JASA EXPRESS LETTERS ARTICLE asa.scitation.org/journal/jel

Dynamical model parameters from ultrasound tongue kinematics

Sam Kirkham[1,a] and Patrycja Strycharczuk[2]
[1]Phonetics Laboratory, Lancaster University, Lancaster LA1 4YL, United Kingdom
[2]Linguistics and English Language, University of Manchester, Manchester M13 9PL, United Kingdom

Abstract: The control of speech can be modeled as a dynamical system in which articulators are driven toward target positions. These models are typically evaluated using fleshpoint data, such as electromagnetic articulography (EMA), but recent methodological advances make ultrasound imaging a promising alternative. We evaluate whether the parameters of a linear harmonic oscillator can be reliably estimated from ultrasound tongue kinematics and compare these with parameters estimated from simultaneously recorded EMA data. We find that ultrasound and EMA yield comparable dynamical parameters, while mandibular short tendon tracking also adequately captures jaw motion. This supports using ultrasound kinematics to evaluate dynamical articulatory models.

[Editor: Douglas D. O'Shaughnessy] https://doi.org/10.1121/10.0039769

Received: 24 June 2025 **Accepted:** 20 October 2025 **Published Online:** 3 November 2025

1. Introduction

A major goal in the study of speech communication is understanding the nature of articulatory control. A common approach is to cast this problem in terms of a dynamical system with point attractor dynamics, where a small number of parameters drive the vocal tract to a stable equilibrium position (Fowler, 1980; Saltzman and Munhall, 1989; Browman and Goldstein, 1986; Gafos, 2006; Tilsen, 2016). A standard model in this framework is the linear harmonic oscillator,

$$m\ddot{x} + b\dot{x} + kx = 0, \tag{1}$$

where m is mass (typically $m = 1$), k is a stiffness coefficient, and b is a damping coefficient, usually set to critically damped $b = 2\sqrt{mk}$. Gestural activation can be governed by step activation, with gestural parameters changing instantaneously at the point of activation and remaining constant over the activation interval.

In this study, we focus on whether the parameters of a linear harmonic oscillator can be estimated from ultrasound tongue imaging data, which we compare with the more common method of fitting to electromagnetic articulography (EMA) data. A major barrier to this goal is that the linear harmonic oscillator is known to be a poor fit to empirical articulatory trajectories, as it predicts overly short time-to-peak velocity, meaning that it is inappropriate for evaluating how the model can fit different data modalities. There are three common solutions to this issue. The first allows gestural activation to vary over time (Byrd and Saltzman, 1998), which adds extrinsic complexity to the model. The second is a nonlinear model, such as adding a cubic term to the linear model (Sorensen and Gafos, 2016; Kirkham, 2025b), or novel nonlinear models (Stern and Shaw, 2025). The third is to abandon oscillatory models and develop new time-dependent (i.e., non-autonomous) models (Elie *et al.*, 2023). All three approaches add significant complexity, but we take an alternative route, which is to retain the simple linear oscillator with step activation, but simply relax the critical damping constraint. This allows for a simple autonomous model that is generally more accurate than critically damped models (Kirkham, 2024, 2025a). We note that all of the above models focus only on piecewise dynamics, such as the movement between articulatory targets, so our decision to relax critical damping only adds a small level of complexity compared with nonlinear or non-autonomous models.

An important aspect of adjudicating between different models is evaluating their fit to empirical data, which allows us to establish prospective parameters for articulatory control. In other words: given an empirical articulatory trajectory, which model parameter values would be required to reproduce its dynamics? To date, the majority of dynamical articulatory model development has focused on fleshpoint tracking data, such as x-ray microbeam and EMA (Iskarous, 2016; Elie *et al.*, 2023; Kirkham, 2025a; Stern and Shaw, 2025), with some applications to MRI data (Lammert *et al.*, 2013). Such data has a number of shortcomings, including limited information on the tongue posterior (EMA), invasive data collection (EMA, MRI), and limited portability (EMA, MRI). Ultrasound imaging largely overcomes these issues and provides

[a]Corresponding author: s.kirkham@lancaster.ac.uk

good imaging of the tongue, as well as hyoid and mandibular short tendon (Wrench and Balch-Tomes, 2022), but suffers from lower frame rates and noisy images. Despite this, recent advances suggest that it is possible to derive kinematics from ultrasound, either via tracking manually identified fanlines (Strycharczuk and Scobbie, 2015) or anatomically defined landmarks using deep learning (Wrench and Balch-Tomes, 2022). For example, Wrench and Balch-Tomes (2022) trained a deep learning model on human-labeled data, where anatomically defined landmarks were placed along the tongue. The accuracy of landmarking was comparable to between-human differences, allowing for automated frame-to-frame tracking of fleshpoint-like trajectory data, which also showed reasonable agreement with EMA.

