

Articulatory strategy as a source of variation in acoustic vowel dynamics

Patrycja Strycharczuk,^{1,a)}  Justin J. H. Lo,²  and Sam Kirkham² 

¹Linguistics and English Language, University of Manchester, Manchester, United Kingdom

²Linguistics and English Language, Lancaster University, Lancaster, United Kingdom

ABSTRACT:

Acoustic vowel dynamics have some speaker-identifying characteristics, which have been ascribed to individual properties of articulatory strategies: formant transitions have a particular shape because speakers move their articulators, using specific and practised movements. However, there is little existing evidence that different articulatory strategies systematically affect formant dynamics. The present study corroborates the link between the two. Ultrasound tongue imaging data from 36 speakers of Northern-Anglo English are used to identify distinct articulatory strategies for the production of palatal vowel /i/. Tongue shape in /i/ is found to be a significant predictor of formant dynamics in diphthongs with a palatal offglide. The observed relationships can be explained by the characteristics of articulatory movement conditioned by vocal tract shape. Greater articulatory displacement of tongue root and/or dorsum produces greater distortion from the mean tongue shape in palatal vowels, and it also requires higher articulatory velocities, resulting in relatively earlier and steeper formant transitions. The results contribute to the conceptual understanding of individuality in speech, by illuminating the regularising and individual aspects of articulatory compensation. © 2026 Acoustical Society of America.

<https://doi.org/10.1121/10.0043814>

(Received 29 August 2025; revised 7 April 2026; accepted 15 April 2026; published online 14 May 2026)

[Editor: Steven M. Lulich]

Pages: 4068–4078

I. INTRODUCTION

A. Speaker-specificity of vowel dynamics

Vowel dynamics are highly individual. Multiple studies show that measures of inherent change in formant values contain some speaker-identifying information, and the performance of speaker-recognition models tends to improve when dynamic information is included, compared to models based on static formant measurements (Heeren, 2020; Hughes *et al.*, 2016; McDougall, 2006; Morrison, 2009; Rhodes, 2012; Rose *et al.*, 2006). Similar improvements are observed when acoustic information is enriched with articulatory information (Kirchhoff *et al.*, 2002). Furthermore, it has been observed that the contribution of dynamic information to speaker-specificity is relatively greater for more diphthongal vowels and more limited when there is less inherent change within the vowel (Heeren, 2020; Hughes *et al.*, 2016).

A prominent line of explanation for the speaker-specific characteristics of formant dynamics focuses on individual differences in articulatory movement used to achieve specific acoustic targets (McDougall, 2004, 2006; McDougall and Nolan, 2007; Nolan and Grigoras, 2005). According to Nolan's (1983) model of sources of between-speaker variation, given the same set of phonetic representations, speakers have the flexibility to reach their primary auditory goals through different implementational rules, which are

themselves informed and constrained by the configurational and dynamic properties of the vocal tract and articulators. The articulatory strategy for vowel production varies from speaker to speaker (Johnson *et al.*, 1993), reflecting both anatomical differences and practised articulatory routines. It has been hypothesised that these types of individual characteristics are more directly observable in movement between articulatory targets. Possible reasons for the increased variability of transitions include the perceptually privileged status of targets (McDougall and Nolan, 2007; Morrison, 2009) and quantal properties of speech (Stevens, 1989), whereby phonological systems preferably select phonetic variants that are characterised by acoustic stability in the presence of articulatory variance. The same preference does not extend to movements between targets, hence the prediction that transitions will allow for greater acoustic variability (Nolan and Grigoras, 2005).

The hypothesis that articulatory strategy contributes to individual variation in acoustic vowel dynamics has inspired a large body of forensic phonetic research, but it has not been directly tested. Some indirect support for the articulatory basis of individual variation in vowel dynamics come from phonetic studies of identical twins. Weirich (2012) reports different levels of similarity between monozygotic and heterozygotic twins and that these interact with the type of measure: static vs dynamic. Greater similarity is found for monozygotic twins but only for dynamic measures (consonant-vowel transitions) and not for static measures. These findings can be attributed to a shared articulatory strategy in

^{a)}Email: patrycja.strycharczuk@manchester.ac.uk

twins, under the assumption that such strategy is shaped by anatomical differences. Furthermore, since differences are observed in transitions but not in static targets, this lends support to the hypothesis that acoustic transitions are more likely to reflect individual articulatory variation than target measures.

B. Relationship between articulatory strategy and acoustic variation

There is scarce evidence in the literature that could explicitly link specific articulatory strategies to specific acoustic differences in vowels. [Noiray et al. \(2014\)](#) show that differences in tongue body height for the production of the /i/-e/ contrast in American English translate into systematic differences in F1. Several studies observe articulatory differences between males and females, such as increased jaw opening in females ([Strycharczuk and Kirkham, 2025](#); [Weirich et al., 2016](#)) and tendency for articulatory undershoot in males ([Weirich and Simpson, 2018](#)), which are linked to acoustic differences, especially larger acoustic vowel spaces in females. However, the relevant examples are relatively few, whereas a large body of articulatory literature emphasises the relative acoustic invariance produced by diverging articulatory strategies. General plasticity of articulatory strategy is evidenced, for instance, by findings on articulatory compensation triggered by mechanical or auditory perturbation. Speakers whose normal articulation is disrupted, such as by the presence of a bite block, adapt their articulation to approximate their habitual acoustic target ([Gay et al., 1981](#); [McFarland and Baum, 1995](#)). Similar adaptation is observed, in some speakers, when the acoustic feedback received by the speaker is altered experimentally ([Guenther, 2016](#); [Houde and Jordan, 1998](#)). These types of adaptation highlight the importance of acoustic invariance over articulatory habit. Classical theories of speech perception postulate a key role of acoustic invariance in phonetics-phonology mappings and establishing a connection between sound and meaning ([Blumstein and Stevens, 1979](#); [Lisker, 1985](#)), although identifying such regular mappings has often proven challenging ([Redford and Baese-Berk, 2023](#)). From the sociolinguistic point of view, production of similar acoustic targets serves an indexical role, such as to signal social category membership ([Eckert, 2008](#); [Foulkes and Docherty, 2006](#)).

Acoustic invariance is thought to be the main factor driving the individual differences in articulatory strategy: speakers need to overcome the individual differences in their vocal tract shape and size in order to produce similar sounds. The existence of compensatory mechanisms in speech is evidenced by the relationship between vocal tract morphology and articulatory strategy. [Serrurier and Neuschaefer-Rube \(2024\)](#) establish several such correlations. For example, they find a relationship between the horizontal and vertical size of the vocal tract and the primary vector of tongue movement used for achieving formant manipulations. Speakers with relatively larger vocal tracts tend to rely more on anterior-posterior tongue displacement,

whereas smaller vocal tracts are associated with more upward-downward displacement. Furthermore, speakers with relatively flatter and more posterior palates tend to achieve F1 manipulations by jaw movement, whereas speakers with more domed and more anterior palates rely more on tongue movement. A similar relationship between palate shape (specifically doming) and jaw movement is also observed by [Johnson \(2023\)](#).

Compensation for palate shape is well documented in the production of palatal vowels, especially /i/.¹ [Hasegawa-Johnson et al. \(2003\)](#) demonstrate that the height of the tongue is closely correlated with palate height for palatal vowels /i, e, ɪ, ε/, but there is no such correlation in velar, pharyngeal, or uvular vowels. Tongue height and palatal vault height were defined as the vertical distance from the line connecting the maxillary gingival margins. Furthermore, the area between the palate and tongue is more stable across speakers in palatal vowels, compared to other vowels.² These findings suggest that controlling the degree of vocal tract narrowing is important for achieving the acoustic target in high vowels.

