ASHA American Speech-Language-Hearing Association

JSLHR

Research Article

Articulatory Strategies in Male and Female Vowel Production

Patrycja Strycharczuk[a] and Sam Kirkham[b]

[a] Department of Linguistics and English Language, The University of Manchester, United Kingdom [b] Department of Linguistics and English Language, Lancaster University, United Kingdom

ARTICLE INFO

Article History:
Received March 11, 2025
Revision received June 25, 2025
Accepted July 21, 2025

Editor-in-Chief: Jessica E. Huber
Editor: Jun Wang

https://doi.org/10.1044/2025_JSLHR-25-00185

ABSTRACT

Purpose: This study investigates the articulatory strategies used by male and female speakers to produce vowel sounds.
Method: Secondary data analysis of a preexisting articulatory corpus was carried out. Dynamic midsagittal ultrasound and acoustic data from 36 speakers of Northern English (21 females and 15 males) were analyzed, representing 17 vowel phonemes in a controlled phonetic environment. Articulatory landmarks corresponding to the tongue root, dorsum, and mandibular short tendon were automatically labeled in the ultrasound image, and their dynamic displacement was analyzed using generalized additive mixed modeling. Dynamic formant trajectories for the first formant (F1) and the second formant (F2) were analyzed using the same method.
Results: Significant articulatory differences were found between male and female speakers for several vowels. Increased tongue dorsum fronting and lowering were found for female GOOSE and GOAT vowels. Greater jaw opening was found in female TRAP, START, SQUARE, DRESS, MOUTH, and LOT, accompanied by greater dorsal retraction, compared with male speakers. For STRUT/FOOT, there was greater retraction of the tongue dorsum in males. For some vowels, for example, TRAP and DRESS, corresponding differences were detected in normalized formant trajectories, but the magnitude of the acoustic differences was typically very small, and in some cases, such as MOUTH and LOT, no differences in normalized F1 or F2 were detected, despite underlying articulatory differences.
Conclusions: Many of the differences we find point to increased jaw opening and greater involvement of the jaw as an articulator in female speakers. This wider strategy affects the production of multiple vowels, but it only manifests acoustically in some cases, suggesting the role of generalization mechanisms. Clinical implications for gender-affirming speech therapy are discussed.

There is a well-established empirical observation that female vowel resonances are, on average, higher than male resonances, but the magnitudes of the mean resonance differences are not uniform across vowels. The nonuniformity was first reported by Fant (1966) for American English and Swedish. Fant noted that male–female differences in the first formant (F1) are relatively greater in open vowels and relatively smaller in rounded back vowels. Since female F1 is relatively greater for open vowels, which have high F1 values, female acoustic vowel spaces are typically overall larger, compared with male vowel spaces. A similar pattern has since been replicated in a range of languages, including Danish, Dutch, multiple varieties of English, Estonian, French, German, Hungarian, Icelandic, Italian, Japanese, Korean, Norwegian, and Polish (Fant, 1975; Henton, 1995; Johnson, 2006). The observed differences concern cisgender females and cisgender males. While this is rarely explicitly acknowledged, existing studies typically use either sex or gender as the relevant conditioning variable, and similar results are obtained, as would be expected when the speaker samples are primarily or exclusively cisgender. We assume a distinction between "sex" and "gender," in which "sex" is aligned with biological factors, whereas "gender" is

Correspondence to Patrycja Strycharczuk: patrycja.strycharczuk@manchester.ac.uk. *__Disclosure:__ The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

associated with behavioral, social, and cultural factors (Munson & Babel, 2019). In our summary of previous findings below, we follow the original study in using the terms "sex" or "gender," but effectively, the synthesized picture of available evidence represents cisgender males and cisgender females.

There has been much discussion on the relative contribution of sex and gender in conditioning nonuniformity of vowel formant differences. Sexual dimorphism is associated with systematic differences in vocal tract anatomy, which could, in principle, produce some acoustic differences, even assuming no underlying difference in speaker behavior. Fant (1966) explores this line of explanation, proposing that relatively lower F1 values in low vowels for male speakers are primarily due to the proportionally longer pharynx and proportionally larger laryngeal cavity. Traunmüller (1984) also advances an explanation that focuses on the properties of the vocal tract as a filter, proposing that male speakers develop their articulatory strategies for vowel production in childhood and continue to rely on the same strategies despite changes in the proportions of the vocal tract following larynx descent during puberty. However, it is now well established that these types of anatomical explanations are incomplete at best.

Limitations of anatomical accounts become apparent when we consider the magnitude of gender differences in formant values, as well as variation concerning said differences. Early vocal tract modeling studies showed that observed differences between male and female formants are greater than those predicted by differences in vocal tract anatomy and that anatomical factors alone fail to produce nonuniform scaling patterns (Goldstein, 1980; Nordström, 1977). Another issue is that developmental patterns of vowel formant changes through childhood and puberty to adulthood do not align well with changes in vocal tract size (Lee et al., 1999; Vorperian & Kent, 2007; Whiteside, 2001). Furthermore, average male–female differences in vowel formants are language specific, as shown by Johnson (2006). Johnson observes that this cross-linguistic variation is not predicted by differences between male and female vocal tract lengths, as approximated by speaker height. This suggests that speakers target language- and gender-specific resonances to express their gender within a particular sociocultural context. A crucial role of speaker control in producing gender differences in vowel resonances is underscored by the findings that such differences are socially salient. Larger vowel spaces are associated with a perception of femininity, and conversely, smaller vowel spaces index masculinity (Heffernan, 2010; Leung et al., 2018; Weirich & Simpson, 2018a). Furthermore, Weirich and Simpson (2018a) show that F1 lowering in low vowels (a masculine trait) is more prominent in male speakers with a stronger perception of own masculinity. This suggests that vowel-specific formant values can be actively targeted in speech production to express gender-related traits. This finding aligns with the proposal by Johnson (2006) that speakers can manipulate their speech to a fine degree in performing gender.

The currently available evidence suggests that speakers actively produce specific formant values in order to sound more or less masculine or feminine, but the target values themselves may be partially shaped by both anatomical and social factors (Munson & Babel, 2019; Simpson, 2009). This pattern of variation can be understood as a result of probabilistic sound changes, whereby specific biases can be combined and amplified to a different degree in different communities. The presence of some systematic bias is supported by the cross-linguistic regularity of nonuniform differences in male and female formant values. A similar pattern has been found in multiple languages, and it is not reversed in any language that we know of. This suggests the involvement of anatomical factors, although it has to be acknowledged that the bulk of the available evidence comes from Indo-European languages, spoken in communities that may share cultural gender norms. Potential anatomical biases that might be relevant here include the diverging resonating properties of the male and female vocal tracts, as discussed above, as well as other factors. Weirich et al. (2016) propose that differences in the anatomy of the vocal tract may trigger a preference for a different articulatory strategy in males and females. Using vocal tract models based on features extracted from magnetic resonance imaging (MRI), they show that the same degree of jaw opening may lead to linguopharyngeal closure in a male vocal tract but not in a female vocal tract. This factor may limit the degree of jaw opening in male speakers as confirmed by the articulatory data from American English and German presented in the same article (Weirich et al., 2016). Social factors favoring nonuniform formant differences associated with gender have also been proposed, in particular the tendency of female speakers to produce clear speech, linked to traditional gender roles involving females as primary caregivers. Larger vowel spaces in female speakers are associated with greater acoustic contrasts between vowel phoneme categories, and therefore, the contrasts are thought to be easier to perceive. This explanation is proposed by Goldstein (1980) and also acknowledged by Traunmüller (1984). Several other properties of female speech align with the general characteristic of female speech being relatively clearer, such as slower speech rate and relative avoidance of phonetic reduction (Byrd, 1994).

Much of the discussion in the literature has concerned the question of why speakers produce gender-specific vowel resonances. In contrast, the question of how they produce such difference has received a lot less attention. All the behavioral accounts summarized above lead to a common prediction that there is an underlying difference in articulatory behavior between male and female speakers. If nonuniform vowel formant differences are controlled by the speaker, as opposed to being a passive consequence of vocal tract size and shape, then it follows that male and female speakers must diverge somehow in their vocal tract movement to produce the relevant differences for different vowels. This is sometimes taken for granted, and often, articulatory interpretations are offered to describe observed acoustic differences. For example, Labov (1990) uses the term "closed-mouthed" to refer to male speech. Leung et al. (2018) discuss male–female differences in vowel space size in terms of differences in articulation and mention wider jaw opening as a strategy for producing more feminine-sounding speech. However, these types of interpretations are typically inferred from acoustics rather than directly evidenced by articulatory data. Only a handful of studies to date have analyzed the effect of gender on articulation. This is a major gap because the relationship between articulation and acoustics is not linear, even though it is characterized by some regularities (Fant, 1971; Stevens & House, 1955). The quantal theory of speech production predicts that articulatory and acoustic variation may not be comparable in all cases, such that acoustic stability may, in some cases, conceal underlying articulatory variation (Stevens, 1989). These predictions are in line with the findings that speakers may pursue different articulatory strategies in their production of the same vowel phonemes (Johnson, 2023; Johnson et al., 1993).

Weirich and Simpson (2018b) is one of the few articulatory studies on male–female differences in speech, reporting articulometry data from 40 speakers of American English and nine speakers of German. The American English data in the study came from the X-ray microbeam corpus by Westbury et al. (1998). The data analysis focused on low vowels—/oʊ/ in American English and /aː/ in German—and found a greater angle of jaw opening in female speakers compared with males in both languages. Serrurier and Neuschaefer-Rube (2024) also report gender-related differences in articulatory strategies based on MRI data from French, German, and English (41 speakers in total). The key differences are related to the vector of tongue movement used to achieve variation in the second formant (F2). Relatively greater vertical displacement is associated with F2 manipulation in female speakers, whereas male speakers tend to rely more on horizontal tongue displacement. Otherwise, several studies analyze gender-specific differences in the articulation of diphthongs. Simpson (2001) analyzes gender differences in the realization of /aɪ/ in the X-ray microbeam corpus and reports greater sensor displacement and higher velocities in male speaker production. Simpson (2002) analyzes the vowel sequence in the phrase "they all," taken from the same corpus. Similarly to the /aɪ/ data, greater sensor displacement and higher velocity were observed for male speakers. The difference in sensor displacement and velocity is potentially attributable to vocal tract size: As male vocal tracts are larger, the tongue needs to traverse a greater physical distance moving between a low-back to a high-front constriction. However, a different view of gender comparisons emerges once the size of the vocal tract is normalized. Weirich and Simpson (2018b) report gender-conditioned differences in the length of the diphthong trajectory in German /aɪ/, such that females show larger normalized displacements of the tongue body in the diphthong production. Additionally, their study analyzed the effect of prominence: The target word was under focus (accented condition), post-focal (unaccented condition), or read out from a list (control condition). The difference between genders was observed in unaccented and control conditions, but there was no difference in the accented condition. The authors interpret this result as increased undershoot in males: Females attain relatively more extreme articulatory positions in the control condition, but males produce some articulatory reduction.

While the existing articulatory studies document some sex- and gender-specific vowel production strategies, multiple gaps remain in the overall empirical picture. One of the unresolved questions is how general the documented strategies are. Most previous studies focus on a few selected vowels, and the data they relied on do not allow for systematic comparisons of gender effects across different vowels, since the segmental and prosodic environment was not controlled for. However, it is important to consider the role of gender not only in the production of individual vowels but also in the production of vowel contrasts. This is motivated by previous explanations of gender (and sex) differences in vowel formants, some of which focus on specific vowels (Fant, 1966; Traunmüller, 1984; Weirich et al., 2016), while others are inherently more systemic, stressing factors such as overall clarity of speech (Goldstein, 1980).

The present study aims to expand our empirical understanding of the role of gender in vowel production by comparing male–female differences in the articulation of all possible vowel phonemes (excluding unstressed vowels) in Northern Anglo-English. The data represent a dialectally coherent sample of speakers producing vowels in a stable segmental and prosodic environment. We analyze ultrasound data from 36 speakers of Northern English in order to identify potential differences in tongue and jaw movement involved in male and female vowel production. In addition, we analyze the acoustic data from the same speakers in order to understand whether the observed articulatory and acoustic differences correspond to one another and also to relate our findings to previous acoustic results.

