のピッチ上昇および 100 cent のピッチ下降を設定した．

抑揚は，DAVID の INFLECTION パラメータの操作により変換する．抑揚を増加させる設定を**図 2** に，抑揚を減少させる設定を**図 3** に示す．なお，これらの設定は，著者の主観に基づき変化が知覚可能な程度に設定された．

ラウドネスは，コンピュータのサウンド出力ボリュームの操作により変換する．ラウドネスの増加においては出力

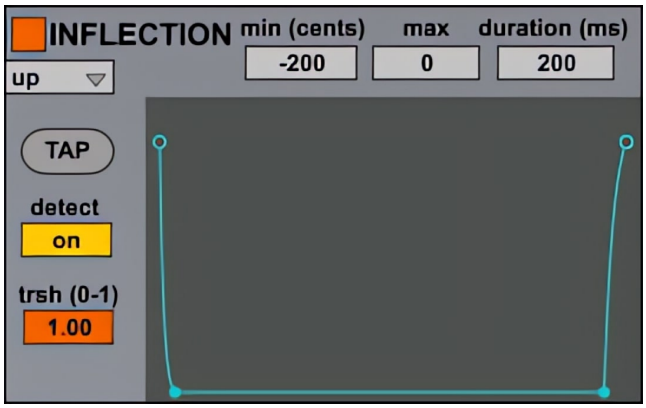

図 3 抑揚を減少させる設定
Fig. 3 Settings that decrease inflection

ボリュームを 5 上昇させ，ラウドネスの減少においては出力ボリュームを 5 下降させる．なお，これらの数値設定も，著者の主観に基づき通常時から音量が 50% 程度増減するように設定された．

4. 実験

音響特性の変換が話者の発話特性に与える影響を調査するため，実装したシステムを用いて実験を行った．本章では，実験条件，仮説，実験参加者，実験タスク，実験環境，および評価指標について述べる．なお，本実験は筑波大学システム情報系の倫理審査委員会の承認を受けて実施した．

4.1 実験条件

実験条件は以下の 6 条件で，すべて参加者内配置で設計した．

- c1. ピッチ上昇
- c2. ピッチ下降
- c3. 抑揚増加
- c4. 抑揚減少
- c5. ラウドネス増加
- c6. ラウドネス減少

すべての条件において，実験者は参加者のタスク実施中に音響特性のパラメータ変換を行う．

c1 では 100 cent 上昇させ，c2 では 100 cent 下降させる．c3 では抑揚増加のパラメータ変換を行い，c4 では抑揚減少のパラメータ変換を行う．c5 では，実験で使用したラップトップコンピュータのサウンド出力ボリュームを上昇させ，c6 ではサウンド出力ボリュームを下降させる．具体的な変換方法に関しては 3.2 節にて述べた通りである．

4.2 仮説

本研究では，以下の 3 点の仮説を設定した．

- H1. フィードバック音声のピッチの変換と逆向きに発話のピッチが変化する．

- H2．フィードバック音声の抑揚を変換すると発話音声が震える．
- H3．フィードバック音声のラウドネスの変換と逆向きに発話のラウドネスが変化する．

H1 については，Burnett ら [6] の研究にて，話者はフィードバック音声のピッチ変換と逆向きに発話のピッチを調節したと報告していることから，本研究においても同様の傾向がみられると考えた．

H2 ついては，抑揚を増加させた場合フィードバック音声のイントネーションが変化するため，話者が日常的に聞いているイントネーションとの乖離に話者の意識が向き，部分的に声が震えることが予想される．抑揚を減少させた場合においても，発話のイントネーションが変化する点は抑揚増加と共通しているため，抑揚増加と同様の変化が発生すると考えた．

H3 については，Lane ら [11] の研究にて，フィードバック音声の音量を増加させると発話の音量が減少すると報告していることから，本研究においても同様の傾向がみられると考えた．また，フィードバック音声の音量を減少させた際にも，ピッチ操作と同様の効果として，発話の音量が増加すると考えた．