[Lammert et al. \(2013a,b\)](#) provide further evidence for the systematic relationship between tongue shape in high vowels and the shape of the palate. They identify three parameters that systematically capture individual variation in palate shape: palate concavity (height of the palate apex), anteriority (position of the palate apex), and sharpness (shape at the palate apex). According to modelling in [Lammert et al. \(2013a\)](#), this type of morphological variation has the scope to affect the first three vowel formants substantially. For example, the frequency of F1 increases with greater palate concavity, while the frequency F2 decreases. However, the corresponding acoustic variation is not detected in the speech of individuals with different palate shapes. Instead, speakers adjust the shape of their tongue to compensate for the variation in palate shape; for example, speakers with more concave palates produce more tongue raising in the palatal region. A correlation between palate shape and mean tongue shape is confirmed by [Serrurier and Neuschaefer-Rube \(2023\)](#) in a larger sample of 41 speakers. [Brunner et al. \(2009\)](#) observe a systematic relationship between articulatory and acoustic variability and palate shape. Specifically, they find that speakers with (cross-sectionally) flat palates show less articulatory variability compared to speakers with domed palates. [Brunner et al. \(2009\)](#) propose that this is because the same degree of tongue movement affects the area functions of the vocal tract to a greater degree when the palate is flatter, due to a proportionally smaller vocal tract. Thus, acoustic invariance appears to be driving variation in articulatory strategy through the mechanism of compensation.

C. Present study

As we have seen, the speaker-specificity of vowel dynamics is an established empirical observation, and it has been attributed to individual differences in articulatory

strategy, where such differences may be partially due to compensation for individual differences in vocal tract morphology. On the other hand, existing research on articulatory variation and articulation-acoustics relationship points in a rather different direction, emphasising that articulatory variation largely serves to reduce individual acoustic differences. However, evidence for acoustic compensation is mainly based on static measurements. While compensatory behaviour is undoubtedly an important factor in shaping the speaker's strategy for achieving a specific acoustic target, little is known about its effect on the dynamic properties of the acoustic signal. Dynamics are crucial to consider, given the forensic findings that acoustic dynamics contribute to speaker-specificity and given the underlying role of articulatory movement. Individual articulatory strategies involve varying degrees of displacement for different articulators. Articulators vary inherently according to their velocity, due to differences in plasticity and mass. Thus, we can expect that acoustic transitions towards a fixed target might reflect different dynamic properties of the key articulators involved.

The present study examines the role of articulatory strategy as a factor conditioning vowel dynamics, through a case study of I-diphthongs in Northern-Anglo English /i, ei, ai, oi/. Note that although we use the /i/ notation for convenience, we consider /i/ to be a diphthong, because it can have a diphthongal quality [ii] and because it behaves like a diphthong structurally (Strycharczuk *et al.*, 2024). Our study is based on ultrasound and acoustic data from 36 speakers in the TarDiS (Targets and Dynamics in Speech) corpus (Strycharczuk *et al.*, 2024, 2025).

In general, diphthongs offer a compelling case study because they are inherently dynamic. Furthermore, English diphthongs form a coherent subsystem, which creates an opportunity to examine a wide range of different articulatory displacements, from a small degree of change in [ii] to traversing a large portion of the vocal tract in [ai]. At the same time, all these vowels have an offglide with a similar quality, which allows us to compare whether transitions towards the offglide vary in a similar way across different starting points.

We investigate whether the speaker's formant dynamics in /i/ vowels and diphthongs are systematically affected by their articulatory strategy for /i/ production. As a proxy for this strategy, we use a normalised measure of midsagittal tongue shape in /i/. Specifically, we test whether tongue shape in /i/ is a significant predictor of formant trajectory shape. This is only one of many possible predictors. For example, formant dynamics as well as articulatory strategy are systematically affected by gender (Strycharczuk and Kirkham, 2025). Here, we focus on /i/ as a predictor because its relationship to vocal tract morphology is well described (Hasegawa-Johnson *et al.*, 2003; Lammert *et al.*, 2013a). As a high front vowel, /i/ is systematically affected by the size of the vocal tract, as well as by palate shape (see Sec. IB). Thus, we are able to rely on previous literature in interpreting variation in /i/ shape in terms of vocal tract morphology.

Another factor that makes /i/ a reasonable predictor is its similarity to the diphthong offglide in other I-diphthongs. Because of the similarity, speakers presumably employ an /i/-like strategy in their transition from the diphthong onglide to the offglide (we test this assumption in Sec. III A). Acoustic compensation may be harder to achieve in such cases compared to /i/, in which there is less inherent change and the tongue shape is relatively stable over a longer time window.

The reason why we use the tongue shape for /i/ to predict formant trajectories and not the actual tongue shape produced in any given vowel at any given time point is to help us tease apart strategy from inter-token variability. We assume that strategy relates to a practised articulatory movement that is relatively stable within speaker (for similar sounds) and different between speakers. However, different speakers may produce different degrees of diphthongisation even when performing the same task, and articulatory variation may also arise within speaker (e.g., some speakers may produce more vowel reduction as the experiment goes on). The presence of such variation can result in significant correlations between articulatory and acoustic measures, which are, however, relatively trivial (e.g., greater articulatory displacement associated with greater formant displacement). Using a by-speaker mean measure of tongue shape instead helps us isolate the effect of speaker-specific articulatory habit.

In summary, we evaluate the hypothesis that articulatory strategy conditions individuality in vowel dynamics by addressing the following research questions:

- (1) What are the different articulatory strategies used by speakers in their production of palatal vowel /i/?
- (2) Are these articulatory strategies systematically related to formant dynamics in the I-diphthongs, as produced by the same speakers?

II. METHODS

A. Corpus

We present a secondary analysis of a pre-existing corpus, TarDiS. Details of the recording procedure and data processing are described in Kirkham *et al.* (2023) and Strycharczuk *et al.* (2025). Key information relevant to the current study is summarised in the following.

The corpus comprises time-aligned midsagittal ultrasound and audio recordings from 40 speakers of English from Northern England (Greater Manchester and Lancashire). The ultrasound data were acquired using the EchoB (for ten speakers) and the Telemed Micro ultrasound speech system (for 30 speakers) from Articulate Instruments Ltd (Musselburgh, UK). The sampling rate ranged between 59.5 and 101 frames per second. The median sampling rate was 81.3 frames per second. Ultrafit was used to stabilise the probe (Spreafico *et al.*, 2018). In order to image the midsagittal plane, we positioned the transducer on the midline of the body under the participant's chin, and we checked

that the tongue contour appeared as single (not parallel) structure in the corresponding ultrasound image.

B. Stimuli

The corpus includes a full set of stressed vowels of Northern Anglo-English produced within a /b_d/ environment: *bad, bade, bard, bared, bead, beard, bed, bid, bide, bird, bod, bode, booed, bored, bowed, bud, and buoyed*. We used all the stimuli in one part of the analysis (Sec. III A). For the remainder of the analysis (Sec. III B and Sec. III C), we focused on the I-diphthong series, represented by *bead, bade, bide, and buoyed*. The stimuli were embedded within two types of carrier phrases: *She says X* and *She says X eagerly*. Up to six repetitions (typically five) were produced for each combination of stimulus and carrier phrase by each speaker.

