

Task-dependence of articulator synergies^{a)}

Tanner Sorensen,^{1,b)} Asterios Toutios,¹ Louis Goldstein,² and Shrikanth Narayanan¹

¹Signal Analysis and Interpretation Laboratory, Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, California 90089, USA

²Department of Linguistics, University of Southern California, Los Angeles, California 90089, USA

(Received 18 May 2018; revised 15 February 2019; accepted 19 February 2019; published online 25 March 2019)

In speech production, the motor system organizes articulators such as the jaw, tongue, and lips into synergies whose function is to produce speech sounds by forming constrictions at the phonetic places of articulation. The present study tests whether synergies for different constriction tasks differ in terms of inter-articulator coordination. The test is conducted on utterances [apa], [ata], [aia], and [aka] with a real-time magnetic resonance imaging biomarker that is computed using a statistical model of the forward kinematics of the vocal tract. The present study is the first to estimate the forward kinematics of the vocal tract from speech production data. Using the imaging biomarker, the study finds that the jaw contributes least to the velar stop for [k], more to pharyngeal approximation for [a], still more to palatal approximation for [i], and most to the coronal stop for [p]. Additionally, the jaw contributes more to the coronal stop for [t] than to the bilabial stop for [p]. Finally, the study investigates how this pattern of results varies by participant. The study identifies differences in inter-articulator coordination by constriction task, which support the claim that inter-articulator coordination of the active articulator synergy.

© 2019 Acoustical Society of America. https://doi.org/10.1121/1.5093538

Pages: 1504-1520

coordination by constriction task support the task-dependence

articulator synergies by quantifying the percent contribution

of each articulator to narrowing or widening the vocal tract

at the synergy's place of articulation. In the task dynamics

model of speech production (Saltzman and Munhall, 1989),

the percent contribution of each articulator in a synergy is

determined by assigning weights to the articulators. In con-

trast to studies that manually assign weights to the articula-

tors based on theoretical considerations (for example, see

Simko and Cummins, 2010, for an assignment of weights

based on articulator mass), the present study is the first to

obtain a quantitative readout of these weights from speech

achieved a balance among the competing factors of tempo-

ral resolution, spatial resolution, and signal-to-noise ratio

that provides a rich source of speech production data for

the present study (Scott *et al.*, 2014). Real-time MRI pulse sequences and reconstruction techniques allow for the capture and visualization of the motion of the jaw, tongue, and

lips with 12 ms temporal resolution (Lingala et al., 2017;

Toutios and Narayanan, 2016). The present study uses

real-time MRI to quantify articulator synergies in terms of

the percent contribution of each articulator to producing

constrictions. The proposed measure of articulator syner-

gies is derived from *in vivo* MRI as a descriptor of articula-

tor synergies (i.e., a quantitative imaging biomarker, "an

objective characteristic derived from an in vivo image

measured on a ratio or interval scale as an indicator of nor-

mal biological processes," Kessler et al., 2015; Sullivan

Advances in magnetic resonance imaging (MRI) have

The present study investigates the task-dependence of

of articulator synergies (Latash, 2008).

I. INTRODUCTION

[ZZ]

An articulator synergy is a functional grouping of articulators such as the jaw, tongue, and lips, whose coordinated movements produce constrictions during speech, and which instantiates a reduction in the number of independent degrees of freedom for controlling a vocal tract movement (Turvey, 1977). Phonetic studies have shown that the coordination of articulators differs depending on where in the vocal tract the synergy produces a constriction. For instance, mechanical perturbations to jaw position during a bilabial stop induce compensatory lip movement with no compensatory tongue movement, whereas mechanical perturbations to jaw position during a lingual constriction induce compensatory tongue movement with no compensatory lip movement (Kelso et al., 1984). Our previous study on unperturbed speech suggests that healthy adult speakers of American English may use the jaw more for bilabial stops, coronal stops, and palatal approximations than for velar stops and pharyngeal approximations (Sorensen et al., 2016). Differences in inter-articulator

1504 J. Acoust. Soc. Am. **145** (3), March 2019

et al., 2015).

production data.

^{a)}Portions of this work were presented in "Factor analysis of vocal-tract outlines derived from real-time magnetic resonance imaging data," International Congress of Phonetic Sciences, Glasgow, UK, 2015; "Characterizing vocal tract dynamics using real-time MRI," LabPhon 2016, Ithaca, NY, 2016; "Characterizing vocal tract dynamics across speakers using real-time MRI," Proceedings of Interspeech, San Francisco, CA, 2016; "Decomposing vocal tract constrictions into articulator contributions using real-time MRI," Proceedings of the 7th International Conference on Speech Motor Control, Groningen, the Netherlands, 2017; and "Test-retest repeatability of articulatory strategies using real-time magnetic resonance imaging," Proceedings of Interspeech 2017, Stockholm, Sweden, 2017.