Method

Data

The data in our study come from the TarDiS corpus (Kirkham et al., 2023; Strycharczuk et al., 2024, 2025). The corpus comprises time-synchronized acoustic, ultrasound, and (for a small subset of speakers) articulometry data from 40 speakers of Northern Anglo-English. We focus on the ultrasound and acoustic data. Head stabilization (UltraFit; Spreafico et al., 2018) was used for recording the ultrasound data, and the speaker's occlusal plane was recorded to standardize the rotation of ultrasound images. For more details on the recording procedure, see Kirkham et al. (2023) and Strycharczuk et al. (2025).

Gender information is available for 39 speakers in the corpus, including 23 female, 15 male, and one nonbinary speaker. No information about the speakers' sex is available. The gender information was provided by the participants in response to an open question (i.e., no response categories were given). The data from the nonbinary speaker were not included, as no group-level generalization could be made based on one speaker. Furthermore, data from two speakers were excluded because of audible rhoticity in some words, which is a recessive feature in present-day Lancashire English (Turton & Lennon, 2023). This left 36 speakers who were included in the final analysis, of whom 21 were female and 15 were male. The mean age of the female speakers was 21.5 (range: 18–41) years, whereas the mean age of the male speakers was 30 (range: 18–48) years.

The corpus includes items representing most vowel phonemes in Northern English in a controlled segmental and prosodic environment. In the current study, we focus on the vowels in the "b_d" context. These vowels are listed in Table 1, along with their corresponding lexical sets, as defined by Wells (1982). Note that speakers included in our study were nonrhotic. The test items were embedded in two types of carrier phrases: *She says X* and *She says X eagerly*. Typically, four to six repetitions are available for each item in each prosodic context (corresponding to carrier phrase) as pronounced by each speaker. The total number of tokens included in the analysis was 5,643.

The study was exempt from ethical approval, as it constitutes secondary data analysis. The data collection for the original corpus received ethical approval, and the participants provided informed consent, as reported in Strycharczuk et al. (2024, 2025).

Table 1. Test items used in the study and the lexical sets to which they belong, representing typical phoneme systems in the North of England.

Item	Lexical set	Phonetic transcription
bad	TRAP	a
bade	FACE	eɪ
bard	START	ɑː
bared	SQUARE	ɛː
bead	FLEECE	iː
beard	NEAR	ɪə/ɪː
bed	DRESS	ɛ
bid	KIT	ɪ
bide	PRICE	aɪ
bird	NURSE	ɜː
bod	LOT	ɒ
bode	GOAT	əʊ
booed	GOOSE	ʉː
bored	NORTH	ɔː
bowed	MOUTH	aʊ
bud	STRUT	ʊ
buoyed	CHOICE	oɪ

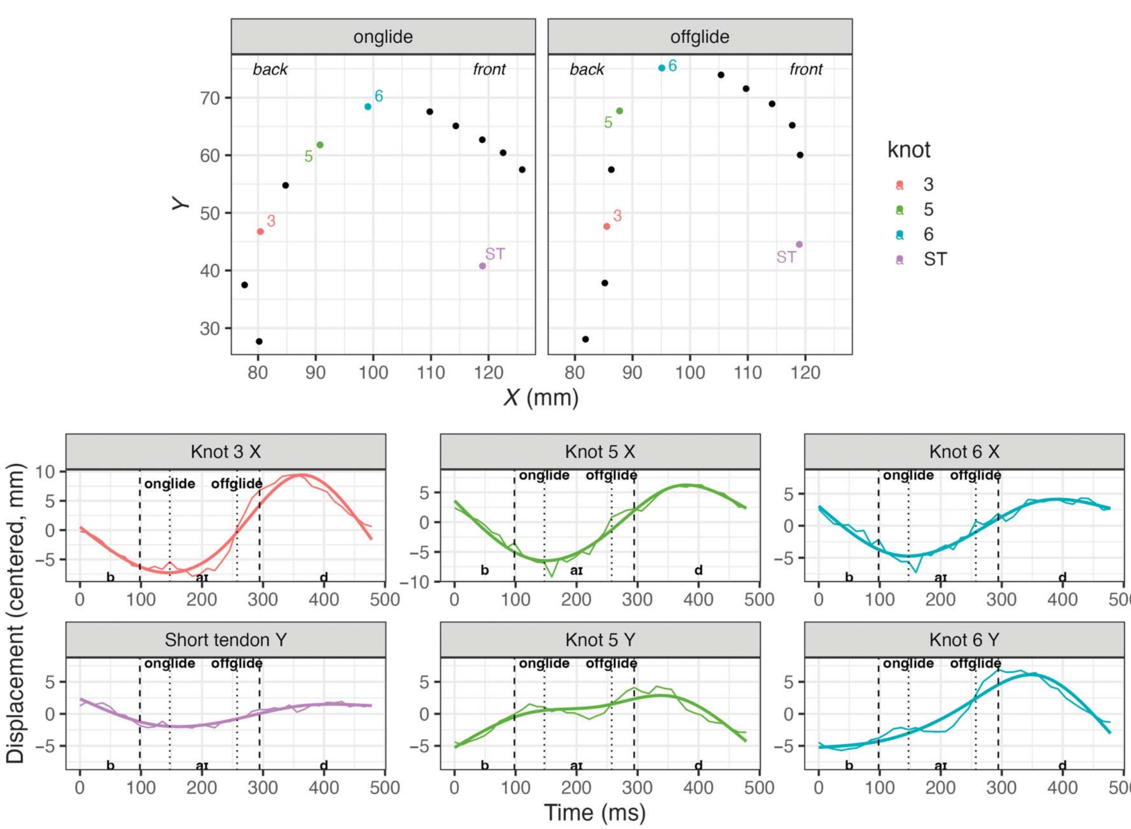
Note. The transcription represents the most common phonetic quality of each vowel in our data.

Data Processing

The ultrasound data were sampled at 59.5–101 (median = 81.3) frames per second and rotated to the speaker's occlusal plane. For the dynamic articulatory analysis, the entire word was included, from the acoustic onset of the initial /b/ to the offset of closure for the final /d/. Additionally, we labeled vowel boundaries: The vowel onset was taken to correspond to the offset of burst for the preceding /b/, whereas the vowel offset was placed at the onset of closure for the final /d/. The tongue contour visible in the ultrasound image was tracked automatically using DeepLabCut (DLC; Mathis et al., 2018; Wrench & Balch-Tomes, 2022), as implemented in Articulate Assistance Advanced Version 2.20 (Articulate Instruments Ltd., 2014). The procedure identifies 11 consistent points on the tongue contour, from the vallecula to the tongue tip. Strycharczuk et al. (2025) demonstrate that most information about vowel contrasts can be reduced to the displacement of several key knots, Knot 3, representing the tongue root; Knot 5, representing the posterior part of the tongue dorsum and typically the constriction point for back vowels; and Knot 6, representing the mid part of the tongue dorsum, and, typically, the DLC knot representative of constriction in nonback vowels. In our analysis, we focus on the dynamic displacement of these knots, as they represent the movement of key articulators involved in dorsal vowel production. Figure 1 shows the location of the 11 DLC knots on the tongue surface for an example token of *bide* taken at two time points: 25% and 80% into the vowel (these time points approximate the diphthong onglide and offglide). In addition to the 11 knots on the tongue contour, DLC also tracks the position of the hyoid and that of the mandibular short tendon (Wrench & Balch-Tomes, 2022). The displacement of the short tendon is controlled by the jaw, and we therefore analyze the short tendon displacement as a proxy for jaw movement.

The bottom panel of Figure 1 shows the dynamic displacement of the key knots for the same example token of *bide*. The displacement is measured in millimeters. To improve figure legibility, we centered each displacement value, such that 0 corresponds to the mean displacement value for each particular knot in either plane (horizontal *X* or vertical *Y*). The figure shows both unsmoothed and smoothed displacement trajectories (the smoothing here is for illustration only; the data were in fact smoothed with generalized additive mixed modeling [GAMM], as described in the Statistical Analysis section). As we can see from the figure, the dynamic displacement of key DLC knots patterns largely as expected given the articulation of /baɪd/. The tongue root (Knot 3) initially retracts until it reaches its most extreme posterior position, at around 150 ms (ca. 25% into the vowel). The tongue root then moves forward until it reaches its most anterior extremum at around 350 ms, at which point the tongue root returns to its mean position. The overall temporal and spatial patterns of tongue dorsum fronting are similar when looking at the fronting of Knots 5 and 6, although the range of movement is smaller than for the tongue root. In terms of dorsum raising, this token/speaker shows a pattern of steady raising of the posterior part of the tongue dorsum (Knot 5), followed by lowering once the full closure for /d/ is achieved (at around 375 ms). The mid part of the tongue dorsum (Knot 6) also shows a pattern of gradual raising until maximum raising during /d/ closure, although the raising trajectory is different from that in Knot 5. The short tendon shows a pattern of initial lowering in the diphthong onglide, followed by raising. The vertical displacement of the short tendon is more limited in spatial terms compared with the other articulators. The pattern

Figure 1. (Top) Location of DeepLabCut knots for an example token of *bide*, representing diphthong onglide and offglide. Key knots analyzed in the study are highlighted. (Bottom) Dynamic displacement of the key knots in the same token of *bide*. Horizontal displacement (*X*) and vertical displacement (*Y*) are plotted separately. Both unsmoothed and smoothed displacements are shown.



of lowering and raising reflects the expected movement of jaw opening and closing, but it is likely that the displacement of the short tendon is overall more limited than the displacement of the jaw.

The displacement values of the key DLC knots were extracted and centered within the speaker; that is, for each displacement trajectory within each speaker, we calculated the mean value and subtracted the mean from each measurement. No further spatial normalization was done, as the exploratory analysis suggested that the mean male and female displacement values were comparable. We discuss this further in the Discussion section. The time domain was linearly normalized within phoneme. The vowel onset was taken and marked as 0, and the vowel offset was marked as 1. The offset of final /d/ was marked as 2, and the onset of initial /b/ was marked as −1. This normalization procedure is centered on the vowel, but it includes flanking consonants to obtain and compare a wide range of articulatory displacement.

The acoustic analysis focused on the vocalic portion of the word only. Formant measurements were extracted dynamically using FastTrack (Barreda, 2021), a Praat add-on that optimizes linear predictive coding analysis settings for each speaker. This analysis was implemented in Praat Version 6.2.14 (Boersma & Weenink, 2009). The maximum formant range was set to 4500–6500 Hz for male speakers and 5000–7000 Hz for female speakers. The remaining settings were as follows: number of steps = 20, coefficients for formant = 5, number of formants = 3, number of bins = 5, and statistic = median. The formant measurements were sampled at every 2 ms (equivalent to 500 Hz) and subsequently downsampled to match the temporal resolution of the ultrasound data.

The formant values were normalized using the ΔF method (Johnson, 2020). This method normalizes for speaker vocal tract by estimating its length, based on formant spacing. The specific ΔF method we used employs a simplified equation from Lammert and Narayanan (2015), as reproduced in Equation (1).

$$\Delta F = 0.56^{*}\ F1 + 0.20666^{*}\ F2 + 0.188^{*}\ F3 \qquad F_{norm} = F/\Delta F \tag{1}$$

The advantage of the ΔF method is that it takes vocal tract length into account, thus eliminating some potential anatomical aspects of gender differences that might arise passively from the mean differences in vocal tract length correlated with gender. According to Johnson (2020), the vowel spaces resulting from ΔF normalization are similar to those obtained by log-mean vowel normalization (Nearey, 1989; Nearey & Assmann, 1986). Both approaches tend to preserve some aspects of nonuniformity in male–female formant differences, especially the relatively higher F1 in female low vowels. This has been argued to be more representative of perceptual normalization (Barreda, 2021). The normalized F1 and F2 trajectories representing the entire vowel were submitted to statistical analysis, as described in the Statistical Analysis section below.

Statistical Analysis

The data were analyzed dynamically using GAMM (Wood, 2017) with maximum likelihood (ML) estimation. Models were run individually for each of the dependent variables. These were the articulatory variables exemplified in Figure 1, as well as normalized F1 and F2. The models were run separately for each vowel. Each model had the following predictor structure:

- Main effect of gender,
- Smooth term for normalized time,
- Smooth term for normalized time by gender,
- By-speaker random intercept
- A random smooth for normalized time by each token

Random smooths were included to reduce the likelihood of Type I error (Soskuthy, 2021). Preferably, we would have run the models across all items, testing for an interaction between vowel and gender. However, we have found that, given the number of vowel categories (17), fitting such a model would exceed the computational resources available to us. Given these limitations, we had to choose between leaving out random smooths and running multiple by-vowel models. We chose the latter approach, because reducing the random part of the model produced a much larger range of significant results, suggesting that models without random smooths are less conservative. We are mindful of the multiple testing issue, and we focus on those effects that are consistent across a range of vowels.