C. Data processing

The acoustic data were forced aligned using the Montreal Forced Aligner (McAuliffe et al., 2017), using the US English MFA language model and the American English dictionary. We acknowledge that there is a difference between the training model and the variety in question, which was mitigated by subsequent manual correction.

Formant measurements were extracted at every 2 ms using FastTrack (Barreda, 2021), running on Praat version 6.2.14 (Boersma and Weenink, 2022). These measurements were then normalised using the ΔF method (Johnson, 2020). The advantage of using this normalisation method is that it takes into account the estimated length of the speaker’s vocal tract, which makes it well suited for investigating aspects of articulation-acoustics relationships.

The tongue contour imaged in the ultrasound recordings was automatically labelled using the DeepLabCut (DLC) algorithm (Mathis et al., 2018), based on a pre-trained tongue model, described in Wrench and Balch-Tomes (2022). DLC identifies 11 anatomical landmarks in each tongue contour, from the vallecula (most posterior) to the tongue tip (most anterior), which can serve as a basis for speaker normalisation (Strycharczuk et al., 2025).

D. Speakers

The speakers in the TarDiS corpus represent two localities in the North of England: Greater Manchester and Lancashire. Note that, traditionally, the Lancashire dialect and some parts of Greater Manchester may show monophthongisation of selected vowels, particularly the FACE vowel, which may be realised as [e] (Wells, 1982, p. 357). In contrast, FACE is typically an upgliding diphthong [eɪ] in the city of Manchester (Baranowski and Turton, 2015) and in General Northern English, the pan-regional standard spoken in the North of England (Strycharczuk et al., 2020; Watt, 2002). Since our hypothesis concerns the realisation of I-diphthongs, we excluded speakers with FACE monophthongisation, whom we identified using the following procedure.

We applied a discrete cosine transform (DCT) to normalised F2 trajectories (Watson and Harrington, 1999). This procedure parametrises the formant trajectory using a number of coefficients. The first three coefficients are, respectively, related to the mean formant value, the slope of the trajectory, and the trajectory’s curvature. Since the second coefficient (k_1) corresponds to the slope of the trajectory, it can be taken as a proxy for diphthongisation. If the absolute k_1 values are close to zero, this indicates a relatively flat formant trajectory, whereas values further from 0 suggest increasing diphthongisation. Figure 1 shows the by-speaker mean k_1 values for the second formant in our data. These are always negative, which means the diphthong is upgliding. However, we note a range of values, with some of them approaching 0, as would be expected in the presence of variable FACE monophthongisation. Based on visual inspection of the means, we identify an elbow in the data, which separates four speakers with greatest monophthongisation: LNCS7, MCR4, MCR17, and LNCS4. We excluded these speakers from further analysis, as they cannot be said to produce an /i/ offglide in *bade*.

The remaining analysis reflects data from 36 speakers. Of these, 19 were female, 15 male, one was non-binary, and one of unknown gender (no response to the gender question). The mean age was 25 (ranging from 18 to 48). The total number of different I-diphthong tokens analysed was 1503.

E. Analysis

Our primary research question concerns the effect of speaker’s tongue shape in /i/ on formant trajectories. To address this question, we need to identify systematic measures of tongue shape that are comparable across speakers.

In order to derive the relevant measure, we used the method previously described in Lo et al. (2025). First, we applied a generalised Procrustes analysis to the tongue contour data. As the input to the analysis, we used the tongue

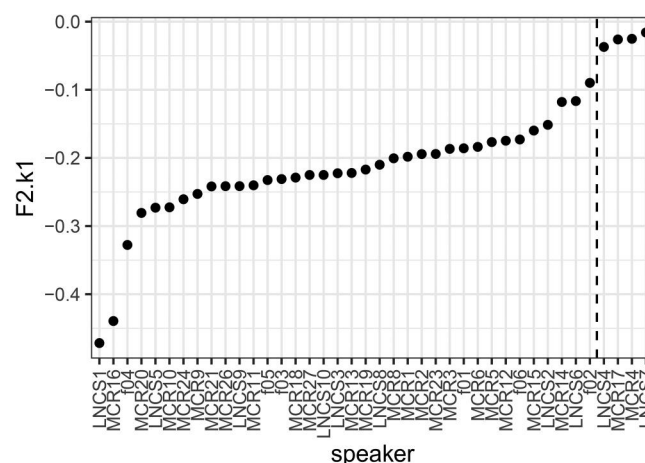


FIG. 1. By-speaker mean of the k_1 DCT coefficient extracted from the normalised F2 trajectory. Values closer to 0 indicate relatively flatter trajectories (more monophthongal), whereas values further from 0 indicate relatively steeper trajectories (more diphthongal).

contour data representing all individual ultrasound frames corresponding to the vocalic portion in the acoustic signal (e.g., all of the available ultrasonic frames for the acoustic portion of /i/ in *bead*). We used the full set of Northern English vowels here to capture the full possible array of vocalic tongue shapes. The analysis performs translation, rotation, and scaling to minimise the Procrustes distance between the tongue contours. As such, the analysis normalises for individual differences in tongue size, ultrasound probe placement, probe rotation, and movement. The sole feature that differentiates the normalised individual contours is tongue shape. We then performed a principal component analysis (PCA) in the tangent space (Kent, 1992) to identify the main orthogonal dimensions of variation in the tongue shape. This involved projecting the Procrustes-aligned landmark configurations onto the tangent space at the mean shape, which provides a linear approximation to Kendall's shape space. PCA was then applied to these tangent coordinates so that the resulting components represent interpretable modes of tongue shape variation.

For each speaker, we summarised the tongue shape for /i/ by calculating the mean of the first three principal components (PCs) (PC1–PC3) across all of their /i/ tokens, averaging over the initial 80% of each vowel. We excluded the final 20% of the vowel from the mean calculations, in order to ensure that the mean value is not overly influenced by the following /d/. We then used these by-speaker means as predictors in the statistical modelling of formant trajectories.

The potential acoustic effects of /i/ tongue shape on diphthong articulation were analysed statistically using generalised additive mixed models (GAMMs) (Wood 2017) with maximum likelihood estimation. ΔF -normalised F1 and F2 trajectories were modelled separately for each vowel in the I-diphthong series (*bead*, *bade*, *bide*, and *buoyed*), based on the same following predictor structure:

- smooth term for duration;
- smooth term for normalised time;
- smooth terms for the three /i/ PCs (i-PC1, i-PC2, and i-PC3);
- tensor product smooth interactions (τ_i smooths) for the three /i/ PCs over normalised time;
- by-speaker random intercept;
- by-speaker random smooth for normalised time;

- by-token random smooth for normalised time.

Random smooths for time-by-token were included to keep the models maximally conservative (Sóskuthy, 2021). Our main question in analysing the models is whether speakers with different mean /i/ shapes produce significantly different formant trajectories in *bead*, *bade*, *bide*, and *buoyed*.

Significance was established based on model comparison. We ran model comparisons between the full model and a nested model without i-PC1 predictors, a model without i-PC2 predictors, and a model without i-PC3 predictors. An i-PC was considered significant if its inclusion led to a model improvement at $\alpha = 0.05$.

III. RESULTS

A. PCA results

The first three PCs combined captured 79.8% of variation in tongue shape (recall that this variation is across all vowels). The first three PCs, illustrated in Fig. 2, capture variation in the degree of tongue concavity/convexity at different locations on the tongue contour, similar to Lo *et al.* (2025).

