

# 視線とジェスチャにより会話進行が変化する VR 英会話システム

植田智裕<sup>1</sup> 田辺弘子<sup>2</sup> 小宮山撰<sup>2</sup>

**概要**：近年、VR の発展に伴いアバタとコミュニケーションをとるコンテンツが多く増加している。中でも、会話相手のアバタや会話場面の空間を再現し、現実に近い状況で会話練習を行う VR 英会話は、英会話教材に対するニーズの高い日本において有用性が期待できる。しかし、現状の VR 英会話は内容や発音評価などに重点が置かれており、実際の会話で重要となる非言語コミュニケーションの要素がユーザ側に実装されておらず、自然な状態で会話が出来るとは言い難い。そこで VR 空間のアバタとの英会話練習において、非言語情報が持つ効果を評価する目的で、ユーザおよびアバタの視線、首振りや傾き、指差しの身体動作が会話の進行に変化を与える英会話練習システムを試作し、そのユーザビリティを評価する。

**キーワード**：バーチャルリアリティ、英語学習（支援）、非言語コミュニケーション、アバタ

## 1. はじめに

近年、VR の発展に伴いアバタとコミュニケーションをとるコンテンツが増加している。中でも、会話相手のアバタや会話場面の空間を再現し、現実に近い状況で会話練習を行う VR 英会話は、英会話教材に対するニーズの高い日本において有用性が期待できる[1][2]。しかし、現状の VR 英会話は内容や発音評価などに重点が置かれており、実際の会話で重要となる非言語コミュニケーションの要素がユーザ側に実装されていない。現実の会話では視線を合わせる、傾く、等の非言語情報が会話進行上重要であり[3]、VR 空間内でのアバタとのコミュニケーションにおいても、非言語情報が重要になると考えられる[4]。そこで VR 空間のアバタとの英会話練習において、非言語情報が持つ効果を評価する目的で、ユーザおよびアバタの視線、首振りや傾き、指差しの身体動作が会話の進行に変化を与える英会話練習システムを試作し、そのユーザビリティを評価する。

## 2. 先行研究

### 2.1 非言語的コミュニケーションについて

我々は、言語情報のみならず、身振りや身体的特徴などの非言語情報に基づいてコミュニケーションを図っており、人間同士の対面による会話の 65-93%がノンバーバルによるものだと言われている[5][6]。図 1 に先行研究で報告された、メッセージ伝達に占める言語、非言語情報の割合について示す。

非言語情報は、視線や身振り、体型、声質、対人距離など数多くある。その中でも身体動作は多種多様で、次々とメッセージを発信する特徴を持ち、言語的機能が特に大きい。他にも、音声を補完したり、対話の枠組みを伝えたり、会話の進行を調整する機能を持つ[3]。そこで Ekman は

Efron[7]が作った用語を援用し、身体動作を分類した[8]。Ekman による身体動作の分類を表 1 に示す。

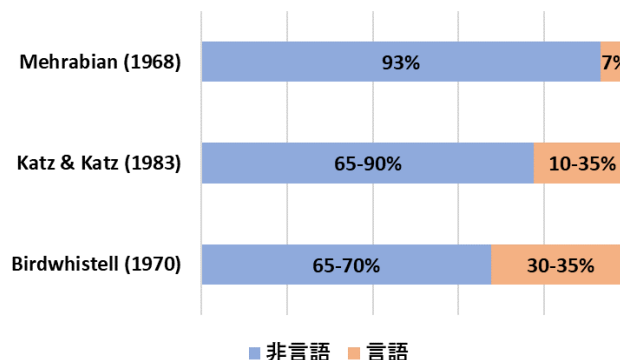


図 1 メッセージ伝達に占める言語、非言語情報の割合

表 1 Ekman による身体動作の分類

| 標識   | 音声語句に翻訳可能      |
|------|----------------|
| 例示子  | 発話内容を強調、精緻化、補足 |
| 情感表示 | 情動に伴う表情、身振りなど  |
| 調整子  | 会話の流れを円滑にする動作  |
| 適応子  | 状況に適応するための動作   |

### 2.2 VR 空間のアバタとの英会話に関する研究

VR 空間の臨場感を活かした英会話学習システムが登場している中で、鈴木らは VR 空間のアバタとの音声対話を利用した英会話を、アバタの有無で比較し、アバタの存在が好意的に評価されることを示した[4]。一方で、非言語情報のない VR 空間でのアバタとのコミュニケーションは、良質な対話とは程遠いという成宮らの主張もある[9]。

そこで近年はノンバーバルな要素を取り入れた外国語会話学習の検討が増えてきている。斎藤らは、人間のモーションキャプチャデータをアバタに取り入れ、アバタ発言からの返答時間に制限を設けることでユーザにプレッシャ

<sup>1</sup> 青山学院大学大学院  
<sup>2</sup> 青山学院大学

を与える VR 空間におけるスペイン語の対話学習システムを提案した[10]. このシステムを用いることで、アバタと会話するという違和感が少なくなることが確認されたとしている. このように、アバタに対して様々な非言語情報を取り入れる VR 英会話の研究が行われている. しかし、ユーザのノンバーバルな反応は考慮されていない. VR 空間内でのアバタとのコミュニケーションにおいても、現実の会話と同様、非言語情報が重要になると考えられるため、ユーザの非言語情報も取り入れる必要があると言える.

