

アバターの視線方向に応じた話者選択と音声制御手法の検討

柿沼大輝^{†1} 山上 華^{†2} 俵 綾那^{†3} 佐藤 玲美^{†3} 外山 祥平^{†3} 益子 宗^{†2}

概要：複数人が参加するメタバース空間での音声会話において、同時に複数のアバターが発話した場合に音声混在し、対話者の発話内容が聞き取りづらいことや、どのアバターに話しかけられているかが分からないという問題があった。本研究では、アバターの視線方向に応じて発話音量を制御することで発話者の音声を聞き取りやすくなる話者選択手法を提案し、音量の変化量や音声の切り替え速度といった音声制御の度合いが対話体験へ与える影響を調査した。実験の結果、アバターの視線方向に応じて聞こえる音声を素早く大きく切り替えることで発話内容が聞き取りやすくなる傾向がみられ、提案手法の有用性が確認できた。

1. はじめに

近年メタバース（仮想空間プラットフォーム）を利用したコミュニケーションはエンターテインメントからビジネスまで様々な分野で活用されており、今後も利用が拡大することが期待される。メタバースでは遠く離れた相手とも同じ空間を共有して対話を行うことが可能である。しかし、メタバースでのコミュニケーションでは複数人での音声会話を行うと同時に発言による音の混在によって発話者の音声が聞き取りにくく、どのアバターが自分に向かって話しかけているのかが分かりにくくなるという課題がある。そこで、本研究ではメタバースにおける複数人での音声会話にアバターの視線方向に応じて音声の音量を制御することによってより発話者の声が聞き取りやすくなる話者選択と音声制御手法を提案する。また音量の変化量や音声の切り替え速度といった音声制御の度合いが対話体験へ与える影響を調査し、提案手法がメタバースでの対話体験において有効であるか検証した。

2. 関連研究

仮想環境でアバターの視覚イメージからずれた音像の効果に関する研究がある。この研究から、アバターの音像（音の位置）と映像のずれは 50~75%程度ずれても受話者に許容されることが分かっている。また、仮想環境上では相手との対人距離（パーソナルスペース）は視覚的な情報が優位であり、音の位置は対人距離にあまり影響を与えないことが分かっている。[1] このことから、実際にアバターが立っている位置よりも音声が近くで聞こえるような効果を使用しても仮想環境では違和感が少ないと考えられる。これらを踏まえて、本研究ではアバター間の距離により音声の大小を制御する一般的な音響モデルではなく、アバターの視線向きに応じて話者を選択し、対象の発話者の音声が通常よりも少し大きく聞こえるような音声制御を実装した。



(赤と青グループで聞こえる音量が異なる)
図 1 会話グループごとに音量が変化の様子



図 2 音声制御が可能なメタバース環境

3. アバターの視線方向に応じた話者選択

本研究では、図1のようなメタバース上の複数人での音声会話でアバター同士の距離や目線の向きから対話対象を限定して動的に音声制御を行うシステムの開発を目指す。対面会話では話し手は受話者の方向にたびたび視線を向けるが、受話者はそれ以上に話者を注視していることが明らかにされている [2]。このことから発話者、受話者の視線方向を利用することで互いが対話者であることが認識しやすくなると考える。これを踏まえて、発話者と受話者のアバターが向かい合っている際に、発話者の音声が受話者に向けてより鮮明に聞こえるような音声制御を行うことによって受話者がより自分に向かって話しかけられていると感じられる音声制御ができる話者選択機能を Unity 上で開発

^{†1} 芝浦工業大学 大学院 理工学研究科

^{†2} 芝浦工業大学 デザイン工学部

^{†3} 大日本印刷株式会社

した。自身が操作するアバターのカメラ視点方向を画面中央に赤い点で示し、対話者に視線を向けると周囲の音量が下がり、音声が大きく聞こえるように Unity の Audio Mixer の機能を活用して対象のアバターから発せられる音量を制御できる機能を実装した。また、音声制御されている状態では視線を向けている相手に対して「Select」の文字が出るように設定を行った（図 2）。

