

- 1 Articulatory strategies in male and female vowel production
- 2
- 3 Patrycja Strycharczuk (corresponding author)
- 4 *Linguistics and English Language*
- 5 *University of Manchester*
- 6 *patrycja.strycharczuk@manchester.ac.uk*
- 7
- 8 Sam Kirkham
- 9 *Linguistics and English Language*
- 10 *Lancaster University*
- 11
- 12 Conflict of interest statement
- 13 *The authors have no competing interests to declare.*

Abstract

Purpose: This study investigates the articulatory strategies used by male and female speakers to produce vowel sounds.

Method: Secondary data analysis of a pre-existing articulatory corpus was carried out. Dynamic midsagittal ultrasound and acoustic data from 36 speakers of Northern English (21 females and 15 males) were analysed, representing 17 vowel phonemes in a controlled phonetic environment. Articulatory landmarks corresponding to the tongue root, dorsum and mandibular short tendon were automatically labelled in the ultrasound image, and their dynamic displacement was analysed using Generalised Additive Mixed Modelling. Dynamic formant trajectories for F1 and F2 were analysed using the same method.

Results: Significant articulatory differences were found between male and female speakers for several vowels. Increased tongue dorsum fronting and lowering was 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 to male speakers. For STRUT / FOOT, there was greater retraction of the tongue dorsum in males. For some vowels, e.g. TRAP and DRESS, corresponding differences were detected in normalised 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 normalised 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 involvement of generalisation mechanisms. Clinical implications for gender affirming speech therapy are discussed.

Keywords: vowels; articulation; gender

1 Introduction

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 non-uniformity was first reported by Fant (1966) for American English and Swedish. Fant noted that male-female differences in 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 to 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 obtain, 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 associated with behavioural, social and cultural factors (Munson and Babel, 2019). In our summary of previous findings below, we follow the original study in using the terms 'sex' or 'gender,

58 'but effectively the synthesised picture of available evidence represents cisgender males
59 and cisgender females.

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

72 Limitations of anatomical accounts become apparent when we consider the
73 magnitude of gender differences in formant values, as well as variation concerning said
74 differences. Early vocal tract modelling studies showed that observed differences
75 between male and female formants are greater than those predicted by differences in
76 vocal tract anatomy, and that anatomical factors alone fail to produce non-uniform scaling
77 patterns (Nordström, 1977; Goldstein, 1980). Another issue is that developmental
78 patterns of vowel formant changes through childhood and puberty to adulthood do not
79 align well with changes in vocal tract size (Lee et al., 1999; Whiteside, 2001; Vorperian
80 and Kent, 2007). Furthermore, average male-female differences in vowel formants are
81 language-specific, as shown by Johnson (2006). Johnson observes that this cross-
82 linguistic variation is not predicted by differences between male and female vocal tract

length, as approximated by speaker height. This suggests that speakers target language- and gender- specific resonances to express their gender within a particular socio-cultural 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 and 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 (Simpson, 2009; Munson and Babel, 2019). 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 non-uniform 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, which suggests that anatomical factors might be at play, though it has to be acknowledged that bulk of the available evidence comes from Indo-European languages, spoken in communities that may share multiple cultural gender norms. Potential anatomical biases that might be relevant here include the diverging resonating properties of the male and female vocal tract, 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 MRI imaging, they show that the same degree of jaw opening may lead to linguo-pharyngeal closure in a male, 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 paper (Weirich et al., 2016). Social factors favouring non-uniform 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 behavioural accounts summarised above lead to a common prediction that there is an underlying difference in articulatory behaviour between male and female speakers. If non-uniform 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) discusses male-female differences in vowel space size in terms of differences in articulation, and mentions 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 analysed 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 characterised by some well understood regularities (Stevens and House, 1955; Fant, 1971). 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 et al., 1993; Johnson, 2023).

The previously discussed study by 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: /ou/ in American English, and /a:/ in German, and found greater angle of jaw opening in female speakers, compared to 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. 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 analyse gender-specific differences in the articulation of diphthongs. Simpson (2001) analyses gender differences in the realisation of /aɪ/ in the X-Ray Microbeam corpus and reports greater sensor displacement and higher velocities in male speaker production. Simpson (2002) analyses the vowel sequence in the phrase *they all*, taken from the same corpus. Similarly to the /aɪ/ data, greater sensor displacement and higher velocity was observed for male speakers. The difference in sensor displacement and velocity are 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 normalised. Weirich and Simpson (2018b) report gender-conditioned differences in the length of the diphthong trajectory in German /aɪ/, such that females show larger normalised displacements of the tongue body in the diphthong production. Additionally, the study analysed 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 conditions, 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 were 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 analyse 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 analyse the acoustic data from the same speakers, in order to understand whether the observed articulatory and acoustics differences correspond to one another, and to relate our findings to previous acoustic results.

2 Method

2.1 Data

The data in our study come from the TarDiS corpus (Kirkham et al., 2023; Strycharczuk et al., 2024, 2025). The corpus comprises time-synchronised 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 stabilisation was used for recording the ultrasound data, and the speakers' occlusal plane was recorded to

standardise the rotation of ultrasound measurements. For more detail 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 non-binary speaker. No information is available about the speakers' sex. 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 non-binary speaker were not included, as no group-level generalisation 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 and 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 female speakers was 21.5 (range 18–41), whereas the mean age of male speakers was 30 (range 18–48).

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 non-rhotic. 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 original data collection received ethical approval as reported in Strycharczuk et al. (2024, 2025).

2.2 Data processing

The ultrasound data were sampled at 59.5–101 frames per second (median = 81.3) 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 labelled 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 and 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 tongue root, knot 5 representing the posterior part of the tongue dorsum and typically the constriction point for back vowels, and knot 6, the mid part of the tongue dorsum, and typically the DLC knot

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
bard	START	ɑ:
bared	SQUARE	ɛə
bead	FLEECE	i:

beard	NEAR	ɪə / ɪː
bed	DRESS	ɛ
bid	KIT	ɪ
bide	PRICE	aɪ
bird	NURSE	ɜː
bod	LOT	ɒ
bode	GOAT	oʊ
booed	GOOSE	uː
bored	NORTH	ɔː
bowed	MOUTH	aʊ
bud	STRUT	ʊ
buoyed	CHOICE	ɔɪ

representative of constriction in non-back 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 and Balch-Tomes, 2022). The displacement of the short tendon is controlled by the jaw, and we therefore analyse 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 millimetres. To improve figure legibility, we centred 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 GAMM, as described in Section 2.3). As we can see from the figure, the dynamic displacement of key DLC knots patterns largely as expected given the articulation of /baid/. The tongue root (knot 3) initially retracts until it reaches its most extreme posterior position, at around 150ms (ca. 25% into the vowel). The tongue root then moves forward until it reaches its most anterior extremum at around 350ms, at which point the tongue root returns to its mean position. The overall temporal and spatial pattern of tongue dorsum fronting is 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

