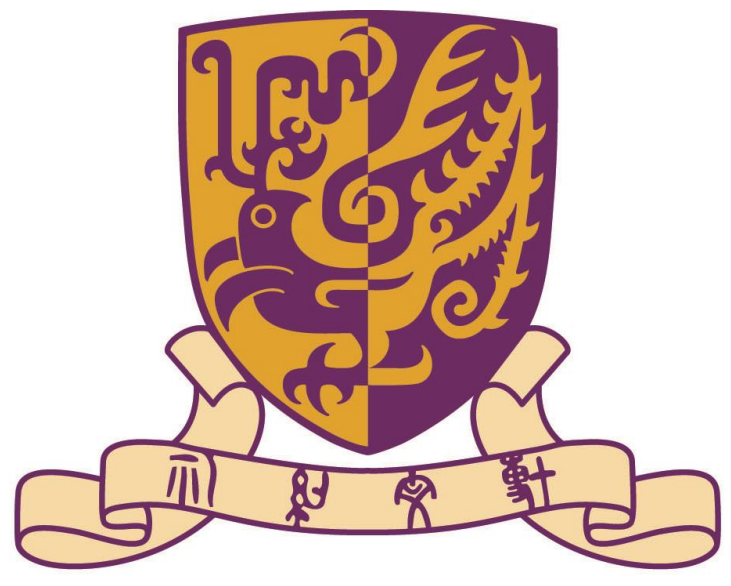


The impact of initial phonetic distance on L2 speech accommodation in Human- and AI-directed speech



Manson Chun Man Wong ^a, Xinyi Chen ^a, Yitian Hong ^a, Yusheng Tian ^a,
Grace Wenling Cao ^b, Peggy Pik Ki Mok ^a

^a The Chinese University of Hong Kong; ^b University College Dublin

mansonwong@link.cuhk.edu.hk, xinyichen001@cuhk.edu.hk, yitianhong@cuhk.edu.hk, ystian@cuhk.edu.hk, grace.cao@ucd.ie, peggymok@cuhk.edu.hk



Introduction

➤ Phonetic Accommodation

- = Speakers' tendency to modify their speech patterns during **interaction**
- 3 types of accommodation: 1. Convergence, 2. Divergence; 3. Maintenance

➤ The type and degree of phonetic accommodation depends on...

- 1. **phonetic difference** between speaker and model talker
 - E.g., larger → more likely to converge [1]
- 2. **sociolinguistic** factors
 - E.g., gender, race, perceived attractiveness [2]

➤ Theories on Phonetic Accommodation

- Interactive Alignment Model (IAM) [3]
 - An automatic process caused by priming solely on **linguistic level**
- Communication Accommodation Theory (CAT) [4]
 - A **socially motivated** behaviour to adjust the social distance

➤ However, previous studies mostly...

- 1. examine phonetic distance from a **cross-variety perspective** [5]
 - this study analyzes phonetic distance **within** a language variety
 - focus on **individual variation** within a linguistic community
 - ✓ L2 varieties, such as Hong Kong English (HKE), provides an **ideal** testing ground due to their high individual variability [6]
- 2. focus on **human-directed speech** (human-human **interaction**) [7]
 - this study examines **AI-directed speech** (AI-human **interaction**)
 - assess whether accommodation is **socially motivated**
 - ✓ AI voices that **closely resemble human speech** offer a **unique** chance to examine this theoretical debate, as they differ primarily in their sociolinguistic characteristics rather than their acoustic properties

➤ Research Questions

- 1. Is the type of accommodation **mediated** by the **initial phonetic distance** between the L2 speaker and the model talker of L1 English?
- 2. Does accommodation differ between interactions with a human and an **acoustically matched AI interlocutor**?

Method

➤ Participants: 30 HKE speakers (12M, 18F; age: 21.60±2.62)

- Age of English Acquisition: 1-6 (2.83±1.15)
- IELTS or equivalent: 6-8 (7.29±0.59)