4. 音声制御による対話体験の評価

3 章にて提案した話者選択手法を用いることで、アバターの視線方向に応じた音声の制御が可能となる。しかし、システムを用いた簡易的な予備調査を事前に行った結果、音声の変化量や切り替えタイミングにより不自然さが発生することが分かった。そのため本章では、音量の変化量や音声の切り替え速度といった音声制御の度合いが対話体験へ与える影響の調査の実験について述べる。まず、4. 1 にて音声制御が対話体験に与える影響を調べ、4. 2 にて音声の変化量による印象の調査、4. 3 にて適切な音声の切り替えタイミングを調べる実験を行った。

4.1 実験 1: 音声制御が対話体験に与える影響の調査

メタバース空間での音声会話における音声制御が対話体験に与える影響を確かめるために、アバターの視線方向に対応した音声制御を行った場合の印象調査を行った。

実験は被験者 4 人に対して二人一組になってもらい、開発したメタバース環境（図 2）で会話を行ってもらい。等距離に配置した円の上にそれぞれアバターを立たせ、アバターに被験者を割り振り目の前の相手（円の色が赤の人は黄色と、青の人は黄緑の相手）と会話を行ってもらい（図 2）。視線方向を示す赤い点は、予めそれぞれのアバターが対話者に視線方向が向いている状態に設定しておく。この状態で以下の 2 回に分けて実験を行った。

- 1 回目：被験者同士の声が 4 人とも全て同じ音量で聞こえる環境で目の前にいるペアの相手と「自己紹介と趣味」の話題について 2 組同時に 5 分間会話を行ってもらい。その後 Google フォームで作成したアンケート 1 に回答してもらう。
- 2 回目：視線方向が向いている目の前のペアの音声より大きく聞こえ、そのほかの音声はほとんど聞こえない環境に変更して「アルバイトと最近出かけた場所のエピソード」について 2 組同時に 5 分間会話を行ってもらい。その後 Google フォームで作成したアンケート 2 に回答してもらう。

4.2 実験 2: 音声の変化量による話しやすさの検討

適切な音声の変化量を調べるために、4 段階の音声制御音量を用意してそれぞれの印象調査を行った。音声制御状



図 3 開発したシステムの UI パネル



図 4 実験の様子

態では、対話者に視線を向けると周りの音量が下がり、対話者の音声が大きく聞こえるような設定を行っている。この時の対話者の音声制御の強さを 4 段階に分けて、どの程度音声制御を行った状態がより話しやすいと感じるかを調査した。4 段階の音声制御の強さの幅は予備実験を数名に行い、それぞれの音声制御の強さが①小、②中、③大、④特大の変化量と被験者が感じるように設定を行った。音声制御が①に近いほど周りの音声聞こえ、④に近いほど音声制御が強く、周りの音声小さく聞こえるような設定となっている。

音声制御に関する設定は UI パネルから変更できるようになっている。（図 3）被験者の操作時には UI パネルは非表示にし、実験に携わる設定を変更する際に利用する。提案手法による音声制御の有無や音声制御の上昇量、減少量、それらの音声制御値に至るまでにかかる変化速度といった項目を実験用メタバース空間にいる状態で変更できる。この実験では Unity の音量ミキサー上での音量の増加 dB 数を 10 dB で固定し、音量の減少 dB 数を 4 段階設定する。予備実験よりそれぞれの値を①0dB、②-5dB、③-10dB、④-15dB とした。

被験者をそれぞれ周囲に誰もいない個室に案内し、実験を行った（図 4）。実験ではラップトップ PC とヘッドホンを用い、視点操作はマウスを使用するものとする。4 体のアバターに対して実験進行者と被験者 3 人を割り当てて二人一組になってもらい、会話を行ってもらい。等距離に配置した円の上にそれぞれアバターを立たせ、目の前の相手（円