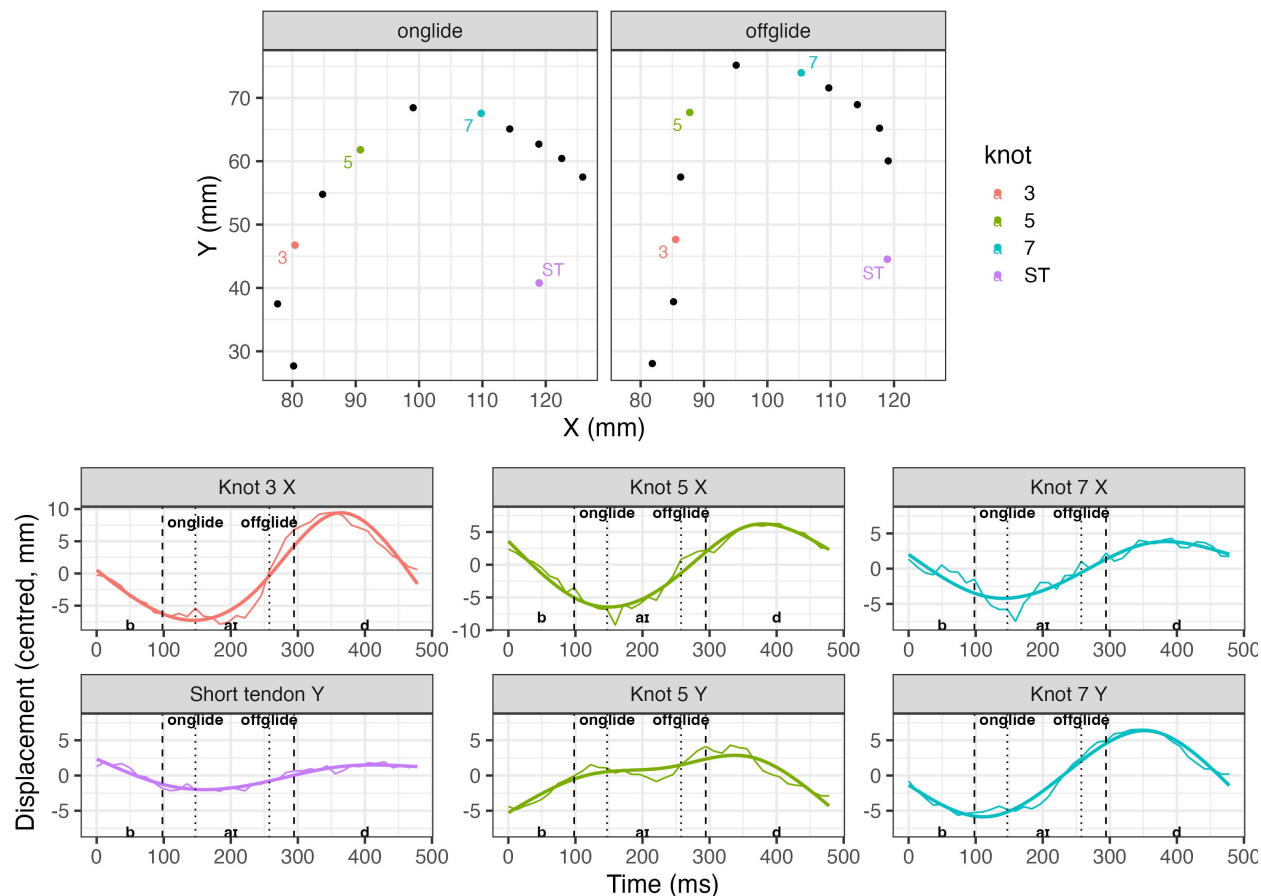


Figure 1: Top: location of DLC knots for an example token of *bide*, representing diphthong onglide and offglide. Key knots analysed in the study are highlighted. Bottom: dynamic displacement of the key knots in the same token of *bide*. Horizontal displacement (X) and vertical displacement (Y) are plotted separately. Both unsmoothed and smoothed displacement is shown.

dorsum (knot 5), followed by lowering once the full closure for /d/ is achieved (at around 375ms). 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 compared to 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 to the other articulators. The pattern of lowering and raising reflects 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 centred within speaker; i.e. for each displacement trajectory within each speaker, we calculated the mean value, and subtracted the mean from each measurements. No further spatial normalisation was done, as the exploratory analysis suggested that the mean male and female displacement values were comparable. We discuss this further in Section 4. The time domain was linearly normalised within phoneme. The vowel onset was taken marked as 0, and the vowel offset as 1. The offset of final /d/ was marked as 2, and the onset of initial /b/ as -1. This normalisation procedure is centred 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 optimises LPC analysis settings for each speaker. This analysis was implemented in Praat version 6.2.14 (Boersma and Weenink, 2009). The maximum formant range was set to 4500–6500 Hz for male speakers and to 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, Statistic = median. The formant measurements were sampled at every 2ms (equivalent to 500 Hz), and subsequently downsampled to match the temporal resolution of the ultrasound data.

The formant values were normalised using the ΔF method (Johnson, 2020). This method normalises 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 (1).

$$(1) \quad \Delta F = 0.56 * F1 + 0.20666 * F2 + 0.188 * F3$$

$$\text{norm} = F / \Delta F$$

The advantage of the ΔF method is that it explicitly 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 normalisation are similar to those obtained by log-mean vowel normalisation (Nearey and Assmann, 1986; Nearey, 1989). Both approaches tend to preserve some aspects of non-uniformity in male-female formants differences, especially the relatively higher F1 in female low vowels. This has been argued to be more representative of perceptual normalisation (Barreda, 2021).

The normalised F1 and F2 trajectories representing the entire vowel were submitted to statistical analysis, as described in Section 2.3 below.

2.3 Statistical analysis

The data were analysed dynamically, using Generalised Additive Mixed Modelling (GAMM, Wood 2017) with Maximum Likelihood estimation. Models were run individually for each of the dependent variables. These were the articulatory variables exemplified in Figure 1, as well as normalised 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 normalised time;
- smooth term for normalised time by gender;
- by-speaker random intercept;
- a random smooth for normalised 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, or 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 anti-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 Maximum Likelihood 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 $\alpha = 0.05$. We did not investigate systematically whether the gender differences were linear or not (i.e. whether the by-gender smooth for normalised 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 `plot_diff()` 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 focusing on visual representation of model predictions and significance testing.

3 Results

3.1 Dynamic articulatory results

Figure 2 shows the prediction of by-vowel GAMM 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. Grey-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 or end of any bVd interval. 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 to males. In *booed*, we find the opposite: males have a more retracted tongue position, compared to females.

Figure 3 shows the prediction of GAMM modelling of tongue dorsum advancement / retraction, as captured by the horizontal displacement of DLC knot 5. This DLC 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 to males 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, i.e. greater tongue root retraction in males compared to 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.

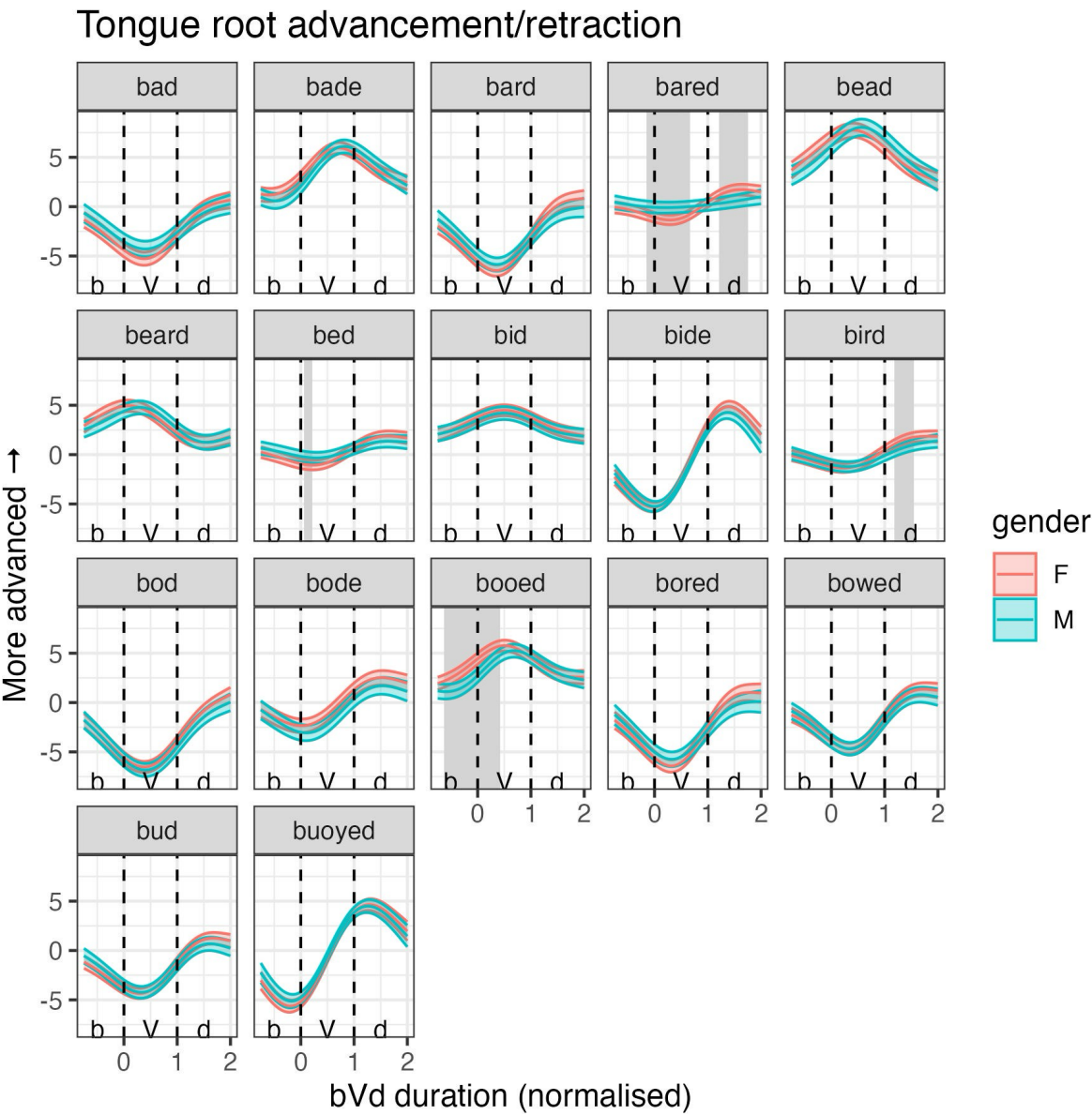
Very similar results emerge when looking at the horizontal displacement of knot 6 (Figure 4), which represents the mid part of the tongue dorsum, and which typically captures the constriction location for mid and front vowels (though some front vowels may have a more anterior constriction). All the gender-based differences affecting position of knot 5 also affect position of knot 6: this part of 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 to males.

