Gender-specific behaviour in vowel articulation

Patrycja Strycharczuk[1], Sam Kirkham[2], Emily Gorman[2], Takayuki Nagamine[2], and Adrian Leemann[3]
[1]University og Manchester (UK), [2]Lancaster University (UK), [3]Bern University (Switzerland)

Gender is a known important source of variation in the acoustic quality of vowels, but the nature of its influence remains under debate. Some gender differences likely originate from sex-based differences in the size and shape of the vocal tract. However, based on previous literature [1, 2], we may expect that there are cases when speakers of different genders systematically diverge in their vocal tract movement to produce vowel sounds. Our study evidences such a case, documenting gender-specific articulatory strategies for producing lax vowel contrasts in Northern Anglo-English.

We collected synchronised acoustic and midsagittal ultrasound data from 36 speakers of English from the North of England (23 F and 13 M speakers; mean age = 24.5, *SD* = 8). The speakers produced 8-10 repetitions of five English lax vowels in a fixed segmental environment (*bid, bed, bad, bod, bud*), in two prosodic contexts elicited via different carrier phrases. DeepLabCut, a markerless pose estimation algorithm was applied to the data to identify 11 landmarks on the surface of the tongue, and the short tendon attached to the mandible [3]. The positions of these landmarks were rotated on the occlusal plane, exported as x-y Cartesian coordinates and scaled within speaker. Horizontal displacement of the tongue root was used to capture lingual retraction / fronting, and vertical position of the tongue body was used to capture the raising / lowering of the tongue.

The left panel of Fig. 1 shows by-gender mean values for the tongue retraction and raising in the five vowels. Non-high vowels, *bed, bad and bod*, appear to be shifted clockwise in female speakers, although the shift is only significant for *bed* and *bad*, based on lmer modelling. This shift is in line with F1 raising and F2 lowering in females, such that males and females become more distinct along the F1 dimension, but more alike along the F2 dimension, as far as raw Hertz values are concerned (middle panel of Fig. 1). Interestingly, some of the observed differences are also apparent after formant normalisation is applied. The right panel of Fig. 1 shows *z*-scored formant values for the same vowels, indicating some acoustic retraction and lowering in *bad* and *bed* for female speakers, which we found to be statistically significant.

The presence of an acoustic difference in normalised formant space raises the possibility that we are dealing with an ongoing vowel shift led by female speakers. In order to evaluate this hypothesis, we investigated acoustic differences in male and female DRESS and TRAP vowels, using corpus data from 266 speakers (145 F, 121 M) from the same geographical area, with a richer age span (mean age=33, *SD* = 15). The data were extracted from the English Dialect App Corpus [4] In the EDAC data (Fig. 2), we replicated the statistical difference between male and female normalised F1 in both DRESS and TRAP (but not the difference in normalised F2). However, we did not find any effect of apparent time: the observed gender differences persist across speakers of different ages, consistent with a diachronically stable gender difference in the articulatory / acoustic vowel target.

We propose that the differences we observe between male and female speech are related to, though not necessarily directly caused by biological constraints on jaw-opening, as reported by [2]. We investigate jaw movement in our speakers with a further set of analyses, by tracking dynamic displacement of the short tendon. The results are consistent with increased jaw opening in females, especially in DRESS and TRAP and they suggest a pattern of jaw-tongue coordination, in which jaw opening correlates with tongue root retraction [5]. We argue that relatively increased jaw opening is a pervasive pattern for low vowels in female speech, and that it leads to increased tongue root retraction in clear speech.

