Loquens 10(1-2)
December 2023, e103
eISSN 2386-2637
https://doi.org/10.3989/loquens.2023.e103

Beyond the average: embracing speaker individuality in the dynamic modeling of the acoustic-articulatory relationship

Carolina Lins Machado[1], Lei He[1,2]

[1]Dept. Computational Linguistics, University of Zurich, Zurich, Switzerland
[2]Dept. Phoniatrics & Speech Pathology, Clinic for Otorhinolaryngology, Head & Neck Surgery, University Hospital Zurich (USZ), Zurich, Switzerland
cmachado@ifi.uzh.ch ORCID: 0000-0003-1001-7052
lei.he@uzh.ch ORCID: 0000-0002-9552-9075

Enviado: 31/10/2023; Aceptado: 26/12/2023; Publicado en línea: *01/02/2024*

Citation / Cómo citar este artículo: Carolina Lins Machado, Lei He (2023). Beyond the average: embracing speaker individuality in the dynamic modeling of the acoustic-articulatory relationship. *Loquens, 10*(1-2), e103, https://doi.org/10.3989/loquens.2023.e103

ABSTRACT: This paper explores the acoustic-articulatory relationship while considering individual differences in speech production. We aimed to determine whether there is a causal relationship between tongue movements and the contours of the first and second formant frequencies (F_1 and F_2) employing a hierarchical Bayesian continuous-time dynamic model, which allows for a more direct connection between the acoustic and articulatory measured variables and theories involving dynamicity. The results show predictive tendencies for both formants, where the anteroposterior and vertical tongue movements may predict changes in F_1, with rising predicting an increase and retraction a decrease; and with tongue fronting and tongue height inversely predicting F_2. Further, the modeled individual differences showed similar global tendencies, except for the rate of change of F_2. Overall, this study provides valuable insights into the relationship between tongue articulatory variables and formant contours, while accounting for between-speaker variability.

Keywords: individual differences, continuous-time modeling, formants, tongue kinematics

RESUMEN: *Más allá del promedio: abarcando la individualidad del hablante en la modelización dinámica de la relación acústico-articulatoria*. Este estudio explora la relación acústico-articulatoria de las diferencias individuales en la producción del habla. Nos propusimos determinar si existe una relación causal entre los movimientos de la lengua y los contornos del primer y segundo formantes (F_1 y F_2) empleando un modelo dinámico jerárquico bayesiano de tiempo continuo, lo que permite una conexión más directa entre las variables acústicas y articulatorias medidas y las teorías que implican dinamicidad. Los resultados muestran tendencias predictivas para ambos formantes, donde los movimientos anteroposteriores y verticales de la lengua pueden predecir cambios en F_1, con la elevación prediciendo un aumento y la retracción una disminución; y con el adelantamiento y la elevación de la lengua prediciendo inversamente F_2. Además, las diferencias individuales modeladas mostraron tendencias globales similares, excepto en el caso de la tasa de cambio de F_2. En general, este estudio presenta información relevante sobre la relación entre las variables articulatorias de la lengua y los contornos de los formantes, sin olvidar la variabilidad entre hablantes.

Palabras clave: diferencias individuales, modelización en tiempo continuo, formantes, cinemática lingual

1. INTRODUCTION

Formant dynamics are believed to carry important acoustic information pertaining to vowel identity, where differences in the trajectories of the first and second formant frequencies (F_1 and F_2, respectively) are shown to be important cues for the perception of vowels in a particular language (Nearey & Assmann, 1986; Hillenbrand & Nearey, 1999). For instance, the English vowel /æ/, is characterized by a slow and steady upward F_1 increase followed by a rapid decrease, and by a steady decrease in F_2 movement (Nearey, 2013). Anatomically, F_1 is more closely related to the back and F_2 to the front cavities of the oral tract, where a constriction in the vocal tract caused by the position of articulators, such as the tongue, dictates the shape of these cavities consequently affecting the values of both frequencies (Fry, 1979). Given this indirect relationship between articulatory position and formant values, the modulation of F_1 is broadly interpreted as the result of the vertical displacement of the tongue, where vertical tongue position is negatively correlated with this formant. Similarly, changes in the frequency of F_2 are believed to be more closely related to the anteroposterior tongue movement, where a more fronted tongue position results in higher F_2 values.

In addition, the inherent spectral changes, occurring in the formants as vowels are being produced, are believed to be a product of co-produced articulatory gestures in constant motion (Carré et al., 2017). As such, formant trajectories are thought to be the direct results of the dynamic nature of speech production and should be regarded and investigated as a dynamic process (Carré, 2009; Carré et al., 2017). However, although formant transitions have been shown to reflect, to some extent, articulatory motion (Lee, 2014; Dromey et al., 2013; Gorman & Kirkham, 2020), more often than not, the relationship between the movement of different articulators and the resulting dynamic acoustic output is proven difficult to be captured (e.g. Wieling, 2016), therefore, not always conforming with the acoustic-articulatory assumptions previously mentioned.

