

Will AI shape the way we speak?

The emerging sociolinguistic influence of synthetic voices

Éva Székely¹ Jūra Miniota¹ Míša (Michaela) Hejná²

¹Department of Speech, Music and Hearing, KTH Royal Institute of Technology, Sweden

²Department of English, Aarhus University, Denmark

szekely@kth.se, jura@kth.se, misa.hejna@cc.au.dk

Abstract

The growing prevalence of conversational voice interfaces, powered by developments in both speech and language technologies, raises important questions about their influence on human communication. While written communication can signal identity through lexical and stylistic choices, voice-based interactions inherently amplify socioindexical elements – such as accent, intonation, and speech style – which more prominently convey social identity and group affiliation. There is evidence that even passive media such as television is likely to influence the audience’s linguistic patterns. Unlike passive media, conversational AI is interactive, creating a more immersive and reciprocal dynamic that holds a greater potential to impact how individuals speak in everyday interactions. Such heightened influence can be expected to arise from phenomena such as acoustic-prosodic entrainment and linguistic accommodation, which occur naturally during interaction and enable users to adapt their speech patterns in response to the system. While this phenomenon is still emerging, its potential societal impact could provide organisations, movements, and brands with a subtle yet powerful avenue for shaping and controlling public perception and social identity. We argue that the socioindexical influence of AI-generated speech warrants attention and should become a focus of interdisciplinary research, leveraging new and existing methodologies and technologies to better understand its implications.

1 Introduction

This position paper proposes that the increasing scale and quality of verbal interactions with AI has the potential to influence people’s habitual voice and speaking style on an unprecedented scale. Recent advancements in large language models (LLMs) and text-to-speech (TTS) technology now enable realistic, expressive, human-like conversations. Moreover, breakthroughs in conversational

AI systems, such as naturalistic turn-taking (Arora et al., 2025) and interruption handling (Cao et al., 2025), are expected to drastically increase the scale of spoken interactions with AI. While both written and spoken language can convey aspects of identity, they do so through different channels. In writing – especially in informal settings – word choice, grammar, and style can reflect social traits such as age, gender, or cultural affiliation (e.g., Rubin 1995). However, spoken interaction inherently and therefore unavoidably conveys such extralinguistic traits through the voice itself. This means that the societal impacts of increased voice-based interactions with AI are likely to differ considerably from those of text-based interactions.

2 Sociointeractionality in spoken AI interaction

2.1 Spoken language and social identity

One particularly relevant concept in this context is socioindexicality, which refers to how features of communication signal social identity and group affiliation (Silverstein, 2003; Eckert, 2019). In spoken language, socioindexical elements, such as accent, intonation, and speech style, play a crucial role in conveying these social cues. A wide range of identity-related aspects – including personality and wellbeing – can be signalled, and indexed, through linguistic variation, including phonetic and phonological variation (e.g. Campbell-Kibler 2009; Pharo et al. 2014; Podesva and Callier 2015; Paladino and Mazzurega 2019; Guy et al. 2022; Hope and Lilley 2023; Grammon 2024).

Given the increasing realism and human-likeness of synthetic voices, socioindexical elements embedded in AI-generated speech may extend the role of conversational AI beyond functionality. These elements could potentially become socially influential, producing tangible effects on users’ perceptions and behaviors through specific vocal traits.

2.2 Acoustic-prosodic entrainment and linguistic accommodation

Entrainment (also called alignment, accommodation, or convergence) refers to the tendency of dialogue partners to become more similar in their communicative behaviors (Levitan and Hirschberg, 2011; Wynn and Borrie, 2022). In human-human conversations, people naturally align on various levels – choice of words, sentence structures, speech rate, intonation, etc. – which can foster rapport (Miles et al., 2009), signify cooperation (Pellegrino and Dellwo, 2023) and reinforce social bonds between speakers.

