

卒業論文

メタバース上の多言語音声コミュニケーションによる プロテウス効果の検証

指導教官 村上 陽平 教授

立命館大学 情報理工学部
先端社会デザインコース
2600210054-0

岩野 智総

2024 年度（秋学期）卒業研究 3（CH）
令和 7 年 1 月 31 日

メタバース上の多言語音声コミュニケーションによるプロテウス効果の検証

岩野 智総

内容梗概

メタバースは、現実世界の制約にとらわれずに活動できるという特性を持つ。例えば、ユーザはアバターを使用することで、その大きさや服装などの外見的特徴を自由に変更して活動ができる。また、メタバースはインターネットを介してどこからでもアクセス可能であり、翻訳機能を活用することで容易に多言語でのコミュニケーションが行える。このような特徴から、メタバースはグローバルな活動のコミュニケーションプラットフォームとして期待されている。したがって、メタバースの特性がユーザやコミュニケーションに与える影響について理解することは重要である。その代表的な影響が「プロテウス効果」である。プロテウス効果とは、アバターの特徴がユーザの行動や思考に影響を与えるという心理的効果を指し、多くの研究がなされている。しかしながら、これらの研究の多くは外見的特徴に焦点を当てており、使用する言語による影響については明らかにされていない。一方、現実世界では、言語を変換することで、その言語の構造や文化的背景によって話者の性格特性が変化することが知られている。この現象はメタバース上での言語変換でも生じる可能性がある。

そこで本研究では、メタバース上の多言語音声コミュニケーションにおける言語の変換によりプロテウス効果が生じるかを検証する。具体的には、音声認識、機械翻訳、音声合成の技術を組み合わせて、ユーザを擬似的にマルチリンガル化する多言語音声コミュニケーション環境を構築する。その後、メタバース上で交渉タスク実験を行い交渉態度と性格特性の変化を測定する。本手法の実現にあたり、取り組むべき課題は以下の2点である。

翻訳合成音声に対する身体化感覚

プロテウス効果は、ユーザがアバターを自分の身体のように感じることで引き起こされる。したがって、言語の変換によるプロテウス効果を引き起こすためには、ユーザが翻訳合成音声での会話を自分の行為として認識できる必要がある。

翻訳合成音声による性格特性の変化の測定

言語の変換によるプロテウス効果の発生を評価するには、ユーザの行

動や思考の変化を定量的に測定する必要がある。

1つ目の課題に対しては、合成音声のフィードバック、および音声認識結果と翻訳後のテキストの表示を行った。自分がその行為を行なっていると感じる感覚は「行為主体感」と呼ばれ、自分の行動と予測された結果が一致したときに生じる。行為主体感を付与するために、合成音声と発話に合わせてテキストを表示するという一定のフィードバックを与えることを行なった。2つ目の課題に対しては、メタバース上の多言語音声コミュニケーション環境を用いて、日本語話者と中国語話者間で最後通牒ゲームを実施し、利他性を評価した。最後通牒ゲームは経済学やゲーム理論における実験で、個人の利他性を測定できる。日本語での会話と合成音声を使用した中国語での会話のそれぞれにおいて、ユーザの分割額の平均値を調べ、比較した。また、性格の評価には、5つの性格特性（神経症傾向、外交性、経験への開放性、誠実性）を測定するビッグファイブ性格特性診断を用いた。この診断をそれぞれ会話の後に実施し、数値を比較することで性格特性の変化を評価した。

翻訳合成音声に対する身体化感覚

実験後のアンケートにおいて、8人中5人の被験者から「翻訳合成音声で話している感覚を得た」という回答が得られた。また、全ての被験者から「音声認識結果と翻訳後のテキストの表示が自分の行為の結果であると感じた」という回答が得られた。これらの結果から、ユーザに行為主体感を付与できたと判断した。

翻訳合成音声による性格特性の変化の測定

最後通牒ゲームでは、日本語での会話に比べ、合成音声を使用した中国語での会話において提示額が増加する傾向がみられた。この結果が統計的に有意な差であるかを対応のあるt検定で調べたところ、1回目では「 $p=.07$ 」、2回目では「 $p=.0.31$ 」という結果となった。また、効果量d（Cohen's d）を測定したところ1回目では「 $d=1.05$ 」、2回目では「 $d=0.55$ 」という結果となった。これらの結果から、合成音声を使用した中国語での会話をした場合、交渉タスクにおいてより強気になる可能性が示唆された。一方、ビッグファイブテストの結果では、全ての性格特性において微量な変化はあったものの、有意とよべる差はみられなかった。

Examining the Proteus Effect in Multilingual Voice Communication in the Metaverse

Tomoaki Iwano

Abstract

In the metaverse, individuals can engage in activities beyond the constraints of the physical world. For example, by using avatars, one can freely change their physical characteristics. Additionally, the metaverse is accessible from anywhere via the internet, and multilingual communication can easily be facilitated through translation functions. Due to these features, the metaverse is expected to serve as a global communication platform for various activities. Therefore, understanding the impact of the metaverse on users and communication is crucial. A key influence in this context is the "Proteus effect," which refers to the psychological effect in which an avatar's characteristics influence the user's behavior. Many studies have been conducted on this effect. However, most of these studies have focused on visual characteristics, with little attention given to the influence of language. In contrast, in the real world, it is known that changing languages can alter a speaker's personality traits. This study aims to examine whether the Proteus effect occurs in the metaverse as a result of language change, leading to changes in the speaker's personality traits. Specifically, by combining speech recognition, machine translation, and speech synthesis technologies, a multilingual voice communication environment will be constructed. Then, negotiation task experiments will be conducted in the metaverse to investigate the occurrence of the Proteus effect. To achieve this, the following two key challenges must be addressed:

Embodiment Sensation Toward Synthetic Speech

To induce the Proteus effect through language conversion, users must be able to perceive conversations conducted with synthesized speech as their own actions.

Measuring Personality Trait Changes Induced by Synthetic Speech

To evaluate the occurrence of the Proteus effect due to language conversion, it is necessary to quantitatively measure changes in the user's behavior. In particular, it is important to assess whether changes in

personality traits occur as a result of language switching, as seen in bilingual speakers.

For the first challenge, we provide feedback consisting of synthesized speech, speech recognition results, and translated text. The sensation of perceiving one's actions as their own is called "agency." To provide this sense of agency, we make the synthesized speech audible to the user and display the text. For the second challenge, we conduct an ultimatum game between Japanese and Chinese speakers. The ultimatum game, a common experimental tool in economics, can measure an individual's altruism. We examine and compare the average amounts offered by users in Japanese voice conversation and Chinese synthesized voice conversation. Additionally, we use the Big Five personality traits assessment to evaluate personality. This assessment was administered after both the Japanese voice conversation and the Chinese synthesized voice conversation. **Embodiment Sensation Toward Synthetic Speech**

In the post-experiment questionnaire, five out of eight participants report that they "had the sensation of speaking through the translated synthesized voice." Additionally, all participants respond that they "felt that the displayed speech recognition results and translated text were the outcome of their own actions." Based on these results, it is determined that a sense of agency was successfully imparted to the users.

