

音量制御のための音声フィードバック手法の検証

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が小さくなってしまいう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、Lombard 効果を活用し、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した。評価実験では、平常状態 (聴取音量を変化させない状態) から聴取音量を 12dB 分大きくすると、発話音量が小さくなり、逆に聴取音量を 12dB 分小さくすると、発話音量が大きくなるという結果が得られた。しかし、先行研究では、被験者に 12dB 分の聴取音量しか与えておらず、聴取音量の大きさと発話音量の大きさの関係性について調査されていない。そこで、本研究では 4 パターンの聴取音量を設定し、聴取音量の変化量によって、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、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。

キーワード: 発話音量, 聴取音量, 音声フィードバック

1. はじめに

近年、ウェアラブルコンピューティング環境における、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、災害など予想外の出来事が起きたときにパニック状態になってしまう状況、プレゼンテーションの際に緊張し声が小さくなってしまいう状況など自分自身の音量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。

筆者らの研究グループでは、Lombard 効果 [2] を活用し、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 [1] を提案した。評価実験では、平常状態 (聴取音量を変化させない状態) から聴取音量を 12dB 分大きくすると、発話音量が小さくなり、逆に聴取音量を 12dB 分小さくすると、発話音量が大きくなるという結果が得られた。しかし、先行研究 [1] では、被験者に 12dB 分の聴取音量しか与えておらず、聴取音量の大きさと発話音量の大きさの関係性について調査されていない。

そこで、本研究では 4 段階の聴取音量を設定し、聴取音量の増幅量と発話音量の関係性を調査することを目的とする。

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 [3] 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、聴覚フィードバックに特化した発声モデルとして Lombard 効果 [2] がある。例えば、我々は周囲の雑音が多い環境では、自然と声が大きくなる傾向があるが、このように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。Lombard 効果

¹ 公立はこだて未来大学
Futere University Hakodate

a) g2117040@fun.ac.jp

b) yoshi@fun.ac.jp

c) hirata@fun.ac.jp

についてはさまざまな研究結果があり、例えば、程島ら [5] 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。本研究では Lombard 効果において、特に自身の声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、VR (Virtual Reality) による視覚提示による食欲制御 [7]、振動モータによる振る舞い制御 [8] [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 が挙げられる [10]。Speech Jammer は発話の制御という点において本研究とは関わりが深い。SpeechJammer 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、不随意的かつ非知覚的にユーザの発話を制御するという点で異なる。

3. 設計

3.1 声量制御の原理

ユーザの視覚や聴覚で得られる情報を操作することで、ユーザの声量を変化させられると考えられる。聴覚的な情報による声量の変化の要因の1つとして環境音と自身の声の大きさの関係が考えられる。例えば、ロンバート効果 [2] により、ライブ会場や工事現場など、周りで大きな音が常に響き続けている場所で会話をしようとするときには、人は普段よりも声は大きくなる。逆に図書館など静まり返った場所においては、人は普段よりも声を小さくする。これらは周囲の環境音の大きさと自身の声の大きさを比較し、2つの音の大きさの差を一定に保つことで声量を適切にコントロールしているからである。上記の例を見ると、前者の場合は聴覚情報として得られる周囲の環境音が大きく、自身の通常時の声量と周囲の環境音とを比較し、自身の声量が小さいため、声が大きくなる。後者の場合はその逆で、周囲の環境音が小さく、通常時の声量と環境音とを比較し、自身の声量が大きいいため、声が小さくなる。

次に、視覚的なアプローチとして、遠近が関係していると考えられる。人は遠くにいる人物に対しては大きな声で話しかけるが、近くにいる人物に対しては小さな声で話しかける。これを利用して基本的な声量制御手法とシステム構成を考案する。