表 1 音声制御の有無と対話体験の評価

アンケート1					
評価	会話相手の声は聞き取りやすかったか	会話相手との会話に集中できたか	自分に向かって話しかけられていると感じたか	会話をスムーズに行うことができたか	他のペアの会話音声は気になったか
とてもそう思う	17%	8%	18%	8%	58%
ややそう思う	17%	25%	42%	25%	25%
あまりそう思わない	33%	50%	25%	42%	8%
全くそう思わない	33%	17%	18%	25%	8%

アンケート2					
評価	会話相手の声は聞き取りやすかったか	会話相手との会話に集中できたか	自分に向かって話しかけられていると感じたか	会話をスムーズに行うことができたか	他のペアの会話音声は気になったか
とてもそう思う	83%	67%	67%	67%	8%
ややそう思う	8%	25%	33%	33%	0%
あまりそう思わない	0%	0%	0%	0%	67%
全くそう思わない	8%	8%	0%	0%	25%

表 2 対話体験に関する代表的な意見

アンケート1	・他のグループの声がかぶって相手の声が聞こえない時があったのが気になった。 ・まず自分と話しているのが誰かわからなかった。 ・自分が話している相手のアバターと声を一致させないと、ただ電話してるだけのように感じられた。
アンケート2	・他の2人の声が少し小さく、会話相手の声が大きく聞こえたため話しやすかった。 ・初めよりもオフラインの会話に近かったように感じる。 ・自分が会話に集中できるようになったため、自分が今置かれている環境が気になるようになった。静かな場所だったため話しにくかった。

の色が赤の人は黄色と、青の人は黄緑の相手）とそれぞれの音声制御設定で2分間会話を行ってもらい、以下の手順で実験を進める。（図2）。

手順（1）：4体のアバターに実験進行者と被験者それぞれ割り振った後、音声制御のない状態で操作方法の説明を聞きながら会話を交えてアバターを操作してもらい、自分のアバターと相手のアバターを認識してもらう。この手順による被験者は自分が操作しているアバターの所有感を高めることができる。

手順（2）：音声制御パネルの「視線での音声変化」にチェックを入れてもらい、音声制御が行われている状態に設定してもらう。この時の数値は音量増加値が10、音量減少値が-10となっており、音声制御が行われている状態を体感してもらう。

手順（3）：音声制御音量の値（①，②，③，④）を指定し、指定した設定の制御状態で2分会話を行ってもらう。この会話をしている間に、対話者に対して視線を向けたり外したりして音声制御の強さの度合いを体感しながら会話を行ってもらう。

手順（4）：2分間の会話が終了したら Google フォーム作成したアンケートに回答してもらう。アンケートが終わったら次の音声制御音量の値でまた2分間会話を行ってもらい、終了したら再度アンケートを行う。

手順（5）：4段階すべての音声制御音量を試してもらった後に、それぞれ話しやすかったものを順位付けしてもらう。

4.3 実験3:音声の切り替え速度の検討

適切な音声の切り替え速度を調べるために、4段階の音

声制御速度を用意してそれぞれの印象調査を行った。音声制御音量が指定した上昇値，減少値に達するまでの速度を4段階に分けて、どの速度がより話しやすい環境だと感じるかを調査した。4段階の音声の切り替え速度の幅は事前に予備実験を数名に行い、それぞれの音声制御速度が①遅い，②少し遅い，③少し速い，④速い，の変化速度となるように、1秒ごとの Unity の音量ミキサー上での dB 数の変化量の設定を行った。その数値は①4dB，②9 dB，③14 dB，④19 dB とした。音声制御速度が①に近いほど音声制御音量が指定した値に達するまでの速度が遅く，④に近いほど瞬時に音声が目標の音量に変化する設定となる。なお，この実験では音声制御音量の音量増加値を 10dB，音量減少値を実験2の評価より-15dB に設定した。

被験者をそれぞれ無人の個室に案内し、4.2の実験2と同様の環境で実験を行った。4体のアバターに対して実験進行者と被験者3人を割り当て二人一組になってもらい会話を行ってもらい。等距離に配置した円の上にそれぞれアバターを立たせ、目の前の相手(円の色が赤の人は黄色と、青の人は黄緑の相手）とそれぞれの音声制御設定で3分間会話を行ってもらい、以下の手順で実験を進める。（図2）。