Measuring Personality Trait Changes Induced by Synthetic Speech

In the ultimatum game, there was a tendency for the offered amounts to increase during conversations using synthetic speech. A t-test is conducted on this result, yielding ' $p = .07$ ' for the first trial and ' $p = .31$ ' for the second trial. Additionally, when measuring Cohen's effect size (d), the results are ' $d = 1.05$ ' for the first trial and ' $d = 0.55$ ' for the second trial. These findings suggest that the use of synthetic speech in Chinese may lead to a more assertive behavior. On the other hand, the results of the Big Five personality test show no significant differences in any of the personality traits.

メタバース上の多言語音声コミュニケーションによるプロテウス 効果の検証

目次

第1章	はじめに	1
第2章	関連研究	3
2.1	プロテウス効果	3
2.2	言語による自己意識の形成	3
2.3	学術的な問い	4
第3章	多言語音声コミュニケーション環境	6
3.1	システム要件	7
3.1.1	3D 空間での一人称視点	7
3.1.2	マルチプレイおよび音声会話	8
3.1.3	音声間翻訳会話	8
3.1.4	テキストの表示	9
3.2	システム構成	9
3.3	ユーザインターフェース	11
第4章	実験	13
4.1	実験概要	13
4.2	最後通牒ゲーム	13
4.3	実験参加者	13
4.4	実験手順	14
4.5	指標	15
4.5.1	身体化感覚	15
4.5.2	人格特性	15
4.5.3	交渉態度	16
第5章	結果	17
5.1	交渉態度への影響	17
5.2	人格特性への影響	18
5.3	身体化感覚の評価	19

第6章 考察	21
6.1 身体化感覚の獲得.....	21
6.2 交渉態度の変化	21
6.3 人格特性の変化	21
第7章 おわりに	23
7.1 実験の拡張	23
7.2 メタバース環境の改善	24
謝辞	25
参考文献	26

第1章 はじめに

個人の行動はどのように決定されるのか. 行動の意思決定に影響を及ぼす要素は性格, 経験, 社会的環境など多岐にわたる. 意思決定の仕組みを明らかにすることは, 政治, 教育, ビジネスなど多くの分野において有用であり, これまでさまざまな研究が行われてきた.

メタバースにおいても, ユーザの意思決定や行動の要因を明らかにする研究が行われている. その代表的な研究が「プロテウス効果」に関する研究である. プロテウス効果とは, ユーザの行動がデジタル上の表現に従って変化するという心理効果のことを指す. 言い換えると, ユーザの自己（自分たらしめる要素）をデジタルで表現することで, 仮想的な自分を作り出し, その結果, それに合わせて行動が変化するという心理効果である. プロテウス効果の代表例としては, 「アバター」が挙げられる. 例えば, 力強い外見のアバターを使用したユーザは, その外見に影響されて自信に満ちた行動をするように変化する. このプロテウス効果の研究はこれまで, アバターの身長や服装, 肌の色といった外見的特徴に着目したものが中心であった. しかしながら, メタバースでデジタルで表現できる要素は外見だけに限られない. 音声もそのうちの一つである. 音声は, 自己を形成する要素の一つであり, 私たちは普段, 発話と同時に自身の声を聞くことで, その声によって自分らしさを感じている. その音声が進化メタバース上で変換された場合, 変換後の合成音声の特徴に従って発話者の行動が変化する可能性がある. 合成音声を使うことで変換可能な要素は音の高さや速さなどがあるが, 本研究では特に言語の変換に着目する.

言語社会学の分野においては, 言語が社会的, 文化的アイデンティティを形成する上で重要な役割を果たすことがよく知られている. つまり, メタバース上で言語の変換は仮想的な自己を作り出すことであると捉えられる. 現実世界において, 言語の変換によって生まれる性格への影響はすでに明らかになっており, バイリンガルが話す言語を切り替えた際の性格特性の変化を調査した研究では, 話す言語の特徴や文化的背景に応じて性格特性が切り替わるという現象が知られている [1]. このような現象は, メタバースで仮想的に言語を切り替えた状況でも生じる可能性がある.

現在, メタバースはビジネス, フィットネス, ゲームなど様々な分野で新たなプラットフォームとして注目されており, その需要は世界的に増加している. さら

に、インターネットを介して世界中からアクセス可能なメタバースでは、グローバルな活動が今後さらに増加すると考えられる。グローバルな活動において、多言語コミュニケーションは重要な役割を果たす。そのような多言語コミュニケーションにおいて、言語が与えるユーザへの影響を明らかにすることには大きな意義がある。

そこで本研究では、メタバース上の多言語音声コミュニケーションにおける言語の変換によりプロテウス効果が生じるかを調査する。具体的には、日本語と中国語間で発話音声の言語を対話相手の言語の合成音声に変換すること（以下、音声間翻訳と呼ぶ）が可能な多言語音声コミュニケーション環境を構築する。その後、交渉タスクおよび性格特性診断テストによって、音声間翻訳が与えるユーザの行動への影響を計測する。それらの結果から、メタバースで合成音声での言語の変換によるプロテウス効果の発生を検証する。本手法の実現にあたり、取り組むべき課題は以下の2点である。

翻訳合成音声の身体化感覚

プロテウス効果は、ユーザがアバターを自分の身体のように感じることで引き起こされる。したがって、言語の変換によるプロテウス効果を引き起こすためには、ユーザが翻訳合成音声での会話を自分の行為として認識できる必要がある。

翻訳合成音声による性格特性の変化の測定

言語の変換によるプロテウス効果の発生を評価するには、ユーザの行動や思考の変化を定量的に測定する必要がある。

以下、2章では、プロテウス効果を検証するにあたり、その基盤となる研究を説明する。次に、3章では作成した多言語音声コミュニケーション環境のシステム構成について説明する。次に4章では、実施した実験の内容を説明する。続いて、5章ではその実験結果を説明する。そして、6章において実験結果の考察を述べ、7章で今後の展望やを述べて結論とする。

第2章 関連研究

本章では, 多言語音声コミュニケーションの変換によるプロテウス効果を検証するにあたり, その基盤となる研究について述べる.

2.1 プロテウス効果

プロテウス効果は Yee らによって初めて提唱された [2]. Yee らは, アバターの外見がユーザの行動に与える影響について研究し, 顔が魅力的なアバターを使用したユーザは, 魅力的ではないアバターを使用したユーザと比べて, 他者との距離を縮め, 自信を持って行動したことを報告している. また, 背の高いアバターを使用したユーザは, 背の低いアバターを使用したユーザと比べて, 交渉タスクにおいてより自信を持って行動したことを報告している. これらの結果から, ユーザの行動はデジタル上の自己表現に影響を受けるとして, この心理的效果をプロテウス効果と呼称した.

プロテウス効果は自己表現の変換が中核をなす要素であり, 本研究においてプロテウス効果を扱うには「自己」とは何かを明らかにする必要がある. すなわち, 自己意識の構造について理解する必要がある. そのために, Gallagher のミニマル・セルフという概念を説明する. Gallagher は, 「最小限の自己」として, 自己主体感 (self-agency) と自己所有感 (self-ownership) によって構成されるミニマル・セルフ (minimal self) を提唱した [4]. ミニマル・セルフは自己の最小の単位であり, 最も基本的な自己意識とされている. ミニマル・セルフは身体的側面に限られないものの, メタバース分野では身体の所有感が着目されてきた. そのため, 操作する体を自分のものと感じる身体所有感 (sense of embodiment), その行為の主体が自分であると感じる行為主体感 (sense of agency), 自身がその位置に存在していると感じる自己位置感覚 (sense of self-location) が自己意識の構成とされることが多く, これらは総称して「身体化感覚」と呼ばれている. よって, プロテウス効果はデジタル上の自己に対して身体化感覚を得ることで生じる効果と言い換えられる.

