Do You See What I See? Vocal Cues to Visual Acuity Discrepancies in VR-Based Stargazing

Sora Iida[0009-0009-7117-8870]1 and Satoshi Nakamura[0000-0003-3492-7093]1

Meiji University, 4-21-1 Nakano, Nakano-ku, Tokyo, Japan
sorabun.iida@gmail.com

Abstract. Stargazing often involves conversation about celestial objects, but perceptual differences such as visual acuity can cause misalignments in what participants see, making communication difficult. As a preliminary investigation, this study examined how visual acuity differences influence conversational behavior during collaborative stargazing. In a VR-based constellation search task, we compared pairs with matched and unmatched acuity. Although results were not statistically significant, consistent trends emerged: more clarification requests, higher question frequency, and longer response latency under acuity differences. These findings suggest that perceptual asymmetry may affect mutual understanding and point to the potential of conversation-based support systems.

Keywords: Stargazing · Conversation · Visual acuity difference · VR

1 Introduction

In collaborative tasks, perceptual differences in what participants see can create communication challenges. To our knowledge, this study is among the first to investigate, through vocal cues, how discrepancies in visual acuity affect conversation during the specific activity of stargazing. While stargazing is a shared activity, communication can be difficult because stars are distant and hard to distinguish. Such perceptual misalignments can disrupt conversation and shared recognition.

Existing tools attempt to reduce these misalignments. Mobile applications such as Star Walk2 [16] and Sky Guide [5] align perceptions by displaying the same screen. In addition, prior research has developed stargazing support systems [10, 12, 11]. Among them, AR-based systems overlay constellation information or support recognition tasks [14, 17]. Furthermore, physical means such as laser pointers are also used to point at stars. However, both digital and physical aids are often impractical in natural stargazing: screen brightness impairs dark adaptation, and laser pointers are unsuitable for outdoor use.

To address this gap, we adopt a conversation analysis perspective. Research shows that clarification questions, repair initiators such as "Eh?" [7] or "Huh?" [4], and turn-taking patterns [3, 13, 15] are key to maintaining mutual understanding. Yet, prior work has largely focused on linguistic or cognitive asymmetries, with little attention to the *perceptual asymmetries* arising from vision differences. Although vision research has used VR/AR to study the impact of impairments such as low vision or color deficiencies [2, 6, 1], these investigations emphasized individual performance rather than collaborative dialogue [8, 9].

Against this background, we investigate how visual acuity differences influence conversation during collaborative stargazing. Using a VR-based constellation search task, we analyze conversational features such as clarification requests, occurrences of "Eh?", and response latency. Our study serves as an initial step toward conversational support systems that can detect and mediate perceptual asymmetries in real time. We are guided by the following research question:

How do visual acuity differences between partners influence conversational behavior during a shared stargazing task?

The contributions of this study are summarized as follows:

1. We introduced a novel perspective by focusing on conversation content during collaborative stargazing. This area had been overlooked in previous research, which primarily addressed visual aids or learning outcomes.
2. We proposed and implemented a VR-based simulation environment in which visual acuity could be systematically varied between two participants while preserving a shared visual field. This enabled controlled experiments on perceptual asymmetry in naturalistic dialogue.
3. We conducted an empirical study with manipulated visual acuity differences and suggested that such differences tended to bring about measurable changes in conversational behavior. Specifically, we observed tendencies toward increased clarification requests such as "Eh?", higher question frequency, and longer response latency.
4. We provided initial evidence that perceptual asymmetry affects mutual understanding in visually grounded dialogue. This finding lays the groundwork for future conversational support systems, for example, voice assistants that intervene when misalignments are detected.

2 Experiment

To examine how visual acuity differences influence conversational behavior, we designed a VR-based constellation search task. Participants collaborated under conditions where their visual acuity was either matched (Equal) or different (Unequal). We analyzed conversational indicators such as clarification questions, "Eh?" utterances, and response latency.