3.2 提案する声量制御手法

声量制御手法を提案するため、人はより大きな聴取声量を受け取ると発話声量は小さくなり、反対に受け取る聴取声量が小さくなると発話声量が大きくなるという仮説を立

てた。以下にその理由を示す。

3.2.1 声量増幅

ユーザの声量を増幅させるためには、「普段の聴取声量」より、「現在の聴取声量」を小さくすれば、現在の聴取声量を普段の聴取声量と同じにしようとし、結果、発話声量が大きくなると考えた。ユーザが発話したとき、聴取声量には骨導音と気導音の2種類あることが知られており、骨導経路の聴取声量を減衰させることは難しい。したがって、気導経路の聴取声量を減衰させるために、普段自分の耳で聞いている環境音の音量(聴覚環境音量)より、現在の聴覚環境音量を大きくすればよい。これは Lombard 効果 [3] により、ノイズを受け取ることで発話声量が大きくなることは証明されている。そこで、今回は、聴取声量が増幅した状態から、聴取声量を減衰させることで、発話声量は増幅すると考えた。

3.2.2 声量減衰

声量減衰においては、声量増幅を反転させればよい。すなわち、現在の聴取声量を増幅させれば、不随意的に声量が減衰すると考えた。

3.3 利用シーン

提案する声量制御手法は、ユーザの声量を不随意的かつ非知覚的に制御できる。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。提案する声量制御手法は、次のようなシーンでの利用が考えられる。

3.3.1 声を大きくする場合

就職活動やアルバイト、受験などでの面接において、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声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。また、面接の場合と同様で、適切な声量へと制御することで、聴取者に効果的にプレゼンテーションをすることができる。マイクなど声を大きくする装置もあるが、発話者の元の声が小さい場合や、装置の向きが適切でない場合には効果を得ることができない。一方、本研究の提案手法は、人の声量自体を制御する。

3.3.2 声を小さくする場合

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。本研究の提案する声量制御手法は、ユーザが不随意的に声量を制御するため、配慮を忘れていても雰囲気には水を差し、ユーザの意識に変化を与えることなく声量を抑えることが期待される。

4. 実験

4.1 実験目的

4 パターンの音声フィードバックの増幅量を与えた時の聴取音量と発話音量の関係を調査するための実験を実施した。被験者は大学生7名(男性6名, 女性1名)である。

4.2 調査項目

聴取音量と発話音量の調査項目として、以下の2つがある。

調査項目 1: 聴取音量の増幅量を 4 パターンで変えた時における発話音量の影響

竹川らの先行研究 [1] では、普段の聴取音量より、大きな聴取音量を与えたとき、普段の聴取音量と同じにしようとし、発話音量が小さくなることがわかった。しかし、被験者に 12dB 分の聴取音量しか与えておらず、聴取音量の大きさと発話音量の大きさに関係性があるかは検証されていない。よって、本実験では 4 パターンの聴取音量を設定し、聴取音量の増幅量が大きくなるほど、発話音量は小さくなることを調査する。

調査項目 2: 聴取音量を増幅した状態から、聴取音量の増幅がない状態に変えた時における発話音量の影響

竹川らの先行研究 [1] では、普段の聴取音量より大きな聴取音量を与えれば、発話音量が小さくなるという結果を得られた。しかし、聴取音量を増幅している状態から聴取音量の増幅がない状態(普段の聴取音量)に戻った時に、発話音量がどのように変化するか調査できていない。本実験では、聴取音量を増幅させた状態から、聴取音量を増幅させない状態にした場合における発話音量の変化について調査する。

4.3 実験システム

提案手法での実験を行うため、図 1 に示す実験システムを実装した。聴取音量の変化に伴う発話音量の変化を調べるため、マイクから出力された音声信号を分析する。被験者は防音室のディスプレイに表示された文章を発話する。その音声をマイクで取得し、パソコンで音声信号を録音し、オーディオインターフェースで音声信号を増幅し、編集済み音声信号として被験者のヘッドフォンへ出力する。これにより、被験者は自身の聴取音量が変化すると体感できる。

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、イヤフォンは iPhone の純正イヤフォンである EarPods を使用した。オーディオインター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 を使用し、聴取音量の変化は 4 パターンとした。実験の様子を図 2 に示す。