In order to examine gender differences within dorsum height, we analysed vertical displacement of DLC knots 5 and 6. In line with other measurements, the displacement values were centralised within 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 tongue dorsum in *bard*, *bide* and *bod* for males. In *bode*, the opposite pattern is found: there is more raising in males, compared to 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 to males. Increased dorsum raising in males, compared to females, is found in *bead* and the offglide of *booed*.

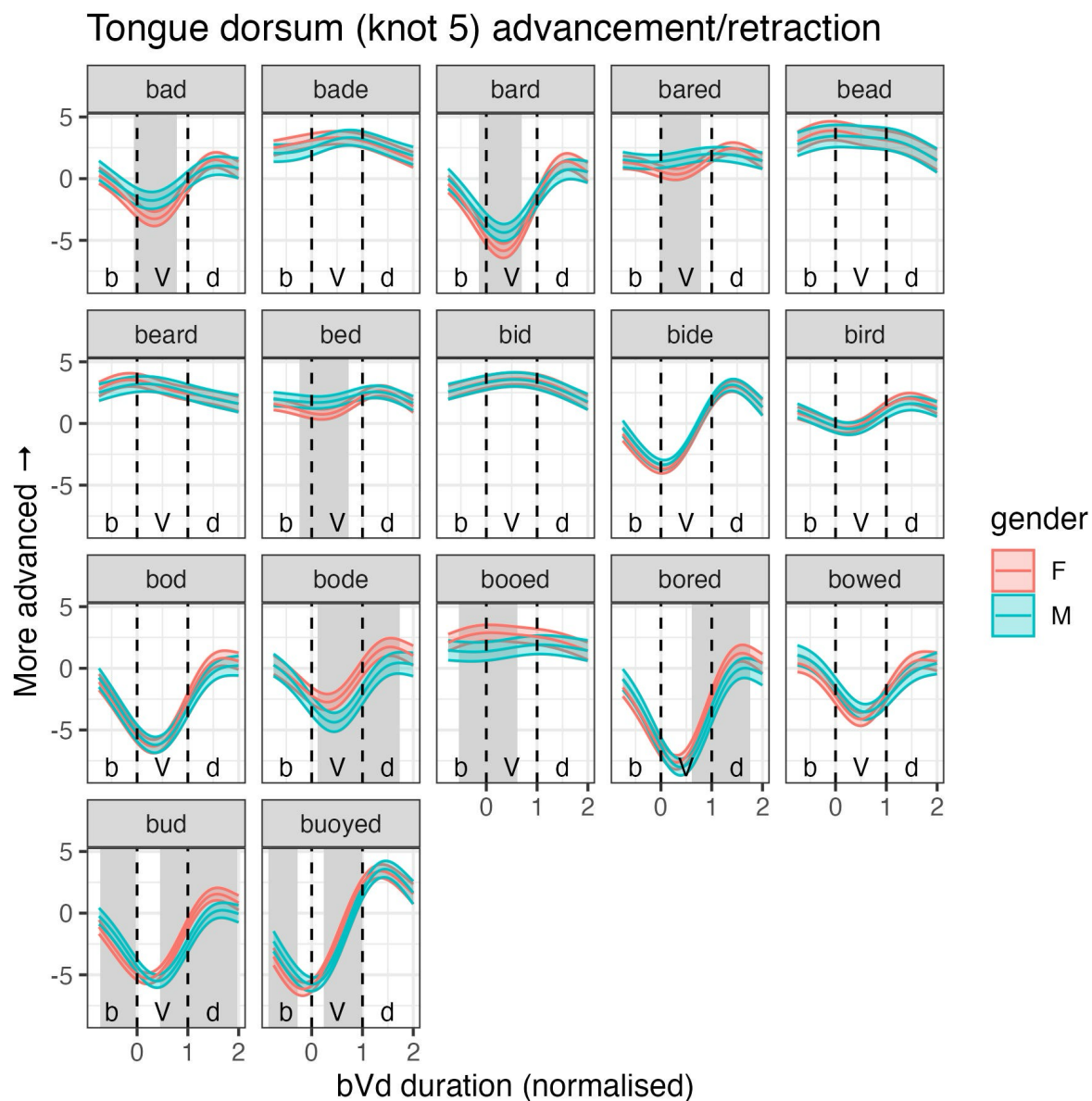
As a proxy for jaw opening and raising, we analysed the vertical displacement of the short tendon, as tracked in the ultrasound image by the DLC algorithm. This measure was modelled using GAMM, analogously to the tongue-related measures. The model predictions for short tendon raising and lowering are visualised 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 to males. The opposite pattern is found in *booed*, and *bode*, where there is more short tendon lowering (greater jaw opening) in males, compared to females.

411



412

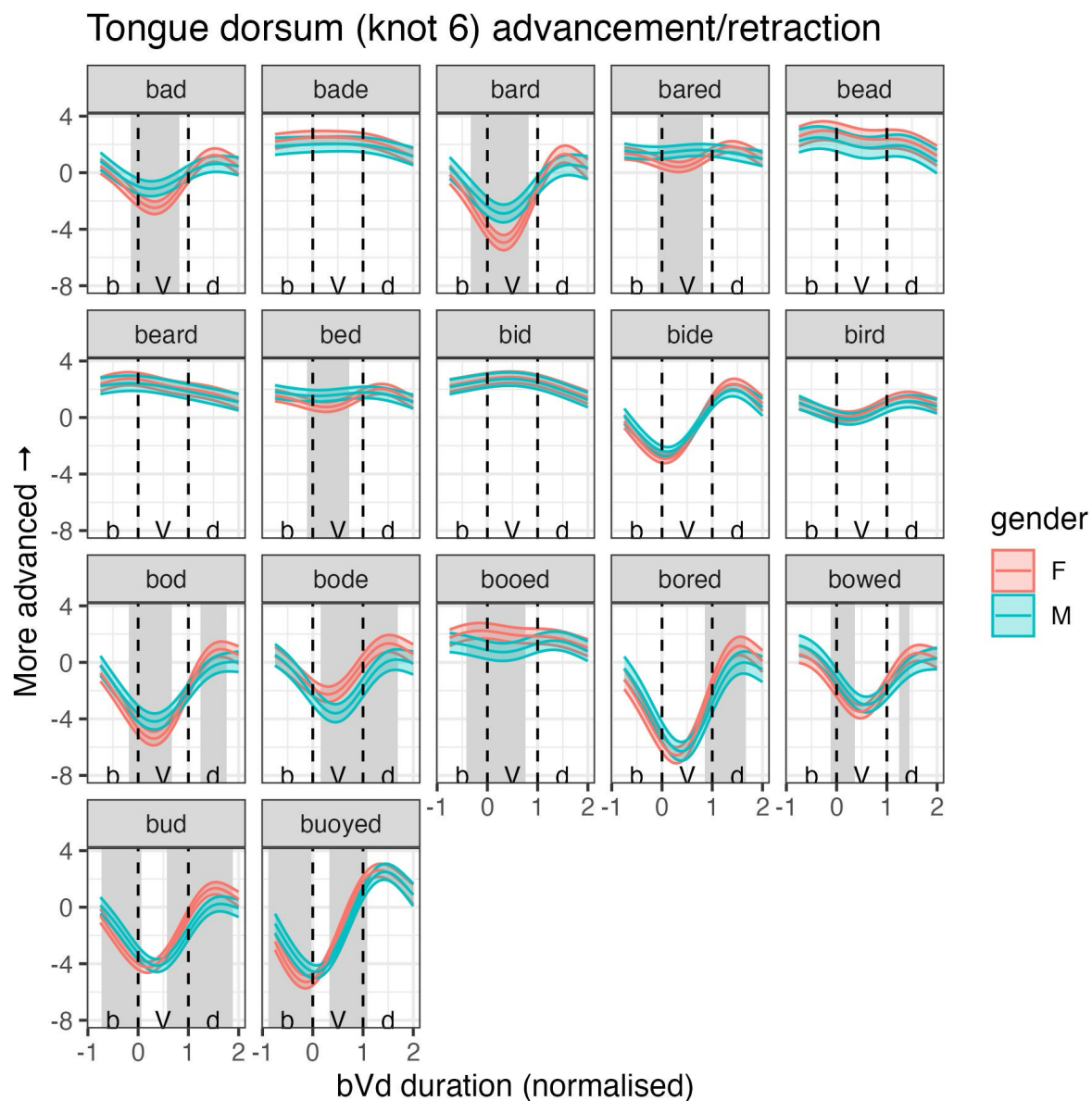
413 Figure 2: GAMM predictions for horizontal tongue root displacement (knot 3) in normalised
414 time, as a function of item and gender



