

HMD上のリップシンクアニメーションが 会話の聞き取りに与える影響の調査

磯山 直也¹ 寺田 努^{2,3} ロペズ ギョーム¹

概要：音声を用いた情報提示は、スピーカなどを用いて大衆へとアナウンスしたり、個人ではイヤホンなどの小型デバイスによりハンズフリーで利用したりでき、他の作業への影響が小さいことから、情報提示方式として有効な手段である。しかし騒音などの周辺状況の影響を受けやすく、ユーザが提示情報を聞き取れない場合が多い。単に音量を大きくするだけでは、他の周囲の音声が聞き取れなくなる可能性が高くなり、不必要に大きいことでユーザが不快感を得る。そこで本研究では、音声情報に視覚情報を加えることにより聞き取りやすくなる情報提示手法を提案する。提案手法では、HMD上に取得したい音声とリップシンクしたアニメーションを見せることで、音声を聞き取りやすくなることを狙う。本稿では、2種類の視覚情報が与える影響についての実験を行ない、提案手法において聞こえやすくなる被験者が多く見られることを確認した。

1. はじめに

駅構内でのアナウンス、パーティ会場での人との会話は、その音声情報を聞き取ることが重要であるが、周囲の音がうるさく必要な情報が聞き取れなかったり、他人との会話の成立が困難であったりする。また、音楽演奏のコンサートでは特定の楽器の音を中心に聴きたいこともある。これらのような状況は日常的に起こり、目的の音声以外の周囲音を聞くことも重要でもあるため、常時イヤホンを付けて、耳を塞いだ状態でそこから情報提示を受けるということは難しい。

ここで、人間の音声情報取得は、聞き取りたい音声・その他の周囲の音声、それらの音量のみによって決められるものではなく、聴覚以外の感覚に影響を受けている。特に視覚刺激による効果は複数研究されており、例えば、McGurkらが、マガーク効果 [1] について行なった研究では、視覚情報として唇を開いて発話する非唇音の/ga-ga/を提示し、聴覚情報として唇音の/ba-ba/を提示すると、被験者の多くが視覚情報でも聴覚情報でもない/da-da/という反応を示している。腹話術効果 [2], [3], [4] では、腹話術師の口元から発せられているはずの台詞が、腹話術師が手に持った人形から発せられているかのように聞こえることが確認されている。このように、視覚刺激は聴覚情報の取得に影響

を与えていることが分かる。これらの知見から、特定の視覚刺激を用いることで、それに関連する音声情報を取得しやすくなるのではないかと考えられる。

そこで本研究では、常時情報閲覧環境において視覚刺激を与えることによって、特定の聴覚情報に注目を集める手法を提案する。視覚刺激を与える手段についてであるが、近年、計算機の小型化・軽量化により、ウェアラブルコンピューティング環境が現実のものとなっており、頭部装着型ディスプレイ (HMD: Head Mounted Display) を装着すれば、視覚情報をいつでも提示できる。本提案手法では、HMDを用いて、ユーザが取得したい音声情報に関連した視覚情報を常時閲覧可能にし、音声情報取得をしやすくする。

本稿では、HMD上に取得したい音声とリップシンクしたアニメーションを見せることで、その音声が聞き取りやすくなることを狙い、それを確かめるために2種類の実験を行なう。実験はそれぞれ、(1) 特定の音声の内容を把握しやすくなったか、(2) 特定の音声の主観的に感じる音量に変化があったか、について調査する。(1)の実験では、複数の音声を同時に聞かせ、その後、特定の音声の内容に関する問いに答えさせる。(2)の実験では、複数の音声を同時に聞かせ、特定の音声の音量をその他の音声の音量と同じになるように被験者に調整させる。これらについて、リップシンクアニメーションの有無で、その正解数・音量の大きさの変化について評価を行なう。

