

# 星空観察における起点となる星のズレが会話に及ぼす影響： 仮想環境での検証

飯田 空<sup>1,a)</sup> 中村 聡史<sup>1</sup>

**概要：**2人で星空観察を行い他者と特定の星や星座について会話する際、観察者間で同じ対象を見ているつもりでも参照対象が一致しないまま会話が進行し、互いにどの星について話しているか混乱することがある。そのため観察者間で見ている星を一致させる支援システムが必要である。そこで本研究では、星座説明の初手で参照される起点星の不一致に着目し、この不一致が会話に与える影響を調査した。実験では星空環境を仮想空間で再現し、起点星が一致する条件と一致しない条件を設けた2人1組による星座探索タスクを実施した。分析では、「え？」などのフィラーや発話間隔といった会話進行に伴う非言語的特徴を中心に扱った。その結果、起点星が一致しない条件では会話序盤において沈黙や発話交代間隔が増加する傾向が確認された。また、最終的な正答率も低下した一方、タスクによっては会話途中で参照対象の不一致が解消される場合があることが示された。

**キーワード：**星空、星空観察、星座、ズレ、会話、発話間隔、VR

## 1. はじめに

星空観察は星や星座に関する知識を深めることができるだけでなく、実際の星空を見ながら会話をするという楽しみ方もあるため、人気のアクティビティのひとつである。一方、星は手の届かない距離に存在するため、星空観察における会話では、自分が注目している星や星座の位置を相手に正確に伝えることは難しい。このとき、自分と相手と同じ星を見ていると誤認したまま会話が進行すると、途中でズレに気づいた際に互いに何を指していたのかが分からなくなり、会話の修正が難しくなる。その結果、会話が成立して見えるにもかかわらず、誤った理解が共有されたまま終わる可能性がある。つまり、星空観察における星座探索および会話をより楽しむために、観察者同士で星の認識のズレを気づかせることは重要である。

星空観察を支援する手段としては、「Star Walk2」<sup>\*1</sup> といったスマートフォンアプリケーションが広く利用されている。これらのアプリケーションでは、複数人で星空観察を行う際に画面を共有することで、話題となっている星や星座についての認識を容易に一致させることができ、自分が見ている星を相手に伝える難しさを軽減できる。しか

し、画面を注視することにより、画面の明るさによって目が夜空に順応しづらくなり、本来見えるはずの暗い星が見えにくくなるという問題がある。また、プラネタリウムや天文台が開催する観覧会などでは、レーザーポインターを用いて特定の星を指し示すことで、大人数の観察者間で注視対象を一致させる方法が用いられることがある<sup>\*2</sup>。これは視覚的に直感的で有効である一方、他の観察者グループが存在する場合には使用を控える必要があるなど、利用できる状況が限られている。このように、視覚的な補助情報に強く依存した支援手法には、一定の制約が存在する。

以上のことより、星空観察における星座探索をより楽しむために、視覚的な追加情報に頼らず、それ以外の情報で見ている星のズレに気づけるよう支援することが重要であると考えられる。こうした問題を踏まえ、音声アシスタントの介入により、複数人で行う星空観察における会話内容から、観察者間で見ている星のズレを推定し、観察者同士で認識を一致させることを目指している。この支援手法実現に向けて、これまでの研究においてズレが起こりやすい観察者間の視力の差が星空観察の会話に与える影響について調査を行ってきた [1]。その結果、視力の差がある環境において、「え？」や発話間隔などが増加する傾向が見られ、会話指標から見ている星のズレの推定可能性が示唆さ

<sup>1</sup> 明治大学  
4-21-1 Nakano, Nakano-ku, Tokyo, 164-8525 Japan

<sup>a)</sup> sorabun.iida@gmail.com

<sup>\*1</sup> Vito Technology: Star Walk2, <https://starwalk.space/ja>.

<sup>\*2</sup> GIGA LASER: 星空観察・天体観測に適したおすすめ  
のレーザーポインター, <https://giga-laser.com/by-purpose/for-stargazing.html>.

れた。

ここで、観察者間の視力差は見ている星のズレを生み出す要因の1つであるものの、実際にズレが起きている際の会話指標と実際のズレを対応づけて推定できるかは十分に検討できていない。音声アシスタント実現のためには実際にズレが起きている環境での会話を分析する必要がある。ここで、星座探索では、ある星を起点として星同士の関係をたどり、星座全体の形を組み立てていく。このとき、観察者間で見ている星が一致していない場合でも、説明の流れ自体は一見整合して進むことがあり、対話者が「理解できた」と感じたまま、誤った星座構造が共有される可能性がある。実際に、我々が研究室の同僚と星空観察をしていた際、最初に着目していた星のズレが起きていたにも関わらず、一致していると思ったまま会話が進んでしまい、混乱した結果その星座についての会話が終わってしまったことがある。つまり、見ている起点星の一致・不一致はズレの有無を明確に操作でき、意図的にズレを起こすための重要な指標である。