2.2 言語による自己意識の形成

多言語音声コミュニケーションによるプロテウス効果の検証は, 言語の変換によって話者の行動が変化するかの検証であり, その検証には言語が自己を構成す

る一部であるということが前提となる。そのため、言語と自己の関係について触れる必要がある。普段の思考や会話において言語を使用していることから、言語が自己を形成する一部であるということは自明であるが、その概念は18世紀ごろから学問として存在している。その中でも基盤をなすものが言語的相対論（サピア・ウォーフの仮説）である。Whorfは言語が思考や行動に対してどのような影響を与えるかについて議論し、英語と他の言語で時間の捉え方が異なるという例を挙げ、言語が話者の習慣的な思考の枠組みを形成していると述べた[3]。この研究は、言語が話者の認知や世界観に影響を及ぼしているということを示した。

言語が自己を形成する一部であると認められる場合、言語の変換によってどのような影響が生じるのかを理解し、検証する必要がある。現実世界において、バイリンガルの人々が言語を切り替えるときに、その言語圏の話者の特徴に沿って性格を変化させる現象は、文化的フレーム・スイッチングとして知られている。Ramírez-Esparzaらの研究では、英語話者とスペイン語話者に対して、Big Five Inventory (BFI) と呼ばれる性格特性を数値で測定するテストを行い、それぞれの言語における話者の性格特性の特徴を調査した。その後、英語とスペイン語のバイリンガルに対して、それぞれの言語でBFIを実施し、話す言語ごとの性格特性の数値を測定した。その結果、バイリンガルはその言語話者の性格特性の特徴に合わせて性格特性が変化すること明らかにした。この研究は、言語が個人の自己認識や行動に影響を与える可能性を示唆した。

2.3 学術的な問い

ここまで、プロテウス効果はデジタル上で自己表現を変換することで身体化感覚を獲得し生じるものあり、また、言語が自己を表現する一部であるということについて述べた。では、メタバースで言語を変換した場合、変換語の音声に身体化感覚が得られ、プロテウス効果は発生するのだろうか。そこで本研究では「翻訳合成音声に身体化感覚が得られるのか？」、また身体化感覚が得られた場合、「メタバース上での言語変換でプロテウス効果は発生するのか？」という2つの問いについて明らかにする。

1つ目の問いについて、身体化感覚のうち、生じる可能性のあるものとしては「合成音声での会話（発話）に対する行為主体感」が挙げられる。行為主体感とは、自分がその行為をしている主体であるという感覚を指し、その自覚の有無に関わらず、我々が行う全ての日常的な行為において生じている。行為主体感は、予

測された結果と実際の結果が一致することで生じ、それらの予測のずれに大きく依存することが知られている [5]. 例えば、部屋の照明を点灯させるという行為では、「スイッチを押すことで照明が点灯する」という予測と、「スイッチを押したことにより照明が点灯した」という実際の結果が一致することで行為主体感が生じる。では、発話行為において我々はどのように行為主体感を得ているのか。発話は口や喉の運動であり、発せられた声には自分らしさを感じられ、我々はその声を聞くことで行為主体感を得ている。発話における行為主体感については、自分らしい声を聞くことが行為主体感を高めることが知られており、音声の高さや音声聞こえるまでの時間を変化させて行為主体感への影響を調べた研究では、通常条件において行為主体感が強く感じられることが報告されている [6]. これは、自分自身の声に対して行為主体感を強く感じるという当たり前のことではあるが、その当たり前の状況を再現することが本研究の課題となる。

2つ目の問いについて、プロテウス効果の発生の検証には、交渉タスクにおける交渉態度の変化を観察する。交渉タスクでは最後通牒ゲームを行い、変化の指標として話者が自身に割り当てる金額の差を比較する。また、本章で詳しく述べたように、バイリンガルの言語切り替えと同様にメタバース上での言語変換においても話者の性格特性が変化すると仮定し、音声間翻訳の有無による性格特性の変化を観察する。性格特性の変化の測定には BigFive テストを行い、5つの性格の程度を表すスコアの比較を行う。

第3章 多言語音声コミュニケーション環境

本研究で作成した多言語コミュニケーション環境は,Meta 社の VR ヘッドセットである Meta Quest 3 向けに開発した Android アプリケーションである. 本アプリケーションは,異なる言語を話すユーザ間でのマルチプレイ環境を想定し,ユーザの発話に対するフィードバックに重点を置いて設計した.

本アプリケーションの概要を説明する. 本アプリケーションでは,ユーザは VR ヘッドセットを装着し,メタバースに生成されたルームに入室する. メタバース内ではユーザは1人称視点で自由に移動することができる. また,コントローラーのボタンを押下することで,自身の声を録音し,その音声を他のユーザに送信できる. この音声通信は通常の音声会話と音声間翻訳を用いた会話の2種類が用意されており,押下するボタンを切り替えることで,2種類の会話を切り替えることができる.

通常の音声会話では,ユーザの発話内容をそのまま対話相手に配信する. この際,ユーザの視界には,自身の発話内容とそれに対応する翻訳結果のテキストが表示される(図1).

音声間翻訳を用いた会話では,ボタンの押下を感知すると,まずユーザの発話音声を音声認識によりテキストデータに変換する. 次に,そのテキストデータを機械翻訳によって対話相手の言語に翻訳する. 最後に,翻訳されたテキストデータを音声合成により音声データに変換し,対話相手に送信する. この際ユーザの視界には発話テキストと翻訳テキストが表示される. また,発話内容を基に生成された合成音声は発話者も聞くことができる(図2).

