

音量制御のためのホワイトノイズ手法と 音声フィードバック手法の構築と検証

原 史也^{1,a)} 竹川 佳成^{1,b)} 平田 圭二^{1,c)}

概要: 公共の場で友人との会話が盛り上がり、声が大きくなってしまい周りに迷惑をかけてしまう状況や、プレゼンテーションの際に緊張し声小さくなる状況など意識的に音量制御ができない場合がある。この問題を解決するために、筆者らの研究グループでは、周囲の雑音大きい騒音下では、人は不随意的に声が大きくなるというロンバード効果 [1] を応用し、2つの音量制御手法を提案する。音量増幅のために、ホワイトノイズを変化させユーザに聞かせることによって発話時のユーザの音量を制御するホワイトノイズ手法を提案する。また、音量減衰のために、ユーザ自身が聞き取る自身の音量 (以下:聴取音量) を変化させることで、発話時のユーザの音量 (以下: 発話音量) を制御する音声フィードバック手法を提案する。評価実験では、4パターンの増幅量を設定し、ホワイトノイズまたは音声フィードバックの変化量によって、発話音量がどのように変化するか調査することを目的とした。その結果、ホワイトノイズ手法ではホワイトノイズが大きくなるにつれ、発話音量は大きくなり、音声フィードバック手法では聴取音量が大きくなるにつれ、発話音量は小さくなっていくという結果が得られた。

1. はじめに

近年、ウェアラブルコンピューティング技術を活用し、ユーザの状況を考慮した情報提供手法に注目が集まっている。例えば、Google Glass や Apple Watch などのウェアラブル端末を用いて、目的地までのナビゲーション、買い物時のセール情報、歩数や心拍に基づくヘルスマネジメントなど、ユーザにとって有用な情報を提供することでユーザの行動を支援している。このように、既存の多くの情報提供手法は、ユーザに行動の選択肢を与える情報を提示し、特定の行動を促すものが多い。しかし、公共の場で騒いでいる子どもを静かにさせたい場合や、プレゼンテーションや面接など緊張している状態で萎縮せず明瞭に話したい場合など、ユーザが自分自身の行動を直ちに制御できない場合がある。したがって、ウェアラブルシステムにより、ユーザに行動を促すだけでなく、ユーザの行動を制御することは重要である。

この問題を解決するために、筆者らの研究グループでは、周囲の雑音大きい騒音下では、人は不随意的に声が大きくなるというロンバード効果 [1] を応用し、2つの音量制御手法を構築した。音量増幅のために、ホワイトノイズを

変化させユーザに聞かせることによって発話時のユーザの音量を制御するホワイトノイズ手法を提案する。また、音量減衰のために、ユーザ自身が聞き取る自身の音量 (以下: 聴取音量) を変化させることで、発話時のユーザの音量 (以下: 発話音量) を制御する音声フィードバック手法を提案する。

2. 関連研究

人の発声モデルに関する研究は古くから多数行われており、その代表的な例として、Speech Chain モデル [2] がある。耳に聞こえてくる周囲の音などの聴覚情報や相手との距離などの視覚情報をもとに発話をコントロールしている。また、音声フィードバックに特化した発声モデルとしてロンバード効果 [1] がある。例えば、我々は周囲の雑音大きい環境では、自然と声が大きくなる傾向があるが、騒音やホワイトノイズを聞かせると声が大きくなるだけでなく、基本周波数やフォルマント周波数が高くなるなど、様々な音響的特徴が変化することが知られている。ロンバード効果についてはさまざまな研究結果があり、例えば、程島ら [3] は、静かな環境・雑音のある環境・残響のある環境において発声の明瞭さを調べている。さらに、ロンバード効果を計算機システム上で実現する研究もある [4]。本研究ではロンバード効果において、特に自身の音量をリアルタイムに発話者にフィードバックしたときの効果について