415

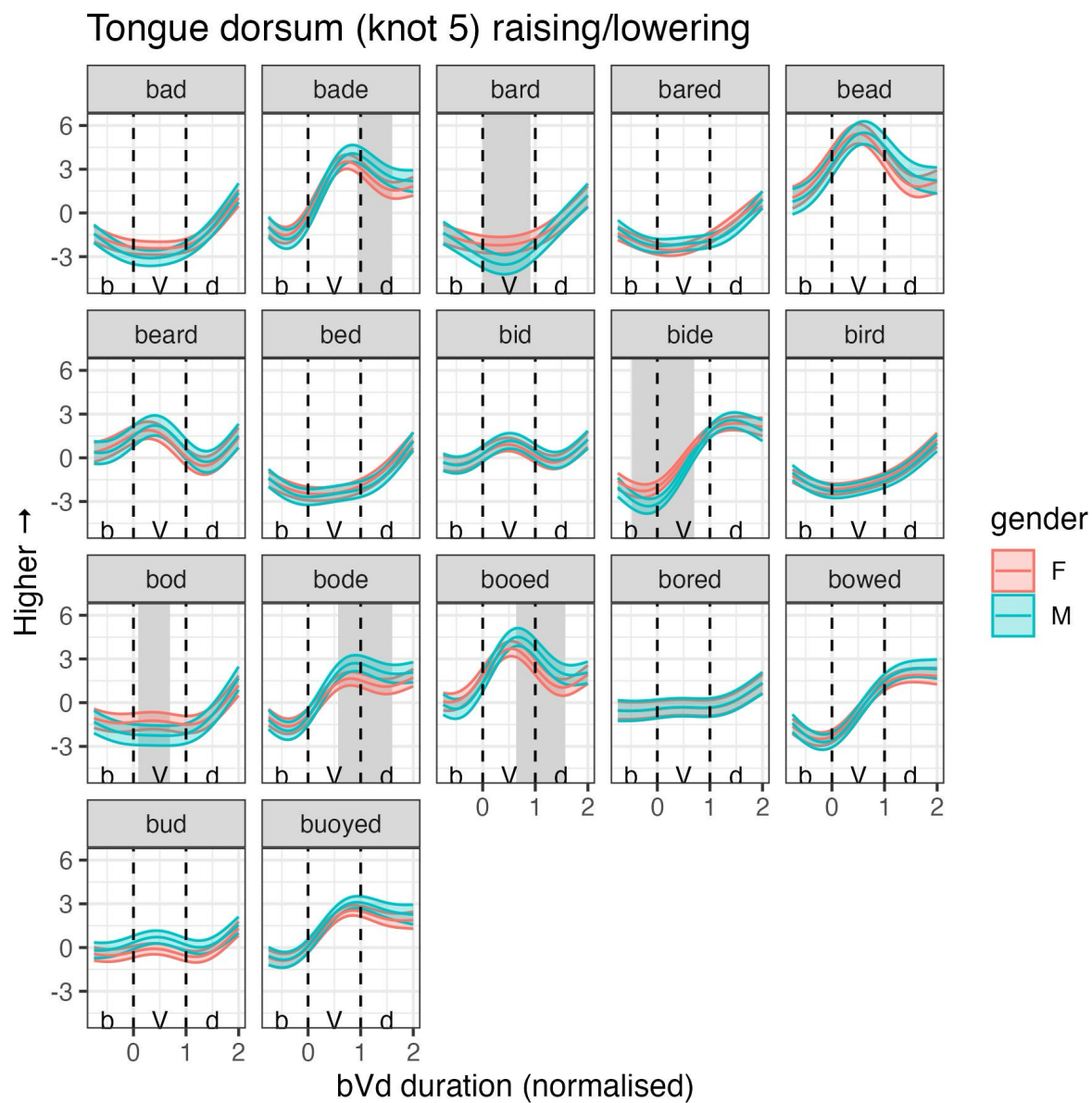
416 Figure 3: GAMM predictions for horizontal displacement of the posterior part of the tongue

417 dorsum (knot 5) in normalised time, as a function of item and gender



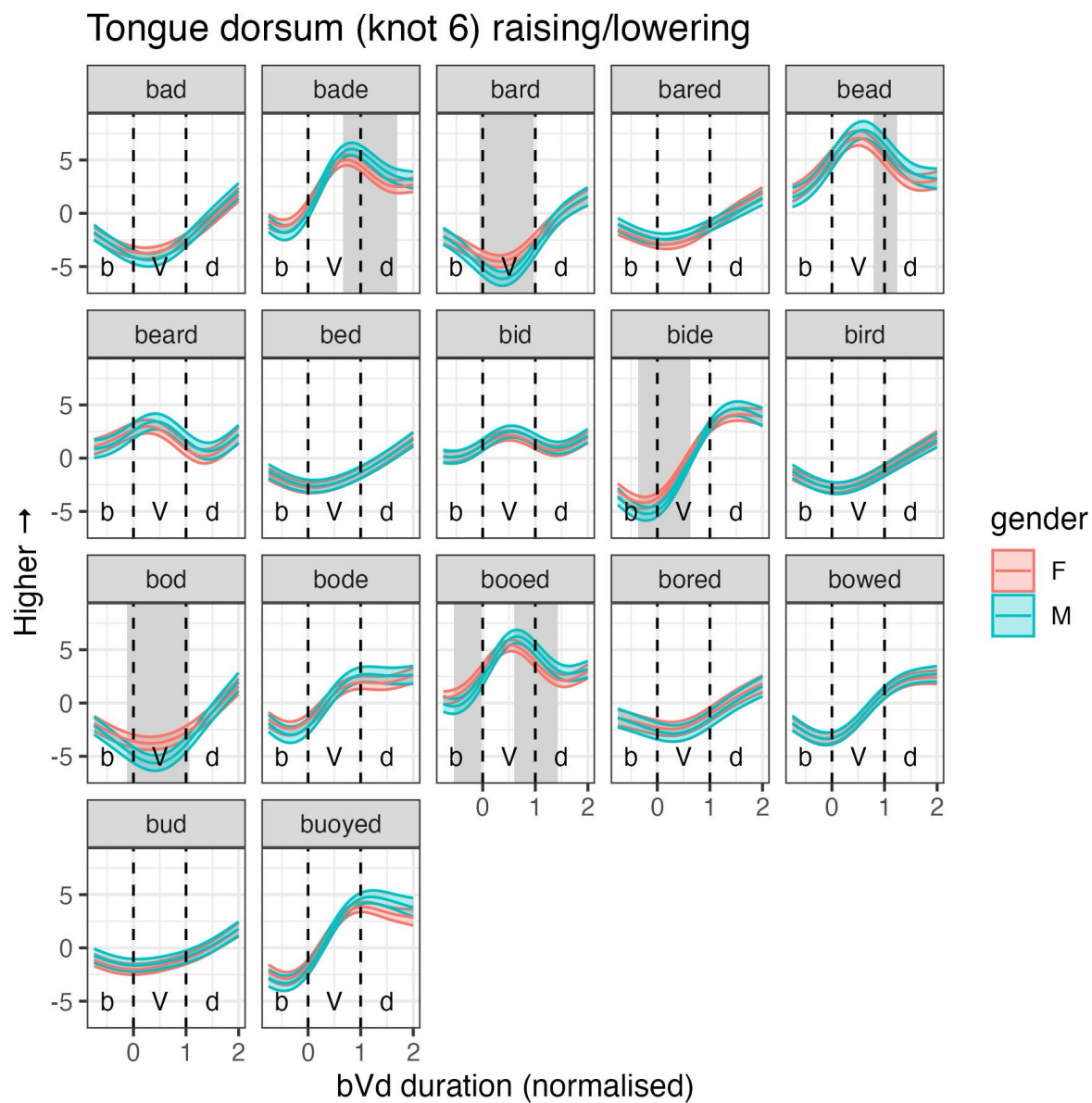
418

419 Figure 4: GAMM predictions for horizontal displacement of the mid part of the tongue
 420 dorsum (knot 6) in normalised time, as a function of item and gender



421

422 Figure 5: GAMM predictions for vertical displacement of the posterior part of the tongue
 423 dorsum (knot 5) in normalised time, as a function of item and gender



424

425 Figure 6: GAMM predictions for vertical displacement of the posterior part of the tongue
 426 dorsum (knot 6) in normalised time, as a function of item and gender