そこで本研究では、将来的に会話の進行から認識のズレを捉えて補完する支援を実現するために、観察者間の見ている星のズレを意図的に生じさせ、その状況が星座探索過程の会話に与える影響を調査する。具体的には、星空環境を仮想環境で再現し、見ている起点星のズレがない条件とズレがある条件において、2人1組による星座探索タスクの会話実験を行い、条件間の会話の違いについて明らかにする。分析では、聞き返しとして現れる「え？」やフィルターである「うーん」などの出現頻度、発話間隔といった会話進行に伴う時間的特徴を中心に扱い、見ている星のズレが会話過程にどのように現れるかを検討する。

## 2. 関連研究

### 2.1 星空観察に関する研究

これまで、複数人での観察を支援するシステムに関する研究は多く行われてきた。Ohamaら [2] [3] は星空観察における注視点を用いた星座の共同学習環境システムを提案している。この研究では、星空を見ている学習者間でそれぞれの星空の注視点を共有可能なシステム構築をし、実験よりシステム使用後の星座に関するテストの結果が向上することを明らかにしている。また、AR技術を用いた星空観察の支援に関する研究も行われており、Planeyら [4] は、ARを用いた星空観察において、指差しなどのジェスチャが観察対象の共有や協調作業を促進させることを明らかにしている。Tianら [5] は、プラネタリウムの季節の星座コンテンツをもとにスマートフォン向けのARシステムを開発し、実験の結果、システムがユーザの星空への興味を高めることを示している。

さらに、仮想空間を用いた星空観察の応用についての研

究も行われており、蓮ら [6] は、HMDを用いてその時間に観察できる星空を仮想空間に表示し、それに対しての気づきや感想を手書きアノテーションで共有できるコミュニケーションツールを提案している。また、システムの手書きアノテーションを用いることで、自分が今見ている星や話している箇所について簡単に伝えることができることを明らかにしている。

これらの研究は、注視点の共有や提示情報の拡張など、視覚的な補助情報を用いて対象の共有を支援するものが中心である。一方、星空観察では参照対象がズレたまま会話が進んでしまうことがあり、そのような状況を会話の進み方に現れる特徴から捉える枠組みは、まだ十分に整理されていない。本研究では追加の視覚情報に頼らず、起点星の一致・不一致という操作可能な要因を導入し、会話特徴からズレの兆候を捉えるための観点を整理する。

### 2.2 協調タスクにおける会話に着目した研究

協調タスクにおいて、対話から状況や対話者の状態を推定する研究が行われている。Saekiら [7] は、オンライン会話における対話者の困惑状態を対象に、沈黙や独り言、視線の逸れなどのマルチモーダルな兆候を特徴量として抽出し、分類モデルにより困惑を推定する手法を提案している。また、豊田ら [8] は、2者対話における発話状態時間長に着目し、雰囲気推定に有効な発話時間特徴を示している。本研究で扱う発話交代間隔や沈黙時間といった時間的特徴は、この研究のような「会話進行に現れる非言語的特徴」として位置づけられる。

さらに、相互理解の手がかりに関する研究として、Hayashiら [9] は、日本語の「え」が単なる驚きではなく、理解や期待・知識のズレへの気づきを相手に提示する相互行為的な資源になり得ることを示している。また、Micklosら [10] は、会話中の「言い直して」などの修復発話に着目し、それが相互理解のズレに気づいて調整するための手続きであることを示している。これらは、ズレに対する気づきや修正の局面を捉えるうえで重要であると考えられる。

一方、視覚的对象を共有している状況において、参照対象のズレが生じた状態を、会話特徴の組み合わせとして定量的に捉え、認識の一致・不一致を推定することに着目した研究は少ない。本研究では、協調タスク等で扱われてきた会話分析の知見を踏まえつつ、星座探索会話を対象に、フィルターや発話間隔などの会話特徴から起点星のズレを推定するための観測指標を検討する。

## 3. 実験

### 3.1 概要

本実験では、介入型音声アシスタントの実現に向けて、星空観察会話支援において参加者間で見ている星が一致し

ている状態と一致していない状態で、会話がどのように異なるかを明らかにすることを目的とする。そのため、フィルター（「え」「うん」等）および参加者間の発話間隔に着目して分析を行う。本実験では以下の仮説を検証する。

- H1: 見ている星のズレがある状態では、見ている星のズレがない状態よりも会話中のフィルターが増加する。
- H2: 見ている星のズレがある状態では、見ている星のズレがない状態よりも参加者間の発話間隔が長くなる。

なお、本実験における見ている星のズレの操作方法は3.2節で述べる。本研究ではその実現手段として起点星に着目した。

### 3.2 見ている星のズレを起こすためのデザイン

参加者間の見ている星のズレが生じた状態では、参加者が「同じ星を見ている」と思い込んだまま会話が進行することがある。このような状況を実験的に再現するため、本研究ではズレを意図的に生じさせる設計として、星座説明の最初に用いられる起点星に着目した。

