University of Łódź
Institute of English Studies
Department of English Language and Applied Linguistics

Accents 2024

Accents in various contexts

17th International Conference on Native and Non-native Accents of English

Łódź, 12 Dec – 14 Dec, 2024

www.filolog.uni.lodz.pl/accents

THE BOOK OF ABSTRACTS

edited by:

Aleksandra Matysiak

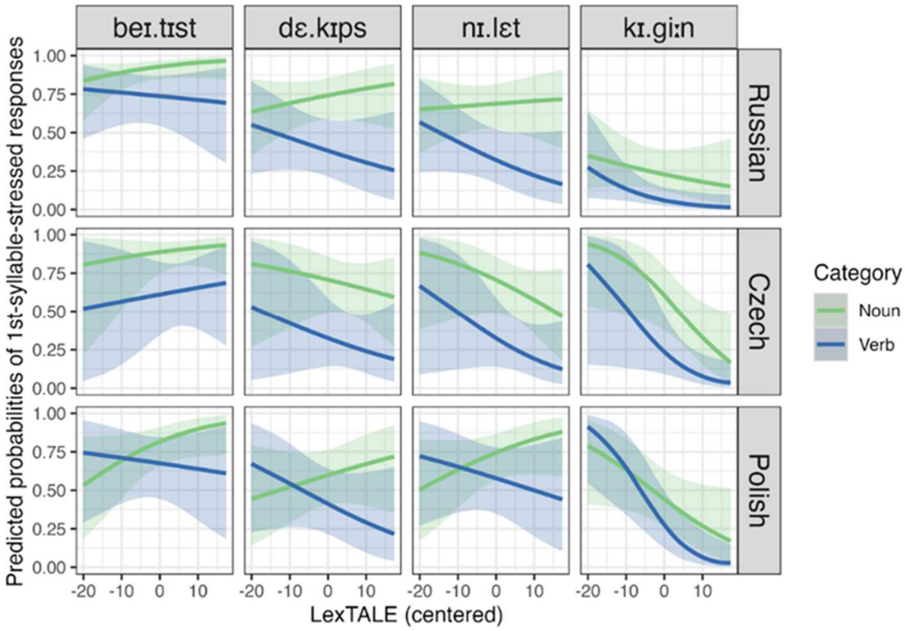

Figure 1. Probabilities of initial-stress responses predicted by a mixed-effect logistic regression model

References:

Guion, S. G., Clark, J. J., Harada, T., & Wayland, R. P. (2003). Factors affecting stress placement for English nonwords include syllabic structure, lexical class, and stress patterns of phonologically similar words. *Language and Speech*, *46*(4), 403-426.

Kijak, Anna. 2009. "How Stressful Is L2 Stress? A Cross-Linguistic Study of L2 Perception and Production of Metrical Systems." PhD diss., Universiteit Utrecht.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. Behavior Research Methods, 44, 325-343.

Peperkamp, S., & Dupoux, E. (2002). A typological study of stress 'deafness'. *Laboratory phonology*, *7*(2000), 203-240.

ASSESSING SMARTPHONE SPEECH RECOGNITION ACROSS DIVERSE ENGLISH ACCENTS: A PRELIMINARY STUDY

Claudia Soria
"A. Zampolli" Institute for Computational Linguistics, National Research Council, Italy
Rosalba Nodari
University of Siena, Italy
Silvia Calamai
University of Siena, Italy

Voice-activated artificial intelligence in smartphones is making spoken human-device interactions increasingly common, with many users utilizing these systems for everyday tasks such as creating shopping lists, dictating messages, or querying information (Ammari et al., 2019). The success of these interactions relies heavily on the accuracy of speech recognition technology embedded in devices, which can be significantly affected by accents and dialects.

Recent advancements have improved the recognition of various accents beyond standard British or American English, driven by the need to ensure equitable service and representation for diverse communities (Choe et al., 2022; Koenecke et al., 2020). Although some automatic speech recognition (ASR) systems embedded in smartphones offer recognition for certain second language (L2) English accents (Lai, 2021), research on their performance remains limited (Chan et al., 2022; Del Rio et al., 2023; Tadimeti et al., 2022).

This work presents preliminary findings from a study assessing the performance of common smartphone speech recognition systems with respect to a range of L1 (native) and L2 (non-native) English accents. The study utilized 36 audio clips from the CIRCE corpus[1], which consisted of the same short text read aloud by male and female speakers of four L1 and nine L2 English accents. The L1 accents included Standard American, African American, Standard British, and Multicultural London English, while the L2 accents covered Indian, Nigerian, Bosnian, Italian, Turkish, Ukrainian, Chinese, German, and Russian. Each clip averaged 0.32 seconds in length.

To simulate typical user experiences, the research evaluated Apple's Siri voice recognition for two everyday tasks: message/note dictation and voice search. The audio clips were played from a laptop with voice recognition activated on an iPhone using the Notes app. Siri's different English locales[2] (USA, UK, Australia, Canada, Japan, India, New Zealand, Singapore, and South Africa) were tested for each accent. Each clip was played three times, resulting in a total of 702 transcripts.