Among the reasons for this lack of clarity in the acoustic-articulatory relationship are the well demonstrated uncertainty related to the contribution of each articulator, or the different parts of a single articulator (e.g. tongue blade and dorsum) in the modulation of formant frequencies, the lack of a one-to-one mapping between acoustics and articulation, tied to the quantal theory of speech (Stevens, 1989), and the individual differences in the acoustic and articulatory domains, pertaining, for instance, to a speaker's anatomical and behavioral characteristics (Yang et al., 1996; McDougall, 2006; He et al., 2019). Nonetheless, shared associations between formant transitions and articulatory movements were demonstrated by means of correlation coefficients (e.g. Dromey, 2013; Lee et al., 2016), linear and non-linear regression models (e.g. Yunusova et al. 2012; Wieling, 2016), and Gaussian graphical models (Lins Machado et al., 2022), to name a few. Although these methods revealed some observed relationships between the acoustic and articulatory domains, causality between the two cannot be determined. One would think that failing to determine causation in statistics may be due to the variety of ways of thinking about causal relations, or the lack of a statistical syntax and semantics for expressing causality. However, theories such as "causal calculus" proposed by Judea Pearl (2009) offer a formal vocabulary and a collection of mathematical principles that allows the inference of causal relationships from observational and interventional data. Moreover, once a definition of causality is accepted, inferences about the causation between variables can be carried out (Granger, 1980).

In the context of this study, causality is defined and consequently investigated as "temporal (or Granger) causality", where time is the necessary structure for the definition of causality to hold. Under this structure "the present is caused by the past", based on the principles that causes occur before their effects and contain specific information about future consequences (Granger, 1980). The fundamental assumption of this definition is that if a time series X "Granger-causes" another time series Y, then past values of X should have information that helps predict Y beyond the information contained in past values of Y alone. To put it more simply, if X causes Y, then changes in X should occur before changes in Y.

When we examine Granger causality in the relationship between formant contours and the movements of different articulators (and parts thereof), we begin to consider that changes in acoustic may not be simultaneous but instead preceded by changes in articulation, even if by an extremely short amount of time. In fact, this is not such a far-fetched notion. The quantal theory of speech (Stevens, 1989) proposes that there are quantal regions in the vocal tract, where the acoustic signal is quite sensitive to relatively small changes in articulation. Thus, as an articulator continuously moves to achieve a certain acoustic output associated with these regions, the movement towards a quantal region can inform the expected changes in the acoustic signal. This, for example, would lead us to expect that tongue movements Granger-cause changes on vowel formants, since the tongue movement towards a vocal tract quantal region can indicate that at that point formants will undergo expected changes.

The problem remaining when trying to account for causality between the acoustic and articulatory domains is that of individual variability in these processes. Depending on the research context (e.g. investigating sociolectal differences or general theories of speech production), these differences tend not to be incorporated in the analyses. That may be due to the higher degree of variability found in the acoustic and articulatory processes, which could be the result of motor equivalences (Hughes & Abbs, 1976) tied to speaker-specific preferred articulatory strategies in the production of a particular linguistic sound (Johnson et al., 1993; McDougall, 2006; Y. Ji et al., 2017; Lins Machado et al., 2022). Yet, considering individual differences in speech production may provide valuable in-

sights into how language is used by individuals, subsequently exposing the underlying structures and patterns of a language not despite individual differences but by considering them (Josserand et al., 2021).

Therefore, the current study seeks to investigate whether a causal relationship between tongue movements and the contours of F_1 and F_2 can be found while incorporating the idiosyncratic information present in the articulatory movements and the acoustic output. The extent to which previous assumptions suggesting that tongue height may be considered the primary articulatory movement driving the changes in F_1, and tongue anteroposterior movement strongly modulating F_2 may likely be a result of previous investigations overlooking the dynamic element of speech or regarding individual differences as "noise". Thus, besides investigating a potential causal relation, the secondary aim of this study is to assess the stability of the previous assumptions, while considering the individual differences inherent to both processes. With regard to a possible causal link between articulatory tongue displacement and formant movement, we believe that causality in this link can, to some extent, be associated with tongue movements. However, the strength of this causal relationship will likely be influenced by individual differences pertaining to characteristic articulatory behaviors.

To explore causality between tongue movement and changes in F_1 and F_2 while considering individual differences, we adopted a hierarchical Bayesian continuous-time dynamic model. By modeling theories as continuous-time dynamic systems, this approach allows for a more direct connection between parameters and theories, formulating changes in terms of predicted transitions over time rather than direct consequences, and allowing for the representation of theories in a causal sense while taking into consideration the limited knowledge of process dynamics and potential model complexity updates (Driver & Tomasik, 2023). The benefit of this strategy is tied to how time and individual differences are handled. The following section is dedicated to explaining this method in further detail.

2. HIERARCHICAL BAYESIAN CONTINUOUS-TIME DYNAMIC MODELLING

In studies investigating dynamic information, the data are usually repeated measurements of the same constructs (concepts and variables under study). For instance, formant contours are characterized by extracting acoustic measurements at multiple time points over the course of a vowel. This sort of measurement allows us to gain insights of our constructs (formant contours) at each temporal interval. However, in many theories of change, it is assumed that the variables under study exist and develop continuously over time, and not solely at the measured occasions (Lohmann et al., 2022). Thus, by statistically modeling these continuously developing constructs we are able to more closely connect models with theories of change and to investigate how dynamic effects may develop (ibid.). The analysis of continuous-time processes and dynamics within and between individuals, is made possible through hierarchical Bayesian continuous-time dynamic models, where the constructs measured repeatedly over time yield a time series that when analyzed in this framework reveal information about a construct's continuous-time dynamics and trends.

Since continuous-time models treat time as continuous rather than discrete, information on dynamics and trends are not limited by time-interval dependency, but rather, processes are represented on a continuous-time scale and parameters are independent of specific intervals (Lohann et al., 2022). This means that parameter estimates are not solely related to a particular interval, but can be generalized to other time intervals, accounting for the continuous nature of the process under study and eliminating bias related to unequal intervals (Driver & Voelkle, 2018). This can be particularly advantageous when investigating acoustic and articulatory time series, since intervals between the measured instances vary due to differences in the length of a particular sound, or to individual differences, for instance.