そこで、筆者らはこれまでに、ユーザおよびアバタの視線、首振りや傾きの身体動作が会話の進行に変化を与える英会話練習システムを試作し、アバタに傾き首振りを取り入れることは人間らしさを生み出すこと、ユーザにアバタに視線を合わせるよう促すことが人間同士の会話に近づくことを示した[11]. この研究では、個々の非言語情報の効果が検討されており、非言語情報の言語的役割と感情表現の役割が分離されておらず、それぞれが会話の自然さにどの程度影響するかについては明らかでない. そこで、本研究の英会話練習システムでは、視線、相槌、プレッシャ表現の感情的情報と、ユーザおよびアバタによる手の動作や傾き、首振りによる言語的情報を実装し、非言語情報の効果の総合的評価を可能とする.

### 3. VR 英会話システムへの非言語情報の実装

#### 3.1 概要

人間同士の対面による会話では、言語のみでなく、非言語情報が重要とされている[3]. ここでは非言語情報のうち身体動作に着目する.

Ekman による身体動作の分類[8]に従い、今回は英会話練習で重要と考えられる標識の「首振り」と「傾き」、例示子の「手の動作」、情感表示、調整子として「相槌」と「視線」に着目する. このうち首振り、傾き、手の動作は言語に置き換え可能な身体動作であり、視線、相槌は置き換えが難しいという性質を持つ.

以上のユーザの非言語情報を用いて会話進行を変化させる英会話練習システムを作成した. 本システムの全体構成図を図 2 に、会話の変化内容は表 2 に示す.

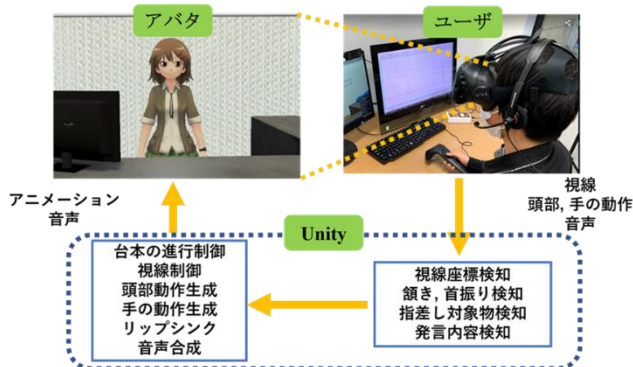


図 2 システムの全体構成図

表 2 会話の変化内容

|      |                                        |
|------|----------------------------------------|
| 視線   | ユーザがアバタに視線を合わせるか否かで、アバタの発言の有無に変化がある    |
| 頭部動作 | ユーザの傾きと首振りが Yes/No の意思表示となり、アバタの発言が変わる |
| 手の動作 | ユーザの指差しが、会話の中で参照している対象物を示す動作となる        |

以下に VR システムへの具体的な実装方法を示す.

#### 3.2 視線

ユーザの視線検出は Tobii Pro の搭載された HTC Vive[12]で行う. 計測した視線 (X,Y) 方向に、VR 空間内で Ray を生成し、Ray が最初に衝突するオブジェクトとの交点を視線座標 (X,Y,Z) と定義している. 実験の始めに、キャリブレーションを行うことで、被験者の正確な視線を取得する. 視線座標の計測方法について図 3 に示す.

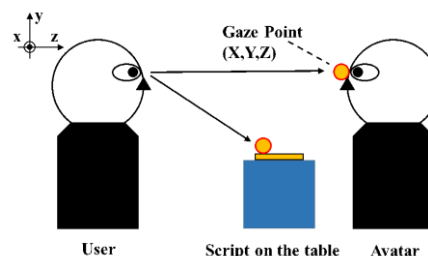


図 3 ユーザの視線検知

ユーザがアバタに視線を 2 秒以上合わせることで、アバタは会話を始め、ユーザに視線を向け発言に反応する. またアバタから視線を 2 秒離すと、アバタは会話を止め、ユーザを見ることなく発言に反応しない. 再度アバタに視線を 2 秒合わせると、ユーザに視線を向けて会話は進行する. 2 秒とした理由は、予備実験により、自然な会話進行が可能と考えられたためである. 他にも、ユーザが指差しや口頭でオブジェクトを選択する場合に、アバタはそのオブジェクトに短時間視線を向けることとした.

#### 3.3 頭部動作

ユーザの傾きや首振り検出は、HTC Vive で計測したユーザの首の角度と回転軸から角速度を計算し、Y 軸方向に一定時間変化があると、これを「傾き」として検出する. また、X 軸方向に一定時間変化があれば、「首振り」とする. 傾きと首振りの検出について図 4 に示す. アバタもユーザと同様に Yes/No の意思表示をするときに、傾きや首振りを用いる. また、ユーザが発言している際に、アバタは傾きを用いて相槌を打つ. ただし、傾き続けると不自然なため、ユーザの 1 回の発言に対して、アバタは 1 度のみ傾く.