PC1 (40.5% of variance) captures variation in the convexity of tongue shape around the dorsum, centred at DLC knot 5. Higher PC1 scores are associated with more dorsal convexity, whereas lower PC1 scores denote less dorsal convexity and a somewhat flatter tongue shape. PC2 (33.2% of variance) captures variation in convexity in the anterior part of the tongue dorsum (DLC knots 6 and 7), as well as some tongue root advancement. High PC2 scores are associated with a small degree of concavity in the pre-dorsum, which is accompanied by retraction of the tongue root. In contrast, low PC2 scores are associated with shapes that have arching in the anterior part of the dorsum. Variation in PC3 (6.2% of variance) is related to variation between convexity (low PC3 scores) and concavity (high PC3 scores) in the tongue mid, centred at DLC knot 8. PCA predictions for tongue variation in /i/ are plotted in Fig. 3. While the general interpretation of the PCs remains the same, the figure visualises the range of variation [within 1 standard deviation (SD)] associated with different articulatory strategies for /i/ in our data.

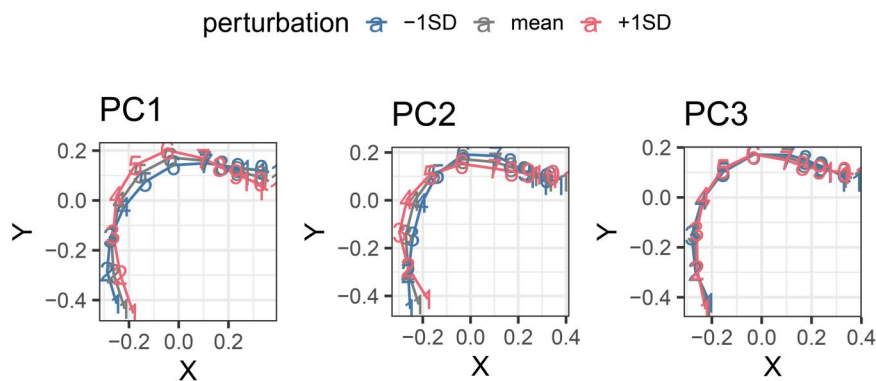


FIG. 2. The effect of PC scores on the variation in tongue shape. The gray line shows the mean tongue shape across all speakers and all vowels. Shapes with relatively higher PC scores (by 1 SD) are plotted in red. Shapes with relatively lower PC scores (by 1 SD) are plotted in blue. The numbers on the tongue contours indicate the location of DLC knots.

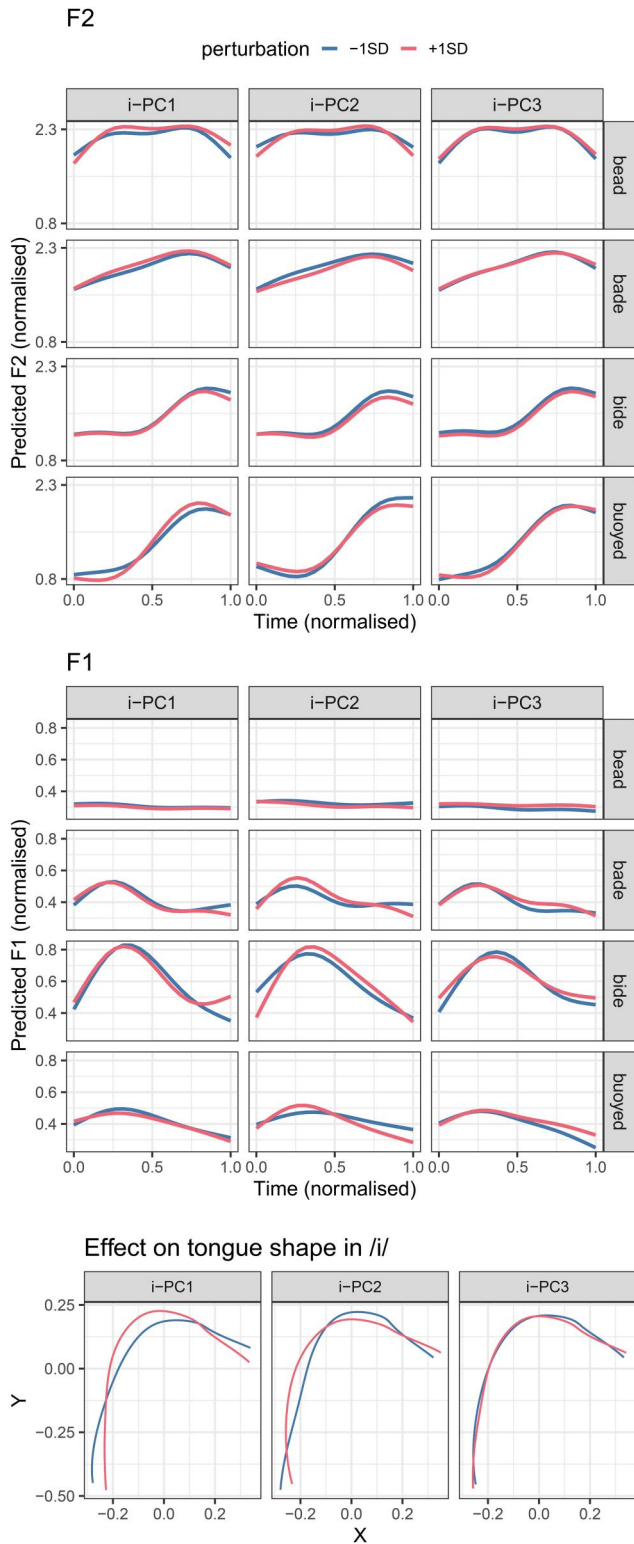


FIG. 3. Top: GAMM predictions for normalised F1 and F2 trajectories, depending on tongue shape (i-PC value) and vowel. Bottom: Visualisation of predicted tongue shape for the speakers with high (red) and low (blue) i-PC values.

Table I summarises the means and SDs of the PC scores in the dataset across all vowels and in /i/ and reports the standard errors of the mean (SEMs). These summary statistics are calculated over the initial 80% of the vowel in order

TABLE I. Summary statistics for PC scores across all vowels and within /i/.

PC	All vowels			/i/		
	Mean	SD	SEM	Mean	SD	SEM
PC1	0.008	0.089	0.0003	0.021	0.078	0.0011
PC2	0.004	0.082	0.0003	-0.101	0.061	0.0008
PC3	-0.003	0.035	0.0001	0.001	0.034	0.0005

to exclude tongue shapes that are likely strongly influenced by the following /d/. Overall variance is greater across vowels than within /i/, which is expected, since /i/ variance spans only inter-speaker and token variation, whereas across vowels, variation is induced by both vowel and speaker. However, even within /i/, we find substantial variation in PC scores, suggesting considerable speaker variation in tongue shape. On average, /i/ has higher PC1 than the grand mean PC1 value. This is expected given the interpretation that PC1 is a correlate of increased doming at tongue dorsum—/i/ being a high vowel has a considerable degree of dorsal doming. Mean PC2 score in /i/ is negative, that is, lower compared to the grand mean PC2 score. This aligns with the interpretation of low PC2 values representing a combination of tongue doming and anteriority (more anterior location of the tongue constriction). In contrast to PC1 and PC2, the central tendency of PC3 is similar in /i/ as across all vowels, suggesting that the tongue shape information carried by PC3 is similar in /i/ and in the speaker’s overall mean tongue shape. The SEM values for /i/ are larger than for the whole dataset, reflective of the fact that the /i/ tokens represent a subset of the data.