The above suggests that ultrasound is a candidate for estimating dynamical model parameters from data. This would be a substantial step forward for evaluating dynamical models, as ultrasound is cheaper, less invasive, and provides richer information about lingual motion. It stands to reason that being able to accurately estimate dynamical parameters from ultrasound would open up a new range of applications for fieldwork and clinical data, which would facilitate model evaluation across more diverse samples and languages. In this study, we compare task dynamic parameters derived from simultaneous EMA and ultrasound data during vowel production. We focus on estimating the parameters of an undamped linear harmonic oscillator [i.e., Eq. (1), but without the critical damping constraint]. We use this model because it is simple and has known characteristics, which makes it an attractive case study for comparing model parameters estimated from ultrasound and EMA data. We expect that the same principles should apply to more complex models, but we use the simple model to establish a straightforward comparison without too many degrees of freedom.

2. Methods

2.1 Speakers and stimuli

The dataset comprises simultaneous electromagnetic articulography and ultrasound tongue imaging data, which was recorded concurrently from six female speakers of Northern Anglo British English. The materials comprised the full set of British English vowels in /bV/ and /bVd/ contexts in two carrier phrases: "She said X" and "She said X eagerly." Each speaker produced four repetitions of 29 words in two carrier phrases, except for one speaker who produced five repetitions. We excluded some blocks from two speakers due to excessive ultrasound probe movement. In total, we analyzed 1095 tokens.

2.2 Instrumentation

EMA data were recorded using a Carstens AG501, with sensors placed on the tongue tip, tongue mid, tongue dorsum, upper/lower lip, and lower teeth, with reference sensors located on the maxilla, nasion, and mastoids. The EMA data were recorded at 1250 Hz, filtered using a 50-Hz low-pass Kaiser-windowed filter (5 Hz for reference sensors), head corrected, and rotated to the occlusal plane. Ultrasound data were recorded in Articulate Assistant Advanced (Wrench, 2022) at ~81 Hz using a Telemed MicrUS scanner with a 20-mm radius, 64-element, 2-MHz probe. The ultrasound probe was stabilized using a headset (Spreafico *et al.*, 2018). Audio was recorded at 48 kHz using a Beyerdynamic Opus 55 microphone and pre-amplified using a Grace Designs m101 preamplifier. Audio, EMA, and ultrasound data were time synchronized by aligning a transistor-transistor logic pulse that was triggered at the time of each prompt presentation and recorded onto each system. For further details of temporal synchronization and analysis of probe motion, see Kirkham *et al.* (2023).

2.3 Data processing

Acoustic data were forced-aligned using Montreal Forced Aligner (McAuliffe *et al.*, 2017) and subsequently hand-corrected, which was used to segment the articulatory data based on the labeled consonant-vowel velocity interval. Anatomical landmarks in the ultrasound images were tracked in each frame using DeepLabCub (DLC) (Mathis *et al.*, 2018), which is a deep learning algorithm for markerless pose-estimation. We specifically used a pre-trained tongue model with 11 points at anatomical landmarks along the tongue, as well as points corresponding to the short tendon and hyoid bone (Wrench and Balch-Tomes, 2022). Figure 1 (left) shows that knot 1 is located at the tongue root and knot 11 is located at the tongue tip, with all knots specified for x/y dimensions (Wrench and Balch-Tomes, 2022; Strycharczuk *et al.*, 2025). Knots were exported as Cartesian coordinates (in millimeters) and rotated parallel to the occlusal plane (estimated using a bite plate recording for each speaker). The EMA data were downsampled to the ultrasound frame rate, and EMA/ultrasound measures were then projected to a shared origin by centering, but without scaling in order to retain dimension-specific variation in movement range. The ultrasound data are noisy compared with the EMA data, so the ultrasound and EMA signals were both smoothed using the fifth-order Discrete Cosine Transform. See Fig. 1 (right) for an example, which clearly shows the necessity of smoothing the ultrasound position data. Note that the EMA and ultrasound signals appear to capture the same underlying signal, but with a small time lag. This is likely a consequence of the EMA and ultrasound spatial points capturing slightly different locations on the tongue dorsum. Unfortunately, it is not straightforward to quantify specific relations between the EMA sensor locations and DLC knot locations, because the EMA sensors are not visible in the ultrasound image, so our selection of equivalent locations is a rough approximation.