4.3 実験参加者

実験参加者は計 18 名（男性 15 名，女性 3 名，平均 22.2 歳，標準偏差 1.18 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 6 条件すべてを実施した．実験の所要時間は 1 時間程度であった．実験参加者には筑波大学の規定に基づく謝金が支給された．

4.4 実験タスク

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．音量調整においては，ヘッドフォンから再生される音声の大きさと参加者の骨導音の大きさが，参加者の主観で同じ大きさになるよう指示した．

その後，参加者は実験者が用意した原稿を音読した．原稿を読む速度や抑揚等，読み方に関する指示は一切行わず，参加者は原稿を自由に音読した．なお，本研究は最終的に音声通話等他者との会話中に使用することを想定しているが，本実験ではその前段階として音響特性の変換が話者の発話特性に与える影響について詳細に調査するため，先行研究 [2] を参考に音読タスクを選定した．原稿は条件毎に異なる内容で，すべて村上春樹の短編小説「ノルウェイの森」および「羊をめぐる冒険」から 900 字程度を抜粋したものである．原稿および条件は対応しており，条件の実施順序はランダム化した．

参加者が原稿を音読している間，実験者はパラメータ変換の反映量を操作した．原稿を文字数で 3 分割し，最初の 300 字はパラメータ変換を行わず，続く 300 字でパラメータ変換の効果を最大の状態まで線形に増加させ，最後の 300 字ではパラメータ変換の効果が最大の状態で固定した．

図 4 実験環境
Fig. 4 Experimental environment

それぞれの条件に対し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．

4.5 実験環境

ヘッドフォン（SONY WH-1000XM4）およびモニタ（BenQ GW2480T）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．ラップトップコンピュータはスクリーンを消灯した状態で実験参加者の正面に配置した．実験者は，参加者の視界に入らない位置に配置したモニタでシステムの監視および操作を行った．参加者はヘッドフォンを装着し，ラップトップコンピュータ内蔵のマイクに向かって声を出した．実験環境を**図 4** に示す．

4.6 評価指標

評価指標として，実験参加者の発話音声に含まれる基本周波数（f0），ジッタ（PPQ5），シマ（APQ5），およびラウドネスレベルを用いる．このうち基本周波数，ジッタ，およびシマは，Jean らの先行研究 [2] で用いられていた評価指標のうち，発話の音響特性に関連する項目として選出された．また，ラウドネスレベルは，特徴量抽出に用いた Python のライブラリである Surfboard[13] において抽出可能な特徴量から選出された．

5. 実験結果

本章では，実験の各評価指標についての結果を示す．各評価指標について，パラメータ変換を行わない状態および変換効果が最大の状態の音声から特徴量を抽出し，比較および分析を行った．特徴量抽出後に等分散性および正規性を確認した．等分散性の検定には F 検定を，正規性の検定にはシャピロウィルク検定を行った．その後，正規性が認

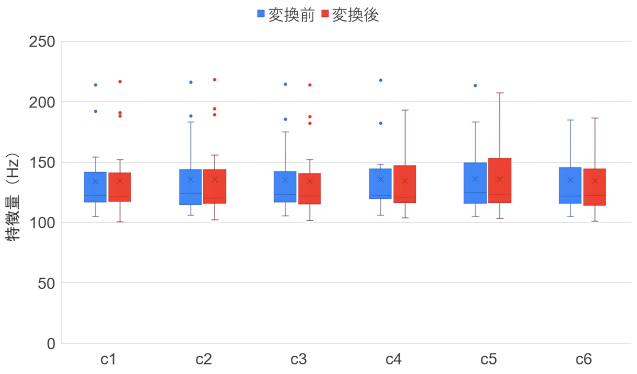

図 5 パラメータ変換前および変換後の基本周波数.