^{b)}Also at Department of Linguistics, University of Southern California, Los Angeles, California 90089, USA. Electronic mail: tannersorensen9991@ gmail.com

The algorithm for computing the articulator synergy biomarker involves a statistical model of the forward kinematics of the vocal tract. The forward kinematics relates articulator parameters to constriction task variables, as in the task dynamics model of speech production (Saltzman and Munhall, 1989; see Lammert et al., 2013a, for a procedure of estimating the forward kinematics of the vocal tract from synthetic data). The forward kinematics has two parts: the direct and differential kinematics. The direct kinematics expresses the degree of constriction (i.e., constriction task variable, measured in millimeters) at the phonetic places of articulation as a function of the position and shape of articulators. This function is called the forward kinematic map. The differential kinematics expresses change in the constriction task variables as a function of small increments of articulator movement. This function is the Jacobian matrix of the forward kinematic map. The algorithm uses the Jacobian matrix of the forward kinematic map to compute the percent contribution of each articulator to narrowing or widening the vocal tract at the synergy's place of articulation.

The research goals of the present study are (i) to estimate and evaluate the forward kinematics of the vocal tract from MRI, (ii) to design and evaluate a biomarker of articulator synergies based on the forward kinematics, and (iii) to use the articulator synergy biomarker to test the taskdependence of articulator synergies by determining whether the relative contribution of the jaw, tongue, and lips differs by constriction task.

The paper is organized as follows. Section II describes the MRI experiment, scanner sequence, participant characteristics, and method for manually annotating the start and end time-points in the real-time MRI videos. Sections III and IV describe the segmentation of articulator contours in the images, and use the segmentation results to estimate constriction task variables and parameters of articulator shape and position, which are related by the forward kinematics of the vocal tract. Section V estimates the forward kinematics and evaluates the model through crossvalidation. Section VI defines the articulator synergy biomarker. It evaluates the articulator synergy biomarker with respect to bias and precision. Section VII tests the effect of constriction task on the articulator synergy biomarker. Confirming this effect shows that inter-articulator coordination differs by constriction task, supporting the taskspecificity of articulator synergies. Section VIII investigates differences by participant in the effect of constriction task on the articulator synergy biomarker as well as intraparticipant variability in biomarker values. Sections IX and X offer discussion and conclusions.

II. MRI

A. Experiment

The data-set included eight (four male, four female) speakers of American English (Töger *et al.*, 2017). Five participants were native speakers of American English. None of the participants reported speech pathology or abnormal hearing. Table I provides participant characteristics. Each

TABLE I. Participant characteristics of the test-retest data-set.

Identification	Age	Gender	State of origin	Native language		
F1	25	F	Rhode Island	American English		
F2	28	F	Texas	American English		
F3	24	F	Nebraska	American English		
F4	29	F	Korea	Korean		
M1	29	М	Iowa	American English		
M2	27	М	United Arab Emirates	American English		
M3	26	М	Germany	German		
M4	39	М	Greece	Greek		
	Median: 28 Range: 24–39	4 male 4 female		5 American English 3 other		

participant took part in one session. The authors explained the nature of the experiment and the protocol to the participant before each scan. The participant lay on the scanner table in a supine position. The head was fixed in place by foam pads inserted between the temple and the receiver coil on the left and right sides of the head. The participant read visually presented text from a paper card taped to the scanner bore in front of the face. The speech corpus included real-time MRI videos of the isolated utterances [apa], [ata], [aka], and [aia] produced in an unrandomized sequence. Although participants were instructed to produce a low back unrounded vowel [a] and high front unrounded vowel [i], there was some variability in whether the low vowel was the front [a] or back [a] and whether or not the high vowel was produced as a glide [j]. Participants produced the sequence of utterances ten times. The authors removed the participant from the scanner for a short break, and then repeated the experiment. After completing the session, the speaker was paid for participation in the study. The University of Southern California (USC) Institutional Review Board approved the data collection procedures.

Vocal tract constrictions were manually identified in the real-time MRI videos. The video frames were inspected on a computer monitor. Guided by graphical presentation of real-time MRI video frames and auditory presentation of a denoised speech audio signal recorded in the scanner bore, the authors manually identified the intervals of time during which the vocal tract produced the bilabial stop in [apa], coronal stop in [ata], palatal approximation in [aia], velar stop in [aka], and pharyngeal approximation in [aia] (second [a] used for pharyngeal approximation). The authors annotated the frame number of the first and last frames in which there was visible movement of the lips (for [apa]) or tongue (for [ata], [aka], and [aia]).