以下、2章で関連研究を説明し、3章で提案システムに

¹ 青山学院大学

² 神戸大学

³ 科学技術振興機構さきがけ

ついて説明する。4章で視覚情報提示が与える影響に関する実験と考察を行ない、最後に5章で本研究をまとめる。

2. 関連研究

視覚などの聴覚以外の感覚や、人間がもつ知識や先入観によって聴覚の知覚が変化する効果がこれまでも研究されている。例えば、マガーク効果 [1] は、視覚情報として唇を開いて発話する非唇音の/ga-ga/を提示し、聴覚情報として唇音の/ba-ba/を提示すると、視覚情報でも聴覚情報でもない/da-da/聞こえるという効果である。腹話術効果 [2], [3], [4] は、実際に見えないスピーカなどから生じている音や音声、テレビ画面に映っている口や腹話術師の持っている人形などのような適切な動きをしているように見える映像などから生じているように誤って知覚される効果である。カクテルパーティー効果 [5], [6], [7] は、多くの人がそれぞれに雑談しているなかでも、自分が興味のある人の会話、自分の名前などは、自然と聞き取ることができる効果であり、これらの効果は視覚や人間があらかじめもつ知識や先入観によって聴覚の知覚が変化する例である。その他にも、長田らは、特定の音を聴くと特定の音が見えるという共感覚をもつ人がいることに着目し、それらが一般の人に対しても差異が無いことを確認し、人間の聴覚知覚が視覚と密接に関わっていることを示している [8]。SmartVoice[9]では、講演者の口の動きに合わせて音声データを出力することにより、講演者本人が直接話しているように見せることを可能にしている。

音声を適切に伝える手法や、主観的な印象を変化させる試みも多く行なわれている。矢高らは、ユーザ状況に応じて変化する、ユーザが感じる音量を主観的音量と定義し、聴覚情報の変化によって主観的音量がどう変化するかを調査し、ユーザの状況に合わせて適切な音量で音声提示を行なう手法を提案している [10]。佐久間らは、オーケストラなどのリアルタイムでの演奏の音量をユーザ側から制御するため、ビデオシースルー型 HMD を用いて、特定の楽器の演奏者のみ見えづらくすることにより主観的な音量が変化するかについて調べている [11]。Okazaki らは、触覚提示により主観的な音量の変化を狙い、聴覚刺激に触覚刺激を付加した場合とそうでない場合で主観的音量が変化するかを調査し、聴覚刺激単体と聴覚刺激に触覚刺激を付加した刺激を聞き比べた時、聴覚刺激の強度が物理的に等しくても、触覚刺激を付加した聴覚刺激のほうが主観的に強度が増すことを確認している [12]。本研究では、これらと同じように取得したい音声情報に関連した視覚情報を与えることにより、主観的な音量が変化するとともに、実際に情報の内容を把握しやすくなることを調査する。

3. 提案システム

3.1 想定環境

本提案システムでは、ユーザがコンピュータを身につけて生活するウェアラブルコンピューティング環境を想定している。ウェアラブルコンピューティング環境において情報提示を受けるには、イヤホン等の個人用のデバイスから音声情報提示を受けることが考えられるが、耳を塞ぐイヤホンは周囲の音が聞こえにくくなるため危険が伴う。アナウンス情報を文字として、HMD 上に視覚情報を提示することも考えられるが、文字を読むためには HMD を注視する必要があり、周囲の状況への注意が散漫となり、危険であることや、ある程度長文であれば文字送りをする必要があり、操作を伴うことなど、文字提示が情報提示手法としてふさわしくない状況がある。そこで本研究では、指向性の弱さなどから幅広く使用されている駅構内でのスピーカからのアナウンスや、他者との会話等といった環境からの音声情報提示を対象とし、その情報取得の補助を、単眼で閲覧可能であり日常的に使用できるタイプの HMD 上に視覚情報提示を行なうことで、特定の音声情報を取得しやすくなることを狙う。

利用シーンとしては、

- 駅構内などでの、特定の案内アナウンスを聞き取りやすくする。
 - 商業施設などでの、広告アナウンスを特定の人に聞かせたい際に、その音声情報の存在に気づかせる。
 - パーティ会場などの想像しい場所での、特定の人との会話を聞き取りやすくし、会話をスムーズにする。
- 等を想定する。

