

令和 6 年度
修士学位論文

VR アバターの視覚的特徴や認知負荷が 視聴覚統合に基づく音源定位に及ぼす影響

Effect of Visual Features of VR Avatars and
Cognitive Load on Audiovisual Integration-Based
Sound Localization

1275103 末延 洸龍

指導教員 繁榎 博昭

2025 年 2 月 28 日

高知工科大学大学院 工学研究科 基盤工学専攻
情報学コース

要 旨

VR アバターの視覚的特徴や認知負荷が 視聴覚統合に基づく音源定位に及ぼす影響

末延 洗龍

本研究では、アバターの音声の音源定位において、男女の違いや口の動きによるアバターの視覚的特徴の影響及び視覚・聴覚による認知負荷の影響について検討した。実験1では、男女アバターと男女の声を使用し、音声の音源定位において、その声と同性・異性のアバター位置の影響や音声と同期してアバターの口が動くことによる影響を検討した。実験の結果、声とアバターの性別の一致・不一致によらず、片方のアバターのみ口が動く条件において、口が動くアバター方向に音源位置がずれて知覚されることが示された。実験2では視覚・聴覚による認知負荷課題を行いながら、アバターに対する音源定位に影響を及ぼすかについて検討した。実験の結果、マガーク効果を用いた先行研究で見られた多感覚情報の影響の減衰は見られず、認知負荷条件の効果は生じなかった。

キーワード VR, アバター, 視聴覚統合, 認知負荷, 音源定位

Abstract

Effect of Visual Features of VR Avatars and Cognitive Load on Audiovisual Integration-Based Sound Localization

Kouryu Suenobu

In this study, we examined the effects of the visual characteristics of avatars due to the difference between male and female avatars and the movement of their mouths, as well as the effects of visual and auditory cognitive load on the sound source localization of the avatar's voices. In Experiment 1, we used male and female avatars and voices to examine the effects of the position of the avatar of the same and opposite gender as well as the effect of synchronous avatar mouth movements with the voice. Results showed that regardless of whether the voice and avatar gender matched or not, when only one avatar's mouth moved, the perceived sound source location was biased toward the direction of the moving mouth. In Experiment 2, we examined the effects of visual and auditory cognitive load on sound source localization to avatars. Results showed no attenuation of multisensory integration effects as observed in previous studies using the McGurk effect, indicating that cognitive load conditions did not affect sound source localization.

key words Virtual Reality, Avatar, Audiovisual Integration, Cognitive Load, Sound Localization

目次

第1章	はじめに	1
1.1	背景	1
1.2	音源定位	2
1.3	腹話術効果	3
1.4	バーチャルリアリティ (VR)	5
1.5	認知負荷	6
第2章	実験1	7
2.1	目的	7
2.2	実験参加者	7
2.3	実験装置	7
2.3.1	使用ソフトウェア	7
2.3.2	視覚刺激呈示装置	8
2.3.3	聴覚刺激呈示装置	8
2.3.4	コントローラー	10
2.4	呈示刺激	10
2.4.1	アバター	10
2.4.2	聴覚刺激	11
2.4.3	VR画面上の実験環境	12
2.5	現実空間上の実験環境	13
2.6	実験条件	15
2.7	実験手続き	16
2.8	アンケート	17
2.9	アンケート結果	18

目次

2.10	分析方法	18
2.11	結果	19
2.12	考察	21
第3章	実験2	23
3.1	目的	23
3.2	実験参加者	23
3.3	実験装置	23
3.3.1	使用ソフトウェア	23
3.3.2	視覚刺激呈示装置	24
3.3.3	聴覚刺激呈示装置	24
3.3.4	コントローラー	26
3.4	呈示刺激	26
3.4.1	アバター	26
3.4.2	聴覚刺激	28
3.4.3	視覚負荷課題における視覚刺激	28
3.4.4	VR画面上の実験環境	29
3.5	現実空間上の実験環境	29
3.6	実験条件	31
3.7	実験手続き	32
3.8	アンケート	34
3.9	アンケート結果	34
3.10	負荷課題の正答率	35
3.11	分析方法	35
3.12	結果	36
3.13	考察	37

目次

第 4 章	まとめ	39
	謝辞	40
	参考文献	41
付録 A	音声内容	44
付録 B	音声内容	45
付録 C	視覚負荷課題の単語内容	47
付録 D	実験 1 のアンケート項目	48
	D.0.1 このアバターは男性に見えますか？	48
	D.0.2 このアバターは女性に見えますか？	48
	D.0.3 この声は男性に聞こえますか？	49
	D.0.4 この声は女性に聞こえますか？	49
付録 E	実験 2 のアンケート項目	50
	E.0.1 課題の難易度は難しかったか	50
付録 F	実験 1 の分析方法の具体例	51

目次

1.1	ILD の例	3
1.2	ITD の例	3
1.3	腹話術師と人形	5
2.1	HMD	8
2.2	単一指向性の小型スピーカー	9
2.3	スピーカーセレクタ	9
2.4	ヘッドホンアンプ	9
2.5	コントローラー	10
2.6	男性アバター	11
2.7	女性アバター	11
2.8	呈示されたアバター位置	11
2.9	単一指向性のコンデンサーマイク	12
2.10	VR 環境	13
2.11	注視点位置	13
2.12	実験風景	14
2.13	スピーカー配置	14
2.14	スピーカーを隠した状態の風景	15
2.15	聴覚刺激呈示条件 (deg)	16
2.16	実験開始時の VR 空間	17
2.17	逆方向への回転時の VR 空間	17
2.18	回転停止時の VR 空間	17
2.19	アバター呈示時の VR 空間	17
2.20	音源位置回答用のバー	17

図目次

2.21	アバターに対する呈示された音源位置	19
2.22	アバターの口が動く条件ごとの知覚された音源位置	20
3.1	HMD	24
3.2	単一指向性の小型スピーカー	25
3.3	スピーカーセレクタ	25
3.4	ヘッドホンアンプ	25
3.5	コントローラー	26
3.6	男性アバター	27
3.7	参加者正面から+5 deg のアバター位置	27
3.8	参加者正面から-5 deg のアバター位置	27
3.9	参加者正面のアバター位置	27
3.10	単一指向性のコンデンサーマイク	28
3.11	視覚負荷条件における単語呈示位置	29
3.12	VR 環境	29
3.13	実験風景	30
3.14	スピーカー配置	31
3.15	スピーカーを隠した状態の風景	31
3.16	実験開始時の VR 空間	33
3.17	逆方向への回転時の VR 空間	33
3.18	回転停止時の VR 空間	33
3.19	アバター呈示時の VR 空間	33
3.20	音源位置回答用のバー	33
3.21	視覚負荷条件におけるアバター呈示時の VR 空間	33
3.22	予測される結果図	36
3.23	負荷条件ごとの知覚された音源位置	37

図目次

B.1	母音が「あ」の場合の口の形	45
B.2	母音が「い」の場合の口の形	45
B.3	母音が「う」の場合の口の形	45
B.4	母音が「え」の場合の口の形	45
B.5	母音が「お」の場合の口の形	46
F.1	右側が男性かつ男性の声が呈示されている場合	51
F.2	右側が男性かつ女性の声が呈示されている場合	51
F.3	右側が女性かつ男性の声が呈示されている場合	52
F.4	右側が女性かつ女性の声が呈示されている場合	52

表目次

3.1 実験2 アンケート結果	34
3.2 n-back 課題の正答率	35

第 1 章

はじめに

1.1 背景

近年，コロナ禍を端緒としてリモート形式での会議が一般的になった．そのなかでも VR 環境を用いたリモート会議は，より対面の会議に近い状態で行うことができるため注目されている [1]．会議では複数のアバターが同時に呈示され，かつその音源位置は必ずしも正確に呈示されるとは限らない．そのため，複数のアバターが呈示されている状態での音源定位の特性を明らかにすることは重要である．音源定位における視聴覚統合の特性としてよく知られている現象に腹話術効果がある．腹話術効果とは，視覚刺激と音源位置が異なる位置に呈示されている場合に，音源の位置が視覚刺激位置に知覚される現象である [2][3]．腹話術効果は人の顔と人の声のように，視覚刺激と音声の意味的，文脈的に一致する場合，その効果が大きくなることが報告されている [4]．アバターを用いた先行研究においても，音声の音源位置がアバター位置に向かってシフトすることが報告されている [5]．しかし，先行研究では，複数のアバターが同時に呈示されている場合において，男女の声質や音声と同期した口の動きによる一致・不一致が音源定位に及ぼす影響は明らかにされていない．そこで実験 1 では，男女のアバターを同時に呈示し，男女の見た目の一致・不一致及び音声と同期したアバターの口の動作の一致・不一致を操作し，どの要因がより音源定位に影響を与えるか検討した．

また，会議中は，会話内容やホワイトボード等にかかれた内容を記憶したり，メモを取ったりするなど，会話以外の負荷のかかるタスクを行うことも多い．そのため，音源定位とは異なる課題を同時に行うことが，アバターの音源定位に及ぼす影響を明らかにすることも重