Moreover, in a Bayesian hierarchical approach, the model structure is shared across all individuals and model parameters are allowed to vary, enabling subject-specific parameters estimation while fully utilizing participants' data to improve model estimates (Driver & Voelkle, 2021). These models take into account variations between individuals while employing shared characteristics to improve model estimates. This allows for the understanding of how parameters vary across a population, since the estimation of population-level parameters while accounting for individual differences is supported (Driver & Voelkle, 2018). Model parameter population distributions serve as a prior distribution for subject-level parameters. With this strategy, previous knowledge from all other subjects is used to aid in the parameter estimate for each unique individual. The key advantage of this technique is that variance and mean of the population distribution can be estimated alongside subject-level parameters, offering a good scope for random-effects over all model parameters (ibid.).

Mathematically, hierarchical Bayesian continuous-time dynamic models require differential calculus. Differential equations are the mathematics of continuous change limiting time to infinitesimally small values. This enables the usage of a temporal effects matrix that reflects the impact of a system's current state on the process' direction of change (Driver, 2022). In this study, the basic stochastic differential equation used in the statistical analysis can be represented as follows:

$$(1)\quad \mathrm{d}y(t) = (\mathrm{A}y(t) + b)\mathrm{d}t + G\mathrm{d}W(t)$$

The derivative $\mathrm{d}y(t)$ provides information on how the latent processes in the vector y are changing at the moment. On the right-hand side, this rate of change is explained by a deterministic term, describing trend components, and a

stochastic part, reporting the random fluctuations around the trends. In the deterministic part the drift matrix A represents how the latent state of the system changes over time characterizing the temporal dynamics of the processes under study. This matrix contains auto effects on its diagonals and cross effects on the off-diagonals. Auto effects describe how each system process determines its own future values and cross effects between processes explain how one process affects the future values of another. The continuous intercept b provides a constant fixed input to y specifying the long-term level around which the process fluctuates. Lastly, dt can be thought as a very small step in time.

In the stochastic part, allowing for uncertainty in the direction of change (Driver & Tomasik, 2023), d$W(t)$ represents the stochastic error term in continuous time (i.e. random fluctuations) and G the effect of this system noise on the change in $y(t)$, the process under study. The corresponding variance-covariance (or diffusion) matrix consists of the process error variances on the main diagonal as well as the process error covariances on the off-diagonals. For further conceptual and technical details see Driver and Voelkle (2018).

When the underlying system under study is believed to be continually changing and interacting, a continuous-time method is essential for its investigation. To illustrate this, consider the act of producing the vowel /æ/, where interactions occur continuously between the different parts of the tongue and its directions of movement: For instance, as the tongue moves backwards (x) and downwards (y) these connected movements affect each other (given the hydrostatic nature of the tongue) and in turn affect F_1 values (z). In this constructed example, the continuous-time temporal matrix A would be:

$$\begin{array}{c} \\ x \\ y \\ z \end{array} \begin{array}{c} \begin{array}{ccc} x & y & z \end{array} \\ \begin{bmatrix} -1 & -0.5 & 0 \\ -0.5 & -1 & 0 \\ 0 & -0.6 & -1 \end{bmatrix} \end{array}$$

Where the negative diagonal coefficients indicate that increases in any of the variables exerts a downwards pressure on the same variables in the future. This happens because systems tend to fluctuate around a range instead of stretching to infinity (Driver, 2022). The off-diagonals show where a change in one variable (determined by the column) leads to a change in another (determined by the row). Translating to our example, these cross-effects would indicate that a backward movement of the tongue (x) would elevate its dorsum, and the simultaneous jaw opening anatomically coupled with the tongue (y) would increase the first formant (z). Considering the relationship between these three continuous-time variables in this scenario allows us to analyze the 'Granger causality' of these relationships. That is, present formant values are caused by past articulatory movements.

3. METHOD

3.1. Materials

Productions of the vowel /æ/ in single-word citation form by twenty native speakers of U.S. English (10 M, 10 F) with an upper Midwest American English dialect background were selected from the EMA-MAE corpus (A. Ji et al., 2014). Selected materials and steps of acoustic and kinematic analysis are the same as Lins Machado et al. (2022). Measurements of F_1 and F_2 (in Herz) and of tongue movement displacement in x (anteroposterior) and y (superior-inferior) directions (in mm) of four kinematic variables TBx and TBy (relative to tongue blade), and TDx and TDy (relative to tongue dorsum) were extracted at nine equidistant points relative to vowel duration. However, only the five innermost analysis points were preserved for further analysis in an effort to reduce the impact of coarticulation from the neighboring consonants (Schwartz, 2021). Moreover, vowel tokens produced in the context of nasal, rhotic, lateral, and approximant syllable onset and codas were excluded from the analysis, since coarticulatory effects related to these consonants have been shown to affect vowel formants in a complex manner (Labov et al. 2006). It is important to mention that acoustic and kinematic measurements were manually inspected prior to extraction. In the case of formants, spectrograms and formant tracks were inspected and extraction parameters were adjusted per speaker and vowel token whenever necessary. The datasets consisted of 1240 data points, with token average duration of 0.254 s (sd = 0.0064 s; median = 0.246 s). Prior to statistical analysis the data was normalized (centered and scaled) per variable, and time was zero-shifted so the first analysis point is always 0.

Important to mention is that tongue displacements can contain contributions of the active tongue movement and the jaw passively moving the tongue, since anatomically the tongue and jaw are coupled. Consequently, the kinematic variables represent compound tongue-jaw movements, congruent to the tract variable of *tongue body constriction location* and *degree* in Articulatory Phonology (Browman & Goldstein, 1989).