3.2 提案手法

本研究では、特定の音声情報を取得する際に、単眼式の HMD 上に、音声情報にリップシンクしたアニメーションを視覚情報として提示することにより、音声情報を聞き取りやすくすることを狙う。ここで、リップシンクアニメーションとは、表示された口が、特定の音声に同期して動き、その口が話しているように見えるアニメーションのことである。システムの利用イメージを図 1 に示す。HMD を使用することにより、ユーザはいつでもどこでもハンズフリーで視覚情報を閲覧することが可能である。リップシンクアニメーションを提示することにより、発声のタイミングが分かるだけでなく、その視覚刺激による影響から主観的な音量も変化し、実際に音声の内容を把握しやすくなることを狙っている。リップシンクアニメーションは、聞き取りたい音声情報がスピーカ等から出力されると同時に環境側から開始信号が HMD へと無線で送信されて開始されることや、特定の話者の音声聞き取りたい際には話者に

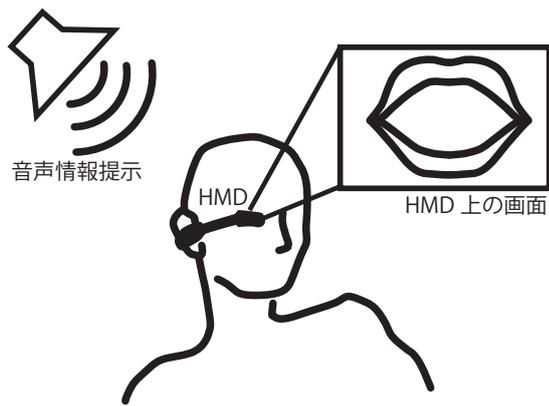


図 1 システムの利用イメージ

骨伝導マイク等が装着されており、その発声タイミングと音量に合わせてアニメーションが変化することを想定している。

4. 視覚情報提示が与える影響に関する実験

音声情報取得の際に、HMD 上のリップシンクしたアニメーションの提示が与える影響についてまだわかっていない。そこで、周囲で様々な音声情報が飛び交う中で特定の音声情報を聞き取りたい状況を想定し、リップシンクアニメーションを閲覧しながら音声情報取得させることにより、音声の内容が聞き取りやすくなるか、主観的な音量に変化はあるかについて実験を行なう。

4.1 内容把握に関する実験

被験者に複数の音声情報データを同時に聞かせ、特定の1つの音声情報に対しリップシンクアニメーションの有無に応じて、聞き取れた内容に違いがあるかを調べる。

4.1.1 実験方法

実験のために、3種類の音声情報データ(天気予報、プレゼンテーション、英語学習)が重なって同時に再生される音声データを作成し、全て異なる内容で5つ(それぞれデータ *a, b, c, d, e* とする)用意した。それぞれのデータは約1分半の長さで、ノーマライズすることで音量は差をなくし、全て女性の声で日本語である。被験者には天気予報の内容を聞き取らせるものとし、それぞれの音声データの天気予報の音声とリップシンクしたアニメーションを作成した。アニメーションの作成には Live2D 社 [13] の CubismAnimator を使用した。アニメーションのキャラクターにはスタンダードなモデルであるイブシロンを使用し、口元のみが映るアニメーション映像を作成した。被験者は、映像が無い状態、PC、または HMD 上でリップシンクアニメーションを閲覧、の3つの状態で音声情報を PC に接続したスピーカから聞き取る。実験の様子を図 2, 3, 4 に示す。実験に使用した HMD は Vuzix 社の M100 である。被験者には実験の前に、3つの音声と同時に聞こえて



図 2 実験中の様子(映像無し)



図 3 実験中の様子(PCでの閲覧)

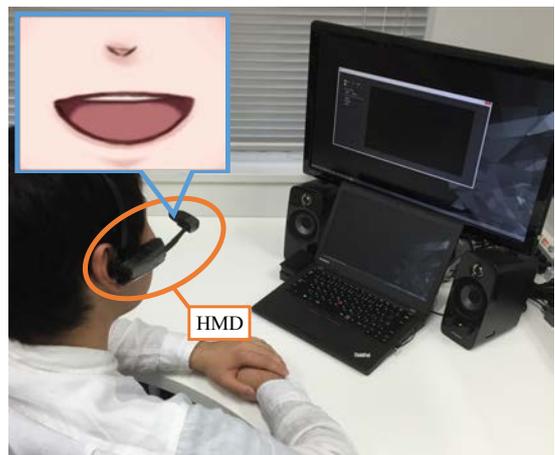


図 4 実験中の様子(HMDでの閲覧)