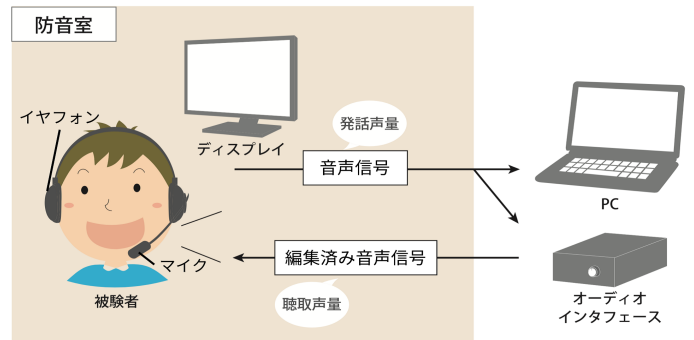


図 1 実験システム

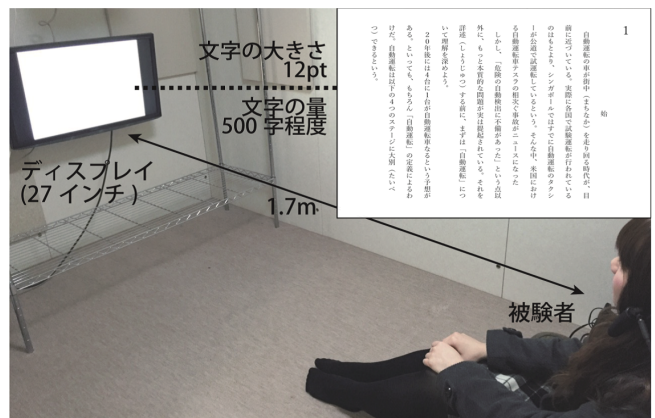


図 2 実験の様子

表 1 適用した聴取音量の増幅量

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

4.4 実験手順

実験の手順を以下に示す。提案手法では「聴取音量が変化しながら、文章を音読する」という試行を行う。被験者は 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。その際に、被験者にはイヤフォンとマイクロフォンを装着してもらい、聴取音量を変化させる。

音読文章は 1 分 30 秒ほどで読み切れるニュース記事 [11] である。同じ文章を計 6 回繰り返し読んでもらった。なお、文章の母音の数などの違いによって発話音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、6 回中最初の 2 回および最後の 2 回は聴取音量を変化させず (0dB)、3 回目および 4 回目の聴取音量を変えた。これを 1

表 2 評価項目

評価項目	説明
汎用性	聴取音量変化後、発話音量が変化した人数
効力	聴取音量増幅有、聴取音量増幅無 1、聴取音量増幅無 2、それぞれの平均発話音量の比較
即効性	聴取音量増幅有に変化した場合、聴取音量増幅無 2 に変化した場合、それぞれ何秒後に発話音量が変化するか
持続性	聴取音量増幅有に変化した場合、聴取音量増幅無 2 に変化した場合、それぞれの発話音量変化後の持続時間
副作用	内容の理解、4 パターンの聴取音量増幅それぞれに対する違和感についてのアンケート調査

