頷き誇張により会話中の「間」を合わせるオンライン会話支援システム

窪田 太一[1] Monica Perusquía-Hernández[1] 磯山 直也[1] 内山 英昭[1] 清川 清[1]

概要：本研究では，非対面コミュニケーションにおいて話者の頷きを誇張し，引き込みのきっかけを与えるシステムの実現を目指す．本稿ではその第一歩として我々が提案するリアルタイム頷き誇張システムがコミュニケーションに与える影響を調査した．リアルタイム頷き誇張システムとは，オンライン会話システム同様に発話者固有の顔の特徴や表情変化をリアルタイムで反映しながら，頷きのピッチ角度を指定した倍率に誇張できるシステムである．実験では，頷き誇張あり・なしの 2 条件を用意し，被験者 11 組に対して対話実験を行った．評価は 14 項目のアンケート及び自由回答により実施した．その結果，全ての項目において 2 条件間で有意差は認められなかった．なお，会話の教示に不備のあった 2 組を除き，9 組のデータのみで検定した場合は Familiarity / 親近感に有意傾向 ($p = 0.062$) が見られた．今後，実験設定や教示方法を見直すことで，提案システムの有効性が示される可能性がある．

Proposal for an Online Conversation Support System to Match “Pauses” in Conversation by Nodding Exaggeration

TAICHI KUBOTA[1] MONICA PERUSQUÍA-HERNÁNDEZ[1] NAOYA ISOYAMA[1]
HIDEAKI UCHIYAMA[1] KIYOSHI KIYOKAWA[1]

1. はじめに

人は対面での会話において言葉によるバーバルな情報だけでなく，頷きや身体動作などのノンバーバルな情報を相互に同調させることで間の合ったコミュニケーションを構築している [1]．このようにコミュニケーションにおいて引き込み現象が起きることが知られている．引き込み現象とは，異なる振動のリズムがそろっていく現象であり，人においても，複数人の間で別々の身体的リズムで行われている行動が，徐々にリズムが揃っていくことが確認されている．間のあう会話では頷きや瞬き，相槌，交替潜時などの同調がよく観察されることが知られている．そして，人はこの引き込み現象により，本来は共有されない主観的な時間や空間である「間」を共有し，合わせることで円滑なコミュニケーションを実現している [2]．その例として，互いの発話タイミングが重ならない会話がある．

しかし，これらの現象をビデオ通話など，互いの身体的な動作が伝わりづらい非対面コミュニケーションで引き起こすことは困難であるとされている [3]．また，引き込み現象に関する研究は，人と人工物（ロボットなど）の会話を対象にした研究が多く，実社会の人同士の会話改善を対象とした研究は少ない．

そこで，我々は非対面コミュニケーションにおいて話者の頷きを誇張し，引き込みのきっかけを与えるシステムの実現を目指す．本稿では，その第一歩として我々が提案するリアルタイム頷き誇張システムがコミュニケーションに与える影響を調査した実験について報告する．

2. 関連研究

円滑で協調的対話において，強い信頼関係を形成する話者同士ほど，姿勢を模倣したり，動きが同期したりしやすいことが知られている [4]．この引き込み現象はあらゆるノンバーバル情報に見られ，どのような情報に起こるか，またどのようなメカニズムがあるのか調査が行われている．小松らは，対話における音声の発話速度（話速）に着目し調査を行った．その結果，話速に引っ込み現象が観察され

[1] 奈良先端科学技術大学院大学

ることが明らかとなった [5]．三宅らは，引き込み現象の仕組みを明らかにするためには，人の主観領域における予測的時間の知覚と処理の機構が解明されなければならないと考え，最も単純な実験系として同期タッピング課題を用いて，人の予測的な振る舞いの 機構を調べた [6]．その他，ロボットやエージェントとの円滑な会話を実現するため，それらへの引き込み現象の適用も検討されている．李らは瞬きの引き込み現象を援用したエージェントを提案し，被験者がエージェントの話を聞くときに引き込み現象が起こることを示している [7]．このように，引き込み現象に関する研究は，人と人工物（ロボットなど）の会話を対象にした研究が多く，実社会の人同士の会話改善を対象とした研究は少ない．