A substantial body of work shows that humans do adjust their speech and language when interacting with machines. Even in early studies of human-machine dialogue, researchers observed entrainment effects that parallel those found in human-human conversation. Users adapt their speech to align with artificial interlocutors in both lexical and prosodic domains. For instance, speakers converged on the vocabulary used by spoken dialogue systems (Parent and Eskenazi, 2010) and conversational agents (Ostrand et al., 2023). Prosodic convergence has also been documented in interactions with animated personas (Oviatt et al., 2004), social robots (Cohn et al., 2023), and virtual tutors (Tsfasman et al., 2021). Participants modulated features like pitch, amplitude, and speech rate to more closely match the agent’s delivery. Speakers even adjusted their speaking rate when addressing early spoken dialogue systems (Bell et al., 2003), and entrained to turn-taking rhythms in expressive humanoid robots (Breazeal, 2002). More recent findings show that the degree of prosodic entrainment can vary based on the agent’s politeness and perceived humanness (Horstmann et al., 2024; Tsfasman et al., 2021). These findings indicate that entrainment in HCI is not limited to functional adaptation, but it also reflects socially grounded mechanisms that operate similarly with both artificial and human interlocutors.

2.3 From alignment to identity expression

Linguistic accommodation is commonly viewed in sociolinguistics as a key mechanism that may influence how linguistic variation evolves into dialect formation and, eventually, language change (Hinskens and Auer, 2005). In other words, short-term accommodation during repeated conversational exchanges can, over time, lead to long-term

changes both at the individual level (Nguyen and Delvaux, 2015; Lee, 2010), as well as at the community level, where it can lead to the spread and adoption of innovative linguistic variants (Hinskens and Auer, 2005). Perceived prestige – often associated with artificial intelligence – has been shown to amplify this effect (Lev-Ari and Peperkamp, 2014). Linguistic accommodation being a reciprocal process, the rise of adaptive conversational AI (Brandt and Hazel, 2025; Pollmann et al., 2023) can be expected to reinforce this phenomenon even further.

This suggests that people could begin to absorb AI-influenced speech patterns in general contexts, potentially shaping their everyday language and, with it, their expression of identity. Evidence of a similar influence is already emerging with text-based chatbots, where users adopt words or phrases commonly generated by language models and subsequently use them in their spoken language, as observed in YouTube videos (Yakura et al., 2024).

3 Potential societal influence

3.1 Lessons from media

Over the past three decades, sociolinguistic research has explored how media influences speech patterns and linguistic performance (Tagliamonte, 2014) and how it contributes to language change (Kristiansen, 2014). Studies show that exposure to media can diffuse linguistic features, both on the lexical (Trudgill, 2014) and on the phonological level (Oviatt et al., 2004). While most of the research in this broader area has targeted either written language (Crystal, 2006; Tagliamonte, 2016) or the potential effects of modes of communication such as Instant Messaging on spoken language (Tagliamonte and Denis, 2008), one of the most notable endeavours in the area of speech influence is presented by Sayers (2014), who proposes a mediated innovation model to operationalise the role of media exposure and engagement on ‘everyday’ linguistic and speech changes.

Regarding phonetic and phonological features, few sociolinguistic studies are available. One prominent example of media influence on the acoustic-phonetic level is the phonological shifts observed in Glaswegian speech linked to psychological engagement with a popular London-based TV drama (Stuart-Smith et al., 2013). The researchers found that TH-fronting and L-vocalisation can be linked to psychological engagement with characters on the *EastEnders* soap opera.

Beyond linguistic variation, [Kristiansen \(2014\)](#) explores how the media shape language change through *ideology* – that is, socially shared beliefs about which ways of speaking are desirable, appropriate, or prestigious – and calls for further research into the media’s role in shaping such perceptions. As conversational AI becomes a more common mode of media engagement, it participates in these ideological processes, subtly reinforcing or shifting language attitudes through ongoing, interactive exposure and perceived prestige ([Xi, 2024](#)).

3.2 The rise of an engineered language change?

As shown by prior research, even before the widespread adoption of AI voices, media had already demonstrated its potential to influence how people speak and express themselves – often in ways that extend beyond direct interpersonal interaction and diffuse across distant geographic regions. Generative AI introduces an interactive dimension that is likely to amplify such influences. Speakers could actively – yet often unconsciously – incorporate socioindexical traits exhibited by conversational agents in their habitual speaking style. Through this process, companies, political movements, and other organizations may gain a new avenue for subtle influence. By designing AI voices with specific socioindexical characteristics – such as accents, speech styles, or voice quality features – these actors may encourage users to adopt speech patterns that signal affiliation with a brand, ideology, or social group. This influence could shape social identity markers and foster subconscious associations with particular movements or subcultures. Likely emerging examples of this phenomenon include AI companions ([Zhang and Li, 2025](#)), AI-powered interactive virtual influencers ([Yu et al., 2024](#)), and chatbot versions of human influencers.