Fig. 5 Fundamental frequency before and after parameter conversion

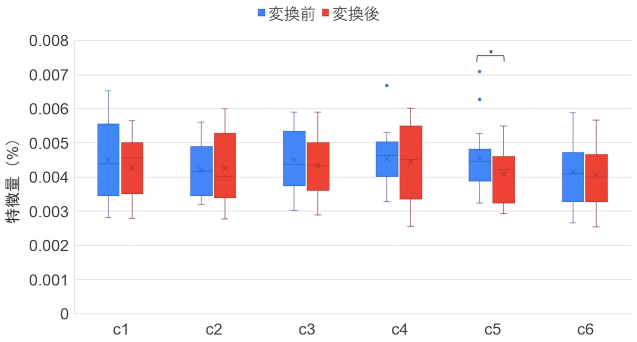

図 6 パラメータ変換前および変換後のジッタ．有意差が出た条件に*を示す.

Fig. 6 Jitter before and after parameter conversion. Indicate * for condition with significant difference.

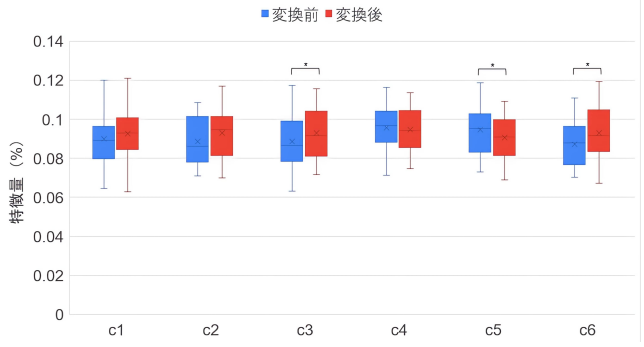

図 7 パラメータ変換前および変換後のシマ．有意差が出た条件に*を示す.

Fig. 7 Shimmer before and after parameter conversion. Indicate * for conditions with significant differences.

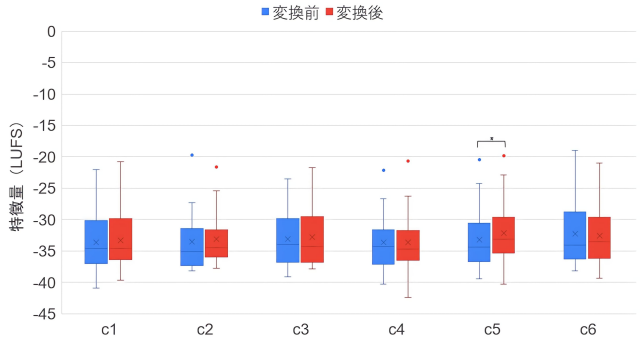

図 8 パラメータ変換前および変換後のラウドネスレベル．有意差が出た条件に*を示す.

Fig. 8 Loudness levels before and after parameter conversion. Indicate * for condition with significant difference.

められた場合は対応のある t 検定を行った．正規性が認められなかった場合はウィルコクソンの符号順位検定を行った．また，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．なお，統計分析には R[19] を用いた．

5.1 基本周波数

パラメータの変換前と変換後における，抽出された基本周波数を条件ごとに箱ひげ図としてまとめた結果を**図 5** に示す．ウィルコクソンの符号順位検定を行った結果，すべての条件において有意差が見られなかった（c1:$Z = 1.43, p = 0.15$, c2:$Z = 0.08, p = 0.93$, c3:$Z = 1.43, p = 0.15$, c4:$Z = 1.12, p = 0.26$, c5:$Z = 0.38, p = 0.70$, c6:$Z = 1.07, p = 0.28$）.