また，オンライン会話を支援する方法として，頷きが着目されている．徳原らはオンライン会話において発話者が表出する頷き反応をカウントし表示することで，自分や他人の仕草を意識するようになり，発話者に自身の振る舞いに対する気づきを与えることができることを示した [8]．その他，渡辺らは身体動作と会話音声の同期に着目し，音声入力からコミュニケーション動作を自動生成する音声駆動型身体引き込みインタラクションシステム InterActor を開発し，会話支援の有効性を示している [9]．石井らは仮想空間においてアバタの影に対話者の身体動作と自動生成による頷き反応を重畳合成する身体的引き込みアバタ影システムを提案し，アバタ影の頷き反応提示により対話者の頷き反応が誘発することを示した [10]．しかし，これらの研究ではアバタを用いるため，オンライン会話システム同様に発話者固有の顔の特徴や表情変化をリアルタイムで反映することはできない．また音声により頷きを予測生成する方法では，同意できない内容に対し，相手話者に同意したと誤って捉えられる可能性がある．さらに，アバタの表現方法に関しても検討が進められている．Yu らの研究によると点群とバーチャルキャラクターを印象を比較した結果，点群はバーチャルキャラクターより，共同存在感 (co presence)，社会的存在感 (social presence)，行動の印象 (behavioral impression)，人間らしさの認識 (humanness) が優れていることが示されている [11]．

従って，オンライン会話システム同様に発話者固有の顔の特徴や表情変化をリアルタイムで反映しながら，頷きを誇張できるシステムは提案されていない．

3. リアルタイム頷き誇張システム

図 1 にリアルタイム頷き誇張システムの概要を示す．本システムはオンライン会話において，映像上でリアルタイムに発話者の頷きを誇張する．発話者は互いに頷きが誇張された相手の映像を見ながら会話を行う．これにより，オンライン会話等の非対面コミュニケーションにおいて伝わりづらいとされている頷きが伝わりやすくなり，引き込みのきっかけを提示できると考える．

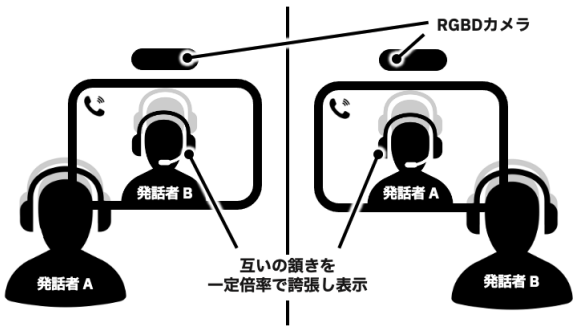

図 1 リアルタイム頷き誇張システムの概要.

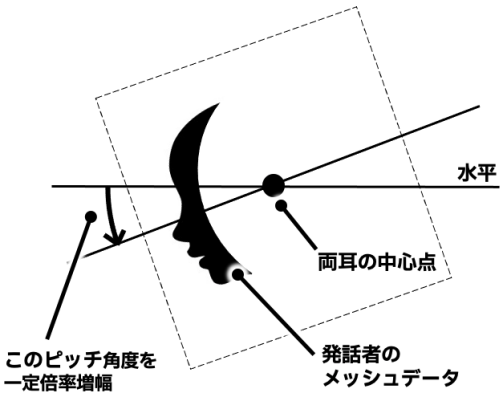

図 2 リアルタイム頷き誇張の概要.

このシステムを実現するためには，まず，既存のオンライン会話システム同様に発話者固有の顔の特徴や表情変化をリアルタイムで反映する必要がある．さらに，それに加え頭部のピッチ角度を増幅した映像を常に生成する必要がある．例えば，下を向いた場合はより下を向いた映像，上を向いた場合はより上を向いた映像が必要になる．そこで，RGBD カメラで取得した三次元点群から構成した発話者頭部の 3D メッシュを用いることで実現する．

図 2 にリアルタイム頷き誇張の概要を示す．顔の中心を検出してそこを中心にピッチ角度を増幅するように話者頭部の 3D メッシュを回転させる．

3.1 システム構成

本システムは 3D メッシュ表示モジュール，頭部検出モジュール，頷き誇張モジュールの 3 つのモジュールにより構成される．まず，3D メッシュ表示モジュールにより発話者の頭部の 3D メッシュを取得する．次に，頭部検出モジュールにより頭部の位置取得し，頭部の領域のみを切り出す．そして，頷き誇張モジュールにより頭部のピッチ角度を増幅するように頭部領域の 3D メッシュデータを回転させることで頷きを誇張表現する．以降では各モジュールの詳細について述べる．

図 3 生成した 3D メッシュ.