1.2 音源定位

要である。腹話術効果の他に視覚情報が音声に影響を与える効果としてマガーク効果がある。マガーク効果とは「ga」と発音している映像に「ba」と発音している音声を組み合わせて呈示すると、「da」として聞こえるという、映像が音声の知覚を変化させる現象である [6]。このマガーク効果に対して負荷課題を行う状態での影響について先行研究で検討されており、マガーク効果の音声回答課題の前後に 8 桁の数字の記憶課題を行った場合、視聴覚の統合が減少することが報告されており [7]、音の判断とは別に触覚課題を行った状態においても、視聴覚の統合が減少し、視覚による影響が減少することが報告されている [8]。しかし、視覚情報が音声知覚に影響を与える現象であるマガーク効果で見られた、認知負荷が生じたことによる視覚による影響の減衰が、腹話術効果においても見られるかは明らかになっていない。そこで、実験 2 では視覚的な負荷課題、聴覚的な負荷課題がそれぞれアバターに対する音源定位に影響を与えるか検討した。

1.2 音源定位

私たち人間は、音が発せられたときにその音が発生した距離や方位を特定することが可能である。このような音の 3 次元的な位置を把握することを音源定位という。音源定位には音が持つ様々な情報が手がかりとして音源位置を把握している。水平方向への音源定位の手がかりとして、両耳間音圧差 (interaural level difference: ILD)、両耳間時間差 (interaural time difference: ITD) がある。ILD とは、左右それぞれの耳で感じられる音の音圧が異なることを手がかりとして音源位置を把握することである (図 1.1) [9]。音声が発せられたのち、耳に音が到達する際、左右どちらかの耳に到達する音は人の頭部によって遮蔽される。これによって、左右の耳に到達する音の音圧に差が生じ、音圧の違いを手がかりとして音源位置を特定している。ITD とは、左右それぞれの耳に音が到達する時間が異なることを手がかりとして音源位置を把握することである (図 1.2) [10]。人の耳は左右に離れていることから、音が発せられた際、それぞれの耳に音が到達する時間が異なる。音が到達する時間の違いを手がかりとして、水平方向の音源位置を推定することが可能になっている。このよ

1.3 腹話術効果

うに ILD と ITD の 2 種類の両耳情報により、音源位置を知覚する事を音源定位の二元説と呼ぶ [11][12].

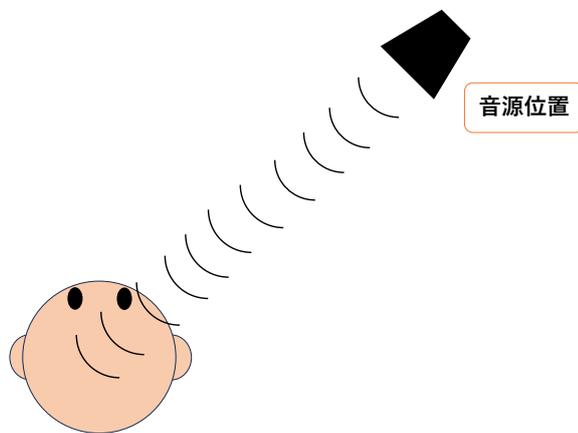


図 1.1: ILD の例

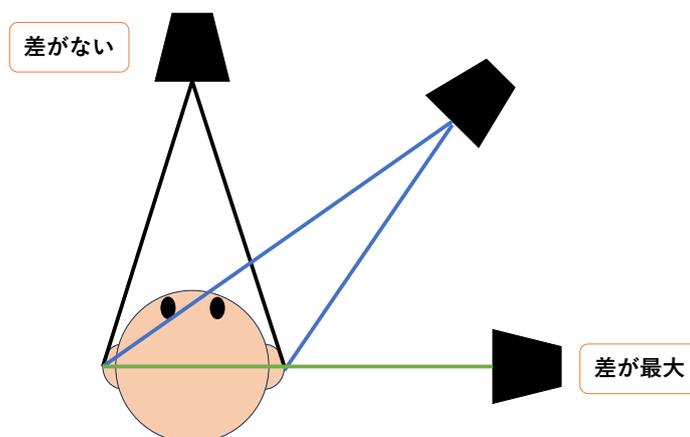


図 1.2: ITD の例

1.3 腹話術効果

音が発せられたとき、音源の位置と視覚刺激位置が異なる場合に、音源位置が実際の音源位置ではなく、視覚情報によって視覚情報が呈示されている位置に知覚される現象がある。

1.3 腹話術効果

このような現象を腹話術効果という [2][3]. この現象は一般的に腹話術師が人形を使用して、人が声を発しているが、人ではなく人形から音声が発せられるように感じさせることが腹話術効果による影響となる (図 1.3). 腹話術効果は古くから知られている現象であり、現在に至るまで多くの研究が行われてきた. 腹話術効果は視覚刺激と音源位置が 8~10deg 以内の範囲で離れている場合に音源位置に知覚され、11deg 以上の場合は腹話術効果が生じないことが報告されている [13][14]. また、腹話術効果は視覚刺激と音源を呈示する時間的特性においても重要視されており、腹話術効果は視覚刺激と音源を呈示するタイミングが同時であるほど生じやすいことが報告されている [15]. さらに、このような腹話術効果における視覚情報の影響には日常生活等での経験の有無によっても左右される. 例えば、「人の口から声が聞こえる」「テレビから音が聞こえる」といったことも、それまでに体験してきた経験によって違和感を生じることなく、視覚情報の影響を受け、テレビのスピーカーとは異なる位置に人が映っている場合においてもその人から音声が流れているように感じる [16]. 腹話術効果は人の顔と人の声のように意味的・文脈的に一致する刺激の場合に生じやすいことが明らかになっており [4], 腹話術効果で使用される人形から目や鼻等の人間的特徴が強いものを取り除いた状態で音源定位を行うと、腹話術効果の影響が減少することも報告されている [17]. VR 環境における音源定位では、アバターが呈示されている際に、アバター方向に向かって音源の位置が知覚されることが報告されている [5].

1.4 バーチャルリアリティ (VR)



図 1.3: 腹話術師と人形

1.4 バーチャルリアリティ (VR)

バーチャルリアリティ (VR:人工現実感)とは、コンピュータで作成された人工的な3次元空間を、視覚や聴覚による感覚によって現実空間と同様な空間として疑似的に体験することができる技術である。VRを構成する際に重要となる要素が3つ存在する。1つ目は、人工的な3次元空間が自然な3次元空間として知覚される「3次元の空間性」である。2つ目は、VR環境が現実空間での人の動作に対応してVR環境内も動作する「実時間の相互作用性」である。3つ目は、VR環境内に実際の空間と同様の感覚モダリティが存在し、VRの体験者自身がVR環境内に存在しているように感じる「自己投射性」である [18]。これら3つの要素を実現するために用いられるものがヘッドマウントディスプレイ (HMD) である。HMDを頭部に装着することで、HMDに搭載されているセンサーによって頭部の動作を検知し、HMDに映像をリアルタイムで映すことが可能となっている。近年、様々な種類のHMDが発売されており、より身近に接する機会が多くなっている。

1.5 認知負荷

短時間で情報を記憶した状態で、同時に記憶した情報を用いて計算や思考といった処理を行うことをワーキングメモリといい [19]、このワーキングメモリの情報保持と情報処理にかかる負担のことを認知負荷という [20]。認知負荷は3つの異なる負荷として分類される [21]。1つ目は課題内在性負荷である。これは、記憶する情報そのものの難易度や複雑さによって発生する負荷のことである。2つ目は課題外在性負荷である。これは、記憶する情報とは異なる不必要な情報や周辺環境によって発生する負荷のことである。3つ目は学習関連負荷である。これは、情報処理を行う人が学習や思考を行うことで発生する負荷のことである。会議にて会話内容やホワイトボード等にかかれた内容を記憶する際には、これらの認知負荷が重要となり、記憶する内容の難解さ、記憶する際の周辺の聴覚的・視覚的な情報過多、メモを取る等、3つに分類された負荷それぞれに関連している。本研究では、視覚情報が音声位置の知覚に影響を与える腹話術効果に対して、認知負荷を与えた際の影響を検討するため、視覚情報の記憶課題、聴覚情報の記憶課題を負荷として使用した。

第 2 章

実験 1

2.1 目的

実験 1 では、男女のアバターを同時に呈示し、男女アバターの見た目の一致・不一致及び音声と同期したアバターの口の動きの一致・不一致を操作し、どの要因がより音源定位に影響を与えるか検討した。声と性別が一致しないアバター位置に音源が呈示された場合においても、声と性別が一致するアバター位置に音源が知覚されると考えられる。また、アバターの口が動くことで、より音源定位に強い影響を与える可能性も考えられる。そのため、本実験では、声と性別が一致するアバター方向への角度を正として知覚された音源位置を算出した結果を用いて、各条件間で比較した。

2.2 実験参加者

24 名（平均年齢 22.08, SD = 1.998, 男性 12 名, 女性 12 名）の大学生が実験に参加した。いずれも正常な視力（矯正を含む）及び聴力を有していた。

2.3 実験装置

2.3.1 使用ソフトウェア

VR 環境の構築は、Unity (2020.3.40f1) を使用した。実験で得られたデータは Unity 用の C# プログラムを用いて抽出し、Microsoft Excel 2019 を用いて整理、分析を行った。各種データの検定には、R 言語 (Ver.4.2.1) [22] を使用し、分散分析を行う関数である

2.3 実験装置

ANOVAKUN[23] を用いて検定を行った。本実験で使用したアバターの作成は 3D アバター制作ソフトウェアである VRoid Studio[24] を使用した。聴覚刺激として使用した音声の変換にはボイスチェンジャーソフトウェアである VC Client[25] を使用した。