本実験における起点星とは、星座探索タスクにおいて星座説明を開始する際に最初に着目する星（説明の起点となる星）を指す。星座の説明は起点星に着目し、星同士の位置関係を順にたどる形で進行することが多い。ここで、起点星が参加者間で一致しない場合にも、星空は点の集合であるため会話が成り立っていると思い込んで、参照する星が一致しないまま進行することが多い。

そこで本実験では、タスク開始時に起点星を実験参加者2人に明示的に提示し（図1）、以下の条件として定義した。

- ズレなし条件：参加者間2人に提示される起点星が一致している。
- ズレあり条件：参加者間2人に提示される起点星が一致していない。

### 3.3 実験設計

本実験では、我々の過去の研究 [1] と同様に、参加者2人で仮想環境内で星座探索タスクを行った。星座探索タスクでは、3.2節の条件下においてペア内で役割を事前に決定し、一方を「伝達役」、もう一方を「理解役」とした。具体的な役割は下記のとおりである。

- 伝達役：星座の名前や構成する星の位置・形状を理解役に対して説明する。
- 理解役：伝達役の説明をもとに仮想環境内で星座を探索し、回答する。

本実験で使用される星座は、事前に筆頭著者が作成したオリジナルの複数種類の星座（計8種類）を用い、試行ごとに使用する星座を切り替えた。各ペアは異なる条件のもとで複数回の星座探索タスクを行い、各試行の会話ログか

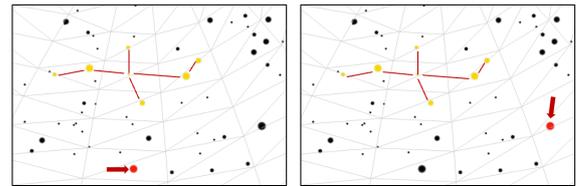


図1: 実験で用いた星座および起点星の提示例（星座：ワシ座）。Webサイト上で3次元の星空視点を再現し、対象星座の構成星を黄色、起点星を赤色で表示している。左図と右図はいずれも同一の星座を示すが、赤色で提示される起点星の位置が異なる。ズレあり条件では、このように参加者間で異なる起点星が提示される。



図2: 実験の様子。実験は空きの教室を用いて行い、外部からの干渉を一切受けない環境下で行った。

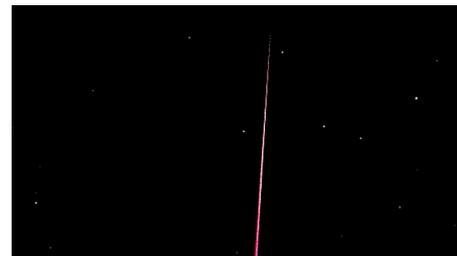


図3: 実験の視点。赤い線は実験参加者がもつコントローラから発生しているもの。実験中は、自分の注視している星をこの線で指差しながら行われる。

らフィルターおよび発話間隔を算出して条件間で比較した。

また、実験者が認識状態を会話途中も含めて把握するため、理解役に対して星座の認識が一致しているかの答え合わせをタスク開始3分時点と試行終了時に設けた。これにより、実験者が認識のズレの有無を客観的に観測できるとともに、試行中にズレが修復されたか否かを区別できると期待される。さらに、各試行後に主観アンケートを実施し、参加者の主観的な認識一致度も取得した。

### 3.4 実験システム：星空観察環境

起点星のズレが星座探索の会話に与える影響を調査するために、参加者間の見ている起点星のズレなし・ズレありを意図的に操作しつつ、星空の配置や、環境光を一定に保つ必要がある。そこで、本研究では先行研究 [1] のシステムを改良し、仮想環境内に起点星を提示できる星空観察環

境を構築した。これにより、星配置や環境条件・起点星の提示方法を統一化することで、参加者の行動および会話を同一条件下で記録することを可能とする。実験システムを用いた際の視点イメージを図3に示す。

星空観察環境はUnityで構築し、HMDとしてMeta Quest 3を使用した。星空の再現には半球状のドームを用い、参加者位置（視点）を中心として同一距離となるように星を配置した。星の見かけの大きさは複数段階を用意し、実際の星空を模倣した。前回実験では、山や月などの周辺物体を手がかりとした説明が多く、実際の星空観察に近い会話表現を十分に再現できなかったため、本実験ではそれらを削除した。本システムでは、会話音声に加えて、HMDの頭部向きとコントローラの指差し方向をログとして取得した。これにより、各試行において理解役が回答時に指していた星が星座の星かどうかを識別できる。

### 3.5 実験手順

実験の様子を図2に示す。実験開始前に、実験参加者に対して実験概要およびタスク内容、録音を含む記録項目について説明をし、同意取得後に実験を開始した。次に、参加者属性に関する事前アンケートに回答してもらい、HMDの装着およびシステムの操作方法を説明した。また、条件推測による会話への影響を低減するため、タスク開始直後に「起点星が一致しているか否か」を確認する行為や、「実験的に〜」といった発言は控えるよう説明した。さらに、タスクにおける起点星の違いについて、「伝達役と理解役は50%の確率で異なる星が提示される」と説明をした。実験説明の後、練習セッションとして短い探索タスクを1回実施し、星の指し示し方や回答方法に慣れてもらった。