3.2 3D メッシュ表示モジュール

3D メッシュ表示モジュールでは Azure Kinect (Microsoft 社) を用いて発話者の 3D メッシュを作成し，画面に表示する．3D メッシュを生成する手順について説明する．まず Azure Kinect により，Depth 画像 (解像度: 640×576) と RGB 画像 (解像度: 1280×720) を取得する．次に，RGB 画像に深度情報を重ね合わせるように変換し，三次元点群を得る．最後に，得られた頂点からポリゴンを生成する．この時，頂点が三角形を紡ぐように繋ぐ順番を指定する．この処理は高さ方向，幅方向に順に実行されるため，隣り合う頂点同士で三角形が作られる．描画は Unity (Unity Technologies 社) により行う．図 3 に Unity のゲームウィンドウに出力した 3D メッシュの例を示す.

3.3 頭部検出モジュール

頭部検出モジュールでは 3D メッシュにおける発話者の頭部位置を検出し，頭部のみの 3D メッシュを抽出する．頭部位置の検出には Azure Kinect Body Tracking SDK (Version 1.1.1) を利用する．Azure Kinect Body Tracking SDK は Microsoft 社が提供する Azure Kinect DK を使用したボディトラッキングアプリケーションを構築するためのヘッダーとライブラリである．このライブラリで取得可能なスケルトンは階層を持つ 32 の関節を含んでいる．この SDK により，RGB 画像中の鼻の座標を取得する．取得した鼻の座標を中心に半径 150 ピクセルの円を定義し，その外側の 3D メッシュの深度情報を 0 に書き換えることで頭部領域の 3D メッシュを切り出した．頭部領域を切り出した 3D メッシュの例を図 5 に示す.

3.4 頷き誇張モジュール

頷き誇張モジュールでは発話者の頭部のピッチ角度を算出し，指定した倍率で 3D メッシュをピッチ方向に回転させる．頭部の角度は Azure Kinect のボディトラッキングにより，得られたスケルトンの両耳の中点と鼻を繋ぐベクトルと水平ベクトルの為す角度 θ_0 とした．角度は水平方向を 0 度とし，ピッチ軸において上向きを正方向，下向きを負方向とした．しかし，図 6 に示すように一般的に鼻は両耳の中心よりも下にあるので，正面を向いている際に角度が常に負の値になる．この角度 θ_0 を補正するため，取得した角度から θ_0 を引いた値を現在角度 θ_n と設定した．なお，この補正は被験者毎に行う必要がある.

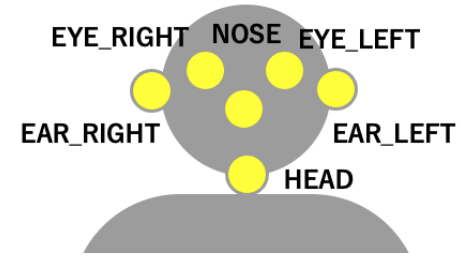

図 4 Azure Kinect Body Tracking SDK で取得可能な関節位置.

図 5 頭部領域を切り取った 3D メッシュ.

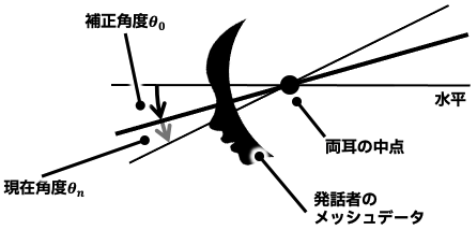

図 6 頭部の角度取得と補正.

誇張がない場合には図 7 に示すように，先述のずれが補正された θ_n が現在角度かつ，目標角度になる．n 倍に誇張する際には，目標角度 θ_e が補正された角度 θ_n の n 倍となる．そして，目標角度まで Unity 内で 3D メッシュを回転させた．なお，回転中心は両耳の中心とした.

4. 違和感を感じない最大の頷き誇張倍率に関する予備調査

本システムにおいて適切な誇張倍率を調査するため，予備調査を行った．調査はオンライン上で実施した.