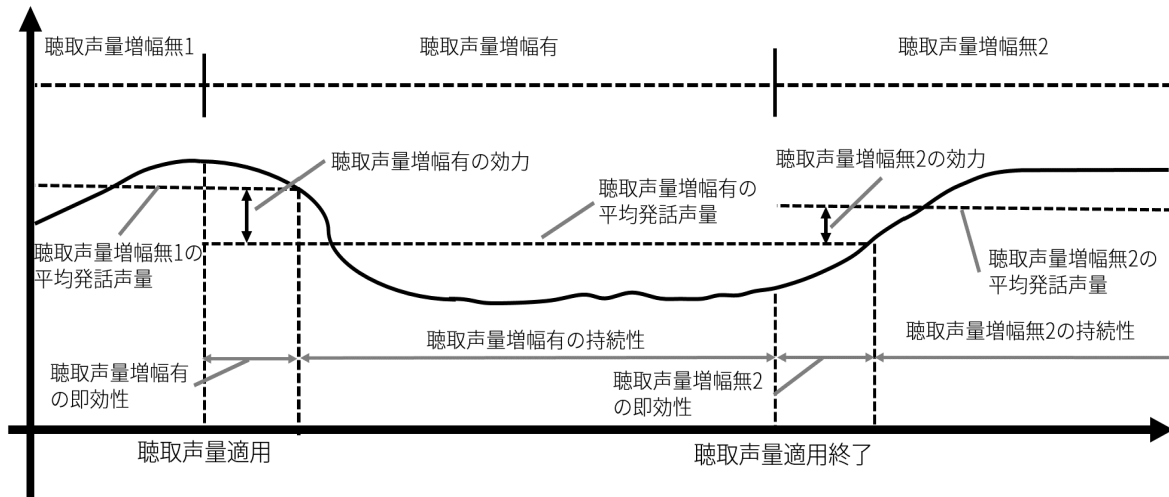


図 3 評価項目のモデル

試行として、試行ごとに聴取音量の増幅量を変え、計 4 試行を行った。1 試行にかかった平均時間は約 10 分 36 秒であった。また、表 1 のように被験者ごとに適用する聴取音量をランダムに変えた。さらに実験終了後、アンケートに回答してもらった。

4.5 被験者への指示

被験者の発話が独り言(他者に伝わる必要のない、音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。また、聴取音量を意識して発話してもらうため、「普段自分が聞いている自分自身の音量になるように音量を調整しながら発話してください」と被験者に指示した。

4.6 評価項目

本実験の評価項目は表 2 に示す下記の 5 つである。副作用以外の評価項目は図 3 を用いて説明する。

4.6.1 効力

効力は、聴取音量増幅無 1、聴取音量増幅有、聴取音量増幅無 2、それぞれの平均発話音量を比較することで得られる値である(図 3)。

聴取音量増幅有の効力は、聴取音量増幅無 1 の平均発話音量から聴取音量増幅有の平均発話音量を引いた値であ

る。また、聴取音量増幅無 2 の効力は、聴取音量増幅有の平均発話音量から聴取音量増幅無 2 の平均発話音量を引いた値である。

4.6.2 即効性

即効性は聴取音量増幅有に変化した場合、聴取音量増幅無 2 に変化した場合、それぞれ何秒後に発話音量が変化したかを表す値である。

聴取音量増幅有の即効性は、聴取音量増幅無 1 から聴取音量増幅有に変化してからの発話音量が、聴取音量増幅無 1 の平均発話音量を下回った時間である。また、聴取音量増幅無 2 の即効性は、聴取音量増幅有から聴取音量増幅無 2 に変化してからの発話音量が、聴取音量増幅有の平均発話音量を下回った時間である。

4.6.3 持続性

持続性は、聴取音量増幅有に変化した状況、聴取音量増幅無 2 に変化した状況、それぞれの発話音量変化後の持続時間を示す値である。

まず、聴取音量増幅有の持続性は、聴取音量増幅無 1 から聴取音量増幅有に変化してからの発話音量が、聴取音量増幅無 1 の平均発話音量を下回った地点から、平均発話音量を上回った地点までの持続時間を求める。次に、聴取音量増幅無 2 の持続性は、聴取音量増幅有から聴取音量増幅無 2 に変化してからの発話音量が、聴取音量増幅有の平均発話音量を上回った地点から、平均発話音量を下回った地

点までの持続時間を求める。

4.6.4 副作用

本論文での副作用とは、文章の内容理解、4パターンの聴取音量増幅それぞれに対する違和感を指し、アンケート調査により評価する。具体的には、以下の3点について、リッカート尺度に基づき、5パターンで評価した。以下にアンケートの内容を示す。

- (1) 内容理解：文章の内容を理解できたか (1:理解できない～5:理解できた)
- (2) 違和感：聞こえる自分の声に対して不満を感じるがあったか (1:感じなかった～5:感じた)

5. 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)。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。

