

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Hearing Speech or Doing Inference? Diagnosing Speech Perception in Speech Large Language Models

Permalink

<https://escholarship.org/uc/item/4q75p7cc>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Author

Chen, Tianfang

Publication Date

2026

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Hearing Speech or Doing Inference?

Diagnosing Speech Perception in Speech Large Language Models

Tianfang Chen (tianfang.chen@ling-phil.ox.ac.uk)¹

¹Faculty of Linguistics, Philology and Phonetics, University of Oxford

Abstract

Speech LLMs with native audio input can produce fluent text from raw acoustic signals, but it remains unclear whether this reflects speech perception as such. I use the sine-wave speech paradigm to probe responses across free description, forced-choice discrimination with silent-audio controls, and open-ended transcription without alternatives. A qualitative split emerges across models in behavior. One class shows little sensitivity to the sine-wave speech, failing to use the signal in free description and transcription while maintaining high forced-choice accuracy even under silence. A second class shows limited stimulus sensitivity, with instruction-dependent improvements that disappear when acoustic input is removed, resembling a weak analogue of human perceptual reorganization in part. I suggest that these differences track access to speech-relevant acoustic information at the front-end rather than downstream language modeling.