手順（1）：4体のアバターに実験進行者と被験者それぞれ割り振った後、音声制御のない状態で操作方法の説明を聞きながら会話を交えてアバターを操作してもらい、自分のアバターと相手のアバターを認識してもらう。

手順（2）：音声制御パネルの「視線での音声変化」にチェックを入れてもらい、音声制御が行われている状態に設定してもらう。この時の数値は音量増加値を 10dB，音量減少値を-15dB と設定し，音声制御が行われている状態を体感

表 3 音声の変化量と対話体験の評価

音声①0					
評価	会話相手の声は聞き取りやすかったか	会話相手との会話に集中できたか	自分に向かって話しかけられていると感じたか	会話をスムーズに行うことができたか	他のペアの会話音声は気になったか
とても思う	27%	18%	55%	36%	36%
やや思う	46%	36%	9%	36%	46%
あまり思うわない	9%	46%	27%	18%	9%
全く思うわない	18%	0%	9%	9%	9%

音声②-5					
評価	会話相手の声は聞き取りやすかったか	会話相手との会話に集中できたか	自分に向かって話しかけられていると感じたか	会話をスムーズに行うことができたか	他のペアの会話音声は気になったか
とても思う	36%	27%	36%	18%	18%
やや思う	36%	46%	55%	55%	64%
あまり思うわない	27%	27%	9%	27%	18%
全く思うわない	0%	0%	0%	0%	0%

音声③-10					
評価	会話相手の声は聞き取りやすかったか	会話相手との会話に集中できたか	自分に向かって話しかけられていると感じたか	会話をスムーズに行うことができたか	他のペアの会話音声は気になったか
とても思う	55%	46%	64%	46%	9%
やや思う	46%	46%	27%	46%	36%
あまり思うわない	0%	9%	9%	9%	46%
全く思うわない	0%	0%	0%	0%	9%

音声④-15					
評価	会話相手の声は聞き取りやすかったか	会話相手との会話に集中できたか	自分に向かって話しかけられていると感じたか	会話をスムーズに行うことができたか	他のペアの会話音声は気になったか
とても思う	91%	91%	91%	82%	18%
やや思う	9%	9%	9%	18%	0%
あまり思うわない	0%	0%	0%	0%	36%
全く思うわない	0%	0%	0%	0%	46%

してもらふ。

手順 (3) : 音声制御速度の値 (①, ②, ③, ④) を指定し, 指定した設定の制御状態で 3 分会話を行ってもらふ。この会話をしている間に, 対話者に対して視線を向けたり外したりして音声制御速度を体感しながら会話を行ってもらふ。

手順 (4) : 3 分間の会話が終わったら Google フォームで作成したアンケートに回答してもらふ。アンケートが終わったら次の音声制御速度の値でまた 3 分間会話を行ってもらふ, 終了したら再度アンケートを行う。

手順 (5) : 4 段階すべての音声制御速度を試してもらった後に, それぞれ話しやすかったものを順位付けしてもらふ。また, 一番不快に感じた音声の変化速度を①~④の中から一つ選んでもらい, その理由を回答してもらふ。

5. 実験結果

5.1 実験 1 の結果

男性 2 人, 女性 10 人の大学生計 12 人の被験者に実験を行った。実験の結果, 音声制御を行った場合 (アンケート 2 の結果) の方が総じて評価が良くなる傾向があることが分かった (表 1)。特に, 「自分に向かって話しかけられていると感じたか」「会話をスムーズに行うことができたか」の項目では音声制御前 (アンケート 1 の結果) では評価にばらつきがみられたが, 音声制御後の評価では「とてもそ