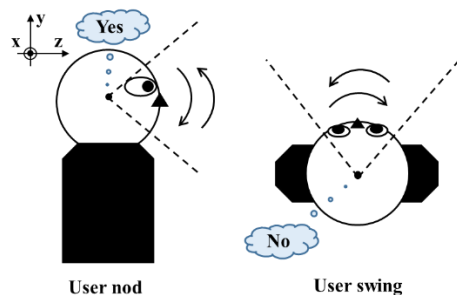


図 4 ユーザの頷き首振り検出

### 3.4 手の動作

ユーザの指差し検出は、HTC Vive のコントローラを用いて、手のトラッキングを行うことで可能になる。コントローラが、VR 空間では手の形をしたモデルの代わりとなる。ユーザが会話の中で参照したい対象物の (X,Y) 方向に、VR 空間の手の人差し指を向けると、人差し指から生成された Ray が最初に衝突する交点 (X,Y,Z), つまりオブジェクトを指差ししたとして検知する。なお今回は、5 本の指を動かす必要はなく、指示動作のみに限定されるため、VR 空間における手の形は常に指差しの形とする。ユーザが対象物に人差し指を 2 秒以上合わせることで、会話の中でオブジェクトを選択したとする。2 秒とした理由は、視線と同様で予備実験により、自然な会話進行が可能と考えられたためである。図 5 にユーザの指差し対象検出について示す。

そして、アバタもユーザと同様に、会話の中でオブジェクトを参照したい場合には、オブジェクトを指差ししながら、会話を進める。

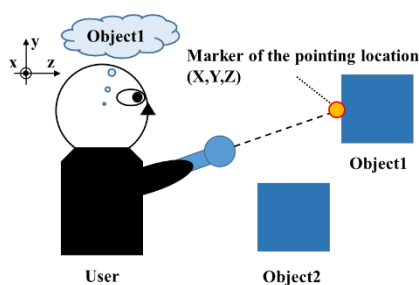


図 5 ユーザの指差し対象検知

### 3.5 ユーザの手の表示

ユーザはコントローラを手を持ち、会話の中で対象物を指差す動作が求められる場面がある。この時に、VR 空間の指が実際にどこを指しているかがわかりやすいように、人差し指から Spot Light を飛ばして指差しポイントとしている。ユーザがこのポイントをオブジェクトに 2 秒以上当ててすることで、会話の中でオブジェクトを選択したとして検知される。なおユーザの手は、Tobii XR SDK を用いた Unity Example の 3D キャラクターモデル[13]を用いた。図 6 に指差し動作の例を示す。



図 6 Unity による指差し動作例

### 3.6 アバタの動作

前述にもあるように、アバタもユーザと同じように、頷きや首振り、指差しの身体動作を行う。そこで本システムのアバタとしては、ボーンが組み込まれていることで自由に動かすことができる「プロ生ちゃん MMD Ver.3 モデル」[14]を用いる。このアバタに人間らしい自然な動きを再現するため、モーションキャプチャを使用した。動作の作成にあたり、実際に筆者自身のジェスチャを HTC Vive より頭部動作を、Vive コントローラ 2 つを用いて指差し動作をキャプチャし、それぞれのアバタの動きとして用いる身体動作アニメーションを作成した。

### 3.7 タイムプレッシャ

斎藤らの手法[10]を踏襲し、アバタの発言後に、タイムバーが進行することで、15 秒以内に返答するようユーザにプレッシャを与える。返答できない場合には、ユーザが 2 秒アバタに視線を合わせなくても、アバタは一つ前の会話をもう一度発言する。その後もう一度タイムバーが進行し、ユーザは 15 秒以内に返答することが求められる。これは、ユーザが返答するまで繰り返行われる。15 秒としたのは、実際の対話場面において違和感のない時間を考慮し、これ以上沈黙が続くと自然な会話ではないと考えたためである。なお、タイムバーはアバタの頭の上に表示した (図 7)。

## 4. 評価実験

### 4.1 目的

前述の身体動作を組み合わせた 3 つの条件で、英会話練習を想定し非言語コミュニケーションの実験を行う。タスク達成時間、ユーザの視線対象とそれを見た時間を測定し、3 種類のアンケートと合わせて、VR 空間内でのアバタとの会話においても、現実の会話と同様に、ユーザが用いる非言語情報が重要といえるかどうか、またそのユーザビリティを調査する。

### 4.2 実験システム

ユーザは常に視線機能を搭載した HMD (HTC Vive) と、ヘッドセット (SC 60 USB ML) を装着し、VR 空間を観察する。VR 空間の開発環境には Unity, Visual Studio2015 を利用し、開発言語には C#を用いた。アバタとしては、ニコ

ニコ立体が無償で提供している「プロ生ちゃん MMD Ver.3 モデル」[14]を用いる。本実験ではリップシンクを導入し、アバタは口を動かしながら、ttsmp3.com[15]を用いて作成した合成音声を発する。また、音声認識については IntelRealsense SDKI[16]によって、マイクから入力されたものを文字に変換している。

なお、本システムは教材である台本をユーザが事前に記憶することなく会話進行中に確認できるように、アバタから離れた場所にあるモニタに台詞を表示し、会話を行う。図 7 に被験者が見る 3 種類の VR 世界を示す。



図 7 Unity による実験環境

また、アバタ側に実装した非言語情報を表 3 に、またユーザ側の非言語情報のうち検出されるものを表 4 に示す。このうち首振り、傾き、手の動作は言語に置き換え可能な身体動作であり、視線、相槌は置き換えが難しいという性質を持つ。

システム全体のフローチャートを図 8 に示す。

表 3 アバタが表出する非言語情報

|      |                                      |
|------|--------------------------------------|
| 視線   | ユーザに視線を向ける/向けない<br>ユーザの指差し対象物に視線を向ける |
| 頭部動作 | ユーザの発言に相槌を打つ<br>傾き、首振りによる Yes/No の表出 |
| 手の動作 | 発言の対象を指差す                            |