2.3.2 視覚刺激呈示装置

視覚刺激及び VR 空間の呈示にヘッドマウントディスプレイ (HMD) である Meta Quest2 (Meta 社) (図 2.1) を使用した。



図 2.1: HMD

2.3.3 聴覚刺激呈示装置

聴覚刺激は、単一指向性の小型スピーカー NS-B210 (YAMAHA 社) (図 2.2) を使用して呈示した。また、呈示するスピーカーの切り替えに 8 チャンネルスピーカーセレクタ MS-8 (MONOPRICE 社) (図 2.3) を使用し、聴覚刺激の音圧レベルの制御にヘッドホンアンプ FX-AUDIO-FX202A/FX-36A PRO (ノースフラットジャパン社) (図 2.4) を使用した。

2.3 実験装置



図 2.2: 単一指向性の小型スピーカー



図 2.3: スピーカーセレクタ



図 2.4: ヘッドホンアンプ

2.4 呈示刺激

2.3.4 コントローラー

実験中，参加者がシーン遷移及び音源位置の回答による操作を行うために，Meta Quest2 コントローラー（Meta 社）（図 2.5）を使用した。



図 2.5: コントローラー

2.4 呈示刺激

2.4.1 アバター

アバターは男性のアバターと女性のアバターの 2 種類を使用した。男性アバターは主観的に男性に感じられる見た目に設定するため，サンプルモデルをもとに VRoid Studio で作成したアバターを使用した（図 2.6）。女性アバターは VRoid Studio に付属されているサンプルモデルを使用した（図 2.7）。

アバターは男女同時に呈示し，実験参加者の受聴位置から半径 150 cm の円周上に並べて配置した。2 体のアバター間の距離は，先行研究より，腹話術効果は 10 deg 以内であると視覚刺激方向に音源が定位することが報告されているため [13]，常に 10 deg の間隔で呈示した（図 2.8）。2 体のアバターの左右位置は，1 試行ごとにランダムな順で入れ替えて呈示した。アバターの目線の高さは，回転椅子に座った実験参加者の目線とほぼ同じとなる 120 cm とした。

2.4 呈示刺激



図 2.6: 男性アバター

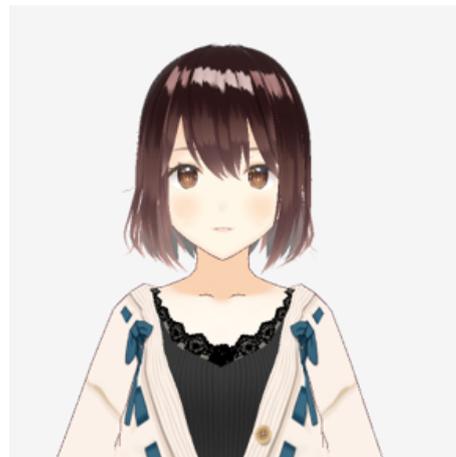


図 2.7: 女性アバター



図 2.8: 呈示されたアバター位置

2.4.2 聴覚刺激

聴覚刺激として男性の声と女性声を呈示した。男性の声は実験者の声を録音したものを使用した。女性の声は実験者の声を録音した音声を VC Client を使用して女性に聞こえる音声に変換したものを使用した。音声の録音には単一指向性のコンデンサーマイク HyperX SoloCast (HyperX 社) (図 2.9) を使用した。呈示する音声の音圧に差が生じないようにするため、音声の音圧レベルは受聴位置で約 60 dB となるように呈示した。また、聴覚刺激は 6 個のスピーカーのうち、ランダムな順に 1 個のスピーカーを選択して呈示した。

2.4 呈示刺激

音声内容は関連性のない3文字の単語10種類を使用した。単語は母音AIUEOの中からランダムな組み合わせの母音で組み合わせられた単語を使用した。使用した単語10種類の内容については付録Aに記述する。



図 2.9: 単一指向性のコンデンサーマイク

2.4.3 VR 画面上の実験環境

HMD に呈示される3次元空間は音源知覚の位置に影響を与えないために、回りを見渡した際に、地面と空の境界線、壁同士の境界線を認識できないほど遠い位置に設定し、アバター以外のVR環境を手がかりとして知覚されないように呈示した(図 2.10)。

また、片方のアバターに注視することを防ぐため、注視点を2体のアバターの中心かつアバターの目線と同じ高さの120 cmに呈示した(図 2.11)。

2.5 現実空間上の実験環境



図 2.10: VR 環境



図 2.11: 注視点位置

2.5 現実空間上の実験環境

本実験では、実験参加者は HMD を装着し、右手にコントローラーを持った状態で回転椅子に座り、音源定位課題を行った（図 2.12）。スピーカーは受聴位置を中心とした半径 150 cm の円周上に 10 deg ずつずらした状態で 6 個配置した（図 2.13）。6 個のスピーカーの内ランダムな順で一つのスピーカーを選択して音声を呈示した。プログラムの円滑な動作のため、プログラムを制御する PC との接続は有線で接続し、ケーブルは実験中、実験参加者の動作に影響を与えないために、上から吊るした状態で接続した。

スピーカーの個数や配置を実験参加者に把握させないため、スピーカーをボードで隠した状態で参加者に HMD を装着し、実験終了後は、再度スピーカーをボードで隠した後、参加者から HMD を外した（図 2.14）。

2.5 現実空間上の実験環境



図 2.12: 実験風景



図 2.13: スピーカー配置

2.6 実験条件



図 2.14: スピーカーを隠した状態の風景

2.6 実験条件

アバターの口が動く際、条件に応じて音声と同期してアバターの口が動いた。アバターの口の動きは呈示された音声内容の母音を発話する口の形と同様の口の形に動いた。それぞれの母音に対応する口の形については付録 B の通りである。

アバターの口が動く条件として以下の 4 条件を設定した。

- 両アバターの口が動く条件
- 男性アバターのみ口が動く条件
- 女性アバターのみ口が動く条件
- 両アバターの口が動かない条件

聴覚刺激呈示条件として以下の 5 条件を設定した (図 2.15)。

- 2 体のアバターの中心位置 (0 deg)
- 片方のアバターと同じ位置 (± 5 deg)
- 片方のアバター位置から外側へ 5 deg ずらした位置 (± 10 deg)

2.7 実験手続き

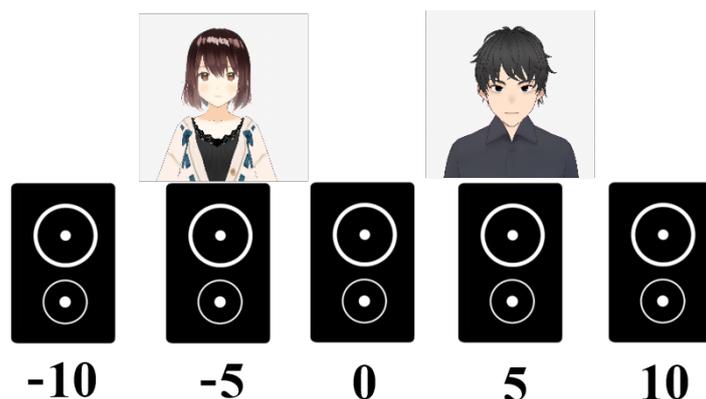


図 2.15: 聴覚刺激呈示条件 (deg)

2.7 実験手続き

実験開始時、画面上には上部に矢印と参加者の顔の向きと同じ方向に伸びているバーが表示された (図 2.16)。実験開始後、参加者が参加者自身とスピーカーとの位置関係を把握することを防ぐために、参加者自身が矢印方向に椅子を半回転させた後、特定の位置で矢印が逆方向になり、逆方向になったタイミングで参加者自身が椅子を逆方向に回転させた (図 2.17)。VR 空間上の特定の位置にオブジェクトが呈示され、顔の向きと同じ方向に伸びているバーとオブジェクトが同じ位置となる正面位置で停止し、注視点を注視した状態でコントローラーのボタンを押した (図 2.18)。その後、アバターが呈示され、音声再生された (図 2.19)。音声の再生後アバターが消失し、音源位置回答用のバーが表示された。コントローラーのスティック入力により、バーを左右方向に移動させて音源位置を回答した。そして再び椅子を半回転させる手順に戻る。この流れを 1 試行とした。また、試行開始時に呈示される矢印方向は 1 試行ごとに入れ替えて呈示した。

2.8 アンケート

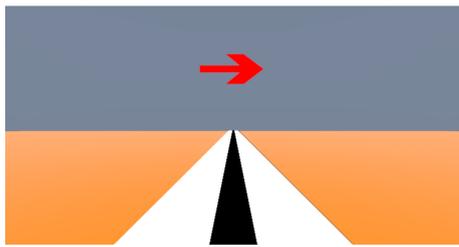


図 2.16: 実験開始時の VR 空間

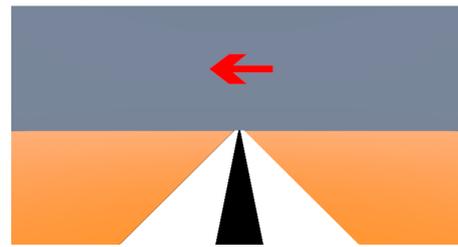


図 2.17: 逆方向への回転時の VR 空間

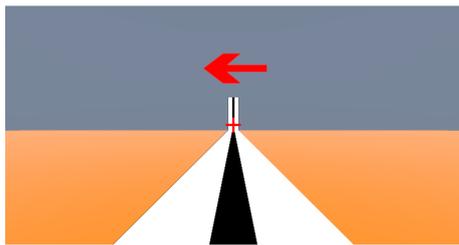


図 2.18: 回転停止時の VR 空間



図 2.19: アバター呈示時の VR 空間

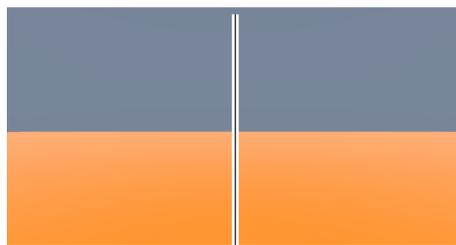


図 2.20: 音源位置回答用のバー

2.8 アンケート

本実験において、声と性別が一致・不一致による影響を検討するため、呈示したアバターや音声の設定した通りの性別に感じられたかについて明らかにする必要がある。そこで実験終了後、呈示した男女のアバターや男女の声が、設定した通りの性別に感じられたかを、5段階リッカート尺度アンケートを行った。実際のアンケート項目については付録 D に記述する。