3.3 Societal implications and ethical risks

It is further pertinent to ask to what extent any linguistic profiling within AI voices might contribute to linguistic discrimination, which is an established phenomenon: linguistic variants can be and have been utilised to classify speakers into social categories and to mistreat these speakers as a result ([Purnell et al., 1999](#); [Gluszek and Dovidio, 2010](#); [Lippi-Green, 2012](#); [Krahé and Papakonstantinou, 2020](#)). On a more general level, AI may reinforce normative biases by defaulting to voices representative of the unmarked and commonly employed

white, cis-gender, heterosexual, and able-bodied speakers. This may reinforce already existing dominant norms. In this light, the potential influence of conversational AI on speech production is not merely a linguistic curiosity but could pose actual ethical harm ([Hutiri et al., 2024](#)).

We believe that the potential societal impacts of AI-driven socioindexical influence on speech patterns and identities can be rather substantial. While empirical evidence is still emerging, we identify socioindexical influence as an under-explored area with significant potential for societal impact. Understanding this phenomenon now, while it is nascent, can be an opportunity to shape ethical design and governance before its effects become widespread. However, whether this is the case, and to what extent, remains unexplored. Considering the rapid advances in relevant technology and the widespread engagement with conversational AI, it is important to develop methods for understanding which speech characteristics may become influential or habitualised through interaction with synthetic voices.

4 Research opportunities and challenges

4.1 Studying short- and long-term effects

While short-term accommodation to AI voices in interaction is established, it remains unclear whether and how these immediate, conversation-specific adaptations carry over into one’s long-term speech habits outside the interaction. Most studies to date examine alignment within an interaction; they do not test if a person’s baseline speaking style changes after repeated exposures. Studying the nature of the long-term potential and topical influences requires methodologies that extend beyond but include traditional sociolinguistic approaches, particularly when considering the interactive nature of modern media and conversational systems ([Sayers, 2014](#)). Indeed, the still highly unique study by [Stuart-Smith et al. \(2013\)](#) presents a tour de force which, among other things, demonstrates the methodological complexities and challenges of investigating the role of the media on phonetic speech variation outside of a laboratory setting. Individual variation in susceptibility to such influence should also be considered. Not all speakers will accommodate to synthetic voices the same way – some may even actively resist alignment. Future work should explore who adapts, who resists, and why.

4.2 Experimental approaches with TTS and Conversational AI

The same technologies that raise questions about socioindexical influence – speech synthesis and conversational AI – also bring new methodological possibilities. Advances in speech synthesis provide researchers with unprecedented control over acoustic-prosodic features, enabling experimental designs that isolate individual variables such as pitch, speech rate, and voice quality. These systems can also affect features such as formality, allowing the development of methodologies that use TTS trained on spontaneous speech data as a research tool (Székely et al., 2024; O’Mahony et al., 2024). Moreover, recent developments in large-scale neural TTS systems trained on thousands of hours of speech have dramatically lowered the threshold for high-fidelity voice replication (Casanova, 2024). Fine-tuning these models on as little as ten minutes of in-the-wild speech material makes it possible to reproduce sociolects without requiring extensive recordings. In addition, zero-shot TTS and voice conversion (Lameris et al., 2024) enable the transfer of these speech patterns to different voice identities, which facilitates experimental comparisons across demographic categories like gender and age and even vocal characteristics. While such manipulations must be approached with care to preserve indexical plausibility and perceptual coherence (Seaborn et al., 2025), this capacity for decoupling linguistic features from speaker identity expands the range of testable hypotheses in experimental sociolinguistics.

Such synthetic stimuli can be used in both perception and production studies, including shadowing tasks (Laycock, 2021; Pardo et al., 2018), to estimate phonetic convergence to emerging sociolects. Interactive experimental designs also become feasible through research-grade conversational agents equipped with controllable TTS (Wang et al., 2024), enabling A/B testing of entrainment during dialogue. These methods can be further complemented by sociolinguistic interviews or ethnographic observations on social media materials (Yakura et al., 2024).