3.2. Statistical analysis

Despite the fact that we only maintained 5 analysis points, continuous-time dynamic modeling supports the representation of continuous phenomena with a few data points by utilizing latent variables, which are constructs inferred from observed variables. This enables the representation of complex, continuous constructs allowing relationships between latent variables and their observable indicators to be established. Moreover, the continuous nature of a given phenomenon can be captured by mathematical equations in the model, which are able to represent how the latent constructs interact with one another and with the observed variables, providing insights about the underlying continuous processes. By including

latent variables and their interactions with observable indicators, continuous-time dynamic models may efficiently capture and model continuous phenomena even with a relatively small amount of time points (Oud & Voelkle, 2014).

To analyze changes in F_1 and F_2, the impact of the articulatory variables on both formants, and individual differences therein, a hierarchical Bayesian continuous-time dynamic model was implemented in R using the ctsem package (Driver et al., 2017). The model was set up using the `ctModel` function with the following arguments: The type of model `stanct`, allowing for a continuous time model for Bayesian fitting; `n.manifest`, defining the number of variables (measurement instances) to be analyzed in a given model and `n.latent` determining the number of process components we need to analyze the variables under study. In this study manifest and latent variables have the same name, since we want to see the direct effect of variables on each other.

Next, the model matrices `DRIFT`, related to the temporal dependencies of latent processes, and `DIFFUSION`, containing system noise, were automatically specified. `CINT`, the continuous-time random intercept vector and `T0MEANS`, a free parameter vector with random effects, were manually specified. Two additional arguments `MANIFESTMEANS` and `MANIFESTVAR`, were used to specify manifest components such as residuals. The final matrix, `LAMBDA`, relates the observed scores to the process components of the model, where all off-diagonals were set to 0 and diagonals to 1. The complete specification of the model is available at https://osf.io/tk56g

4. RESULTS

4.1. Continuous time parameter estimates

Continuous drift parameters describe how a process is changing. Autoregressive (AR) effects describe fluctuations in future time points carried over from a previous time point, describing how each process influences itself. In the context of this study AR effects describe how long deviations from the trend influence articulatory and acoustic variable values.

Figure 1 represents the AR effects of the acoustic and kinematic parameters and how they vary over time. Overall, the high absolute AR coefficients (Table 1) indicate the instability of these constructs, suggesting that when the system deviates from its expected deterministic trend a high downward pressure pushes it to return to the baseline levels. Group level AR effects showed that changes in F_1 are less persistent than for F_2 (drift_F1 = -12.58, 95% CI [-18.42, -6.69]; drift_F_2 = -8.59, 95% CI [-13.46, -4.03]). Regarding the articulatory variables, changes in the anteroposterior direction are less persistent than in the superior-inferior direction of both, tongue blade and dorsum, where, relatively speaking, TDx changes were the least persistent (drift_TDx = -10.47, 95% CI [-15.36, -5.46]) and TDy changes the most persistent (drift_TDy = -7.26, 95% CI [-12.58, -2.49]).

Figure 1: Discrete-time autoregressive effects of acoustic and articulatory variables for varying time intervals.

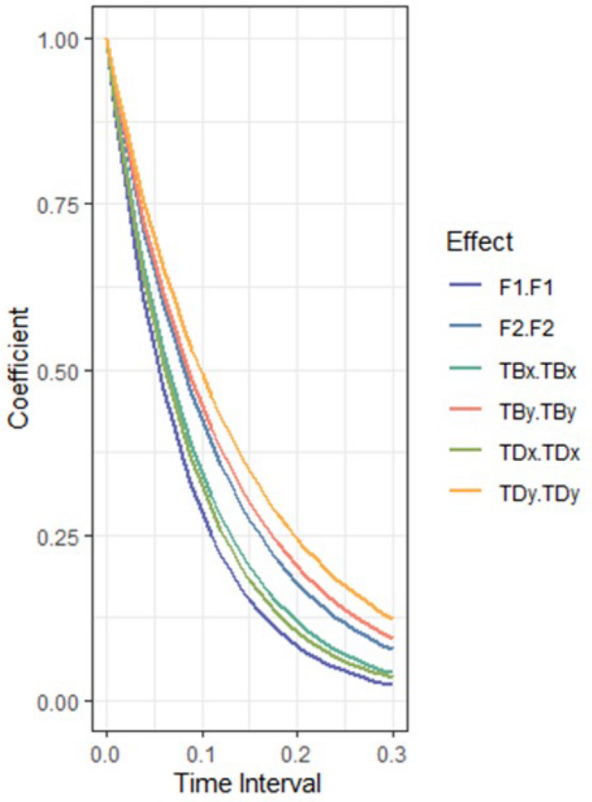


Table 1: Continuous auto-regressive and cross-lagged drift parameter estimates (Est.) and 95% Confidence Intervals (CI) of both formants and the four tongue variables. Effects not including the value of zero in the 95% CI were significant at the level .05.

