An exploratory study of phonetic accommodation of Cantonese sound change in human-human and human-AI interactions

Grace Wenling Cao[1], Ann Wai Huen To[2], Karen Ka Ki So[2] and Peggy Mok[2]

[1]University College Dublin [2] The Chinese University of Hong Kong

Introduction

Sound change in the new AI era

Exemplar-based theories：(Johnson, 2006; Pierrehumbert, 2003)

- long-term sound change can be seen as the result of accumulation of short-term accommodation

New source of speech input：

- AI voice and speech also become part of speech input that can trigger changes

Research questions

1. Do speakers accommodate differently in **human-human vs. human-AI interactions**?
2. How does it affect the ongoing Cantonese sound change in Hong Kong?

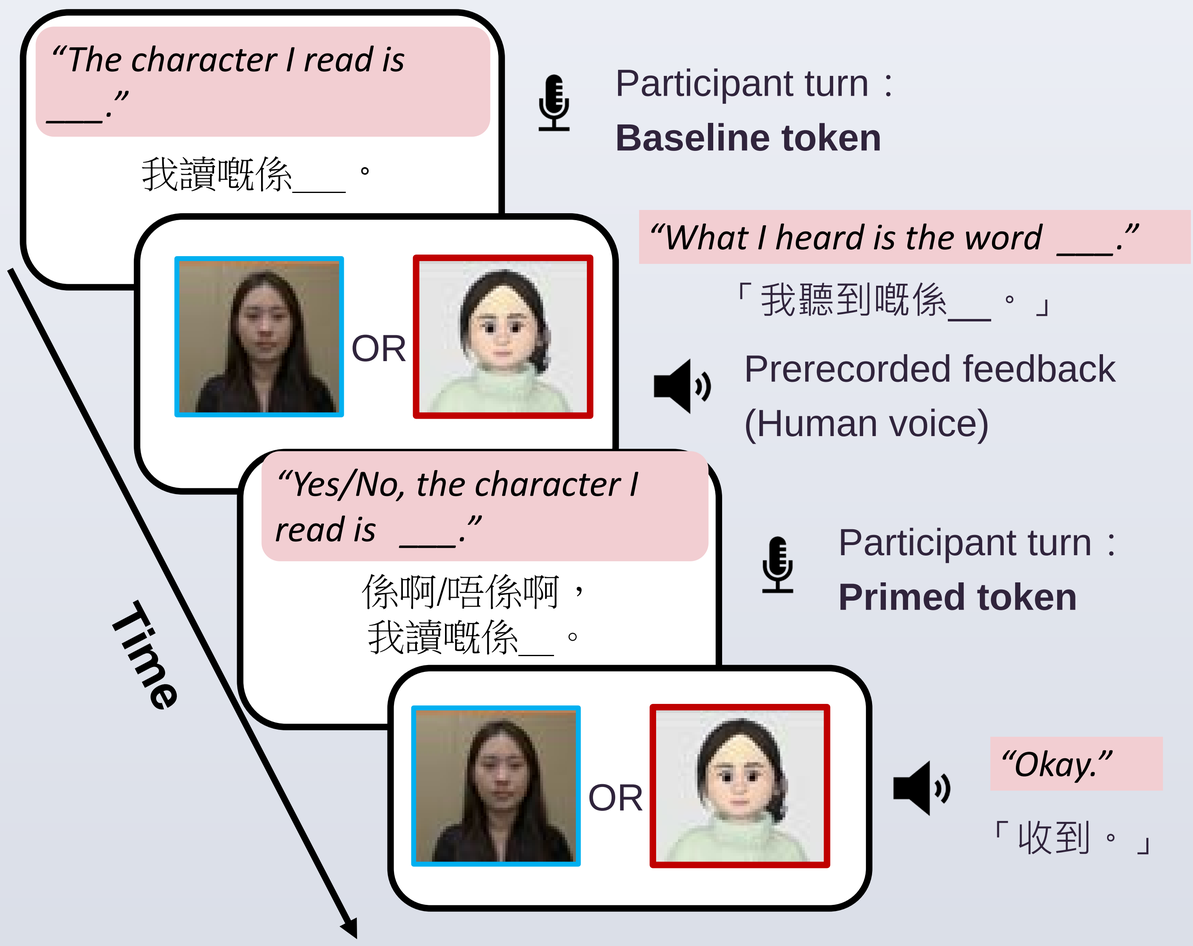
Study Design

Participants & experiment design

- 20 young native speakers of Hong Kong Cantonese
- Pilot 1: 📷 Image; Pilot 2: 📹 Video