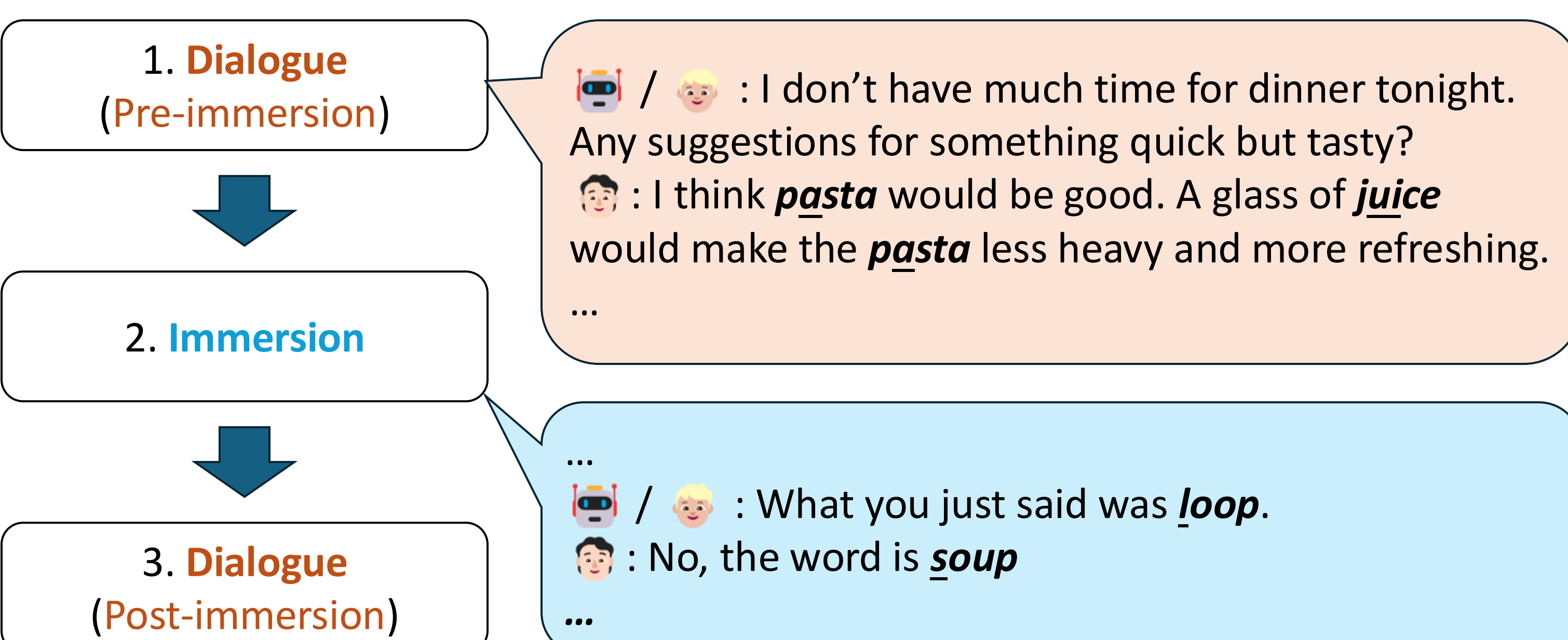
➤ Model talker (MT): A male L1 speaker of Standard Southern British English

➤ AI voice of MT was generated using CosyVoice 2

- Prosodic characteristics are preserved
- No significant difference in F0, F1, F2, F3 values
- Difference in voice quality attributable to the speech synthesis process

➤ Experimental Design

- Between-subjects design: AI Condition OR Human Condition
- Dialogue**: 3 words are selected to elicit the corner vowels /i, a, u/ respectively
- Immersion**: An error-repair conversation with the same speaker



➤ Data Processing

- Midpoint F1 & F2 of the dialogues' vowels were extracted and normalized
- Acoustic distance between each participant's corner vowel and the corresponding vowel produced by the model talker was calculated using **Euclidean distance** in the F1-F2 space

Results

1. The slope (solid line) was **positive**

- = a degree of **consistency** in speakers' production before & after immersion

2. The slope was also **different from the maintenance line** (dashed line) of 1

- = evidence that **initial phonetic distance** affects the **type** of accommodation