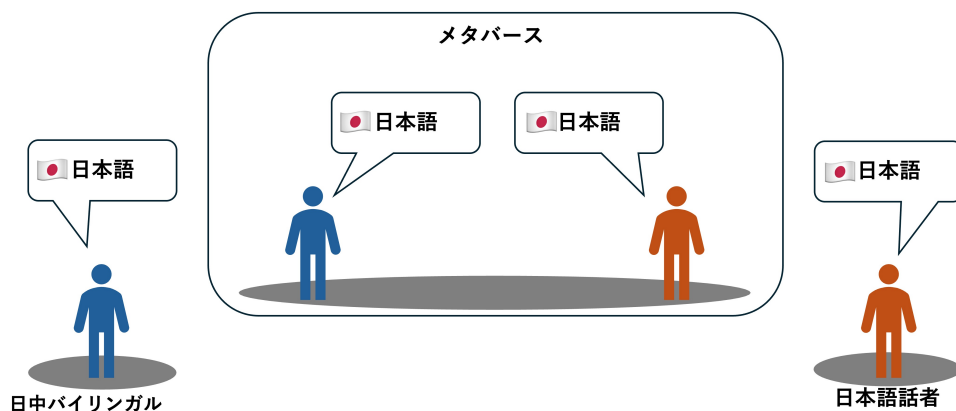


図1: 通常の音声会話（日本語）の例

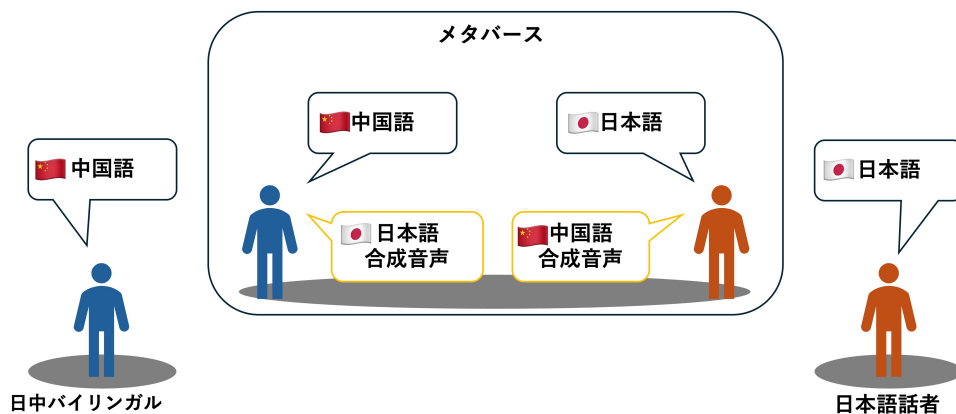


図 2: 日中間の音声間翻訳会話の例

3.1 システム要件

作成するメタバースアプリケーションの機能要件は次のとおりである。

- 3D 空間の 1 人称視点での移動
- マルチプレイ
- 通常の音声会話
- 音声間翻訳会話
- テキスト表示

以下で要件を満たすために使用した各モジュールについて詳しく説明する。

3.1.1 3D 空間での一人称視点

本アプリケーションは没入型のメタバースであるため、3D 空間をアバターを使用して一人称視点で移動できるように設計する必要がある。3D の Android アプリケーションを開発するプラットフォームとして Unity を使用した。Unity は公式に運営するオンラインマーケットプレイスである Unity Asset Store を通じて、ゲームやアプリ開発に役立つ多種多様な素材やツールを提供しており、比較的容易にメタバースを作成できる。

本アプリケーションで使用するアバターは、その Unity Asset Store で公開されている無料のアセットを採用した (図 3)。すべてのユーザはこのアバターを利用してメタバース空間を移動する。アバターを使用する際は、その外見によるプロテウス効果の影響に配慮する必要がある。今回、メタバース内には鏡が配置されておらず、ユーザが自身のアバターの外見を直接確認することはできない。そのため、ユーザ自身のアバターの外見によるプロテウス効果の影響は小さいと考えらるが、対話相手のアバターの外見は確認できるため、その影響を軽減する

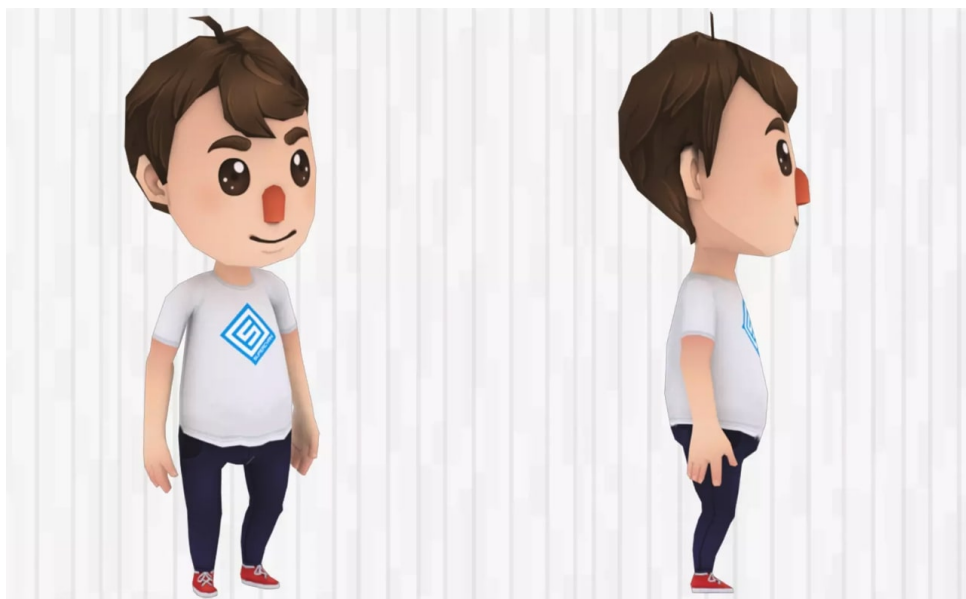


図 3: 使用したアバター [7]

目的で, アバターは人型であり, 表情や服装などが平凡なデザインのものを選択した.

3.1.2 マルチプレイおよび音声会話

マルチプレイおよび音声会話には, Photon が提供する Photon Voice 2 を使用した. Photon Voice 2 は, Unity アプリケーションに高品質なボイスチャット機能を簡単に追加できる SDK である. Photon Voice 2 には, 認証やマッチメイキング, 高速通信などの機能が備わっている. Photon Voice 2 によってユーザはルームに参加し, 音声データを送信するストリームを作成する. このストリームは, ルーム内の他のすべてのユーザに受信されるよう設定した.

3.1.3 音声間翻訳会話

音声翻訳会話を実現するためには, 音声認識, 機械翻訳, 音声合成のサービスを組み合わせる必要がある. 本アプリケーションでは, Google Cloud が提供する「Cloud Speech-to-Text(STT) API」, 「Cloud Translation API」, 「Cloud Text-to-Speech(TTS) API」を使用した.

Cloud STT API は, 125 を超える言語に対応しており, さまざまな地域と言語の音声をテキストに変換できる. また, 短時間の音声クリップ, 長時間の音声クリップ, ストリーミング音声の入力にも対応している. 本アプリケーションでは, エンコーディング形式が LINEAR16, サンプルレートの 16,000Hz, モノラル

音声の短時間の音声クリップを入力として使用した。

Cloud Translation API は、テキストを動的に翻訳するための RESTful API であり、数千の言語ペアでのテキスト翻訳が可能である。

また、Cloud TTS API は 40 を超える言語と方言に対応し、220 種類以上の音声から選択可能である。音声合成には Google の機械学習技術が活用されており、ニューラルネットワークアルゴリズムを用いることで、自然な発音の音声を合成することができる。

リアルタイムの音声翻訳会話を実現するには、ストリーミングでの翻訳処理が理想である。しかし、Cloud Translation API はストリーミング処理に対応しておらず、また Cloud TTS API はストリーミング入力に英語のみ対応しているため、本アプリケーションでは短い音声クリップを逐次処理する設計を採用した。

本研究においては日本語と中国語間の会話を想定しアプリケーションを作成したが、本アプリは認識言語を追加することによる、多言語拡張が可能である。

3.1.4 テキストの表示

発話テキストはユーザが自身の発話内容が正しく認識され、翻訳されているかを確認できるようにするために表示を行った。翻訳テキストは、ユーザの合成音声での発話に対する行為主体感を高めることを目的に表示を行った。発話テキストはユーザの 1.5 メートル前方の左側に表示され、翻訳テキストは右側に表示される。これらのテキストは、音声認識および翻訳が行われるたびに内容が随時更新される (図 4)。