Significance was established using ML comparison. For each model, we fitted a corresponding model from which the gender effects were removed. Gender was considered to have a significant effect if including gender predictors led to a significant increase in ML at $\alpha = .05$. We did not investigate systematically whether the gender differences were linear or not (i.e., whether the by-gender smooth for normalized time was significant on its own). In cases where there was a significant effect of gender, we identified the significant time intervals using difference curves as implemented in the plotdiff() function in the itsadug package (van Rij et al., 2015). Due to the large number of models fitted, we do not present the individual model results in detail, instead we focus on visual representation of model predictions and significance testing. Individual model summaries are reported in the online data repository at https://osf.io/zcet3/.

Results

Dynamic Articulatory Results

Figure 2 shows the prediction of by-vowel generalized additive mixed models of tongue root advancement/retraction, as captured by the horizontal displacement of DLC Knot 3. The shading around the mean curves represents 95% confidence intervals. Gray-shaded areas represent time intervals where the difference between males and females is significant (only for models with a significant difference in ML comparisons). The values were centered, such that 0 corresponds to the mean position of Knot 3 (within each gender). Displacement is measured in millimeters. As we can see from the figure, the mean tongue root displacement trajectories are typically comparable between males and females, and typically, no significant differences are observed between genders at the start of /b/ or at the end of /d/. Focusing on the differences in the vowels, females have a significantly more retracted tongue root during the vowel in *bared*, and for a small portion of *bed*, compared with males. In *booed*, we find the opposite: Males have a more retracted tongue position, compared with females.

Figure 3 shows the prediction of GAMM of tongue dorsum advancement/retraction, as captured by the horizontal displacement of DLC Knot 5. This DLC

Figure 2. Generalized additive mixed modeling predictions for horizontal tongue root displacement (Knot 3) in normalized time, as a function of item and gender. F = female; M = male.

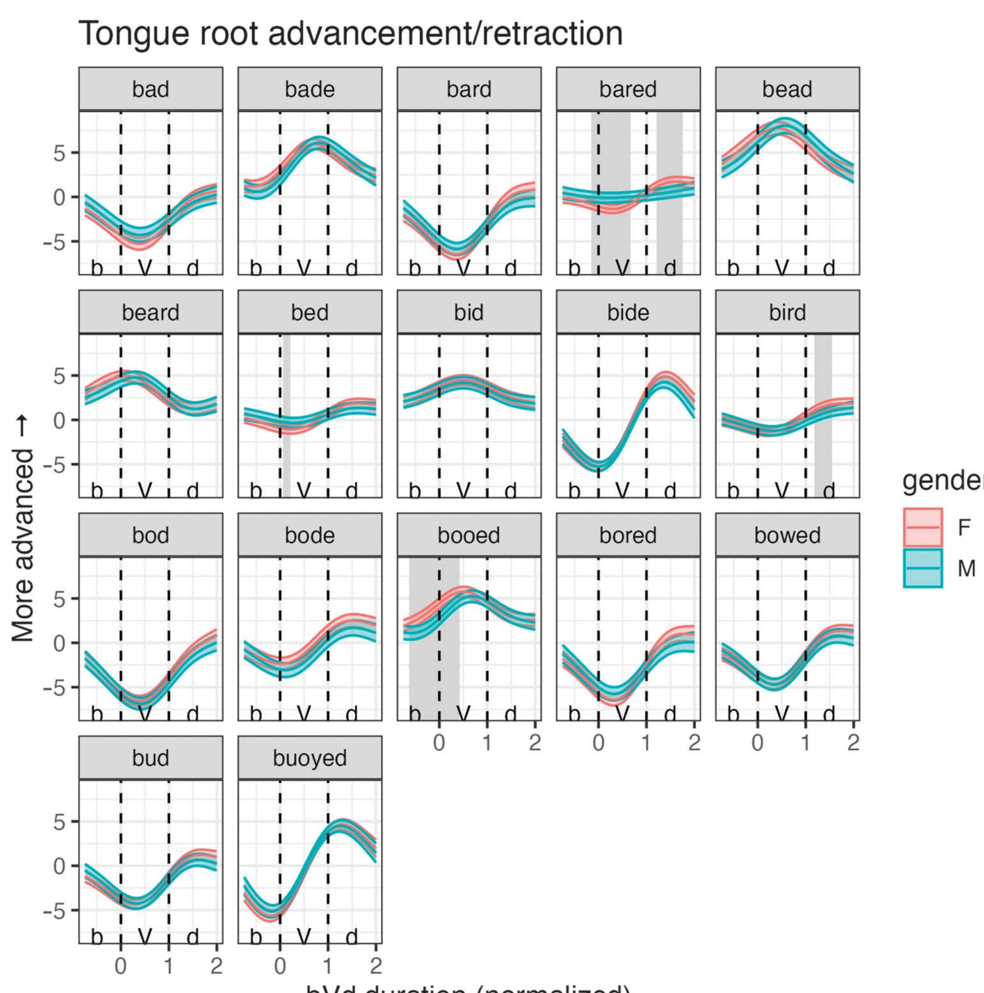
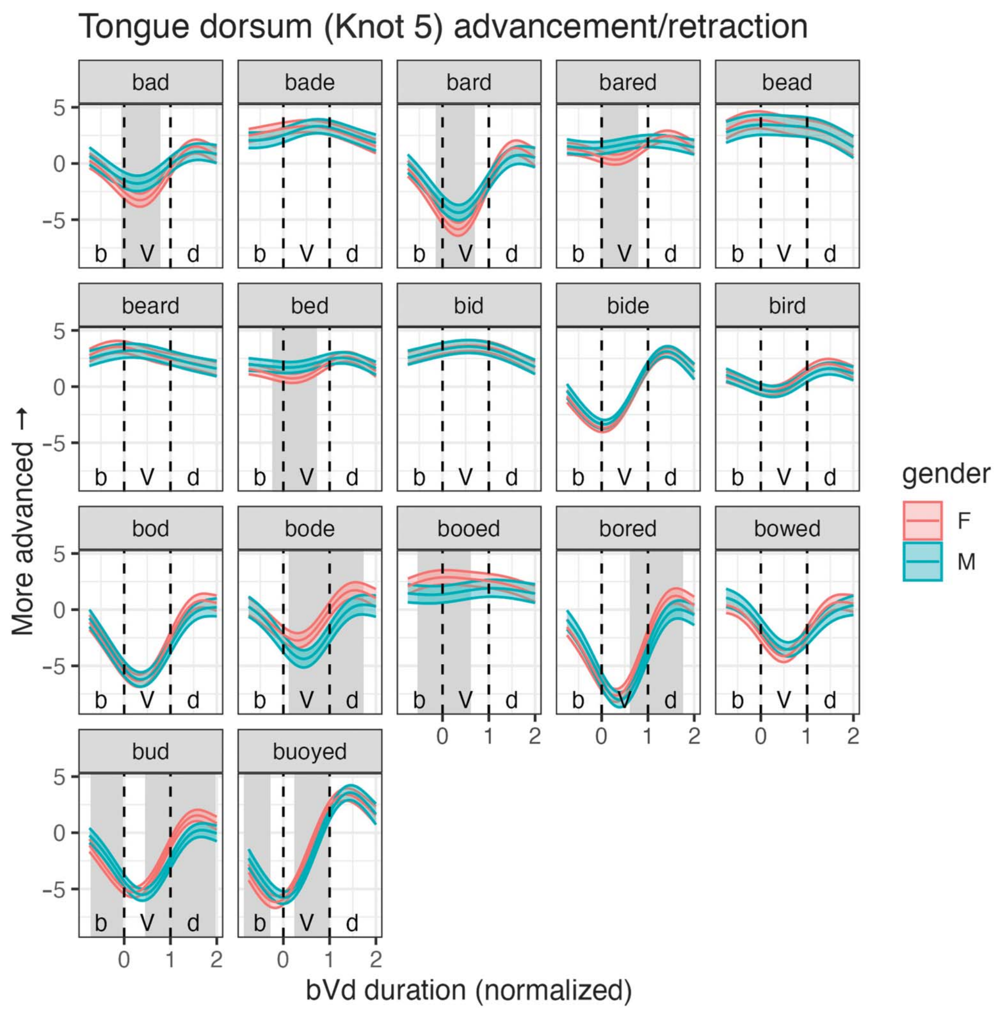


Figure 3. Generalized additive mixed modeling predictions for horizontal displacement of the posterior part of the tongue dorsum (Knot 5) in normalized time, as a function of item and gender. F = female; M = male; V = vowel.

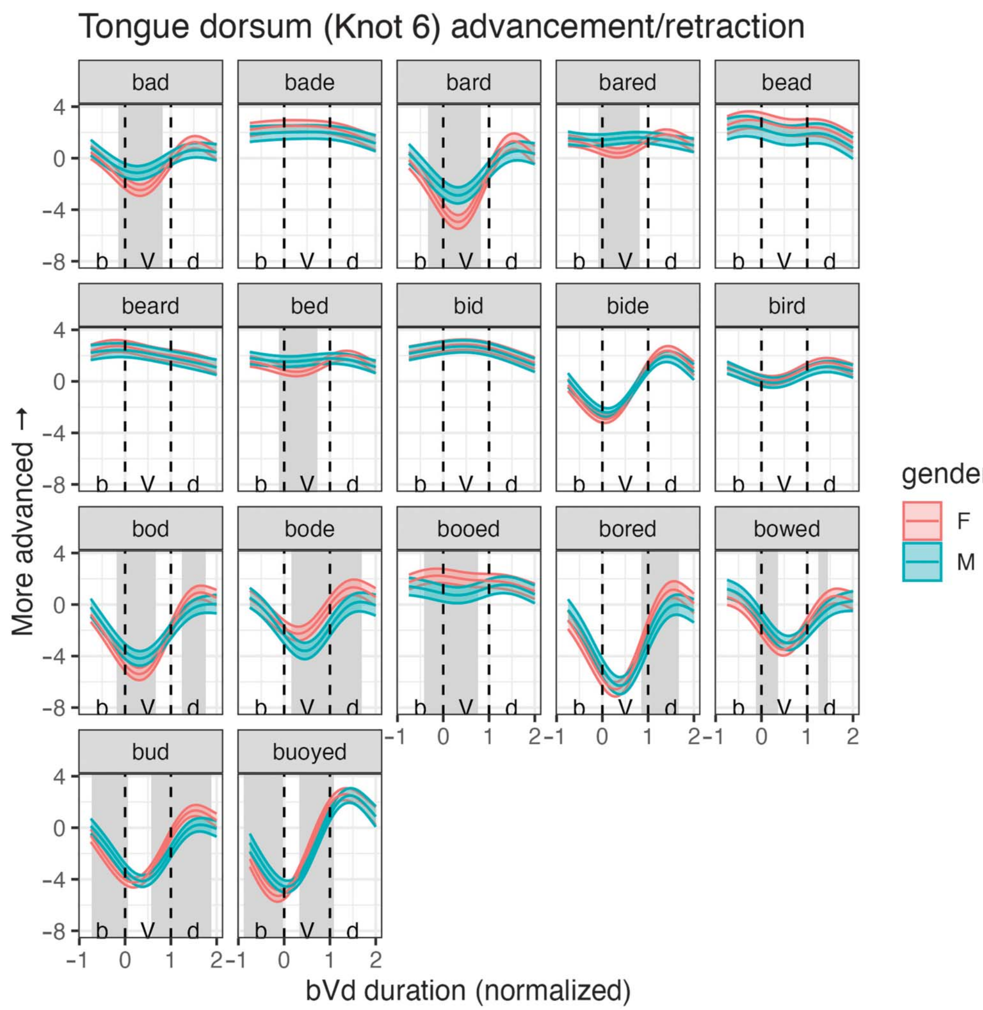


knot is typically representative of the location of dorsal constriction for back vowels (Strycharczuk et al., 2025). Vowels in *bared* and *bed*, which showed greater tongue root retraction for female speakers compared with male speakers, also show greater tongue dorsum retraction at Knot 5 for female speakers. Additionally, vowels in *bad* and *bard* also show significantly greater tongue dorsum retraction in female speakers. Five vowels display the opposite pattern, that is, greater tongue root retraction in males compared with females during the vowel. These are *bode*, *booed*, *bored*, *bud,* and *buoyed*. Out of these, *bode* and *booed* show a difference in the same direction for the tongue root, while the others do not.