We tested the following hypotheses:

H1. Clarification questions will occur more frequently in the Unequal condition than in the Equal condition.
H2. "Eh?" utterances will occur more frequently in the Unequal condition than in the Equal condition.
H3. Response latency between speaker turns will be longer in the Unequal condition than in the Equal condition.
H4. Subjective ratings of conversational smoothness will be lower in the Unequal condition than in the Equal condition.

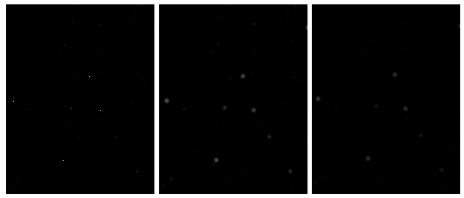

Fig. 1. Simulated views of the star field under each visual acuity condition, from left to right: Good, Normal, Poor.

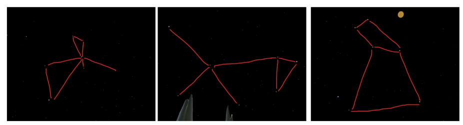

Fig. 2. Examples of synthetic constellations used in the experiment, from left to right: Cat, Butterfly, Moai.

System. The system was implemented in Unity and presented using Meta Quest 3. A star field of several hundred stars, created for this study, was arranged on a hemispherical dome, with varied sizes to approximate natural appearance. Three visual acuity levels (Good, Normal, Poor) were simulated using different levels of blur (Fig. 1). To provide common reference points, simple landmarks such as mountains and planets were included. Avatars were omitted, so participants could not directly see each other's gaze or pointing gestures.

Task and Design. Each pair was assigned fixed roles: a *Describer*, who explained the target constellation verbally, and an *Identifier*, who attempted to locate it. Four star field patterns containing 14 synthetic constellations were used (Fig. 2). To support explanations, the Describer was given supplementary information about constellation shapes. In the experiment, we compared two conditions: the Equal condition, in which both participants had the same level of visual acuity, and the Unequal condition, in which their visual acuity levels differed. Combinations of Good and Poor visual acuity levels were excluded.

Participants and Procedure. Before the experiment, participants completed a questionnaire and received instructions on roles and procedures. In each trial, the Describer studied the target constellation and explained it, while the Identifier attempted to locate it. Each trial lasted up to 10 minutes, followed by a brief questionnaire on conversational smoothness and recognition alignment. This sequence was repeated across conditions. Eight participants (5 male and 3 female), all members of the same laboratory, took part in the experiment. They were undergraduate and graduate students aged 20 to 23. All had a visual acuity of 0.7 or higher (with or without corrective lenses) and no astigmatism that affected daily activities. Each participant was paired with others two or three times and participated in three or four conversation trials per pairing. To prevent overfamiliarization with all three visual acuity conditions within a single star field, only two acuity levels were assigned per star pattern.

3 Results

3.1 Analysis

To analyze question frequency, we followed previous research on utterance classification and divided both Describer and Identifier utterances into individual Idea Units (IUs). One IU was defined as a single simple sentence. In cases involving compound or multiple sentences, we counted them as two IUs. We then classified all IUs into the following nine categories:

– **Explanation**: Describes what the speaker sees or explains the shape and position of a constellation.
– **Confirmation**: Checks whether the listener shares the same perception.
– **Question**: Seeks clarification or repetition.
– **Response**: Provides answers to questions, confirmations, or requests.
– **Request**: Affects conversation flow, such as "Please wait" or "Can you repeat that?".
– **Backchannel**: Listener feedback like nods or utterances such as "uh" or "Eh?".
– **Incomplete**: Utterances that stop mid-sentence without continuation.
– **Gaze instruction**: Specifies directions for finding a constellation.
– **Other**: Unrelated talk or self-directed speech.

Response latency was measured as the interval between speaker turns, using the *pyannote.audio* speaker diarization model and excluding within-speaker pauses.

In addition, we focused on the Japanese token "Eh?", which functions as a repair initiator indicating recognition misalignment. Following prior work, we treated the frequency of "Eh?" utterances as an objective measure of recognition misalignment.