5.2 ジッタ

パラメータの変換前と変換後における，抽出されたジッタを条件ごとに箱ひげ図としてまとめた結果を**図 6** に示す．対応のある t 検定を行った結果，c5 において有意差が見られた（c1:$t(17) = 1.61, p = 0.13$, c2:$t(17) = -0.41, p = 0.68$, c3:$t(17) = 1.58, p = 0.13$, c4:$t(17) = 0.58, p = 0.57$, c5:$t(17) = 2.83, p = 0.01, d = 0.67$, c6:$t(17) = 0.42, p = 0.68$）.

5.3 シマ

パラメータの変換前と変換後における，抽出されたシマを条件ごとに箱ひげ図としてまとめた結果を**図 7** に示す．対応のある t 検定を行った結果，c3，c5，および c6 において有意差が見られた（c1:$t(17) = -1.33, p = 0.20$, c2:$t(17) = -1.72, p = 0.10$, c3:$t(17) = -2.28, p = 0.03, d = 0.53$, c4:$t(17) = 0.43, p = 0.67$, c5:$t(17) = 2.49, p = 0.02, d = 0.59$, c6:$t(17) = -2.81, p = 0.01, d = 0.66$）.

5.4 ラウドネスレベル

パラメータの変換前と変換後における，抽出されたラウドネスレベルを条件ごとに箱ひげ図としてまとめた結果を**図 8** に示す．正規性が認められた c1，c4，c5，および c6 にて対応のある t 検定を行った結果，c5 において有意差が見られた（c1:$t(17) = -0.88, p = 0.39$, c4:$t(17) = -0.07, p = 0.94$, c5:$t(17) = -4.26, p = 0.0005, d = 1.00$, c6:$t(17) = 1.47, p = 0.16$）．また，正規性が認められなかった c2 および c3 にてウィルコクソンの符号順位検定を行った結果，それぞれの条件において有意差が見られなかった

(c2:$Z = 1.29, p = 0.20$, c3:$Z = 0.90, p = 0.37$).

6. 議論および考察

本章では，4.2 節で設定した仮説の検証を行う．その後，5 章で得られた一部の結果に対して議論を行う．

6.1 仮説の検証

「H1．フィードバック音声のピッチの変換と逆向きに発話のピッチが変化する.」について，c1 および c2 において基本周波数に有意差は見られなかった．したがって H1 は支持されなかった．半構造化インタビューにて「自分の声を聞くことに慣れておらず，原稿を音読するだけで精一杯だった」(P1, P13, P14) という意見を得た．そのため，実験にトレーニング条件を追加することで結果に変化が見られると考えられる．

「H2．フィードバック音声の抑揚を変換すると発話音声が震える.」について，c3 におけるシマのみ有意に増加した．したがって H2 は部分的に支持された．しかし，シマの増加が抑揚の変化ではなくラウドネスレベルの変化に起因する可能性がある．そのため H2 の検証には慎重な判断が求められる．

「H3．フィードバック音声のラウドネスの変換と逆向きに発話のラウドネスが変化する.」について，c5 においてラウドネスレベルが有意に増加した．また，c5 以外においてラウドネスレベルに有意差は見られなかった．したがって H3 は支持されなかった．半構造化インタビューにて，「遅延があるため骨導音を集中して聞いていたが，スピーカ側の音が大きくなったため話しにくかった」(P7) という意見を得た．この意見から，参加者が話しにくさを解消するために骨導音を大きくすることを試み，発話のラウドネスレベルが増加したと考えられる．そのため，遅延を短縮することで結果に変化が見られる可能性があり，システム設計を見直したうえでの再調査が望まれる．

6.2 ジッタ・シマおよびラウドネスレベルの関係

実験の結果，c5 においてジッタおよびシマの減少とともにラウドネスレベルの増加が確認された．ジッタおよびラウドネスレベル，シマおよびラウドネスレベルの間にそれぞれ負の相関があることが報告されており [5]，本実験においても報告を再現したものとして考えられる．なお，報告で分析の対象とされていたのはラウドネスレベル (LKFS) ではなく音圧レベル (dB) であるが，LKFS および dB は同等のものとして扱うことが可能である [16].