2.9 アンケート結果

実験後アンケートの結果、男性のアバター及び女性のアバターともに、実験参加者 24 名のうち、すべての参加者が 5 段階のリッカート尺度で一番高い数値の性別通りに感じたとは回答した。また男性の声及び女性の声においても、すべての参加者が 5 段階のリッカート尺度で一番高い数値の性別通りに感じたとは回答した。このことから、本実験においては、男女の見た目の違いは有効であったといえる。

2.10 分析方法

アンケート結果より、男女の見た目の違いが有効であったため、男女の違いをまとめて音声とアバターの性別の一致・不一致を指標として分析を行った。また、左右位置の違いもまとめて扱った。

アバターに対する音源位置

- 2 体のアバターの中央に音声を呈示した場合を 0 deg
- 声と見た目の性別が一致するアバターの位置に音声を呈示した場合を 5 deg
- 声と見た目の性別が一致しないアバターの位置に音声を呈示した場合を -5 deg
- 5 deg 条件よりも外側に 5 deg ずらした位置に音声を呈示した場合を 10 deg
- -5 deg 条件よりも外側に 5 deg ずらした位置に音声を呈示した場合を -10 deg

これら 5 つを音源が呈示された位置の要因として分析を行った（図 2.21）。具体的な要因の分け方については付録 F に記述する。

2.11 結果

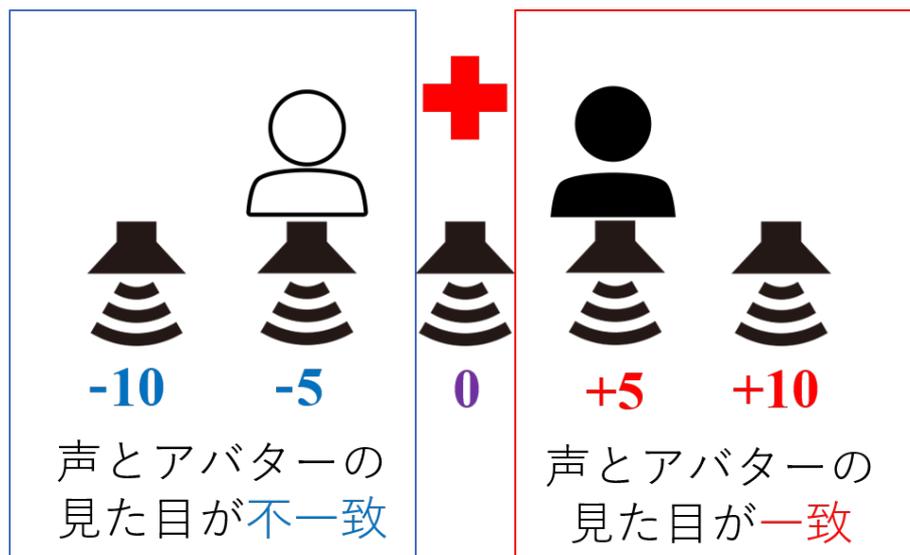


図 2.21: アバターに対する呈示された音源位置

口の動き要因

- 両アバターの口が動く条件
- 声の性別と一致するアバターの口が動く条件
- 声の性別と一致しないアバターの口が動く条件
- 両アバターの口が動かない条件

これらの4条件をアバターの口の動きの要因として分析を行った。この二つの要因によって、知覚された音源位置 (deg) がどのように変化するかを検討した。音源位置の結果は、2体のアバターの中心を 0 deg として、声の性別と一致するアバター方向への角度を正とした。

2.11 結果

知覚された音源位置の平均値を図 2.22 に示す。各条件でのエラーバーは標準誤差を表す。点線は現実空間上での音源位置と知覚された音源位置が同じ位置であった場合の基準線である。一方で、音源位置が音声と性別が同じアバターの位置に知覚された場合、5 deg の横線と同じ位置となる。アバターの視覚情報によって音源位置が完全にではなく、不十分ながら

2.11 結果

も変調された場合はこれらの線の間位置することになる。

実験の結果,

両アバターの口が動かない条件の各平均値は-10 deg 条件から順に,

-8.04, -4.36, 2.66, 5.56, 10.22 deg であった。

声の性別と一致するアバターの口が動く条件の各平均値は-10 deg 条件から順に,

-4.60, -0.17, 4.44, 6.54, 9.61 deg であった。

声の性別と一致しないアバターの口が動かない条件の各平均値は-10 deg 条件から順に,

-8.96, -5.23, -1.25, 1.17, 9.18 deg であった。

両アバターの口が動く条件の各平均値は-10 deg 条件から順に,

-7.87, -3.24, 3.04, 6.82, 8.67 deg であった。

この結果から、両アバターの口が動く条件及び両アバターの口が動かない条件においては、実際に呈示された音源位置に近い位置に音源が知覚される傾向がみられた。声の性別と一致するアバターの口が動く条件においては、口が動くアバター方向に音源がずれて知覚される傾向がみられた。声の性別と一致しないアバターの口が動く条件においては、5 deg 条件では声の性別が一致しないアバター方向に音源がずれて知覚され、その他の条件では呈示された音源位置に音源が知覚される傾向がみられた。

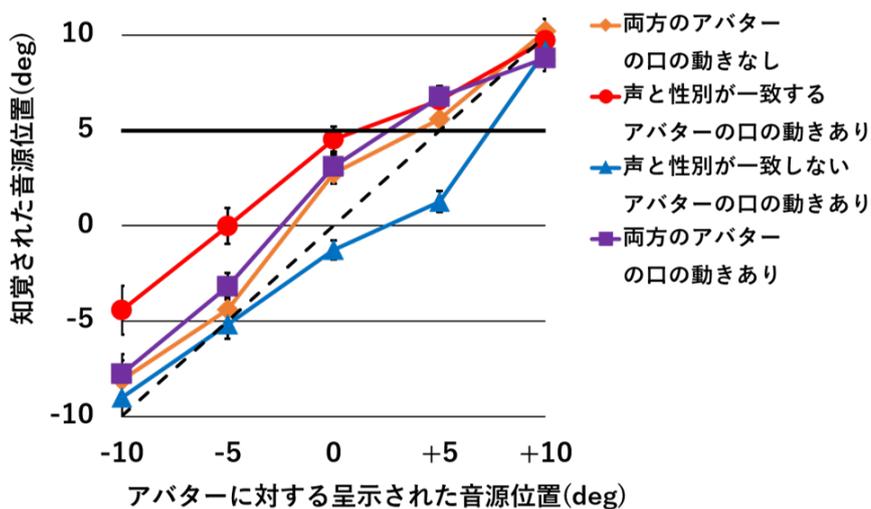


図 2.22: アバターの口が動く条件ごとの知覚された音源位置

2.12 考察

アバターに対する音源位置と口の動きを要因とする対応ありの 2 要因分散分析を行った結果、アバターに対する音源位置 ($F(1.16, 26.64)=128.32, p=0.00, \eta^2=0.67$)、口の動き ($F(1.39, 31.92)=22.60, p=0.00, \eta^2=0.04$) の各主効果及び、これらの交互作用いずれも有意であった ($F(5.07, 116.65)=12.68, p=0.00, \eta^2=0.02$)。

Shaffer の方法による多重比較の結果、5 deg 条件において、声と性別が一致しないアバターの口が動く条件と、他の 3 条件（両アバターの口が動く条件、声と性別が一致するアバターの口が動く条件、両アバターの口が動かない条件）との間に有意な差がみられた ($p<0.01$)。また、-5 deg 条件において、声と性別が一致するアバターの口が動く条件と、他の 3 条件（両アバターの口が動く条件、声と性別が一致しないアバターの口が動く条件、両アバターの口が動かない条件）との間に有意な差がみられた ($p<0.01$)。一方で、両アバターの口が動く条件と両アバターの口が動かない条件間には有意な差はみられなかった ($p=0.23$)。

-5 deg 条件、0 deg 条件、5 deg 条件において、知覚された音源位置が実際の音源位置から有意に差があるかを検定するため、各条件で 1 標本の t 検定を行った。その結果、声と性別が一致するアバターの口が動く条件において、知覚位置の値が-5 deg より有意に大きく ($t=5.11, p=0.00$)、声と性別が一致しないアバターの口が動く条件において、知覚位置の値が 5 deg より有意に小さかった ($t=-6.83, p=0.00$)。また、0 deg 条件においては、声と性別が一致するアバターの口が動く条件における知覚位置の値が 0 deg より有意に大きかった ($t=6.74, p=0.00$)。

2.12 考察

先行研究では、呈示した音源位置と空間的に一致するという低次な情報の知覚処理よりも、聴覚刺激と視覚刺激が文脈的に一致するかという高次な知覚処理の方が強く音源定位に影響を与えることが報告されている [5]。この先行研究の結果によれば、声と性別が異なるアバターの位置に音声呈示される本研究の-5 deg 条件においては、声と性別が一致する