Drift Parameters	Est.	95% CI	
		92.5%	97.5%
Auto-regressions			
drift_F1	-12.59	-18.42	-6.69
drift_F2	-8.59	-13.46	-4.03
drift_TBx	-10.13	-15.53	-4.90
drift_TBy	-8.28	-12.77	-3.92
drift_TDx	-10.47	-15.36	-5.46
drift_TDy	-7.26	-12.58	-2.49
Cross-regressions			
drift_F1_TBx	0.22	-1.78	2.22
drift_F1_TBy	0.05	-1.86	1.89
drift_F1_TDx	0.25	-1.63	2.20
drift_F1_TDy	-0.31	-2.24	1.58
drift_F2_TBx	-0.39	-2.20	1.47
drift_F2_TBy	-0.20	-2.00	1.56
drift_F2_TDx	-0.33	-2.33	1.55
drift_F2_TDy	-0.53	-2.47	1.45

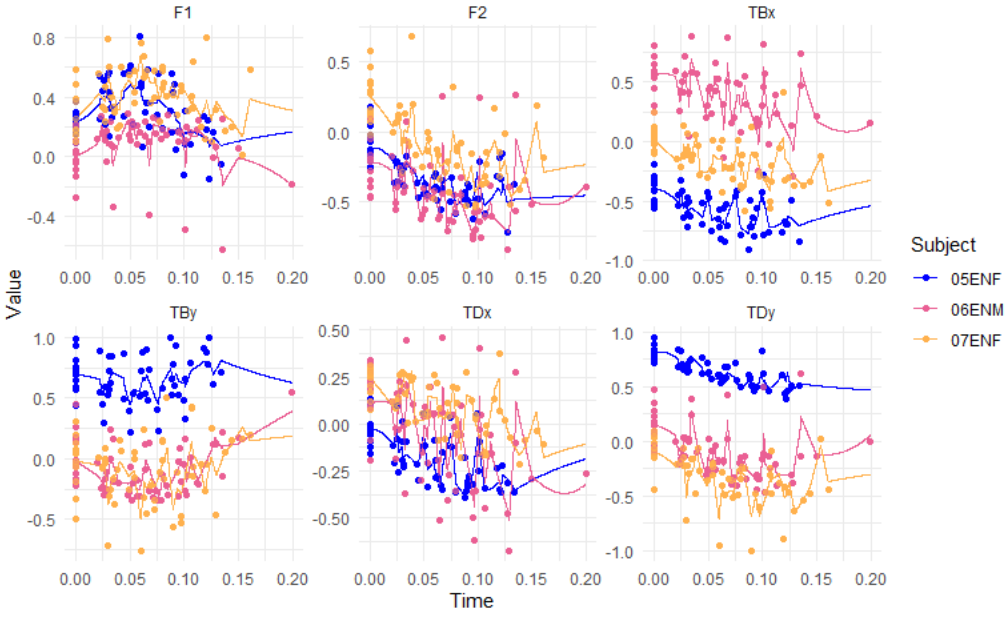
Cross-regressive (CR) effects illustrate the temporal dependencies and potential causal linkages between variables by showing how variables affect one another over time. An effect closer to (or of) zero reflects little to no influence of one variable on another. The direction of interaction between variables is given by the sign of the parameter estimates, where a positive coefficient indicates the same direction and a negative sign reflects opposite directions. CR effects between articulatory variables and the formants F_1 and F_2 indicate how each tongue variable predicted each formant. In the present analysis, there were no significant (at the α level = .05) CR effects between the articulatory variables and both formants. Nevertheless, non-significant results should not be deemed useless and unimportant, since they do not suggest the absence of an effect; rather they imply the lack of a statistically significant effect. Therefore, additional insights into possible effects of the tongue kinematic variables on these formants can still be provided, such as the robustness (or stability) of the associations between them. The following results provide a deeper understanding of the probabilistic behavior of these effects.

The results suggested that tongue raising negatively predicts changes in F_1; i.e. a higher tongue position likely decreases F_1, with an effect from the tongue dorsum (drift_F1_TDy = -.31, 95% CI [-2.24, 1.58]). The anteroposterior movement of the tongue seems to indicate that fronting the tongue positively predicts changes in F_1, that is, a more fronted tongue position likely increases this formant. In this direction, the tongue blade seemed to show a stronger effect on F_1 (drift_F1_TBx = .22, 95% CI [-1.78, 2.22]). Regarding changes on F_2, all tongue variables seem to negatively predict changes in this formant, with the strongest effects being from the vertical and anteroposterior displacement of tongue dorsum (drift_F2_TDy = -.53, 95% CI [-2.47, 1.45]; drift_F2_TDx = -.33, 95% CI [-2.33, 1.55]) and the anteroposterior displacement of the tongue blade (drift_F2_TBx = -.39, 95% CI [-2.20, 1.47]). These results suggest that when the tongue moves backwards and tongue dorsum lowers, F_2 likely increases.