The study measured transcript accuracy using the Word Error Rate (WER) to compare and evaluate the performance of ASR systems. This new and unique comparable speech corpus provided insights into which L1 and L2 English accents are best recognized by common smartphones, as well as a comparative analysis of different automatic recognition models of local Englishes. Additionally, these preliminary results were compared with existing literature on human intelligibility of L1 and L2 accents (Verbeke and Simon, 2023).

References:

Ammari, T., Kaye, J., Tsai, J. Y., and Bentley, F. 2019. Music, search, and IoT: how people (really) use voice assistants. *ACM Transactions on Computer Human Interaction* 26(3), 1–28. doi: 10.1145/3311956

Besacier, L., Barnard, E., Karpov, A., and Schultz, T. 2014. Automatic speech recognition for under-resourced languages: A survey. *Speech Communication* 56. 85-100. doi:10.1016/j.specom.2013.07.008.

Chan, M.P.Y., Choe, J., Li, A., Chen, Y., Gao, X., and Holliday, N. 2022. Training and typological bias in ASR performance for world Englishes. Proceedings of Interspeech 2022, 1273-1277. doi: 10.21437/Interspeech.2022-10869

Choe, J., Chen, Y., Chan, M., Li, A., Gao, X., Holliday, N. 2022. Language-specific Effects on Automatic Speech Recognition Errors for World Englishes. *Proceedings of the 29th International Conference on Computational Linguistics*, pages 7177–7186.

Del Río, M., Miller, C., Profant, J., Drexler-Fox, J., Mcnamara, Q., Bhandari, N., Delworth, N., Pirkin, I., Jetté, M., Chandra, S., Ha, P., and Westerman, R. 2023. Accents in Speech Recognition through the Lens of a World Englishes Evaluation Set. *Research in Language*, *21*(3), 225–244. https://doi.org/10.18778/1731-7533.21.3.02

[1] https://www.circe-project.eu

[2] https://www.apple.com/ios/feature-availability/#quicktype-keyboard-dictation

Koenecke, A., Nam, A., Lake, E., Nudell, J., Quartey, M., Mengesha, Z., Toups, C., Rickford, J. R.,

Jurafsky, D., and Goel, S. 2020. Racial disparities in automated speech recognition. *Proceedings of the National Academy of Sciences*, *117*(14), 7684–7689. https://doi.org/10.1073/pnas.1915768117

Lai, A. 2021. Supported languages. *Otter.ai Help Center.*

Levis, J., and Suvorov, R. 2012. Automatic Speech Recognition. In Chapelle, Carol A. (ed.), *The Encyclopedia of Applied Linguistics*. Oxford, UK: Blackwell. doi:10.1002/9781405198431.wbeal0066.

Li, J., Deng, L., Haeb-Umbach, R., and Gong, Y. 2016. *Robust automatic speech recognition: a bridge to practical applications*. Waltham, MA: Academic Press.

Tadimeti, D., Georgila, K., and Traum, D. 2022. Evaluation of Off-the-shelf Speech Recognizers on Different Accents in a Dialogue Domain. *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, 6001–6008. https://aclanthology.org/2022.lrec-1.645

Verbeke, G., and Simon, E. 2023. Listening to accents: Comprehensibility, accentedness and intelligibility of native and non-native English speech. *Lingua*, *292*, 103572. https://doi.org/10.1016/j.lingua.2023.103572

SYSTEMATIC MEASUREMENTS OF FORMANT VALUES ACROSS THE CONTINUUM OF VOWEL ARTICULATION BASED ON RENDERINGS OF CARDINAL VOWELS BY DIFFERENT PHONETICIANS

Łukasz Stolarski
Jan Kochanowski University in Kielce

Attempts to create a reference system for vowel articulation, which began in the nineteenth century, were systematized by Jones (1914, 1917a, 1917b), who proposed the Cardinal Vowel System. Initially, the system included 16 items, but in the current version of the IPA Vowel Diagram, it encompasses 28 reference vowels. Acoustic measurements of these vowels available in the literature are quite limited in scope, typically focusing only on rough estimates of the first two or three formants for the Primary, and sometimes Secondary, series articulated by one or two male phoneticians (Koffi, 2018; Lindblom & Sundberg, 1969; Thomas, 2017; Vaissière, 2009, 2011).

This study presents a comprehensive acoustic analysis of Cardinal Vowels pronounced by a diverse group of 20 phoneticians, including both male and female speakers. The primary aim was to obtain reliable measurements of the first three formants across the entire continuum of vowel articulation. The methodology was designed to achieve the highest possible precision. This involved carefully selecting recordings based on articulatory quality, evaluating measurement outcomes within the broader context of other vowels, and applying "average spectra" and "average quasi-spectrograms" to correct any inaccuracies that are difficult to detect objectively when analyzing individual recordings in Praat.

The results provide the most accurate formant measurements to date across the vowel articulation continuum for both men and women. The data reveal consistent patterns across speakers, with expected gender differences in absolute values but similar relative formant structures. Additionally, the study highlights the significance of the formant merger phenomenon in acoustic analysis, identifying specific regions in the vowel space