Speech rate of short vs. long interrogative sentences in human-directed vs. device-directed dialectal Arabic speech

Yahya Aldholmi, Arwa Alsahafi, Reema Alshiddi, May Al-Sager
King Saud University, Kingdom of Saudi Arabia

https://doi.org/10.36505/TheLinguisticProceedings/2025/16/01/002/000662

Abstract

This paper examines the speech pattern adjustments made by female dialectal Arabic speakers when addressing an AI voice assistant compared to a close interlocutor and a stranger across different age groups. Each participant was presented with a set of ten interrogative sentences, controlling for syllabic length (short vs. long). They directed each set to the three addressees, after which the speech rate was measured. The results showed that the overall speech rate averaged 5.41 syllables per second (SpS), with adults speaking slightly faster than teenagers. Speech was fastest with a familiar partner (SpS = 5.54), followed by an unfamiliar one (SpS = 5.48), and slowest with the AI (SpS = 5.21). Linguistic complexity, namely, utterance length, matters, with shorter utterances being articulated more rapidly than longer counterparts. These findings call for future research into additional acoustic features and gender-related differences.

Keywords: device-directed speech; human-directed speech; dialectal Arabic; speech rate; age differences

Introduction

Individuals across languages and dialectal varieties adapt their speaking styles on the basis of several social and contextual factors (e.g., Cohn et al., 2021). Speakers make distinct acoustic–phonetic adjustments for speech directed to machines vs. to humans. Various articulatory features, such as speech rate, pitch, intensity, and duration, have been examined to differentiate between device-directed speech (DDS) and human-directed speech (HDS) (Cohn and Zellou 2021; Cohn et al. 2021; Cohn et al. 2022; Song et al. 2022; Christenson et al. 2023; Cohn et al. 2024a, 2024b).

Studies have amply documented that both speech rate and prosody shape human–AI interactions. Speech rate, articulation clarity, and pitch adjustments are evident and are further influenced by other factors such as context and age, with children exemplifying exaggerated prosody more than adults, likely to guarantee intelligibility (Cohn et al., 2024b). To date, however, research has predominantly considered English varieties (e.g., Cohn et al., 2024a), with far less attention to non-Western varieties. This paper thereby addresses this gap, examining adjustments in speech rate in HDS vs. DDS by native speakers of Najdi Arabic (NA) and whether differences are statistically significant by age

© The International Linguistic Society

group. Accordingly, we hypothesized that NA speakers will produce slower utterances when directed to an AI voice assistant.

Methodology

The current study tested speech rate under three conditions (ChatGPT–DDS, familiar and unfamiliar human–HDS) and compared two age groups (adolescents and adults). Across these interlocutors, participants produced identical sets of prompts of interrogative utterances.

Participants

A total of $N = 20$ NA participants were recruited through the native Najdi society. Recruitment criteria included dialect nativeness (i.e., including both familiar "experimenter" and unfamiliar interlocutors) and aged between 13–18 (M = 15.5, SD = 1.71) and 30–45 (M = 37.5, SD = 4.61) years old for teenagers and adults, respectively. All the participants were female, had no speech impairments, and were familiar with AI usage.

Procedure

Participants completed the experiment in person with a familiar "experimenter" and remotely with unfamiliar human interlocutors. First, participants were asked to introduce themselves. Then, they asked 10 interrogative sentences of varying length—short (7/8 syllables, M = 7.5, SD = 0.5) and long (14/15 syllables, M = 14.5, SD = 0.5)—per addressee in a counterbalanced order. For each interrogative utterance, speech rate was measured (mean number of syllables per second: SpS) via *Praat* (version 6.4.27), and the extracted data were then analysed in a linear mixed effects model.

Results

The aggregate results revealed an average speech rate of 5.41 SpS. Relative to ChatGPT–DDS, the speech rate was faster in familiar–HDS [$coef = 0.299$, $p < 0.049$]. Although there was no statistically significant effect of age [$coef = -0.248$, $p > 0.269$], there was an effect of sentence length; over the course of the experiment, participants tended to accelerate speech rate when producing short interrogatives [$coef = 0.484$, $p < 0.001$].