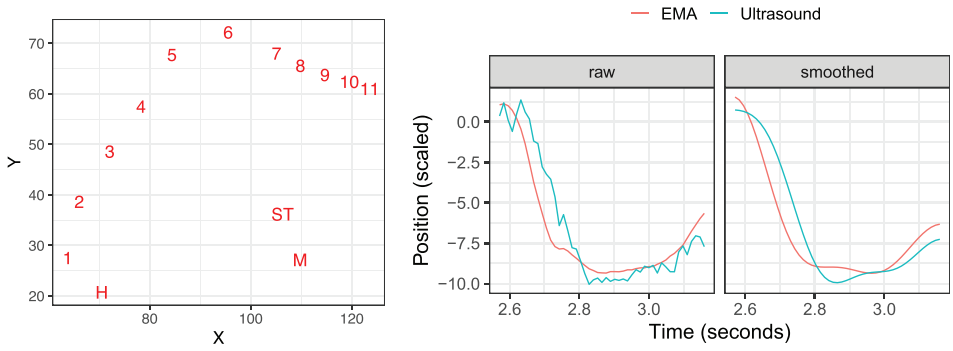

Fig. 1. Left: Location of DLC knots estimated for each ultrasound frame (knot 1 is tongue root, knot 11 is tongue tip; H, hyoid; M, mandible; ST, short tendon). Right: Raw and smoothed data for TD horizontal position from EMA and ultrasound in the word "bar."

2.4 Feature extraction

We analyzed horizontal and vertical movements of the tongue dorsum (TD) and jaw (JAW). TD is a standard measurement dimension in the EMA literature, while JAW is an additional dimension that can be tracked using EMA and ultrasound. In the EMA data, TD is defined by the horizontal and vertical coordinates of the tongue dorsum sensor, while JAW is defined by the horizontal and vertical coordinates of the lower teeth sensor. In the ultrasound data, TD is the horizontal and vertical coordinates of DLC knot 5 and JAW is the mandibular short tendon knot (Strycharczuk *et al.*, 2025). We use these knots as possible ultrasound correlates of TD and JAW, but it was not possible to verify that these represent identical physical locations as the EMA sensors. Our analysis instead focused on how each signal captures the relative distances between vowels, rather than raw comparisons. Position and velocity trajectories were segmented into separate gestures, defined as an interval bounded by two zero-crossings in the velocity signal. Diphthongs and some long monophthongs can have two distinct velocity peaks (Strycharczuk *et al.*, 2024), and we also included closure and release gestures. We only retained trajectories for which there exists a matching EMA/ultrasound pair within a given sensor/dimension (e.g., TDx). In total, we analysed 2093 trajectories (630 TDx, 549 TDy, 504 JAWx, and 410 JAWy). The different trajectory counts for x/y dimensions are a consequence of their different movement dynamics, which is not an issue for our present analysis, where we compare parameter estimation separately within dimensions.