B. Imaging parameters

Data were acquired on a Signa Excite HD 1.5 T scanner (General Electric Healthcare, Waukesha WI) with gradients capable of 40 mT/m amplitude and 150 mT/m/ms slew rate. A custom eight-channel upper airway coil was used for radio frequency signal reception. The coil had two four-channel arrays. A real-time MRI pulse sequence based on a spiral fast gradient echo pulse sequence was used. The real-time MRI pulse



FIG. 1. (Color online) (a) Subjectspecific templates for $[\alpha]$, [i], [p], [t], [k], which are automatically deformed to fit the articulator contours in the real-time magnetic resonance images. (b) Articulator contour segmentation of a sequence of real-time magnetic resonance images acquired in the transition from $[\alpha]$ to [i] in $[\alpha i\alpha]$. Frame rate downsampled for presentation.

sequence parameters were the following: $200 \text{ mm} \times 200 \text{ mm}$ field of view, $2.4 \text{ mm} \times 2.4 \text{ mm}$ reconstructed in-plane spatial resolution, 6 mm slice thickness, 6 ms repetition time (TR), 3.6 ms echo time (TE), 15° flip angle, 13 spiral interleaves for full sampling. The scan plane was manually aligned to the head. Images were retrospectively reconstructed to a temporal resolution of 12 ms (6 ms TR times two spirals per image, 83 frames per second). Reconstruction was performed using the Berkeley Advanced Reconstruction Toolbox (Uecker *et al.*, 2015).

III. CONSTRICTION TASK VARIABLE MEASUREMENT

The contours of articulators were identified in the realtime MRI videos and tracked automatically during vocal tract constrictions (Bresch and Narayanan, 2009). The algorithm was manually initialized with templates matching vocal tract contours during the sounds [0], [i], [p], [t], [k] (Fig. 1). If visual inspection revealed clear errors, then the algorithm initialization was manually corrected and the contours were re-submitted to the algorithm. This was repeated as needed until no clear contour tracking errors remained. See Fig. 1 for example segmentation results.

An algorithm automatically measured constriction task variables at the phonetic places of articulation in each video frame. As we use the term in this study, a constriction task variable is defined as the shortest distance between opposing structures at a given place of articulation. The opposing structures were the upper and lower lips for [p] (bilabial constriction task variable), tongue and coronal place for [t] (coronal constriction task variable), tongue and palatal place for [i] (palatal constriction task variable), tongue and soft palate for [k] (velar constriction task variable), and tongue and rear pharyngeal wall for [α] (pharyngeal constriction

task variable). The contour tracking algorithm automatically identified the upper lip, lower lip, tongue, hard palate, soft palate, and rear pharyngeal wall (Bresch and Narayanan, 2009). The anterior 1/4 of the hard palate was the coronal place of articulation. The posterior 1/2 of the hard palate was the palatal place of articulation. The velar place was bounded anteriorly by the anterior edge of the soft palate and extended posteriorly over 1/8 of the total soft palate contour, which included the oral, uvular, oropharyngeal, and nasal surfaces of the soft palate (cf. gray soft palate contours in Fig. 1). The pharyngeal place was the posterior pharyngeal wall, bounded superiorly by the velopharyngeal port and inferiorly by the larynx. Figure 2 illustrates the constriction task variable measurements at the phonetic places of articulation.

IV. GUIDED FACTOR ANALYSIS OF VOCAL TRACT SHAPES

A. Objective of the guided factor analysis

The guided factor analysis of the present study was motivated by the analysis of Maeda (1990), which extracts factors of the jaw, tongue, and lip articulators (Maeda's "elementary articulators") and factor scores that parameterize how the position and shape of these articulators change over time (Maeda's "elementary gestures"). In this framework, vocal tract movements are the sum of a few elementary gestures, which are factors in our analysis.

The objective of the guided factor analysis was to parameterize the vocal tract contours $\mathbf{X} \in \mathbb{R}^{n \times 2p}$, where *n* is the number of images and *p* is the number of contour vertices, as the linear combination of factors $\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_q \in \mathbb{R}^{2p}$ such that each factor characterizes spatial variation in the



FIG. 2. (Color online) Constriction task variable at the phonetic places of articulation in the transition from [a] to [i] in the sequence [aia]. The phonetic places of articulation (blue lines) are bilabial place, coronal place, palatal place, velar place, and pharyngeal place. Frame rate downsampled for presentation.

position and shape of an articulator (specifically, the jaw, tongue, lips). Prior to the factor analysis, the vocal tract contours **X** were centered on zero. The time-varying coefficients $\mathbf{w}_{\cdot,1}, \mathbf{w}_{\cdot,2}, \dots, \mathbf{w}_{\cdot,p} \in \mathbb{R}^n$ of the linear combination are factor scores that characterize temporal variation in the position and shapes of the articulators. Factor scores change from one image to the next as the articulators move and change shape. Thus, changes in the factor scores parameterize articulator motion. The guided factor analysis is based on the approach of Toutios and Narayanan (2015).