¹ 公立はこだて未来大学

^{a)} g2117040@fun.ac.jp

^{b)} yoshi@fun.ac.jp

^{c)} hirata@fun.ac.jp

ても調査する点で既存研究と異なる。

一方、ウェアラブルデバイスを用いて人の行動を制御する研究事例も多数存在する。例えば、虚偽情報による整体情報制御 [5]、VR (Virtual Reality:バーチャルリアリティ) による視覚提示による食欲制御 [6]、振動モータによる振る舞い制御 [7] [8]、心拍情報の資格定時による演奏テンポの制御 [9] などさまざまなアプローチで、不随意的な行動制御を実現している。栗原らは、人工的に自分の発話を遅らせて聞かせる聴覚遅延フィードバックにより、肉体的な苦痛を与えることなく話し手の発話を阻害するシステム SpeechJammer を提案している [10]。発話の制御という点において本研究と Speech Jammer は関わりが深い。Speech Jammer を適用されたユーザは発話を直接的に阻害されることで発話を止めてしまう。本研究のアプローチは、音量に着目し、発話を阻害することなく発話音量の制御を行うという点で異なる。

3. 設計

3.1 提案する音量制御手法

本研究の音量制御手法は、ロンバード効果 [1] を応用した 2 つの手法を提案する。

ホワイトノイズ手法 ホワイトノイズ手法は、ユーザの発話音量を増幅させるための手法である。ユーザの発話音量を増幅させるためには、普段聞こえる自分自身の声より現在聞こえる自分自身の声を減衰させる。これにより、普段聞こえる自分自身の声と同じ発話音量を保とうとし、発話音量が増幅すると考えられる。この手法を構築するためにホワイトノイズを用いる。雑音に近い音であるホワイトノイズをユーザに与えることによって、ユーザ自身の声を相対的に小さく聞こえさせることで発話音量を制御する。

音声フィードバック手法 音声フィードバック手法は、ユーザの発話音量を減衰させるための手法である。ユーザの発話音量を減衰させるためには、普段聞こえる自分自身の声より現在聞こえる自分自身の声を増幅させる。これにより、普段聞こえている自分自身の声と同じにしようとし、結果、発話音量が減衰すると考えられる。つまり、普段自分の耳で聞いている環境音の音量より、現在自分の耳で聞いている環境音の音量を減衰させれば良い。そこで、聴取音量が減衰された状態から聴取音量を増幅された状態にすることで、発話音量は減衰すると考えられる。

3.2 利用シーン

提案する音量制御手法は、ユーザの音量を不随意的に制御できるという特徴がある。したがって、緊張などで萎縮してしまう状況や、逆にリラックスしつつも周囲への配慮を必要とする状況など、特殊な場合においても効果が期待

される。提案する音量制御手法は、次のようなシーンでの利用が考えられる。

ホワイトノイズ手法 就職活動や人前でのプレゼンテーションなどにおいて、緊張しないという人は珍しく、特に面接の重要度が増すにつれて感じる緊張は大きくなる。その後、時間が経過するとともに声が小さくなってしまふ。これらの問題を防ぎ、聴取者である面接官に伝わる音量へと制御することで声が聞こえない、伝わらないというケースを排除することができる。

音声フィードバック手法 自宅でのパーティや図書館、機内など、他人が近くにいる場合には大きな声を出すのはマナー違反である。しかし、話が盛り上がってしまうと、つい配慮を忘れてしまうことがある。本研究の提案する音量制御手法は、ユーザが不随意的に音量を制御するため、気を配ることができない状態であっても、その場の雰囲気を壊さずに音量を抑えられると期待される。

4. 評価実験

増幅量 (ホワイトノイズ手法ではホワイトノイズ、音声フィードバック手法ではイヤフォンから流れる自分自身の声) を適用した時における音量の変化を検証することで提案するホワイトノイズ手法と音声フィードバック手法の有用性を検証する。

被験者は大学生 7 名 (男性 6 名、女性 1 名) である。全ての被験者は本研究の詳細や実験目的を理解しておらず、被験者を募集する際には、文章を音読するというタスクを実施してもらおうとだけ伝えた。