4.3 Multidisciplinary opportunities

The complexity of media influence, which is shaped by engagement, identity, and context, calls for a transdisciplinary approach (Androutsopoulos, 2014). Studying the sociolinguistic impact of syn-

thetic voices, and especially the conditions under which local adaptation might evolve into long-term language change, will require collaboration across multiple disciplines. Building research-purpose TTS and Conversational AI systems, analysing subtle language and speech variations, and interpreting social impact are key components of this research direction and will necessitate close collaboration between engineers, linguists, social scientists, and ethicists. We anticipate that the increasing presence of speech AI in society will lead to further research areas becoming increasingly multidisciplinary. This may require rethinking research infrastructures or even education programs.

5 Conclusion

This position paper calls attention to the need for a concerted effort to address the socioindexical influence of AI-generated voices in interaction. First, it is imperative to establish the existence and extent of this emergent phenomenon. This requires a foundational understanding of how AI interaction may impact speech patterns and identity expression among users. Second, the development of robust methodologies is critical for systematically studying and measuring these influences. Finally, we must begin to consider the broader implications, including ethical, societal, and legal dimensions. Addressing these priorities will help us prepare to understand and manage the implications of voice-based conversational AI for human speech, communication, social identity, and its potential role in driving language change.

Acknowledgements

This research is supported by the Swedish Research Council project Perception of speaker stance (VR-2020-02396), and the Riksbankens Jubileumsfond project CAPTivating (P20-0298). We thank the anonymous reviewers for their insightful comments which helped improve the final version of this paper. We are also deeply grateful to the many colleagues and friends who generously shared feedback on earlier drafts, including Alyssa Allen, Matthew Aylett, Jens Edlund, Emer Gilmartin, Maxwell Hope, David House, Tanya Karoli Christensen, Johannah O’Mahony, Amir H. Payberah, André Pereira, Graham Pullin, Fin Tams-Grey, Iliaria Torre and Marcin Włodarczak. Their thoughtful suggestions and encouragement were invaluable in shaping this work.