2.5 Parameter estimation and evaluation

We estimate the coefficients for the parameters b, k, and T of a linear harmonic oscillator

$$\ddot{x} + b\dot{x} + k(x - T) = 0, \tag{2}$$

using constrained least squares. We optimize over the generic objective function in Eq. (3), where $\dot{X}$ is a time series of derivatives, $\Theta(X)$ is a feature library comprised of the model parameters in Eq. (2), and Ξ is the coefficient matrix to be optimized,

$$\min_{\Xi} \frac{1}{2} ||\dot{X} - \Theta(X)\Xi||^2 \text{ subject to } C\,\xi = d. \tag{3}$$

We specifically solve for the acceleration of the system and integrate to obtain position and velocity trajectories that are evaluated against empirical data. We split the second-order differential equation into two first-order Eq. (1) ($y = \dot{x}$); and Eq. (2) ($\dot{y} = -by - kx$), with the first equation subject to the linear constraint $y \overset{!}{=} 1.0\dot{x}$ to reduce model complexity (Champion *et al.*, 2020). We use a maximum of 30 iterations to allow convergence of the optimization algorithm. After discovering optimal coefficients, we generate a simulated trajectory by solving a linear harmonic oscillator using the discovered coefficients and quantify fit between the modelled and empirical trajectories using R^2 values. This essentially follows the same process as in Kirkham (2025a), but without any thresholding parameters, meaning that all model terms are used in fitting to data.

3. Results

3.1 Model fit and parameter comparisons

Table 1 shows R^2 summary statistics for the fit between data and model predictions, with all variables at $R^2 \geq 0.9$. This suggests that good model fits can be achieved and that fitting accuracy is comparable between EMA and ultrasound, but that TD models fit slightly more accurately than JAW models. We visualize example velocity fits for TDx from each modality in Fig. 2, which represents three tokens selected using a fixed random seed. It is apparent that the fits are qualitatively similar between EMA and ultrasound, with some small errors in the model predictions for each trajectory. We also

Table 1. R^2 model fit statistics for each variable, which summarizes the accuracy of model fits to empirical data.

Variable	Modality	N	$\bar{R}^2$	R^2 σ	R^2 min, max
TDx	EMA	630	0.95	0.09	0.23, 1.00
	US	630	0.95	0.09	0.32, 1.00
TDy	EMA	549	0.94	0.09	0.38, 1.00
	US	549	0.95	0.10	0.37, 1.00
JAWx	EMA	504	0.90	0.15	0.27, 1.00
	US	504	0.93	0.13	0.29, 1.00
JAWy	EMA	410	0.92	0.12	0.36, 1.00
	US	410	0.93	0.12	0.32, 1.00

note slight variation in the underlying data between modalities. This is likely to arise from similar sources as in Fig. 1, where we observe small time lags or slight durational differences.

Parameter values from EMA and ultrasound were then compared using the Bayesian hierarchical regression model: $y_i \sim \mathcal{N}(\alpha + \alpha_s[s_i] + (\beta + \beta_w[w_i]) \cdot \text{modality}_i, \sigma)$, where y_i is an observation of the outcome variable, α is the intercept, β is the effect of modality (EMA/ultrasound), $\beta_w \sim \mathcal{N}(0, \tau_\beta)$ is a by-word random slope for the effect of modality, $\alpha_s \sim \mathcal{N}(0, \tau_\alpha)$ is a speaker-level random intercept, and all other priors are weakly informative $\mathcal{N} \sim (0, 2)$. We ran MCMC sampling for 1000 warm-up iterations and 2000 sampling iterations using 4 chains, with the step size initialized at 0.1. Models were fitted using Stan version2.36 (Stan Development Team, 2024). In all cases, EMA is the baseline variable, so the values represent how ultrasound differs from EMA.

Table 2 shows the mean effect of measurement modality on parameter estimation, along with 95% credible intervals. In summary, when $\beta < 0$, it means that the ultrasound-estimated parameter is on average lower than the EMA-estimated parameter, whereas when $\beta > 0$ the ultrasound-estimated parameter is on average higher than the EMA-estimated parameter. We find that the credible intervals cross-zero across every variable for k and b, suggesting no systematic difference between EMA and ultrasound in these parameters, largely due to a high degree of variability. The estimated T difference is much narrower, where TDx, JAWx, and JAWy have systematically lower estimated T values in ultrasound (i.e., all 95% CIs are below zero). The TDy ultrasound T values are on average higher than the EMA values ($\overline{\beta} = 0.29$), but this is the one case for T where the credible interval includes positive and negative values, indicating high uncertainty and no systematic differences.