We can also use the PC scores to verify one of the key assumptions of this study, namely, that the mean tongue shape in /i/ is closely related to the shape of I-diphthong offglides. Table II summarises the results of a series of Pearson’s correlation tests, examining the correlation between by-speaker mean PC scores for /i/ and by-speaker mean scores for other I-diphthong offglides. The mean PC scores for /i/ are calculated over the initial 80% of the vowel, as in Table I. The offglide was defined as the ultrasound frame closest to the 85% of vowel duration. This time point was selected based on our previous research into diphthong timing in the same dataset (Strycharczuk et al., 2024). As we can see from the table, the strength of the correlation was very high or high for PC1 and PC2 (correlation coefficients ranging from 0.85 to 0.95). In comparison, PC3 shows somewhat weaker, though still quite substantial, correlations between /i/ and diphthong offglides: 0.77 for the offglide of

TABLE II. Pearson’s correlations between by-speaker mean PC for each I-diphthong offglide and the corresponding i-PC (all $p < 0.001$).

PC	bade	bide	buoyed
PC1	0.93	0.88	0.89
PC2	0.95	0.85	0.88
PC3	0.77	0.65	0.67

bade, 0.65 for the offglide of *bide*, and 0.67 for the offglide of *buoyed*. From this, we can conclude that the mean tongue shape in /i/ is strongly isomorphic with the tongue shape of the same speaker’s diphthong offglide, especially in terms of palate doming and anteriority. Having verified this assumption, let us consider whether the variation in /i/ tongue shape affects formant trajectories in I-diphthongs.

B. GAMM results

Table III summarises the results of Maximum Likelihood comparison testing the significance of individual i-PCs as predictors of time-varying F1 and F2. An i-PC was considered significant if its inclusion in the model significantly improved the model fit.

As we can see from the table, i-PC1 significantly predicted the F1 trajectory shape for *bade*, *bide*, and *buoyed* ($p < 0.001$) but not for *bead* ($p = 0.950$). For F2 trajectory shape, i-PC1 was significant only for *buoyed* and *bead* ($p < 0.001$), while the effects for *bade* and *bide* were not significant. i-PC2 significantly predicted F1 for all four vowels. i-PC3 showed significant effects on F1 trajectory for all vowels, including *bead* ($p = 0.044$), but for F2, it was significant only for *buoyed* and *bead* ($p < 0.001$).

C. Follow-up exploratory analysis

In order to understand the nature of the i-PC effects summarised in Table III, we visualised the model predictions, comparing speakers with high and low i-PC values, where “high” and “low” are defined as being ± 1 SD from the mean i-PC value. Figure 3 illustrates predicted formant trajectories for normalised F1 and F2, depending on tongue shape and vowel. The predicted formant trajectories for speakers with high (1 SD above the mean) and low i-PC values (1 SD below the mean) are plotted in red and blue, respectively. To aid interpretation, the predicted tongue shapes in /i/, depending on PC perturbation, are included in the bottom panel of the same figure.

While the effect of i-PC1 on F1 trajectories in *bade*, *bide*, and *buoyed* was significant (see Table III), it is very small in scope. In comparison, the effect of i-PC1 on the F2 trajectory in *buoyed* is more readily observable. Nevertheless, across these four cases, we can generalise that the acoustic shift from onglide to offglide starts earlier in

speakers with high i-PC1 (speakers with more dorsal doming).

Variation in i-PC2 was significantly associated with trajectory shape for all vowels and trajectories, according to Table III. According to Fig. 3, speakers with low i-PC2 scores (more domed tongue shape in /i/ with more anterior constriction) generally produce earlier and faster formant transitions. The differences are larger for F1 than for F2. The F1 trajectory in *buoyed* is somewhat exceptional here in that speakers with low i-PC2 maintain a relatively more stable F1 over the whole duration of the vowel compared to speakers with high i-PC2 scores, such that the offglide is slightly acoustically undershot.

Variation in i-PC3 mainly affects the offglide F1. For *bade*, *bide*, and *buoyed*, speakers with high i-PC3 values achieve a slightly higher minimum F1 in the offglide compared to speakers with low i-PC3. The effect of i-PC3 on F2 trajectories is overall very small, although it was significant for *bead* and *buoyed*.

In order to further illuminate the observed differences, we conducted an analysis of vowel duration. Our rationale for analysing duration is that we observe differences in the relative timing of formant trajectories in normalised time. It is, therefore, important to ensure that the differences are not an artifact of time normalisation and fundamentally caused by systematic differences in vowel duration. To test the effect of i-PCs on duration, we fitted a linear mixed effects regression model with vowel duration as dependent variable, and i-PCs 1, 2, and 3 as predictors. Vowel (/i/, /eɪ/, /aɪ/, and /oɪ/) and speaker were included as random intercepts. *p*-values for coefficients were obtained using Satterthwaite’s approximation (Kuznetsova et al., 2017). The model summary is in Table IV. As we can see from the table, none of the i-PCs was significant. This suggests that the differences we identify in formant dynamics conditioned by i-PCs are not related to any underlying differences in vowel duration.

Finally, we conducted an analysis of changes in tongue shape and articulatory displacement, as a function of the /i/-PCs. Our aim was to understand the nature of movement produced in diphthongs by speakers with different tongue shapes in /i/. In the articulatory analysis, we used a range of articulatory variables, measured dynamically, as dependent variables in GAMMs. The variables were: dynamic change in PC1, PC2, and PC3, fronting of the tongue root (horizontal displacement of DLC knot 4), fronting of the tongue dorsum (horizontal displacement of DLC knot 5), raising of the posterior and anterior parts of the tongue dorsum (vertical displacement of DLC knots 5 and 7), raising of the tongue

TABLE III. Summary of *p*-values returned by model comparisons between a full GAMM model of formant trajectory and a nested model without a specific i-PC.

Predictor	DV	<i>bade</i>	<i>bide</i>	<i>buoyed</i>	<i>bead</i>
i-PC1	F1	<0.001	<0.001	<0.001	0.950
i-PC1	F2	0.186	0.083	<0.001	<0.001
i-PC2	F1	<0.001	<0.001	<0.001	0.003
i-PC2	F2	0.010	0.021	<0.001	0.005
i-PC3	F1	<0.001	<0.001	<0.001	0.044
i-PC3	F2	0.529	0.576	<0.001	<0.0001

TABLE IV. Model summary for a linear mixed effects model predicting vowel duration.