う思う」「やや思う」に評価が集中した。これらのアンケート結果から, 音声を制御することでメタバース上での対話体験の向上に一定の効果があることが分かった。

また, それぞれのアンケートから得られた代表的な意見を表 2 に記す。音声制御前 (アンケート 1 の結果) では他のペアの会話音声に気になるなどの会話のしにくさについての意見が多くみられた。音声制御後 (アンケート 2 の結果) では相手の声が聞き取りやすく話しやすいといった肯定的な意見が見られる一方, 他のペアの声が小さく静かになったことによって現実世界の自分が置かれている環境が気になってしまったという意見が見られた。

5.2 実験 2 の結果

男性 6 人, 女性 5 人の大学生計 11 人の被験者に実験を行った。実験の結果, 音声制御が強くかかるほど総合的に評価が良くなる傾向があることが分かった (表 3)。特に「対

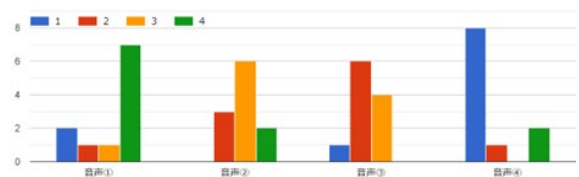


図 5 音声の変化量と話しやすさの順位

表 4 音声の変化速度と対話体験の評価

速度①					
評価	音声制御による音の変化を感じたか	音声制御による音量の変化にストレスを感じたか	会話相手との会話に集中できたか	会話をスムーズに行うことができたか	自分に向かって話しかけられていると感じたか
とても思う	44%	11%	33%	22%	67%
やや思う	22%	22%	33%	56%	11%
あまり思わない	22%	22%	33%	22%	22%
全く思わない	11%	44%	0%	0%	0%

速度②					
評価	音声制御による音の変化を感じたか	音声制御による音量の変化にストレスを感じたか	会話相手との会話に集中できたか	会話をスムーズに行うことができたか	自分に向かって話しかけられていると感じたか
とても思う	56%	11%	33%	22%	56%
やや思う	33%	44%	56%	78%	44%
あまり思わない	11%	22%	11%	0%	0%
全く思わない	0%	22%	0%	0%	0%

速度③					
評価	音声制御による音の変化を感じたか	音声制御による音量の変化にストレスを感じたか	会話相手との会話に集中できたか	会話をスムーズに行うことができたか	自分に向かって話しかけられていると感じたか
とても思う	33%	11%	67%	78%	67%
やや思う	44%	33%	33%	22%	33%
あまり思わない	11%	33%	0%	0%	0%
全く思わない	11%	22%	0%	0%	0%

速度④					
評価	音声制御による音の変化を感じたか	音声制御による音量の変化にストレスを感じたか	会話相手との会話に集中できたか	会話をスムーズに行うことができたか	自分に向かって話しかけられていると感じたか
とても思う	78%	11%	67%	56%	67%
やや思う	11%	22%	11%	44%	33%
あまり思わない	0%	33%	22%	0%	0%
全く思わない	11%	33%	0%	0%	0%

話者の声は聞き取りやすかったか」「対話者との会話に集中できたか」「自分に向かって話しかけられていると感じたか」の項目は音声制御が強かかればかかるほど評価が良くなっていく傾向が見られた。また、①～④の音声制御音量の順位付けを図5に示す。周りの音声小さく聞こえる④が最も話しやすいと感じる設定で、①の音声制御が最も弱い状態が一番好まれない結果となった。