2.12 考察

アバター位置 (5 deg) の方に音源位置が知覚されると予測され、声と性別が一致するアバターの口の動きがある条件では、そのアバターの位置に知覚されると予測される。しかし、実際の結果は、-5 deg 条件及び-10 deg 条件で、どの口の動きの条件においても音源位置が声と性別が一致するアバター位置である 5 deg 位置には知覚されず、声と性別が一致するアバターの口が動く条件では、口が動くアバター方向に有意に近づく結果となったが、その他の条件においては現実空間上に呈示された音源位置に近い位置に知覚されていた。このことから、本研究で検討したような、音声とアバターの性別が一致しているという高次な文脈的一致の効果が腹話術効果に与える影響は小さいことが考えられる。声と性別が一致するアバターの口が動く条件とその他の条件間には有意な差がみられ、声と性別が一致するアバターに有意に近づいたことから、音声と同期したアバターの口の動きの一致の効果は腹話術効果に与える効果が大きいことが考えられる。また、声と性別が一致するアバターの位置に音声呈示される 5 deg 条件においても、声と性別が一致しないアバターの口が動く条件と他の条件間に有意な差がみられ、声と性別が一致しないアバターのほうに有意に近づいたことから、音声と同期したアバターの口の動きの一致の効果は、音声とアバターの性別が一致している文脈的一致の効果よりも、腹話術効果に強く影響を与えることが考えられる。なお、この一致は視聴覚刺激が生じる時間タイミングとしては低次な情報でもあり、さらに、視覚的な変化が注意を誘発するため、ボトムアップ的注意の効果も生じる。これらの相互作用によって声と見た目的一致・不一致に関わらず、口が同期して動いたほうに音源が知覚されやすくなると考えられる。以上の結果より、VR 環境において話者から発話されているように知覚させるためには、同期した口の動きを表示することが効果的であることが示された。

第 3 章

実験 2

3.1 目的

実験 2 では、視覚的な負荷課題、聴覚的な負荷課題がそれぞれアバターに対する音源定位に影響を与えるか検討した。触覚による注意不可をかけた場合、マガーク効果に対して視覚刺激による効果が減少する結果が、先行研究で報告されている [8]。この先行研究の結果から、本研究の視覚的な負荷課題と聴覚的負荷課題においては負荷がない状態よりも、アバター位置ではなく実際の音源定位に近い位置に知覚されると考えられる。そのため、本実験では、アバター方向に偏移した角度を正として、知覚された音源位置を条件ごとに比較した。

3.2 実験参加者

23 名（平均年齢 22.08, SD = 1.886, 男性 12 名, 女性 11 名）の大学生が実験に参加した。いずれも正常な視力（矯正を含む）及び聴力を有していた。

3.3 実験装置

3.3.1 使用ソフトウェア

VR 環境の構築は、Unity (2020.3.40f1) を使用した。実験で得られたデータは Unity 用の C# プログラムを用いて抽出し、Microsoft Excel 2019 を用いて整理、分析を行った。各種データの検定には、R 言語 (Ver.4.2.1) [22] を使用し、分散分析を行う関数である

3.3 実験装置

ANOVAKUN[23] を用いて検定を行った。本実験で使用したアバターの作成は 3D アバター制作ソフトウェアである VRoid Studio[24] を使用した。

3.3.2 視覚刺激呈示装置

視覚刺激及び VR 空間の呈示にヘッドマウントディスプレイ (HMD) である Meta Quest2 (Meta 社) (図 3.1) を使用した。



図 3.1: HMD

3.3.3 聴覚刺激呈示装置

聴覚刺激は、単一指向性の小型スピーカー NS-B210 (YAMAHA 社) (図 3.2) を使用して呈示した。また、呈示するスピーカーの切り替えに 8 チャンネルスピーカーセレクタ MS-8 (MONOPRICE 社) (図 3.3) を使用し、聴覚刺激の音圧レベルの制御にヘッドホンアンプ FX-AUDIO-FX202A/FX-36A PRO (ノースフラットジャパン社) (図 3.4) を使用した。

3.3 実験装置



図 3.2: 単一指向性の小型スピーカー



図 3.3: スピーカーセレクトタ



図 3.4: ヘッドホンアンプ

3.4 呈示刺激

3.3.4 コントローラー

実験中，参加者がシーン遷移及び音源位置の回答による操作を行うために，Meta Quest2 コントローラー（Meta 社）（図 3.5）を使用した。



図 3.5: コントローラー

3.4 呈示刺激

3.4.1 アバター

アバターは男性アバターのみを使用した（図 3.6）。男性アバターは VRoid Studio で作成したアバターを使用した。アバターは実験参加者の受聴位置から半径 150 cm の円周上に配置し，アバターの目線の高さは，回転椅子に座った実験参加者の目線とほぼ同じとなる 120 cm とした。また，アバター位置はアバターと音源位置の位置関係を把握しづらくさせるため，参加者正面位置（図 3.9），参加者正面から ± 5 deg ずらした位置（図 3.7，図 3.8）からランダムに呈示した。

3.4 呈示刺激



図 3.6: 男性アバター

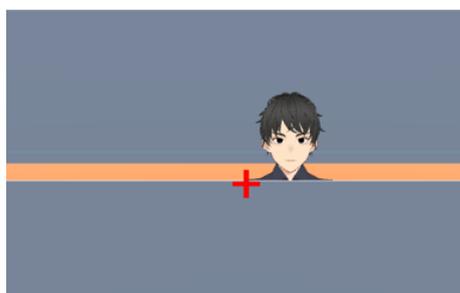


図 3.7: 参加者正面から+5 deg
のアバター位置

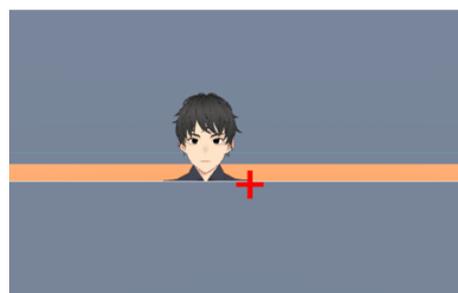


図 3.8: 参加者正面から-5 deg
のアバター位置



図 3.9: 参加者正面のアバター位置

3.4 呈示刺激

3.4.2 聴覚刺激

聴覚刺激として男性の声として実験者の声を録音したものを呈示した。音声の録音には単一指向性のコンデンサーマイク HyperX SoloCast (HyperX 社) (図 3.10) を使用した。呈示する音声の音圧に差が生じないようにするため、音声の音圧レベルは受聴位置で約 60 dB となるように呈示した。また、聴覚刺激は 6 個のスピーカーのうち、ランダムな順に 1 個のスピーカーを選択して呈示した。

音声内容は関連性のない 3 文字の単語 10 種類を使用した。使用した単語 10 種類の内容については付録 A に記述する。



図 3.10: 単一指向性のコンデンサーマイク

3.4.3 視覚負荷課題における視覚刺激

視覚負荷課題として呈示する刺激として、聴覚刺激として呈示される音声内容とは異なる、関連性のない 3 文字の単語 10 種類を使用した。使用した単語 10 種類の内容については付録 C に記述する。単語の呈示位置はアバターの目線から 20 cm 下となる位置に呈示した (図 3.11)。

3.5 現実空間上の実験環境



図 3.11: 視覚負荷条件における単語呈示位置

3.4.4 VR 画面上の実験環境

HMD に呈示される 3 次元空間は音源知覚の位置に影響を与えないために、回りを見渡した際に、地面と空の境界線、壁同士の境界線を認識できないほど遠い位置に設定し、アバター以外の VR 環境を手がかりとして知覚されないように呈示した (図 3.12)

また、アバター及び視覚負荷条件にて呈示される単語に注視することを防ぐため、注視点をアバターと単語の間となるアバターの目線から 10cm 下の位置に呈示した (図 3.11)。



図 3.12: VR 環境

3.5 現実空間上の実験環境

本実験では、実験参加者は HMD を装着し、右手にコントローラーを持った状態で回転椅子に座り、音源定位課題及び認知負荷課題を行った (図 3.13)。スピーカーは受聴位置を中

3.5 現実空間上の実験環境

心とした半径 150 cm の円周上に 10 deg ずつずらした状態で 6 個配置した (図 3.14) . プログラムの円滑な動作のため, プログラムを制御する PC との接続は有線で接続し, ケーブルは実験中, 実験参加者の動作に影響を与えないために, 上から吊るした状態で接続した.

スピーカーの個数や配置を実験参加者に把握させないため, スピーカーをボードで隠した状態で参加者に HMD を装着し, 実験終了後は, 再度スピーカーをボードで隠した後, 参加者から HMD を外した (図 3.15) .