3.2 Word-specific differences

We now investigate word-level effects to compare how the estimated parameters pattern between different words/vowels. This is important because EMA and ultrasound track points on the tongue in different ways, so we expect systematic effects of vowel height and anteriority. We visualize the difference between EMA and ultrasound modalities using the model's intercept and random slope coefficients, where EMA is the baseline. Note that each variable has a different range, so we focus on within-variable differences rather than between-variable comparisons.

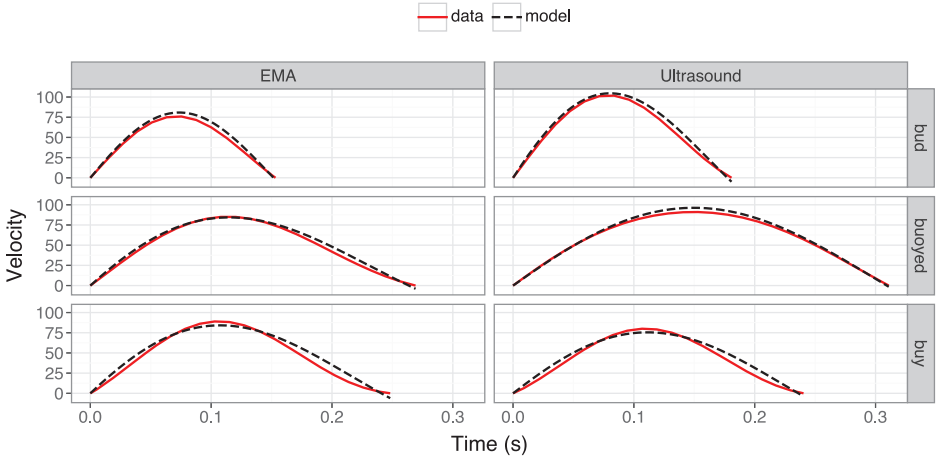

Fig. 2. A random sample of three example velocity fits between EMA and ultrasound for TDx. Tokens were selected using a fixed random seed, and each word represents the same underlying token produced by a speaker. All fits are $R^2 > 0.92$.

Table 2. Bayesian mean and 95% CIs for each parameter, which represents how much the ultrasound-estimated parameters deviate from the EMA-estimated parameters.

Parameter	Variable	$\bar{\beta}$	95% CI
T	TDx	−1.20	[−2.35, −0.06]
	TDy	0.29	[−0.32, 0.92]
	JAWx	−0.45	[−0.74, −0.17]
	JAWy	−1.36	[−1.81, −0.91]
k	TDx	−0.49	[−3.96, 2.74]
	TDy	0.86	[−2.38, 4.13]
	JAWx	2.29	[−0.81, 5.22]
	JAWy	0.38	[−2.88, 3.69]
b	TDx	−0.11	[−3.25, 3.07]
	TDy	0.07	[−3.18, 3.29]
	JAWx	1.91	[−1.21, 5.08]
	JAWy	0.13	[−3.00, 3.34]

Figure 3 shows word-level effects for the target parameter *T* in x/y space for TD and JAW, which represents the magnitude and direction of the difference between EMA and ultrasound. TD shows a systematic effect where front and high vowels, such as "bee," "bead," "booed," and "beer," have a lower and more posterior target for ultrasound parameters than EMA. Notably, these differences are consistent with previous research on how different ultrasound knots estimate vowel articulation, whereby the tongue dorsum knot underestimates the height and anteriority of front vowels (Strycharczuk *et al.*, 2025). The results for JAW also show a systematic difference, although over a smaller range. Ultrasound underestimates the JAW target relative to EMA for front vowels (e.g., "beed," "booed," "bid") and overestimates it in some low vowels (e.g., "bar"). This is likely an artifact of probe stabilization. In an ultrasound experiment, the probe is placed under the chin in its neutral position. When the jaw is lowered for the production of a low vowel, the soft tissue is squeezed against the probe, which can underestimate the distance between the short tendon and the probe, which leads to overestimation of the vertical JAW target.

Figure 4 shows word-level random slope coefficients for the stiffness parameter *k* and damping parameter *b*. In both cases, the majority of words cluster around zero with wide credible intervals. This indicates higher variability in measurement and a lack of systematic differences between EMA and ultrasound, except for a higher average *k* and lower average *b* in TDy for "bore." The JAW results show similar patterns, with near complete overlap, although note that "bar" was removed from the JAW plots (but not the modeling) due to extremely wide credible intervals that skewed the plotting range. Note that variability in *k* and *b* estimation occurs in both EMA and ultrasound data, so it is not necessarily the case that only one modality produces extreme estimates. Overall, this suggests that estimation of *k* and *b* is not systematically different between EMA and ultrasound, but that estimates are much more variable than for the target (*T*) parameter.