Term	β	SE	<i>t</i>	<i>p</i>
(Intercept)	0.27	0.03	10.49	<0.001
i-PC1	0.02	0.15	0.17	0.868
i-PC2	0.09	0.20	0.45	0.658
i-PC3	-0.47	0.43	-1.08	0.287

Posterior tongue dorsum raising

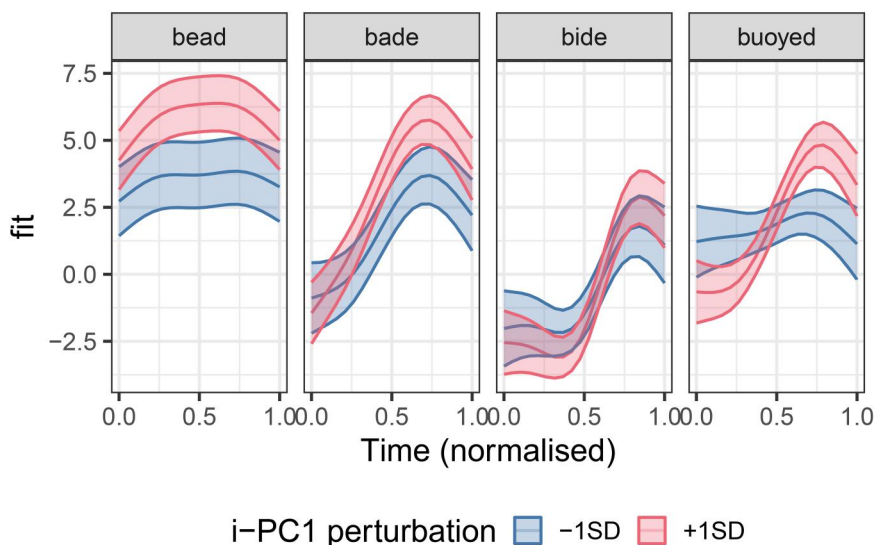


FIG. 4. GAMM predictions for patterns of raising of the posterior part of the dorsum (DLC knot 5) in speakers with high and low values of i-PC1.

mid (knot 8), and raising of the tongue blade (knot 10). The choice of these variables was informed by our previous research on parametrising tongue shape variation in vowels (Strycharczuk *et al.*, 2025). All of the displacement variables were rotated to the speaker’s occlusal plane and centred within speaker. Note that the tongue coordinates were not corrected relative to any hard structures in the vocal tract, such as the palate. Consequently, it is possible that some of the height measures are overestimated in low vowels and somewhat underestimated in high vowels. We do not consider this issue to be severe, based on good overall agreement between DLC-based measures of tongue position and height, compared to corresponding electromagnetic articulometry results (Kirkham and Strycharczuk, 2025).

The predictor structure in the articulatory models was identical to the acoustic models reported here, except we used fast restricted maximum likelihood estimation. The full

details of all of the models are reported in the online repository (see Data Availability). Here, we report selected key results, focusing on the articulatory variables that we expect to co-vary systematically with tongue shape in /i/, based on our interpretation of the PCA presented in Sec. III A.

Figure 4 shows predicted patterns of tongue dorsum raising in speakers with high and low i-PC1. Speakers with relatively higher i-PC1 scores show considerably more dorsum raising in *bead* and in *bade* and *buoyed* offglides, consistent with greater doming of the tongue dorsum as captured by PC1. Additionally, the tongue dorsum raising is more rapid in *bade* and *buoyed*.

Figure 5 illustrates the predicted trajectory of tongue root fronting in I-diphthongs, depending on variation in i-PC2. Low i-PC2 scores are associated with a more domed tongue shape and a more anterior constriction location (Fig. 3). In accordance with this, speakers with relatively

Tongue root fronting

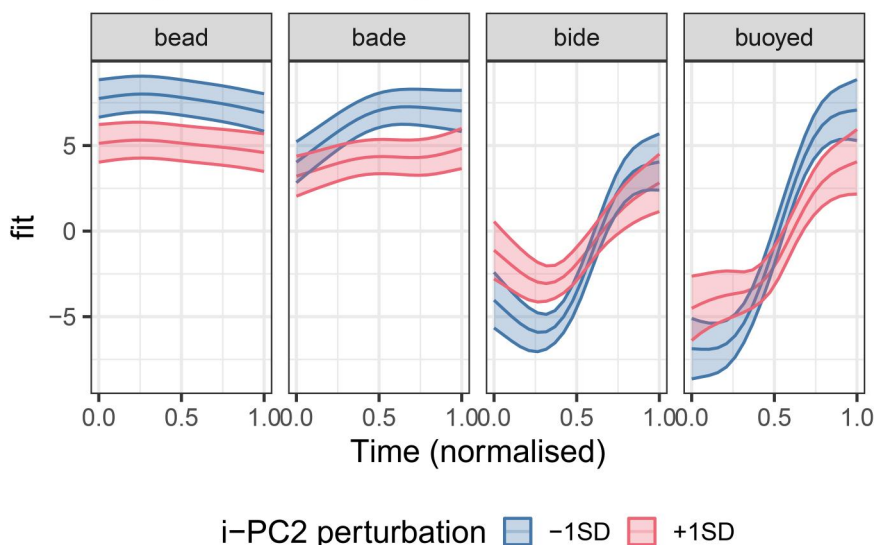


FIG. 5. GAMM predictions for patterns of tongue root fronting in I-diphthongs (DLC knot 4) in speakers with high and low values of i-PC2.

lower PC2 scores have more tongue root fronting in *bead* and in I-diphthong offglides compared to speakers with high i-PC2 scores. Greater tongue-root advancement may be necessary for speakers with highly domed palates, allowing those speakers to raise the dorsum closer to the palate. Low i-PC2 scores are also associated with more inherent change in tongue root fronting in *bide* and *buoyed*. This suggests a greater horizontal articulatory range in speakers with low i-PC2 scores.

Compared to i-PCs 1 and 2, variation in i-PC3 is not linked to robust differences in articulatory displacement. We return to this point in Sec. IV.

IV. DISCUSSION

We have presented evidence that speakers who share specific tongue shape features in their /i/ production also share some features of their acoustic formant trajectories in selected I-diphthongs. This result provides support to the hypothesis that articulatory strategy has a systematic influence on formant transitions. More importantly, our findings illuminate the nature of this influence, by linking specific types of articulatory strategy for palatal vowel production with specific acoustic dynamic changes. In general, we find that speakers who have a more anterior constriction, or more domed tongue shape, in /i/ produce earlier formant transitions between onglide and offglide in selected I-diphthongs.

The relationship between tongue shape and formant dynamics can be explained by the properties of articulatory movement associated with different tongue shapes and their interaction with palate shape. Some tongue shapes involve more articulatory displacement than others: more domed tongue shape is associated with greater tongue dorsum raising (Fig. 4), and producing a constriction in the front section of the tongue involves more tongue root fronting (Fig. 5). This means that speakers with high PC1 and speakers with low PC2 have to traverse a larger distance within the same time window. Note that we find no differences in vowel duration depending on tongue shape (Table IV). Given the need to cover greater distance and a stable time interval, these speakers need to produce higher velocity, as well as faster movement, at the onset of the gesture. This is consistent with the predictions of a linear harmonic oscillator model for tract variable motion (Saltzman and Munhall, 1989). Specifically, under the assumption of identical activation durations, damping and stiffness parameters, movements across different distances will converge to the target at the same time point, with larger movements having a much steeper velocity slope and higher peak velocity. The differences observed in our data can, therefore, be understood as a simple consequence of variable movement distances in differently sized vocal tracts and potentially with tongues of differing volume, with similar timing of target achievement being privileged across speakers.

Having discussed why the specific articulatory strategies condition the observed changes in formant dynamics, let us also consider why we find the specific strategies in the

first place. Variation in i-PC1 and i-PC2 aligns closely with previous findings on articulatory strategy and vocal tract morphology (Lammert *et al.*, 2013a; Serrurier *et al.*, 2019; Serrurier and Neuschaefer-Rube, 2023, 2024). Variation in i-PC1 is characterised by different degrees of dorsal doming. Increased doming of the tongue dorsum involves more dorsal raising. Additionally, more domed /i/ shapes are associated with a greater vertical range of dorsal displacement. This type of articulatory strategy is possibly conditioned by greater vocal tract size in the vertical dimension, and possibly also by greater palate doming. The tongue root differences between speakers with low and high i-PC2 are also most likely driven by articulatory compensation. Previous research suggests that increased range of tongue root movement and greater palate doming are characteristic of relatively longer vocal tracts (Serrurier and Neuschaefer-Rube, 2024).