図 3.13: 実験風景

3.6 実験条件



図 3.14: スピーカー配置



図 3.15: スピーカーを隠した状態の風景

3.6 実験条件

認知負荷条件として、聴覚負荷条件、視覚負荷条件、負荷なし条件の計 3 条件を設定した。実験順序は参加者間でカウンターバランスを取り、よいや疲労の効果を低減させるため、30 分以上の時間を空けて実験を行った。本研究では負荷課題として n-back 課題 ($n=2$) を使用

3.7 実験手続き

した。聴覚負荷条件では、呈示された音声は2試行前の音声と同一か判断する n-back 課題 (n=2) を行った。視覚負荷条件では、アバターの下に呈示された単語が2試行前に呈示された単語と同一か判断する n-back 課題 (n=2) を行った。

聴覚刺激呈示条件として、聴覚刺激位置をアバターと同一位置条件、アバターに対して左右方向に 5 deg ずらした条件、10 deg ずらした条件の計 3 条件を設定した。

3.7 実験手続き

実験開始時、画面上には上部に矢印と参加者の顔の向きと同じ方向に伸びているバーが表示された (図 3.16)。実験開始後、参加者が参加者自身とスピーカーとの位置関係を把握することを防ぐために、参加者自身が矢印方向に椅子を半回転させた後、特定の位置で矢印が逆方向になり、逆方向になったタイミングで参加者自身が椅子を逆方向に回転させた (図 3.17)。VR 空間上の特定の位置にオブジェクトが呈示され、顔の向きと同じ方向に伸びているバーとオブジェクトが同じ位置となる正面位置で停止し、注視点を注視した状態でコントローラーのボタンを押した (図 3.18)。その後、アバターが呈示され、音声は再生された (図 3.19)。音声の再生後アバターが消失し、音源位置回答用のバーが表示された (図 3.20)。コントローラーのスティック入力により、バーを左右方向に移動させて音源位置を回答した。そして再び椅子を半回転させる手順に戻る。この流れを1試行とした。また、試行開始時に呈示される矢印方向は1試行ごとに代入して呈示した。

聴覚負荷条件では、再生された音声内容が2試行前に再生された音声内容と同じ場合に、コントローラーのボタンを押した。視覚負荷条件では、呈示された単語が2試行前に呈示された単語と同じ場合に、コントローラーのボタンを押した (図 3.21)。

3.7 実験手続き

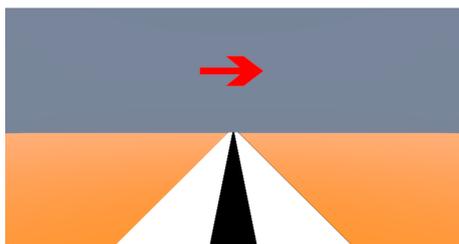


図 3.16: 実験開始時の VR 空間

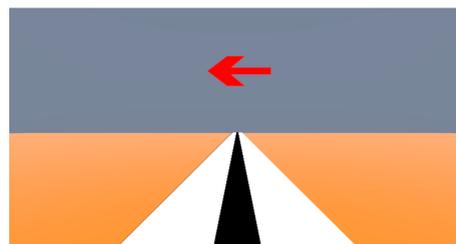


図 3.17: 逆方向への回転時の VR 空間

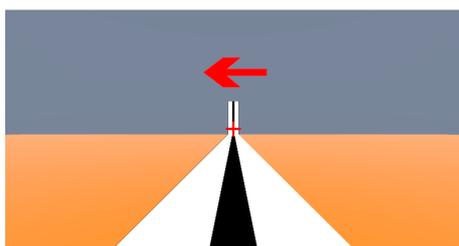


図 3.18: 回転停止時の VR 空間

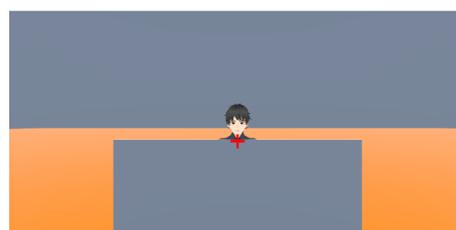


図 3.19: アバター呈示時の VR 空間

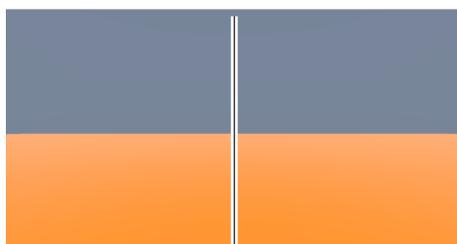


図 3.20: 音源位置回答用のバー



図 3.21: 視覚負荷条件におけるアバター呈示時の VR 空間

3.8 アンケート

本実験において、負荷課題の負荷の高さに応じて、腹話術効果に与える影響が変化する可能性があるため、実験参加者ごとにどの程度負荷がかかっていたかについて明らかにする必要がある。そこで実験終了後、負荷課題の難易度について、5段階リッカート尺度アンケートを行った。実際のアンケート項目については付録 E に記述する。

3.9 アンケート結果

実験後アンケートの結果は表 3.2 の通りである。実験参加者 23 名の内、聴覚負荷条件において n-back 課題が難しいと回答した参加者は 7 名、n-back 課題が少し難しいと回答した参加者は 13 名であり、視覚負荷条件において n-back 課題が難しいと回答した参加者は 9 名、n-back 課題が少し難しいと回答した参加者は 13 名であった。このことから、二重課題による負荷の操作は主観的には一定程度は達成されたと考えられる。

回答内容	聴覚負荷条件	視覚負荷条件
難しい	7 名	9 名
少し難しい	13 名	13 名
どちらとも言えない	2 名	1 名
少し簡単	1 名	0 名
簡単	0 名	0 名

表 3.1: 実験 2 アンケート結果

3.10 負荷課題の正答率

聴覚負荷条件において、n-back 課題の正答率は 93.46% (SD=5.01)、視覚負荷条件において、n-back 課題の正答率は 92.65% (SD=5.62) であった。このことから、実験参加者が十分に集中した状態で、認知負荷課題を行うことができたといえる。

聴覚負荷条件	視覚負荷条件
93.46%	92.65%

表 3.2: n-back 課題の正答率

3.11 分析方法

認知負荷条件

- 聴覚負荷条件
- 視覚負荷条件
- 負荷なし条件

これらの 3 条件を認知負荷条件の要因として分析を行った。

アバターとの相対的な音源位置

- アバターと同じ位置に音声を呈示した場合を 0 deg 条件
- アバター位置から 5 deg ずらした位置に音声を呈示した場合を 5 deg 条件
- アバター位置から 10 deg ずらした位置に音声を呈示した場合を 10 deg 条件

これらの 3 条件をアバターとの相対的な音源位置の要因として分析を行った。

この二つの要因によって、知覚された音源位置 (deg) がどのように変化するかを検討した。音源位置の結果は、左右位置の違いはまとめて扱い、呈示された音源位置を 0deg とし、アバター方向へ偏位した角度を正とした。また、負荷のかかった状態のデータのみを扱うため、全ての条件において 3 試行目以降のデータを用いた。

3.12 結果

横軸を音源位置に対して呈示されたアバター位置とし、縦軸を実験参加者に知覚された音源位置として図 3.22 に示す。腹話術効果が生じ、アバター位置に音源が知覚された場合には斜めの実線位置となり、腹話術効果が生じず、呈示された音源位置に知覚された場合には縦軸の 0 deg の点線位置となる。

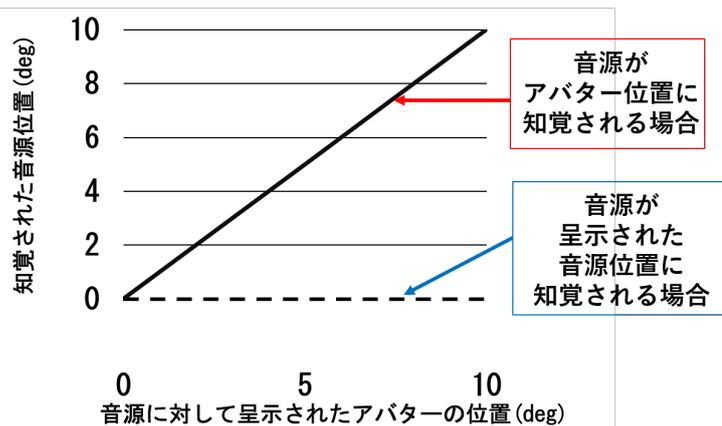


図 3.22: 予測される結果図

3.12 結果

負荷条件ごとの知覚された音源位置の平均値を図 3.23 に示す。各条件でのエラーバーは標準誤差を表す。

実験の結果、

聴覚負荷条件の各平均値は 0 deg 条件から順に、-1.17, 6.19, 3.33 deg,

視覚負荷条件の各平均値は 0 deg 条件から順に、-0.98, 4.84, 3.61 deg,

負荷なし条件の各平均値は 0 deg 条件から順に、-0.30, 4.20, 1.54 deg であった。

3.13 考察

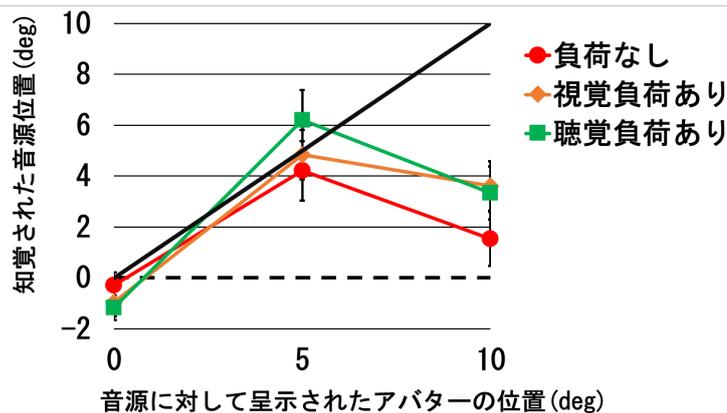


図 3.23: 負荷条件ごとの知覚された音源位置

視覚負荷条件および聴覚負荷条件ともに音源位置には知覚されずアバター方向に知覚され、視覚および聴覚による負荷課題を与えた状態では腹話術効果を減衰させる結果とはならなかった。また、10 deg 条件においてはどの条件においてもアバター位置には知覚されず、呈示された音源位置に近い位置に知覚される結果となった。アバターの相対的位置と認知負荷条件を要因とする対応ありの 2 要因分散分析をおこなった結果、アバターの相対的位置の主効果は有意であったが ($F(2, 44)=15.02, p=0.00, \eta^2=0.23$)、認知負荷条件の主効果 ($p>0.05$) 及びこれらの交互作用は有意ではなかった ($p>0.05$)。

