Perceived foreign accent predicts intelligibility of foreign-accented speech

Yevgeniy Melguy & Keith Johnson

Previous literature has found that listeners' social expectations about a speaker can either inhibit or facilitate the intelligibility of speech (Babel & Mellesmoen, 2019; Babel & Russel, 2015; McGowan, 2015). For example, Babel and Russel (2015), found that listeners transcribed a native-accented Chinese-Canadian English talker less accurately in an audio-visual condition than in an audio-only condition, whereas McGowan (2015) found improved intelligibility of a Mandarin Chinese-accented voice when it was paired with an image of an Asian face (compared to a white face). This literature suggests a match-mismatch explanation: listeners tend to expect non-white individuals to be accented, and if their expectations are correct, speech intelligibility increases—if incorrect, intelligibility decreases. However, the mechanism underlying successful adaptation to accented speech remains unclear. If adaptation is general (i.e., not accent-specific), listeners should see a benefit simply from expecting to hear a non-native accent (e.g., if the speaker is non-white). If it is targeted, requiring "tuning in" to accent-specific phonetic patterns, listeners should *only* benefit when talker ethnicity is congruent with the perceived accent.

In experiment 1, monolingual speakers of American English (n = 139) were recruited via Amazon's Mechanical Turk to complete a transcription in noise task with short English sentences read by an L1 Mandarin speaker of English. All listeners heard the same Chinese-accented voice, but were randomly assigned to one of 4 visual conditions: they saw either a blank silhouette, a European face, an East Asian face, or a South Asian face. Results showed that while purported talker ethnicity was not a significant predictor of transcription accuracy, perceived accent was: subjects who did not report hearing a foreign accent displayed lower transcription accuracy than those who did (β = -1.20, z-score = -4.18, $p < 0.001$). However, it did not matter whether subjects correctly detected a Chinese accent or reported some other foreign accent – both groups showed identical performance (β = -0.03, z-score = -0.10, $p = 0.92$). In experiment 2, a different set of Mturk subjects (n = 104) completed the same transcription task, but this time preceded by a short training block where they were presented with faces corresponding to each visual condition in the test phase, paired with congruent voices (a South Asian face paired with Hindi-accented English, an East Asian face paired with Mandarin-accented English, and a European face paired with an American English accent). Despite the social priming introduced in the training phase, the pattern of post-training transcription results was largely consistent with the findings of experiment 1 – talker ethnicity did not significantly affect transcription accuracy, but perceived talker accent did: subjects who reported hearing a native accent (English) generally showed significantly worse transcription accuracy than those who reported a Chinese accent or another foreign accent. (β = -1.11, z-score = -3.46, $p < 0.001$).

Taken together, these results suggest that adaptation to foreign-accented speech is the result of a general rather than a targeted mechanism – detection of accent (whether "correct" or not) predicted improved perceptual performance. Crucially, however, lack of an overall effect of visual condition (purported speaker ethnicity) suggests that listeners may vary in terms of their perceptual flexibility and propensity to draw on stereotyped associations between a talker's ethnicity and language background to guide adaptation to non-natively accented speech.

References

Babel, M., & Mellesmoen, G. (2019). Perceptual adaptation to stereo-typed accents in audio-visual speech. In Sasha Calhoun, Paola Escudero, Marija Tabain & Paul Warren (eds.) *Proceedings of the 19th International Congress of Phonetic Sciences, Melboune, Australia 2019* (pp. 1044-1048). Canberra, Australia: Australasian Speech Science and Technology Association Inc.

Babel, M., & Russell, J. (2015). Expectations and speech intelligibility. *The Journal of the Acoustical Society of America, 137*(5), 2823-2833.

McGowan, K. B. (2015). Social expectation improves speech perception in noise. *Language and Speech, 58*(4), 502-521.

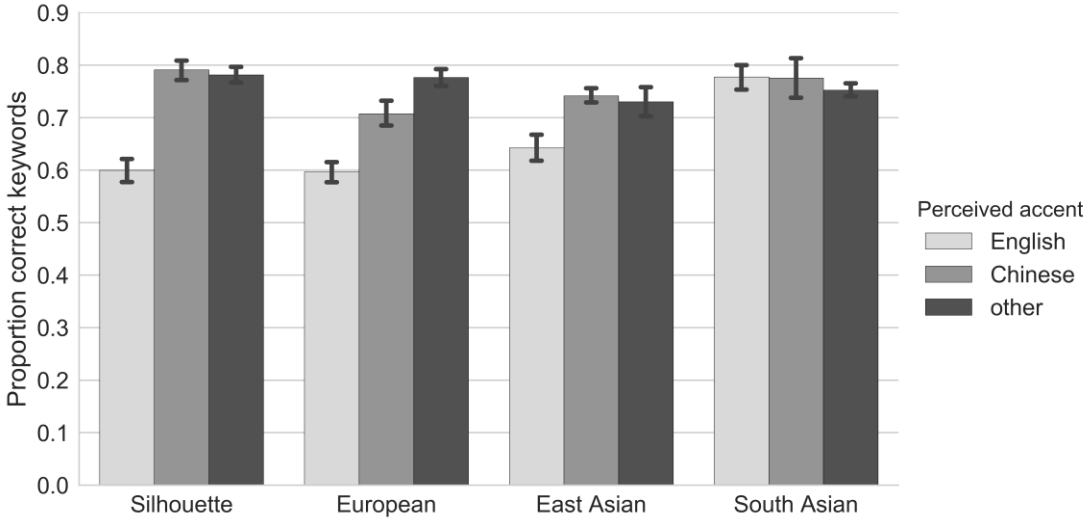

Figure 1. Mean proportion correctly transcribed keywords in experiment 1, by purported talker ethnicity and perceived accent.

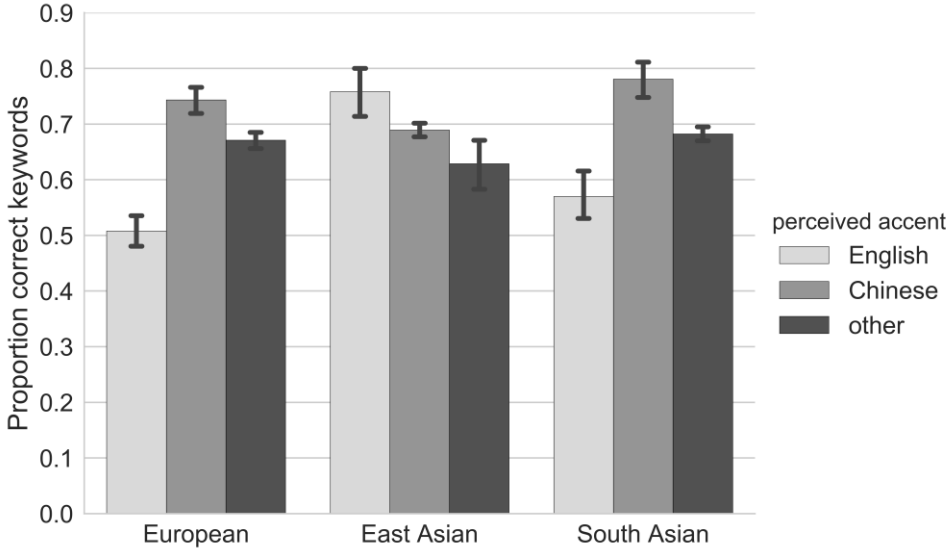

Figure 2. Mean proportion correctly transcribed keywords in experiment 2, by purported talker ethnicity and perceived accent.