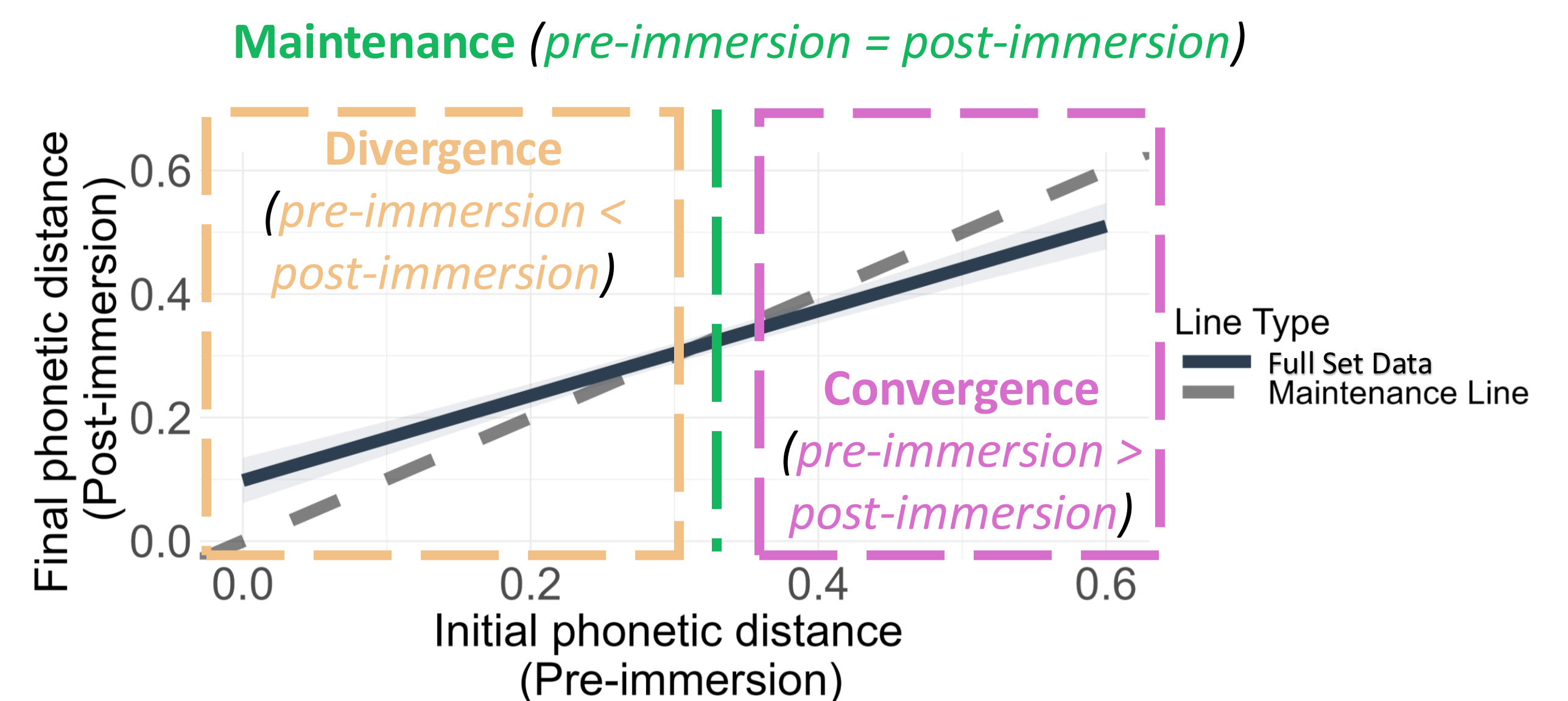


Fig. 1. Effect of the initial phonetic distance on the final phonetic distance.

3. For the same initial phonetic distance, **AI-directed speech** yielded a smaller final phonetic distance, showing a **greater degree in convergence** and a **reduced degree in divergence**

4. **Interlocutor type** affects the **threshold** (= $y\text{-intercept}/(1\text{-slope})$)

- A **lower convergence threshold** in **AI-directed speech** (AI: 0.25; Human: 0.38)
 - Easier to elicit phonetic convergence in **AI-directed speech**