Very similar results emerge when looking at the horizontal displacement of Knot 6 (see Figure 4), which represents the mid part of the tongue dorsum and typically captures the constriction location for mid and front vowels (although some front vowels may have a more anterior constriction). All the gender-based differences affecting the position of Knot 5 also affect the position of Knot 6: This part of the tongue dorsum is more retracted for females in *bad*, *bard*, *bared,* and *bed* and more retracted for males in *bode*, *booed*, *bored*, *bud,* and *buoyed*. Additional differences emerge in the tongue dorsum fronting in the onglide of *bowed* and in *bod*, which patterns with *bad*, *bard,* and *bed*, such that females show more retraction, compared with males.

In order to examine gender differences within dorsum height, we analyzed vertical displacement of DLC Knots 5 and 6. In line with other measurements, the displacement values were centralized within the speaker, separately for each knot. Figure 5 shows the GAMM predictions for the vertical displacement of Knot 5, depending on vowel and gender. There is significant lowering of the

Figure 4. Generalized additive mixed modeling predictions for horizontal displacement of the mid part of the tongue dorsum (Knot 6) in normalized time, as a function of item and gender. F = female; M = male; V = vowel.



tongue dorsum in *bard*, *bide,* and *bod* for males. In *bode*, the opposite pattern is found: There is more raising in males compared with females.

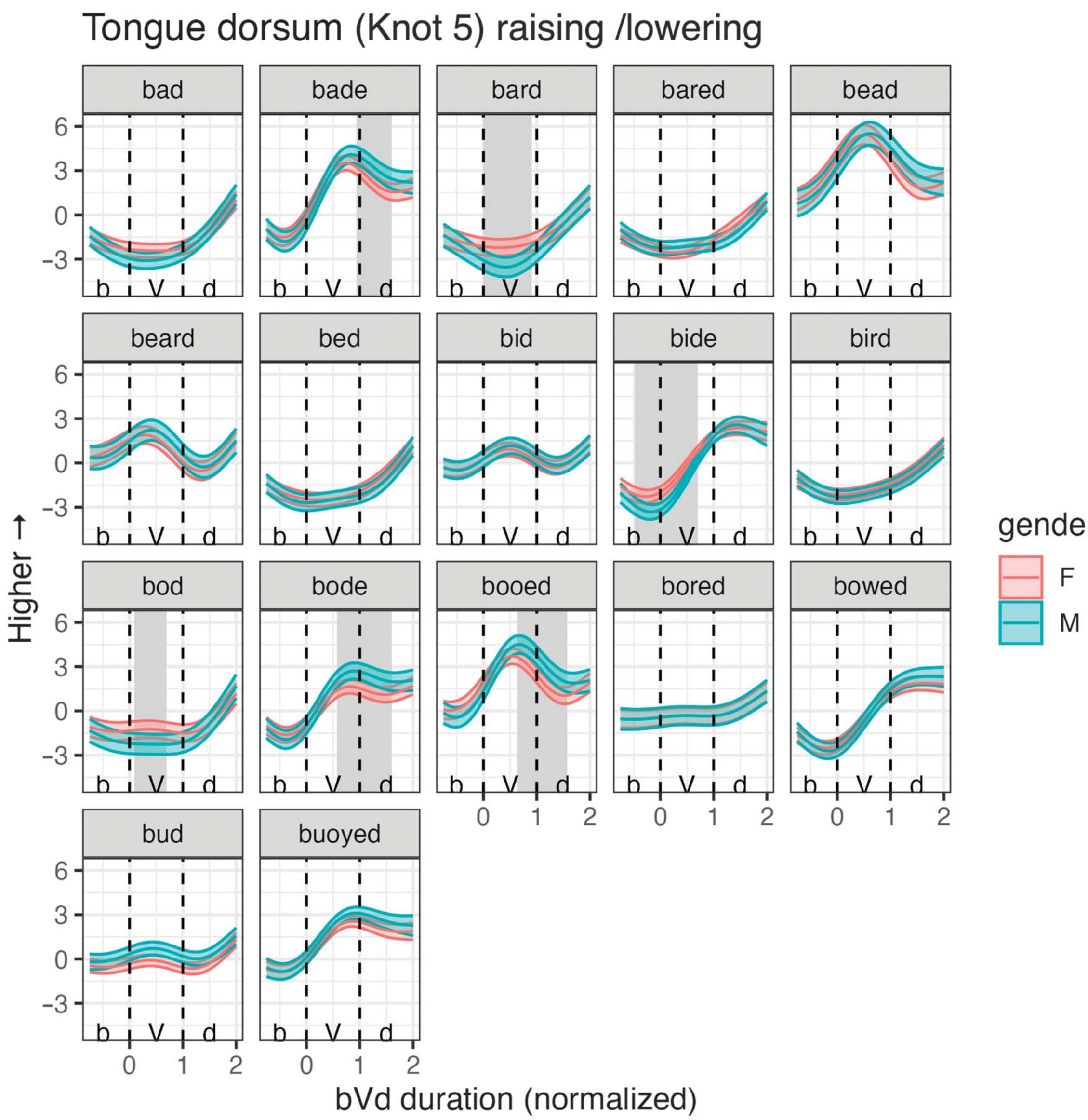
The vertical displacement of Knot 6 (mid part of the tongue dorsum) patterns in a similar way, as illustrated in Figure 6. Once again, we see significant lowering of the tongue dorsum at Knot 6 in *bard*, *bide,* and *bod* for females compared with males. Increased dorsum raising in males compared with females is found in *bead* and the offglide of *booed.*

As a proxy for jaw opening and raising, we analyzed the vertical displacement of the short tendon, as tracked in the ultrasound image by the DLC algorithm. This measure was modeled using GAMM, analogously to the tongue-related measures. The model predictions for short tendon raising and lowering are visualized in Figure 7. We find significantly more short tendon lowering in female *bad*, *bard*, *bed*, *bared*, *bid*, *bod*, *bade,* and *bowed*, which suggests that female speakers produce greater jaw opening in these vowels, compared with males. The opposite pattern is found in *booed* and *bode*, where there is more short tendon lowering (greater jaw opening) in males compared with females.

Dynamic Acoustic Results

In the acoustic part of the analysis, we focus on two measures: F1 and F2. These measures were normalized for vocal tract length using the F method (Johnson, 2020) and analyzed dynamically with GAMM. Figure 8 illustrates model predictions for F1 trajectories, depending on vowel and gender. F1 is significantly greater for females in *bared*, *bed*, *bird*, *bad*, and *bide*, although the magnitude of the difference varies, depending on the vowel. We find the opposite pattern, that is, greater F1 in males compared with females for *bead*, *booed,* and *beard*; however, these

Figure 5. Generalized additive mixed modeling predictions for vertical displacement of the posterior part of the tongue dorsum (Knot 5) in normalized time, as a function of item and gender. F = female; M = male; V = vowel.



differences only affect a small portion of the formant transitions. We remain cautious about interpreting these differences.

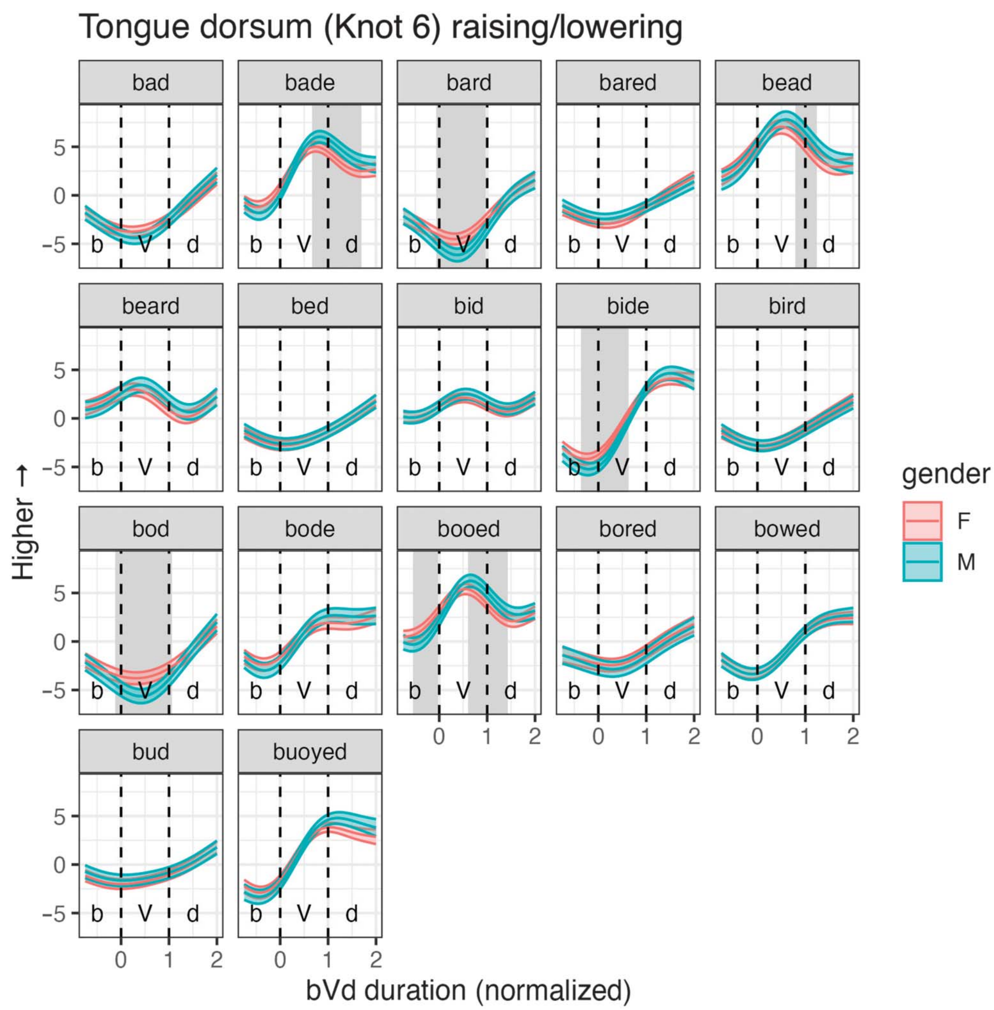
When it comes to F2, as shown in Figure 9, we find significantly lower normalized F2 for female *bared*, *bed*, *bad*, and *bard*, as well as the onglide of *bored*. F2 is significantly greater in females for *bode*, *bead*, and *buoyed*.

Summary of the Results

Based on the analysis presented above, there are multiple differences between males and females, both in the dynamic displacement of key articulators and in normalized F1 and F2. Table 2 presents a summary of the significant differences we have found. We have excluded differences that only affect C1 or C2 (/b/ or /d/) in the articulation or acoustic differences limited to a narrow portion of CV transitions. We use males as a reference level in the table. This is done for brevity, and it should not be taken to mean that the male articulation is seen to be the default one.

We have ordered the vowels in Table 2 in such a way that vowels representing a similar pattern are grouped together. We can identify three groups of vowels that

Figure 6. Generalized additive mixed modeling predictions for vertical displacement of the posterior part of the tongue dorsum (Knot 6) in normalized time, as a function of item and gender. F = female; M = male; V = vowel.