4.1 調査方法

予備調査では，第一著者が異なる倍率で頷く様子をあら

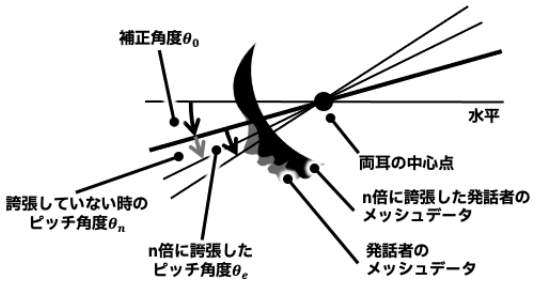

図 7 頭部角度の誇張.

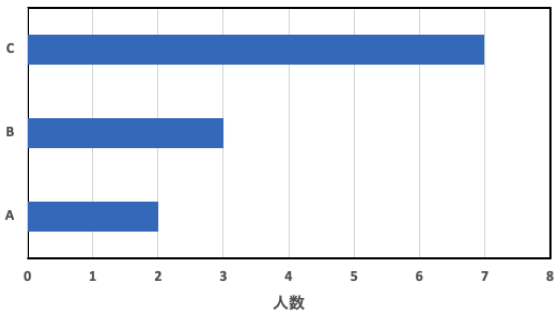

図 8 違和感を感じない最大の頷き誇張倍率の調査結果．A (1.2 倍) を 2 名，B (1.4 倍) を 3 名，C (1.7 倍) を 7 名がそれぞれ選択した.

かじめ撮影した動画を参加者に見比べさせた．そして，「違和感を感じない最大の誇張倍率」を選択させ，その回答理由を記述させた．倍率は初項 1，公比 1.2 の等比数列に従い 1.2, 1.4, 1.7, 2.1, 2.5, 3.0, 3.6, 4.3, 5.2（小数点第二位を四捨五入）の 9 段階とした．なお，数値による先入観を排除するため，選択肢にはそれぞれ小さい倍率から順に A〜I の名前をつけた．動画では第一著者がおよそ 25 度ほど上向きに頭を上げ，水平より少し下に落とす「大きな頷き」を 2 回，そして，水平方向に対し 10 度程度小刻みに上下に頭を動かす「小さい頷き」を 5 回，それぞれの倍率で録画している.

4.2 調査結果

参加者は 22 歳から 51 歳（中央値 24 歳）の男女 12 名（内 7 名が男性，6 名が女性）であった．文化による影響も考慮するために，人生の中で最も長く過ごした国を回答させており，その回答結果は，8 名が日本，3 名が中国，1 名がメキシコであった.

アンケートによる調査結果を図 8 に示す．A（1.2 倍）が違和感を感じない範囲で一番大きな誇張倍率であると答えたのは参加者の内 2 名，B（1.4 倍）が 3 名，C（1.7 倍）が 7 名であった．また女性参加者 5 名の内，4 名が A もしくは B を選択していた．自由記述欄では「D 以降は頷く際に顎の裏まで見えてしまい違和感を感じた」，「D 以降は動作がわざとらしく、同意や承諾を表していないように見えた」，「B 以降はカメラから視線をそらしているように見えた」などの記述があった．これらの結果から，B と C を合わせると 1.4 倍の誇張であれば 83.3%の参加者が違和感を感じないことがわかる.

4.3 考察

調査結果を見ると，女性参加者 5 名の内，4 名が A，B を選択していたことから，頷きの誇張に関しては女性の方が敏感である可能性がある．回答が A（1.2 倍）〜C（1.7 倍）に集中していたことから 1.2 倍刻みの選択肢は粗すぎた可能性がある．他に，「人生の中で最も長い時間を過ごした国」に日本以外と回答した 4 名の回答は，中国と回答した 3 名（内 2 名が男性，1 名が女性）が共に C，メキシコと回答した 1 名が B を選択した．これらの回答は日本と回答した他の参加者と似通っており，文化による差はあまりない可能性がある．しかし，本調査ではデータが十分でないため，追加での調査が必要である.

5. リアルタイム頷き誇張システムの評価実験

実装したリアルタイム頷き誇張システムにおいて，頷きを誇張する場合と誇張しない場合を比較し，コミュニケーションがどのように変化したのかをアンケートを通じて定性的に評価する.

