

AI-cent judgements: Accent stereotypes persist with AI-generated voices

Jun Liu¹, Leah Bradshaw^{1,2}, and Eleanor Chodroff¹

¹Department of Computational Linguistics, University of Zurich, Switzerland

{jun.liu|leah.bradshaw|eleanor.chodroff}@uzh.ch

²JP French International, UK.

{leah.bradshaw}@jpfrench.com

Recent advances in neural Text-to-Speech (TTS) systems have enabled the creation of highly realistic synthetic voices capable of reproducing human-like intonation, rhythm, and conversational fluency (Do et al., 2022). While existing research has largely focused on the technological development and detection of synthetic speech (Almutairi & Elgibreen, 2022; Khurana et al., 2023), AI-generated voices also provide a valuable framework for investigating how listeners evaluate social identity, credibility, and authenticity through speech. Because synthetic voices allow accent, gender, and vocal style to be manipulated while holding semantic content constant, they create new opportunities for examining whether traditional sociolinguistic stereotypes persist in human-AI interaction. This study explores how regional accent and perceived gender shape listener evaluations of synthetic speech. Previous studies demonstrate that listeners rapidly attribute personality traits and social identities based solely on accent, often overriding objective acoustic evaluation (Shamina, 2016). Indeed, empirical evidence demonstrates that listeners form several biases around a speaker based on their accent alone (e.g., Garrett et al., 2003; Coupland & Bishop, 2007; Paver et al., 2025). AI-generated speech therefore provides a useful model for investigating whether entrenched social evaluations persist even when listeners know that a voice is synthetic. The study additionally considers how listeners judge “humanness” and “naturalness” in AI-generated speech. Understanding which cues listeners prioritise has important implications for forensic speech science, where synthetic speech is becoming an increasingly important consideration in fraud investigations, disputed recordings, and the evaluation of voice evidence.

To investigate these questions, 40 listeners ages 21-40 participated in a Likert-scale rating task, comprising 20 British English and 20 American English participants (21 F, 18 M, 1 preferred not to report gender) in which they were informed that all voices were AI-generated. Participants evaluated 16 AI-generated voices created using ElevenLabs; two male and two female voices for four accent backgrounds (Yorkshire English, SSBE, General American, Southern American). Participants rated each voice from 1 to 6 on perceived trustworthiness, professionalism, human-likeness of the voice and naturalness of the overall audio, alongside providing qualitative comments regarding their judgements. Findings demonstrate that traditional accent stereotypes persist within evaluations of synthetic speech. Across listener groups, Yorkshire English voices received the highest overall trustworthiness ratings, while SSBE voices were consistently ranked the most professional (Figures 1 and 2). Furthermore, ratings of “human-like” voices and “natural” audio quality were found to be highly correlated ($r = 0.88$, $p < 0.001$). Qualitative analysis indicated that listeners frequently relied on the same vocal cues when making both judgements, focusing primarily on characteristics of the voice itself rather than on broader audio artefacts. This suggests that lay listeners may evaluate AI speech authenticity through socially conditioned perceptions of vocal performance rather than cues that reliably distinguish synthetic from human speech. These findings have important implications for forensic speech science, highlighting that intuitive judgements about whether a voice is “real” may be influenced by accent stereotypes and perceptions of humanness rather than diagnostic acoustic evidence. As AI-generated speech becomes increasingly prevalent in criminal investigations, fraud, and disputed recordings, understanding these perceptual biases will be essential for interpreting listener judgements and informing best practice in the evaluation of voice evidence. Although forensic speech science increasingly relies on acoustic and computational analyses, human perceptual judgements remain central to many stages of casework, including the assessment of recordings, forensic transcription, earwitness testimony, and the interpretation of disputed voice evidence.

