

## Introduction

Because people often see, hear, and interact with the world in similar ways, we can map symbols, such as words, onto **shared meanings** [1]. This makes communication intuitive. But what happens when perception differs entirely? In real-world settings, both humans and artificial agents may rely on different sensory inputs. For example, one robot might use vision while another uses sound. Similarly, people with sensory impairments may interpret the world differently. This research explores a significant question: *how can communication emerge without a shared way of perceiving the world?*

## Methodology

To explore how agents communicate without shared perception, we use a **multi-turn referential game** (Figure 1) based on existing emergent communication research [2], where agents create their own language through interaction, rather than being pre-programmed [3]. One agent (*sender*) hears an object and sends a binary message. The other agent (*receiver*), seeing multiple objects, is tasked with guessing the object referred to by the sender. We test this in two setups: **multimodal**, where one agent sees and the other hears, and **unimodal**, where both share the same input.

## Key Findings

Our results reveal key differences in how agents communicate based on perceptual alignment. Unimodal agents, with shared input, performed better using **shorter, more structured messages**, showing **lower entropy** (less uncertainty) and **greater within-class message similarity**. In contrast, multimodal agents needed **longer, denser, and more abstract messages**, struggling to compensate for their lack of shared perception. Interestingly, bits that frequently flipped ( $0 \leftrightarrow 1$ ) carried more meaning in unimodal systems. Finally, both systems **grounded their messages in low-frequency features** (Figure 2), suggesting that even abstract communication retains ties to perceptual input