3.2 Subjective Evaluations of Conversation

Table 1 summarizes participants' questionnaire ratings. Overall, conversational smoothness and comprehension were consistently rated higher in the Equal condition than in the Unequal condition. This suggests that matched visual acuity facilitated smoother interaction and better mutual understanding. Participants also reported less conversational fatigue when visual acuity was matched.

Regarding strategies, Describers tended to begin with the constellation's name and shape, often referring to bright landmarks such as Mercury or Mars. Identifiers reported repeatedly checking information and attempting to visualize the constellation shape. These findings suggest that participants actively adapted their explanation and listening strategies depending on their partner's condition.

3.3 Objective Measures of Recognition Misalignment

Time to Constellation Recognition Alignment. We analyzed thirty conversations, equally divided between the Equal and Unequal conditions. For each conversation, we measured the time required for both participants to reach recognition alignment of the constellation. As shown in Fig. 3, the mean alignment time was 345 seconds in the Equal condition and 465 seconds in the Unequal condition. A paired-samples t-test revealed no significant difference, but there was a marginal trend toward longer alignment times

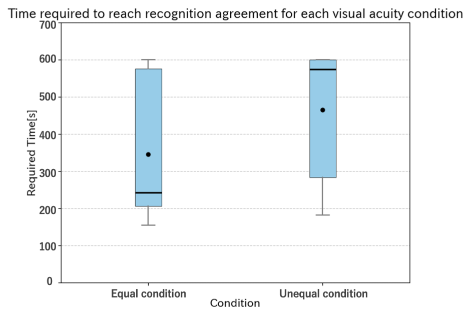

Fig. 3. Time to recognition alignment under Equal and Unequal conditions.

Table 1. Evaluation Results by Vision Level (Good, Normal, Poor) under visual acuity conditions

Item	Good		Normal		Poor	
	Unequal	Equal	Unequal	Equal	Unequal	Equal
Smoothness of conversation	2.75	2.50	2.63	3.00	2.33	2.90
Difficulty understanding partner's utterances	2.63	3.00	2.00	1.20	1.83	1.20
Agreement on constellation recognition	0.75	0.70	0.69	1.00	0.67	1.00
Conversation fatigue	2.13	2.50	2.31	2.60	3.17	2.22

in the Unequal condition ($t(14) = -1.94$, $p = 0.073$). Furthermore, the number of trials that exceeded the 600-second time limit and were thus considered unsuccessful was higher in the Unequal condition (7 trials) compared to the Equal condition (4 trials). When trials exceeding the 600-second limit were excluded, the mean times were 252 seconds (Equal) and 347 seconds (Unequal). These results suggest that transmitting constellation information tended to take longer when visual acuity differed.

Response Latency. We also examined turn-taking and response latency. The Equal condition yielded an average response latency of 0.87 seconds, whereas the Unequal condition yielded 0.98 seconds. Although not statistically significant, participants tended to respond more quickly when their visual acuity was matched.

Number of "Eh?" Utterances. We counted "Eh?" utterances in 30 trials (15 Equal, 15 Unequal), excluding self-directed tokens. A total of 40 occurred in the Equal condition and 61 in the Unequal condition. Although a paired-samples t-test showed no significant difference ($t(14) = -0.94$, $p = 0.362$), the tendency indicates that clarification requests were more frequent when visual acuity differed.

Comprehension Test Results. After each conversation, the Identifier completed a brief test by identifying stars and drawing the constellation shape (max = 2 points). The Equal condition yielded an average score of 1.60, compared to 1.33 in the Unequal condition. Although the difference was not statistically significant ($t(14) = 0.89$, $p = 0.389$), the trend suggests that matched visual acuity supported better comprehension (as shown in Fig. 4 for an example).