4.1 実験システム

評価実験を実施するため、図 1 に示す実験システムを構築した。聴取音量の変化に伴う発話音量の変化を調べることが目的のため、マイクから取得した音声信号を分析する。被験者は防音室に入り、防音室内の 1.7 メートル離れているディスプレイに表示される文章を音読する。マイクで被験者の発話音量を取得し、マイクが生成した音声信号はパソコンとオーディオインタフェースに分配される。パソコンで音声信号を録音し、オーディオインタフェースを通じてホワイトノイズまたは編集済みの音声信号を被験者のヘッドフォンへ出力し、被験者はそれを聴取音量として認識する。

ディスプレイは三菱電機社の RDT27WLM-A、パソコンは TOSHIBA 社の dynabook を使用した。マイクはオーディオテクニカ社のモノラルマイクロホン AT9913iS、ヘッドフォンは TIMERY 社のヘッドセットマイク HEADSET-A010BK を使用した。オーディオインタフェースはオーディオテクニカ社のマイクロホンアンプ AT-MA2 を使用

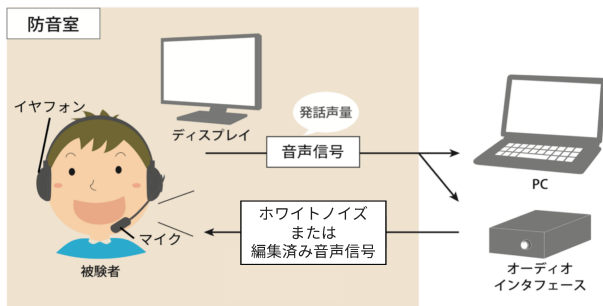


図 1 システム構成

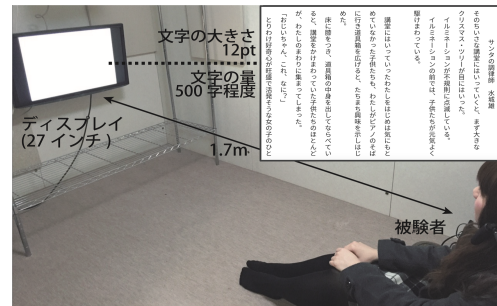


図 2 評価実験の様子

表 1 ホワイトノイズの増幅量

被験者	試行 1	試行 2	試行 3	試行 4
A	4dB	16dB	8dB	24dB
B	16dB	8dB	24dB	4dB
C	8dB	16dB	4dB	24dB
D	24dB	8dB	16dB	4dB
E	8dB	4dB	16dB	24dB
F	4dB	24dB	8dB	16dB
G	16dB	24dB	4dB	8dB

表 2 音声フィードバックの増幅量

被験者	試行 1	試行 2	試行 3	試行 4
A	4dB	16dB	8dB	24dB
B	16dB	8dB	24dB	4dB
C	8dB	16dB	4dB	24dB
D	24dB	8dB	16dB	4dB
E	8dB	4dB	16dB	24dB
F	4dB	24dB	8dB	16dB
G	16dB	24dB	4dB	8dB

した。図 2 に実験の様子を示す。

4.2 実験手順

増幅量 (ホワイトノイズ手法ではホワイトノイズ, 音声フィードバック手法では自分自身の声) が変化の中で, 文章を音読するという試行を行う。

音読してもらう文章は 1 分 30 秒ほどで読み切れるニュース記事 [11] である。なお, 文章の母音の数などの違いによって発話声量に差が生まれることを防ぐため, 同じ文章を利用した。また, 十分な発話時間を確保するため, 用意した文章を計 6 回繰り返し読んでもらった。6 回中最初の 2 回および最後の 2 回は増幅量を変化させず (0dB), 3 回目および 4 回目の増幅量を変えた。これを 1 試行として, ホワイトノイズ手法と音声フィードバック手法それぞれで 4 試行ごとに行い, 計 8 試行を行った。また, 各手法で適用する増幅量の試行順番をホワイトノイズ手法は表 1, 音声フィードバックは表 2 にそれぞれ示す。1 試行にかかった平均時間は約 10 分 36 秒であった。

さらに, 被験者の発話が独り言 (他者に伝わる必要のない, 声量調整の必然性が薄い発話) にならないように, 「ディスプレイの位置に人がいると想定して音読するようにしてください」と被験者に指示した。