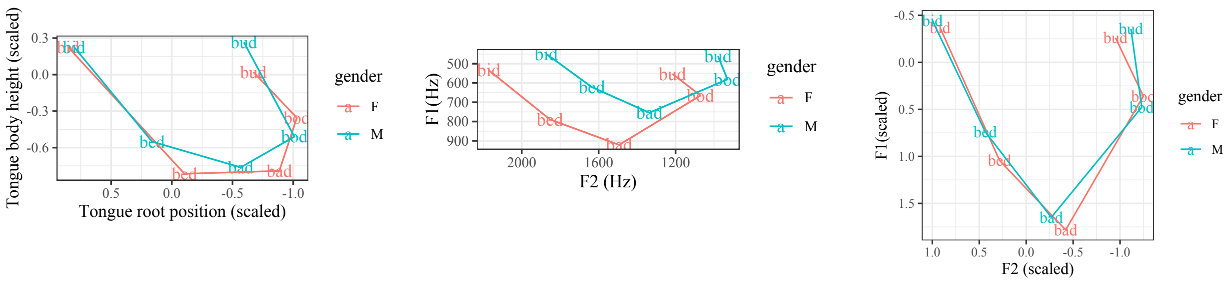

Fig. 1. The lax vowel subsystem conditioned by gender in three sets of dimensions. Left: lingual articulation (scaled); Middle: formant values (in raw Hertz), Right: formant values (scaled)

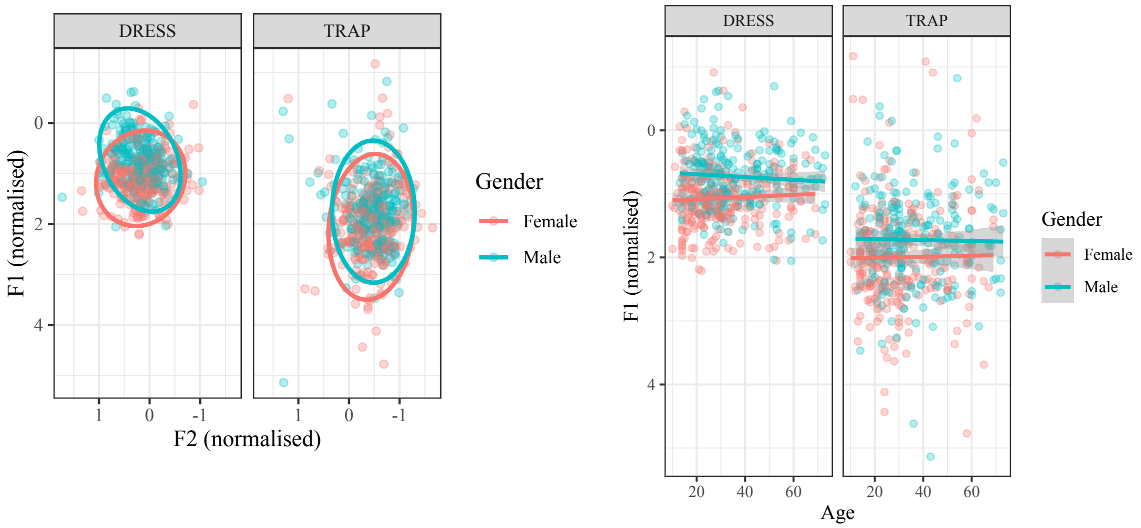

Fig. 2. EDAC data. Left: Male vs. female normalised formant values for DRESS and TRAP. Right: normalised F1 for DRESS and TRAP by age and gender

References

[1] Fant, B. G. (1966). A note on vocal tract size factors and non-uniform F-pattern scalings". In: *Speech Transmission Laboratory Quarterly Progress and Status Report* 1.

[2] Weirich, M., Fuchs, S., Simpson, A., Winkler, R., & Perrier, P. (2016). Mumbling: Macho or morphology?. *Journal of Speech, Language, and Hearing Research*, *59*, S1587-S1595.

[3] Wrench, A., & Balch-Tomes, J. (2022). Beyond the edge: markerless pose estimation of speech articulators from ultrasound and camera images using DeepLabCut. *Sensors*, *22*, 1133.

[4] Leemann, A., Kolly, M. J., & Britain, D. (2018). The English Dialects App: The creation of a crowdsourced dialect corpus. *Ampersand*, *5*, 1-17.

[5] Johnson, K., Ladefoged, P., & Lindau, M. (1993). Individual differences in vowel production. *The Journal of the Acoustical Society of America*, *94*, 701-714.