試行は以下の手順で実施した。

- (1) 星座説明準備：試行で扱う星座について、伝達役へのみ星座情報（形状・構成）をスライドで提示し、説明準備を行ってもらった。
- (2) 起点星提示：試行開始時に、伝達役および理解役に起点星を提示した。提示方法および提示時間は全試行で統一し、提示時間は10秒とした。提示中、起点星は赤色で強調表示し、大小に変化するアニメーションを付与した。
- (3) 星座探索タスク：探索時間は最大6分とし、参加者同士で認識が一致したと判断した時点で申告してもらい、試行を終了した。時間内に回答に至らなかった場合は、6分経過時点で終了した。また、3分経過時点で理解役に星座だと思える星をコントローラで指してもらい、暫定回答として記録した。
- (4) 試行後アンケート・答え合わせ：試行終了後、理解役に星座だと思える星をコントローラで指してもらい、最終回答として記録した。続いて、主観評価（会話

表1: 分析項目の一覧（試行ごと）

区分	指標（要約）
タスク達成度	正答率（3分時点、最終）：理解役の指差し回答に基づき算出。
会話指標	ターン交代率 [回/分]、発話率 [回/分]、平均発話交代間隔 [秒]（オーバーラップを含む）、平均沈黙時間 [秒]（オーバーラップ区間を除外）、全フィルター率 [回/分]、「え」系フィルター率 [回/分]。
主観評価	試行後アンケート：円滑さ、3分時点の一致感、最終時点の一致感、起点星の使いやすさ、楽しさ、ストレス。

の円滑さ、認識の一致感、疲労等）に関するアンケートに回答してもらった。

- (5) 休憩：試行の負担を低減するため、3分の休憩を設けた。

参加者には固定ペアで参加してもらい、上記試行を計8回2日間に分けて取り組んでもらった。1日目には練習タスク1回に加えて本試行4回（計5試行）に取り組んでもらいい、2日目には練習を行わず本試行4回に取り組んでもらった。8試行終了後、実験全体に関する事後アンケートに回答してもらい、実験を終了とした。

### 3.6 分析指標

本研究では、起点条件（ズレなし・ズレあり）が星座探索会話に与える影響を検討するため、試行ごとに会話音声、行動ログ、およびアンケート回答を取得し、表1の項目を分析した。なお、タスク達成度および会話指標は、「タスク開始から3分まで」と「タスク全体」の2区間で算出した。

会話指標は文字起こしおよびタイムスタンプから算出した。ターン交代率は1分当たりのターン交代回数（回/分）、発話率は1分当たりの発話回数（回/分）とした。平均発話交代間隔は直前発話の終了時刻から次発話の開始時刻までの差（秒）とし、発話の重なりが生じた場合は負の値（オーバーラップ）も含めた。平均沈黙時間はオーバーラップ区間を除外し、発話間の無音区間のみを集計した（秒）。フィルター率は観測されたフィルター総数をタスク所要時間で正規化し、1分当たり（回/分）に換算した。

## 4. 結果

### 4.1 実験参加者と分析

本実験の参加者は、裸眼または視力矯正（コンタクトレンズ、眼鏡）を施した状態で視力が0.7以上の大学生および大学院生16名（平均年齢22.8歳、男性13名、女性3名）であった。なお、全参加者は2名1組の計8ペアとして実験に割り当てられた。

また、本実験では、同一ペアによる複数試行のデータが含まれるため、個々のデータが独立していないという問題がある。この問題に対処するため、分析ではペア固有の違

表 2: 起点条件別の正答率の平均と標準偏差

起点条件	3分時点 平均 (SD)	最終 平均 (SD)
ズレなし条件	0.45 (0.30)	0.65 (0.37)
ズレあり条件	0.13 (0.21)	0.26 (0.29)

いを統計的に調整したうえで、ズレなし条件およびズレあり条件の効果を検証できる線形混合モデル (LMM) を採用した。また、星座の違いによるばらつきも分析に取り入れることを検討したが、本研究のデータでは星座ごとの試行数が限られておりモデル推定が安定しない可能性がある。そのため、本実験ではペア差のみを考慮して条件の効果を分析した。なお、全試行で音声記録および各種ログに欠損はなかったため、実施した全試行を分析対象とした。

#### 4.2 タスクにおける理解役の星座認識について

タスクの達成度を測る客観的指標として、星座を構成する全星数のうち、理解役が正解の星を正しく選択できた割合 (以下、正答率) を算出した。本研究では、最終的な達成度ではなく、会話途中における到達状況も把握するため、タスク開始3分時点における正答率 (以下、3分時点正答率) も合わせて算出した。最終正答率及び3分時点正答率について、全8ペアによる64会話を対象に各条件が達成度に与える影響を調査した。分析結果を表2に示す。