表 4 検出対象のユーザの非言語情報

|      |                    |
|------|--------------------|
| 視線   | アバタを見ている/見ていない     |
| 頭部動作 | 傾き (Yes), 首振り (No) |
| 手の動作 | ユーザの指差し対象物         |

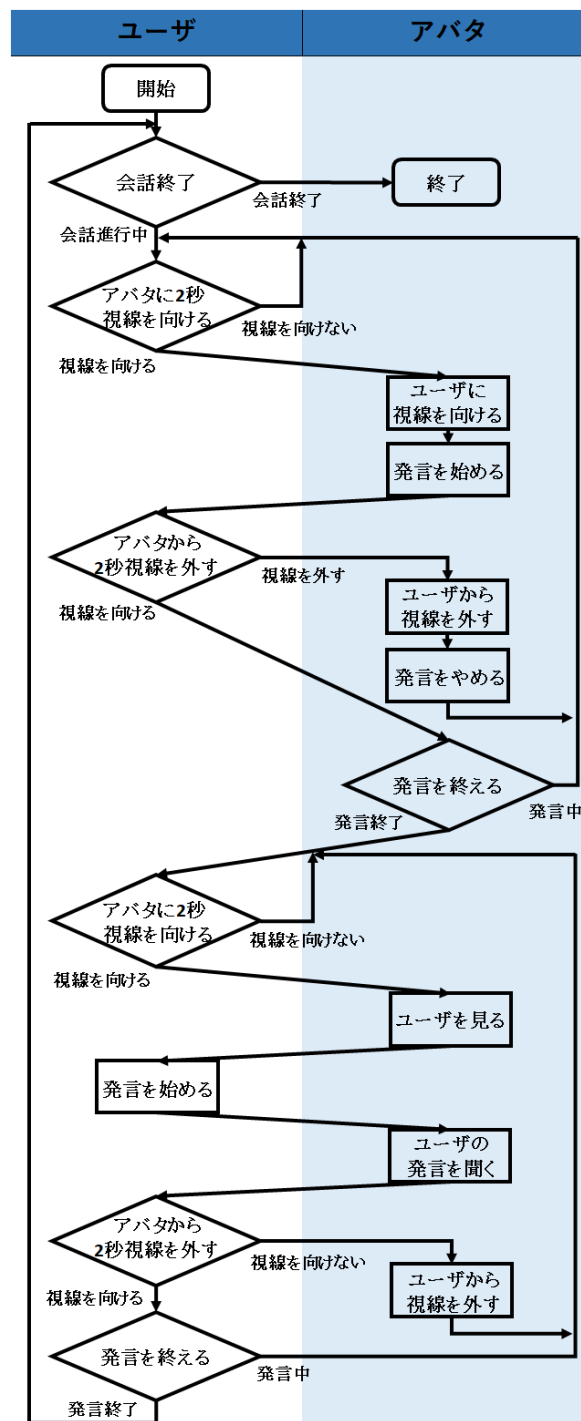


図 8 システム全体のフローチャート

#### 4.3 実験条件

実験条件は、3 節で紹介した視線と頭部動作、手の動作、タイムプレッシャの 4 つの非言語情報を用いて、① Gaze&Gesture 条件、②Gesture 条件、③None 条件、の 3 つ

を用意した(表 5)。また、英会話練習システムとして観光案内所での相談場面、レンタカー屋で車を借りる場面、不動産会社での物件相談場面の 3 つの環境とその台本を設定した。本実験では、被験者ごとに 3 つの条件に対し、それぞれ異なる 3 つの台本をランダムに割り当て、実験を行った。なお、実験開始前に視線計測のキャリブレーションを行わせ、実験中にも計測のズレが気になり、会話の進行に影響がある場合には、再度キャリブレーションを行い、システムを最初から始めることを許可した。

表 5 各条件の非言語情報の組み合わせ

|           | Gaze&Gesture |      | Gesture |      | None   |      |
|-----------|--------------|------|---------|------|--------|------|
|           | Avatar       | User | Avatar  | User | Avatar | User |
| 頷き<br>首振り | ○            | ○    | ○       | ○    |        |      |
| 相槌        | ○            |      |         |      |        |      |
| 視線        | ○            | ○    |         |      |        |      |
| 指差し       | ○            | ○    | ○       | ○    |        |      |
| タイム<br>バー | ○            |      |         |      |        |      |

#### 4.4 実験タスク

今回の実験では、作成した台本に従いアバタとの会話を終わることをタスクとした。ノンバーバルな要素に重点を置いた実験のため、台本とは異なる発言や正しい発音ができない場合においても、会話は進行するように設定した。なお、ユーザおよびアバタに頭部動作がある条件では、どちらも Yes/No を伝える際に頷きや首振りをする。また手の動作がある条件では、どちらも発話の対象を指差す。

本実験は観光案内所での相談場面、レンタカー屋で車を借りる場面、不動産会社での物件相談場面の 3 つの場面で会話を行うため 3 つの台本を用意した。そして、話している箇所をわかりやすくし会話を円滑にできるように、机の上にあるモニタ上にユーザの台詞だけを表示した(図 7)。また、台本間で実験結果に違いが生じないように、ユーザおよびアバタが頷きや首振りをする箇所は同じポイントに設定した。