一方，c6 においてシマのみが有意に増加している．これはシマおよびラウドネスレベルの相関に反する結果である．この原因を考察するためには情報が不足しており，シマおよびラウドネスレベルの関係や両者に影響を与える要因についての詳細な調査が求められる．

6.3 考察のまとめ

本研究の最終目標は，AAF の特性を活用し，ボイスチェンジャを用いることなく話者の発話特性を変化させるシステムを実現することである．システムの実装にあたり，今回の実験で得られた知見より，以下のような機能の実現が考えられる．

- 話者の発話が不明瞭な場合， AAF で発話のラウドネスを減少させることにより，発話のジッタおよびシマを減少させて発話を明瞭にする．
- 話者の発話の音量が小さい場合，AAF で発話のラウドネスを増加させることにより，発話のラウドネスレベルを増加させて発話の音量を上げる．

ただし，実際にコミュニケーションで用いる上では，上記の機能のみでは効果が限定的であると考えられる．そのため，AAF を用いた場合の発話特性の変化について，今後より多くの知見を明らかにしていく必要がある．

今回の実験で得られた知見が限定的であった理由として，実験に使用したシステムや実験設計に課題が存在したことが考えられる．7 章にて，本研究の制約および今後の課題について述べる．

7. 本研究の制約および今後の課題

7.1 フィードバックの遅延

半構造化インタビューの中で一部の参加者から，「システムで再生される音声の遅れにより発話が阻害された」(P2, P6, P9, P12, P18) という意見を得た．聴覚フィードバックにおいてフィードバック音声に 30 ms から 300 ms 程度の遅延が存在する場合，発話の流暢性が低下することが報告されている [4], [12], [22]．本システムにおける遅延は，3.1 節にある通り著者の環境で 120 ms 程度であり，遅延が原因となって参加者の発話が阻害された可能性がある．今後はシステムの遅延を，発話の阻害が発生しない程度に短縮する必要がある．

7.2 変換パラメータの選定

3.2 節で説明した通り，本研究では変換する音響特性としてピッチ，ラウドネス，および抑揚を設定した．しかしこれらのパラメータは DAVID の仕様に大きく依存している．今後はピッチ，ラウドネス，および抑揚以外の音響特性の調査を想定し，DAVID に依存しないシステムの設計が必要とされる．

また，パラメータの変換量に関しても再考を要する．本研究では抑揚およびラウドネスの変換量を著者の主観で設定した．そのため AAF において発話特性に変化を与える変化量の閾値については考慮されていない．今後は変換パラメータの選定とともに，パラメータの変換量についても精査が求められる．

7.3 評価指標の選定

4.6 節で説明した通り，本研究では実験の評価指標として発話音声の基本周波数，ジッタ（PPQ5），シマ（APQ5），およびラウドネスレベルを用いた．しかしこれらのパラメータが評価指標として必要十分であるか否かは不明である．そのため評価指標に用いる音響特性について精査する必要がある．

8. おわりに

本研究では，AAF において発話の音響特性を個別に変換した際の，話者の発話特性への影響について調査した．変換する音響特性としてピッチ，ラウドネス，および抑揚を設定し，発話の基本周波数，ジッタ（PPQ5），シマ（APQ5），およびラウドネスレベルを分析した．

調査の結果，フィードバック音声の抑揚を増加させると発話のシマが増加した．フィードバック音声のラウドネスを増加させると発話のジッタおよびシマが減少し，ラウドネスレベルが増加した．また，フィードバック音声のラウドネスを減少させるとシマが増加した．なお，どの音響特性を変換しても発話の基本周波数に有意差は見られなかった．

今後はシステム設計の見直しおよび評価指標の精査を行い，再度話者の発話特性への影響について調査する予定である．

参考文献

[1] Arakawa, R., Kashino, Z., Takamichi, S., Adrien, V. and Masahiko, I.: Digital Speech Makeup: Voice Conversion Based Altered Auditory Feedback for Transforming Self-Representation, *Proceedings of the 2021 International Conference on Multimodal Interaction*, ICMI '21, No. 9, New York, NY, USA, Association for Computing Machinery, pp. 159–167 (online), DOI: 10.1145/3462244.3479934 (2021).