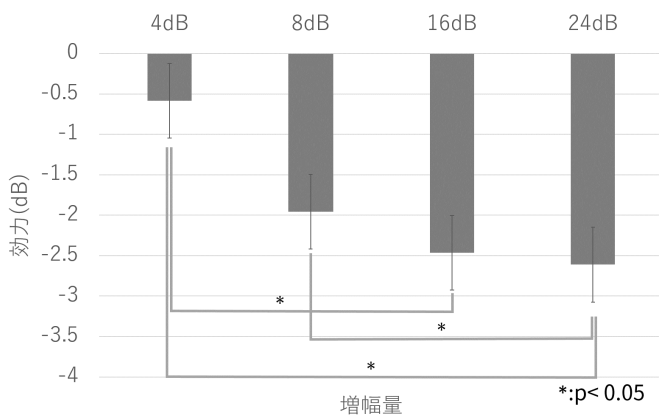


図4 聴取音量増幅有の発話音量平均 - 聴取音量増幅無1の発話音量平均

5.1 汎用性

提案手法において、聴取音量増幅変化後に発話音量が変化した人数を調べることで評価できる。

今回の実験では、被験者7名全ての発話音量は変化したことから、提案手法に汎用性はあったといえる。

5.2 効力

図4および図5効力の結果を示す。図4より聴取音量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、発話音量が小さくなっていった。また、聴取音量の増幅量ごとに Holm の多重分析を行った

