

Modeling Disrupted Conversation Dynamics: A Real-Time VR and Active Inference Approach

Cumhur Erkut *Aalborg University, Copenhagen, Denmark* cer@create.aau.dk

Abstract

Disrupted conversational dynamics are a hallmark of both auditory verbal hallucinations and hearing impairment, where the brain struggles to balance prior expectations against unreliable sensory evidence. This project addresses the scientific challenge of modeling these disruptions by combining Active Inference (AIF) with real-time simulation in immersive virtual reality. We implement a discrete, message-passing formulation of AIF, where likelihood precision, policy confidence, and transition dynamics can be manipulated to generate either in-context or out-of-context disruptions. Our scientific AI approach thus bridges computational psychiatry, speech technology, and VR engineering. The potential impact lies in advancing mechanistic understanding of hearing loss or psychosis, creating novel therapeutic VR interventions, and fostering collaborations across AI, clinical neuroscience, and human-computer interaction.

Keywords: Neuroscience, AI in Science, Active Inference, Virtual Reality.

1 Introduction

Hearing impairment (HI) and Auditory verbal hallucinations (AVH) both disrupt conversational flow, albeit through different mechanisms. While AVH introduces internally generated, intrusive content (Benrimoh et al. [2018, 2019]), HI distorts or limits access to external speech cues (Friston et al. [2021]). Despite their differences, both challenge **active listening**—the brain’s inferential process of tracking and interpreting communicative signals under uncertainty.

This study introduces a novel system that integrates real-time voice conversion, embodied avatars, and Active Inference modeling within immersive virtual reality (VR). Active Inference, a theory on how the brain predicts and updates beliefs from sensory input [Parr et al., 2022], is applied both to HI and AVH to simulate conversational dynamics under active listening in real time.

After outlining our scientific motivation, Sec. 2 introduces our light-weight, explainable, and gradient-free AI approach that can be personalized across patient conditions and their conversation interaction history. We conclude with the potential for impact and collaboration possibilities facilitated by our approach.

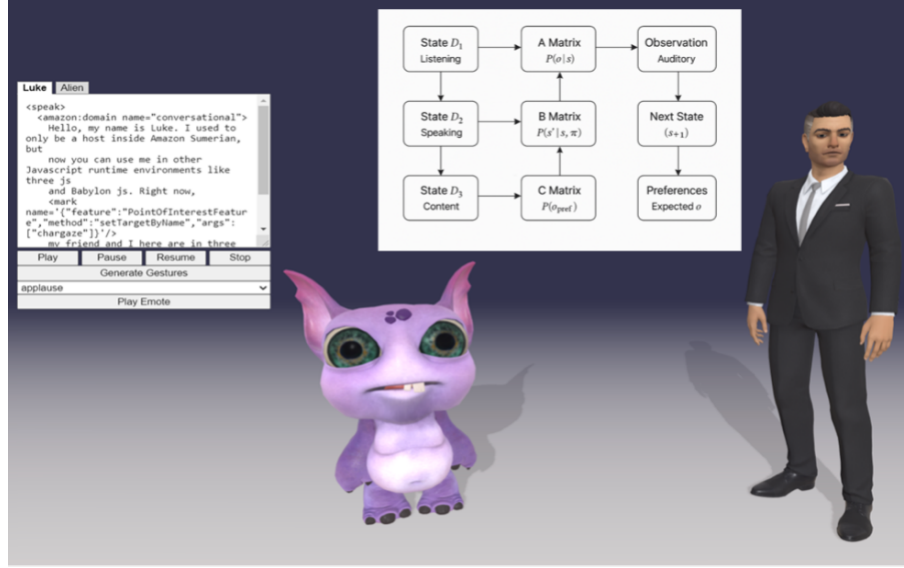


Figure 1: The embodied avatars of actual (right) versus hallucinated (left) voices. The Active Inference Model passes messages to the system to switch the avatars, their voices, emotions, and gestures.

2 Methods

Through real-time scientific simulations and surrogate modeling, we aim to enable therapeutic / training applications through avatar-based interactions in three powerful, complementary ways:

1. Externalizing Internal Beliefs, by augmenting the generative model with a controlled, simulated voice. For AVH, this makes the hallucinations and coping mechanisms as policies interactive and testable. For HI, this illustrates the balance between sensory weighting and prior policies.
2. Manipulating Policy Beliefs via Dialog: shifting maladaptive policy priors (e.g., “I must obey the voice”), which allows for therapeutic restructuring of agency, and
3. Real-time audio-visual feedback, so that the avatar becomes part of the generative process, enabling precision shaping and graded exposure.

We aim for various voice-driven avatar models in popular game engines, such as Unity 3D or Unreal Engine. Currently we prototype our development using AWS family of voice synthesis models, driven by Amazon hosts, whose switching is governed by the policy inference mechanism framework with three hidden state factors: listening/not-listening, speaking/not-speaking, and content (see Fig. 1).