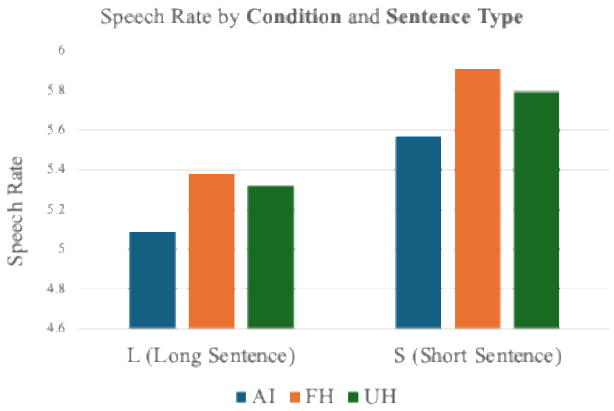

Figure 1. Condition and Sentence Type

Figure 1. Condition and sentence type.

With respect to interaction effects, the model showed that none achieved statistical significance [all $p > 0.05$]. As illustrated in Fig. 1, the fastest speech rate occurred for short sentences (S) under all conditions. The largest speed gap was identified in the familiar human (FH) condition (5.9 vs. 5.4 SpS) and the smallest in AI (5.55 vs. 5.1 SpS), with the unfamiliar human (UH) condition falling in between. Adults outpaced teenagers in sentence length. That is, adults articulated at 5.0 and 5.6 SpS, whereas teenagers articulated at 4.8 and 5.35 SpS for long and short sentences, respectively. However, differences by age were not statistically significant.

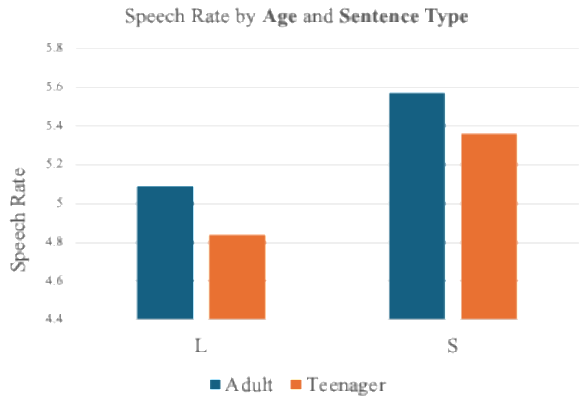

Figure 2. Age and Sentence Type

Figure 2. Age and sentence type.

Discussion and conclusion

This study tested adjustments that emerged when NA speakers interacted with a voice assistant compared to familiar and unfamiliar human interlocutors. There were consistent adaptations for the voice assistant: Speech directed toward the device was slower, which postulates a style shift in comparison to human interlocutors. In related work, a faster speech rate has been observed for an unfamiliar addressee compared to a familiar human (Cohn et al., 2024b). In the current study, however, speakers produced faster speech when talking to a familiar human. In contrast to Cohn et al. (2024a), who identified longer and higher-pitched utterances among children overall compared to adults, this study identified nearly no statistically significant age-based differences within teenagers and adults. This experiment has several limitations that can be addressed in future work. First, the present study examined one variety of dialectal Arabic, Najdi Arabic, but many other related varieties are comparably rare-to-unattested. Second, the current examined acoustic feature was speech rate, aiming to probe the speaking speed within speech directed to devices and NA interlocutors. However, NA exhibits distinct features, such as pitch variation and duration, that speakers might alter when addressing voice assistants.

References

Cohn, M., Zellou, G. 2021. Prosodic Differences in Human- and Alexa-Directed Speech, but Similar Local Intelligibility Adjustments. Frontiers in Communication 6.

Cohn, M., Liang, K.-H., Sarian, M., Zellou, G., Yu, Z. 2021. Speech Rate Adjustments in Conversations with an Amazon Alexa Socialbot. Frontiers in Communication 6.

Cohn, M., Segedin, B. F., Zellou, G. 2022. Acoustic-phonetic properties of Siri- and human-directed speech. Journal of Phonetics 90.

Cohn, M., Mengesha, Z., Lahav, M., Heldreth, C. 2024a. African American English speakers' pitch variation and rate adjustments for imagined technological and human addressees. JASA Express Letters 4.

Cohn, M., Barreda, S., Graf Estes, K., Yu, Z., Zellou, G. 2024b. Children and adults produce distinct technology- and human-directed speech. Scientific Reports 14.

Christenson, B., Ringler, C., Sirianni, N. J. 2023. Speaking fast and slow: How speech rate of digital assistants affects likelihood to use. Journal of Business Research 163.

Song, J. Y., Pycha, A., Culleton, T. 2022. Interactions between voice-activated AI assistants and human speakers and their implications for second-language acquisition. Frontiers in Communication 7.