Speech adaptation in interactive scenarios: Talking with adults, babies, and robots

Angelica Lim[1], Paige Tuttösí[1], Henny Yeung[2],Yue Wang[2]
[1]School of Computing Science, [2]Department of Linguistics
Simon Fraser University, Burnaby, BC, Canada

Information theoretic features in speech typically measure the amount of information being transmitted across different units. For example, syllable information density (ID), syllable information rate (IR), and the amount of syllable reduction all help to quantify the amount of information being transmitted over time. Across languages, there is evidence of a constant information rate (~39 bits/s), which manifests as trade-offs between language-specific information density and speech rate (Coupé 2019). Recently, information-theoretic approaches have suggested that this information rate may also vary by *producer*: L1 speakers have higher information rates, compared to a non-native L2 speakers (Bradlow, 2022). We expand on that framework to further explore how information rates may also vary by *audience*, and the interplay between the *producer* and the *audience* to achieve a common communication goal.

Producers use distinct speech styles or registers, depending on audience, environments, or communicative goals. For example, we adopt a clear speech style when addressing non-native or hearing-impaired adults, when speaking to infants or children, or when chatting with Siri. In this position paper, we address three key challenges for an information theoretic account in understanding how producers adapt to audience needs, with a focus on commonalities between clear speech, infant-directed speech (IDS), and speech directed at AI systems.

One key finding in these three literatures is that adaptive speech emerges from the **dynamics** of conversational interactions with each audience type. Speakers adjust their speech over the course of a conversation to converge to their listeners' communicative needs (Pardo et al., 2017; Wagner et al., 2021). For example, both the pitch (Smith & Trainor, 2008; Nencheva et al., 2021) and spectral properties (Lam & Kitamura, 2012) of IDS are well-tuned to infant behaviour through the course of a conversational interaction, which, in turn, echoes research on speech directed to AI systems (Thomason et al., 2013; Zellou et al., 2021). A first challenge to integrate into information theoretical approaches is how to quantify efficient communication of information amid time-sensitive conversational dynamics in these kinds of situations.

A second key finding is that adaptive speech is directly related to the producer's **intent** along a set of diverse and varied dimensions. For example, using clear speech (Smiljanic, 2021), IDS (Wang et al., 2022), and speech to AI systems (Zellou et al., 2021) have all been linked to improving intelligibility, yet each manifests in a different style. A second challenge for information theoretic accounts is to consider which types of adaptations could effectively enhance intelligibility and communicative efficiency across different audience types.

A third key finding is to understand how to quantify information beyond measures of intelligibility, as producers have other **para-linguistic aims** in these scenarios such as: directing attention (Fernald, 1989; Spinelli et al. 2017), learnability (Eaves et al., 2016), and conveying non-verbal emotions (Benders, 2013) and social stances (Schachner & Hannon, 2011). A final challenge for information theoretic accounts is to develop measures of communicative aims that extend beyond intelligibility, into a broader concept of information transfer.

Text-to-speech systems may benefit from addressing these challenges. For example, studies on robot voices find acceptability can depend on gender, naturalness or accent coupled with robot appearance or task (Torre, 2020), speech style in an ambient context (Hughson, 2022), vocal empathy for healthcare (James, 2020), or volume proportional to the distance to the interlocutor (Fischer, 2021). An information theoretic framework will be essential to conceptualize how AI speech systems should adapt *simultaneously* to non-linguistic contexts *and* interlocutors, as well as the interplay between those adaptations.

Benders, T. (2013). Mommy is only happy! Dutch mothers' realisation of speech sounds in infant-directed speech expresses emotion, not didactic intent. *Infant Behavior and Development*, *36*(4), 847–862.
Bradlow, Ann R. (2022). Information encoding and transmission profiles of first-language (L1) and second-language (L2) speech. *Bilingualism: Language and Cognition* 25.1: 148-162.
Coupé, C., Oh, Y. M., Dediu, D., & Pellegrino, F. (2019). Different languages, similar encoding efficiency: Comparable information rates across the human communicative niche. *Science Advances*, 5(9)
Eaves, B. S., Feldman, N. H., Griffiths, T. L., & Shafto, P. (2016). Infant-directed speech is consistent with teaching. *Psychological Review*, *123*(6), 758–771.
Fernald, Anne.. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message?. *Child development*: 1497-1510.
Fischer, K., Naik, L., Langedijk, R. M., Baumann, T., Jelínek, M., & Palinko, O. (2021). Initiating human-robot interactions using incremental speech adaptation. *Companion of HRI* (pp. 421-425).
Tuttosi, P., Hughson, E., Matsufuji, A., Zhang, C. & Lim, A. (2023). Read the Room: Adapting a Robot's Voice to Ambient and Social Contexts. *IEEE IROS (to appear).*
James, J., Balamurali, B. T., Watson, C. I., & MacDonald, B. (2021). Empathetic speech synthesis and testing for healthcare robots. *International J. of Social Robotics*, 13, 2119-2137.
Lam, C., & Kitamura, C. (2012). Mommy, speak clearly: Induced hearing loss shapes vowel hyperarticulation. *Developmental Science*, *15*(2), 212–221.
Nencheva, M. L., Piazza, E. A., & Lew‐Williams, C. (2021). The moment‐to‐moment pitch dynamics of child‐directed speech shape toddlers' attention and learning. *Developmental Science*, 24(1)
Pardo, J., Urmanche, A., Wilman, S., & Wiener, J. (2017). Phonetic convergence across multiple measures and model talkers. *Attention, Perception and Psychophysics* 79, 637-659.
Schachner, A., & Hannon, E. E. (2011). Infant-directed speech drives social preferences in 5-month-old infants. *Developmental Psychology*, *47*(1), 19–25.
Smiljanić, R. (2021). Clear speech perception: Linguistic and cognitive benefits. In J. S. Pardo, et al. (Eds), *The Handbook of Speech Perception* (pp.177-205). John Wiley & Sons, Inc.
Smith, N. A., & Trainor, L. J. (2008). Infant-directed speech is modulated by infant feedback. *Infancy*, *13*(4), 410–420.
Spinelli, M., Fasolo, M., & Mesman, J. (2017). Does prosody make the difference? A meta-analysis on relations between prosodic aspects of infant-directed speech and infant outcomes. *Developmental Review*, *44*, 1–18.
Thomason, J., Nguyen, H. V., & Litman, D. (2013). Prosodic entrainment and tutoring dialogue success. *International Conference on Artificial Intelligence in Education*, 750–753.
Torre, I., Latupeirissa, A. B., & McGinn, C. (2020). How context shapes the appropriateness of a robot's voice. *IEEE RO-MAN* (pp. 215-222).
Wagner, M.A., Broersma, M., McQueen, J.M., Dhaene, S., & Lemhöfer, K. (2021). Phonetic convergence to non-native speech: Acoustic and perceptual evidence. *J of Phonetics*, 88, 1-20.
Wang, L., Kager, R., and Wong, P.C.M. (2022). The effect of tone hyperarticulation in Cantonese infant-directed speech on toddlers' word recognition in the second year of life. *First Language* 1-23.

Zellou, G., Cohn, M. and Kline, T. (2021). The influence of conversational role on phonetic alignment toward voice-AI and human interlocutors. *Lan, Cogn. and Neuro.*, 36, 1298-1312.