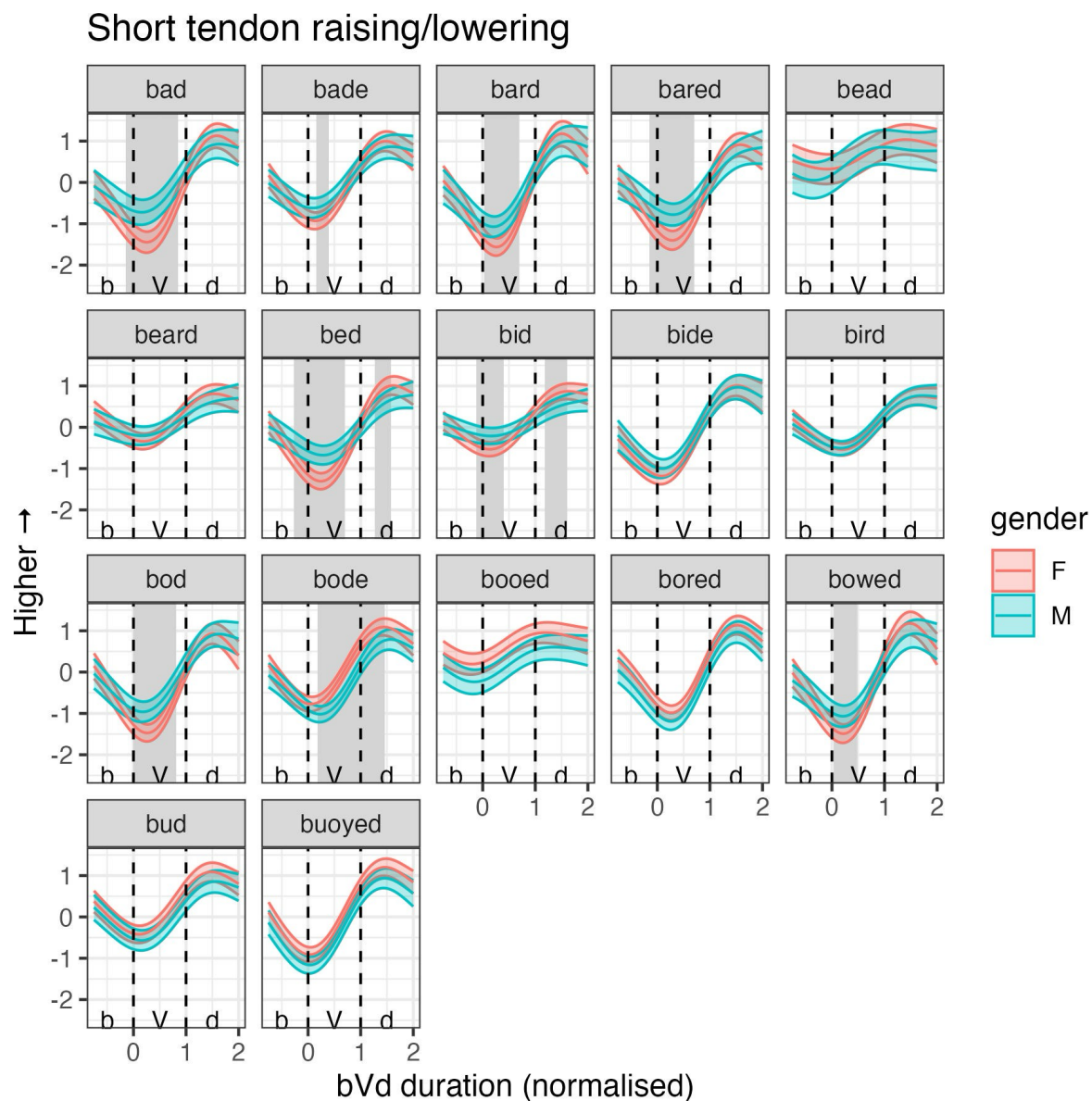


Figure 7: GAMM predictions for vertical displacement of the short tendon, as a function of item and gender

3.2 Dynamic acoustic results

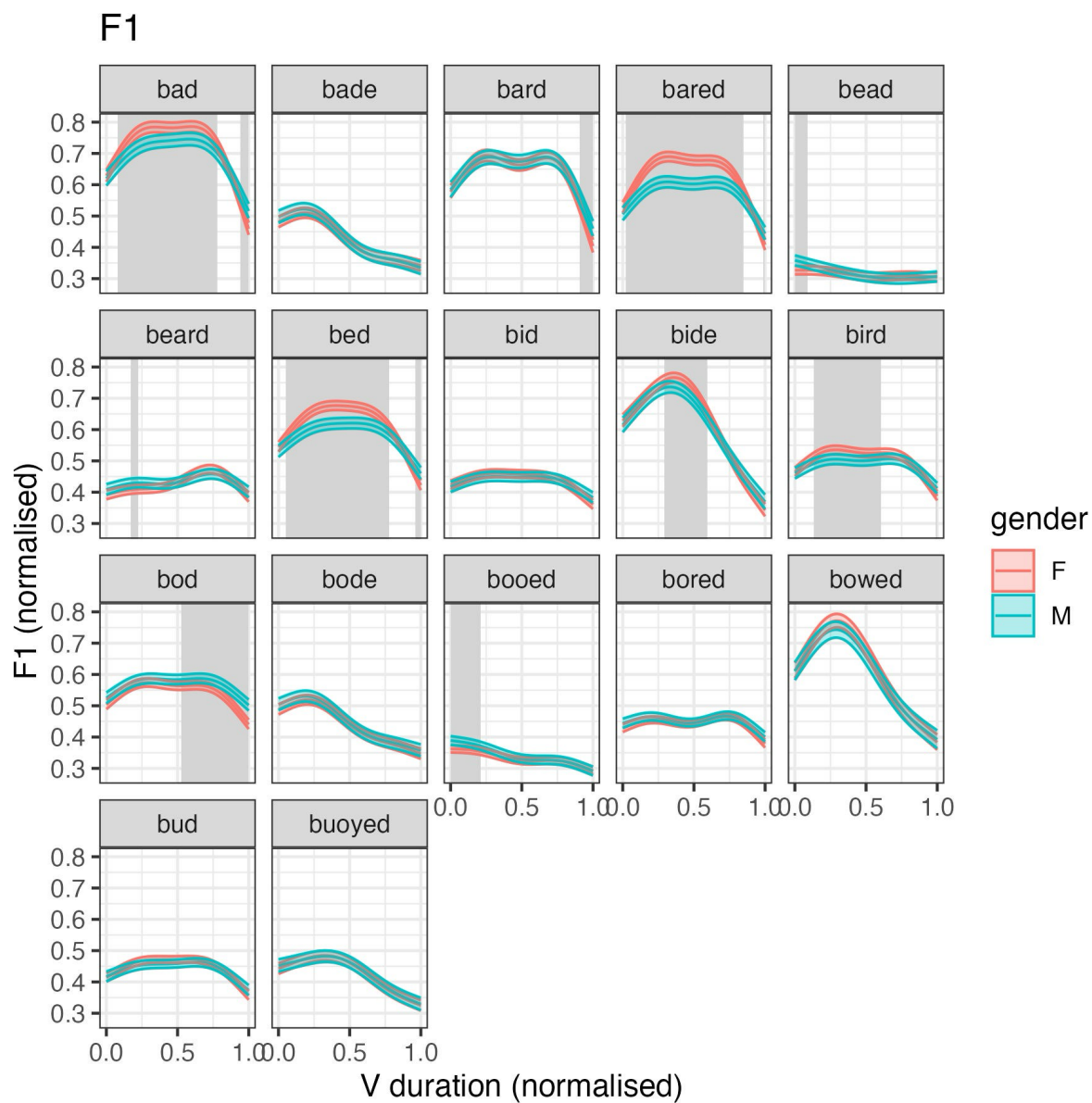
In the acoustic part of the analysis, we focus on two measures: F1 and F2. These measures were normalised for vocal tract length, using the ΔF method (Johnson, 2020), and

analysed 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, i.e. greater F1 in males compared to females for *bead, booed* and *beard*, however these differences only affect a small portion of the formant transitions. We remain cautious about interpreting these differences.

When it comes to F2, 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, buoyed*.