Utterance Protocol. We coded all conversational data into utterance protocol, with the full breakdown shown in Table 2. The proportion of question-category utterances

Table 2. Proportion of utterance protocol by Visual Acuity Condition

	Unequal (%)	Equal (%)
Explanation	25.1	27.7
Backchannel	21.0	26.2
Response	17.1	15.1
Question	10.8	9.2
Confirmation	11.1	10.1
Gaze instruction	1.2	1.7
Request	0.4	0.2
Incomplete utterance	0.2	0.3

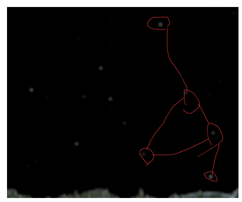

Fig. 4. An example of participant annotation from the comprehension test.

was 9.2% in the Equal condition and 10.8% in the Unequal condition. This difference was not statistically significant ($t(14) = 0.99, p = .337$), but the pattern suggests that participants in the Unequal condition initiated slightly more clarification questions.

4 Discussion

Summary of Findings. This study investigated how visual acuity differences affect conversational behavior during collaborative stargazing. Although most differences did not reach statistical significance, consistent tendencies were observed. Unequal visual acuity pairs required longer time to align constellation recognition, produced more clarification questions and “Eh?” utterances, and reported lower smoothness and comprehension in post-trial questionnaires. These findings suggest that perceptual asymmetry can complicate mutual understanding, even in a shared visual environment.

Interpretation of Key Metrics. Among the indicators we examined, question-category utterances and “Eh?” utterances provide useful signals of recognition misalignment. Even though the differences were not significant, their higher frequency in the Unequal condition suggests that perceptual asymmetry increased the need for repair. These conversational cues may serve as real-time markers for detecting when partners are looking at different stars.

Furthermore, subjective ratings revealed a gap in perceived alignment. The difference between Describer and Identifier responses to the question “Did your recognition match?” was larger in the Unequal condition. This suggests that Describers may have overestimated how well their explanations were understood, a form of overconfidence that can hinder timely clarification and exacerbate recognition mismatches. This highlights that not only visual access but also a mutual awareness of each other's understanding plays a critical role in collaboration.

Interestingly, another finding emerged regarding visual complexity. Participants with Good visual acuity in the Equal condition reported more difficulty understanding their partner compared to those in the Normal or Poor conditions. One possible explanation is that an abundance of visible stars introduced ambiguity when selecting or referring to specific constellations. This suggests that higher visual clarity does not always benefit communication, especially when the visual scene becomes overly complex, and implies that support systems must manage visual density, not just enhance clarity. In addition, constellation complexity may interact with perceptual asymmetry, underscoring the need to balance visual scene design in future experiments.

Implications for Support Systems. The observed tendencies suggest that conversational cues such as question-category utterances and "Eh?" can be leveraged to estimate recognition misalignment in real time. Rather than relying on intrusive visual aids, a support system could monitor these cues during stargazing conversations and infer when partners are likely to be looking at different stars. Such a system could then provide minimal interventions, for example by suggesting relevant landmarks or prompting clarification, without disrupting the natural flow of dialogue. This approach highlights the potential of conversation-based monitoring for designing lightweight and context-sensitive support tools in collaborative astronomy tasks.

Limitations. This study has several limitations. The small, homogeneous sample of participants from the same laboratory limits the generalizability of our findings. Furthermore, the analysis of conversational features, such as "Eh?" utterances and question-category utterances, by a single coder raises concerns about reliability. The presence of prominent landmarks in the experimental environment may also have influenced conversation strategies and distorted the nature of the task. Future work must address these issues by recruiting a more diverse participant pool, ensuring robust annotation through multiple coders, and refining the experimental environment to better simulate real-world conditions. By overcoming these challenges, we aim to develop a real-time system that detects conversational cues to mediate recognition misalignments.

5 Conclusion

This study investigated how visual acuity differences influence conversational behavior during collaborative stargazing in VR. Across multiple measures, pairs with unequal acuity tended to show conversational friction: they appeared to require more time to achieve recognition alignment, produced more clarification questions, and reported lower smoothness and comprehension. These tendencies suggest perceptual asymmetry is reflected in measurable vocal cues, the central finding of this work.