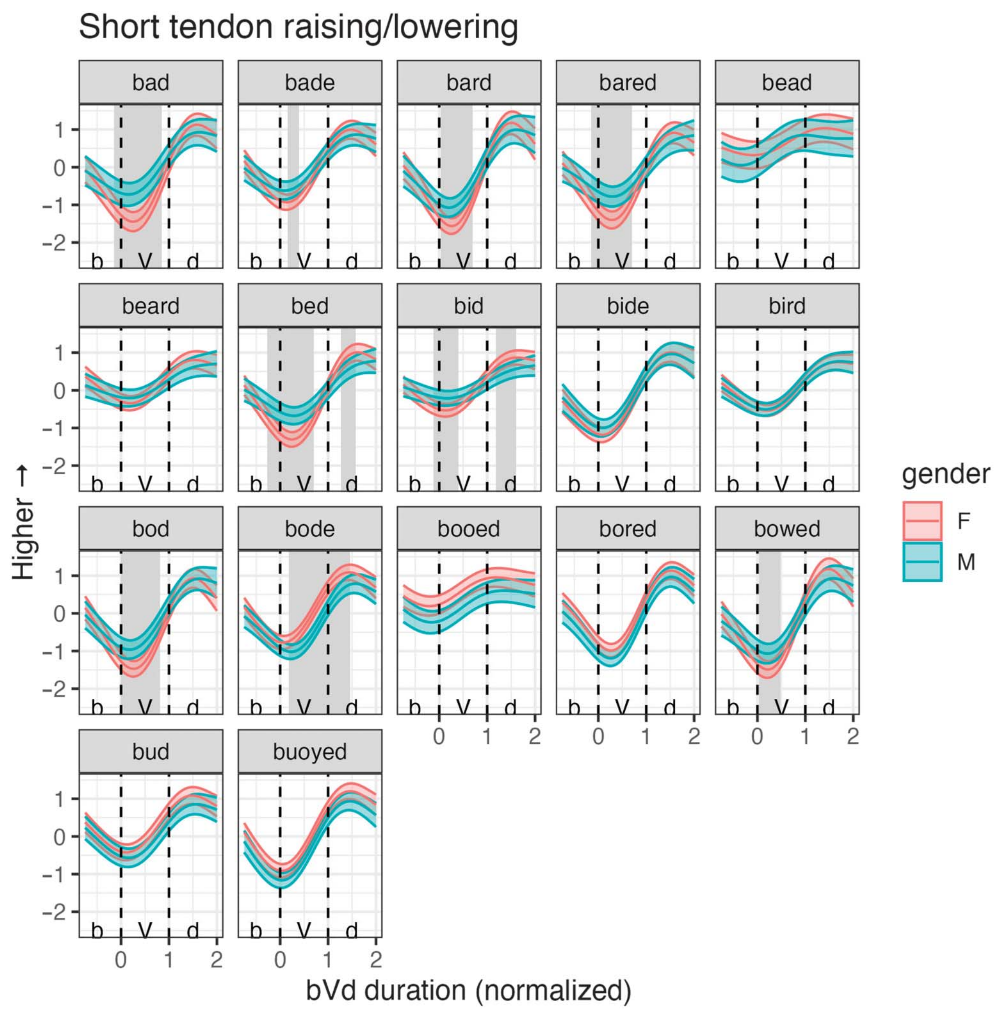
share several articulatory and acoustic characteristics and that show systematic gender differences.[1]

The first group that we can distinguish comprises the vowels GOOSE and GOAT (*booed* and *bode*). GOOSE and GOAT pattern together sociolinguistically, as both vowels are known to participate in the same sound change: Fronting of the GOOSE vowel frequently triggers subsequent fronting of GOAT (Labov, 1994, p. 208). These two vowels show systematic gender differences such that the tongue dorsum is more advanced in females and more retracted in males. Additionally, they both show some dorsum lowering in females compared with males, and *bode* also has more short tendon lowering in males compared with females. With respect to acoustics, *bode* shows relatively higher F2 in females and lower F2 in males, in line with the articulatory differences in tongue root and tongue dorsum fronting.

The second group is composed of low front and mid vowels, including a diphthong with a low onglide. The vowels in this group are *bad*, *bed*, *bard*, *bared*, and *bowed*. In terms of gender differences, these vowels typically show greater short tendon lowering, suggesting greater jaw opening in female speakers—this is the case for all the vowels in this group. All the vowels show tongue dorsum retraction in females compared with males when looking at the mid part of the tongue dorsum (Knot 6) and typically also when looking at the posterior part (Knot 5). Additionally, in *bared* and *bed,* there is a significant

[1]We did not attempt to systematically relate the effect of articulation on acoustics as a function of gender due to complexity of articulation–acoustics relationships and associated statistical complexity. See Strycharczuk et al. (2025) for some discussion on the relationship between vowel articulation and the first two formants informed by the same corpus.

Figure 7. Generalized additive mixed modeling predictions for vertical displacement of the short tendon, as a function of item and gender. F = female; M = male; V = vowel.

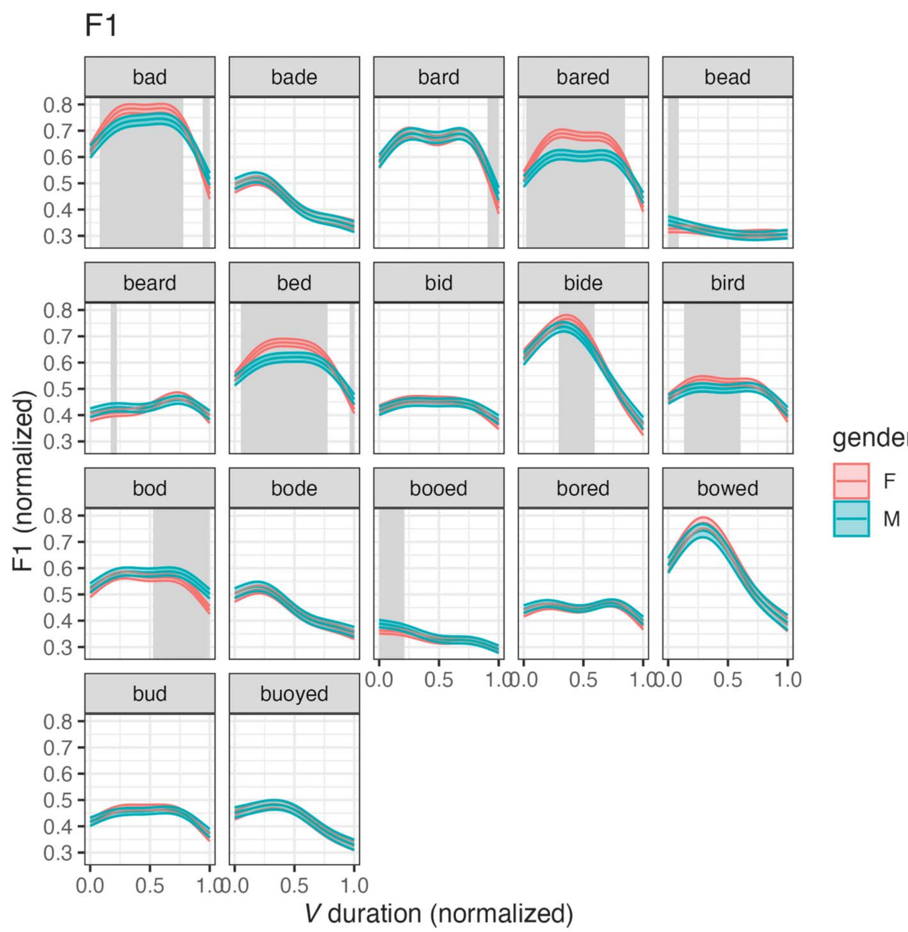


difference in the position of the tongue root, such that the tongue root is more retracted in females and more advanced in males. The gender differences are less pronounced when it comes to tongue dorsum height. The tongue dorsum is higher for female speakers in *bard* but not in other vowels. These articulatory differences generally have some acoustic effect that persists despite vocal tract normalization. We see higher normalized F1 in females for *bad*, *bared*, and *bed*. Furthermore, all vowels in this group, except *bowed*, have lower normalized F2 in females compared with males.

The third group we can discern includes the vowels in *bod* and *bud*. The rationale for grouping them together is that both these vowels belong in the lax vowel subsystem and both are back vowels. In terms of gender differences, females have more short tendon lowering in *bod*, and some significant differences between males and females can be observed for *bod* at Knot 6: This part of the dorsum is higher and more retracted for females. For the vowel *bud*, the posterior part of the tongue dorsum is significantly more advanced in females. These articulatory differences do not correspond to any systematic acoustic differences: Normalized F1 and F2 are not significantly different between males and females in either *bod* or *bud*.

In order to understand the gender difference affecting articulation of *bod* and *bud*, it is helpful to consider the overall tongue contours, which complements our understanding of the kinematic data that focused on the displacement of individual articulators. Figure 10 illustrates mean tongue contours for females and males for the five lax vowels; it includes *bod* and *bud*, as well as the remaining lax vowels for reference. These tongue contours represent the acoustic midpoint of the vowel, and they were rotated to the occlusal plane and centered within the

Figure 8. Generalized additive mixed modeling predictions for F1 change over vowel duration, as a function of item and gender. F = female; M = male; V = vowel.



speaker. In this case, 0 corresponds to the mean value of the vowel space across the 11 DLC knots. The by-gender, by-item means were obtained using multivariate GAMMs (Coretta & Sakr, 2024). The main takeaway from this figure is that the male *bud* is higher relative to other vowels: It is close in dorsum height to *bid*. In comparison, the female *bud* is lower. The posterior part of the dorsum is very similar for female *bud* and *bod*, whereas in the anterior part, *bud* is only slightly higher than *bod*.

Figure 11 focuses more closely on the comparison of tongue shape in male and female *bod* and *bud*. In this case, the data were centered, scaled, and rotated to the same plane using Procrustes analysis, using the procedure in the study by Dryden and Mardia (2016). This normalization approach obliterates some differences in overall tongue position and tongue height, aligning the different tongue contours as closely as possible. However, some differences persist despite normalization, and these are related to tongue shape. As we can see from the figure, the mean tongue shapes for *bod* and *bud* are very similar in female speakers. In contrast, these two vowels are clearly differentiated by tongue shape in males.

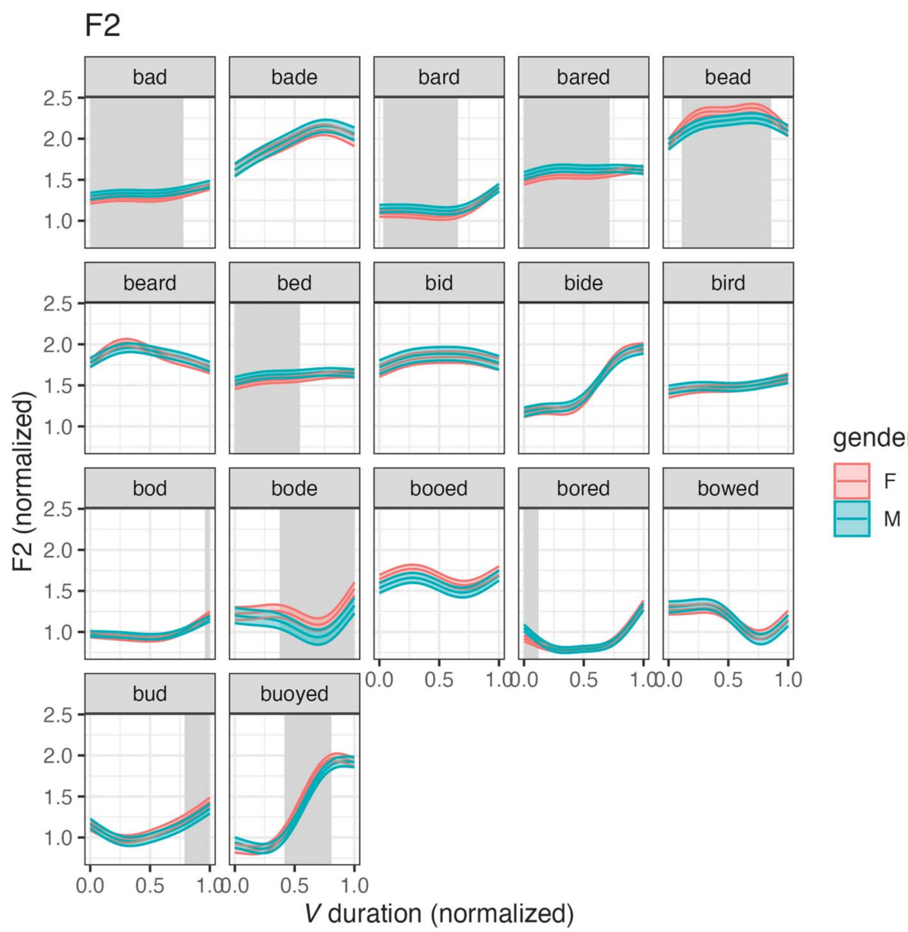
In the Discussion section below, we discuss the possible explanations for the patterns of gender differences affecting the three groups of vowels, as identified above. Additionally, we have observed some differences in several other vowels. For example, we observe some short tendon lowering in female *bid* (a high–mid vowel), which has no reflection in F1 or F2. These types of differences, summarized in Table 2, do not form any coherent pattern that we can identify, so we shall refrain from interpreting them.