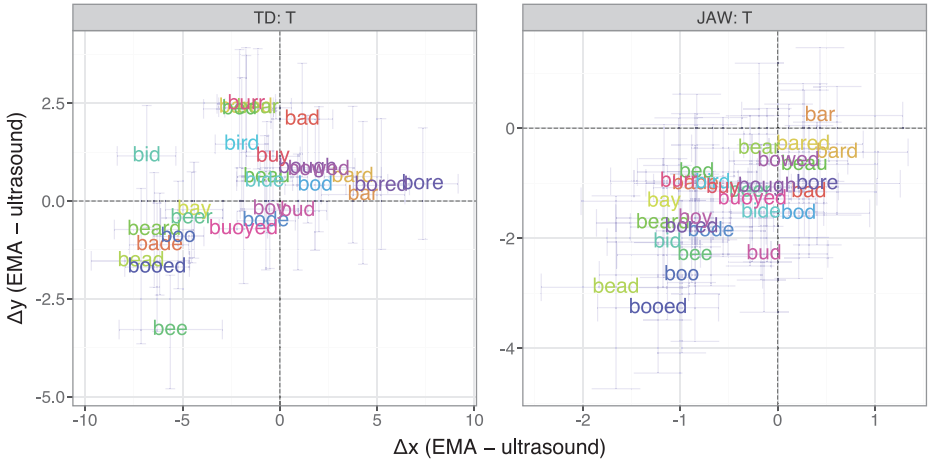

Fig. 3. By-word effects showing how ultrasound-estimated *T* (target) values differ from EMA-estimated values. Word labels show estimated means. Blue lines indicate 95% credible intervals. A value of zero indicates that ultrasound parameters do not differ from EMA parameters.

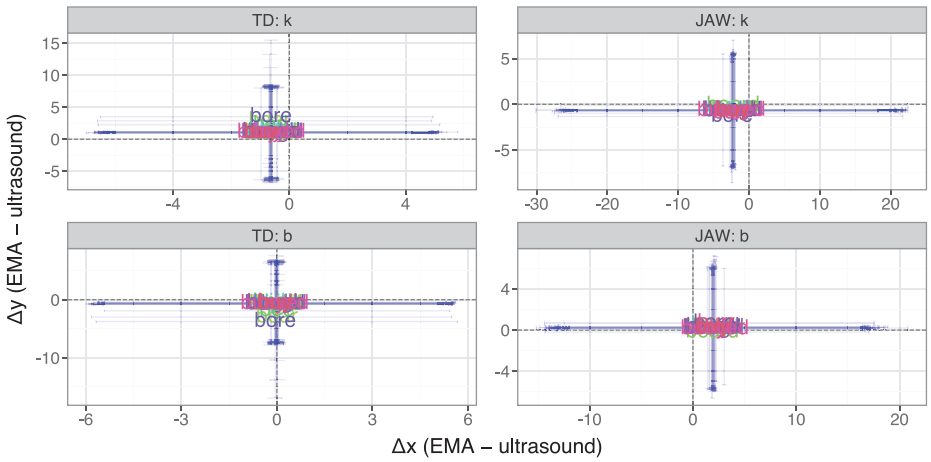

Fig. 4. By-word effects showing how ultrasound-estimated k (stiffness) and b (damping) values differ from EMA-estimated values. Word labels show estimated means. Blue lines indicate 95% credible intervals. A value of zero indicates that ultrasound parameters do not differ from EMA parameters. Note that the word "bar" has been removed from the JAW plots due to excessively large confidence intervals (crossing zero in both dimensions) that distort the axis ranges.