実験の結果、3分時点正答率においてズレあり条件が大きく低下していた。条件別の平均値は、3分時点正答率ではズレなし条件が0.45、ズレあり条件が0.13であった。LMMによる分析の結果、ズレあり条件は3分時点正答率を有意に低下させることが示された。

また、最終正答率についても同様の傾向が確認された。条件別の平均値は、ズレなし条件で0.65、ズレあり条件で0.26であり、ズレなし条件が有意に高い結果となった。

さらに、星座ごとの難易度差を確認するため、条件を考慮せずに星座別の正答率を算出した (表3)。その結果、星座によって正答率にばらつきが見られた。例えば3分時点では「イカ」の平均正答率が0.48で最も高く、「ウマ」は0.18と低かった。最終時点においても同様に、「イカ」は0.64と高い一方、「ペンギン」は0.33と低かった。

#### 4.3 会話分析

表4に、開始3分間およびタスク全体における発話指標の比較結果を示す。以下では、LMMにより推定した条件効果 ( $\beta$ ) と95%信頼区間 (95% CI) に基づき記述する。

分析の結果、開始3分間では平均発話交代間隔がズレあり条件の方が増加傾向の効果が推定されたが、95% CIが0を含むため不確実性が残る。一方、平均沈黙時間はズレなし条件に対してズレあり条件の方が増加した ( $\beta = +0.13$ ,

表 3: 星座ごとの正答率の平均と標準偏差

星座	3分時点 平均 (SD)	最終 平均 (SD)
イカ	0.48 (0.29)	0.64 (0.33)
ワシ	0.38 (0.31)	0.59 (0.41)
イルカ	0.32 (0.30)	0.43 (0.27)
コウモリ	0.29 (0.33)	0.36 (0.43)
イチゴ	0.27 (0.38)	0.42 (0.50)
ペンギン	0.25 (0.31)	0.33 (0.35)
ザリガニ	0.19 (0.25)	0.47 (0.38)
ウマ	0.18 (0.25)	0.43 (0.43)

95% CI [0.04, 0.23])。ターン交代率及び発話率については条件間で差はみられなかった。言語的指標について、全フィルター率はズレなし条件に対してズレあり条件で増加傾向の効果が推定されたが、95% CIが0を含むため不確実性が残る。また、「え」系フィルター率については差が認められなかった。

次にタスク全体では、ズレあり条件において、平均発話交代間隔が増加した ( $\beta = +0.12$ , 95% CI [0.02, 0.22])。また、平均沈黙時間についてもズレあり条件の方が増加している ( $\beta = +0.12$ , 95% CI [0.05, 0.20])。ターン交代率はズレあり条件で低下傾向が示されたが、95% CIが0を含むため不確実性が残る。一方、発話率については差は認められなかった。言語的指標では、全フィルター率および「え」系フィルター率のいずれも差が認められなかった。

#### 4.4 タスクごとの主観評価の結果

本実験で行われた全8ペアによる64会話で、主観アンケートは各試行で役割の異なる2名分の回答を得たため計128件の回答を対象に、各条件による主観評価への影響を検討した。分析結果を表5に示す。なお、主観評価については条件間差の有無を簡潔に示すことを優先し、条件効果の検定結果 ( $p$  値) を中心に報告した。

その結果、「円滑さ」において、ズレあり条件でスコアが有意に低下した ( $p < .001$ )。特に認識の一致に関わる評価項目で影響が大きく、「3分認識一致感」と「最終認識一致感」ではどちらも大幅な低下が確認された ( $p < .001$ )。また、「起点星の使いやすさ」においてもズレあり条件で有意な低下が認められた ( $p = .001$ )。

それに対して、「楽しさ」は条件による有意差が見られなかった。また、「ストレス」についてもズレあり条件で増加する傾向は見られたものの、有意差は認められなかった ( $p = .167$ )。

## 5. 考察

### 5.1 仮説の検証

4.3節より、仮説H1については、開始3分間およびタス

表 4: 開始 3 分間およびタスク全体における発話指標の比較 (LMM: 固定効果の推定値と 95%信頼区間)

指標	開始 3 分間			タスク全体		
	ズレなし	ズレあり	$\beta$ (95% CI)	ズレなし	ズレあり	$\beta$ (95% CI)
構造的指標 (交代・間隔)						
ターン交代率 [回/分]	22.74	22.13	-0.61 [-2.36, 1.13]	23.41	21.60	-1.81 [-3.63, 0.00]
発話率 [回/分]	40.53	39.94	-0.59 [-2.83, 1.65]	40.19	38.50	-1.69 [-3.91, 0.53]
平均発話交代間隔 (重なり含む) [秒]	0.02	0.11	+0.09 [-0.02, 0.21]	0.00	0.12	+0.12 [0.02, 0.22]
平均沈黙時間 (重なり除外) [秒]	0.87	1.00	+0.13 [0.04, 0.23]	0.89	1.01	+0.12 [0.05, 0.20]
言語的指標 (フィルター出現率)						
全フィルター率 [回/分]	15.71	17.48	+1.77 [-0.20, 3.73]	8.26	9.19	+0.93 [-0.05, 1.90]
「え」系フィルター率 [回/分]	3.25	3.90	+0.64 [-0.28, 1.56]	2.32	2.64	+0.31 [-0.28, 0.91]