3.2 システム構成

まず、メタバースへの入室処理について説明する。入室処理ではアプリケーションを起動すると最初に Photon サーバーに接続される。Photon サーバーへの接続が完了すると、指定されたルームへの入室が試みられる。ルームが存在しない場合は新しいルームが生成され、ユーザはそのルームに入室する。ルームへの入室が成功すると、ユーザのアバターがインスタンス化され、メタバース内に表示される。

次に音声処理について説明する。音声処理では、コントローラーの指定のボタンが押下されると録音を開始される。録音はプッシュアンドトーク方式を採用し、ボタンの押下後に 10 秒経過するか、ボタンが離されることで録音が停止する。録音停止後、録音データから AudioClip と呼ばれる Unity で扱う形式の音声

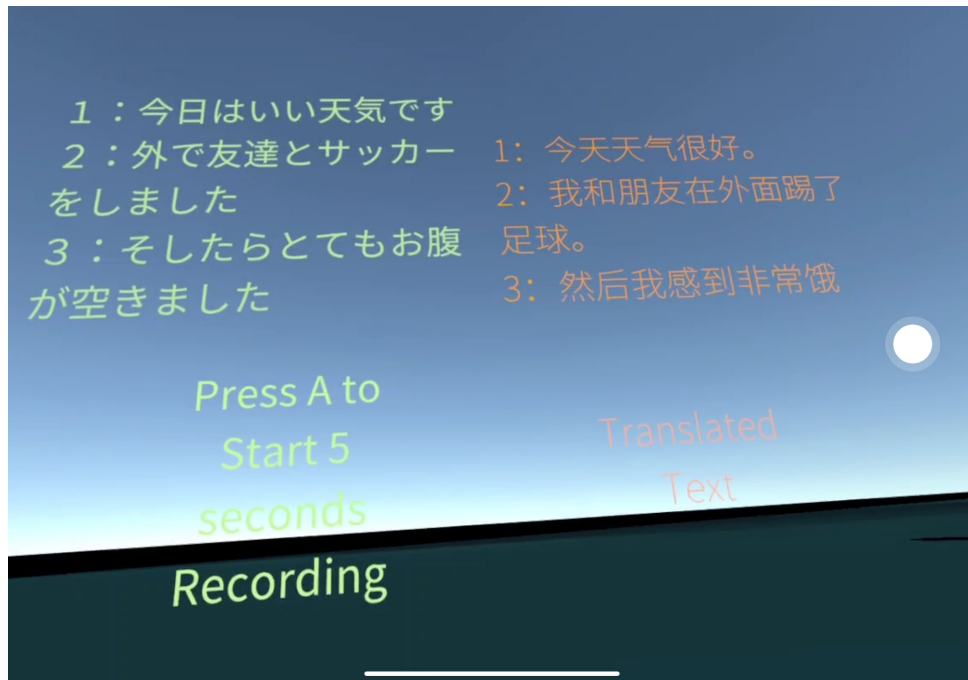


図 4: テキストの表示例

データが作成され, その後, 不必要な部分をトリミングして新たな AudioClip が生成される. 生成された AudioClip は Photon と Google Cloud に送信され, 通常の音声会話用の処理と音声間翻訳会話用の音声認識処理が開始される.

通常の音声会話処理では, AudioClip が Photon の配信用音声データフィールドにセットされ, 他のクライアントに向けて配信される.

音声間翻訳会話用の音声認識処理では, 録音された AudioClip が事前に適切なフォーマットに変換され, Google STT に送信される. Google STT は音声を認識し, 言語を判別し予測精度の高い言語でテキストに変換して返す. 音声認識処理が終わり結果を受け取ると翻訳処理が開始される. Google Translation に音声認識の結果テキストと言語コードが送信され, 翻訳が行われる. この際, ユーザの視界に表示されている音声認識結果のテキストが更新される. 翻訳処理が終わり翻訳テキストを受け取ると音声合成処理が開始される. Google TTS に翻訳テキストが送信され, 音声データが生成される. 同時に, 翻訳結果のテキストが更新される. 音声合成処理が終わり音声データを受け取ると, 音声データが再生処理に渡される. 音声再生処理では, 各クライアントに対して生成された音声データが送信される.

システムのアーキテクチャ図を示す (図 5) .

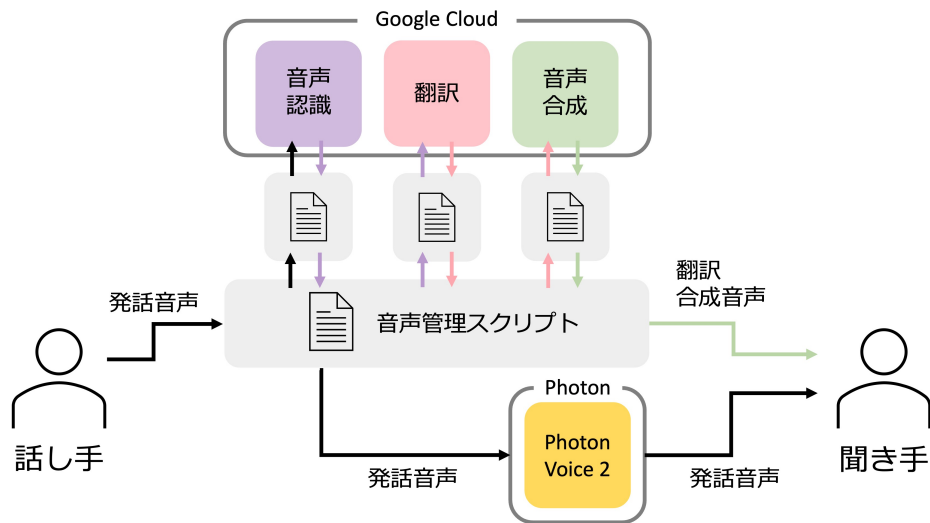


図 5: アーキテクチャ図

3.3 ユーザーインターフェース

ユーザのすべての操作はコントローラーによって行われる。ボタンのマッピングは次のように設定した（図 6）。

- A ボタン：音声間翻訳会話
- X ボタン：音声会話
- Y ボタン：翻訳テスト文の表示・非表示の切り替え
- 右スティック：視点の回転
- 左スティック：前後左右の移動

本アプリケーションでは、ユーザは視点の回転と前後左右の移動のみでアバターを操作する。A ボタンの押下により、音声間翻訳会話のプロセスが開始され

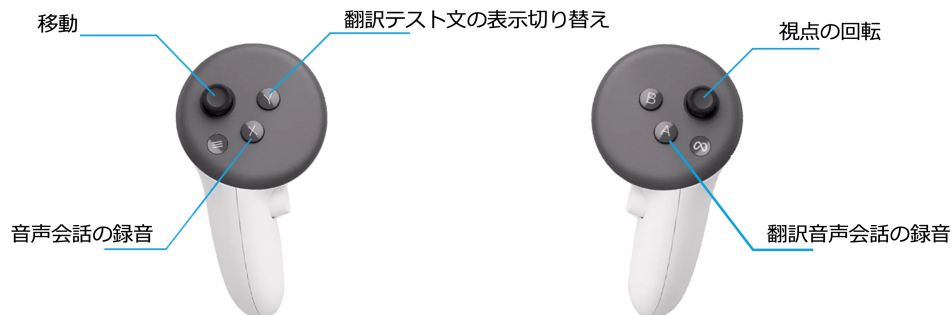


図 6: ボタンのマッピング

る。また、X ボタンを押すことにより、通常の音声会話のプロセスが開始される。

第4章 実験

この実験ではメタバース上で音声会話において、日本語での会話と、音声合成を使用した中国語での会話を比較した場合、発話者の行動に変化が現れるのかを調査することが目的である。