きて、聞き取り後に天気予報に関する問いを行なうので、天気予報の内容について聞き取るように指示をした。実験の際には、被験者は3つの内のどれかの状態で、まずデータ *e* を練習として聞き、その内容について回答させ、その後、他の4つの音声データのどれか1つを聞く。聞き取り後、天気予報に関する3択の問題を6問回答させ、その正誤により評価を行なう。選択問題であることから問いに勘

で答えることができるが、内容を聞き取れたかが重要であるため、勘で答えることなく聞き取れた問にだけ答えるよう回答前に指示した。1人の被験者につき、4つの音声データをそれぞれ3つの状態で聞かせる。各被験者12回の実験を行なうが、同じ日には1回しか実験を行なわないものとした。同じ音声データを3度聞くが、同じ音声データを聞く日は2週間以上空くようにし、内容を覚えていないように注意した。全被験者が音声データ a から順に聞くが、実験が進むにつれて実験に慣れることも考えられるため、被験者を3つのグループ(グループ α , β , γ)に分け、3つの状態の順序がグループによって異なるようにした。実験の順序を表1に示す。被験者は20代の男女22人である。

4.1.2 実験結果

表2に実験結果を示す。被験者 $A-G$ がグループ α 、被験者 $H-O$ がグループ β 、被験者 $P-V$ がグループ γ である。各値は、被験者ごとの音声データ $a-d$ の3つの状態でのそれぞれの正解数、誤回答数、未回答数の平均である。また、各状態での全被験者の平均正解数を図5に示す(エラーバーは標準誤差)。正解数について、各閲覧状態によって差があるかを調べるために分散分析を行なった。その結果、有意ではなかった ($F_{(2,21)} = 1.56, p > .05$)。しかし、映像無しとPCでの閲覧の結果について見てみると、11人の被験者が映像無しよりもPCでの閲覧の方が良い結果となり、2人の被験者のみが悪い結果となった。映像無しとHMDでの閲覧の結果について見てみると、9人の被験者については、映像無しよりも、PCでの閲覧、HMDでの閲覧両方の結果が共に良い結果となった。どちらも悪い結果となったのは、被験者 D のみであった。全ての被験者の正解数の平均を見ると、映像無し: 2.5, PCでの閲覧: 2.8, HMDでの閲覧: 2.7であった。音声データ毎の正解数の平均は、 a : 2.7, b : 2.8, c : 2.3, d : 2.7であった。

実験後には、被験者に自由記述を求めたが、その回答の多くが映像があった方がわかりやすいというものであった。発声のタイミングがわかることがその多くの理由に見られた。聞き取り時にメモ等を取ることを禁止していたため、聞き取ることはできるものの、答えを覚えていられないという意見も多くあった。

4.2 音量調整に関する実験

被験者に複数の音声情報データを同時に聞かせ、特定の1つの音声情報に対しリップシンクアニメーションの有無に応じて、主観的な音量に違いがあるかを調べる。

4.2.1 実験方法

4.1節の実験と同じ音声データ a, b, c, d, e 、リップシンクアニメーションを用いる。被験者は、映像が無い状態、PC、またはHMD上でリップシンクアニメーションを閲覧、の3つの状態で音声情報をPCに接続したスピーカから聞き取る。実験に使用したHMDは4.1節と同様に

Vuzix社のM100である。被験者は、音声聞き取り時に、キーボードの上下キーにより天気予報の音声の音量のみ調整することができ、音声聞き取り時に操作することで他の2つの音声と同じ音量になるように調整する。被験者が納得する音量になった時点で実験者に伝えさせ、実験終了となる。実験終了時の天気予報の音声の音量の違いにより評価を行なう。実験開始時の3つの音声の音量はノーマライズすることで同じ音量になっているが、被験者には伝えずに実験を行なった。どの音声データも実験開始時の音量は同じとした。被験者は1日に音声データ5種類を聞くが、最初に音声データ e を練習として聞き、その後でその他の4つの音声データを聞き、その4つの結果を評価に用いる。音声データは、毎回 e, a, b, c, d の順で聞くが、聞き取りの状態の順番が日によって異なるように、被験者を3つのグループ(グループ α, β, γ)に分けた。実験の順序を表3に示す。音声データ e は音声データ a と同じ状態で音量調整を行なう。被験者は20代の男女21名である。