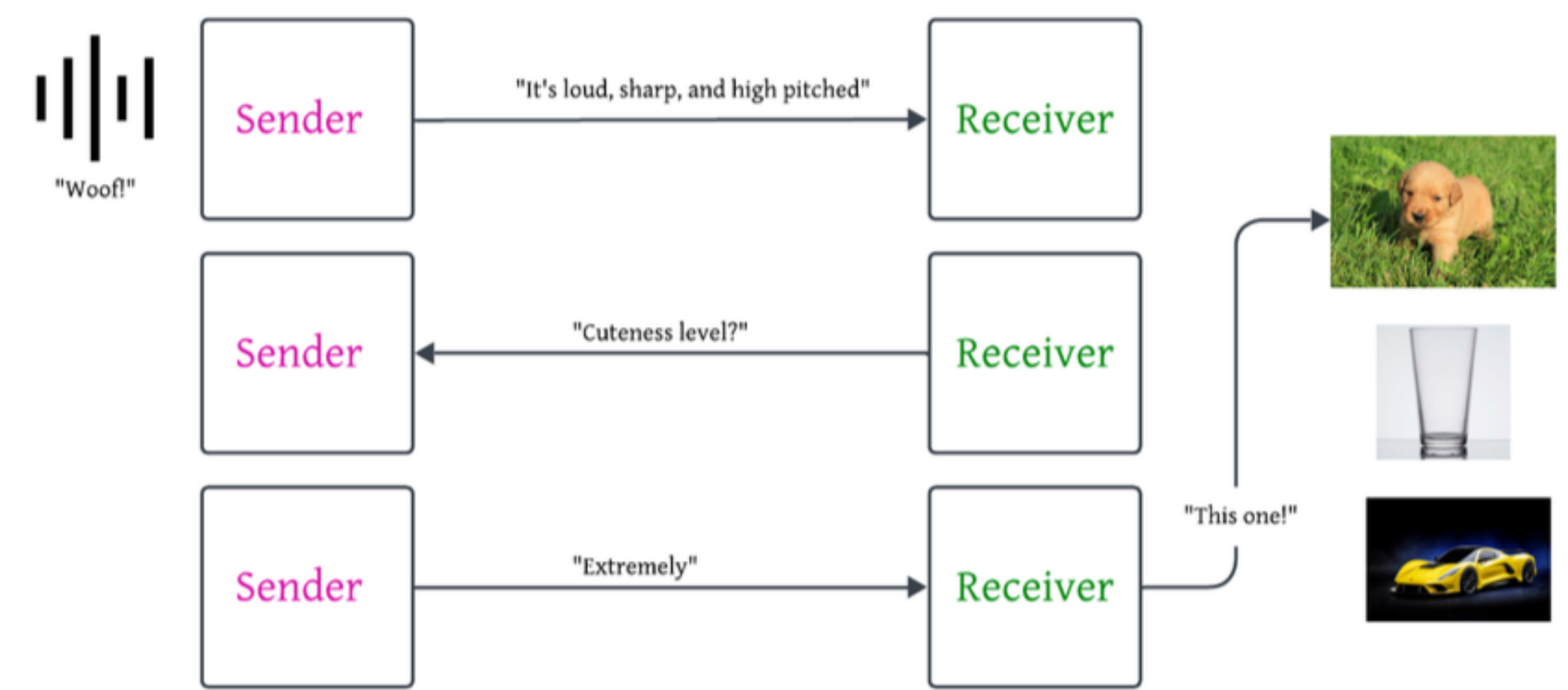


Figure 1 – Multimodal Multistep Referential Game

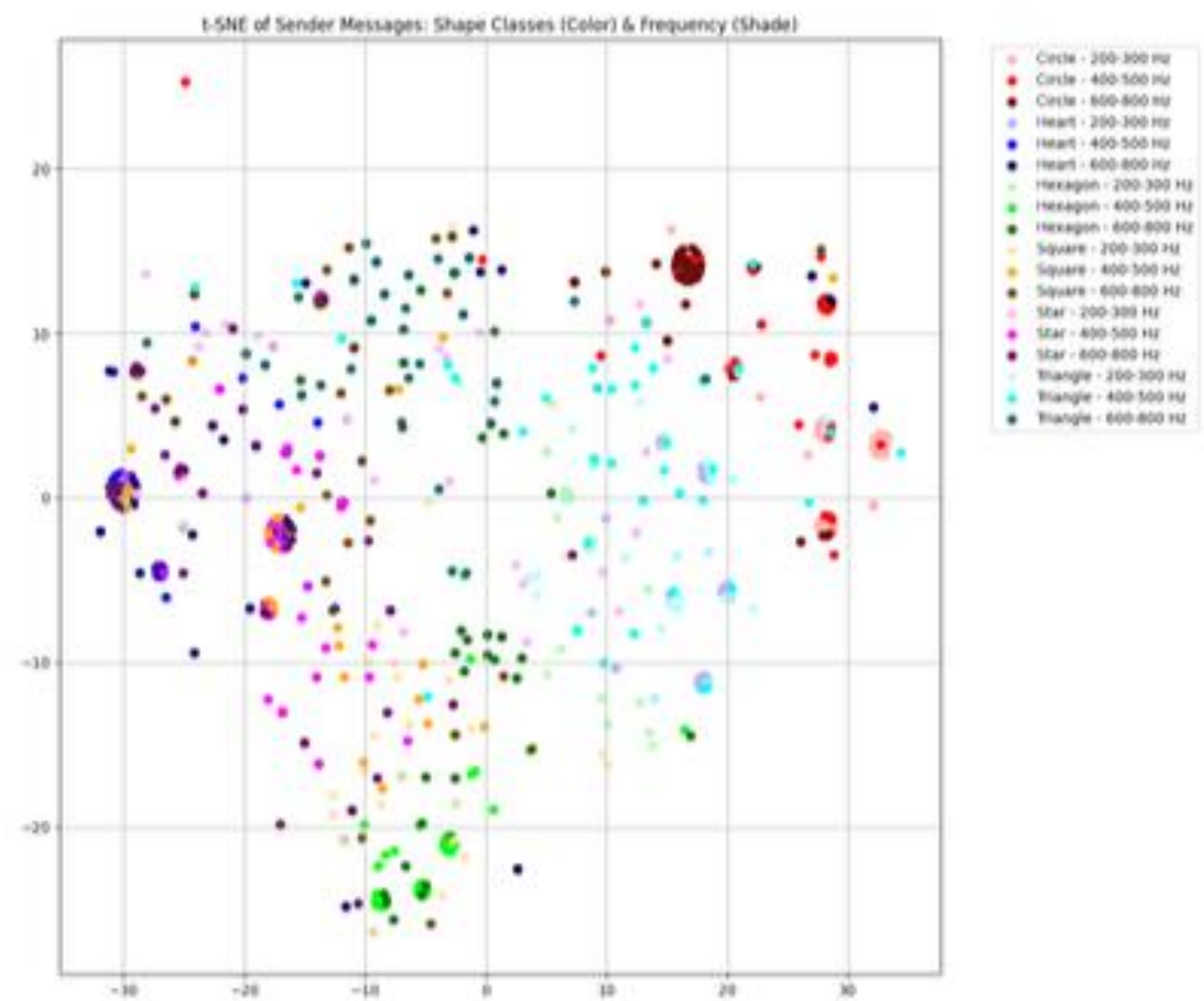


Figure 2 – tSNE embeddings of sender messages for different sound input frequencies

## Future Applications

This research lays a foundation for real-world systems where agents must communicate despite differing perceptual experiences. It supports the development of **heterogeneous robotic teams** for settings like search-and-rescue, and informs the design of **assistive technologies** that help individuals with sensory or cognitive impairments communicate more effectively. Importantly, it offers a computational framework for studying how **meaning emerges across divergent modalities**.

## References

1. Stevan Harnad. 1990. The symbol grounding problem. *Physica D: Nonlinear Phenomena* 42, 1 (1990), 335–346. doi:10.1016/0167-2789(90)90087-6
2. Katrina Evtimova, Andrew Drozdov, Douwe Kiela, and Kyunghyun Cho. 2018. Emergent Communication in a Multi-Modal, Multi-Step Referential Game. arXiv:1705.10369 [cs.LG]
3. Brendon Boldt and David Mortensen. 2024. A Review of the Applications of Deep Learning-Based Emergent Communication. arXiv:2407.03302 [cs.CL]