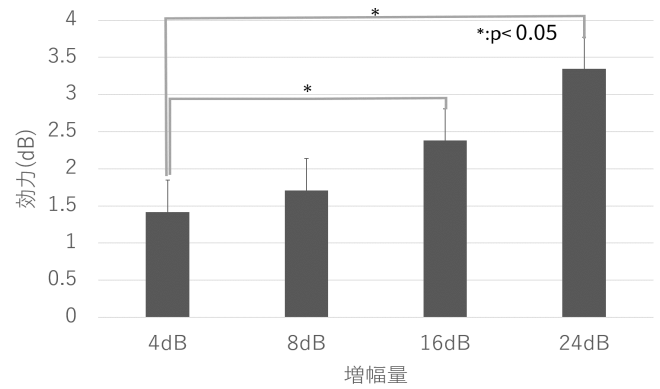


図5 聴取音量増幅無2の発話音量平均 - 聴取音量増幅有の発話音量平均

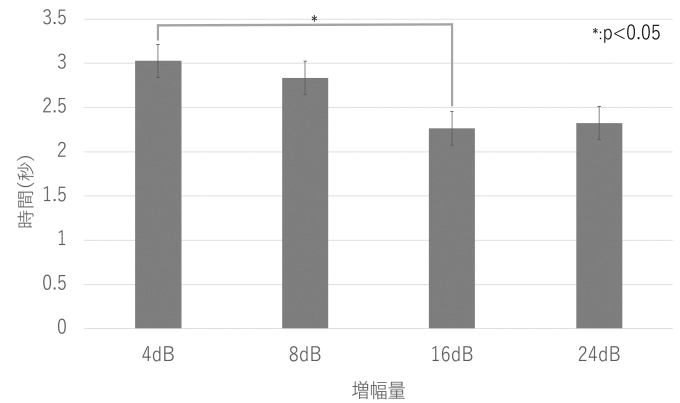


図6 聴取音量増幅有の即効性

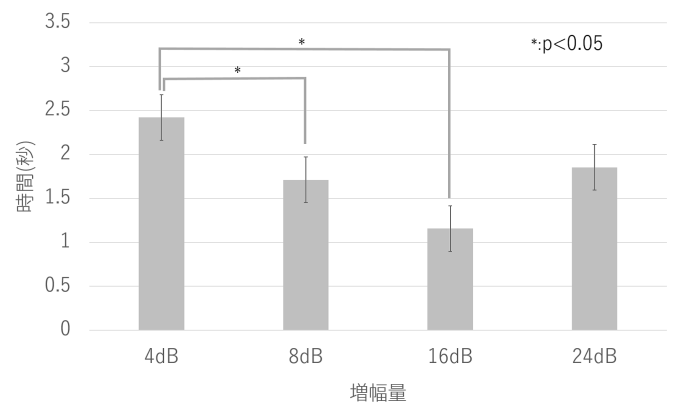


図7 聴取音量増幅無2の即効性

結果、4dBと16dB、4dBと24dB、8dBと24dBに有意な差が見られた ($p < 0.05$)。

次に、図5より聴取音量の増幅量が大きくなっていくと、効力が大きくなっていくことがわかった。すなわち、発話

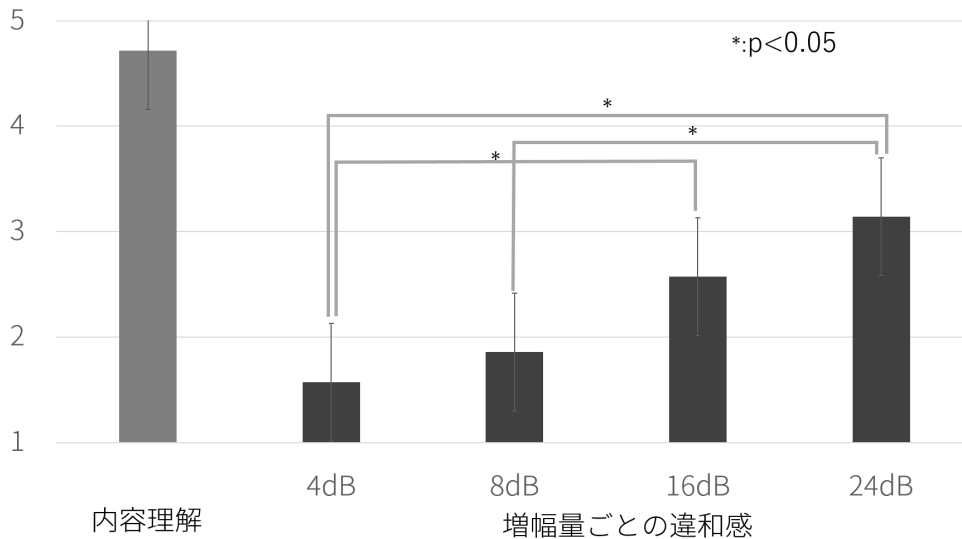


図 8 アンケート結果

声量が大きくなっていった。また、増幅量ごとに Holm の多重分析を行った結果、4dB と 16dB、4dB と 24dB に有意な差が見られた ($p < 0.05$)。

5.3 即効性

図 6 に、聴取音量増幅有の即効性を示す。図 7 に、聴取音量増幅無 2 の即効性を表す。

図 6 から、聴取音量の増幅量が大きくなるほど、即効性が高くなる傾向が見られた。また、Holm の多重分析を行った結果、4dB と 16dB で有意な差が見られた ($p < 0.05$)。次に、図 7) は、図 6) より、全体的に即効性が高い結果となったが、増幅量が大きくなるほど即効性が高くなるという結果は得られなかった。また、Holm の多重分析を行った結果、4dB と 8dB、4dB と 16dB に有意な差が見られた ($p < 0.05$)。

5.4 副作用

副作用は図 8 に表し、文章の内容理解、聴取音量増幅それぞれに対する違和感をそれぞれ表している。内容理解は 4.7、違和感は 4dB では 1.6、8dB の増幅量では 1.9、16dB では 2.6、24dB では 3.1 となった。また、聴取音量増幅それぞれに対する違和感のみ Holm の多重分析を行なった結果、4dB と 16dB、4dB と 24dB、8dB と 24dB でそれぞれ有意差が見られた ($p < 0.05$)。

5.5 考察

本実験結果から、聴取音量の増幅量を 4 パターンで変えた時における発話音量の影響としては、聴取音量の増幅量が大きくなっていくと、発話音量が小さくなっていくということがわかった (図 4)。また、聴取音量を増幅した状態

から、聴取音量の増幅がない状態に変えた時における発話音量の影響としては、聴取音量の増幅量が大きくなっていくと、発話音量が大きくなっていくということがわかった。すなわち、聴取音量増幅有で発話音量が小さくなくても、普段の聴取音量になった場合には、普段の発話音量に戻そうとし、発話音量を大きくすることがわかった (図 5)。

即効性は高く (図 6, 図 7)、人は聴取音量の変化に対して素早く順応し、発話音量を変化できるということがわかった。また、図 7 より、図 6) 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、聴取音量増幅有から聴取音量増幅無する場合の方が、発話音量の変化が早くなるということがわかった。加えて、図 6, 図 7) どちらも 16dB が 1 番即効性が高いという結果になった。しかし、この結果に対して、現在の実験データからでは考察が難しい。そこで、被験者の人数を増やし、表 1 で示す増幅量の順番の偏りをなくす必要がある。

