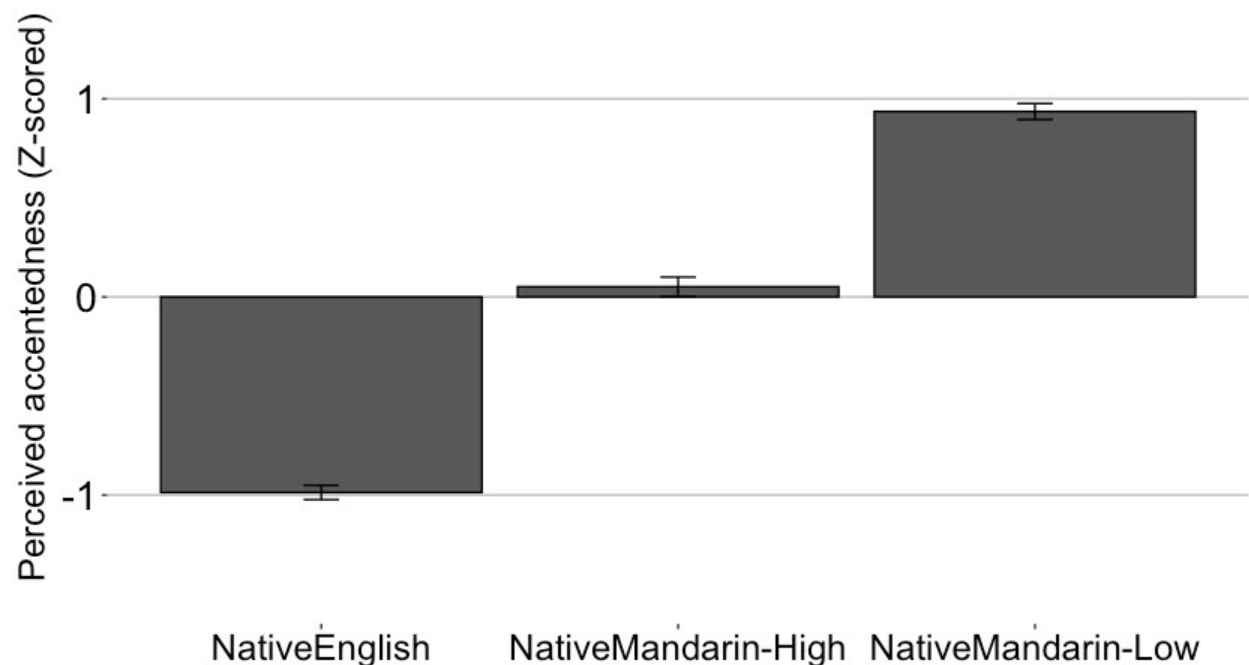


### *Talkers' perceived foreign accentedness*

In order to characterize non-native (native Mandarin) talkers' English proficiency, the recordings of the practice sentences were evaluated for foreign accentedness by native English listeners. The sentence-length files of the practice sentences were RMS normalized to 65dB SPL. Forty native English listeners (13 females, 27 males; age range = 23 - 67 years, mean = 35.8), recruited using Amazon Mechanical Turk ([www.mturk.com](http://www.mturk.com)), participated in the foreign accent rating task evaluating the accentedness of the talkers who recorded the materials. None of the listeners reported a history of speech or hearing impairment. None of the listeners reported experience with Mandarin Chinese. In the accent rating task, conducted via Qualtrics ([www.qualtrics.com/](http://www.qualtrics.com/)), the listeners were told that they would listen to English sentences and evaluate the foreign accent of the speech. In each trial, listeners heard an English sentence without noise and were instructed to rate the accentedness of the speech on a scale of 1 ("a native speaker of English") through 9 ("an extremely strong foreign accent"; similar to Munro & Derwing, 1995). In order to prevent the accentedness ratings from being influenced by the intelligibility of the speech, the transcript of the sentence was displayed while the listeners were evaluating the speech (Gittleman & Van Engen, 2018). Each sentence could not be played more than once, but there was no time limit for responding. Twenty listeners evaluated 6 talkers (i.e., 2 Native English, 2 Native Mandarin-High talkers, 2 Native Mandarin-Low talkers) and another set of 20 listeners evaluated the other 6 talkers. Thus, each listener evaluated 60 sentences (i.e., 10 unique sentences x 6 talkers). The presentation of the sentences was randomized for each listener.

Foreign accent ratings were z-score normalized for each listener in order to account for variation in the listeners' use of the nine-point rating scale. Figure 1 shows accent ratings by talker group (Native English, Native Mandarin-High, and Native Mandarin-Low). In order to

examine whether the accent ratings differed for different talker groups, one-way ANOVA was carried out with z-scored ratings as the dependent variable. The results indicated that the ratings differed significantly by the talker group [ $F(2, 9) = 128.29, p < .001, \eta^2_p = .97$ ]. The post-hoc Tukey comparisons confirmed that all the group comparisons were significant: Native English vs. Native Mandarin-High, Native English vs. Native Mandarin-Low, and Native Mandarin-High vs. Native Mandarin-Low ( $p < .0001$  for all). These results demonstrated that there was a clear difference in the perceived accentedness of the talkers. That is, Native English talkers were perceived to be less accented than Native Mandarin talkers, and Native Mandarin-High talkers were perceived to be less accented than Native Mandarin-Low talkers.



**Figure 1.** Z-score normalized accentedness ratings plotted by talker group. Error bars represent 95% confidence interval of the mean.