3.3 Summary of the results

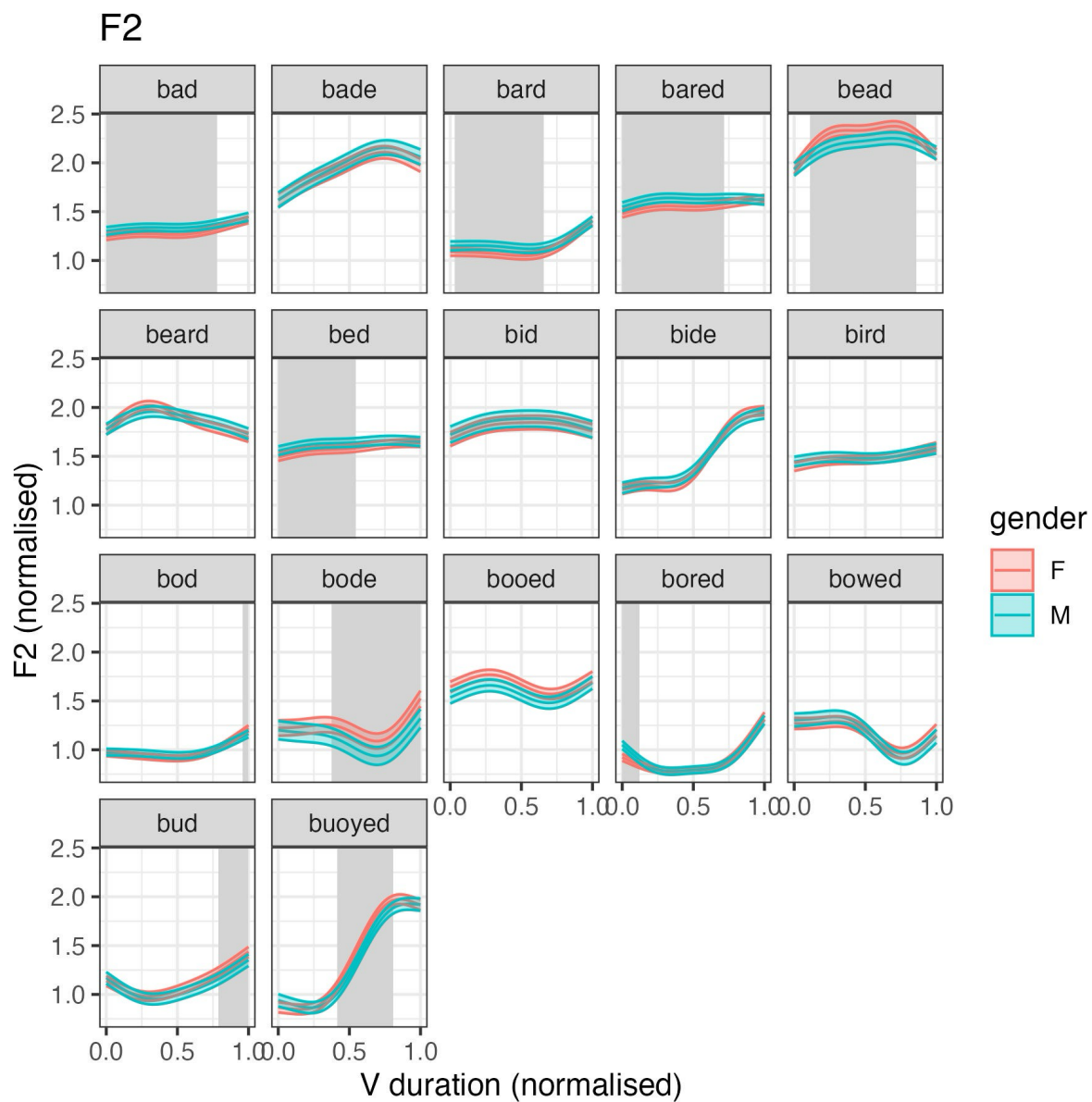
Based on the analysis presented above, there are multiple differences between males and females, both in the dynamic displacement of key articulators and in normalised F1 and F2. Table 2 presents a summary of the significant differences we have found. We have excluded differences that only affect C1 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.



451

452 Figure 8: GAMM predictions for F1 change over vowel duration, as a function of item and

453 gender



454

455 Figure 9: GAMM predictions for F2 change over vowel duration, as a function of item and

456 gender

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
<i>booed</i>	GOOSE and GOAT		advanced in F	advanced in F	lower in F	advanced in F	lower in F		
<i>bode</i>		higher in F		advanced in F	lower in F	advanced in F			higher in F
<i>bad</i>	front mid low vowels	lower in F		retracted in F		retracted in F		higher in F	lower in F
<i>bard</i>		lower in F		retracted in F	higher in F	retracted in F	higher in F		lower in F
<i>bared</i>		lower in F	retracted in F	retracted in F		retracted in F		higher in F	lower in F
<i>bed</i>		lower in F	retracted in F	retracted in F		retracted in F		higher in F	lower in F
<i>bowed</i>		lower in F				retracted in F			
<i>bod</i>	lax back vowels	lower in F			higher in F	retracted in F	higher in F		
<i>bud</i>				advanced in F		advanced in F			
<i>bide</i>	remaining vowels				higher in F		higher in F	higher in F	
<i>buoyed</i>				advanced in F		advanced in F			higher in F
<i>bade</i>		lower in F					lower in F		
<i>bead</i>							lower in F		higher in F
<i>beard</i>									
<i>bid</i>		lower in F							
<i>bird</i>								higher in F	
<i>bored</i>				advanced in F					

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 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 to males, and *bode* also has more short tendon lowering in males, compared to 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 to 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 difference in the position of the tongue root, such that the tongue root is more retracted

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.

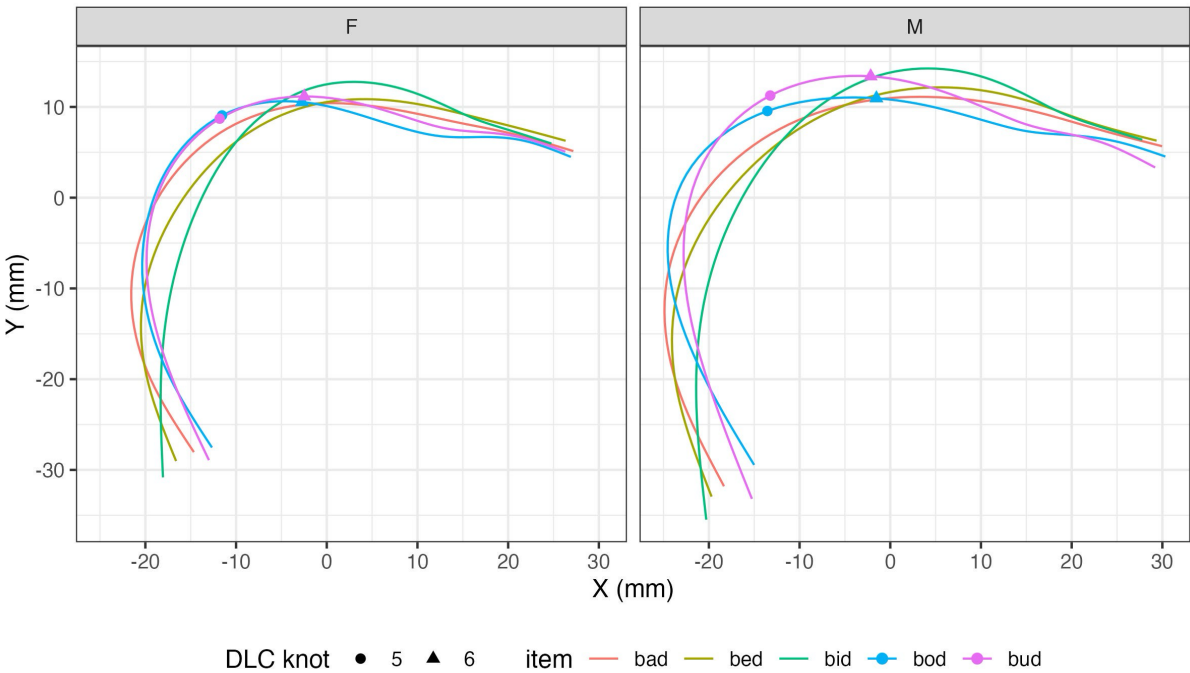
in females and more advanced in males. The gender differences are less pronounced when it comes to tongue dorsum height. The tongue dorsum is higher for female speakers in *bard*, but not in other vowels. These articulatory differences generally have some acoustic effect that persists despite vocal tract normalisation. We see higher normalised F1 in females for *bad*, *bared*, and *bed*. Furthermore, all vowels in this group, except *bowed* have lower normalised F2 in females, compared to 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: normalised F1 and F2 are not significantly different between males and females in either *bod*, or *bud*.

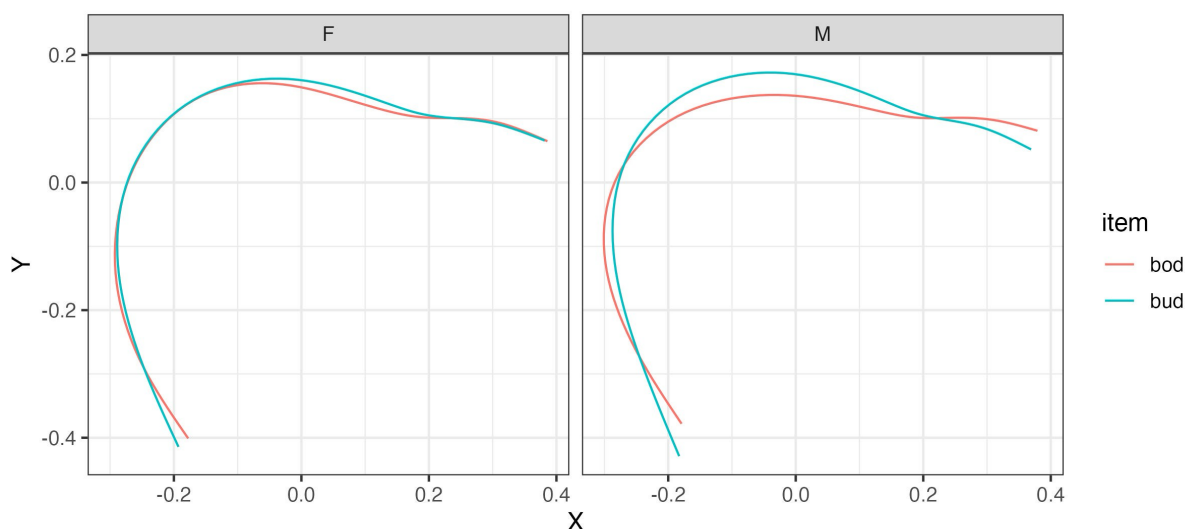
In order to understand the gender difference affecting articulation of *bod* and *bud*, it is helpful to consider the overall tongue contours, which complements our understanding of the kinematic data that focused on the displacement of individual articulators. Figure 10 illustrates mean tongue contours 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 centred within 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 GAMs (Coretta and 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 centred, scaled and rotated to the same plane using Procrustes analysis, using the procedure in Dryden and Mardia (2016). This normalisation 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 normalisation, 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.