4. Discussion

We estimated the parameters of a linear dynamical model from EMA- and ultrasound-derived kinematic measurements. The model is a reasonably good fit to EMA and ultrasound kinematics for the tongue dorsum and jaw, which shows that dynamical models can be fitted to ultrasound kinematic data with comparable accuracy to EMA data, at least in the case of British English vowels. Our second aim was to establish the nature of the estimated parameters. We do find differences between EMA and ultrasound parameters, but these differences are predictable based on known characteristics of how the ultrasound knot tracking captures tongue movement. For example, previous research shows that ultrasound knot 5 underestimates dorsum height in front vowels, but other knots (e.g., anterior knot 7) inaccurately estimate dorsum height in low vowels (Strycharczuk *et al.*, 2025). A *post hoc* comparison shows that the T parameter estimated from TDy is more highly correlated with T estimated from knot 7Y ($r = 0.83$) than from knot 5Y ($r = 0.60$), where knot 7 captures a more anterior part on the tongue. This is because knot 7Y more substantially overestimates the target in low vowels in addition to the knot 5 high vowel differences, making the magnitude of EMA/ultrasound difference more linearly related to vowel height. In summary, while we find some differences in estimated tongue dorsum target parameters, these can be explained by the consequences of selecting a specific articulatory dimension to represent the tongue dorsum. Note also that EMA does not necessarily represent a golden standard in this respect, given flexibility in EMA sensor placement (Rebernik *et al.*, 2021). Moreover, ultrasound offers additional opportunities in tracking multiple points from tongue tip to root. As such, it is of paramount importance to report the precise dimensions used for model fitting.

Flexibility of point selection is less relevant for JAW, where ultrasound only tracks a single point. Despite this, JAW targets also vary between EMA and ultrasound, but in a way that is consistent. Ultrasound underestimates JAW targets in high vowels and overestimates in low vowels compared with EMA. Importantly, this variation is not random and is clearly related to both probe movement and vowel height, suggesting that the mandibular short tendon tracked in the ultrasound image captures systematic aspects of jaw position, but in a different manner to the EMA lower teeth sensor. This goes some way toward validating short tendon tracking as capturing meaningful aspects of vowel articulation, which significantly expands the utility of ultrasound imaging beyond the tongue surface.

The ultrasound kinematic data are noisier than EMA and the underlying signal may be slightly obscured even when smoothed, which could be another source of differences. Wrench and Balch-Tomes (2022) analyzed a subset of the data we used in this study and showed that horizontal tongue tip and blade measures correlate well between EMA sensors and DLC-tracked knots from ultrasound images ($r \geq 0.88$), but other articulatory dimensions have correlation coefficients $r < 0.4$. This suggests that EMA and ultrasound capture similar information in the tongue tip horizontal direction, but may capture differential spatial information for other dimensions. To this end, there is also scope for improving ultrasound kinematic measures, with an obvious area being better frame-to-frame tracking of DLC knots. At present, the DLC model estimates kinematics on a frame-by-frame basis, but inherent noise and measurement inaccuracy mean that knot tracking is unlikely to return a smooth function of time. One solution is to constrain tracking such that sharp divergences between frames are penalized. Finally, we focused on fitting a simple dynamical model to the data: an undamped linear harmonic oscillator. This allowed us to identify unique models for each token, with clearly interpretable parameters, but in future work it would be productive to compare more complex models, such as nonlinear dynamical models (Sorensen and Gafos, 2016; Stern and Shaw, 2025), as well as nonautonomous models with time-dependent parameters (Elie *et al.*, 2023).

5. Conclusion

We show that a linear task dynamic model can be fitted to ultrasound kinematic data with a relatively high degree of accuracy, conditional on sufficient smoothing and segmentation. The estimated model parameters differ in specific ways between EMA and ultrasound, but the differences are systematic and are a consequence of the selected measurement dimensions. As a result, these results broaden the possibilities for lingual kinematic analysis, with ultrasound providing information on the tongue posterior, as well as a greater number of points along the tongue. We also find that tracking the mandibular short tendon allows for meaningful jaw movement dynamics, opening up new directions for the study of inter-articulator coordination in dynamical theories of speech production.

Acknowledgments

This work was supported by AHRC Grants AH/S011900/1 (to P.S. and S.K.) and AH/Y002822/1 (to S.K.), and British Academy Grant MFSS24\40076 (to P.S.).

Author Declarations

Conflict of Interest

The authors have no conflicts of interest to declare.