3.13 考察

数字を記憶する課題を行った場合、マガーク効果に対して視覚刺激による効果が減少する結果が先行研究で報告されている [7]。この先行研究の結果から、本研究の視覚負荷条件および聴覚負荷条件においては、負荷がない条件よりも実際の音源呈示位置に近い位置に知覚されると予測される。しかし、実際の結果、負荷条件の有意な効果はみられず、負荷がかかった視覚負荷条件、聴覚負荷条件ともに呈示された音源位置ではなく、アバターの位置に近い位置に知覚された。このことから、視覚や聴覚による認知負荷は、アバターに対する腹話術効果を減衰させる効果は生じず、VR 環境において音声を話者のアバター位置に音源定

3.13 考察

位させる効果を検討する際には、他のタスクによる認知的負荷の影響はそれほど考慮しなくても良い可能性がある。また、有意な結果は得られなかったが、視覚負荷条件および聴覚負荷条件ともに負荷なし条件よりもアバター方向に知覚される傾向がみられた。これは、視覚や聴覚の課題による負荷が生じている場合には、聴覚の位置情報に注意が向かず、より視覚情報に注意が向き強く腹話術効果がみられた可能性が考えられる。

先行研究において腹話術効果は 10 deg 以内では生じやすいことが報告されていたことから [13]、本研究のアバターと音源位置の距離の最大値は 10 deg に設定した。しかし、実際の結果では、5 deg 条件では腹話術効果がみられ、アバター位置に音源が知覚されたが、10 deg 条件においては 3 つの負荷条件いずれも、アバターの位置に音源が知覚されず、呈示された音源位置に近い位置に知覚される結果となった。これは本研究の実験では、0 deg 条件、5 deg 条件、10 deg 条件をそれぞれランダムな順で呈示しており、0 deg 条件や 5 deg 条件に比べて 10 deg 条件はアバターと呈示された音源位置が離れていると感じたことから、腹話術効果が発生しづらく、呈示された音源位置に近い位置に知覚される結果となったことが考えられる。また、10 deg は腹話術効果が生じやすい角度の限界であることから、本研究の VR を用いた環境においては、アバターと音源位置が 10 deg 離れた場合には腹話術効果が生じなかったと考えられる。視覚刺激と音源位置の距離を 10 deg ではなく、8 deg に設定して腹話術効果がみられるか検討している研究もあり [14]、アバターと音源位置の距離を 8 deg に設定した場合には腹話術効果がみられる可能性がある。

第 4 章

まとめ

本研究では、男女のアバターの視覚的特徴の違いや、視覚・聴覚による認知負荷を与えた状態における視聴覚統合に基づく音源定位への影響について検討した。実験 1 では、音声とアバターの性別が一致・不一致に関わらず、口が動くアバター方向に音源位置が知覚された。実験 2 では、マガーク効果で検討した先行研究とは異なり、認知負荷が生じても視覚刺激の影響が減少する効果は見られなかった。以上の結果より、VR 環境において話者から発話されているように知覚させるためには、同期した口の動きを表示することが効果的であることが示され、音声を話者のアバター位置に音源定位させる効果を検討する際には、他のタスクによる認知的負荷の影響はそれほど考慮しなくても良い可能性がある。

謝辞

本研究の実施及び論文の作成にあたり，多大なご支援とご指導を賜りました繁榊博昭先生には深く感謝致します。また，本論文の副査を務めていただいた門田宏先生，西野裕樹先生及び竹田真己先生にも深く感謝致します。最後に，本研究の実験に参加していただいた47名の皆様にも感謝致します。

参考文献

- [1] S. Choi, K. Lee, C. He, K. Donovan, M. Billinghurst, “Social VR as the New Normal: A Framework for Interactive Virtual Meetings Using Mobile AR and Social Virtual Reality”, *IEEE Transactions on Visualization and Computer Graphics*, Vol.28, No.11, pp.3678-3688, 2022.
- [2] R. B. Welch and D. H. Warren, “Immediate perceptual response to intersensory discrepancy”, *Psychological Bulletin*, vol.88, no.3, pp.638-667, 1980.
- [3] L. Chen, J. Vroomen, “Intersensory binding across space and time: A tutorial review”, *Attention Perception Psychophysics*, vol.75, pp.790-811, 2013.
- [4] R. I. Bermant and R.B.Welch, “Effect of degree of separation of visual-auditory stimulus and eye position upon spatial interaction of vision and audition”, *Percept Mot Skills*, 42, pp.487-493, 1976.
- [5] G. I. Parisi, P. Barros, D. Fu, S. Magg, H. Wu, G. Yang, Z. Li, X. Liu, and S. Wermter, “A computational model of crossmodal processing for conflict resolution”, in *IEEE EPIROB-ICDL*, pp.33-38, 2017.
- [6] H. McGurk and J. MacDonald, “Hearing lips and seeing voice”, *Nature* 264, pp.746-748, 1976.
- [7] J. N. Buchan and K. G. Munhall, “The Effect of a Concurrent Working Memory Task and Temporal Offsets on the Integration of Auditory and Visual Speech Information”, *Seeing and Perceiving* 25, pp.87-106, 2012.
- [8] A. Alsius, J. Navarra, S. S. Faraco, “Attention to touch weakens audiovisual speech integration”, *Exp Brain Res* 183, pp.399-404, 2007.
- [9] J. Blauert, “Spatial Hearing: The Psychophysics of Human Sound Localization (Revised Edition)”, The MIT Press, 1996.

参考文献

- [10] B. C. J. Moore, “An Introduction to the Psychology of Hearing(6th edition)”, Acoustical Society of America, 2014.
- [11] 古川茂人, “音を聞く仕組みを探る”, 計測と制御, 43 巻 4 号, pp.294-300, 2004.
- [12] M. F. ベアー, B. W. コノーズ, M.A. バラディーソ著, 加藤宏司, 後藤薫, 藤井聡, 山崎良彦監訳, “神経科学 脳の探求”, 西村書店, pp.284-286, 2012.
- [13] 中林克己, 辻本廉, 二階堂誠也, “ステレオ音像とテレビ映像の相互作用に関する基礎実験”, 昭 54 春季日本音響学会講演論文集, pp.245-246, 1979.
- [14] G. H. Recanzone, “Rapidly induced auditory plasticity: the ventriloquism aftereffect”, Proc. Natl. Acad. Sci. USA 95, pp.869-875, 1998.
- [15] D. A. Slutsky, G.H. Recanzone, “Temporal and spatial dependency of the ventriloquism effect”, NeuroReport, 12, pp.7-10, 2001.
- [16] 中林克己, 小宮山撰, 溝口章夫, “テレビ映像とステレオ音像の相互作用に関する基礎実験”, テレビジョン学会技術報告, pp55-59, 1982.
- [17] C. E. Jack, W. R. Thurlow, “Effect of degree of visual association and angle of displacement on the “Ventriloquism” effect”, Perceptual and Motor Skills, 37, pp.967-979, 1973.
- [18] 舘すすむ, 佐藤誠, 廣瀬通考, “バーチャルリアリティ学”, コロナ社, 2011.
- [19] M. Osaka, M. Komori, M. Morishita, N. Osaka, “Neural bases of focusing attention in working memory: an fMRI study based on group differences”, Neuroscience and Biobehavioral Reviews, 101, pp.1-12, 2019.
- [20] W. Schnotz and C. Kurschner, “A Reconsideration of Cognitive Load Theory”, Educ Psychol Rev, 19, pp.469-508, 2007.
- [21] J. Sweller, “Cognitive load during problem solving: Effects on learning”, Cognitive Science, 12, pp.257-285, 1988.
- [22] R Core Team, “R: A language and environment for statistical computing.”, The R Project for Statistical Computing, <https://www.r-project.org/>, 2025.

参考文献

- [23] 井関龍太, “anovakun version 4.8.9”, <https://riseki.cloudfree.jp/?ANOVA>
- [24] ピクシブ, “VRoid Studio”, <https://vroid.com/studio>, 2025.
- [25] W. Okada, “voice-changer”, <https://github.com/w-okada/voice-changer>, 2025.

付録 A

音声内容

実験 1 及び実験 2 で使用した音声内容における単語 10 種類を以下に示す.