Future work should validate these findings with larger and more diverse participant samples, improve annotation reliability, and control for factors such as constellation difficulty, which may amplify the effects of acuity differences. Ultimately, these findings may inform the design of lightweight, context-sensitive conversational agents that leverage vocal cues to mediate perceptual asymmetries without disrupting the natural flow of collaborative stargazing.

References

1. Aoki, T., Fujiwara, Y., Nakamura, S.: Validation of game advantage disadvantage control considering color vision characteristics: A basic study on "Among Us" with different color settings. Procedia Computer Science **225**, 2982–2991 (2023)
2. Chen, C.F., Huang, K.C.: Effects of background lighting color and movement distance on reaching times among participants with low vision, myopia, and normal vision. Perceptual and Motor Skills **122**(2), 518–532 (2016)
3. Clark, H.H., Schaefer, E.F.: Contributing to discourse. Cognitive Science **13**(2), 259–294 (1989)
4. Dingemanse, M., Torreira, F., Enfield, N.J.: Is "huh?" a universal word? conversational infrastructure and the convergent evolution of linguistic items. PLOS ONE **8**(11), e78273 (2013)
5. Fifth Star Labs: Sky Guide. https://www.fifthstarlabs.com/, accessed: 31 May 2024
6. Fujiwara, Y., Nakamura, S.: Fundamental study of color combinations by using deuteranope-simulation filter for controlling the handicap of color vision diversity in video games. In: Entertainment Computing–ICEC 2021. pp. 127–138. Springer (2021)
7. Hayashi, M.: Marking a 'noticing of departure' in talk: Eh-prefaced turns in Japanese conversation. Journal of Pragmatics **41**(10), 2100–2129 (2009)
8. Jones, P.R., Somoskeöy, T., Chow-Wing-Bom, H., Crabb, D.P.: Seeing other perspectives: evaluating the use of virtual and augmented reality to simulate visual impairments (OpenVisSim). NPJ digital medicine **3**(1), 32 (2020)
9. Neugebauer, A., Castner, N., Severitt, B., Stingl, K., Ivanov, I., Wahl, S.: Simulating vision impairment in virtual reality: a comparison of visual task performance with real and simulated tunnel vision. Virtual Reality **28**(2), 97 (2024)
10. Ohama, M., Soga, M.: Collaborative constellation learning environment with sharing learners' gazing points in the real night sky. In: 2012 IEEE Fourth International Conference On Digital Game And Intelligent Toy Enhanced Learning. pp. 123–125. IEEE (2012)
11. Soga, M., Matsui, K., Takaseki, K., Tokoi, K.: Interactive learning environment for astronomy with finger pointing and augmented reality. In: 2008 Eighth IEEE International Conference on Advanced Learning Technologies. pp. 542–543. IEEE (2008)
12. Soga, M., Ohama, M., Ehara, Y., Miwa, M.: Real-world oriented mobile constellation learning environment using gaze pointing. IEICE transactions on information and systems **94**(4), 763–771 (2011)
13. Stolcke, A., Ries, K., Coccaro, N., Shriberg, E., Bates, R., Jurafsky, D., Taylor, P., Martin, R., Ess-Dykema, C.V., Meteer, M.: Dialogue act modeling for automatic tagging and recognition of conversational speech. Computational linguistics **26**(3), 339–373 (2000)
14. Tian, K., Urata, M., Endo, M., Mouri, K., Yasuda, T., Kato, J.: Real-world oriented smartphone AR supported learning system based on planetarium contents for seasonal constellation observation. Applied Sciences **9**(17), 3508 (2019)
15. Varonis, E.M., Gass, S.: Non-native/non-native conversations: A model for negotiation of meaning. Applied linguistics **6**(1), 71–90 (1985)
16. Vito Technology: Star Walk 2. https://starwalk.space/ja, accessed: 31 May 2024
17. Zhang, J., Sung, Y.T., Hou, H.T., Chang, K.E.: The development and evaluation of an augmented reality-based armillary sphere for astronomical observation instruction. Computers & Education **73**, 178–188 (2014)