4.3 評価項目

本実験の評価項目は下記の 4 つであり, 表 3 に示す。効力, 即効性, 持続性は図 3, 図 4 を用いて説明する。

汎用性 増幅量適用前から増幅量適用中において, 発話声量が増幅量適用前から増幅量適用中に変化した人数を求めることで評価する。

効力 効力は増幅量適用中の発話声量を求めることで評価する。増幅量適用中の効力 (増幅量適用前の平均発話声量と増幅量適用中の平均発話声量の差分) は, 被験者の発話声量減衰または発話声量増幅それぞれを評価する (図 3, 図 4)。

即効性 即効性は増幅量適用中, 何秒後に被験者の発話声量が増幅量適用前と比べて変化したかを求めることで評価する (図 3, 図 4)。

持続性 持続性は増幅量適用中, 被験者の発話声量が増幅量適用前と比べて変化した後, 何秒間持続するかを評価する (図 3, 図 4)。

4.4 実験結果

マイクから取得した音声のサンプリング周波数は 44.1kHz で, それを 4kHz に変換した。また量子化ビット数は 16bit とした。また, 被験者の音声信号から, 振幅の絶対値を取得し, 100 区間の移動平均を求めた (以下, 振幅絶対値平均と呼ぶ)。振幅は最大 1 から -1 まで変化し, その取得した振幅絶対値平均をデシベルに変換した。その後, 評価項目ごとに分析した。各評価項目の結果を図 5~図 10 に示す。

汎用性 ホワイトノイズ手法と音声フィードバック手法の 4 パターンの増幅量全てにおいて, 被験者 7 名の発話声量は増幅量適用前から増幅量適用中に変化したことから, 高い汎用性を確認できた。

効力 図 5 および図 6 にホワイトノイズ手法と音声

表 3 評価項目

評価項目	手法	説明
汎用性	ホワイトノイズ手法	ホワイトノイズの増幅量変化後、発話声量が変化した人数
	音声フィードバック手法	聴取声量変化後、発話声量が変化した人数
効力	ホワイトノイズ手法	ホワイトノイズの増幅量変化前の発話声量の平均とホワイトノイズの増幅量変化後の発話声量の差
	音声フィードバック手法	聴取声量変化前の発話声量の平均と聴取声量変化後の発話声量の差
即効性	ホワイトノイズ手法	ホワイトノイズの増幅量変化後、何秒後に発話声量が変化するか
	音声フィードバック手法	聴取声量変化後、何秒後に発話声量が変化するか
持続性	ホワイトノイズ手法	ホワイトノイズの増幅量変化後、発話声量が変化してから何秒間持続するか
	音声フィードバック手法	聴取声量変化後、発話声量が変化してから何秒間持続するか