表 5: 条件ごとのアンケート主観評価項目

項目	ズレなし平均	ズレあり平均	p 値
円滑さ	3.92	2.77	< .001
3 分認識一致感	3.94	2.34	< .001
最終認識一致感	4.08	2.89	< .001
起点星の使いやすさ	4.06	2.80	.001
楽しさ	4.34	3.98	.128
ストレス	2.28	2.73	.167

ク全体のいずれにおいても全フィルター率に有意差は認められなかったものの、ズレあり条件で増加する傾向が見られた。フィルターは迷いの表出として直感的に解釈しやすい一方、個人の発話スタイルや対話の進行方略の影響を受けやすい。そのため、本実験ではズレの有無がフィルターの総量差として明瞭に現れにくかったことが考えられる。

次に、仮説 H2 について、開始 3 分間では平均沈黙時間が有意に増加し、平均発話交代間隔も増加傾向を示した。またタスク全体では、平均発話交代間隔および平均沈黙時間の両方が有意に増加し、仮説 H2 は支持された。これは、ズレが生じると互いに見ている星の確認が必要となり、沈黙や交代間隔が長くなる可能性が示唆された。特にタスク全体で差が明確になったことについて、参加者間で見ている星のズレを会話の進行に伴い認識したため、終盤にかけて修正が増えたことが考えられる。

## 5.2 主観評価と正答率の乖離がある会話について

本研究の最終的なゴールは参加者が互いに「認識が一致している」と誤認したまま進行する「無自覚なズレ」を、発話特徴から客観的に推定することである。自覚的なズレの検知は「え？」や「わからない」などを用いることで識別可能であると考えられるが、無自覚なズレに対しては検知が難しい。そこで本節では、ズレあり条件にも関わらず実験参加者のどちらもが 3 分時点と最終時点で星座が一致していると回答し、客観的な正答率が低かったペアのタスク (以下、タスク A) に着目する。表 6 は、開始 3 分間と

タスク全体における主要指標について、タスク A と各条件平均値を比較したものである。

開始 3 分間のタスク A と各条件平均を比較すると、タスク A では平均発話交代間隔が負の値となっており、発話の重なりを伴うテンポの速い対話であった。また、平均沈黙時間も条件平均より短かった。一方、「え」系フィルター率は条件平均より低く、迷いを明示する反応が少ないまま対話が進行していたことが示唆される。つまり、明示的なフィルターが少ない段階でも、発話交代間隔や沈黙時間といったテンポ指標の組み合わせから状態を捉える必要があると考えられる。

次に、会話全体のタスク A と各条件平均を比較すると、タスク A では平均発話交代間隔と平均沈黙時間はいずれも条件平均と同程度であり、開始 3 分間のような顕著な短縮は見られなかった。一方、「え」系フィルター率は条件平均より高く、進行に伴って不確かさが表出していた可能性がある。したがって、将来的な推定では、単一時点の値だけでなく、序盤から全体にかけて会話のテンポ指標とフィルター率の変化として捉えることが重要であると考えられる。

## 5.3 行動指標と発話分析の結びつき

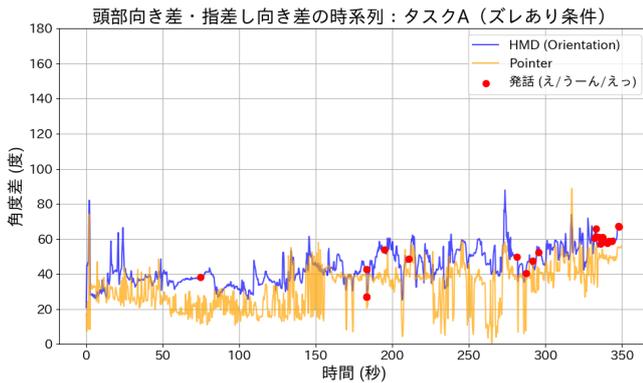
5.2 節で述べたように、発話指標のみでは検知が難しい「無自覚なズレ」を音声情報以外の情報を用いて補足推定することを検討するため、本研究では HMD の頭部向きとコントローラの指差し方向ログから、二者間の参照方向差を算出し、時系列で可視化した。Meta Quest 3 では視線計測が利用できないため、指差し方向を注視点の代理として扱った。図 4 の縦軸は参照方向差 (角度差)、横軸は経過時間であり、ズレあり条件における同一ペアの 2 試行、タスク A (5.2 節と同一) と、修正ができ正答率が高かったタスク (以下、タスク B) を示す。

タスク A では、考察 5.2 節で示したように会話テンポは安定していたにもかかわらず、参照方向差が長時間大きのまま推移しており、発話の円滑さが参照対象の一致を保証しないことが示された。一方タスク B では、「え？」などの修復関連発話の直後に参照方向差が縮小する傾向が見ら