本実験には、近年テレビや新聞のニュースで中国の「富裕層」や「爆買い」などの言葉を目にすることが多く、中国が経済的に成功しているというイメージを持ちやすいという背景から、中国語の合成音声で会話を行った場合、日本語での通常の会話と比較して、交渉タスクにおいてより強気な行動をとるという仮説を立てた。

4.1 実験概要

プロテウス効果を検証するための交渉タスクとして最後通牒ゲームを行った。また、性格特性の変化を測定するために、通常の音声会話と音声合成を使用した中国語での会話の後にそれぞれ BigFive テストを実施し、比較を行なった。順序効果をなくすため、被験者は2グループに分けられ、グループごとに会話の順番を入れ替えて実験を行った。

4.2 最後通牒ゲーム

最後通牒ゲームは行動経済学や実践経済学で頻繁に用いられる実験手法であり、人間の意思決定や公平性の感覚を測ることを狙いとしている。このゲームはシンプルなルールのもとで行われるが、人間の行動に関する深い洞察を得ることができる。最後通牒ゲームの基本的なルールでは、二人の個人が仮想の資金プールをどのように分割するかを決定する。一人が資金プールの分割額の提案を行い、もう一人は提案を受け入れるか拒否するかを決定する。提案が受け入れられた場合は、資金はその分割額に応じて分配される。提案が拒否された場合はどちらも資金を受け取ることができない。

4.3 実験参加者

この実験には実験協力者として1人の中国籍の日中バイリンガルが参加した。また、被験者として1人の女性と7人の男性の日本人が参加した。被験者の年齢は21歳から22歳であった。この実験は一人の実験協力者が全ての被験者と対話

タスクを行なった。

4.4 実験手順

被験者は小部屋に案内された後、デバイスの使い方について一通りの説明を受け、実験を開始した。実験の流れを図7に示す。

被験者はメタバースに入るとまず対話相手との自己紹介を行い、その後、最後通牒ゲームを開始した。本実験での最後通牒ゲームは資金プールを1万円に設定し、4回のラウンドにかけて行われた。1回目と3回目のラウンドは被験者が分割額の提示を行った。実験協力者は、10:90（被験者が有利）以上にならない限り提案を受け入れるように指示されていた。2回目と4回目のラウンドは実験協力者が分割額の提示を行った。2回目のラウンドは常に50:50の分割額を提案するように指示されていた。また、4回目のラウンドは常に75:25（実験協力者側に有利）の分割額を提案するように指示されていた。2つの比率は、公平と不公平な分割を表すために選択された。本実験での最後通牒ゲームの設定はYeeらの研究に基づき設定されている[2]。

最後通牒ゲームの全てのラウンドが終了後、Big Five テストを実施した。Big Five テストは性格特性を診断するという目的を伏せるために、Google Form を用いて質問を転写したものを使用した。Big Five テストの終了後、日本語での会話と中国語での会話を切り替えて2回目の会話を行った。最後に実験後アンケートを行った。

音声間翻訳を使用した会話を行う前には、翻訳操作の練習が行われた。メタバー

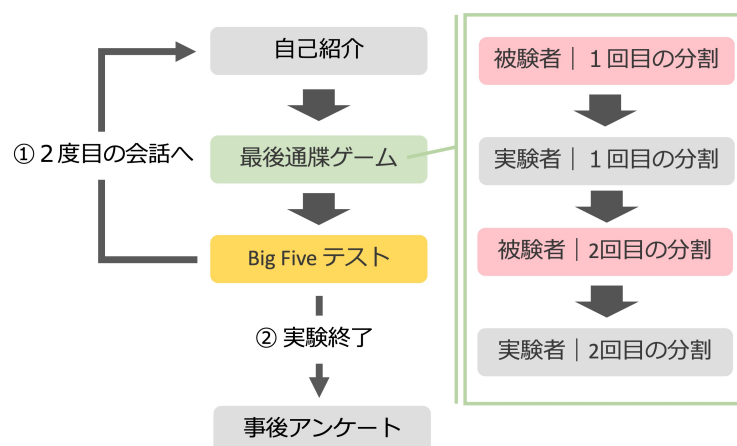


図7: 実験の流れ

ス空間に日本語のテキストとその中国語のテキストのペアで構成された問題文が表示され、被験者は日本語のテキストの読み上げを行った。テキストの読み上げは問題文の中国語テキストと翻訳結果のテキストが一致するか、または1つのテキストを3回読み上げるまで繰り返された。この翻訳操作の練習は合成音声をした発話行為に対して行為主体感の強度を高めることを目的に行われた。問題文の翻訳後のテキストの一部は意図的に翻訳結果が一致しないように変更されており、翻訳操作の回数を増やすよう設定されていた。

4.5 指標

4.5.1 身体化感覚

合成音声を使用した中国語での会話に対して身体化感覚が獲得できたかを判断するために、アンケートを実施し、行為主体感について5段階の尺度で評価を収集した。アンケート項目は次のとおりである。

1. 質問1

- (a) 翻訳後の言語を話しているという感覚はありましたか？
- (b) その理由があれば教えてください

2. 質問2

- (a) 翻訳後の合成音声の再生は自分の行為の結果であると感じましたか？
- (b) その理由があれば教えてください

3. 質問3

- (a) 音声認識の結果と翻訳結果のテキストが表示が更新されたのは自分の行為の結果であると感じましたか？
- (b) その理由があれば教えてください

4.5.2 人格特性

本実験では、性格特性の変化を定量的に計測するため BigFive テストを使用した。BigFive テストは性格特性の5因子（神経症傾向、外交性、経験への開放性、協調性、誠実性）を複数の質問項目により、数値で測定することができる手法である。BigFive テストは性格心理学において最も有力なモデルであり、基本的な枠組みとして定着している。

BigFive テストの評価の指標は2つある。1つ目が、被験者ごとの各性格特性の数値の変化の方向性である。二つ目が、数値の平均値の変化量である。

本実験では無料のオープンソース BigFive テストを提供している Web ページ

(<https://bigfive-test.com/ja>) を用いて評価を行った. この Web ページは高度な性格測定およびその他の個人差測定の開発のための科学共同機関である IPIP (International Personality Item Pool) の公式 Web ページ (<https://ipip.ori.org/>) において, IPIP を使った最高のオンライン実装の一つと紹介されており, 信頼性が高いと判断し, 本実験の評価手段とした. この BigFive テストは 120 項目の質問によって構成されており, 5 段階のリッカート尺度で回答を行う.

4.5.3 交渉態度

最後通牒ゲームで評価対象の指標は 3 つあり, まず第 1 回のラウンドで被験者が提示した分割額 (以降, 第 1 分割と呼ぶ), 次に第 3 ラウンドで被験者が提示した分割額 (以降, 第 3 分割と呼ぶ), 最後に第 4 ラウンドで実験協力者が提示した不公平な分割を被験者が受け入れたかどうか (以降, 最終分割と呼ぶ) である.