副作用 (図 8) は、内容理解は非常に高く、文章の理解に関して聴取音量を変化させることは弊害がないと言える。また、違和感は増幅量が大きくなるごとに不満を感じる人もいたが、不満をととても感じるという結果ではないため、発話を障害するほどの増幅量ではないとわかった。

持続性は、竹川ら [1] の先行研究の結果と大きく異なっていることが分析の結果判明した。よって、本実験の持続性に関してのデータや評価方法に問題がある可能性を考慮し、データや評価方法を再検討し、持続性を改めて求める。

6. まとめ

筆者らの研究グループでは、Lombard 効果を活用し、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した。竹川ら [1] の先行研究では、被験者に 12dB 分の聴取音量しか与えておらず、聴取音量の大きさと発話音量の大きさの関係性について調査されていない。そこで、本研究では 4 パターンの聴取音量を設定し、聴取音量の変化量によって、発話音量がどのように変化するのか調査することを目的とし、汎用性・効力・即効性・持続性・副作用の評価項目をもとに実験を実施した。実験の結果、汎用性・効力・即効性・副作用の効果は高いという結果となった。

本研究で得た知見から、本研究の音量制御手法を適用することで、公共の場で声が大きくなってしまい周りに迷惑をかけてしまうなどの状況を回避できるとわかった。しかし、本実験ではユーザの発話音量を普段の発話音量より大きくさせるための聴取音量の増幅量は与えていない。すなわち、プレゼンテーションや騒音下での発話などの声を大きくすべき状況への対応ができていない。今後は、発話音量を大きくするための聴取音量の与え方を考案する。その調査が終わり次第、ウェアラブルシステムとして、システム設計と実装を行っていく。

参考文献

- [1] 竹川佳成, 平田圭二: 音量制御のための音声フィードバック手法の提案, 研究報告エンタテインメントコンピューティング (EC).
- [2] Lane,H. and Tranel,B. The Lombard Sign and the Role of Hearing in Speech, J. Speech Hear. Res, Vol. 14, pp.677-709, 1971.
- [3] Denes,P. and Pinson,E. The Speech Chain, 2nd Ed, New York : W.H Freeman and Co. 1993.
- [4] S, Hiroya. T. Mochida. M, Kashino.: Articulatory gestures, not auditory frequency resolution, determine formant frequency discrimination thresholds in vowels, Abstracts of The 29th MidWinter Meeting of Assoc. Res. Otolaryngol, 2006.
- [5] Hodoshima, N. Arai, T. Kurisu, K: Intelligibility of speech spoken in noise and reverberation , Proceedings of 20th International Congress on Acoustics (ICA2010), pp. 3632-3635, 2010.
- [6] Kogure, K. Yoshinaga, M. Suzuki, H. Kitahara, T: A Spoken Dialogue System for Noisy Environment, International Conference on Human-Computer Interaction(HCI2014) pp 577-582, 2014.
- [7] Narumi, T. Ban, Y. Kajinami, T. Tanikawa, T. Hirose, M.: Augmented perception of satiety: controlling food consumption by changing apparent size of food with augmented reality, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 109-118, 2012.
- [8] Rekimoto, J.: Traxion: A Tactile Interaction Device-with Virtual Force Sensation, Proceeding of Annual ACM Symposium on User Interface Software and Technology (UIST2013), pp. 427-432. 2013.
- [9] Junji, W. , Ando, H. Maeda, T. : Shoe-shaped Interface for Inducing a Walking Cycle,Proceedings of the 2005 international conference on Augmented tele-existence (ICAT2005), pp. 30-34, 2005.
- [10] Kurihara, K. Tsukada, K: SpeechJammer: A System Utilizing Artificial Speech Disturbance with Delayed Auditory Feedback, The 18th Workshop on Interactive Systems and Software, pp.77-82, 2010.
- [11] 機械に責任を持たせることができるか, 入手先 (<http://webronza.asahi.com/science/articles/2016111600001.html>) (参照 2017-12-12).