4.2. Individual differences

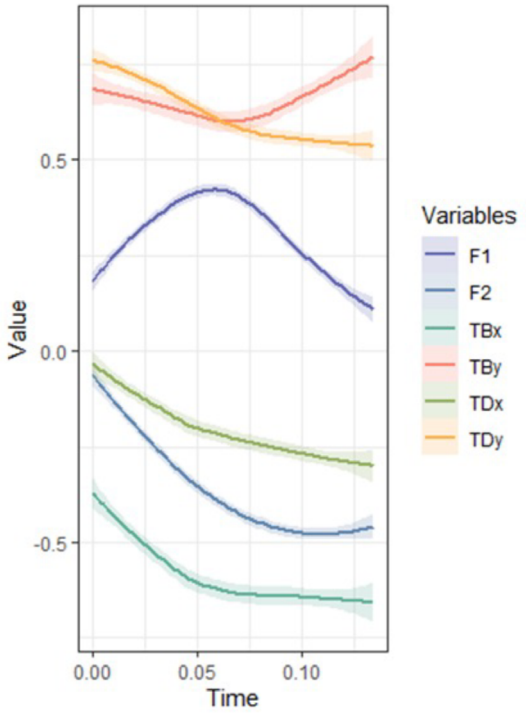
Subject-level parameters related to the initial latent states, or baseline (`T0`), and the continuous intercept, or slope (`CINT`), captured the variation among different subjects. Individual baselines capture the variation in the initial values of each variable across speakers and individual slopes represent person specific rates of change of each variable across time. The correlations between the initial latent states of the tongue variables (TBx_t0, TBy_t0, TDx_t0, TDy_t0) and formants' continuous intercepts (F1_cint and F2_cint) indicate the relationship between the rate of change these formants and the baseline values of articulatory variables across speakers. Regarding F_1, tongue dorsum variables related to the vertical and anteroposterior displacements positively covary with F1_cint ($r_{TDy\_t0\_F1\_cint}$ = .13, z = .47; $r_{TDx\_t0\_F1\_cint}$ = .46, z = 1.99), indicating that a relatively slower increase in F_1 is expected for speakers who start the production of this vowel with a higher and more fronted tongue dorsum position. With respect to F_2, both tongue dorsum variables (TDx and TDy) and tongue blade anteroposterior displacement negatively covary with the slope of this formant ($r_{TDx\_t0\_F2\_cint}$ = -.01, z = -.03; $r_{TDy\_t0\_F2\_cint}$ = -.23, z = -.69; $r_{TBx\_t0\_F2\_cint}$ = -.07, z = -.20), where a slow increase in F_2 is expected for speakers who start their production of this vowel with a lower tongue dorsum and a more retracted overall tongue position.

Figure 2: Observed data points and predicted trajectories (lines) of each acoustic and articulatory variable over the time course of the vowel for three random subjects.



Regarding the relationships between the continuous intercepts of the tongue variables (TBx_cint, TBy_cint, TDx_cint, TDy_cint) and the slope of each formant; i.e., the degree to which changes in tongue movement are associated with changes in the frequency of these formants, the results indicate a negative relationship between the rate of change of all articulatory variables and the slope of F_1. More specifically, given the negative correlation between F_1 and TBy ($r_{TBy\_cint\_F1\_cint}$ = -.08, z = -.24) and TDy ($r_{TDy\_cint\_F1\_cint}$ = -.22, z = -.70), a slow decrease in F_1 is expected for individuals whose tongue height slowly increases. Similarly, the negative correlation between the anteroposterior tongue displacement ($r_{TBx\_cint\_F1\_cint}$ = -.37, z = -1.37) would lead us to expect that F_1 decreases at a relatively slower rate when individuals' tongue blades slowly move forward. However, F_1 would be expected to slowly decrease when individuals slowly retract the tongue dorsum ($r_{TDx\_cint\_F1\_cint}$ = .23, z = .73). As for F_2, the slope of the tongue kinematic variables positively covary with the rate of change of this formant. Here, a slow increase in the slope of F_2 is expected when speakers slowly raise the tongue dorsum ($r_{TDy\_cint\_F2\_cint}$ = .31, z = .88) and slowly move their tongue forwards ($r_{TDx\_cint\_F2\_cint}$ = .29, z = .90; $r_{TBx\_cint\_F2\_cint}$ = .17, z = .52).

Figure 3: Expected trends of acoustic and articulatory variables before taking observations into account.



Individual trajectories of the acoustic and tongue kinematic variables are displayed in Figure 2 representing the observed data for 3 speakers and model predictions. After accounting for individual variation in the expected trajectories of each variable, the model's estimated forward predictions are noticeably less smooth than their expected trends (Figure 3). Further, individual observations were not closely tracked by the model predictions, indicating significantly large measurement error estimates. Nevertheless, although speaker-specific characteristics influenced predictions, resulting in substantial fluctuations in the expected trajectory for these variables, the expected trend shape is still observed.

5. DISCUSSION

The present study used acoustic measures of F_1 and F_2 and kinematic measurements of tongue blade and dorsum displacements in the anteroposterior and superior-inferior directions to investigate a possible causal relation between these acoustic and articulatory variables and the individual dynamics in the production of the vowel /æ/ in a sample of native U.S. English speakers. Although statistically significant indications of causality were not demonstrated, the continuous-time modeling approach provided further insights into the dynamic acoustic-articulatory relationship. Further, by accounting for idiosyncratic information present in both domains the stability of this relationship could be investigated.

Regarding F_1, the model predictions followed the hypothesis that vertical tongue movement has an opposing relationship to this formant. Moreover, the results also suggested that the anteroposterior movement of the tongue blade may have an effect on F_1 of similar magnitude. These findings suggest that not only tongue height but also the anteroposterior tongue movement have a predictive effect on F_1. However, while raising predicts an increase, retraction predicts a decrease in F_1. These results make sense if we consider that in some varieties of U.S. English the vowel /æ/ has a diphthongal quality (Nearey, 2013), with F_1 slowly increasing with a rapid final decrease as the result of the compound raising and retracting movements. Further, after considering individual differences, the F_1–tongue movement relationship remained in line with previous assumptions, suggesting that the dynamic relationship between F_1 and tongue kinematic variables incorporates both height and retraction movements.