4.2.2 実験結果

表4に実験結果を示す。被験者 $A-L, P-T$ は4.1節と同じ被験者である。実験開始時の音量を0として、1回キーボードを押下する毎に値が ± 1 ずつ変化するものとした際の実験終了時の値について、音声データ4つの平均を示している。右3列の結果は、被験者毎に12回の実験結果を正規化した際の平均である。表5に表4の結果について、映像無しとPCでの閲覧、映像無しとHMDでの閲覧に対してそれぞれ差を出した結果を示す。本結果の値がプラスになるのは、映像がある状態で閲覧することにより、主観的に大きな音量で聞こえているように感じているといえる。また、各状態での全被験者の調整された音量の平均値を図6に示す(エラーバーは標準誤差)。調整された音量について、各閲覧状態によって差があるかを調べるために分散分析を行なった。その結果、有意ではなかった ($F_{(2,20)} = 1.57, p > .05$)。しかし、映像無しとPCでの閲覧の結果について見てみると、14人の被験者が映像無しよりもPCでの閲覧の方が実験終了時の音量が小さい結果となり、7人の被験者が大きい結果となった。映像無しとHMDでの閲覧の結果について見てみると、被験者の11人が、映像無しとPCでの閲覧、映像無しとHMDでの閲覧、それぞれどちらも映像無しの状態の方が実験終了時の音量が大きい結果となった。被験者 A, B, D, F, H, K の6人については、映像無しとPCでの閲覧、映像無しとHMDでの閲覧、それぞれどちらも映像無しの状態の方が実験終了時の音量が小さかった。本実験は、HMDでの閲覧時に映像無しよりも音量が小さく設定されるという筆者らの狙いの結果はあまり得られなかった。この理由としては、被験者らがHMDを使用したことがほとんど無い状態で、実験中に音声に集中しながらも、キーボード操作を必要とさせたことで、音声に注意を向けきれなかったことが考えられる。

表 1 内容把握実験日程

グループ	日程	1	2	3	4	5	6	7	8	9	10	11	12
	音声	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>
α	視覚	無	PC	HMD									
β		PC	HMD	無									
γ		HMD	無	PC									

表 2 内容把握実験結果

	リップシンク映像								
	無			PC			HMD		
	正	誤	未	正	誤	未	正	誤	未
A	3.0	0.8	2.3	3.0	0.3	2.8	3.3	1.0	1.8
B	4.0	0.8	1.3	4.0	0.3	1.8	3.5	0.8	1.8
C	2.0	0.5	3.5	2.0	1.0	3.0	3.0	1.0	2.0
D	4.0	1.8	0.3	3.0	1.5	1.5	2.5	2.0	1.5
E	3.8	1.3	1.0	3.8	1.8	0.5	3.0	2.3	0.8
F	1.8	0.3	4.0	2.5	0.0	3.5	2.0	0.5	3.5
G	3.5	1.3	1.3	4.3	1.0	0.8	3.8	1.3	1.0
H	1.8	0.8	3.5	2.5	0.8	2.8	2.0	0.5	3.5
I	3.8	0.8	1.5	4.0	1.0	1.0	4.3	0.8	1.0
J	1.5	0.5	4.0	1.5	1.0	3.5	1.8	0.5	3.8
K	3.5	0.3	2.3	2.3	1.0	2.8	3.5	1.3	1.3
L	1.3	0.3	4.5	2.5	0.3	3.3	2.5	0.0	3.5
M	2.8	0.3	3.0	3.3	0.8	2.0	2.8	0.8	2.5
N	0.8	3.3	2.0	1.0	2.3	2.8	1.3	2.3	2.5
O	1.5	0.8	3.8	2.5	0.0	3.5	2.0	0.3	3.8
P	2.3	0.5	3.3	2.3	0.5	3.3	3.8	0.0	2.3
Q	3.5	1.0	1.5	3.5	0.8	1.8	3.0	0.5	2.5
R	2.3	1.8	2.0	2.5	0.5	3.0	1.3	0.8	4.0
S	2.0	0.5	3.5	2.5	0.8	2.8	2.3	0.0	3.8
T	3.3	1.3	1.5	3.8	0.3	2.0	3.3	1.3	1.5
U	2.8	1.0	2.3	4.3	0.3	1.5	3.8	0.8	1.5
V	0.3	0.0	5.8	0.0	0.3	5.8	0.5	0.3	5.3
平均	2.5	0.9	2.6	2.8	0.8	2.5	2.7	0.9	2.5