#### 4.5 実験方法

実験の被験者には大学生および大学院生計 10 名(平均年齢 22.7 歳の男性 8 名、女性 2 名)を用いた。被験者はシステムに慣れるための練習の会話を一度行ったあと、本実験として台本が異なる 3 つの実験条件でランダムな順番で実験を行った。

今回は 60 フレーム毎にユーザの視線対象とそれを見た時間、またシステムの完了時間を計測した。さらに、各条件の実験終了ごとに印象評価、System Usability Scale (SUS) [17], NASA-TLX[18]の概念に基づいた「主観的メ

ンタルワークロードチェックリスト」[19]の 3 種類のアンケートを実施した。

## 5. 実験結果

### 5.1 タスク完了時間

各条件のタスク完了時間を図 9 に示す。一元配置分散分析を行った結果、主効果がみられた ( $F(2, 27) = 15.09, p < 0.01$ )。Bonferroni 法による多重比較の結果、Gaze&Gesture と Gesture, Gaze&Gesture と None との間に有意差があり(いずれも  $p < 0.01$ )、Gaze&Gesture 条件でタスク時間が有意に長いことがわかった。

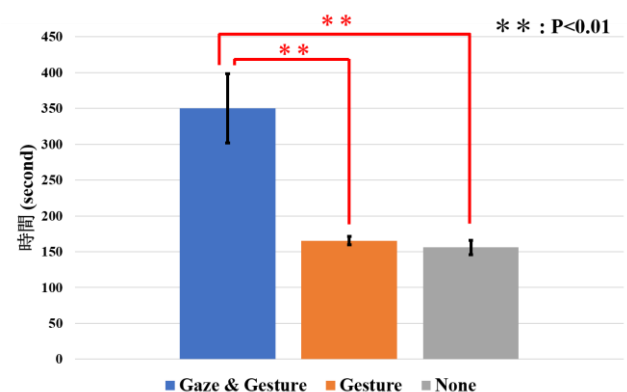


図 9 タスク完了時間

### 5.2 発言時における視線対象

各条件の被験者が発言した際の視線対象を図 10 に示す。二元配置分散分析を行った結果、視線対象の主効果 ( $F(3, 116) = 104.57, p < 0.01$ )、交互作用が有意であった ( $F(6, 113) = 9.72, p < 0.01$ )。Bonferroni 法による多重比較の結果、アバタを見る回数は、Gaze&Gesture 条件が他の 2 条件よりも有意に多いことがわかった (いずれも  $p < 0.01$ )。一方台本を見る回数では、Gaze&Gesture 条件が他の 2 条件よりも有意に少なかった (いずれも  $p < 0.01$ )。

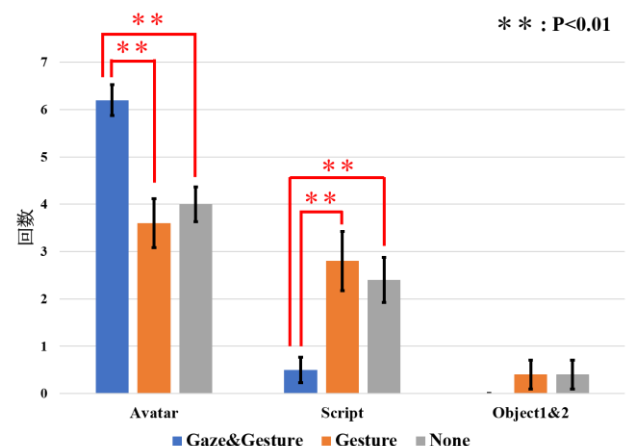


図 10 発言時における視線対象

### 5.3 アバタ発言からの発言、ジェスチャまでの時間

各条件のアバタが発言してから被験者が発言、ジェスチャをするまでの時間を図 11 に示す。被験者が発言するまでの時間について、一元配置分散分析を行った結果、主効果がみられた ( $F(2, 207) = 34.07, p < 0.01$ )。Bonferroni 法による多重比較の結果、Gaze&Gesture と Gesture、Gaze&Gesture と None との間に有意差があり (いずれも  $p < 0.01$ )、Gaze&Gesture 条件でアバタ発言後から被験者が発言するまでの時間が有意に長いことがわかった。

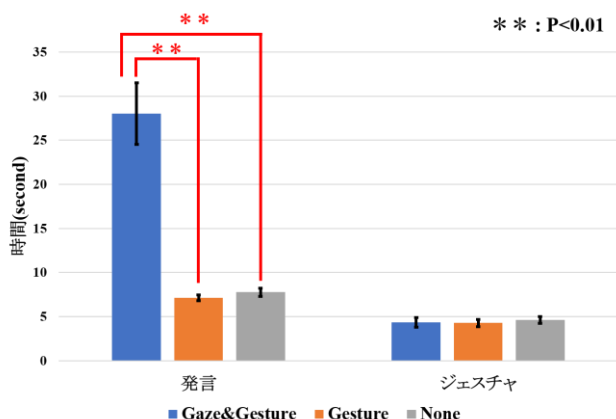


図 11 アバタ発言からの発言、ジェスチャまでの時間

### 5.4 印象評価アンケート

評価項目は-2 から 2 の 5 段階評価で行い点数が高いほど、そう思う、そう感じるという評価に設定した。評価項目は、VR 英会話におけるアバタとのコミュニケーションに対し、人間同士のものと近い印象を持てたか評価する目的で選定した (表 6)。A, B, C, D, E は自然さ, F, G, H は発言やジェスチャの使用感, I, J, K は英会話練習で起きがちな緊張感や精神的負担に関する項目である。