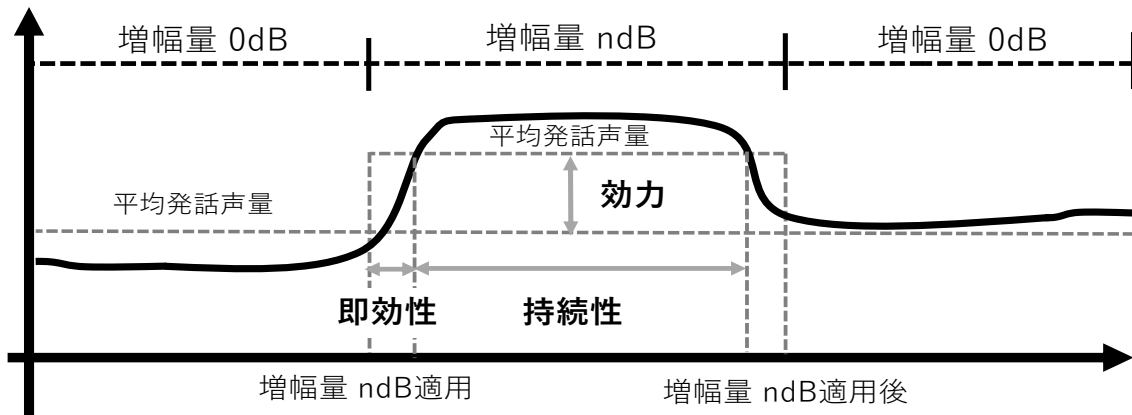


図 3 ホワイトノイズ手法における評価項目のモデル

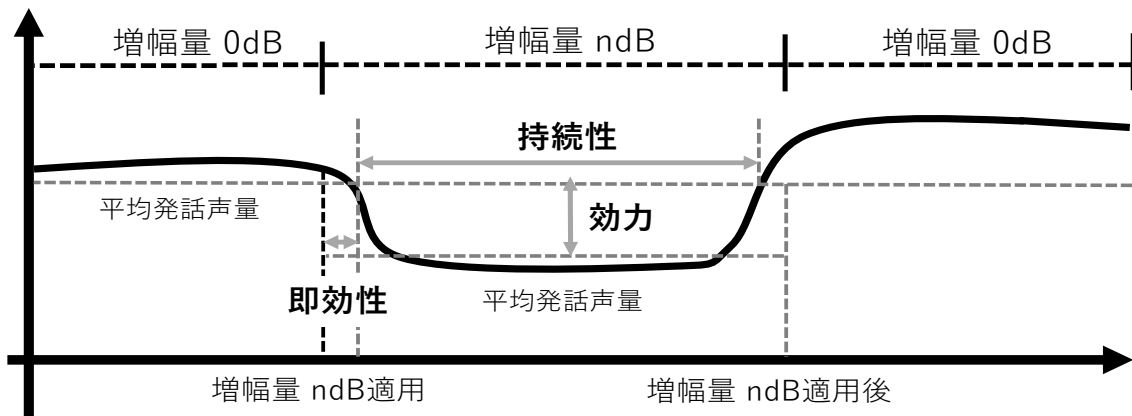


図 4 音声フィードバック手法における評価項目のモデル

フィードバック手法の効力の結果を示す。図 5 よりホワイトノイズの増幅量が大きくなっていくと、効力が大きくなることがわかった。すなわち、発話声量が大きくなっていくことが観測された。次に、ホワイトノイズ手法で有意水準 5% において、Welch の t 検定を適用した結果、4dB と 24dB ($t(7) = -2.39, p < .05$), 8dB と 24dB ($t(7) = -2.02, p < .05$) で有意差が観測された。

図 6 より音声フィードバックの増幅量が大きくなっていくと、効力が小さくなっていくことがわかった。すなわち、発話声量が小さくなっていくことが観測された。次に、音声フィードバック手法で有意水準 5% において、Welch の t 検定を適用した結果、4dB と 16dB ($t(11) = 3.41, p < .05$), 4dB と 24dB ($t(11) = 4.85, p < .05$) で有意差が観測され

た。

即効性 図 7 および図 8 に、ホワイトノイズ手法と音声フィードバック手法の即効性の結果を示す。図 7 から、ホワイトノイズの増幅量が大きくなるほど、発話声量の変化が早くなり、即効性が強くなる傾向が観測された。次に、提案手法で有意水準 5% において、Welch の t 検定を適用した結果、4dB と 24dB ($t(7) = 1.96, p < .05$) で有意差が観測された。

図 8 から、聴取声量の増幅量が大きくなるほど、発話声量の変化が早くなり、即効性が強くなる傾向が観測された。次に、提案手法で有意水準 5% において、Welch の t 検定を適用した結果、8dB と 24dB ($t(11) = 2.74, p < .05$) で

有意差が観測された。

持続性 図9および図10に、ホワイトノイズ手法と音声フィードバック手法の持続性の結果を示す。図9から、ホワイトノイズの増幅量が大きくなるほど、持続性が高くなるということが観測された。次に、ホワイトノイズ手法で有意水準5%において、Welchの t 検定を適用した結果、4dBと16dB($t(11) = -1.96, p < .05$)、4dBと24dB($t(11) = -2.29, p < .05$)で有意差が観測された。