5.1 実験方法

日常の会話を想定した対話を被験者に行わせる実験を行った．実験条件は頷きを誇張する場合と誇張しない場合の 2 条件である．どちらの条件も頭部のみを切り取った 3D メッシュを表示した．頷きを誇張する条件における倍率は前項の予備調査より 1.4 倍とする．図 9 に実験の様子を示す．対話は 2 名 1 組で行わせ，被験者同士は衝立越しに背中を合わせており，お互いの声は衝立越しに聞こえるようになっている．Azure Kinect は互いに交差するように PC に接続されており，ディスプレイには相手の 3D メッシュのみが表示される．会話時間は 1 条件 2 分間とし，開始から 1 分が経過したタイミングで話題を変更するよう説明した．同意が起こりやすいよう一つの答えを出すための討論でなく，様々な答えがあってもいいような日常的でかつ，肯定的な雑談を想定し，以下に示すような複数の話題を用意した.

- What do you like about NAIST? / NAIST のいいところ
- Favorite food / 好きな食べ物
- Favorite animal / 好きな動物
- What I would like to do when I become rich / お金持ちになったらしたいこと
- Places I would like to travel / 旅行に行きたい場所

実験条件ごとに被験者それぞれにこれらの話題を書いた

紙を事前に手渡した．話題を提供する順番は交互に行うように指示した．話題を指定した理由は，どちらかの一方的な会話でなく両者がバランスよく会話を行うよう誘導するためである．会話終了後には条件ごとのアンケートにより以下の 14 項目による 7 段階のリッカート尺度（中立 4）の定性評価をさせた．

- Enjoyable / 楽しさ
- Preference / 好み
- Ease of dialogue / 対話しやすさ
- Sense of unity / 一体感
- The excitement of the place / 場の盛り上がり
- Familiarity / 親近感
- Naturalness / 自然さ
- I felt they were interested in me / 相手が私に興味を持っていると感じた
- The other party isn't taking the conversation seriously / 相手は話を真剣に聞いていない
- The other person agreed with me a lot / 相手が自分によく同意した
- We agreed with each other / 互いの意見が一致した
- The movement was in sync with the other party / 相手と動作が同調した
- Conversation felt smooth / 会話がスムーズに感じた
- It felt like we were breathing together / 互いの間が合うように感じた

条件ごとのアンケート項目は石井ら [10] による尺度を参考に，意見や動作の同調に関する項目を追加した．また，最後に自由記述にて次のような内容を尋ねた．

- Did you notice any differences in the two conversations? / 二つの会話で違いを感じましたか？
- Please describe any other observations or impressions you may have / その他に気づいたことや感想があれば記述お願いします

条件の提示順は順序効果を考慮して組ごとにカウンタバランスをとった．

5.2 実験手順

実験の手順について説明する．まず被験者に対し，2 分間 2 セットの会話をすること，1 セットの会話の中で 1 分ごとに話題を変えること，話題は紙により指定されること，そして 1 セット目は右に座った被験者から，2 回目は左に座った被験者から話し始めることを説明し，事前アンケートに回答するよう指示する．次に，被験者に椅子に座らせ，椅子の位置と高さを調整する．椅子の位置は両者の背もたれが衝立につくように，椅子の高さは互いの顔がディスプレイの中心に映るように調整させる．同時に，実験者が 3.4 節で説明したように，初期角度のキャリブレーションを行う．次に，実験者が話題が書かれた紙をそれぞれの被験者に手渡す．実験者の合図に従い会話を開始させ，1 分が経過したタイミングで話題を変えるよう指示を行う．会話が 1 セット終わるごとにアンケートに回答させる．会話の間に設定を変更するため被験者 2 人には，話さないよう説明した上で数分間待機するよう指示する．次の会話でも同じように話題を提供し合計 2 分間の会話をするよう指示する．最後に事後アンケートに回答させ実験を終了した．

5.3 実験結果

被験者は 22 歳から 36 歳（中央値 24 歳）の男女 22 名 11 組（内 13 名が男性，9 名が女性）であった．また，文化による影響も考慮するために，事前アンケートにて人生の中で最も長く過ごした国を回答させた．結果は，日本が 12 名，インドネシアが 4 名，バングラデシュが 2 名，アルメニア・インド・マレーシア・韓国 1 名ずつであった．言語は日本語または英語とし，被験者の希望言語を考慮し同じ言語で話すよう組を作った．同じ組の被験者とは普段どの程度話すか（知り合いかどうか）も回答させた．結果は 7 組が初対面であり，4 組は毎日もしくは週に 1 度話す程度であった．