表 6 評価項目一覧

| 評価項目一覧 |                         |
|--------|-------------------------|
| A      | 話しやすい                   |
| B      | 人間との会話に近い               |
| C      | コミュニケーションを取れている         |
| D      | アバタに親しみを持てた             |
| E      | アバタがどこを見ているのか理解できた      |
| F      | 発話のタイミングが取りやすかった        |
| G      | 思ったように Yes/No の意思表示ができた |
| H      | 思ったように対象物を選択できた         |
| I      | 緊張しなかった                 |
| J      | 会話することが恥ずかしくなかった        |
| K      | プレッシャを感じた               |

各条件の実験終了後に行ったアンケートの結果を図 12 から図 14 に示す。一元配置分散分析を行った結果、評価項目 A, E, H, I, J, K で主効果がみられた。Bonferroni 法による多重比較の結果、評価項目 A は Gaze&Gesture と Gesture との間に有意差があり ( $p < 0.05$ )、Gesture 条件で有意に話しやすいと感じることがわかった。評価項目 E と J は、Gaze&Gesture と None との間に有意差があり (E:  $p < 0.01$ , J:  $p < 0.05$ )、Gaze&Gesture 条件で有意にアバタがどこを見ているのか理解できた、None 条件で有意に会話することが恥ずかしくなかったと感じることがわかった。評価項目 H は、Gesture と None との間に有意差があり ( $p < 0.01$ )、Gesture 条件で有意に思った通り対象物を選択できたと感じることがわかった。評価項目 K は、Gaze&Gesture と None ( $p < 0.05$ )、Gaze&Gesture と Gesture との間に有意差があり ( $p < 0.01$ )、Gaze&Gesture 条件で有意にプレッシャを感じたことがわかった。

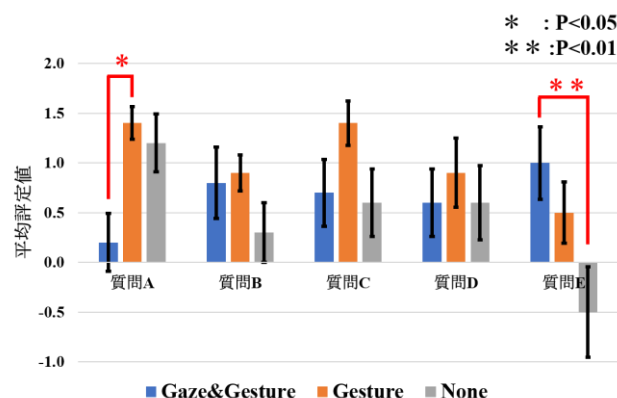


図 12 評価項目 A, B, C, D, E のアンケート結果

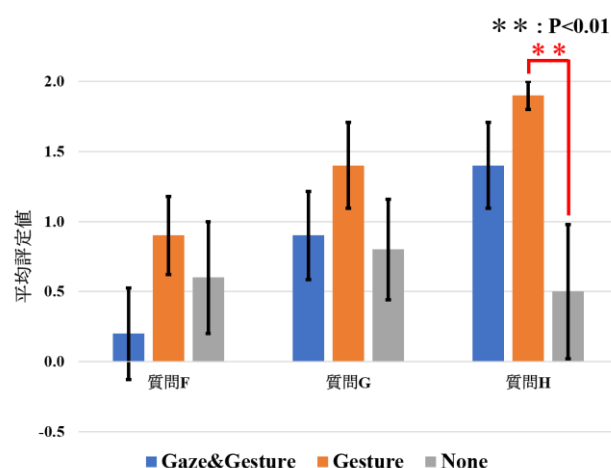


図 13 評価項目 F, G, H のアンケート結果

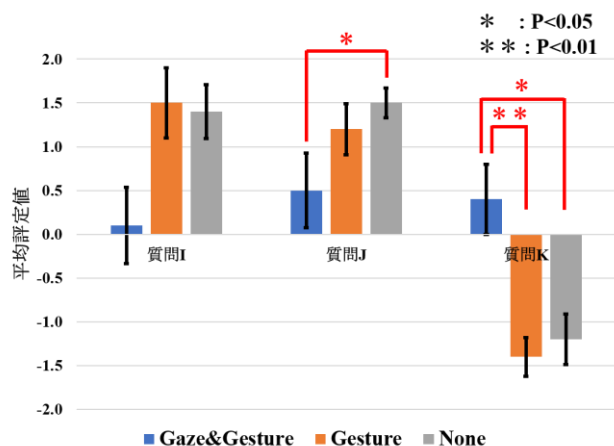


図 14 評価項目 I, J, K のアンケート結果

## 5.5 ユーザビリティ評価のための SUS

各条件のユーザビリティ評価のための SUS を図 15 に示す。一元配置分散分析を行った結果、主効果がみられた ( $F(2, 27) = 20.67, p < 0.01$ )。Bonferroni 法による多重比較の結果、Gaze&Gesture と Gesture, Gaze&Gesture と None との間に有意差が見られ (いずれも  $p < 0.01$ )、Gaze&Gesture が最もユーザビリティ評価が低いとわかった。