References

- Jannis Androutsopoulos. 2014. [Beyond ‘media influence’](#). *Journal of Sociolinguistics*, 18(2):242–249.
- Siddhant Arora, Zhiyun Lu, Chung-Cheng Chiu, Ruoming Pang, and Shinji Watanabe. 2025. Talking turns: Benchmarking audio foundation models on turn-taking dynamics. *arXiv preprint arXiv:2503.01174*.
- Linda Bell, Joakim Gustafson, and Mattias Heldner. 2003. Prosodic adaptation in human-computer interaction. In *Proc. ICPHS*, volume 3, pages 833–836.
- Adam Brandt and Spencer Hazel. 2025. Towards interculturally adaptive conversational ai. *Applied Linguistics Review*, 16(2):775–786.
- Cynthia Breazeal. 2002. Regulation and entrainment in human—robot interaction. *The International Journal of Robotics Research*, 21(10-11):883–902.
- Kathryn Campbell-Kibler. 2009. The nature of sociolinguistic perception. *Language Variation and Change*, 21:135–156.
- Shiye Cao, Jiwon Moon, Amama Mahmood, Victor Nikhil Antony, Ziang Xiao, Anqi Liu, and Chien-Ming Huang. 2025. Interruption handling for conversational robots. *arXiv preprint arXiv:2501.01568*.
- Edresson Casanova. 2024. XTTS: a massively multilingual zero-shot text-to-speech model. In *Proc. Interspeech*, pages 4978–4982.
- Michelle Cohn, Ashley Keaton, Jonas Beskow, and Georgia Zellou. 2023. Vocal accommodation to technology: the role of physical form. *Language Sciences*, 99:101567.
- David Crystal. 2006. *Language and the Internet*. Cambridge University Press.
- Penelope Eckert. 2019. The limits of meaning: Social indexicality, variation, and the cline of interiority. *Language*, 95(4):751–776.
- Agata Gluszek and John F. Dovidio. 2010. [Speaking with a nonnative accent: perceptions of bias, communication difficulties, and belonging in the united states](#). *Journal of Language and Social Psychology*, 18(2):224–234.
- Devin Grammon. 2024. Ideology, indexicality, and the 12 development of sociolinguistic perception during study abroad. *L2 Journal: An Open Access Refereed Journal for World Language Educators*, 16(1):1–17.
- Gregory R. Guy, Livia Oushiro, and Ronald Beline Mendes. 2022. Indexicality and coherence. In *The Coherence of Linguistic Communities. Orderly Heterogeneity and Social Meaning*, pages 53–68.
- FLMP Hinskens and Peter Auer. 2005. The role of interpersonal accommodation in a theory of language change. In *Dialect change. The convergence and divergence of dialects in contemporary societies*, pages 335–357. Cambridge UP.
- Maxwell Hope and Jason Lilley. 2023. Differences in sibilant perception between gender expansive and cisgender individuals. *Seminars in Speech and Language*, 44(2):61–75.
- Aike C Horstmann, Clara Strathmann, Lea Lambrich, and Nicole C Kramer. 2024. Communication style adaptation in human-computer interaction. *Human-Machine Communication*, 8:53–72.
- Wiebke Hutiri, Orestis Papakyriakopoulos, and Alice Xiang. 2024. Not my voice! a taxonomy of ethical and safety harms of speech generators. In *Proc. of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, page 359–376.
- Barbara Krahé and Lida Papakonstantinou. 2020. [Speaking like a man: women’s pitch as a cue for gender stereotyping](#). *Sex Roles*, 82:94–101.
- Tore Kristiansen. 2014. [Knowing the driving force in language change: density or subjectivity?](#) *Journal of Sociolinguistics*, 18(2):233–241.
- Harm Lameris, Joakim Gustafson, and Éva Székely. 2024. Creakvc: a voice conversion tool for modulating creaky voice. In *Proc. Interspeech*, pages 1005–1006.
- Kyle Laycock. 2021. [The influence of socioindexical information on the speech perception-production link: Evidence from a shadowing task](#). *Theses and Dissertations—Linguistics*. 41.
- Chi-Chun Lee. 2010. Quantification of prosodic entrainment in affective spontaneous spoken interactions of married couples. In *Proc. Interspeech*, page 793.
- Shiri Lev-Ari and Sharon Peperkamp. 2014. An experimental study of the role of social factors in language change: The case of loanword adaptations. *Laboratory Phonology*, 5(3):379–401.
- Rivka Levitan and Julia Hirschberg. 2011. [Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions](#). In *Interspeech 2011*, pages 3081–3084.
- Rosina Lippi-Green. 2012. *English with an Accent: Language, Ideology, and Discrimination in the United States*.
- Lynden K Miles, Louise K Nind, and C Neil Macrae. 2009. The rhythm of rapport: Interpersonal synchrony and social perception. *Journal of experimental social psychology*, 45(3):585–589.
- Noël Nguyen and Véronique Delvaux. 2015. Role of imitation in the emergence of phonological systems. *Journal of Phonetics*, 53:46–54.
- Johannah O’Mahony, Catherine Lai, and Éva Székely. 2024. Well, what can you do with messy data? exploring the prosody and pragmatic function of the discourse marker "well" with found data and speech synthesis. In *Proc. Interspeech*, pages 4084–4088.