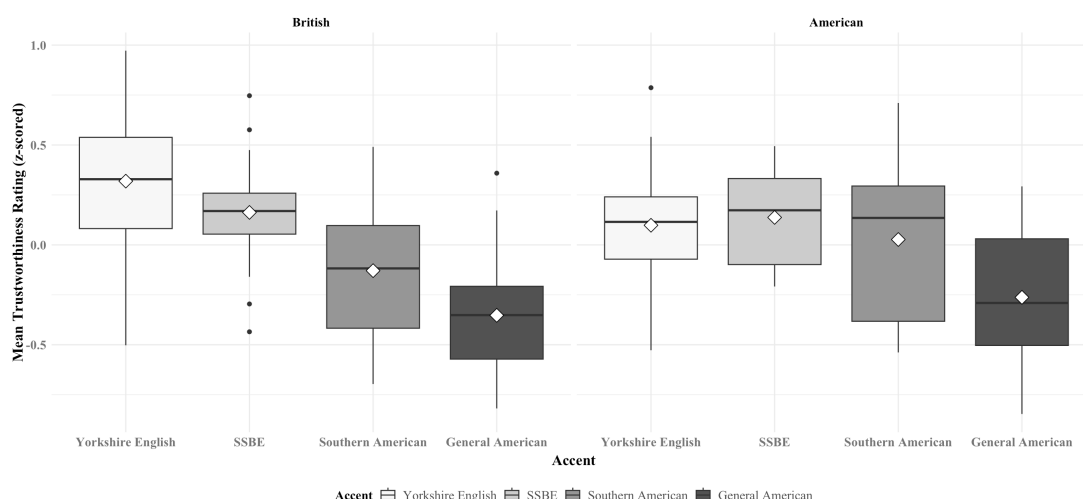


Figure 1. Mean trustworthiness ratings for British and American English listeners faceted by dialect type. Higher ratings indicate higher perceptions of trustworthiness from the voices.

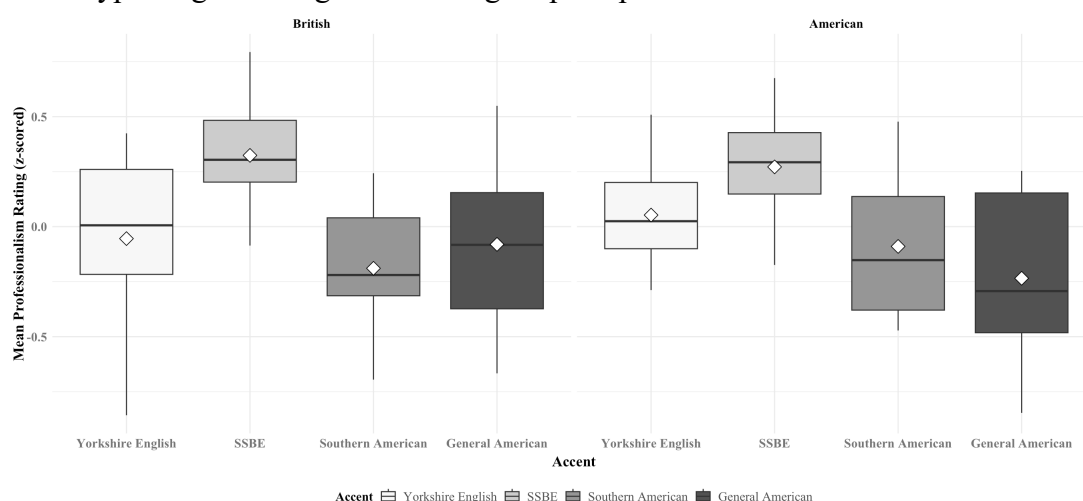


Figure 2. Mean professionalism ratings for British and American English listeners faceted by dialect type. Higher ratings indicate higher perceptions of professionalism from the voice.

References

- Almutairi, Z., & Elgibreen, H. (2022). A review of modern audio deepfake detection methods: challenges and future directions. *Algorithms*, 15(5), 155.
- Chaiken, S., & Trope, Y. (Eds.). (1999). *Dual-process theories in social psychology*. Guilford Press.
- Coupland, N., & Bishop, H. (2007). Ideologised values for British accents 1. *Journal of sociolinguistics*, 11(1), 74-93.
- Do, P., Coler, M., Dijkstra, J., & Klabbers, E. (2022). Text-to-speech for under-resourced languages: Phoneme mapping and source language selection in transfer learning. In *Proceedings of the 1st Annual Meeting of the ELRA/ISCA Special Interest Group on Under-Resourced Languages* (pp. 16–22).
- Garrett, P., Coupland, N., & Williams, A. (2003). *Investigating language attitudes: Social meanings of dialect, ethnicity and performance*. University of Wales Press.
- Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: state of the art, current trends and challenges. *Multimedia tools and applications*, 82(3), 3713-3744.
- Levine, T. R. (2014). Truth-default theory (TDT) a theory of human deception and deception detection. *Journal of Language and Social Psychology*, 33(4), 378-392.
- Paver, A., Wright, D., Braber, N., & Pautz, N. (2025). Stereotyped accent judgements in forensic contexts: listener perceptions of social traits and types of behaviour. *Frontiers in Communication*, 9, 1462013.
- Shamina, E. (2016). An experimental study of English accent perception. *ExLing* 2016, 159.
- Zahn, C. J., & Hopper, R. (1985). Measuring language attitudes: The speech evaluation instrument. *Journal of language and social psychology*, 4(2), 113-123.