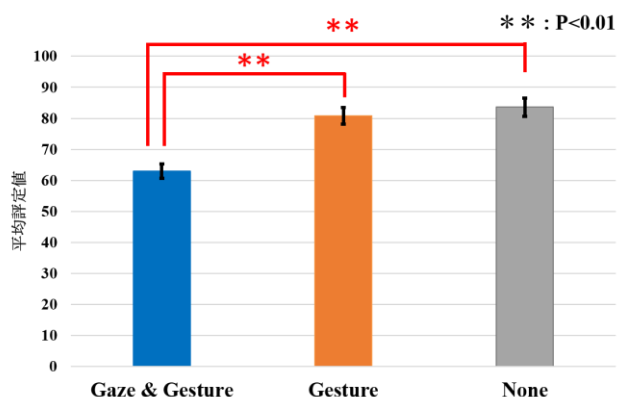


図 15 ユーザビリティ評価のための SUS 結果

## 5.6 主観的メンタルワークロードチェックリスト

各条件の主観的メンタルワークロードチェックリストの結果を図 16 に示す。一元配置分散分析を行った結果、すべての項目で主効果がみられた。Bonferroni 法による多重比較の結果、すべての項目で Gaze&Gesture と Gesture との間に有意差が (精神負担, 身体負担, 努力, フラストレーション:  $p < 0.01$ , 時間圧力, 作業成績:  $p < 0.05$ )、時間圧力を除いた全ての項目で Gaze&Gesture と None との間に有意差があり (精神負担, 身体負担, 努力, フラストレーション:  $p < 0.01$ , 作業成績:  $p < 0.05$ )、Gaze&Gesture 条件で精神的負担, 身体負担, 時間圧力, 作業成績, 努力, フラストレーションが有意に高いという結果が得られた。

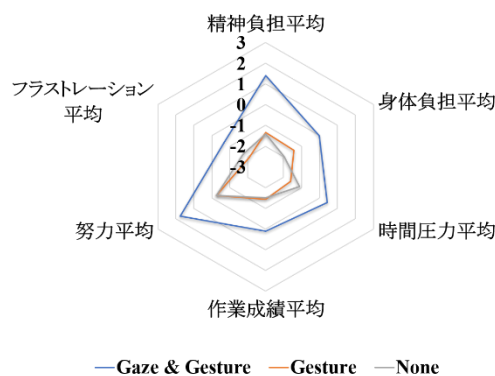


図 16 主観的メンタルワークロードチェックリスト結果

## 6. 考察

図 12 の質問 C,D の結果から、Gesture 条件ではコミュニケーションを取れている、アバタに親しみを持てたと感じたことがわかる。これは、アバタが言葉だけでなく、言語的機能が大きいとされる身体動作の顔き・首振りや指差し動作を行うことが影響していると考えられる。また質問 E から、None 条件ではアバタがどこを見ているのか理解できなかったと感じていたことがわかる。以上のことから、None 条件のアバタは視線も変わらず指差しも行わないために、被験者はアバタが不自然で人間らしくないと感じていることが考えられる。そのため、アバタに人間と同様、言葉に加えてジェスチャをさせる、視線を向けるようにすることで、アバタに人間らしさが与えられると言える。

図 15, 図 16 から、ユーザにとって頭部動作と指差し動作は容易に使いこなすことができ、身体への負担が少ないことがわかる。中でも図 13 の質問 H の結果から、特に指差し動作は思い通りにアバタに意思表示をすることができたことがわかる。このことから、本システムで意思表示として実装した頭部動作と指差し動作は、VR 空間の会話においても現実と同じように違和感なく容易に行うことができ、会話の没入感を高めると言える。

図 11 より、視線が実装された Gaze&Gesture 条件はアバタの発言後からユーザが発言するまでに時間がかかることがわかる。また、図 10 より、被験者はアバタに視線を合わせて発言するという制約がない場合には、台本を見ながら会話することが多いことがわかる。以上のことから、被験者は Gaze&Gesture 条件でも台本を見ながら会話をしたという想いで会話を進めているため、台本とアバタへの視線移動が生じ、何度もアバタに視線を合わせては発言をしようと試行していたことが考えられる。また、図 12 の質問 A から Gaze&Gesture 条件が他の 2 条件よりも話しやすさが低いこと、図 15 や図 16 からユーザビリティが低く、負担が高い結果となったことから、今回設定したアバタに視線を 2 秒合わせて発言するという行為が、被験者にとっ

て困難でストレスを感じるものであることがわかる。図 9 で Gaze&Gesture 条件が他の 2 条件の倍近いタスク完了時間であることから、裏付けが取れる。

これに対し、従来研究[11]では台本に目を通した後に視線を合わせて発言することが困難ではなく、ユーザに視線を合わせるよう促すことは人間同士の会話に近づくという結果が出ている。これは従来研究では練習回数を複数設けたのに対して、本研究は台本が長い実験時間が長くなることを考慮し、練習を Gaze&Gesture 条件の 1 回としたことが影響していると考えられる。また、従来研究では台本をアバタとの間に提示して瞬時に台本を見直すことが可能であったのに対し、本研究ではこれを不自然な動作だと考え、台本をアバタから離れた。それゆえに、台本を無意識に見たときにアバタから視線を外す時間が 2 秒を超えてユーザは繰り返し発言することになった可能性がある。そこで、練習を複数回行うこと、またユーザが VR 空間で台本を手を持つようにすればこれを改善でき、人間同士の会話に近づくと推測できる。