- Rachel Ostrand, Victor S. Ferreira, and David Piorkowski. 2023. Rapid lexical alignment to a conversational agent. In *Proc. Interspeech*, page 2653.
- Sharon Oviatt, Courtney Darves, and Rachel Coulston. 2004. [Toward adaptive conversational interfaces: Modeling speech convergence with animated personas](#). *Proc. TOCHI*, 11(3):300–328.
- Maria Paola Paladino and Mara Mazurega. 2019. One of us: On the role of accent and race in real-time in-group categorization. *Journal of Language and Social Psychology*, 39(1):22–39.
- Jennifer S Pardo, Adelya Urmanche, Sherilyn Wilman, Jaclyn Wiener, Nicholas Mason, Keagan Francis, and Melanie Ward. 2018. A comparison of phonetic convergence in conversational interaction and speech shadowing. *Journal of Phonetics*, 69:1–11.
- Gabriel Parent and Maxine Eskenazi. 2010. Lexical entrainment of real users in the let’s go spoken dialog system. In *Proc. Interspeech*, pages 3018–3021.
- Elisa Pellegrino and Volker Dellwo. 2023. Speakers are more cooperative and less individual when interacting in larger group sizes. *Frontiers in Psychology*, 14:1145572.
- Nicolai Pharao, Marie Maegaard, Janus Spindler Møller, and Tore Kristianse. 2014. Indexical meanings of [s+] among Copenhagen youth: Social perception of a phonetic variant in different prosodic contexts. *Language in Society*, 43:1–31.
- Robert J. Podesva and Patrick Callier. 2015. Voice quality and identity. *Annual Review of Applied Linguistics*, 35:173–194.
- Kathrin Pollmann, Wulf Loh, Nora Fronemann, and Daniel Ziegler. 2023. Entertainment vs. manipulation: Personalized human-robot interaction between user experience and ethical design. *Technological Forecasting and Social Change*, 189:122376.
- Thomas Purnell, William Idsardi, and John Baugh. 1999. Perceptual and phonetic experiments on american english dialect identification. *Journal of Language and Social Psychology*, 18:10–30.
- Donald L. Rubin, editor. 1995. *Composing social identity in written language*. Lawrence Erlbaum Associates.
- Dave Sayers. 2014. The mediated innovation model: A framework for researching media influence in language change. *Journal of sociolinguistics*, 18(2):185.
- Katie Seaborn, Katja Rogers, Maximilian Altmeyer, Mizuki Watanabe, Yuto Sawa, Somang Nam, Tatsuya Itagaki, and Ge ‘Rikaku’ Li. 2025. Unboxing manipulation checks for voice ux. *Interacting with Computers*, page iwae062.
- Michael Silverstein. 2003. Indexical order and the dialectics of sociolinguistic life. *Language & Communication*, 23(3-4):193–229.
- Jane Stuart-Smith, Gwilym Pryce, Claire Timmins, and Barrie Gunter. 2013. Television can also be a factor in language change: Evidence from an urban dialect. *Language*, pages 501–536.
- Éva Székely, Jeff Higginbotham, and Francesco Possemato. 2024. Voice and choice: Investigating the role of prosodic variation in request compliance and perceived politeness using conversational TTS. In *Proc. SIGDial*, pages 466–476.
- Sali Tagliamonte. 2014. [Situating media influence in sociolinguistic context](#). *Journal of Sociolinguistics*, 18(2):223–232.
- Sali Tagliamonte. 2016. [So sick or so cool? the language of youth on the internet](#). *Language in Society*, 45(1).
- Sali Tagliamonte and Derek Denis. 2008. [Linguistic ruin? lol! instant messaging and teen language](#). *American Speech*, 83(1):3–324.
- Peter Trudgill. 2014. [Diffusion, drift, and the irrelevance of media influence](#). *Journal of Sociolinguistics*, 18(2):213–222.
- Maria Tsfasman, Avinash Saravanan, Dekel Viner, Daan Goslinga, Sarah De Wolf, Chirag Raman, Catholijn M Jonker, and Catharine Oertel. 2021. Towards a real-time measure of the perception of anthropomorphism in human-robot interaction. In *Proc. ACM MuCAL*, pages 13–18.
- Siyang Wang, Éva Székely, and Joakim Gustafson. 2024. Contextual interactive evaluation of TTS models in dialogue systems. In *Proc. Interspeech*, pages 2965–2969.
- Camille J Wynn and Stephanie A Borrie. 2022. Classifying conversational entrainment of speech behavior: An expanded framework and review. *Journal of Phonetics*, 94:101173.
- Yipeng Xi. 2024. Navigating technological shifts: An examination of user inertia and technology prestige in large-language-model ai chatbot transition. *International Journal of Human-Computer Interaction*, 0(0):1–17.
- Hiromu Yakura, Ezequiel Lopez-Lopez, Levin Brinkmann, Ignacio Serna, Prateek Gupta, and Iyad Rahwan. 2024. Empirical evidence of large language model’s influence on human spoken communication. *arXiv preprint arXiv:2409.01754*.
- Joanne Yu, Astrid Dickinger, Kevin Kam Fung So, and Roman Egger. 2024. Artificial intelligence-generated virtual influencer: Examining the effects of emotional display on user engagement. *Journal of Retailing and Consumer Services*, 76:103560.
- Shuning Zhang and Shixuan Li. 2025. The real her? exploring whether young adults accept human-ai love. *arXiv preprint arXiv:2503.03067*.