図10から、音声フィードバックの増幅量が大きくなるほど、持続性が高くなる傾向が観測された。次に、音声フィードバック手法で有意水準5%において、Welchの t 検定を適用した結果、4dBと24dB($t(7) = -3.19, p < .05$)で有意差が観測された。

4.5 考察

効力 ホワイトノイズ手法については、すべての増幅量で増幅量適用中の効力は増幅量適用前の効力より大きくなるということが観測された(図5)。4dBの増幅量は、2.3dBの効力があり、増幅量適用中の効力は増幅量適用前の効力に比べて、約1.3倍発話声量が増幅した。8dBの増幅量では、2.9dBの効力があり、増幅量適用中の効力は増幅量適用前の効力に比べて、約1.4倍発話声量が増幅した。16dBの増幅量では、4dBの効力があり、増幅量適用中の効力は増幅量適用前の効力に比べて、1.6倍発話声量が増幅した。24dBの増幅量では、5.1dBの効力があり、増幅量適用中の効力は増幅量適用前の効力に比べて、1.8倍発話声量が増幅した。これらの結果より、全ての増幅量で被験者の声量の増幅を客観的に認識できることがわかった。したがって、ホワイトノイズによって被験者の声量増幅を制御できることがわかった。今後は、実際の生活環境の中で、ホワイトノイズ手法を適用するためのシステム構築やシステムの評価実験を行う。

音声フィードバック手法については、全ての増幅量で、増幅量適用中の効力は増幅量適用前の効力より小さくなるということが観測された(図6)。4dBの増幅量は、-2.1dBの効力があり、増幅量適用中の効力は増幅量適用前の効力に比べて、約0.8倍発話声量が減衰した。8dBの増幅量では、-3.4dBの効力があり、増幅量適用中の効力は増幅量適用前の効力に比べて、約0.67倍発話声量が減衰した。16dBの増幅量では、-4.1dBの効力があり、増幅量適用中の効力は増幅量適用前の効力に比べて、0.39倍発話声量が減衰した。24dBの増幅量では、-6.1dBの効力があり、増幅量適用中の効力は増幅量適用前の効力に比べて、0.25倍発話声量が減衰した。これらの結果より、全ての増幅量で声量の変化を客観的に認識できることがわかった。しかし、本実験の被験者は、全員落ち着いた環境で発話をしているが、周りが騒音環境であったり、ユーザ自身が興奮状態になっている場合

に音声フィードバック手法が適切に適用されるのかを今後確かめる必要がある。

即効性 即効性はどちらの手法も高く(図7、図8)、人は聴取音量の変化に対して素早く順応し、発話できるということがわかった。また、図7、図8よりホワイトノイズの方が全体的に即効性が高いという結果となった。すなわち、ホワイトノイズ手法が音声フィードバック手法より発話音量の変化が早いということが観測できた。この結果から、被験者はホワイトノイズを与えられると、雑音と素早く認識し、その雑音に合わせた音量に変化したと考えられる。また、音声フィードバックは被験者自身の声を用いて音量制御させる。したがって、音声フィードバックの増幅量が増幅した後に自身の声を聞き慣れていることから、若干の認識遅れが発現したと考えられる。これは、急激な音量変化より自然な音量変化の方が他者が違和感を感じにくいことから日常生活を考慮すると効果的であるといえる。

持続性 持続性はどちらの手法も最低でも約2分程度の持続性があることが観測できた(図9、図10)。ホワイトノイズと音声フィードバックは、それぞれ約4分間増幅量を適用する時間があった。その中で一番大きい増幅量である24dBは、どちらの手法も約3分発話音量を維持できたため、長い時間効果が持続するということが観測できた。音声フィードバック手法では、8dBの増幅量の方が16dBの増幅量より持続性が若干高いという結果となった。しかし、この2つの増幅量に有意差が観測みられない、かつ10秒程度の差であることから大きな問題点ではないと結論づけた。