条件ごとのアンケートによる 14 項目の評価結果を図 10 に示す．帰無仮説を条件間に差は無いとして Wilcoxon の符号順位検定によって評価した．2 条件の間で全ての項目において有意差は認められず，帰無仮説は棄却されなかった．

自由記述については「Did you notice any differences in the two conversations? / 二つの会話で違いを感じましたか？」に対して，1 件「さっきよりは相手のうなずきとか動作がみやすく，話しやすかった気がした」と頷きに言及した回答が得られた．しかし，回答の内 11 件が「違いを感じなかった」という内容であった．その他の回答は違いは感じていると回答しているものの本来の趣旨とは異なる表示された顔の大きさ等に触れていた．他に「Please describe any other observations or impressions you may have / その他に気づいたことや感想があれば記述お願いします」に対しては，1 件「頷きを強調するシステム？だと感じたが，少し画面酔いしやすい気がした」という回答が得られた．また，趣旨とは異なるが「表情がよく見えるので普通の通話よりも相手の反応がわかりやすく話しやすかった」という回答も存在した．

5.4 考察

本実験では 2 条件の間で全ての項目において有意差は認められなかった．さらに，多くの被験者が 2 条件の違いについて気づかなかったと回答した．この結果について 2 点の原因が存在すると考えている．

図 9 実験の様子.

1 点目は，今回設定した誇張倍率が適切でなかった可能性である．予備調査では動画では頷きが誇張されているということを説明した上で，参加者に違和感を感じない範囲で「一番大きな誇張倍率」を選択させた．その結果，多くの参加者が 1.2 倍〜1.7 倍で違和感を感じると回答した．しかし，これは「誇張されていること」を知った上での違和感であり，「誇張されていること」を知らなければ違和感は元より気づくこともない可能性がある．実験同様「誇張されていること」を知らない前提で，複数の誇張倍率の動画を一対比較させ，何も知らなくても違いに気づくことができる倍率，そして違和感を感じる倍率を調査する必要がある．

2 点目は，実験設定及び実験時のオペレーションによる不備である．本実験では 11 組の被験者が会話を行なったが，その中で 2 組はオペレーションのミスにより意図しない形式の会話が行われた．具体的にはその 2 組は交互に発話を繰り返す会話でなく，1 分間ずつ一方的に話題について意見を話すというシチュエーションとなった．これは単に最初に話し始めた話者の話が長いことに釣られて，もう一方の話者も合わせたのか，時間いっぱい一方的に意見を話すという課題という風に認識していたかは不明であるが，オペレーションで会話のシチュエーションを一意に定められなかったことは事実である．

なお，会話が一方的なトークになってしまった 2 組を除き，9 組のデータのみで Wilcoxon の符号順位検定を行なったところ，Familiarity / 親近感に有意傾向 ($p = 0.062$) が見られた．また，現在のシステムと現在の実験設定（会話のシチュエーション）では，ほぼ頭部を動かさない被験者が存在した時，そもそも頷きが誇張されることがない．そこで，最低でも何度かは頷くような実験設定を設ける必要があると考えられる．

6. おわりに

本研究では，オンライン会話において伝わりづらいとされるノンバーバル情報である頷きをリアルタイムに誇張表現する「リアルタイム頷き誇張システム」を開発し，コミュニケーションへの影響を調査した．しかし，日常会話を想定した対話実験では全ての項目において有意差を得ることはできなかった．今後は，実験設定を見直して，引き続き本システムが引き込み現象を誘発できるかを調査する．

参考文献

[1] W.S. Condon and L.W. Sander: Neonate Movement is Synchronized with Adult Speech, *Science*, Vol. 183, Issue 4120, pp. 99–101 (1974).

[2] 三宅美博: 共創的コミュニケーションと「間（ま）」—リズム運動とその同調を内側から支援すること—, バイオメカニズム学会誌, Vol. 36, No. 2 pp. 97–103 (2012).

[3] 渡辺富夫, 大久保雅史: コミュニケーションにおける引き込み現象の生理的側面からの分析評価, 情報処理学会論文誌, Vol. 39, No. 5, pp. 1225–1231 (1998).