While the proposed mechanism can account for the effect of i-PC1 (dorsal doming) and i-PC2 (anterior doming and tongue root advancement), the effect of i-PC3 on formant transitions in *bade* is more challenging to account for. Unlike i-PC1 and i-PC2, variation in i-PC3 does not seem to be linked to major differences in articulatory displacement, and so it probably is not conditioned by differences in vocal tract size or shape. We hypothesise that speakers with high i-PC3 have a greater than average tendency for anticipatory coarticulation between the vowel and the following consonant, in this case /d/. The slight concavity in the anterior part of the tongue for these speakers (Fig. 3) could be caused by tongue tip raising in anticipation of the following alveolar closure. This interpretation is consistent with the observation that speakers with relatively high i-PC3 tend to produce higher F1 in their I-diphthong offglides, suggesting a small degree of offglide undershoot, which could be consistent with increased coarticulation. This explanation, however, is only tentative, and it invites follow-up research with more direct evidence on palate shape.

We observe that *bead* consistently shows less robust acoustic differences related to tongue shape, compared to other I-diphthongs (Fig. 3). This aligns with earlier forensic findings that speaker-specificity of formant trajectories interacts with vowel, such that greater inherent vowel change (more diphthongisation) is associated with greater speaker-specificity of formant transitions (Heeren, 2020). Both sets of results suggest that the acoustic effect of articulatory strategy is more readily observable where there is larger articulatory displacement. While speakers are generally able to compensate for differences in their vocal tract shape, instances of large articulatory displacement can occasionally reveal underlying individual strategy variation in the resulting acoustics.

V. CONCLUSION

Our results provide empirical support for the hypothesis that vowel dynamics can systematically differ depending on the characteristics of articulatory strategy in palatal vowel

production, which itself is likely a reflection of vocal tract constraints. The link between the nature of articulation and acoustic vowel dynamics is mediated by systematic properties of articulatory movement. Specifically, greater articulatory displacement is associated with somewhat earlier and more rapid formant transitions. We have argued that these effects arise from the dynamic properties of the gestural organisation of speech, which condition a faster velocity rise when greater displacement is required within a fixed time window. This mechanism explains the link between vocal tract constraints on the one hand and individual differences in vowel dynamics on the other, also highlighting the limits of articulatory compensation: while articulatory variation is largely conditioned by compensation in the first place, some aspects of the relevant variation are acoustically recoverable. As such, our study bridges earlier forensic and articulatory findings and advances the conceptual understanding of the sources of individuality in speech. It also calls for further investigation into the generalisability of the relationship between greater articulatory displacement and formant trajectory shape.

Our study is informed by data from speakers of English from a specific locality. We have no reason to suspect that the relationship between the articulatory strategy and formant dynamics that we observe is specific to English, which is why we explain it in terms of universal principles of the articulation-acoustics relationship. However, we cannot rule out the possibility that our results are influenced, for example, by language-specific articulatory setting (Laver, 1980). Our conclusion that speaker anatomy can influence formant dynamics via articulatory strategy would be greatly strengthened by evidence from other languages and/or from studies focusing on other aspects of anatomy and articulatory strategy.

ACKNOWLEDGMENTS

We thank the JASA editor and reviewers for their feedback on the earlier versions of this manuscript. This research was supported by an AHRC research grant awarded to P.S. and S.K. (AH/S011900/1), a British Academy Mid-Career Fellowship awarded to P.S. (MFSS24\240076), an AHRC Fellowship awarded to S.K. (AH/Y002822/1) and a Royal Society grant also supported by the Leverhulme Trust (APXR1\251102) awarded to S.K. We would like to acknowledge the assistance given by Research IT and the use of the Computational Shared Facility at The University of Manchester.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Ethics approval

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).

DATA AVAILABILITY

The data and code that support the findings of this study are openly available in the OSF repository at <https://osf.io/xtp6q/>.

¹Note that compensation in this case refers to full parasagittal morphology and not just the midsagittal plane.

²A reviewer observes that comparisons of variability between different vowels are complicated by the properties of the vowel inventory in a given language, e.g., variability may be constrained by the presence of multiple vowels in the same area. Another reviewer notes that tongue bracing may add to the relative stability of palatal vowels.

- Baranowski, M., and Turton, D. (2015). "Manchester English," in *Researching Northern Englishes*, edited by R. Hickey (John Benjamins, Amsterdam, The Netherlands), pp. 293–316.
- Barreda, S. (2021). "Fast track: Fast (nearly) automatic formant-tracking using Praat," *Linguist. Vanguard* 7(1), 20200051.
- Blumstein, S. E., and Stevens, K. N. (1979). "Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants," *J. Acoust. Soc. Am.* 66(4), 1001–1017.
- Boersma, P., and Weenink, D. (2022). "Praat: Doing phonetics by computer (version 6.2.14) [computer programme]," <http://www.praat.org/> (Last viewed May 4, 2026).
- Brunner, J., Fuchs, S., and Perrier, P. (2009). "On the relationship between palate shape and articulatory behavior," *J. Acoust. Soc. Am.* 125(6), 3936–3949.
- Eckert, P. (2008). "Variation and the indexical field," *J. Sociolinguist.* 12(4), 453–476.
- Foulkes, P., and Docherty, G. (2006). "The social life of phonetics and phonology," *J. Phonetics* 34, 409–438.
- Gay, T., Lindblom, B., and Lubker, J. (1981). "Production of bite-block vowels: Acoustic equivalence by selective compensation," *J. Acoust. Soc. Am.* 69(3), 802–810.
- Guenther, F. H. (2016). *Neural Control of Speech* (The MIT Press, Cambridge, MA).
- Hasegawa-Johnson, M., Pizza, S., Alwan, A., Cha, J. S., and Haker, K. (2003). "Vowel category dependence of the relationship between palate height, tongue height, and oral area," *J. Speech. Lang. Hear. Res.* 46(3), 738–753.
- Heeren, W. (2020). "The contribution of dynamic versus static formant information in conversational speech," *Int. J. Speech Lang. Law* 27(1), 75–98.
- Houde, J., and Jordan, M. I. (1998). "Sensorimotor adaptation in speech production," *Science* 279(5354), 1213–1216.
- Hughes, V., Wood, S., and Foulkes, P. (2016). "Strength of forensic voice comparison evidence from the acoustics of filled pauses," *Int. J. Speech Lang. Law* 23(1), 99–132.
- Johnson, K. (2020). "The ΔF method of vocal tract length normalization for vowels," *Lab. Phonol.* 11(1), 10.
- Johnson, K. (2023). "Individual differences in speech production: What is 'phonetic substance'?", in *Proceedings of the 20th International Congress of Phonetic Sciences*, edited by R. Skamitzl and J. Volín (International Phonetic Association, London, UK), pp. 1102–1106.
- Johnson, K., Ladefoged, P., and Lindau, M. (1993). "Individual differences in vowel production," *J. Acoust. Soc. Am.* 94(2), 701–714.
- Kent, J. T. (1992). "New directions in shape analysis," in *The Art of Statistical Science*, edited by K. V. Mardia (Wiley, New York), pp. 115–127.
- Kirchhoff, K., Fink, G. A., and Sagerer, G. (2002). "Combining acoustic and articulatory feature information for robust speech recognition," *Speech Commun.* 37(3–4), 303–319.
- Kirkham, S., and Strycharczuk, P. (2025). "Dynamical model parameters from ultrasound tongue kinematics," *JASA Express Lett.* 5(11), 115201.
- Kirkham, S., Strycharczuk, P., Gorman, E., Nagamine, T., and Wrench, A. (2023). "Co-registration of simultaneous high speed ultrasound and