5. 今後の展望

今後の展望として、ホワイトノイズ手法と音声フィードバック手法の応用実験を実施する。そして、実験環境としてプレゼンテーションを想定している。プレゼンテーションで本音量制御手法を用いる理由として、発表者が適切な音量で発表することで聴取者全員に正確な内容を伝えられるようにするためである。また、実験環境を構築しやすく、発表者の発話音量以外の音声がマイクに録音される可能性が低いからである。次に、このプレゼンテーション実験では、ホワイトノイズ手法を採用する。これは、ホワイトノイズ手法は声を大きくすることに特化した手法であるためプレゼンテーション実験に適していること、またノイズを聞かせることにより、発表者は普段発表する時よりリラックスした状態で発表ができると考えたためである。今後はプレゼンテーション実験に用いるための実験システムの構築、実験要項の作成などを行なっていく。

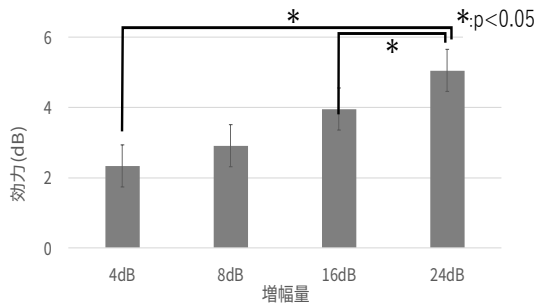


図 5 ホワイトノイズ手法の効力

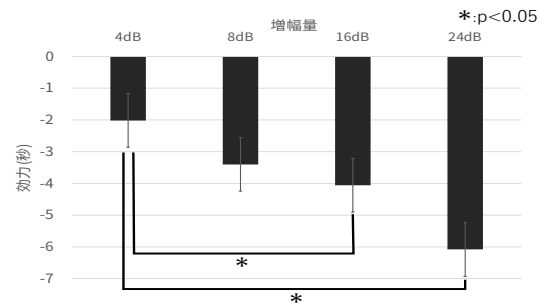


図 6 音声フィードバック手法の効力

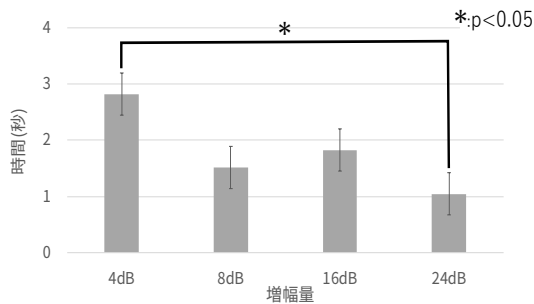


図 7 ホワイトノイズ手法の即効性

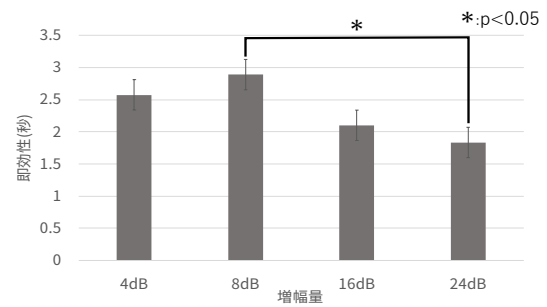


図 8 音声フィードバック手法の即効性

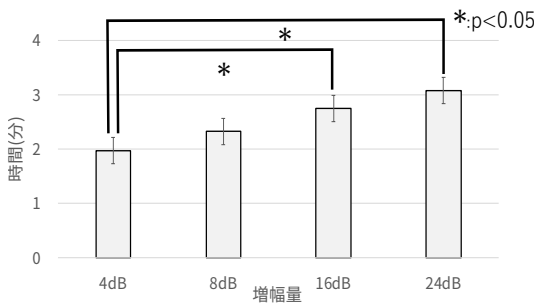


図 9 ホワイトノイズ手法の持続性

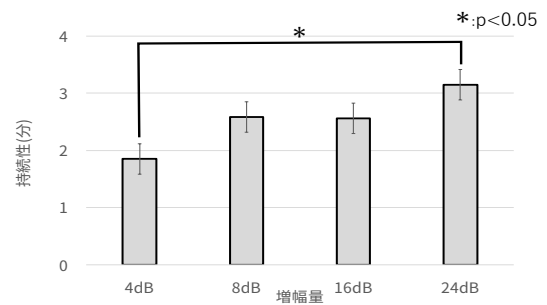


図 10 音声フィードバック手法の持続性

6. おわりに

本研究では、音量制御のためのホワイトノイズ手法と音声フィードバック手法を構築した。

評価実験は、4パターンの増幅量を設定し、ホワイトノイズ手法と音声フィードバック手法の増幅量によって発話音量がどのように変化するかを調査することを目的とし、効力、即効性、持続性、3つの評価項目をもとに実施した。実験の結果、効力に着目すると、ホワイトノイズ手法は、すべての増幅量で2dB以上の音量変化を観測でき、ホワイトノイズの増幅量が大きくなるほど発話音量も大きくなっていくことが観測できた。音声フィードバック手法も、すべての増幅量で2dB以上の音量変化を観測できた。また、増幅量が大きくなっていくと発話音量が小さくなるということが観測できた。即効性と持続性はどちらの手法も高い効果を観測できた。特に、持続性は全ての増幅量で2分以上

上の持続時間となり、どちらの手法も変化した音量を維持しやすいということがわかった。今後の展望としては、音量制御手法の応用実験としてホワイトノイズ手法を用いたプレゼンテーション実験を実施する。

謝辞

評価法について適切なアドバイスをしてくださった寺井あすか准教授(公立はこだて未来大学)に感謝いたします。本研究はJSPS 科研費 JP15K00279 の助成を受けたものです。

参考文献

- [1] Lane, H., Tranel, B.: The Lombard Sign and the Role of Hearing in Speech, Journal of Speech and Hearing Research, Vol. 14, pp. 677-709 (1971)
- [2] Denes, B. P., Pinson, N. E.: The Speech Chain: The Physics and Biology of Spoken Language, W. H. Freeman and Company (1993)