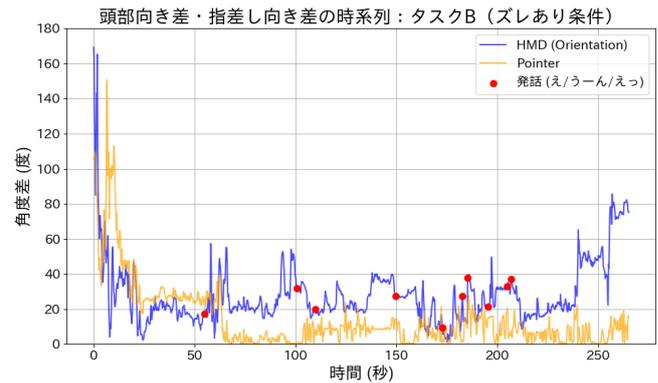
表 6: タスク A と条件平均の主要指標比較 (開始 3 分間 / タスク全体)

指標	開始 3 分間			タスク全体		
	ズレなし平均	ズレあり平均	タスク A	ズレなし平均	ズレあり平均	タスク A
ターン交代率 [回/分]	20.30	18.23	28.67	23.41	21.60	27.68
平均発話交代間隔 (重なり含む) [秒]	0.02	0.11	-0.11	0.00	0.12	0.10
平均沈黙時間 (重なり除外) [秒]	0.87	1.00	0.72	0.89	1.01	0.73
「え」系フィルター率 [回/分]	3.25	3.90	0.67	2.32	2.64	3.08

注: 発話交代間隔はオーバーラップを含むため負値を取り得る。



(a) タスク A (無自覚なズレ: 正答率 0/7)



(b) タスク B (修復を伴う収束: 正答率 6/7)

図 4: 参照方向差の時系列. 青: HMD 由来の二者間方向差, 橙: コントローラの参加者間方向差, 赤点: 修復関連発話 (「え」「うん」「えっ」) の出現時刻.

れ, 確認を契機に参照が調整され収束していく様子が観測された. 以上より, 参照方向差の時系列は, 無自覚なズレの継続や見ている星のズレの修正の生起を補足する手がかりとして機能できる可能性がある.

#### 5.4 起点ズレの推定に向けて

5.2 節で示したように, 会話が滑らかに進行していても, 見ている星の一致が保証されるとは限らず, 無自覚な起点星のズレが表面化しにくい場合がある. また主観評価では, 楽しさとストレスに有意差は見られなかった一方, 円滑さおよび認識一致感はズレあり条件で有意に低下していた. 以上より, 本研究が目指す早期検出は正しさを強制するためではなく, 必要な場面でのみ再確認を促し, 対話の円滑さと納得感を保つ支援につなげるために重要である.

そのため, 基点ズレの早期検出に向けた推定手法としては, 機械学習や LLM などのアプローチが考えられる. 機械学習では, 発話間隔やフィルター出現などの時系列特徴に加え, 5.3 節で述べたような参照方向差系列や角速度・加速度といった行動指標も統合し, 時間推移からズレの兆候を捉えることが考えられる. 一方 LLM では, 発話交代間隔や相槌の多さといった表層的な円滑さに依存せず, 参照対象・位置関係・形状記述の不整合や, 参加者による言い換えや再説明の有無を指標として組み込むことが重要であ

ると考えられる.

ここでは予備検討として, LLM を用いてタイムスタンプを付与した文字起こしからその会話が起点ズレ条件かどうかを推定できるか分析した. 推定には OpenAI 社の GPT-4o を用いた. 入力には, 各試行のタイムスタンプ付き文字起こし (話者・開始時間・終了時間・発話内容) を用いた. これらの情報を与え, LLM に起点ズレの有無を 0/1 で出力させた. プロンプトには, 判定基準として妥協的合意表現 (例: 「一旦合っていることにしよう」), 位置関係に関する不整合, および対話テンポの変化などを含めた.

分析対象は全 64 試行であり, 内訳はズレなし 32 試行, ズレあり 32 試行であった. 評価指標を表 7 に, 混同行列を表 8 に示す. 正解率は 0.70, F1 スコアは 0.75 であった. また, 再現率は 0.90 と高く, ズレあり試行を取りこぼしにくいことが示された一方, 適合率は 0.64 と低い結果となった. これは, 妥協的合意表現や位置関係表現の言い換え, 対話テンポの揺れがズレなし条件の探索過程にも起こるため, タイムスタンプを付与した文字起こしだけではズレあり条件の兆候として識別できなかったことが考えられる.