In terms of F_2, tongue fronting and tongue height inversely predicted this construct. A more overall fronted tongue indicated a subsequent decrease in F_2 and lower tongue dorsum predicted an increase in this formant. Alone, these results do not follow previous accounts postulating that forward and elevating tongue movements increase F_2. However, when interpreted in combination, they may be indicative of a possible shift in cavity association. That is, instead of the common association of F_2 with the front vocal tract cavity, its affiliation is likely to be with the cavity behind the constriction point for this vowel (Fant, 1980). Shifts in cavity affiliation happen due to the change in cavity length and constriction degree. The front and back cavities are connected by a region of significant cross-sectional area making the two interact. The narrower the constriction between these cavities the greater the acoustic impedance "uncoupling" the resonances of each cavity. When the constriction degree is broader, such

as in the vowel /æ/, the coupled cavities influence each other's resonances by reducing or increasing the resonant frequencies. Since these are related to the length of the associated cavity, they tend to be higher for shorter cavities and lower for longer ones, however, acoustic coupling can affect this to some extent. The formants F_1 and F_2 are said to have shifted in cavity affiliation due to a vocal tract configuration lowering the acoustic impedance between cavities. For instance, as the constriction location moves backwards, the back cavity becomes shorter than the front. Consequently, the back cavity resonance frequency rises to a certain level higher than the front cavity; at that level the back cavity resonance results in F_2 and the front cavity resonance results in F_1. Although coherent, this interpretation remains speculative, since an investigation of vocal tract area has not been carried out. Additionally, individual differences followed the same global tendencies except for the rate of change of F_2, which seems to indicate that a slower increase in F_2 is expected for speakers who slowly raise their tongue blade. These individual differences, however, seem to suggest that elevating tongue movements increase F_2 values of speakers in which this formant may be associated with a smaller front cavity.

Overall, the lack of statistically significant effects of tongue kinematic variables on F_1 and F_2 could be due to the effects of other unaccounted articulatory variables that are believed to affect formant values, such as tongue shape (Lee et al., 2015), and laryngeal movement, which most notably either increases or decreases F_1 values (Esling, 2005). Furthermore, the individual differences mostly followed previous assumptions and model predictions related to the relationship between tongue movement and formant outcomes while also highlighting the complexity of the associations between acoustic features and articulatory variables in these relationships, which we believe are the result of individual articulatory strategies essentially driven by speaker-specific anatomical characteristics and behavioral preferences (Hughes & Abbs, 1976; He et al., 2019; Lins Machado et al., 2022).

Finally, the major limitation of this study must be addressed, this being what the constructs F_1 and F_2 actually relate to. Formants are a result of the deformations in the vocal tract area, and although these are primarily done by the tongue, both the lips and the larynx are known to shorten and lengthen the vocal tract, consequently affecting cavity areas and subsequently the values of both formants. Future analysis should, therefore, try to include measurements of these articulators. Notwithstanding this limitation, the present study is a first attempt at explaining possible causal relationships between tongue articulatory variables and the first two formant frequencies, while accounting for its dynamics and the individual differences therein.

6. ACKNOWLEDGEMENTS

This work was supported by the Swiss National Science Foundation (Grant #PZ00P1_193328) to LH. We wish to thank Dr. Charles Driver for the valuable feedback provided on earlier analysis stages. Any remaining errors are our own.

7. REFERENCES

Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. Phonology, 6(2), 201–251. https://doi.org/10.1017/S0952675700001019

Carré, R. (2009). Dynamic properties of an acoustic tube: Prediction of vowel systems. *Speech Communication*, *51*(1), 26–41. https://doi.org/10.1016/j.specom.2008.05.015

Carré, R., Divenyi, P., & Mrayati, M. (2017). *Speech: A dynamic process*. De Gruyter. https://doi.org/10.1515/9781501502019

Driver, C. C. (2022, January 14). *Inference With Cross-Lagged Effects—Problems in Time*. https://doi.org/10.31219/osf.io/xdf72

Driver, C. C., Oud, J. H. L., & Voelkle, M. C. (2017). Continuous Time Structural Equation Modeling with R Package ctsem. *Journal of Statistical Software*, *77*(5), 1–35. https://doi.org/10.18637/jss.v077.i05

Driver, C. C., & Tomasik, M. J. (2023). *Formalizing Developmental Phenomena as Continuous-Time Systems:Relations Between Mathematics and Language Development* [Journal Article]. https://osf.io/szx96

Driver, C. C., & Voelkle, M. C. (2018). Hierarchical Bayesian Continuous Time Dynamic Modeling. *Psychological Methods*, *23*(4), 774–799. https://doi.org/10.1037/met0000168

Driver, C. C., & Voelkle, M. C. (2021). Chapter 34—Hierarchical continuous time modeling. In J. F. Rauthmann (Ed.), *The Handbook of Personality Dynamics and Processes* (pp. 887–908). Academic Press. https://doi.org/10.1016/B978-0-12-813995-0.00034-0

Dromey, C., Jang, G.-O., & Hollis, K. (2013). Assessing correlations between lingual movements and formants. *Speech Communication, 55*(2), 315–328. https://doi.org/10.1016/j.specom.2012.09.001

Esling, J. H. (2005). There Are No Back Vowels: The Larygeal Articulator Model. *Canadian Journal of Linguistics/Revue Canadienne de Linguistique*, *50*(1–4), 13–44. https://doi.org/10.1017/S0008413100003650

Fant, G. (1980). *The Relations between Area Functions and the Acoustic Signal*. *37*(1–2), 55–86. https://doi.org/10.1159/000259983

Fry, D. B. (1979). *The Physics of Speech*. Cambridge University Press. https://books.google.ch/books?id=Ud-8yy-DCZgC

Gorman, E. F., & Kirkham, S. (2020). Dynamic acoustic-articulatory relations in back vowel fronting: Examining the effects of coda consonants in two dialects of British English. *The Journal of the Acoustical Society of America, 148*(2), 724.