Discussion

Gender-Specific Articulatory Strategies in Different Vowels

Our data deliver evidence of multiple differences between male and female strategies for producing vowel

Figure 9. Generalized additive mixed modeling predictions for F2 change over vowel duration, as a function of item and gender. F = female; M = male; V = vowel.



sounds. While not all vowels involve gender differences in production, many of them do. Moreover, the difference we observe is systematic, such that groups of vowels that share specific articulatory properties pattern together with respect to gender differences in production.

GOOSE and GOAT

The first group of gender differences we identify concerns the vowels in *booed* and *bode*. In *booed* and *bode*, we find greater tongue dorsum advancement in female speakers, and in *bode,* we find higher normalized F2 for females. These differences are likely related to GOOSE and GOAT fronting, which are sound changes in progress. Specifically, the differences we find are consistent with more fronting in female speakers. GOOSE fronting has been well documented in Northern English, in both the acoustic and articulatory domains (Lawson et al., 2019; Strycharczuk et al., 2020). However, comparing with recent data on Southern English, as reported in the study by Cole and Strycharczuk (2024), GOOSE fronting is somewhat more limited in the North and presumably still ongoing. Therefore, it is not surprising to find a gender difference whereby females are more advanced in the change, since it is known that females tend to lead sound change (Labov, 2001, p. 321), which manifests phonetically as a greater degree of advancement in ongoing changes. The fronting of the GOAT vowel (*bode*) has also been noted in the North (Haddican et al., 2013), although once again, it is considerably more limited compared with the South of England (Cole & Strycharczuk, 2024; Strycharczuk et al., 2020). Therefore, the gender difference we find here is likely a manifestation of a female-led sound change.

Mid and Front Low Vowels

Second, females typically produce mid and front low vowels (*bad*, *bed*, *bard*, *bared*, and *bowed*) with more jaw opening, as evidenced by the more extreme displacement of the short tendon. Our results align with those of Weirich et al. (2016), who find more jaw opening in female

Table 2. Summary of significant gender differences for different measures, depending on vowel.

Vowel	Vowel group	Short tendon	Tongue root	Knot 5 X	Knot 5 Y	Knot 6 X	Knot 6 Y	F1	F2
booed	GOOSE and GOAT		Advanced in F	Advanced in F	Lower in F	Advanced in F	Lower in F		
bode		Higher in F		Advanced in F	Lower in F	Advanced in F			Higher in F
bad	Front mid low vowels	Lower in F		Retracted in F		Retracted in F		Higher in F	Lower in F
bard		Lower in F		Retracted in F	Higher in F	Retracted in F	Higher in F		Lower in F
bared		Lower in F	Retracted in F	Retracted in F		Retracted in F		Higher in F	Lower in F
bed		Lower in F	Retracted in F	Retracted in F		Retracted in F		Higher in F	Lower in F
bowed		Lower in F				Retracted in F			
bod	Lax back vowels	Lower in F			Higher in F	Retracted in F	Higher in F		
bud				Advanced in F		Advanced in F			
bide	Remaining vowels				Higher in F		Higher in F	Higher in F	
buoyed				Advanced in F		Advanced in F			Higher in F
bade		Lower in F					Lower in F		
bead							Lower in F		Higher in F
beard									
bid		Lower in F							
bird								Higher in F	
bored				Advanced in F					

Note. F1 = first formant; F2 = second formant; F = female.

speakers, and also with those of Johnson (2023), who reports a greater range of jaw opening in speakers with relatively shorter vocal tracts. More generally, our findings are in line with the hypothesis that greater jaw opening is a trait of female speech. Greater jaw opening is the possible root cause of the remaining gender differences we observe for the same vowels. We find that the mid and front low vowels listed above all show greater dorsal

Figure 10. By-gender mean tongue contours for the lax vowels, taken at the acoustic midpoint. The data are centered within speaker and averaged using generalized additive mixed modeling. DLC = DeepLabCut; F = female; M = male.

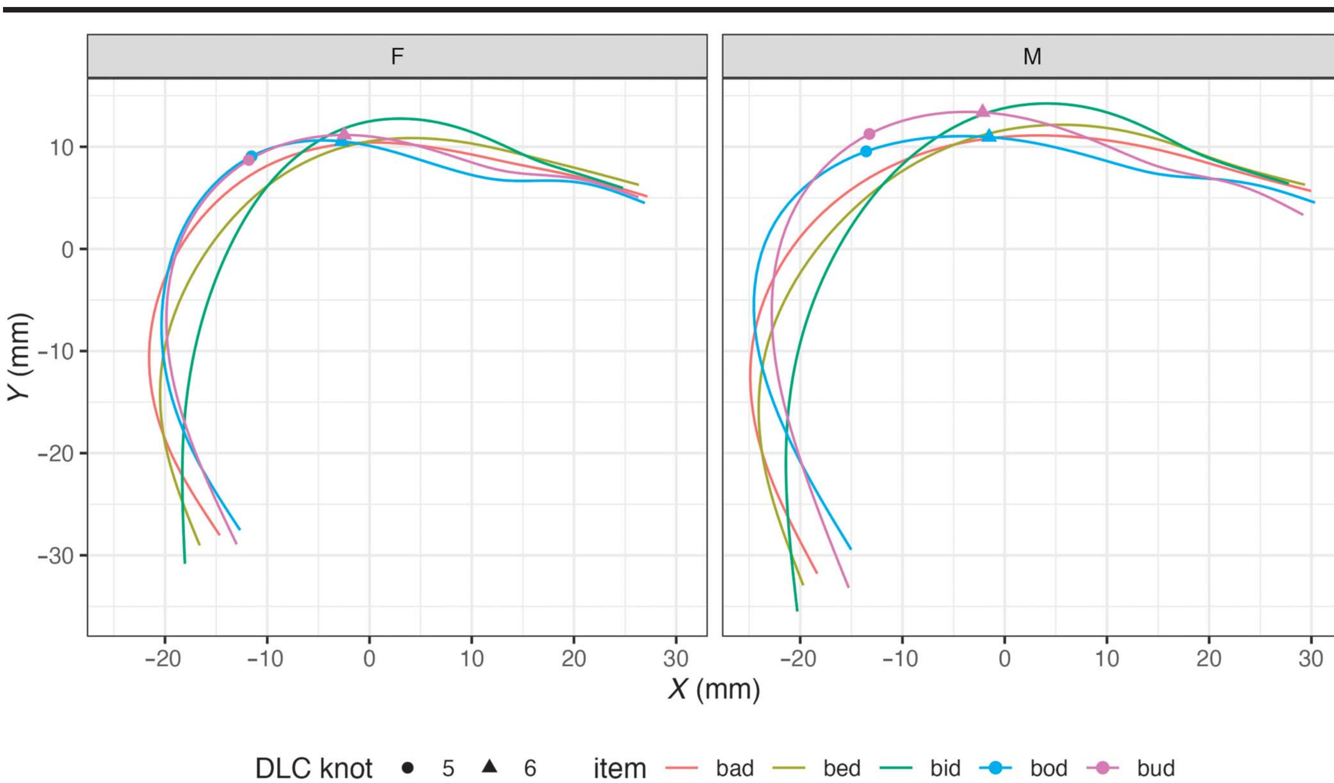
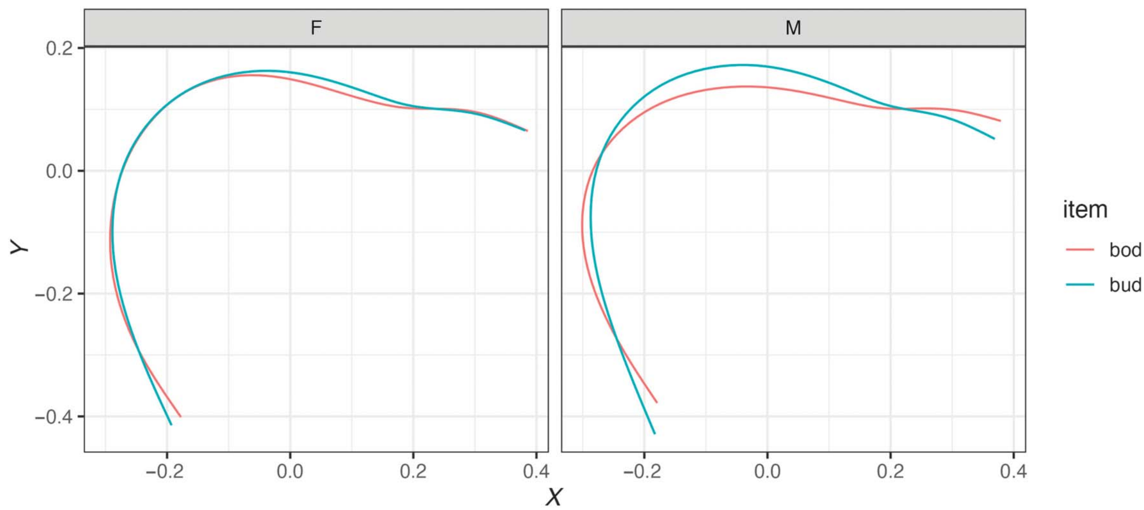


Figure 11. Comparison of female (F) and male (M) mean tongue shape in *bod* and *bud*. The data are rotated, centered, and scaled within speaker and averaged using generalized additive mixed modeling.



retraction in females compared with males. This retraction is potentially a pattern of jaw–tongue coordination, whereby jaw opening leads to retraction of the tongue body even if there is no change in the tongue shape (Lindblom & Sundberg, 1971). The acoustic effects we observe are consistent with well-established patterns of articulatory–acoustic relationship: Jaw lowering is known to produce F1 raising, whereas dorsal retraction lowers F2 (Fant, 1971; Stevens & House, 1955).

It is important to note the scale of the relevant acoustic effects. The F1 differences between male and female *bad*, *bared,* and *bed* are quite robust even following normalization (see Figure 8). We can be fairly confident that these differences arise from distinct articulatory strategies that involve a greater degree of jaw opening in females and a smaller degree in males: Males and females open their jaws to different degrees in order to produce specific resonances. However, for some vowels, the differences are smaller, and the relationship between short tendon lowering and F1 raising is more complex. In *bowed* and *bod*, there is a difference in the displacement of the short tendon, consistent with greater jaw opening in females, but no corresponding difference in F1. Clearly, the jaw opening strategy does not consistently produce the same degree of F1 raising across vowels. The magnitude of F2 lowering in female *bad* and *bed* is very small. It is likely that this type of difference would not be detectable in more naturalistic data where the segmental context is not strictly controlled. We can generalize that the acoustic consequences of greater jaw lowering in females are relatively consistent across low vowels, but they are only robust in some vowels, typically the more front ones.

Lax Back Vowels

Similarly to mid and front low vowels, we find greater short tendon lowering in female *bod* compared with male *bod*, accompanied by some differences in the tongue dorsum: The mid part of the tongue dorsum is lowered and relatively more front in males. However, these articulatory differences do not correspond to any systematic acoustic effects. The only acoustic difference we observe here is increased F1 lowering in the offglide of *bod* in females. This effect is the opposite of what we would expect from increased jaw opening. Since it is limited to the vowel offglide, it is likely a coarticulatory effect from the following coronal consonant. The neighboring lax back vowel *bud* also displays several gender-conditioned articulatory differences: The tongue dorsum is more retracted in males. However, again, no systematic acoustic differences can be found between normalized F1 and F2 trajectories in male and female *bud*.

The combined information from GAMM reported in the Results section and tongue shape comparison in Figures 10 and 11 suggests that males and females differ in their articulatory coordination for the production of *bod* and *bud*. For females, the articulatory contrast between *bod* and *bud* is attained chiefly by jaw lowering in *bod*, while the tongue shape remains relatively similar. For males, the contrast is attained mainly through modification to the tongue shape, while the jaw opening remains relatively similar (compare Figure 7). The tongue shape modifications include lowering and flattening the tongue shape for *bud* and velarization in *bud*. These distinct strategies are compatible with a differentiation between "tongue movers" and "jaw movers" (Johnson, 2023; Johnson et al., 1993). While vowel production clearly involves the contribution of both the tongue and the jaw, tongue movers produce more distinct tongue shapes to produce vocalic contrasts, whereas jaw movers produce a greater range of jaw opening. Apparently, these strategies are acoustically equivalent for the *bod–bud* contrast, producing the same targets in normalized F1 and F2 space. From the articulatory point of view, the wider strategy is similar to what we have already seen: Females rely more on jaw opening to produce low vowels, compared with males. Distinct acoustic goals, however, do not always align with the articulatory differences: In mid and front low vowels, jaw opening produces a discernible acoustic effect, while in the back vowel *bod*, it does not.