Table 1: Active Inference matrices used in the MDP formulation.

Matrix	Represents
Likelihood A	$P(o_t s_t)$: probability of an observation given a hidden state
Transition B	$P(s_{t+1} s_t, \pi)$: probability of transitioning between hidden states under a policy
Prior over Observations C	$P(o_t)$: preferred observations (soft goals)
Prior over States D	$P(s_1)$: prior belief about initial hidden states

Participants interact with avatars producing both self-generated and externally imposed voices that simulate impaired or hallucinated content. By adjusting sensory precision (reliability of input) and policy priors (behavioral expectations), the system explores miscommunication phenomena. It uses probabilistic inference to update beliefs about conversational roles (e.g., speaker/listener), intentions, and content plausibility, by adjusting sensory precision (the reliability of sensory input) and policy priors (expectations about conversational behavior).

In general, an Active Inference workflow starts with determining the type of the inference. In our case, we opt for discrete Markov Decision Process. The next step is to define the model components, as enlisted in Table 1.

Our simulation model currently includes the following matrix components:

- **A**: Encodes the probability of auditory or proprioceptive *observations* given hidden *states*. For instance, hearing "Alien" in Fig. 1 when in the listening and content state. Modulated by sensory precision (ζ).
- **B**: Encodes how states evolve across time (e.g., content progressing normally or abnormally through an intrusive phrase pronounced by "Alien"). Modulated by transition uncertainty.
- **C**: Desired observations (e.g., avoiding hearing hallucinated voices of "Alien"). Influences policy selection via expected free energy.
- **D**: Initial belief over hidden states, e.g., expecting to begin in a quiet or verbal state. It specifies the agent's belief about what state it's in at discrete time step $t = 1$. Often uniform or context-specific.

These matrices are adjusted to simulate low sensory precision ζ , high policy precision γ , and variable transition dynamics ω . Adjusting sensory weighting and priors accounts for HI simulation [Friston et al., 2021], whereas incorporating dynamic belief updates and anomalous **B** matrices enables in-context vs. out-of-context hallucinations [Benrimoh et al., 2019]. The content layer's abnormal transitions (e.g., state 3 jumping to 1) mimic internal voice intrusions such as "Alien wants me to cut your arm".

In inference, we minimize the Variational Free Energy

$$F = \mathbb{E}_{Q(s)}[\ln Q(s) - \ln P(o, s | \pi)], \quad (1)$$

where beliefs $Q(s_\tau)$ approximate the true posterior $P(s_\tau \mid o_{1:\tau}, \pi)$. We drive policy selection by balancing epistemic value (information gain) and pragmatic value (goal-directed outcomes) in Expected Free Energy

$$G(\pi) = \mathbb{E}_{Q(o|\pi)}[D_{\text{KL}}(Q(s|o, \pi) \| Q(s|\pi))] - \mathbb{E}_{Q(o|\pi)}[\ln P(o)], \quad (2)$$

where D_{KL} denotes the Kullback–Leibler divergence between the policy distributions given the observations o , states s , and the policies π .

3 Conclusions

Our study demonstrates how Active Inference matrices (A, B, C, D) can simulate both auditory hallucinations and hearing impairments by altering precision and transition dynamics. By embedding a high-quality real-time AIF model within avatar-based VR, we externalize internal beliefs, restructure maladaptive conversational policies, and provide controlled sensory feedback. This offers a principled computational framework for studying miscommunication phenomena and testing therapeutic protocols. The approach creates fertile ground for collaboration across AI, computational psychiatry, VR engineering, and clinical neuroscience, advancing both mechanistic understanding and translational applications.

We are currently experimenting with a secondary hierarchical control layer (Fig. 1), allowing policy switching based on inferred volatility. This can be extended for real-time voice modulation and externalized control over auditory precision.

References

- D. Benrimoh, T. Parr, R. A. Adams, and K. Friston. Hallucinations both in and out of context: an active inference account. *PLOS ONE*, 14(8):e0212379, 2019. doi: 10.1371/journal.pone.0212379.
- D. A. Benrimoh, T. Parr, P. Vincent, R. A. Adams, and K. Friston. Active inference and auditory hallucinations. *Computational Psychiatry*, 2(0):183, 2018. doi: 10.1162/cpsy_a_00022.
- K. J. Friston, N. Sajid, D. R. Quiroga-Martinez, T. Parr, C. J. Price, and E. Holmes. Active listening. *Hearing Research*, 399:107998, 2021. doi: 10.1016/j.heares.2020.107998.
- T. Parr, G. Pezzulo, and K. J. Friston. *Active Inference*. The MIT Press, 2022. doi: 10.7551/mitpress/12441.001.0001.