Granger, C. W. J. (1980). Testing for causality: A personal viewpoint. *Journal of Economic Dynamics and Control*, *2*, 329–352. https://doi.org/10.1016/0165-1889(80)90069-X

He, L., Zhang, Y., & Dellwo, V. (2019). Between-speaker variability and temporal organization of the first formant. *The Journal of the Acoustical Society of America*, *145*(3), EL209–EL214. https://doi.org/10.1121/1.5093450

Hillenbrand, J. M., & Nearey, T. M. (1999). Identification of resynthesized /hVd/ utterances: Effects of formant contour. *The Journal of the Acoustical Society of America*, *105*(6), 3509–3523. https://doi.org/10.1121/1.424676

Hughes, O. M., & Abbs, J. H. (1976). Labial-Mandibular Coordination in the Production of Speech: Implications for the Operation of Motor Equivalence. *Phonetica*, *33*(3), 199–221. https://doi.org/doi:10.1159/000259722

Ji, A., Berry, J. J., & Johnson, M. T. (2014). The Electromagnetic Articulography Mandarin Accented English (EMA-MAE) corpus of acoustic and 3D articulatory kinematic data. *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 7719–7723. https://doi.org/10.1109/ICASSP.2014.6855102

Ji, Y., Wei, J., Zhang, J., Fang, Q., Lu, W., Honda, K., & Lu, X. (2017). Speech Behavior Analysis by Articulatory Observations. *Procedia Computer Science*, *111*, 463–470. https://doi.org/10.1016/j.procs.2017.06.048

Johnson, K., Ladefoged, P., & Lindau, M. (1993). Individual differences in vowel production. *The Journal of the Acous-*

tical Society of America, *94*(2), 701–714. https://doi.org/10.1121/1.406887

Josserand, M., Allassonnière-Tang, M., Pellegrino, F., & Dediu, D. (2021). Interindividual Variation Refuses to Go Away: A Bayesian Computer Model of Language Change in Communicative Networks. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.626118

Labov, W., Ash, S., & Boberg, C. (2006). *The atlas of North American English: Phonetics, phonology, and sound change: a multimedia reference tool*. Mouton de Gruyter.

Lee, J. (2014). Relationship between the first two formant frequencies and tongue positional changes in production of /aɪ/. *The Journal of the Acoustical Society of America, 135*(4_Supplement), 2294–2294. https://doi.org/10.1121/1.4877541

Lee, S.-H., Yu, J.-F., Hsieh, Y.-H., & Lee, G.-S. (2015). Relationships Between Formant Frequencies of Sustained Vowels and Tongue Contours Measured by Ultrasonography. *American Journal of Speech-Language Pathology*, *24*(4), 739–749. https://doi.org/10.1044/2015_AJSLP-14-0063

Lee, J., Shaiman, S., & Weismer, G. (2016). Relationship between tongue positions and formant frequencies in female speakers. *The Journal of the Acoustical Society of America, 139*(1), 426–440. https://doi.org/10.1121/1.4939894

Lins Machado, C., Dellwo, V., & He, L. (2022). Idiosyncratic lingual articulation of American English /æ/ and /ɑ/ using network analysis. *Interspeech 2022*, 754–758. https://doi.org/10.21437/Interspeech.2022-10397

Lohmann, J. F., Zitzmann, S., Voelkle, M. C., & Hecht, M. (2022). A primer on continuous-time modeling in educational research: An exemplary application of a continuous-time latent curve model with structured residuals (CT-LCM-SR) to PISA Data. *Large-Scale Assessments in Education*, *10*(1), 5. https://doi.org/10.1186/s40536-022-00126-8

McDougall, K. (2006). Dynamic features of speech and the characterization of speakers: Towards a new approach using formant frequencies. *International Journal of Speech, Language and the Law*, *13*(1), 89–126. https://doi.org/10.1558/sll.2006.13.1.89

Nearey, T. M. (2013). Vowel Inherent Spectral Change in the Vowels of North American English. In G. S. Morrison & P. F. Assmann (Eds.), *Vowel Inherent Spectral Change* (pp. 49–85). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-14209-3_4

Nearey, T. M., & Assmann, P. F. (1986). Modeling the role of inherent spectral change in vowel identification. *The Journal of the Acoustical Society of America*, *80*(5), 1297–1308. https://doi.org/10.1121/1.394433

Oud, J. H. L., & Voelkle, M. C. (2014). Do missing values exist? Incomplete data handling in cross-national longitudinal studies by means of continuous time modeling. *Quality & Quantity*, 48(6), 3271–3288. https://doi.org/10.1007/s11135-013-9955-9

Pearl, J. (2009). *Causality: Models, Reasoning and Inference* (2nd ed.). Cambridge University Press.

Schwartz, G. (2021). The phonology of vowel VISC-osity – acoustic evidence and representational implications. *Glossa: A Journal of General Linguistics*, *6*(1). https://doi.org/10.5334/gjgl.1182

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

Wieling, M., Tomaschek, F., Arnold, D., Tiede, M., Bröker, F., Thiele, S., Wood, S. N., & Baayen, R. H. (2016). Investigating dialectal differences using articulography. *Journal of Phonetics, 59*, 122–143. https://doi.org/10.1016/j.wocn.2016.09.004

Yang, X., Millar, J. B., & Macleod, I. (1996). On the sources of inter- and intra- speaker variability in the acoustic dynamics of speech. *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP '96, 3*, 1792–1795 vol.3. https://doi.org/10.1109/ICSLP.1996.607977

Yunusova, Y., Green, J. R., Greenwood, L., Wang, J., Pattee, G. L., & Zinman, L. (2012). Tongue movements and their acoustic consequences in amyotrophic lateral sclerosis. *Folia Phoniatrica et Logopaedica : Official Organ of the International Association of Logopedics and Phoniatrics (IALP), 64*(2), 94–102. https://doi.org/10.1159/000336890