なお台本を暗記した後に注視して発言することが困難であった一方で、英会話練習の観点から見ると、欧米では相手の目を見て話さない人は不誠実な人とされる[20]ことから、本システムは相手を見て話す英会話練習としては有用であると言えるだろう。

## 7. おわりに

本研究では、VR 空間のアバタとの会話においてユーザ側の非言語情報が持つ効果を評価することを目的とした。そこで、ユーザおよびアバタの視線や身体動作が会話の進行に変化を与える英会話練習システムを試作した。そして、この英会話練習システムを用いて、視線や身体動作などを組み合わせた台本が異なる 3 つの条件で、ノンバーバルコミュニケーションの実験を行った。

その結果、VR 空間のアバタとの会話においても、現実でのコミュニケーションと同様、ユーザとアバタの両者において非言語情報が現実の会話を再現する上で重要であること、アバタに人間らしさを与えることがわかった。また意思表示の際に頷き首振り、指差し動作を行うことは、ユーザにとって負担は小さく、ユーザビリティが高いことがわかった。特に指差し動作は意思表示が容易で、会話の没入感を高めることが判明した。

そして、アバタに視線を合わせて発言するという制約をユーザに与えることは、適切な注視時間に設定できれば、相手の顔を見て行う人間同士の会話に近づく可能性がある。一方で適切な時間でない場合には、ユーザにとっては困難でストレスや負担を与えることがわかった。しかし、ユーザの視線でアバタの発言が変化することは、相手の目を見て話す英会話練習となり有用であると考えられる。

今後の課題としては、日本人や欧米人の視線モデルを基

にアバタの視線行動を調整する、適切な注視時間の設定、アバタとユーザの両者に表情や声の大きさなどの様々な非言語情報を追加することで、より人間同士の会話に近づくのか、各非言語情報が VR 空間の会話においてどのような効果を持つのか検証を行う必要がある。

## 参考文献

- [1] “Smart Tutor” . <https://www.plusone.space/>, (参照 2021-2-15).
- [2] “Mondly” . <https://ja.mondly.com/>, (参照 2021-2-15).
- [3] 黒川隆夫. ノンバーバルインタフェース(ヒューマンコミュニケーション工学シリーズ). オーム社, 1994, 221p.
- [4] 鈴木直人, 廣井富, 藤原祐磨, 千葉祐弥, 能勢隆, 伊藤彰則. 英会話学習システムにおける CG キャラクタの効果と学習者の発話タイミング制御のための付加表現に関する検討. 日本音響学会研究発表会講演論文集, 2014, vol.2014 秋季, p. ROMBUNNO.2-8-6.
- [5] Birdwhistell, R. L.. Kinesics and context: Essays on body motion communication. University of Pennsylvania Press, 1970, 352p.
- [6] Merabian, A.. Communication without words, Psychological Today 2, 1968, p.52-55.
- [7] Efron, D.. Gesture and Environment, King's Crown, 1941, 184p.
- [8] Ekman, P. & Friesen, W. V.. The repertoire of nonverbal behavior. Semiotica 1, 1969, p.49-98.
- [9] 成宮良, 河野和宏. 仮想世界内でのノンバーバルコミュニケーションに関する一検討. 電子情報通信学会大会講演論文集, 2017, Vol.2017 総合大会, p.ROMBUNNO.H-2-9.
- [10] 齊藤真生, コルドバユキコ, 櫻井淳. VR 空間におけるスペイン語対話学習システムの提案. 情報処理学会インタラクシオン 2020 論文集, 2020, 3B-41, p.940-945.
- [11] 植田智裕, 小宮山撰. VR 英会話におけるアバタとのノンバーバルコミュニケーションの効果. 第 173 回研究会報告ヒューマンインタフェース・ステップアップキャンプ 2020 (SIGOD-03/SIG-MAS-12, ヒューマンインタフェース学会学会誌, 2020, Vol.22, No.1, p.5-10).
- [12] “Tobii pro VR Integration” . <https://www.tobii.com/siteassets/tobii-pro/product-descriptions/tobii-pro-vr-integration-product-description.pdf?v=1.7>, (参照 2021-2-15).
- [13] “手のモデル” . <https://vr.tobii.com/sdk/develop/unity/unity-examples/social/>, (参照 2021-2-15).
- [14] “プロ生ちゃん” . <https://pronama.jp/2014/08/20/kurei-kei-3d-model-unity-package>, (参照 2021-2-15).
- [15] “音声合成サイト” . <https://ttsmp3.com/>, (参照 2021-2-15).
- [16] “IntelRealsense SDKI” . <https://www.intelrealsense.com/developers/#downloads>, (参照 2021-2-15).
- [17] John Brooke.. SUS-A quick and dirty usability scale. Usability evaluation in industry, 1996, p.189-194.
- [18] Sandra G Hart and Lowell E Staveland.. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In Advances in psychology, 1988, Vol. 52, p.139-183.
- [19] “主観的メンタルワークロードチェックリスト” . <http://acpsy.hus.osaka-u.ac.jp/questionnaire.html>, (参照 2021-2-15).
- [20] 福井康之. まなざしの心理学. 創元社, 1984, 397p.