- [3] Hodoshima, N., Arai, T. and Kurisu, K.: Intelligibility of Speech Spoken in Noise and Reverberation, *In Proceedings of International Congress on Acoustics*, pp. 3632–3635 (2010).
- [4] Kogure, K., Yoshinaga, M., Suzuki, H. and Kitahara, T.: A Spoken Dialogue System for Noisy Environment, *In Proceedings of International Conference on Human-Computer Interaction-Posters' Extended Abstracts*, pp. 577–582 (2014).
- [5] 中村憲史, 片山拓也, 寺田 努, 塚本昌彦: 虚偽情報フィードバックを用いた生体情報の制御手法, *情報処理学会論文誌*, Vol. 54, No. 4, pp. 1433–1441, 2013.
- [6] Narumi, T., Ban, Y., Kajinami, T., Tanikawa, T. and Hirose, M.: Augmented Perception of Satiety: Controlling Food Consumption by Changing Apparent Size of Food with Augmented Reality, *In Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 109–118 (2012).
- [7] Rekimoto, J.: Traxion: A Tactile Interaction Device with Virtual Force Sensation, *In Proceedings of Association for Computing Machinery Symposium on User Interface Software and Technology*, pp. 427–431 (2013).
- [8] Watanabe, J., Hideyuki, Ando. and Maeda, T.: Shoe-shaped Interface for Inducing a Walking Cycle, *In Proceedings of International Conference on Augmented Tele-Existence*, pp. 30–34 (2005).
- [9] Suzuki, D., Takegawa, Y., Terada, T. and Tsukamoto, M.: A Heart Presentation System for Keeping Music Tempo in Live Performance, *In Proceedings of The Institute of Electrical and Electronics Engineers Global Conference on Consumer Electronics*, pp. 177–181 (2013).
- [10] 栗原一貴, 塚田浩二: SpeechJammer: 聴覚遅延フィードバックを利用した発話阻害の応用システム, *WISS インタラクティブシステムとソフトウェアに関するワークショップ論文集*, pp. 77–82 (2010).
- [11] 機械に責任を持たせることができるか, <<https://webronza.asahi.com/science/articles/2016111600001.html>> (参照 2017-12-12).