Ethics Approval

This study received ethical approval from Lancaster University Ethics Committee (Ref: FL18188) and the University of Manchester's Proportionate University Research Ethics Committee (Ref: 2022–13946-22714).

Data Availability

All data and code are available at https://doi.org/10.5281/zenodo.17466633.

References

Browman, C. P., and Goldstein, L. M. (**1986**). "Towards an articulatory phonology," Phonology **3**(1), 219–252.
Byrd, D., and Saltzman, E. (**1998**). "Intragestural dynamics of multiple prosodic boundaries," J. Phonet. **26**(2), 173–199.
Champion, K., Zheng, P., Aravkin, A. Y., Brunton, S. L., and Kutz, J. N. (**2020**). "A unified sparse optimization framework to learn parsimonious physics-informed models from data," IEEE Access **8**, 169259–169271.
Elie, B., Lee, D. N., and Turk, A. (**2023**). "Modeling trajectories of human speech articulators using general Tau theory," Speech Commun. **151**, 24–38.
Fowler, C. A. (**1980**). "Coarticulation and theories of extrinsic timing," J. Phonet. **8**(1), 113–133.
Gafos, A. I. (**2006**). "Dynamics in grammar," in *Laboratory Phonology, Vol. 8: Varieties of Phonological Competence*, edited by L. Goldstein, D. Whalen, and C. T. Best (Mouton de Gruyter, Berlin, Germany), pp. 51–79.
Iskarous, K. (**2016**). "Compatible dynamical models of environmental, sensory, and perceptual systems," Ecol. Psychol. **28**(4), 295–311.
Kirkham, S. (**2024**). "Discovering dynamical models of speech using physics-informed machine learning," in *Proceedings of the ISSP: 13th International Seminar on Speech Production*, pp. 185–188.
Kirkham, S. (**2025a**). "Discovering dynamical laws for speech gestures," Cogn. Sci. **49**(5), e70064.
Kirkham, S. (**2025b**). "Scaling laws for nonlinear dynamical models of articulatory control," JASA Express Lett. **5**(2), 025201.
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 19th International Congress of Phonetic Sciences*, pp. 942–946.
Lammert, A., Goldstein, L., Narayanan, S., and Iskarous, K. (**2013**). "Statistical methods for estimation of direct and differential kinematics of the vocal tract," Speech Commun. **55**(1), 147–161.
Mathis, A., Mamidanna, P., Mury, 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*, pp. 498–502.
Rebernik, T., Jacobi, J., Jonkers, R., Noiray, A., and Wieling, M. (**2021**). "A review of data collection practices using electromagnetic articulograph," Lab. Phonol. **12**(1), 6.
Saltzman, E., and Munhall, K. G. (**1989**). "A dynamical approach to gestural patterning in speech production," Ecol. Psychol. **1**(4), 333–382.
Sorensen, T., and Gafos, A. I. (**2016**). "The gesture as an autonomous nonlinear dynamical system," Ecol. Psychol. **28**(4), 188–215.
Spreafico, L., Pucher, M., and Matosova, A. (**2018**). "UltraFit: A speaker-friendly headset for ultrasound recordings in speech science," in *Proceedings of Interspeech*, pp. 1–4.
Stan Development Team (**2024**). "Stan Reference Manual, version 2.36.0," https://mc-stan.org.
Stern, M. C., and Shaw, J. A. (**2025**). "Nonlinear second-order dynamics describe labial constriction trajectories across languages and contexts," J. Phonet. **111**, 101427.
Strycharczuk, P., Kirkham, S., Gorman, E., and Nagamine, T. (**2024**). "Towards a dynamical model of English vowels: Evidence from diphthongisation," J. Phonet. **107**, 101326–101349.

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**, 689–721.

Strycharczuk, P., and Scobbie, J. M. (**2015**). "Velocity measures in ultrasound data: Gestural timing of post-vocalic /l/ in English," in *Proceedings of the XVIII International Congress of Phonetic Sciences*, pp. 1–5.

Tilsen, S. (**2016**). "Selection and coordination: The articulatory basis for the emergence of phonological structure," J. Phonet. **55**, 53–77.

Wrench, A. (**2022**). "Articulate Assistant Advanced," version 220.04 [software].

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**, 1129–1133.