[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（オンライン），DOI: 10.1073/pnas.1506552113 (2016).

[3] Bailey, H. J.: Open Broadcaster Software, obsproject (online), available from ⟨https://obsproject.com/⟩ (accessed 2023-09-29).

[4] 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).

[5] Brockmann, M., Storck, C., Carding, P. N. and Drinnan, M. J.: Voice loudness and gender effects on jitter and shimmer in healthy adults, *J Speech Lang Hear Res*, Vol. 51, No. 5, pp. 1152–1160 (online), DOI: 10.1044/1092-4388(2008/06-0208) (2008).

[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 (online), DOI: 10.1016/S0892-1997(97)80079-3 (1997).

[7] 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: https://doi.org/10.1145/3173574.3174205 (2018).

[8] Coughler, C., de Launay, K. L. 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).

[9] Discord: Discord, Discord (online), available from ⟨https://discord.com/⟩ (accessed 2023-09-29).

[10] Kim, J., Kong, J. and Son, J.: Conditional Variational Autoencoder with Adversarial Learning for End-to-End Text-to-Speech (2021).

[11] 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).

[12] 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).

[13] Lenain, R., Weston, J., Shivkumar, A. and Fristed, E.: Surfboard: Audio Feature Extraction for Modern Machine Learning (2020).

[14] M, L., A, P. and M, O.: 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).

[15] Microsoft: Microsoft Teams, Microsoft (online), available from ⟨https://www.microsoft.com/microsoft-teams/group-chat-software⟩ (accessed 2023-09-29).

[16] 難波精一郎：知っているようで知らないラウドネス，日本音響学会誌， Vol. 73, No. 12, pp. 765–773 (2017).

[17] 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).

[18] Puts, D. A., Gaulin, S. J. and Verdolini, K.: Dominance and the evolution of sexual dimorphism in human voice pitch, *Evolution and Human Behavior*, Vol. 27, No. 4, pp. 283–296 (online), DOI: https://doi.org/10.1016/j.evolhumbehav.2005.11.003 (2006).

[19] R Core Team: *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria (2018).

[20] Rachman, L., Liuni, M., Arias, P., Lind, A., Johansson, P., Hall, L., Richardson, D., Watanabe, K., Dubal, S. and Aucouturier, 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).

[21] RVC-Project: RVC-Project/Retrieval-based-Voice-Conversion-WebUI, GitHub (online), available from

⟨https://github.com/RVC-Project/Retrieval-based-Voice-Conversion-WebUI⟩ (accessed 2023-09-29).

[22] 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).

[23] Taguchi, W., Nihei, F., Takase, Y., Nakano, Y. I., Fukasawa, S. and Akatsu, H.: Effects of Face and Voice Deformation on Participant Emotion in Video-Mediated Communication, *Proceedings of the 20th International Conference on Multimodal Interaction: Adjunct*, ICMI '18, New York, NY, USA, Association for Computing Machinery, (online), DOI: 10.1145/3281151.3281159 (2018).

[24] Wang, T.-Y., Kawaguchi, I., Kuzuoka, H. and Otsuki, M.: Effect of Manipulated Amplitude and Frequency of Human Voice on Dominance and Persuasiveness in Audio Conferences, *Proc. ACM Hum.-Comput. Interact.*, Vol. 2, No. CSCW, pp. 1–18 (online), DOI: 10.1145/3274446 (2018).

[25] Zoom: Zoom Meetings, Zoom Video Communications (online), available from ⟨https://explore.zoom.us/ja/products/meetings/⟩ (accessed 2023-09-29).