第5章 結果

5.1 交渉態度への影響

実験参加者の提示した分割額の一覧を表1に示す。また、第1分割と第3分割の平均提示額と標準偏差を表2に示す。なお、被験者1人のデータは会話ログの欠落により除外された。

まず、音声間翻訳を使用した場合と使用しなかった場合において、分割額の平均に有意な差があるかを測定するため対応のあるt検定を行った。また、その変化の大きさを測定するためにCohenの効果量dを測定した。その結果の一覧を表3に示す。また、箱ヒゲ図を図8に示す。

第1分割においてはt値;1.98,p値;0.07と統計的に有意とは認められないものの、有意である傾向が示された。また、効果量dは1.05でありCohenの定義に基づくと大きな効果あることが示された。

第3分割においてはt値;1.04,p値;0.31と統計的に有意とは認められなかった。

表1: 第一分割と第三分割の提示額

	音声間翻訳あり		音声間翻訳なし	
	第1分割	第3分割	第1分割	第3分割
1人目	5000	6500	-	-
2人目	10000	7000	5000	5500
3人目	5000	7000	5000	2000
4人目	6000	6000	5000	1000
5人目	6000	7500	5000	8000
6人目	5000	6000	5000	5000
7人目	10000	8000	5000	10000
8人目	5000	5100	5000	6000

表2: 提示額の平均と標準偏差

	第1分割	第3分割
音声間翻訳あり	6,714(2288)	6,657(1003)
音声間翻訳なし	5,000(0)	5,357(3145)

表 3: 最後通牒ゲーム—t 検定の結果及び効果量 d

	第 1 分割	第 3 分割
t 値	1.98	1.04
p 値	0.07	0.31
効果量 d	1.05	0.55

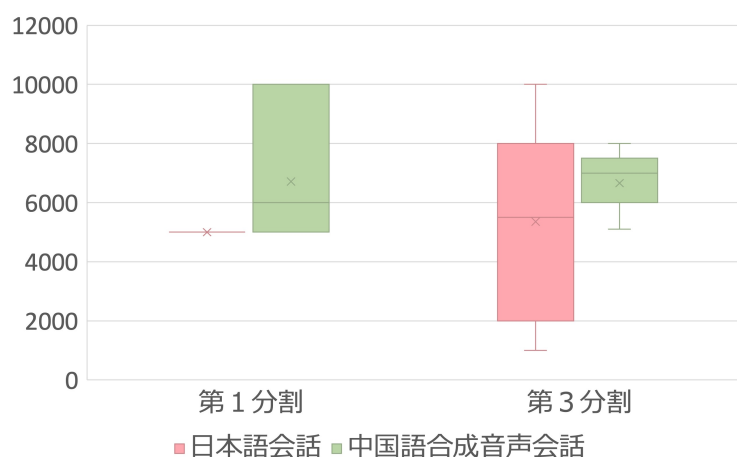


図 8: 箱ヒゲ図 | 最後通牒ゲーム

また, 効果量 d については中程度の効果があることが示された。

最終分割の承諾率は音声間翻訳ありで 85.7%, 音声間翻訳なしで 28.5%であった。

5.2 人格特性への影響

被験者ごとの各性格特性の数値の変化の方向性については, 合成音声を使用した中国語での会話において数値が増加したか, 減少したかを集計し, の方向性の傾向を調べた. その結果を図 4 に示す.

各性格特性の数値の平均値の変化量については, 最後通牒ゲームと同様に対応

表 4: 性格特性の変化の方向性

	神経症傾向	外交性	経験への開放性	協調性	誠実性
増加	4	4	5	3	6
一致	2	0	0	0	1
減少	2	4	3	5	1

表 5: BigFive テスト—t 検定の結果及び効果量 d

	神経症傾向	外交性	経験への開放性	協調性	誠実性
t 値	0.20	0.14	0.37	0.92	0.14
p 値	0.83	0.88	0.71	0.37	0.88
効果量 d	0.11	0.07	0.20	0.49	0.07

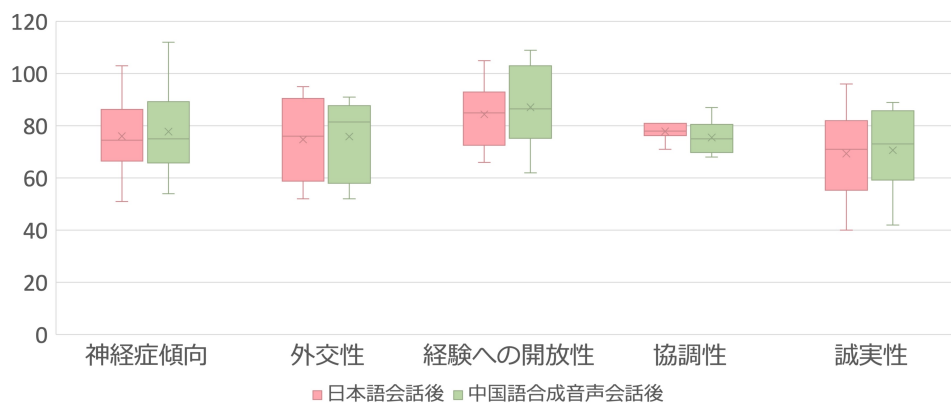


図 9: 箱ヒゲ図：Big Five テスト結果

のある t 検定を行い, その差の変化量を計測するために Cohen の効果量 d を算出した. t 検定と効果量の結果の一覧を表 5 に示す. また, 箱ひげ図を図 9 に示す.

5.3 身体化感覚の評価

実験後のアンケート調査では次のような結果が得られた.

「翻訳後の言語で話しているという感覚はあったか」という質問に対しては, とても感じたが 1 人, 感じたが 4 人, 感じなかったが 2 人, 全く感じなかったが 1 人であった. 感覚を得られた側の理由では, 「音声がすぐに変換されたから」「中国語テキストが表示されたから」「コミュニケーションが成立したから」という理由が挙げられた. 感覚を得られなかった側の理由では「日本語でしか受け答えしていない」「日本語テキストが表示されたから」「合成音声聞き取りづらいから」という理由が挙げられた.

「翻訳後の音声再生されたのは自分の行為の結果であると感じたか」という質問に対しては, とても感じたが 2 人, 感じたが 6 人であった. 「翻訳された音声を聞いたから」「翻訳されたテキストが表示されたから」という理由が挙げら

れた。

「音声認識結果と翻訳結果のテキストが表示されたのは自分の行為の結果と感じたか」という質問に対しては、とても感じたが2人、感じたが5人、感じなかったは1名であった。感覚を感じた理由では「発話に連動していたから」という理由が挙げられた。感覚を感じなかった理由では「意識していない間に変わっていたから」という理由が挙げられた。

第6章 考察

6.1 身体化感覚の獲得

アンケートの結果から, プロテウス効果の検証の前提となる, 合成音声での会話(発話)に対する行為主体感の付与は概ね成功したと評価した. 合成音声が生されるまでの時間の差が少なくしたことが, 行為主体感の強度を高めた要因になったと考えられる. また, 行為主体感の強度を高めるために視覚的なフィードバックとして表示したテキストは, 中国語の翻訳テキストを表示することで中国語で話しているという感覚を高めた一方で, 日本語の音声認識テキストを表示することにより日本語で話しているという感覚も高める結果となった.