518 Figure 10: By-gender mean tongue contours for the lax vowels, taken at the acoustic midpoint.
 519 The data are centred within speaker and averaged using GAM



520
 521 Figure 11: Comparison of female and male mean tongue shape in *bod* and *bud*. The data are
 522 rotated, centred and scaled within speaker and averaged using GAM

523
 524 In Section 4 below, we discuss the possible explanations for the patterns of gender
 525 differences affecting the three groups of vowels, as identified above. Additionally, we have
 526 observed some differences in several other vowels. For example, we observe some short
 527 tendon lowering in female *bid* (a high-mid vowel), which has no reflection in F1 or F2.

These types of differences, summarised in Table 2, do not form any coherent pattern that we can identify, so we shall refrain from interpreting them.

4 Discussion

4.1 Gender-specific articulatory strategies in different vowels

Our data deliver evidence of multiple differences between male and female strategies for producing vowel sounds. While not all vowels involve gender differences in production, many of them do. Moreover, the difference we observe are systematic, such that groups of vowels that share specific articulatory properties pattern together with respect to gender differences in production.

4.2.1 GOOSE and GOAT

The first group of gender differences we identify concern the vowels in *booed* and *bode*. In *booed* and *bode*, we find greater tongue dorsum advancement in female speakers, and in *bode* higher normalised 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, both in the acoustic and articulatory domain (Lawson et al., 2019; Strycharczuk et al., 2020). However, comparing with recent data on Southern English, as reported in Cole and Strycharczuk (2024), GOOSE -fronting is somewhat more limited in the North, and presumably 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 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 to the South of England (Strycharczuk et al., 2020; Cole and Strycharczuk, 2024). Therefore, the gender difference we find here is likely a manifestation of a female-led sound change.

4.2.1 Mid and front low vowels

Secondly, 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 speakers, and also with 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 retraction in females, compared to 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 and Sundberg, 1971). The acoustic effects we observe are consistent with well-established patterns of articulatory-acoustics relationship: jaw lowering is known to produce F1 raising, whereas dorsal retraction lowers F2 (Stevens and House, 1955; Fant, 1971).

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 normalisation (see Figure 8). We can be fairly confident that these differences arise from distinct articulatory strategies that involve greater degree of jaw opening in females and 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 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 generalise 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.

4.2.1 Lax back vowels

Similarly to mid and front low vowels, we find greater short tendon lowering in female *bod*, compared to 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 neighbouring 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 normalised F1 or F2 trajectories in male and female *bud*.

The combined information from GAMM modelling reported in Section 3, 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 velarisation in *bud*. These distinct strategies are compatible with a differentiation between ‘tongue movers’ and ‘jaw movers’ (Johnson et al., 1993; Johnson, 2023). While vowel production clearly involves the contribution of both tongue and 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 normalised 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 to 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 the light of some of the previous explanations as to why male-female differences in jaw opening arise in the first place. As discussed in Section 1, 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 directly 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 linguo-pharyngeal 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 generalise them to a wider context.

4.2 The role of articulatory setting

The idea of a generalisation underlines the concept of articulatory setting, as developed by Laver (1980). Articulatory setting can be broadly described as habitual positioning of articulators that characterises 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, 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, i.e. 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, 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 evidence as a specific range of jaw movement, relatively greater in females (because their setting is relatively open), and relative 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 to back vowels, producing acoustic variation in the presence of articulatory stability.

4.3 Articulatory target uniformity

Another possible mechanism for a generalisation 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 and Wilson, 2017, 2022). It is plausible that speakers develop some specific motor routines for the production of some vowels, and re-use the same routines in a range of related vowel contexts. This type of behaviour is economical from the speech production point of view, as speakers can rely on the same highly practiced movements (Maddieson, 1996). Existing evidence for target uniformity includes within-speaker stability in the acoustic realisation of VOT (Chodroff and Wilson, 2017), and fricatives (Chodroff and 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 realisation of phonological features, and acoustic uniformity is a secondary reflection of that. McAllister Byun, Inkelas, & Rose (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 realisation 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 (2022), in that relative articulatory stability can in some cases be linked to acoustic variation.

4.4 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 to 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 centred values measured in mm). As shown in Figure 10, the mean tongue size for the male speakers was larger than 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*, nor for the most retracted vowel *bored*. In terms of tongue dorsum raising, males show more lowering in the relatively lowest vowel *bard*, potentially signalling a somewhat greater range, but still, the difference is limited.

These observations point to the presence of more undershoot in male speakers, similar as reported in 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 they male speakers produced greater displacement in accented condition, compared to 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.

4.5 Clinical implications

There are several clinical implications that follow from our findings. Gender-specific vowel resonances have been recognised as a potential goal for gender affirming speech

therapy, i.e. therapy that helps transgender people to produce speech that aligns with their gender identity (Leung et al., 2018; Kawitzky and McAllister, 2020; 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 Hirsh (2017) propose more anterior tongue placement and lip spreading as an effective strategy to raise formants and feminise voice. Leung et al. (2018) suggests 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”. In general, these strategies can be expected to raise f_0 , F1 and F2. Larynx raising is expected to raise f_0 , 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 normalised 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 normalised F2 in some low vowels produced by female speakers, compared to 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.

5 Conclusion

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 potentially stable pattern of variation characterised 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 generalisation is at play, which can be conceptualised 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, e.g. 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, and not just individual sounds. Our findings also have possible applications for Gender Affirming Voice Therapy, as they confirm a jaw-based strategy for low vowel production in female speakers.

Data availability statement

The data and code used in the analysis are publicly available through an Open Science Project repository https://osf.io/zcet3/?view_only=b28eb776f37f46ceb5ce7034448128ac (URL to be replaced by a permanent link)

Acknowledgments

The research reported in this paper was funded by the by the Arts and Humanities Research Council grant no. AH/S011900/1 and the British Academy grant no. MFSS24\40076. SK was supported by AHRC fellowship AH/Y002822/1. 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*, 33:27–53.
- Boersma, P. and Weenink, D. (2009). Praat: doing phonetics by computer [Computer programme]. Version 6.2.14.
- Byun, T. M., Inkelas, S., & Rose, Y. (2016). The A-map model: Articulatory reliability in child-specific phonology. *Language*, 92, 141-178.
- Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech communication*, 15:39–54.
- 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*, 21:591-603.
- Chodroff, E. and Wilson, C. (2017). Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. *Journal of Phonetics*, 61:30–47.
- Chodroff, E. and Wilson, C. (2022). Uniformity in phonetic realization: Evidence from sibilant place of articulation in American English. *Language*, 98:250–289.
- Cole, A. and Strycharczuk, P. (2024). The search for linguistically coherent accents: Unsupervised clustering of diphthong variation in southeast England. *English WorldWide*, 45:1–29.

813 Coretta, S. and Sakr, G. (2024). Analysing tongue contours with multivariate Generalised
814 Additive Models. In Wilson, I., Mizoguchi, A., Perkins, J., Villegas, J., and Yamane, N.,
815 editors, *Ultrafest XI: Extended Abstracts*, pages 100–101. Zenodo.

816 Dryden, I. L. and Mardia, K. V. (2016). *Statistical shape analysis: with applications in R*. John
817 Wiley & Sons.

818 Fant, G. (1966). A note on vocal tract size factors and non-uniform F-pattern scalings.
819 *Speech Transmission Laboratory Quarterly Progress and Status Report*, 1:22–30.

820 Fant, G. (1971). *Acoustic theory of speech production: with calculations based on X-ray*
821 *studies of Russian articulations*. Walter de Gruyter.

822 Fant, G. (1975). Non-uniform vowel normalization. *Department for Speech, Music and*
823 *Hearing. Quarterly Progress and Status Report*, 16:1–19.

824 Faytak, M. (2022). Place uniformity and drift in the Suzhounese fricative and apical
825 vowels. *Linguistics Vanguard*, 8:569–581.

