Perception and production of Mandarin /y-u/ contrast by Scottish English speakers

Jonathan Havenhill and Jane Stuart-Smith

Acquisition of /y-u/ contrast is known to be challenging for Anglophones [1]. In many English varieties, /u/ exhibits fronted [y/ʉ] allophones, yielding an F2 range spanning (nearly) the entire high vowel space: [y] to [u̟ɬ]. The articulatory-acoustic quality of fronted /u/ varies, however. Southern British and American varieties realize fronted /u/ with a front-central tongue position, protruded lips, and diphthongization [2,3], while Scottish English /u/ is monophthongal [ʉ] with a lower, retracted tongue position and less protruded lips [4]. Such variation may promote distinct L2 learning outcomes as L1 and L2 systems are postulated to inhabit the same phonetic and phonological space [5,6]. Cross-dialectal differences in the articulation of English /u/ may influence the production of non-native /y/, yet most prior work on L2 phonology relies on acoustic data, so it is not entirely known whether transfer operates on articulatory and/or acoustic targets [5–7]. This study examines perception and production of Mandarin /i-y-u/ by native speakers of Scottish English. More specifically, the study aims to ascertain the extent to which Scottish /ʉ/ productions are transferred to non-native /y/ and /u/.

Twenty-five Scottish English native speakers (11M/12F/2NB), including 5 Mandarin learners, each completed three production (L1 elicitation, L1 & L2 imitation) and two perception (AXB discrimination, categorization/goodness) tasks. Production materials comprised 101 English words containing /i ɪ e ɛ ʉ o ɔ ʌ a aɪ aʊ ɔɪ/ and /jʉ ji wi wʉ jaʊ/, as well as 87 Mandarin words containing /i y u ɤ a/ and /ʐ ʐ̩ ə̹ ei ou ie ye iou uei uo ai au ia iau ua uai/. All Mandarin items were open syllables, while some English items included /t l r/ codas. All items were monosyllabic, with coronal and non-coronal onsets. Pseudorandomly presented items were repeated 3 times in isolation. High-speed ultrasound (81 fps), lip video (60 fps), and audio were simultaneously recorded with AAA. Articulatory data were automatically labeled with DeepLabCut and formant trajectories were extracted with FastTrack. Articulatory and acoustic trajectories for Mandarin /y u/ (relative to English /ʉ/) were modeled using ordered factor reference-difference GAMM smooths with random smooths for speaker and word.

Production results from 20 Mandarin-naïve listener-imitators are given in Fig. 1. Difference smooths show that the F2 of non-native /y/ does not differ significantly from native /ʉ/ ($p = 0.289$), suggesting a cross-linguistic mapping of Scottish /ʉ/ to Mandarin /y/. On the other hand, the F2 of Mandarin /u/ is significantly lower than L1 /ʉ/ ($\beta = –0.46$, $t = –8.31$, $p < 0.001$), indicating a novel production strategy for high back /u/, which has no direct equivalent in Scottish English. This acoustic difference may be driven by lip rounding; lip aperture for non-native /u/ is significantly smaller than that of native /ʉ/ ($\beta = –0.46$, $t = –5.14$, $p < 0.001$), whereas rounding for non-native /y/ is L1-like ($p = 0.409$). Tongue body anteriority for non-native /u/ is likewise significantly retracted relative to L1 /ʉ/ ($\beta = –0.37$, $t = –17.23$, $p < 0.001$), while /y/ does not differ ($p = 0.449$). Individual speakers show substantial variability in the degree to which they distinguish non-native /y-u/ by differences in tongue position and shape. Thus far, these results from Scottish English appear to differ from prior work on American English learners of Cantonese, where listener-imitators produced variants closer to monophthongal L2 /y/ than to diphthongized L1 /u/. Taken together, this suggests that transfer in learner phonology may be influenced by native dialect variation, and hence that more L1 cross-dialectal comparisons are needed to ascertain the mechanisms underlying L1 > L2 transfer.

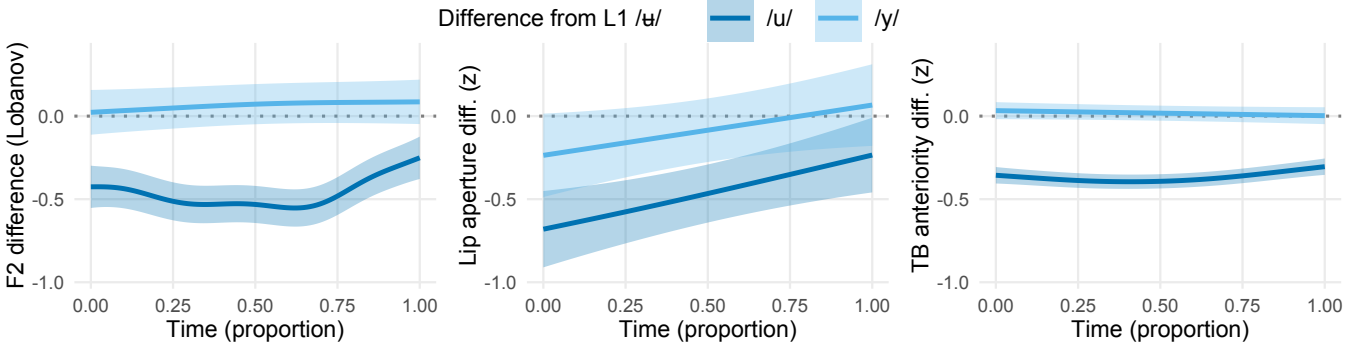

Fig. 1: GAMM difference smooths for Mandarin /y/ and /u/ relative to Scottish English /ʉ/. Left: F2. Center: lip aperture. Right: tongue body anteriority.

[1] E. S. Levy & F. F. Law. “Production of French vowels by American-English learners of French”. *JASA* 128.3 (2010). **[2]** J. Harrington, F. Kleber & U. Reubold. “The contributions of the lips and the tongue to the diachronic fronting of high back vowels in Standard Southern British English”. *JIPA* 41.2 (2012). **[3]** J. Havenhill. “Articulatory and acoustic dynamics of fronted back vowels in American English”. *JASA* 155.4 (2024). **[4]** E. Lawson, J. Stuart-Smith & L. Rodger. “A comparison of acoustic and articulatory parameters for the GOOSE vowel across British Isles Englishes”. *JASA* 146.6 (2019). **[5]** J. E. Flege & O.-S. Bohn. “The Revised Speech Learning Model (SLM-r)”. *Second Language Speech Learning*. Ed. by R. Wayland. Cambridge: Cambridge UP, 2021. **[6]** C. T. Best & M. D. Tyler. “Nonnative and second-language speech perception: Commonalities and complementarities”. *Language experience in second language speech learning: In honor of James Emil Flege*. Ed. by O.-S. Bohn & M. J. Munro. Phil., PA: John Benjamins, 2007. **[7]** M. Oakley. “The Articulatory and Acoustic Representation of Second-Language French Vowels”. *Lang. Speech* (2024).