- electromagnetic articulography for speech production research,” in *Proceedings of the 20th International Congress of Phonetic Sciences*, edited by R. Skarnitzl and J. Volín (International Phonetic Association, London, UK), pp. 918–922.
- Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). “lmerTest package: Tests in linear mixed effects models,” *J. Statist. Software* **82**(13), 1–26.
- Lammert, A., Proctor, M., and Narayanan, S. (2013a). “Interspeaker variability in hard palate morphology and vowel production,” *J. Speech. Lang. Hear. Res.* **56**(6), 1924–1933.
- Lammert, A., Proctor, M., and Narayanan, S. (2013b). “Morphological variation in the adult hard palate and posterior pharyngeal wall,” *J. Speech. Lang. Hear. Res.* **56**(2), 521–530.
- Laver, J. (1980). *The Phonetic Description of Voice Quality* (Cambridge University Press, Cambridge, UK).
- Lisker, L. (1985). “The pursuit of invariance in speech signals,” *J. Acoust. Soc. Am.* **77**(3), 1199–1202.
- Lo, J. J. H., Strycharczuk, P., and Kirkham, S. (2025). “Articulatory strategy in vowel production as a basis for speaker discrimination,” in *Proceedings of Interspeech 2025*, pp. 3504–3508.
- Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., and Bethge, M. (2018). “DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning,” *Nat. Neurosci.* **21**(9), 1281–1289.
- McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., and Sonderegger, M. (2017). “Montreal forced aligner: Trainable text-speech alignment using Kaldi,” in *Proceedings of Interspeech 2017*, pp. 498–502.
- McDougall, K. (2004). “Speaker-specific formant dynamics: An experiment on Australian English /a/,” *Int. J. Speech Lang. Law* **11**(1), 103–130.
- McDougall, K. (2006). “Dynamic features of speech and the characterization of speakers: Toward a new approach using formant frequencies,” *Int. J. Speech Lang. Law* **13**(1), 89–126.
- McDougall, K., and Nolan, F. (2007). “Discrimination of speakers using the formant dynamics of /u/ in British English,” in *Proceedings of the 16th International Congress of Phonetic Sciences*, edited by J. Trouvain and W. J. Barry, pp. 1825–1828.
- McFarland, D. H., and Baum, S. R. (1995). “Incomplete compensation to articulatory perturbation,” *J. Acoust. Soc. Am.* **97**(3), 1865–1873.
- Morrison, G. S. (2009). “Likelihood-ratio forensic voice comparison using parametric representations of the formant trajectories of diphthongs,” *J. Acoust. Soc. Am.* **125**(4), 2387–2397.
- Noiray, A., Iskarous, K., and Whalen, D. (2014). “Variability in English vowels is comparable in articulation and acoustics,” *Lab. Phonol.* **5**(2), 271–288.
- Nolan, F. (1983). *The Phonetic Bases of Speaker Recognition* (Cambridge University Press, Cambridge, UK).
- Nolan, F., and Grigoras, C. (2005). “A case for formant analysis in forensic speaker identification,” *Int. J. Speech Lang. Law* **12**(2), 143–173.
- Redford, M., and Baese-Berk, M. (2023). “Acoustic theories of speech perception,” in *Oxford Research Encyclopedia of Linguistics* (Oxford University Press, Oxford, UK).
- Rhodes, R. W. (2012). “Assessing the strength of non-contemporaneous forensic speech evidence,” Ph.D. thesis, University of York, York, UK.
- Rose, P., Warren, P., and Watson, C. (2006). “The intrinsic forensic discriminatory power of diphthongs,” in *Proceedings of the 11th Australian International Conference on Speech Science and Technology*, pp. 64–69.
- Saltzman, E. L., and Munhall, K. G. (1989). “A dynamical approach to gestural patterning in speech production,” *Ecol. Psychol.* **1**(4), 333–382.
- Serrurier, A., Badin, P., Lamalle, L., and Neuschaefer-Rube, C. (2019). “Characterization of inter-speaker articulatory variability: A two-level multi-speaker modelling approach based on MRI data,” *J. Acoust. Soc. Am.* **145**(4), 2149–2170.
- Serrurier, A., and Neuschaefer-Rube, C. (2023). “Morphological and acoustic modeling of the vocal tract,” *J. Acoust. Soc. Am.* **153**(3), 1867–1886.
- Serrurier, A., and Neuschaefer-Rube, C. (2024). “Formant-based articulatory strategies: Characterisation and inter-speaker variability analysis,” *J. Phonetics* **107**, 101374.
- Sóskuthy, M. (2021). “Evaluating generalised additive mixed modelling strategies for dynamic speech analysis,” *J. Phonetics* **84**, 101017.
- Spreato, L., Pucher, M., and Matosova, A. (2018). “UltraFit: A speaker-friendly headset for ultrasound recordings in speech science,” in *Proceedings of Interspeech 2018* (International Speech Communication Association), pp. 1517–1520.
- Stevens, K. N. (1989). “On the quantal nature of speech,” *J. Phonetics* **17**(1–2), 3–45.
- Strycharczuk, P., and Kirkham, S. (2025). “Articulatory strategies in male and female vowel production,” *J. Speech. Lang. Hear. Res.* **68**(12), 5629–5649.
- Strycharczuk, P., Kirkham, S., Gorman, E., and Nagamine, T. (2024). “Towards a dynamical model of English vowels. Evidence from diphthongisation,” *J. 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,” *Lang. Speech* **68**(3), 689–721.
- Strycharczuk, P., López-Ibáñez, M., Brown, G., and Leemann, A. (2020). “General Northern English. Exploring regional variation in the North of England with machine learning,” *Front. Artif. Intell.* **3**, 48.
- Watson, C. I., and Harrington, J. (1999). “Acoustic evidence for dynamic formant trajectories in Australian English vowels,” *J. Acoust. Soc. Am.* **106**(1), 458–468.
- Watt, D. (2002). “‘I don’t speak with a Geordie accent, I speak, like, the Northern accent’: Contact-induced levelling in the Tyneside vowel system,” *J. Sociolinguist.* **6**(1), 44–63.
- Weirich, M. (2012). “The influence of nature and nurture on speaker-specific parameters in twins speech,” Ph.D. thesis, Humboldt-Universität zu Berlin, Berlin, Germany.
- Weirich, M., Fuchs, S., Simpson, A., Winkler, R., and Perrier, P. (2016). “Mumbling: Macho or morphology?,” *J. Speech. Lang. Hear. Res.* **59**(6), S1587–S1595.
- Weirich, M., and Simpson, A. P. (2018). “Individual differences in acoustic and articulatory undershoot in a German diphthong—variation between male and female speakers,” *J. Phonetics* **71**, 35–50.
- Wells, J. (1982). *Accents of English 1: An Introduction* (Cambridge University Press, Cambridge, UK), Vol. 2.
- Wood, S. (2017). *Generalized Additive Models: An Introduction with R*, 2nd ed. (Chapman and Hall, London, UK).
- Wrench, A., and Balch-Tomes, J. (2022). “Beyond the edge: Markerless pose estimation of speech articulators from ultrasound and camera images using DeepLabCut,” *Sensors* **22**(3), 1133.