#### 5.5 限界と今後

本研究にはいくつかの限界がある. 特に, ペア数・タスク数が限定的であり, 星座刺激の差が結果に影響した可能

表 7: LLM による起点ズレ推定の評価指標

指標	数値
正解率	0.70
適合率	0.64
再現率	0.90
F1 スコア	0.75

表 8: 混同行列 (起点ズレ推定; 行: 真値, 列: 推定値)

真値	推定値	
	ズレあり	ズレなし
ズレあり	29	3
ズレなし	16	16

性を完全には排除できていない。また、文字起こしおよびタイムスタンプの精度には誤差が含まれ得るため、各条件の会話指標の平均値にも不確実性が残る。課題として、条件の統制を進めるとともに、文字起こしの品質検証やログの同期精度の向上を行う必要がある。

今後は、起点ズレ推定に向けて、まず発話交代間隔・沈黙時間・オーバーラップおよびフィルターなど、発話タイミング由来の特徴のみでどこまで推定可能かを検討し、有効な特徴集合を明らかにする。そのうえで、必要に応じてテキスト情報に加え、HMD 頭部向きや指差し方向ログから得られる参照方向差などの行動指標も統合し、誤判定の低減を目指した推定手法へ拡張する。

さらに、推定結果を用いた介入型支援へ拡張する場合には、対話を阻害しにくいタイミングと提示方法の設計が重要となる。例えば、「疑わしい状態が一定時間継続した場合」や「修復関連発話の直後」など、対話の区切れが生じやすい局面に限定して、起点星の再確認を促す短い提示を行うことが考えられる。一方、介入の頻度や強度が過剰になると、会話の自然な流れや探索の集中を損ねる可能性があるため、最小限の提示で修復を促す設計を行う必要がある。

## 6. まとめ

本研究では、星空観察における会話支援に向けた基礎調査として、星座説明の初手で参照される起点星が参加者間で一致しない場合 (ズレあり条件) が、星座探索タスクおよび会話の音声の特徴に与える影響を検討した。仮想環境内に星空観察環境を構築し、起点星が一致する条件と一致しない条件を設定したうえで、参加者による星座探索会話実験を実施した。

実験の結果、ズレあり条件ではズレなし条件に比べて正答率が低下し、主観評価においても円滑さや認識の一致に関する評価が低下した。一方、発話率やターン交代率など

会話のリズムや密度に関わる指標には大きな差がみられず、発話が円滑に進行しているように見えてもズレが潜在する局面が示唆された。また、予備検討として文字起こしとタイムスタンプを入力とした LLM による起点ズレ推定を行った結果、一定の識別可能性が確認できたが適合率が低いことが確認された。今後は、頭部向きや指差し方向などの行動指標を統合したマルチモーダル推定により偽陽性の低減と精度向上を図るとともに、推定結果に基づく介入タイミング・提示方法を含めた会話支援としての統合を検討する。

## 参考文献

- [1] Iida, S. and Nakamura, S.: Do You See What I See? Vocal Cues to Visual Acuity Discrepancies in VR-Based Stargazing, *International Conference on Collaboration Technologies and Social Computing*, Springer, pp. 123–131 (2025).
- [2] Soga, M., Ohama, M., Ehara, Y. and Miwa, M.: Real-world oriented mobile constellation learning environment using gaze pointing, *IEICE transactions on information and systems*, Vol. 94, No. 4, pp. 763–771 (2011).
- [3] Ohama, M. and Soga, M.: Collaborative Constellation Learning Environment with Sharing Learners' Gazing Points in the Real Night Sky, *2012 IEEE Fourth International Conference On Digital Game And Intelligent Toy Enhanced Learning*, IEEE, pp. 123–125 (2012).
- [4] Planey, J., Rajarathinam, R. J., Mercier, E. and Lindgren, R.: Gesture-mediated collaboration with augmented reality headsets in a problem-based astronomy task, *International Journal of Computer-Supported Collaborative Learning*, No. 2, pp. 259–289 (2023).
- [5] Tian, K., Urata, M., Endo, M., Mouri, K., Yasuda, T. and Kato, J.: Real-world oriented smartphone AR supported learning system based on planetarium contents for seasonal constellation observation, *Applied Sciences*, Vol. 9, No. 17, p. 3508 (2019).
- [6] 蓮雄一, 加藤直樹: 星空ビジョン: 仮想空間での星空を介したコミュニケーションを支援するツールの開発, 研究報告コンピュータと教育 (CE), Vol. 2021, No. 18, pp. 1–6 (2021).
- [7] Saeki, M., Miyagi, K., Fujie, S., Suzuki, S., Ogawa, T., Kobayashi, T. and Matsuyama, Y.: Confusion Detection for Adaptive Conversational Strategies of An Oral Proficiency Assessment Interview Agent., *INTERSPEECH*, pp. 3988–3992 (2022).
- [8] 豊田薫, 宮越喜浩, 山西良典, 加藤昇平: 発話状態時間長に着目した対話雰囲気推定, 人工知能学会論文誌, Vol. 27, No. 2, pp. 16–21 (2012).
- [9] Hayashi, M.: Marking a 'noticing of departure' in talk: Eh-prefaced turns in Japanese conversation, *Journal of Pragmatics*, Vol. 41, No. 10, pp. 2100–2129 (2009).
- [10] Micklos, A. and Woensdregt, M.: Cognitive and interactive mechanisms for mutual understanding in conversation, *Oxford research encyclopedia of communication* (2023).