Error-repair paradigm

(Cohn & Zellou, 2021; Zellou et al. 2021)



Selected features

Feature	Conservative form	New variant
Onset [n]-[l] merging	[neɪ] 你 'you'	[leɪ] 'you'
Onset [ŋ] deletion	[ŋɔ:] 我 'I'	[ɔ:] 'I'
Syllabic [ŋ̩]-[m̩] merging	[ŋ̩] 午 'noon'	[m̩] 'noon'
[ŋ]-[n] merging	[kɔŋ] 講 'to say'	[kɔn] 'to say'
Palatalisation of [s]	[sy:] 樹 'tree'	[ʃy:] 'tree'
[kʷ]-[k] merging	[kʷɔk] 國 'country'	[kɔk] 'country'
Pinjam	[tsɐu lɔŋ**21**] 走廊 'corridor'	[tsɐu lɔŋ**25**] 'corridor'
High-falling of T1	[sa:m**55**] 三 'three'	[sa:m**51**] 'three'

- Each feature was represented by a set of monosyllabic characters and a set of disyllabic words, except for the tonal features.

Results

Accommodation patterns

When speakers' primed token differs from its baseline:

- **Convergence**: Primed token = Model token
- **Divergence**: Primed token ≠ Model token

Image condition： Human > Avatar

	Interlocutor	Converge (%)	Diverge (%)
Monosyllabic	Human	**2.93**	2.34
	Avatar	0.39	1.17
Disyllabic	Human	**1.72**	1.56
	Avatar	0.63	0.00

Video condition： Avatar > Human

	Interlocutor	Converge (%)	Diverge (%)
Monosyllabic	Human	3.13	0.39
	Avatar	**11.72**	7.81
Disyllabic	Human	3.13	3.13
	Avatar	**6.56**	3.44

Overall: 📹 Video > 📷 Image

- We did observe accommodation patterns across conditions, but the overall effect is weak.

Accommodation hierarchy

High

- Syllabic [ŋ̩]-[m̩] merging
- Coda [ŋ]-[n] merging
- Onset [n]-[l] merging
- Onset [ŋ] deletion

Low

- **Nasal features** are more likely to change across conditions.
- Tonal features have no change.

Discussion & Conclusion

Videos are conducive to more phonetic accommodation

- Participants tend to accommodate more towards a lively talker compared to static images.

A hierarchy of accommodation pattern was found, which might be related to the status of Cantonese sound change variants (To et al., 2015)

- An ongoing variant: Coda [ŋ]-[n] merging
- An almost-complete variant: Onset [n]-[l] merging
- Sound change completed: Onset [ŋ] deletion

A weak accommodation effect in general

- A new paradigm was used in our subsequent studies.

Acknowledgements

This project is partially supported by the Postgraduate Studentship by the Chinese University of Hong Kong awarded to the third author, and RGC GRF (Project No. 14612023) to the last author.

Key reference

Cohn, M., & Zellou, G. (2021). Prosodic differences in human- and Alexa-directed speech, but similar local intelligibility adjustments. *Frontiers in Communication, 6*, 675704.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, 34(4), 485–499.

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, 46(2–3), 115–154.

To, Carol & Mcleod, Sharynne & Cheung, Pamela. (2015). Phonetic variations and sound changes in Hong Kong Cantonese: Diachronic review, synchronic study and implications for speech sound assessment. Clinical linguistics & phonetics. 29. 1-21. 10.3109/02699206.2014.1003329.

Zellou, G., Cohn, M., & Kline, T. (2021). The influence of conversational role on phonetic alignment toward voice-AI and human interlocutors. Language, Cognition and Neuroscience, 36(10), 1298–1312. https://doi.org/10.1080/23273798.2021.1931372