This observation is interesting in light of some of the previous explanations as to why male–female differences in jaw opening arise in the first place. As discussed in the introduction section, several explanations have been proposed. One set of explanations has to do with acoustic goals. These goals include clearer speech targets and greater acoustic contrasts in case of females, or they can be interpreted as more broadly indexical of gender categories, without an express intent for producing more or less clear speech. This interpretation can account for some, but not all, of the jaw opening differences we find. It cannot directly explain the cases where we find a systematic difference in jaw opening, but these differences are not reflected in F1, as is the case with *bard*, *bowed,* and *bod*. A different explanation, put forward by Weirich et al. (2016), is that anatomical constraints limit the degree of jaw opening in males, so as to avoid complete linguopharyngeal closure. We would expect this constraint to directly affect back vowels, but not front ones. However, as we have seen, front and mid vowels present robust gender-conditioned differences in jaw opening. Our data do not contradict any of the explanations that might contribute to the development of different articulatory strategies in males and females. However, they clearly demonstrate that the strategies we can identify are typically more general than the context in which a particular phonetic constraint or bias may be active. It appears that articulatory and acoustic factors may be relevant to the development of specific strategies under some conditions, but speakers tend to generalize them to a wider context.

The Role of Articulatory Setting

The idea of a generalization underlines the concept of articulatory setting, as developed by Laver (1980). Articulatory setting can be broadly described as habitual positioning of articulators that characterize the speech of an individual. This habitual positioning interacts with the articulatory requirements for the production of individual segments, which may be conflicting. To describe this relationship, Laver (1980, p. 20) introduces the notion of susceptibility: Some segments are more susceptible to a particular setting, making the settings itself more evident, whereas others may be not susceptible, potentially obliterating the setting. In consequence, articulatory settings may be *intermittent* during speech production; that is, it may be observable only in the context of some segments, effectively manifesting as ranges of the relevant parameters. Reduced jaw range is mentioned by Laver (1980, p. 155) as one of the traits associated with "lax voice," a setting typical of Received Pronunciation, the acrolectal variety of British English. Additionally, the acoustic effect of articulatory setting is mediated by the quantal nature of speech, such that articulatory changes may sometimes have a negligible acoustic effect (Stevens, 1989).

The nature of jaw opening in male and female speech that emerges from our data fits very well with the framework described by Laver (1980). We can view jaw opening as a setting, as it is a pervasive characteristic affecting multiple segments. However, it manifests only intermittently, because the need to produce segmental contrasts overrides the setting in most contexts, with the possible exception of low-vowel targets. As a result, the setting is evidenced as a specific range of jaw movement, relatively greater in females (because their setting is relatively open) and relatively smaller in males, whose setting is closer. The acoustic effect of jaw opening varies, depending on the vowel and its associated susceptibility to the jaw opening setting, as well as on the vowel's position. What we observe is that jaw lowering is more strongly linked with F1 raising for front vowels compared with back vowels, producing acoustic variation in the presence of articulatory stability.

Articulatory Target Uniformity

Another possible mechanism for a generalization of an articulatory strategy is target uniformity, a constraint on speech production that links specific phonetic realization with a distinctive feature, producing within-speaker consistency for natural classes of sounds (Chodroff & Wilson, 2017, 2022). It is plausible that speakers develop some specific motor routines for the production of some vowels and reuse the same routines in a range of related vowel contexts. This type of behavior is economical from the speech production point of view, as speakers can rely on the same highly practiced movements (Maddieson, 1995). Existing evidence for target uniformity includes within-speaker stability in the acoustic realization of voice onset time (Chodroff & Wilson, 2017) and fricatives (Chodroff & Wilson, 2022), as well as selected sound changes that affect classes of sounds (Fruehwald, 2019). Some authors have argued explicitly that target uniformity holds at the articulatory level, such that speakers employ the same motor routines in their realization of phonological features, and acoustic uniformity is a secondary reflection of that. McAllister Byun et al. (2016) model a pressure to execute a stable motor plan as one of the main constraints shaping the development of child speech. Ménard et al. (2008) present evidence of within-speaker consistency of F1 values across vowel categories and argue that this consistency arises from stable patterns of tongue height adopted by individual speakers. Faytak (2022) examines a case of consonant–vowel uniformity in the realization of fricatives and apical vowels in Suzhounese. He reports varying degrees of acoustic uniformity for these categories, ascribed to the effect of an ongoing sound change. On the articulatory level, however, uniformity holds much more strongly, in line with the hypothesis that target uniformity predominantly constrains articulation. Our own findings are similar to those reported by Faytak, in that relative articulatory stability can, in some cases, be linked to acoustic variation.

Articulatory Ranges in Males and Females

A potentially interesting incidental finding in our study is that male speakers do not clearly show larger articulatory ranges compared with female speakers. This is somewhat surprising, given the difference we observe in the size of the tongue (recall that the articulatory displacements we compared were centered but not scaled values measured in millimeters). As shown in Figure 10, the mean tongue size for the male speakers was larger than that for female speakers. The average tongue contour length in our study was 70.5 mm for females (SD = 6.94) and 80.2 mm for males (SD = 9.18). All things being equal, we would expect larger articulators to produce larger displacement, but that is not systematically the case in our data. We did not attempt to quantify the area of the articulatory space for the individual speakers, but we can comment on the male–female differences in the displacement of various articulators, focusing on the most peripheral vowels. In terms of tongue root and dorsum retraction, we do not find gender differences for the most front vowel *bead* or for the most retracted vowel *bored*. In terms of tongue dorsum raising, males show more lowering in the relatively lowest vowel *bard*, potentially signaling a somewhat greater range, but still, the difference is limited.

These observations point to the presence of more undershoot in male speakers, similar to that reported in the study by Weirich and Simpson (2018b). Weirich and Simpson (2018b) compared the range of male and female articulatory displacement in three conditions: accented, unaccented, and control. The control condition represented reading sentences from a list, and so it is the most comparable to our corpus. They report comparable displacement ranges with no significant difference between males and females in the control condition, which resembles our data. Additionally, they found that male speakers produced greater displacement in the accented condition compared with the control condition, but there was no difference between these two conditions for female speakers. Their interpretation was that male speakers produce undershoot in the control condition, whereas the female speakers use their full articulatory range. Our findings are also consistent with this interpretation. Male and female speakers having similar articulatory displacements, despite a difference in size, suggests that female speakers use relatively more of the articulatory range available to them. This, however, is somewhat speculative, as we did not specifically elicit hyperarticulated speech.

Clinical Implications

There are several clinical implications that follow from our findings. Gender-specific vowel resonances have been recognized as a potential goal for gender-affirming speech therapy, that is, therapy that helps transgender people to produce speech that aligns with their gender identity (Kawitzky & McAllister, 2020; Leung et al., 2018; Leyns et al., 2024). Vowel resonances constitute a therapy goal based on the observation that they contribute to the perception of masculinity and femininity (Leung et al., 2018). However, it is not yet clear what specific interventions are effective in achieving these goals and what instructions should be given to speech therapy clients in order to help them achieve the desired articulations.

Carew et al. (2007) and Hirsch (2017) propose more anterior tongue placement and lip spreading as an effective strategy to raise formants and feminize voice. Leung et al. (2018) suggest several articulatory strategies for producing more feminine-sounding speech: “a wider jaw opening, constricting the anterior part of the tongue, positioning the tongue tip as close as possible to the incisors and raising larynx height” (p. 291). In general, these strategies can be expected to raise fundamental frequency ($F0$), F1, and F2. Larynx raising is expected to raise $F0$, as well as higher formants; jaw opening is expected to raise F1; and fronting the tongue body is expected to raise F2. Kawitzky and McAllister (2020) describe the use of tongue body fronting to their participants as a strategy for achieving higher F2. Leyns et al. (2024) use a cork exercise, in which participants produced speech while holding a cork in their mouth in order to achieve wider jaw opening and a more fronted tongue body position.

Our data confirm that wider jaw opening is a strategy systematically used by female speakers in low-vowel production, consistent with increased normalized F1 in the same contexts, although the magnitude of the F1 effect can vary. Jaw opening, however, frequently co-occurs with a more retracted tongue body, which is possibly a passive consequence of increased jaw opening. Acoustically, this is reflected in lower normalized F2 in some low vowels produced by female speakers compared with male speakers.

In light of these results, it is likely that targeting a more fronted tongue body setting may lead to a distortion of the overall acoustic vowel space. While fronting of the tongue body is, in principle, a strategy for raising F2, this strategy is limited by the articulatory requirements on producing vowel phoneme contrasts. Furthermore, it may conflict locally with a strategy for producing more peripheral vowel articulations, which is also characteristic of female speech, including more extreme tongue retraction in back vowels. Navigating these conflicting requirements is a complex problem for speech therapy, and it could benefit from a systematic comparison of interventions focusing on different aspects of articulatory strategy, including increased jaw opening, more fronted tongue setting, and more extreme articulatory movement overall. An additional compelling avenue for investigation is incorporating articulatory imaging into clinical efficacy studies to establish whether instructions from a therapist are successfully implemented by the client to achieve a specific shift in articulation.

Conclusions

Our study confirms the presence of systematic articulatory differences between male and female speakers of Northern English. Some of these differences align with ongoing sound changes (GOOSE and GOAT fronting), whereas others are suggestive of a potentially stable pattern of variation characterized by greater jaw opening in female speech. A crucial novel contribution of our study is that the greater jaw opening appears to be a part of a wider strategy for the production of multiple vowels, but the magnitude of jaw opening and its impact on the acoustics vary depending on the vowel. In some vowels, increased jaw opening leads to proportionally greater F1 in female speakers, but in other vowels, there is no such effect. This suggests that some, but not all, differences in male and female vowel production are directly acoustically driven. In addition to the acoustic targets, articulatory generalization is at play, which can be conceptualized as articulatory setting or target uniformity. Our findings shed light on the systemic nature of male–female differences in speech production, where general constraints on speech production, such as target uniformity, interact with other biases, for example, social or anatomical ones, such that differences may arise without the presence of a local trigger. An important consequence for future studies is to consider articulatory variation for whole sound systems or subsystems, not just individual sounds. This finding is also of potential importance to gender-affirming speech therapy, as it highlights that interventions targeting vowel resonances may need to be fine-tuned for different vowels.

Data Availability Statement

The data and code used in the analysis are publicly available through an Open Science Project repository at https://osf.io/zcet3.

Acknowledgments

The research reported in this article was funded by Arts and Humanities Research Council Grant AH/S011900/1 (awarded to P.S. and S.K.), British Academy Grant MFSS24\40076 (awarded to P.S.), and Arts and Humanities Research Council Fellowship AH/Y002822/1 (awarded to S.K.). The authors would like to acknowledge the assistance given by Research IT and the use of the Computational Shared Facility at The University of Manchester.

References

Articulate Instruments Ltd. (2014). *Articulate Assistant Advanced ultrasound module user manual, Revision 2.16*.

Barreda, S. (2021). Perceptual validation of vowel normalization methods for variationist research. *Language Variation and Change, 33*(1), 27–53. https://doi.org/10.1017/S0954394521000016

Boersma, P., & Weenink, D. (2009). Praat: Doing phonetics by computer (Version 6.2.14) [Computer program]. Institute of Phonetic Sciences.

Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech Communication, 15*(1–2), 39–54. https://doi.org/10.1016/0167-6393(94)90039-6

Carew, L., Dacakis, G., & Oates, J. (2007). The effectiveness of oral resonance therapy on the perception of femininity of voice in male-to-female transsexuals. *Journal of Voice, 21*(5), 591–603. https://doi.org/10.1016/j.jvoice.2006.05.005

Chodroff, E., & Wilson, C. (2017). Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. *Journal of Phonetics, 61,* 30–47. https://doi.org/10.1016/j.wocn.2017.01.001

Chodroff, E., & Wilson, C. (2022). Uniformity in phonetic realization: Evidence from sibilant place of articulation in American English. *Language, 98*(2), 250–289. https://doi.org/10.1353/lan.2022.0007

Cole, A., & Strycharczuk, P. (2024). The search for linguistically coherent accents: Unsupervised clustering of diphthong variation in Southeast England. *English World-Wide, 45*(1), 1–29. https://doi.org/10.1075/eww.22054.col

Coretta, S., & Sakr, G. (2024). Analysing tongue contours with multivariate generalised additive models. In I. Wilson, A.