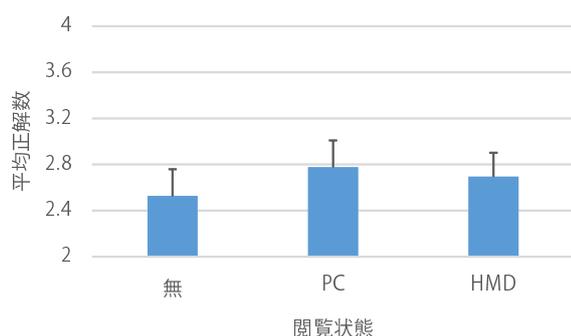


図 5 内容把握実験結果 (平均正解数)

キーボードの場所を探したり、操作に手間取ったりして、音声に集中できていない被験者が多く見られた。正規化した音量について、全ての被験者の平均を見ると、映像無し: 0.12, PCでの閲覧: -0.19, HMDでの閲覧: 0.067となり、映像無しが一番大きく、PCでの閲覧時が一番小さい結果であった。

4.3 考察

本稿では、特定の音声にリップシンクしたアニメーションをユーザに見せることで、音声を聞き取りやすくなるかについて、2種類の実験により調べた。天気予報の音声の内容について聞き取れているかについての実験(4.1節)、天気予報の音声の音量を他の音声の音量と同じに調整させる際に主観的な音量に変化があるかについての実験(4.2節)を行なった。本研究の目的について理想的なのは、4.1節の実験ではPCやHMDでの閲覧時の方が映像無しよりも正解数が多く、4.2節の実験ではPCやHMDでの閲覧時の方が映像無しよりも調整された音量が小さいことである。両方の実験を行なった被験者は17人であり、これらの被験者について見てみると、被験者C, I, J, L, P, S, Tの7人はこちらの理想と逆になることは無かった。被験者B, D, Kの3人はこちらの理想通りの結果となることは無かった。被験者RはPCでの閲覧では理想通りであるが、HMDでの閲覧では理想と逆の結果となった。これらの被験者について、4.2節の実験での正規化された音量について見てみると、被験者C, I, L, P, Sは映像無しと比べて映像有りの時に比較的大きく音量を小さく設定しており、被験者B, Kは映像無しと比べて映像有りの時に比較的大きく音量を大きく設定していた。このことから、映像が有ることにより、聞き取りについて良い影響を受けやすい人と、悪い影響を受けやすい人がいることがわかった。被験者B, D, Kについて、実験後に映像が有ると聞き取りづらくなる等の意見は無く、自覚無く影響を受けている可能性がある。

その他の意見として、HMDでの閲覧時に音声はスピーカから出力されているにもかかわらず、「HMDから音声は流れているように聞こえる」という意見が見られた。このことから、HMD上で表示する口の大きさを変化させることで、話者との距離感を変化させたり、音声アナウンスのスピーカがユーザにとって後ろからであったとしても、音源の向きを変化させて、音声の内容に集中しやすくなる可能性が考えられる。そういった可能性を考慮するためにも、今後は口の大きさを変化させたり、スピーカがユーザに正対していない状態にしたりして、実験を行なう必要がある。その他にも、今回は絵の口を表示したが実際の人の口を表示させたり、表情を変化させたりすることによる影響についても調査が必要である。

上記以外に今後実験して調査する必要があることは、リッ

表 3 音量調整実験日程

グループ	日程 音声	1				2				3			
		a	b	c	d	a	b	c	d	a	b	c	d
α	視覚	無	PC	HMD									
β		PC	HMD	無									
γ		HMD	無	PC									