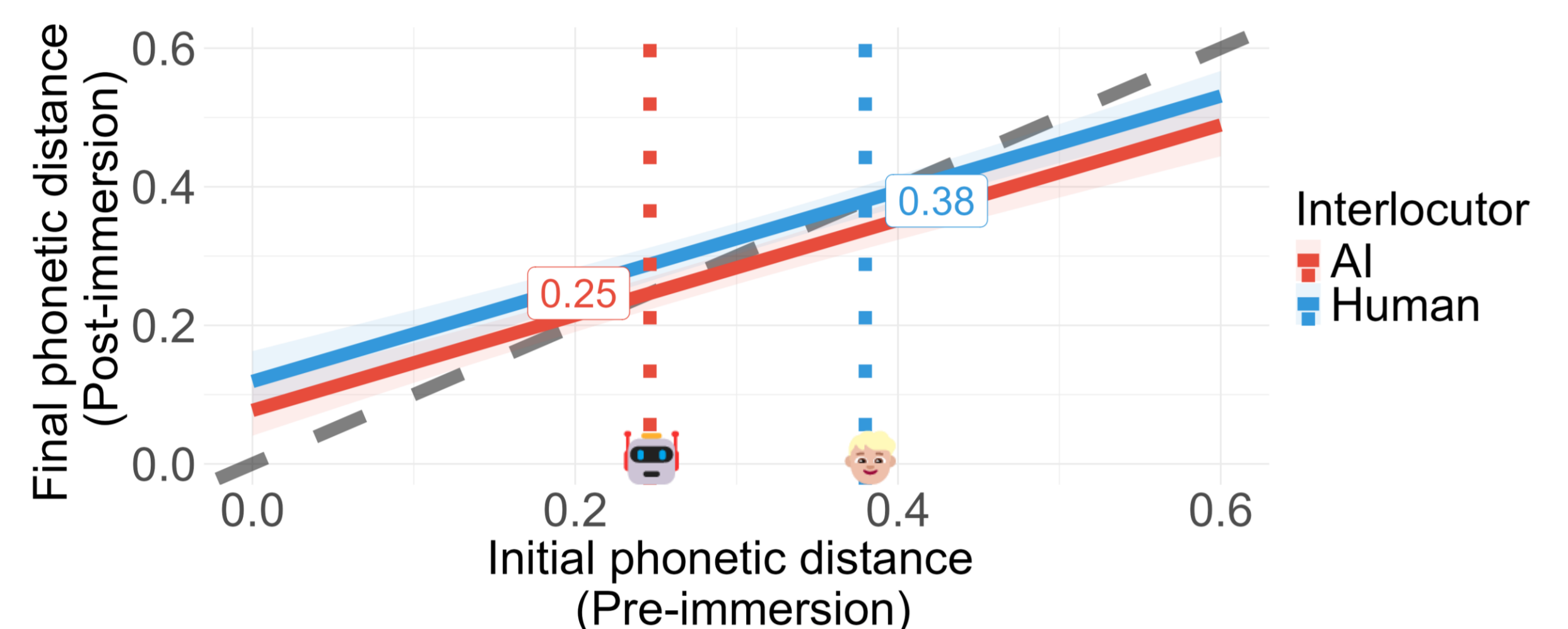


Fig. 2. Comparison of phonetic accommodation towards AI and human interlocutors.

Discussion

➤ Type of phonetic accommodation is **mediated** by the **initial phonetic distance**

- Supporting the view that phonetic difference effect operates **within varieties** as well, rather than being solely accounted for by a cross-variety perspective
- Suggesting convergence may only be **initiated** when the initial distance between interlocutors is **sufficiently large**; otherwise, being **too close** initially may lead to speakers **attempting to differentiate** themselves from one another
 - Whether this behaviour is conscious **remains to be explored**

➤ Phonetic accommodation is affected by **the type of interlocutor**, providing evidence for CAT [3], a socially driven account of phonetic accommodation

- Highlighting the need to **incorporate additional social factors** in accounting for accommodation patterns, particularly in **the era of AI**

➤ However, it remains **unclear** why human speakers are **more likely to converge** in **AI-directed speech** than in human-directed speech

- One **possible** explanation is that AI interlocutors **reduce socially evaluative pressures**, thus increasing speakers' willingness to **reduce social distance**

➤ Ongoing work will further investigate how **awareness of AI** shaped by **visual cues**, **perceived differences**, and **social factors** affect phonetic accommodation

Acknowledgements: The project is supported by the RGC Hong Kong PhD Fellowship Scheme (Reference No.: PF25-16581), the CUHK Vice-Chancellor HKPFS Scholarship Scheme, the Hong Kong RGC Postdoctoral Fellowship, the Hong Kong RGC GRF (Project No. 14612023), and the CUHK Research Committee Postdoctoral Fellowship Scheme 2023-24. **Key References:** [1] A. Walker and K. Campbell-Kibler, "Repeat what after whom? Exploring variable selectivity in a cross-dialectal shadowing task," *Front. Psychol.*, vol. 6, May 2015. [2] M. Babel, "Evidence for phonetic and social selectivity in spontaneous phonetic imitation," *Journal of Phonetics*, vol. 40, no. 1, pp. 177–189, Jan. 2012. [3] H. Giles, A. L. Edwards, and J. B. Walther, "Communication accommodation theory: Past accomplishments, current trends, and future prospects," *Language Sciences*, vol. 99, p. 101571, Sep. 2023. [4] M. J. Pickering and S. Garrod, "The interactive-alignment model: Developments and refinements," *Behavioral and Brain Sciences*, vol. 27, no. 2, pp. 212–225, Apr. 2004. [5] G. W. Cao, "Phonetic Dissimilarity and L2 Category Formation in L2 Accommodation," *Lang Speech*, vol. 67, no. 2, pp. 301–345, Jun. 2024. [6] J. C. Wakefield, "English Speakers in Hong Kong," *International Journal of TESOL Studies*, vol. 3, no. 3, 2021. [7] X. Chen *et al.*, "Modulation of Phonetic Realizations in Cantonese Dialogue with Human and AI Interlocutors," presented at the Proc. Interspeech 2026, Sydney, Australia, Sep. 2026.