[4] F.J. Bernieri, J.S. Gillis, J.M. Davis, and J. G. Grahe: Dyad Rapport and the Accuracy of Its Judgment Across Situations: A Lens Model Analysis, *Journal of Personality and Social Psychology*, Vol. 71, No. 1, pp. 110–129 (1996).

[5] 小松孝徳，森川幸治: 人間と人工物との対話コミュニケーションにおける発話速度の引き込み現象，情報処理学会研究報告，Vol. 2004，No. 105，pp. 71–78 (2004).

[6] 三宅美博, 大西洋平, エルンストペッペル: 同期タッピングにおける 2 種類のタイミング予測，計測自動制御学会論文集, Vol. 38, No. 12, 1114/1122 (2002).

[7] 李明輝，福田悠人，小林貴訓，久野義徳: 瞬きの引き込み現象を援用した対話エージェント，情報処理学会研究報告，Vol. 2018-CVIM-212，No. 33 (2018).

[8] 徳原耕亮, 荒川 豊, 石田繁巳: 頷きのリアルタイムフィードバックによるビデオ会議, マルチメディア, 分散協調とモバイルシンポジウム 2021 論文集, pp. 953–959 (2011).

[9] Watanabe, T., Danbara, R., Okubo, M.: InterActor: Speech- Driven Embodied Interactive Actor, International Journal of Human-Computer Interaction, Vol. 17, No. 1, pp. 43–60 (2004).

[10] 石井 裕, 江崎敬三, 渡辺富夫: アバタを介したコミュニケーションを支援する身体的引き込みアバタ影システム, ヒューマンインターフェース学会論文誌, Vol. 18, No. 3, pp. 249–260 (2016).

[11] K. Yu, G. Gorbachev, U. Eck, F. Pankratz, N. Navab, and D. Roth: Avatars for Teleconsultation: Effects of Avatar Embodiment Techniques on User Perception in 3D Asymmetric Telepresence, *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, Vol. 27, No. 11, pp. 4129–4139 (2021).

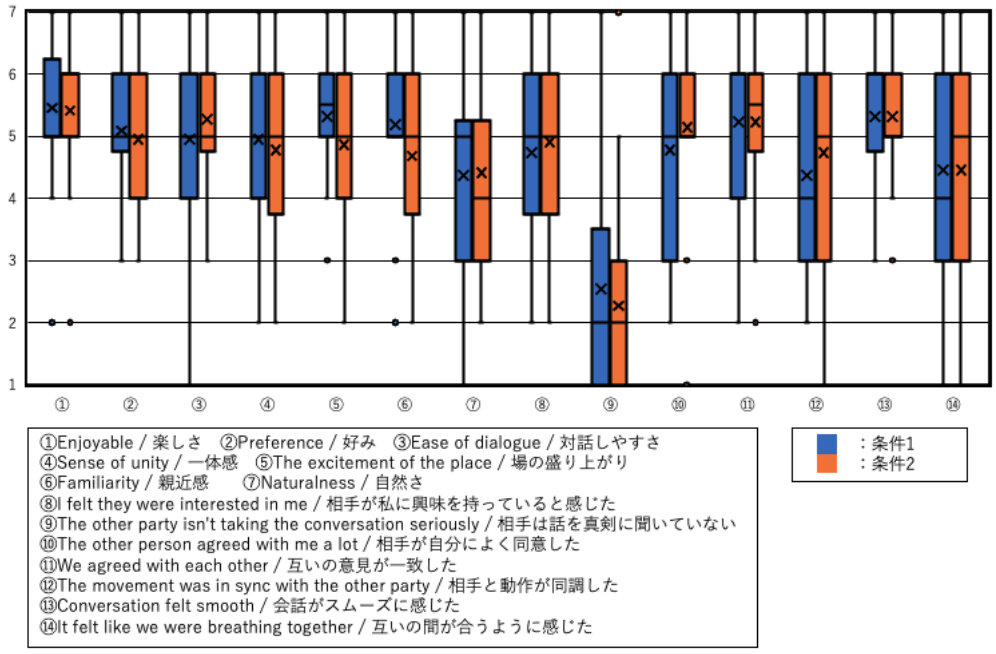

図 10 本システムにおける 7 段階評価結果.