6.2 交渉態度の変化

最後通牒ゲームの結果では, 第1分割と第3分割での中国語の合成音声を使用した会話において, いずれも提示額の平均が増加した. この結果は事前の仮説と一致した. 第1分割については, t 検定での分析において p 値が 0.07 と十分とは言えないが有意である傾向が示された. また, 効果量についても大きな効果が認められた. 第3分割では, p 値が 0.31 と有意ではない結果となったが, 効果量は中程度の効果が認められた.

これらの結果を総評すると, 本実験は被験者数が 8 名と少なく有意な結果であるとはいえないものの, 中国語の合成音声を使用した会話において話者がより強気な行動をとるように変化し, メタバースでの合成音声を使用した言語の変換においてもプロテウス効果が発生する可能性が示唆された.

6.3 人格特性の変化

BigFive テストの結果では, 通常の会話の後と中国語の合成音声を使用した会話の後で行ったテストの比較において各性格特性の数値は微量に変化したものの, t 検定での分析ではいずれの差も有意であるとは認められなかった. また, 効果量については, 「経験への開放性」が小さな効果, 「協調性」で中程度の効果が認められた.

最後通牒ゲームで被験者の行動に変化が見られた一方で, BigFive テストにおいてその数値に微量な変化しか見られなかった理由は, そもそも性格特性が変化しなかったか, BigFive テストの質問文が全て日本語であったことで中国語を使

用している感覚を失ったためであると推察される. 文化的フレームスイッチングの研究 [1] においては, バイリンガルに対してそれぞれの言語で BigFive テストを実施していたが, 本研究においては回答者は日本語のモノリンガルであるため, 日本語での BigFive テストの実施であった. メタバースで中国語の口頭での BigFive テストを行った場合, 中国語を使用している感覚を失わずに異なる結果が得られた可能性がある.

第7章 おわりに

本研究では, メタバースにおける発話言語によるプロテウス効果を検証した. 具体的には, 発話音声为中国語の合成音声に変換する音声間翻訳会話を行った際に, 発話者の行動に変化が現れるかを交渉タスクを用いて確認し, 同時に BigFive テストで性格特性に変化があるかを調べた. 結果として, 交渉タスクにおいて音声間翻訳を使用した場合に強気な行動をとる可能性が示唆されたが, 性格特性の変化は見られなかった.

7.1 実験の拡張

本研究の実験設定では, 被験者側の中国語での会話の要素として「相手の中国語の発話音声を聞くこと」「自身の中国語の合成音声を聞くこと」「中国語のテキストを見ること」の3つの要素があり, どの要素が行動に影響を与えたかを特定することはできなかった. そのため, 音声間翻訳会話と, 音声間翻訳会話の被験者の合成音声を削除した会話 (図 10) を比較して, どの要素が変化をもたらしたのかを特定する必要性が感じられた.

また, 合成音声を使用した中国語での会話において行動が変化する可能性が示唆されたものの, その影響がメタバース特有の現象であるのか, 現実世界での音声翻訳を用いた会話においても発生する現象であるのかは検証できていない.

さらに, 今回は日本語と中国語間のみでの検証であったが, 本研究で作成したアプリケーションを拡張し, 他の言語においてはプロテウス効果が発生するのかを検証を行う必要がある.

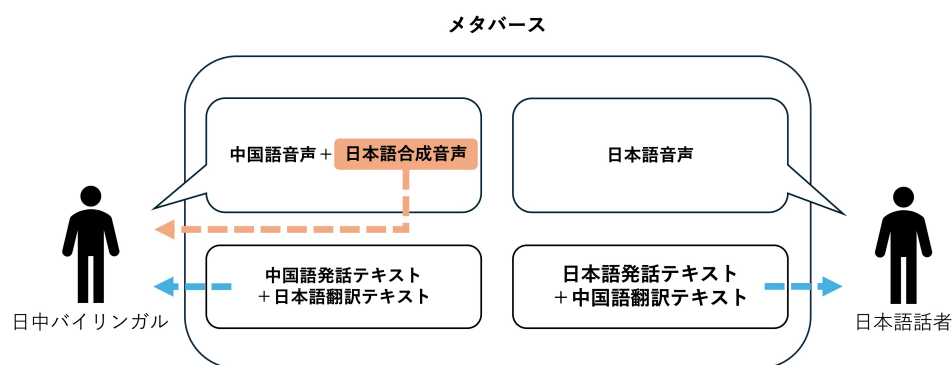


図 10: 追加実験設定

7.2 メタバース環境の改善

本研究でのメタバース環境には、実験後に見つかった多くの改善の余地がある。ひとつは、音声翻訳サービスの精度の問題である。音声認識は概ね正確だったが、翻訳が不適切で、発話の意図を聞き返す場面がたびたび見られた。これは会話の円滑さを損ない、結果に影響を与えたと考えられる。また、音声合成については、今回使用した Google Text-to-Speech のビルトインボイスでは、発話者の声の高さにあった調整ができないため、十分な行為主体感を生み出せなかったと考えられる。今後、追加で実験する場合は、音程を調整可能な音声合成サービスの使用を行うべきである。

さらに、今回の実験ではアバター間の距離や向きの影響を考慮していなかった。今回のメタバース設計では、ユーザが自由に移動可能であったため、会話の際のアバター同士の距離や向きに差異が存在した。対人での会話において、対話相手との距離と向きが話者に対して影響を与えることは自明である。そのため、今後追加で実験を行う際は、アバターの位置と向きについての厳密な指定が必要である。

謝辞

本研究を行うにあたり,指導教官の村上陽平教授と Mondheera Pituxcoosuvann
助教から大変熱心なご指導,多大なる助言を承りました. 厚く感謝申し上げます.

参考文献

- [1] Ramírez-Esparza, N., Gosling, S. D., Benet-Martínez, V., Potter, J., & Pennebaker, J. W.: Dobilinguals have two personalities? A special case of cultural frame switching, *Journal of Research in Personality*, 40, 2, pp. 99-120 (2006).
- [2] Yee, N., & Bailenson, J.: The Proteus effect: The effect of transformed self-representation on behavior, *Human Communication Research*, 33, 3, pp. 271-290 (2007).
- [3] Whorf, B. L.: The relation of habitual thought and behavior to language, *Sociolinguistics: A Reader*, pp. 443-463 (1997).
- [4] Gallagher, S.: Philosophical conceptions of the self: implications for cognitive science, *Trends in Cognitive Sciences*, 4, 1, pp. 14-21 (2000).
- [5] Sato, A., & Yasuda, A.: Illusion of sense of self-agency: discrepancy between the predicted and actual sensory consequences of actions modulates the sense of self-agency, but not the sense of self-ownership, *Cognition*, 94, 3, pp. 241-255 (2005).
- [6] Ohata, R., Asai, T., Imaizumi, S., & Imamizu, H.: I Hear My Voice; Therefore I Spoke: The Sense of Agency Over Speech Is Enhanced by Hearing One ' s Own Voice, *Psychological Science*, 33, pp. 1226-1239 (2022).
- [7] Virtual Frontiers Oy, 無題, <https://assetstore.unity.com/packages/3d/characters/humanoids/pack-free-sample-79870>, 2017