表 4 音量調整実験結果

	リップシンク映像			正規化		
	無	PC	HMD	無	PC	HMD
A	33.0	33.5	34.3	-0.050	-0.0071	0.057
B	-3.3	0.8	1.8	-0.69	0.23	0.46
C	21.8	10.8	18.3	0.49	-0.62	0.13
D	0.5	1.3	1.3	-0.057	0.028	0.028
E	13.3	12.0	19.5	-0.22	-0.39	0.61
F	12.5	15.0	23.0	-0.50	-0.21	0.71
G	-4.5	-3.7	-8.3	-0.12	0.28	-0.16
H	0.8	6.8	11.8	-0.72	0.042	0.68
I	6.8	2.0	2.3	0.52	-0.28	-0.24
J	15.8	14.8	14.0	0.11	-0.010	-0.10
K	0.3	4.3	6.0	-0.88	0.20	0.67
L	7.8	3.8	-1.0	0.78	0.046	-0.83
P	0.5	-5.5	-3.3	0.94	-0.79	-0.14
Q	1.3	0.8	6.8	-0.24	-0.31	0.56
R	-0.5	-3.0	1.3	0.034	-0.30	0.27
S	0.8	-11.3	-11.8	1.09	-0.51	-0.58
T	12.3	10.3	11.0	0.15	-0.13	-0.023
W	2.3	-2.3	-1.0	0.73	-0.54	-0.19
X	2.5	1.3	1.8	0.65	-0.57	-0.081
Y	10.3	3.5	4.5	0.40	-0.25	-0.15
Z	54.3	53.8	48.3	0.16	0.12	-0.28
平均	9.0	7.1	8.6	0.12	-0.19	0.07

表 5 音量調整実験結果の差

	音量の差		正規化	
	無-PC	無-HMD	無-PC	無-HMD
A	-0.5	-1.3	-0.043	-0.11
B	-4.1	-5.1	-0.92	-1.15
C	11	3.5	1.11	0.36
D	-0.8	-0.8	-0.085	-0.085
E	1.3	-6.2	0.17	-0.83
F	-2.5	-10.5	-0.29	-1.21
G	-0.8	3.8	-0.4	0.04
H	-6	-11	-0.762	-1.4
I	4.8	4.5	0.8	0.76
J	1	1.8	0.12	0.21
K	-4	-5.7	-1.08	-1.55
L	4	8.8	0.73	1.61
P	6	3.8	1.73	1.08
Q	0.5	-5.5	0.07	-0.8
R	2.5	-1.8	0.33	-0.24
S	12.1	12.6	1.6	1.67
T	2	1.3	0.28	0.17
W	4.6	3.3	1.27	0.92
X	1.2	0.7	1.22	0.73
Y	6.8	5.8	0.65	0.55
Z	0.5	6	0.04	0.44

プシクスのズレについてである。リップシンクのズレの違和感に関する研究は数多く行なわれているが [14], [15], [16], HMD により常時ズレた状態での閲覧環境における実験は筆者らの知る限り存在しない。ズレに慣れてしまうことにより、実際の日常での会話に支障をきたす可能性がある。リップシンクのズレた状態での閲覧を長時間に渡って行なわせた際の影響について調べることは重要である。本稿での実験は、PC での閲覧と HMD での閲覧をどちらも行なったが、今後 HMD を用いることで常時情報閲覧環境というこれまでに無かった状況での生活が考えられるため、PC での閲覧との違いについて考えていく必要がある。

5. まとめ

本稿では、特定の音声情報を聞き取りたい際に、その音声にリップシンクしたアニメーションを HMD を用いてユーザに見せることで、音声の聞き取りやすくなることを狙った手法について提案し、その有効性について、2種類の実験により調べた。1つ目の実験では、特定の音声の内容を聞き取りやすくなるかについて調べ、映像が無い場合

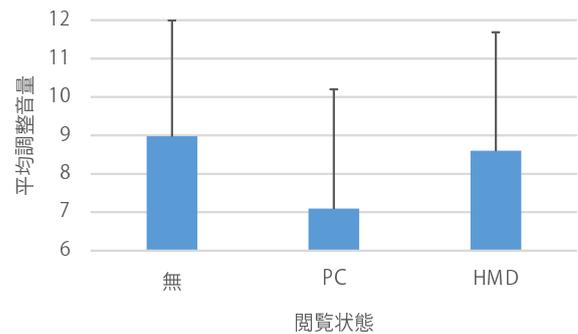


図 6 音量調整実験結果 (平均調整音量)

と比べて、PC 上でリップシンクアニメーションを見ながら聞くことで、聞き取りやすくなる被験者が多く見られた。2つ目の実験では、特定の音声について主観的な音量が変化するかについて調べ、映像が無い場合と比べて、PC 上で映像を見ながら聞くことで、音量が大きく聞こえる被験者が多く見られた。HMD を通じて見るることについては有意差が見られなかったが、被験者が HMD に慣れていないことも原因として考えられるため、今後さらに評価実験

