Voice Preference in German: A Cross-linguistic Comparison of Native and Chinese Listeners

Hongwei Ding[1], Rüdiger Hoffmann[2] and Oliver Jokisch[3]

[1]School of Foreign Languages, Shanghai Jiao Tong University, China
[2]Chair for System Theory and Speech Technology, TU Dresden, Germany
[3]Institute of Communications Engineering, HfT Leipzig, Germany
hwding@sjtu.edu.cn, ruediger.hoffmann@tu-dresden.de, jokisch@hft-leipzig.de

In this study we investigated whether native and non-native speakers have a similar preference in the selection of a pleasant voice, and how about the correlation of their preference rankings with the voice quality and the prosodic features. 50 Chinese without prior knowledge of German and 10 native German listeners participated in a pair comparison test to choose the best voice out of eight candidate speakers for a German speech synthesis. A mixed model was used to handle the unbalanced numbers of listeners in statistics. Pitch change patterns or melody metrics were compared with Momel algorithm (Hirst, 2015), and voice quality was analyzed in terms of phonation types. The test results showed that the ranking scores of German and Chinese listeners were highly correlated. A further investigation revealed that the decisions of German listeners were highly or moderately correlated with the voice quality only, while those of the Chinese listeners were highly or moderately correlated with pitch movements and speech rate as well. It can be concluded that both German and Chinese listeners prefer a breathy voice. Besides, Chinese listeners also vote for speakers who exhibit faster and larger falling pitch movements. The findings suggest that speakers of tone and non-tone languages may rely on similar voice quality cues but different prosodic cues in the selection of their preferred voice.

Previous studies have shown that the average pitch values, prosody and voice quality are supposed to be the main features which can influence the attractiveness of a voice (Xu, 2013). Mean F0 values are mainly determined by speaker's anatomy and physiology, which are related to body size; while the utilisation of prosody and voice quality may be language-dependent. While few researches have been devoted to the investigation of acoustic correlates of voice preference from a cross-linguistic perspective (Hain, 2009), this study endeavours to provide some preliminary understanding in this field. Furthermore, it is also interesting to compare the correlates between listeners of a tone and a non-tone language. We chose German and Chinese listeners to vote for their preferred German voice. Since the Chinese listeners had no previous knowledge of German, they mainly relied on the paralinguistic and non-linguistic information for their selections.

Literature:

[1] Hain, H.-U.; Jokisch, O. & Coelho, L.: "Multilingual voice analysis: Towards prosodic correlates of voice preference". In Proc. Konferenz Elektronische Sprachsignalverarbeitung (ESSV) 2009, ser. Studientexte zur Sprachkommunikation, R. Hoffmann, Ed., vol. 53, Dresden, Germany, September 2009, pp. 215–221.

[2] Hirst, D. & Ding, H.: "Using melody metrics to compare English speech read by native speakers and by L2 Chinese speakers from Shanghai". In Interspeech, 2015, pp. 1942–1946.

[3] Xu, Y.; Lee, A.; Wu, W.-L.; Liu, X. & Birkholz, P.: "Human Vocal Attractiveness as Signalled by Body Size Projection". In: PLOS ONE, 2013. DOI: 10.1371/journal.pone.0062397