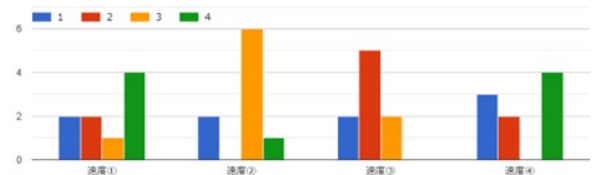


図 6 音声の変化速度と話しやすさの順位

5.3 実験3の結果

男性4人、女性5人の大学生計9人の被験者に実験を行った。実験の結果、「音声制御による音の変化を感じたか」という項目において、音声変化速度が速いほど音声制御による音の変化を感じられるようになることが分かった（表4）。また「対話者との会話に集中できたか」「会話をスムーズに行うことができたか」の項目では速度③が最も評価が良くなることが分かった。また①～④の音声制御速度の順位付けを図6に示す。音声変化速度の話しやすさの評価にはばらつきがあるものの、話しにくく感じるものとして速度①と速度②が多く選ばれる結果となった。4パターンの音声制御速度で一番不快に感じた音声制御速度とその理由を表5に示す。速度①が不快な理由として話し相手に視線を向けているのに音声制御がかかりきるまでの反応が遅く、違和感があったという意見が多かった。速度④が不快な理由として音声の切り替わりがはっきりと分かりやすいため違和感があったという意見が多かった。

6. 考察

実験1の結果から、提案手法による音声制御によって、メタバースにおける会話の混線を防ぎ、会話に集中しやす

表 5 一番不快に感じた音声の変化速度とその理由

速度① 5人	
不快に感じた理由	・音声制御の効果をあまり感じなかった ・変化を感じにくい ・反応が遅かったため ・ゆっくり変化するので違和感があった ・速度がゆっくり過ぎる
速度④ 4人	
不快に感じた理由	・Selectから離れた時、どの人のこえかわからなくなるため ・異次元に飛ばされた気がした ・視線を外した時に周りの声がすぐに大きくなった気がした ・変化が気持ち悪かった

い環境がつくられることが示唆された。また、実験2、実験3の結果から、音声制御の上昇値、減少値の幅は大きく、はっきりと音量変化させる方がユーザーにとって好ましいことがわかった。一方で、音量変化の速度は遅すぎず、早すぎず、速度②や速度③のように緩やかに変化の方が好ましいことがわかった。しかし、双方の結果には傾向がみられたものの、それぞれ個人差があった。

今回の実験は被験者に会話に集中できる静かな個室を用意したため、音声制御が強かかった状態でも現実世界の環境を気にせず会話を行うことができた。しかし、実際は周囲の環境音が聞こえる場合もあり、それらはメタバース上での音声制御の評価にも影響されることが考えられる。これらのことから、音声制御の音量変化の大きさや早さに

関しては、一定に固定するのではなく、不快感のない範囲内で個人ごとに設定できる方が好ましいユーザー体験がつくることができる考える。

7. おわりに

本稿では複数人が参加するメタバース空間での音声会話において、同時に複数のアバターが発話した場合に音声混在し、対話者の発話内容が聞き取りづらいことや、どのアバターに話しかけられているかが分からないという従来の問題に対し、アバターの視線により対象の音声の音量を制御できる話者選択手法を提案した。また実験では、音量の変化量や音声の切り替え速度といった音声制御の度合いが対話体験へ与える影響の調査を行い、メタバース上での音声会話における話者選択による音声制御は、対話体験を向上する効果があることがわかった。一方で、メタバース上での音声制御音量に関しては現実世界の体験環境によって会話のしやすさが左右されるという意見もみられた。また、音量の変化速度に関しては、話しやすいと感じる制御速度に個人差はあるものの、極端に早いもしくは遅い場合に不快に感じるという傾向がみられた。

今後は得られた実験結果を元に、ユーザーに使いやすい音声制御の設定を思案していくことが必要である。例えば、今回は正面の固定された位置関係の相手に対して会話をする実験を行ったが、実際のメタバース空間のように、複数人の複雑な位置関係においても音声制御が有効か検討していく。また、音声制御システムもアバター一体を選択する以外に、画面中央から画面端に向かうにつれて段階的に聞き取りやすい音量変化をするなど、よりユーザーが直感的に使用できる変化方法、操作方法を検討していきたい。

参考文献

- [1] K. Zempo, A. Yamazaki, N. Wakatsuki, K. Mizutani and Y. Okada, "Mouth-in-the-Door: The Effect of a Sound Image of an Avatar Intruding on Personal Space That Deviates in Position From the Visual Image, " in IEEE Access, vol. 10, pp. 125772-125791, 2022.
- [2] Clark, H. H. :Using Language. Cambridge:Cambridge University Press(1996)