Mizoguchi, J. Perkins, J. Villegas, & N. Yamane (Eds.), *Ultrafest XI: Extended abstracts* (pp. 100–101). Zenodo.
Dryden, I. L., & Mardia, K. V. (2016). *Statistical shape analysis, with applications in R*. Wiley. https://doi.org/10.1002/9781119072492.index
Fant, G. (1966). A note on vocal tract size factors and non-uniform F-pattern scalings. *Speech Transmission Laboratory Quarterly Progress and Status Report, 7,* 22–30.
Fant, G. (1971). *Acoustic theory of speech production: With calculations based on X-ray studies of Russian articulations*. Walter de Gruyter. https://doi.org/10.1515/9783110873429
Fant, G. (1975). Non-uniform vowel normalization. *Speech Transmission Laboratory Quarterly Progress and Status Report, 16,* 1–19.
Faytak, M. (2022). Place uniformity and drift in the Suzhounese fricative and apical vowels. *Linguistics Vanguard, 8*(s5), 569–581. https://doi.org/10.1515/lingvan-2021-0071
Fruehwald, J. (2019). Is phonetic target uniformity phonologically, or sociolinguistically grounded. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019* (pp. 681–685). https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2019/papers/ICPhS_730.pdf [PDF]
Goldstein, U. G. (1980). *An articulatory model for the vocal tracts of growing children* [Unpublished doctoral dissertation]. Massachusetts Institute of Technology
Haddican, B., Foulkes, P., Hughes, V., & Richards, H. (2013). Interaction of social and linguistic constraints on two vowel changes in northern England. *Language Variation and Change, 25*(3), 371–403. https://doi.org/10.1017/S0954394513000197
Heffernan, K. (2010). Mumbling is macho: Phonetic distinctiveness in the speech of American radio DJs. *American Speech, 85*(1), 67–90. https://doi.org/10.1215/00031283-2010-003
Henton, C. (1995). Cross-language variation in the vowels of female and male speakers. In K. Elenius & P. Branderud (Eds.), *Proceedings of the XIIIth International Congress of Phonetic Sciences* (Vol. *4,* pp. 420–423).
Hirsch, S. (2017). Combining voice, speech science and art approaches to resonant challenges in transgender voice and communication training. *Perspectives of the ASHA Special Interest Groups, 2*(10), 74–82. https://doi.org/10.1044/persp2.SIG10.74
Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics, 34*(4), 485–499. https://doi.org/10.1016/j.wocn.2005.08.004
Johnson, K. (2020). The ΔF method of vocal tract length normalization for vowels. *Laboratory Phonology: Journal of the Association for Laboratory Phonology, 11*(1), Article 10. https://doi.org/10.5334/labphon.196
Johnson, K. (2023). Individual differences in speech production: What is "phonetic substance"? In R. Skarnitzl & J. Volín (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences* (pp. 1102–1106). International Phonetic Association. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2023/full_papers/867.pdf [PDF]
Johnson, K., Ladefoged, P., & Lindau, M. (1993). Individual differences in vowel production. *The Journal of the Acoustical Society of America, 94*(2), 701–714. https://doi.org/10.1121/1.406887
Kawitzky, D., & McAllister, T. (2020). The effect of formant biofeedback on the feminization of voice in transgender women. *Journal of Voice, 34*(1), 53–67. https://doi.org/10.1016/j.jvoice.2018.07.017
Kirkham, S., Strycharczuk, P., Gorman, E., Nagamine, T., & Wrench, A. (2023). Co-registration of simultaneous high speed ultrasound and electromagnetic articulography for speech production research. In R. Skarnitzl & J. Volín (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences* (pp. 918–922). International Phonetic Association. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2023/full_papers/145.pdf [PDF]
Labov, W. (1990). The intersection of sex and social class in the course of linguistic change. *Language Variation and Change, 2*(2), 205–254. https://doi.org/10.1017/S0954394500000338
Labov, W. (1994). *Principles of linguistic change: Internal factors* (Vol. 1). Blackwell.
Labov, W. (2001). *Principles of linguistic change: Social factors* (Vol. 2). Blackwell.
Lammert, A. C., & Narayanan, S. S. (2015). On short-time estimation of vocal tract length from formant frequencies. *PLOS ONE, 10*(7), Article e0132193. https://doi.org/10.1371/journal.pone.0132193
Laver, J. (1980). *The phonetic description of voice quality*. Cambridge University Press.
Lawson, E., Stuart-Smith, J., & Rodger, L. (2019). A comparison of acoustic and articulatory parameters for the GOOSE vowel across British Isles Englishes. *The Journal of the Acoustical Society of America, 146*(6), 4363–4381. https://doi.org/10.1121/1.5139215
Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America, 105*(3), 1455–1468. https://doi.org/10.1121/1.426686
Leung, Y., Oates, J., & Chan, S. P. (2018). Voice, articulation, and prosody contribute to listener perceptions of speaker gender: A systematic review and meta-analysis. *Journal of Speech, Language, and Hearing Research, 61*(2), 266–297. https://doi.org/10.1044/2017_JSLHR-S-17-0067
Leyns, C., Corthals, P., Cosyns, M., Papeleu, T., Van Borsel, J., Morsomme, D., T'Sjoen, G., & D'haeseleer, E. (2024). Acoustic and perceptual effects of articulation exercises in transgender women. *Journal of Voice, 38*(1), 246.e15–246.e25. https://doi.org/10.1016/j.jvoice.2021.06.033
Lindblom, B. E., & Sundberg, J. E. (1971). Acoustical consequences of lip, tongue, jaw, and larynx movement. *The Journal of the Acoustical Society of America, 50*(4B), 1166–1179. https://doi.org/10.1121/1.1912750
Maddieson, I. (1995). Gestural economy. In K. Elenius & P. Branderud (Eds.), *Proceedings of the XIIIth International Congress of Phonetic Sciences* (Vol. 4, pp. 574–578).
Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., & Bethge, M. (2018). DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience, 21*(9), 1281–1289. https://doi.org/10.1038/s41593-018-0209-y
McAllister Byun, T. M., Inkelas, S., & Rose, Y. (2016). The A-map model: Articulatory reliability in child-specific phonology. *Language, 92*(1), 141–178. https://doi.org/10.1353/lan.2016.0000
Ménard, L., Schwartz, J.-L., & Aubin, J. (2008). Invariance and variability in the production of the height feature in French vowels. *Speech Ccommunication, 50*(1), 14–28. https://doi.org/10.1016/j.specom.2007.06.004
Munson, B., & Babel, M. (2019). The phonetics of sex and gender. In W. F. Katz & P. F. Assmann (Eds.), *The Routledge handbook of phonetics* (pp. 499–525). Routledge. https://doi.org/10.4324/9780429056253-19
Nearey, T. M. (1989). Static, dynamic, and relational properties in vowel perception. *The Journal of the Acoustical Society of America, 85*(5), 2088–2113. https://doi.org/10.1121/1.397861
Nearey, T. M., & Assmann, P. F. (1986). Modeling the role of inherent spectral change in vowel identification. *The Journal of the Acoustical Society of America, 80*(5), 1297–1308. https://doi.org/10.1121/1.394433

Nordström, P.-E. (1977). Female and infant vocal tracts simulated from male area functions. *Journal of Phonetics, 5*(1), 81–92. https://doi.org/10.1016/S0095-4470(19)31115-5

Serrurier, A., & Neuschaefer-Rube, C. (2024). Formant-based articulatory strategies: Characterisation and inter-speaker variability analysis. *Journal of Phonetics, 107,* Article 101374. https://doi.org/10.1016/j.wocn.2024.101374

Simpson, A. P. (2001). Dynamic consequences of differences in male and female vocal tract dimensions. *The Journal of the Acoustical Society of America, 109*(5, Pt. 1), 2153–2164. https://doi.org/10.1121/1.1356020

Simpson, A. P. (2002). Gender-specific articulatory–acoustic relations in vowel sequences. *Journal of Phonetics, 30*(3), 417–435. https://doi.org/10.1006/jpho.2002.0171

Simpson, A. P. (2009). Phonetic differences between male and female speech. *Language and Linguistics Compass, 3*(2), 621–640. https://doi.org/10.1111/j.1749-818X.2009.00125.x

Soskuthy, M. (2021). Evaluating generalised additive mixed modelling strategies for dynamic speech analysis. *Journal of Phonetics, 84,* Article 101017. https://doi.org/10.1016/j.wocn.2020.101017

Spreafico, L., Pucher, M., & Matosova, A. (2018). UltraFit: A speaker-friendly headset for ultrasound recordings in speech science. In *Proceedings of Interspeech 2018, 19th Annual Conference of the International Speech Communication Association, Hyderabad, India* (pp. 1517–1520). International Speech Communication Association.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics, 17*(1–2), 3–45. https://doi.org/10.1016/S0095-4470(19)31520-7

Stevens, K. N., & House, A. S. (1955). Development of a quantitative description of vowel articulation. *The Journal of the Acoustical Society of America, 27*(3), 484–493. https://doi.org/10.1121/1.1907943

Strycharczuk, P., Kirkham, S., Gorman, E., & Nagamine, T. (2024). Towards a dynamical model of English vowels. Evidence from diphthongisation. *Journal of Phonetics, 107,* Article 101349. https://doi.org/10.1016/j.wocn.2024.101349

Strycharczuk, P., Kirkham, S., Gorman, E., & Nagamine, T. (2025). Dimensionality reduction in lingual articulation of vowels: Evidence from lax vowels in Northern Anglo-English. *Language and Speech, 68*(3), 689–721. https://doi.org/10.1177/00238309251320581

Strycharczuk, P., López-Ibáñez, M., Brown, G., & Leemann, A. (2020). General Northern English. Exploring regional variation in the north of England with machine. *Frontiers in Artificial Intelligence, 3,* Article 48. https://doi.org/10.3389/frai.2020.00048

Traunmüller, H. (1984). Articulatory and perceptual factors controlling the age- and sex-conditioned variability in formant frequencies of vowels. *Speech Communication, 3*(1), 49–61. https://doi.org/10.1016/0167-6393(84)90008-6

Turton, D., & Lennon, R. (2023). An acoustic analysis of rhoticity in Lancashire, England. *Journal of Phonetics, 101,* Article 101280. https://doi.org/10.1016/j.wocn.2023.101280

van Rij, J., Wieling, M., Baayen, R. H., & van Rijn, H. (2015). *itsadug: Interpreting time series and autocorrelated data using GAMMs* (R package Version 1.0.3).

Vorperian, H. K., & Kent, R. D. (2007). Vowel acoustic space development in children: A synthesis of acoustic and anatomic data. *Journal of Speech, Language, and Hearing Research, 50*(6), 1510–1545. https://doi.org/10.1044/1092-4388(2007/104)

Weirich, M., Fuchs, S., Simpson, A., Winkler, R., & Perrier, P. (2016). Mumbling: Macho or morphology? *Journal of Speech, Language, and Hearing Research, 59*(6), S1587–S1595. https://doi.org/10.1044/2016_JSLHR-S-15-0040

Weirich, M., & Simpson, A. P. (2018a). Gender identity is indexed and perceived in speech. *PLOS ONE, 13*(12), Article e0209226. https://doi.org/10.1371/journal.pone.0209226

Weirich, M., & Simpson, A. P. (2018b). Individual differences in acoustic and articulatory undershoot in a German diphthong—Variation between male and female speakers. *Journal of Phonetics, 71,* 35–50. https://doi.org/10.1016/j.wocn.2018.07.007

Wells, J. (1982). *Accents of English 1: An introduction* (Vol. 2). Cambridge University Press. https://doi.org/10.1017/CBO9780511611759

Westbury, J. R., Hashi, M., & Lindstrom, M. J. (1998). Differences among speakers in lingual articulation for American English /ɹ/. *Speech Communication, 26*(3), 203–226. https://doi.org/10.1016/S0167-6393(98)00058-2

Whiteside, S. P. (2001). Sex-specific fundamental and formant frequency patterns in a cross-sectional study. *The Journal of the Acoustical Society of America, 110*(1), 464–478. https://doi.org/10.1121/1.1379087

Wood, S. (2017). *Generalized additive models: An introduction with R* (2nd ed.). Chapman & Hall/CRC Press. https://doi.org/10.1201/9781315370279

Wrench, A., & Balch-Tomes, J. (2022). Beyond the edge: Markerless pose estimation of speech articulators from ultrasound and camera images using DeepLabCut. *Sensors, 22*(3), Article 1133. https://doi.org/10.3390/s22031133