826 Fruehwald, J. (2019). Is phonetic target uniformity phonologically, or sociolinguistically
827 grounded. In Calhoun, S., Escudero, P., Tabain, M., and Warren, P., editors, *Proceedings*
828 *of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*,
829 pages 681–85.

830 Goldstein, U. G. (1980). *An articulatory model for the vocal tracts of growing children*. PhD
831 thesis, Massachusetts Institute of Technology.

832 Haddican, B., Foulkes, P., Hughes, V., and Richards, H. (2013). Interaction of social and
833 linguistic constraints on two vowel changes in northern England. *Language Variation*
834 *and Change*, 25:371–403.

835 Heffernan, K. (2010). Mumbling is macho: Phonetic distinctiveness in the speech of
836 American radio DJs. *American Speech*, 85:67–90.

837 Henton, C. (1995). Cross-language variation in the vowels of female and male speakers.
838 In Elenius, K. and Branderud, P., editors, *Proceedings of the XIIIth International Congress*
839 *of Phonetic Sciences*, volume 4, pages 420–423.

840 Hirsch, S. (2017). Combining voice, speech science and art approaches to resonant
841 challenges in transgender voice and communication training. *Perspectives of the ASHA*
842 *Special Interest Groups*, 2:74-82.

843 Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social
844 identity and phonology. *Journal of Phonetics*, 34:485–499.

845 Johnson, K. (2020). The δf method of vocal tract length normalization for vowels.
846 *Laboratory Phonology*, 11.

847 Johnson, K. (2023). Individual differences in speech production: What is “phonetic
848 substance”? In Skarnitzl, R. and Volín, J., editors, *Proceedings of the 20th International*
849 *Congress of Phonetic Sciences*, pages 1102–1106.

850 Johnson, K., Ladefoged, P., and Lindau, M. (1993). Individual differences in vowel
851 production. *The Journal of the Acoustical Society of America*, 94:701–714.

852 Kawitzky, D. and McAllister, T. (2020). The effect of formant biofeedback on the
853 feminization of voice in transgender women. *Journal of Voice*, 34:53–67.

854 Kirkham, S., Strycharczuk, P., Gorman, E., Nagamine, T., and Wrench, A. (2023). Co-
855 registration of simultaneous high speed ultrasound and electromagnetic

856 articulography for speech production research. In Skarnitzl, R. and Vol'in, J., editors,
857 *Proceedings of the 20th International Congress of Phonetic Sciences*, pages 918–922.

858 Labov, W. (1990). The intersection of sex and social class in the course of linguistic change.
859 *Language variation and change*, 2:205–254.

860 Labov, W. (1994). *Principles of linguistic change: Internal factors*, volume 1. Oxford:
861 Blackwell.

862 Labov, W. (2001). *Principles of linguistic change. Social factors*, volume 2. Oxford:
863 Blackwell.

864 Lammert, A. C. and Narayanan, S. S. (2015). On short-time estimation of vocal tract length
865 from formant frequencies. *PloS one*, 10:e0132193.

866 Laver, J. (1980). *The Phonetic Description of Voice Quality*. Cambridge: Cambridge
867 University Press.

868 Lawson, E., Stuart-Smith, J., and Rodger, L. (2019). A comparison of acoustic and
869 articulatory parameters for the GOOSE vowel across British Isles Englishes. *The Journal*
870 *of the Acoustical Society of America*, 146:4363–4381.

871 Lee, S., Potamianos, A., and Narayanan, S. (1999). Acoustics of children's speech:
872 Developmental changes of temporal and spectral parameters. *The Journal of the*
873 *Acoustical Society of America*, 105(3):1455–1468.

874 Leung, Y., Oates, J., and Chan, S. P. (2018). Voice, articulation, and prosody contribute to
875 listener perceptions of speaker gender: A systematic review and meta-analysis. *Journal*
876 *of Speech, Language, and Hearing Research*, 61:266–297.

877 Leyns, C., Corthals, P., Cosyns, M., Papeleu, T., Van Borsel, J., Morsomme, D., T'Sjoen, G., and
878 D'haeseleer, E. (2024). Acoustic and perceptual effects of articulation exercises in
879 transgender women. *Journal of Voice*, 38:246–e15.

880 Lindblom, B. E. and Sundberg, J. E. (1971). Acoustical consequences of lip, tongue, jaw,
881 and larynx movement. *The Journal of the Acoustical Society of America*, 50:1166– 1179.

882 Maddieson, I. (1996). Gestural economy. *Proceedings of the 13th International Congress of*
883 *Phonetic Sciences (ICPhS), Stockholm*, page 574–78.

884 Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., and Bethge, M.
885 (2018). Deeplabcut: markerless pose estimation of user-defined body parts with deep
886 learning. *Nature neuroscience*, 21(9):1281–1289.

887 Ménard, L., Schwartz, J.-L., and Aubin, J. (2008). Invariance and variability in the
888 production of the height feature in French vowels. *Speech communication*, 50:14–28.

889 Munson, B. and Babel, M. (2019). The phonetics of sex and gender. In *The Routledge*
890 *handbook of phonetics*, pages 499–525.

891 Nearey, T. M. (1989). Static, dynamic, and relational properties in vowel perception. *The*
892 *Journal of the Acoustical Society of America*, 85:2088–2113.

893 Nearey, T. M. and Assmann, P. F. (1986). Modeling the role of inherent spectral change in
894 vowel identification. *The Journal of the Acoustical Society of America*, 80:1297–1308.

895 Nordström, P.-E. (1977). Female and infant vocal tracts simulated from male area
896 functions. *Journal of Phonetics*, 5(1):81–92.

- Serrurier, A. and Neuschaefer-Rube, C. (2024). Formant-based articulatory strategies: Characterisation and inter-speaker variability analysis. *Journal of Phonetics*, 107:101374.
- 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):2153–2164.
- Simpson, A. P. (2002). Gender-specific articulatory–acoustic relations in vowel sequences. *Journal of Phonetics*, 30(3):417–435.
- Simpson, A. P. (2009). Phonetic differences between male and female speech. *Language and Linguistics Compass*, 3:621–640.
- Soskuthy, M. (2021). Evaluating generalised additive mixed modelling strategies for dynamic speech analysis. *Journal of Phonetics*, 84:101017.
- Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, 17:3–45.
- Stevens, K. N. and House, A. S. (1955). Development of a quantitative description of vowel articulation. *The Journal of the Acoustical Society of America*, 27:484–493.
- Strycharczuk, P., Kirkham, S., Gorman, E., and Nagamine, T. (2024). Towards a dynamical model of English vowels. Evidence from diphthongisation. *Journal of Phonetics*, 107:101349.
- Strycharczuk, P., Kirkham, S., Gorman, E., and Nagamine, T. (2025). Dimensionality reduction in lingual articulation of vowels: Evidence from lax vowels in Northern Anglo-English. In press in *Language and Speech*.

- Strycharczuk, P., L'opez-Ib'an~ez, M., Brown, G., and Leemann, A. (2020). General Northern English. Exploring regional variation in the North of England with machine learning. *Frontiers in Artificial Intelligence*, page 48.
- Traunmüller, H. (1984). Articulatory and perceptual factors controlling the age-and sex-conditioned variability in formant frequencies of vowels. *Speech Communication*, 3:49–61.
- Turton, D. and Lennon, R. (2023). An acoustic analysis of rhoticity in Lancashire, england. *Journal of Phonetics*, 101:101280.
- van Rij, J., Wieling, M., Baayen, R. H., and van Rijn, H. (2015). itsadug: Interpreting time series and autocorrelated data using gamms. R package version 1.0.3.
- Vorperian, H. K. and Kent, R. D. (2007). Vowel acoustic space development in children: A synthesis of acoustic and anatomic data. *Journal of Speech, Language, and Hearing Research*, 50:1510–1545.
- Weirich, M., Fuchs, S., Simpson, A., Winkler, R., and Perrier, P. (2016). Mumbling: Macho or Morphology? *Journal of Speech, Language, and Hearing Research*, 59.
- Weirich, M. and Simpson, A. P. (2018a). Gender identity is indexed and perceived in speech. *PLoS One*, 13(12):e0209226.
- Weirich, M. and 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.
- Wells, J. (1982). *Accents of English 1: An introduction*, volume 2. Cambridge: Cambridge University Press.

939 Westbury, J. R., Hashi, M., and J Lindstrom, M. (1998). Differences among speakers in
 940 lingual articulation for American English /ô/. *Speech Communication*, 26:203–226.

941 Whiteside, S. P. (2001). Sex-specific fundamental and formant frequency patterns in a
 942 cross-sectional study. *The Journal of the Acoustical Society of America*, 110:464–478.

943 Wood, S. (2017). *Generalized additive models: an introduction with R*. Chapman and
 944 Hall/CRC, 2nd edition.

945 Wrench, A. and Balch-Tomes, J. (2022). Beyond the edge: Markerless pose estimation of
 946 speech articulators from ultrasound and camera images using DeepLabCut. *Sensors*,
 947 22:1133.