- アイス
- バスケット
- ピアノ
- リソウ
- スミレ
- ウオザ
- セカイ
- エノグ
- ホタテ
- トケイ

付録 B

音声内容

母音に対応するアバターの口の形は以下の通りである。

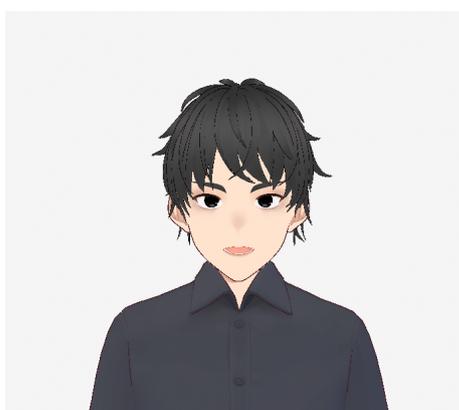


図 B.1: 母音が「あ」の場合の口の形



図 B.2: 母音が「い」の場合の口の形

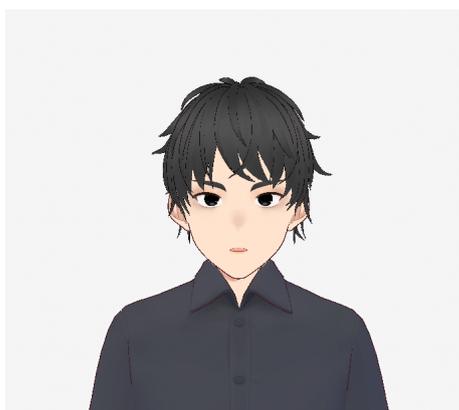


図 B.3: 母音が「う」の場合の口の形

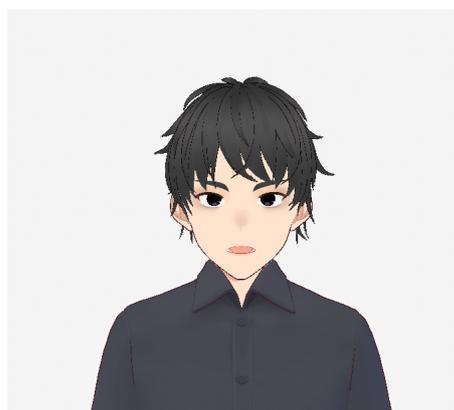


図 B.4: 母音が「え」の場合の口の形

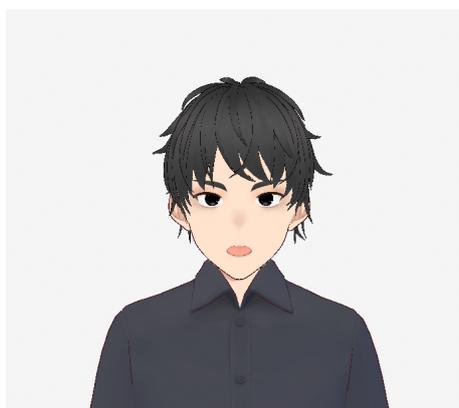


図 B.5: 母音が「お」の場合の口の形

付録 C

視覚負荷課題の単語内容

実験 2 で使用した，視覚負荷課題にて呈示した単語 10 種類を以下に示す．

- キツネ
- コタツ
- スマホ
- モヤシ
- バイト
- ピエロ
- カモメ
- ズケイ
- メイク
- テラス

付録 D

実験 1 のアンケート項目

実験 1 において実験終了後に行ったアンケート項目を以下に示す。D.0.3, D.0.4 において回答時実際に使用した音声を再生したのち、回答を行った。

D.0.1 このアバターは男性に見えますか？

- 見える
- 少し見える
- どちらともいえない
- 少し見えない
- 見えない

D.0.2 このアバターは女性に見えますか？

- 見える
- 少し見える
- どちらともいえない
- 少し見えない
- 見えない

D.0.3 この声は男性に聞こえますか？

- 聞こえる
- 少し聞こえる
- どちらともいえない
- 少し聞こえない
- 聞こえない

D.0.4 この声は女性に聞こえますか？

- 聞こえる
- 少し聞こえる
- どちらともいえない
- 少し聞こえない
- 聞こえない

付録 E

実験 2 のアンケート項目

実験 2 において実験終了後に行ったアンケート項目を以下に示す。

E.0.1 課題の難易度は難しかったか

- 難しい
- 少し難しい
- どちらともいえない
- 少し簡単
- 簡単

付録 F

実験 1 の分析方法の具体例

呈示されたアバター位置に応じた音源位置の具体的な分析例は以下の通りである。

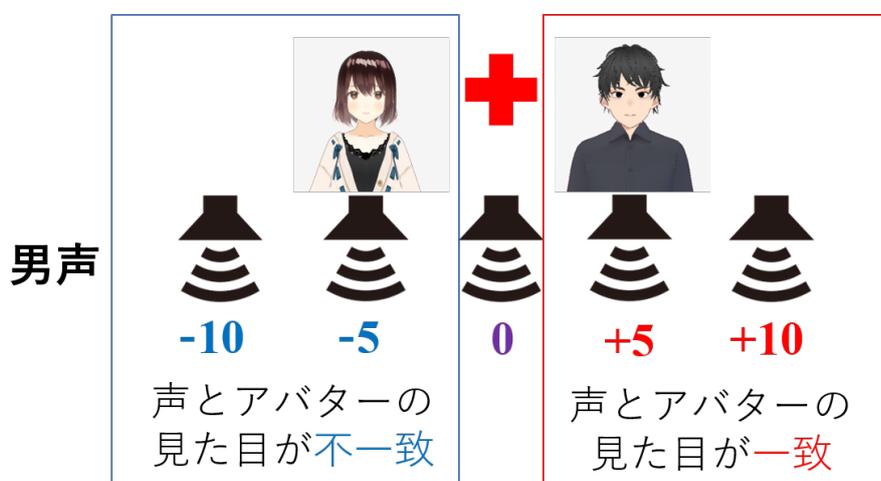


図 F.1: 右側が男性かつ男性の声が呈示されている場合

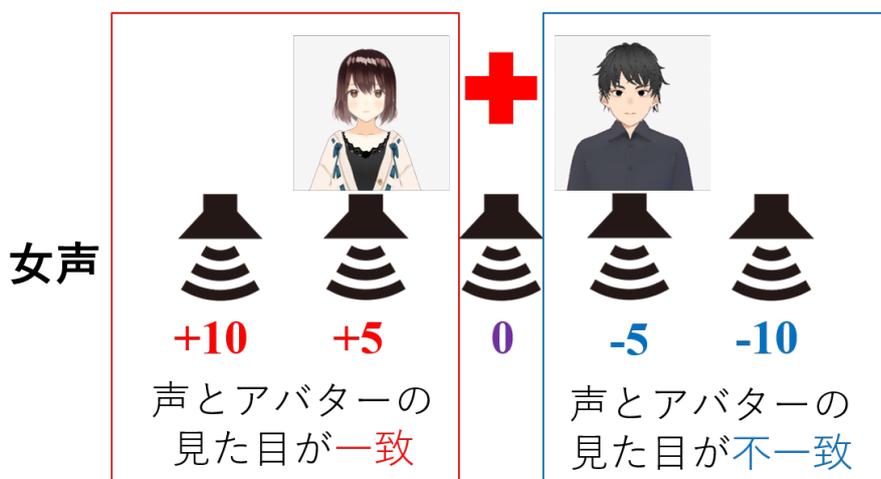


図 F.2: 右側が男性かつ女性の声が呈示されている場合

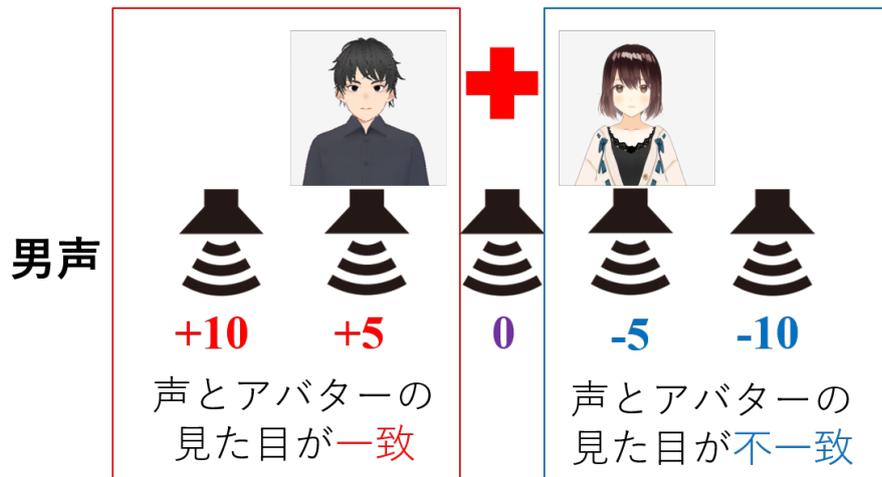


図 F.3: 右側が女性かつ男性の声が呈示されている場合

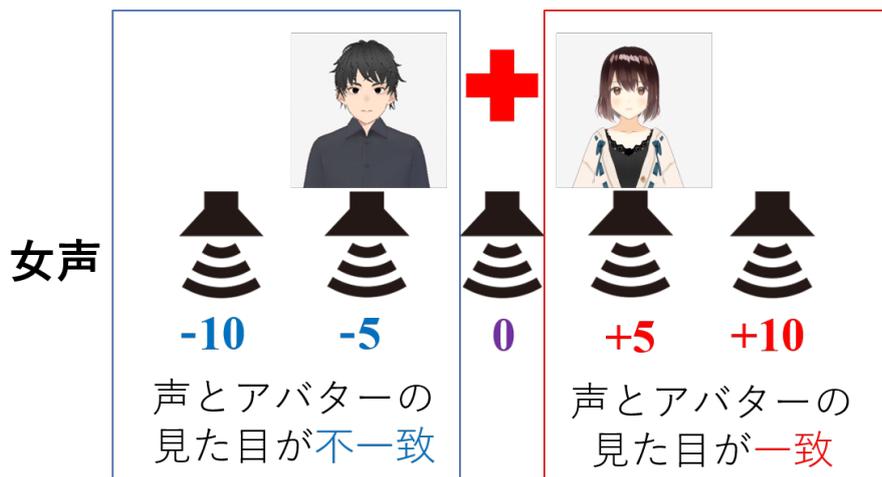


図 F.4: 右側が女性かつ女性の声が呈示されている場合