を進めていく。音声アナウンスに合わせてアニメーションを再生させるシステムのプラットフォームについてや、特定の人との会話時に、相手の発声に合わせてアニメーションを動かす手法についても今後検討していく。

謝辞

本研究の一部は、科学技術振興機構戦略的創造研究推進事業(さきがけ)および文部科学省科学研究費補助金挑戦的萌芽研究(25540084)によるものである。ここに記して謝意を表す。

参考文献

- [1] M. Harry and M. John: Hearing Lips and Seeing Voices, *Nature*, Vol. 264, Issue. 5588, pp. 746–748 (1976).
- [2] R. B. Welch and D. H. Warren: Immediate Perceptual Response to Intersensory Discrepancy, *Psychological Bull.*, Vol. 88, No. 3, pp. 638–667 (1980).
- [3] J. Vroomen and B. D. Gelder: Perceptual Effects of Crossmodal Stimulation: Ventriloquism and the Freezing Phenomenon, *The Handbook of Multisensory Processes MIT Press*, Vol. 3, Issue. 1, pp. 1–23 (2004).
- [4] 木村真弘, 梶井 浩, 高橋 誠, 山本克之: 周辺視野における腹話術効果, *日本バーチャルリアリティ学会論文誌*, Vol. 4, No. 1, pp. 253–260 (1999).
- [5] A. W. Bronkhorst: The Cocktail Party Phenomenon: A Review of Research on Speech Intelligibility in Multiple-Talker Conditions, *Journal of Acta Acustica united with Acustica*, Vol. 86, No. 1, pp. 117–128 (2000).
- [6] E. C. Cherry: Some Experiments on the Recognition of Speech, with One and with Two Ears, *Journal of Acoustical Society of America*, Vol. 25, pp. 975–979 (1953).
- [7] E. C. Cherry and W. K. Taylor: Some Further Experiments upon the Recognition of Speech, with One and with Two Ears, *Journal of Acoustical Society of America*, Vol. 26, pp. 554–559 (1954).
- [8] 長田典子, 岩井大輔, 津田 学, 和氣早苗, 井口征士: 音と色のノンバーバルマッピング: 色聴保持者のマッピング抽出とその応用(ヒューマンコミュニケーション), *電子情報通信学会論文誌*, Vol. 86, No. 11, pp. 1219–1230 (2003).
- [9] X. Li and J. Rekimoto: SmartVoice: A Presentation Support System for Overcoming the Language Barriers, *Proc. of the SIGCHI conference on Human Factors in Computing Systems (CHI 2014)*, pp. 1563–1570 (2014).
- [10] 矢高真一, 田中宏平, 寺田 努, 塚本昌彦, 西尾章治郎: ウェアラブルコンピューティングのための状況依存音声情報提示手法, *情報処理学会論文誌*, Vol. 51, No. 12, pp. 2384–2395 (2010).
- [11] 佐久間一平, 寺田 努, 塚本昌彦: 視覚効果を用いた主観的音量の制御システムの設計と実装, *エンタテインメントコンピューティングシンポジウム 2015 (EC 2015)*, pp. 357–364 (2015).
- [12] R. Okazaki, T. Hachisu, M. Sato, S. Fukushima, V. Hayward, and H. Kajimoto: Judged Consonance of Tactile and Auditory Frequencies. *Proc. of the IEEE World Haptics Conference*, pp. 663–666 (2013).
- [13] Live2D: <http://www.live2d.com/>.
- [14] W. Fujisaki, S. Shimojo, M. Kashino, and N. Nishida: Recalibration of Audiovisual Simultaneity, *Nature Neuroscience*, Vol. 7, No. 7, pp. 773–778 (2004).
- [15] C. Spence, R. Baddeley, M. Zampini, R. James, and

- D. I. Shore: Multisensory Temporal Order Judgments: When Two Locations are Better than One, *Perception & Psychophysics*, Vol. 65, pp. 318–328 (2003).
- [16] M. Zampini, D. I. Shore, and C. Spence: Audiovisual Temporal Order judgment, *Experimental Brain Research*, Vol. 152, No. 2, pp. 198–210 (2003).