Factors $\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_q \in \mathbb{R}^{2p}$ are the columns of the matrix $\mathbf{F} \in \mathbb{R}^{2p \times q}$. Rows $\mathbf{w}_{1,\cdot}, \mathbf{w}_{2,\cdot}, ..., \mathbf{w}_{n,\cdot}$ of the matrix $\mathbf{W} \in \mathbb{R}^{n \times q}$ contain the factor scores for images 1, 2, ..., n of the real-time MRI data-set. Thus, the contour vertices $\mathbf{x}_{i,\cdot}$ for the *i*th image of the real-time MRI data-set are approximately equal to the following linear combination of the factors:

$$\hat{\mathbf{x}}_{i,\cdot} = \mathbf{w}_{i,\cdot} \mathbf{F}^{\mathsf{T}} \tag{1}$$

$$= w_{i,1}\mathbf{f}_{\cdot,1}^{\mathsf{T}} + w_{i,2}\mathbf{f}_{\cdot,2}^{\mathsf{T}} + \dots + w_{i,q}\mathbf{f}_{\cdot,q}^{\mathsf{T}}.$$
 (2)

The motion of the tongue and lips systematically cooccurs with motion of the jaw due to mechanical constraints and regularities in motor commands. For this reason, we seek a factor that corresponds to the motion of the jaw along with the concomitant motion of the tongue and lips.

The set of q_{jaw} jaw factors $\mathbf{F}_{jaw} \in \mathbb{R}^{2p \times q_{jaw}}$ parameterizes motion of the jaw along with concomitant motion of the tongue, lips, and velum. The set of q_{tongue} tongue factors $\mathbf{F}_{tongue} \in \mathbb{R}^{2p \times q_{tongue}}$ parameterizes motion of the tongue that is independent of the jaw. The set of q_{lips} lip factors \mathbf{F}_{lips} $\in \mathbb{R}^{2p \times q_{lips}}$ parameterizes motion of the lips that is independent of the jaw. The set of q_{velum} velum factors \mathbf{F}_{velum} $\in \mathbb{R}^{2p \times q_{velum}}$ parameterizes motion of the velum. The full set of factors can be written as the block matrix $\mathbf{F} \in \mathbb{R}^{2p \times q}$, where $q = q_{jaw} + q_{tongue} + q_{lips} + q_{velum}$,

$$\mathbf{F} = \begin{bmatrix} \mathbf{F}_{jaw} | \mathbf{F}_{tongue} | \mathbf{F}_{lips} | \mathbf{F}_{velum} \end{bmatrix}.$$
(3)

The corresponding set of factor scores can be written as the block matrix $\mathbf{W} \in \mathbb{R}^{n \times q}$, where $q = q_{\text{jaw}} + q_{\text{tongue}} + q_{\text{lips}} + q_{\text{velum}}$,

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_{jaw} | \mathbf{W}_{tongue} | \mathbf{W}_{lips} | \mathbf{W}_{velum} \end{bmatrix}.$$
(4)

Section IVB is a preliminary technical note. Section IVC derives the jaw factors. Section IVD derives the tongue, lip, and velum factors. Section IVE derives the factor scores.

B. Preliminaries to the guided factor analysis

Different steps of the guided factor analysis focus on different articulators. The guided factor analysis uses a projection operator to set to zero the contour vertices of articulators not under analysis in a given step of the analysis. For example, in order to derive the matrix $\mathbf{X}_{jaw} \in \mathbb{R}^{n \times 2p}$, which contains only jaw contour vertices, the guided factor analysis sets to zero the contour vertices (i.e., columns of **X**) of all non-jaw articulators and leaves the contour vertices of the jaw unchanged. Specifically, the non-jaw contour vertices are set to zero by multiplying **X** by the diagonal projection matrix $\mathbf{P}_{jaw} \in \mathbb{R}^{2p \times 2p}$. We have that $p_{i,i} = 1$ if the *i*th column of **X** is a jaw vertex. Otherwise, $p_{i,i} = 0$. This projection operator sets to zero the columns of **X** corresponding to non-jaw contour vertices. If jaw contour vertices are in columns q_1, q_2, \dots, q_ℓ of **X**, then the projection works out to the following:

$$\mathbf{X}_{jaw} = \mathbf{X} \mathbf{P}_{jaw} \begin{bmatrix} \overbrace{\mathbf{x}_{1,q_{1}} \cdots \mathbf{x}_{1,q_{\ell}} \mathbf{0}}^{jaw \text{ contour vertices}} \\ 0 \quad x_{1,q_{1}} \cdots \mathbf{x}_{1,q_{\ell}} \mathbf{0} \\ 0 \quad x_{2,q_{1}} \cdots \mathbf{x}_{2,q_{\ell}} \mathbf{0} \\ \vdots \quad \vdots \quad \vdots \quad \vdots \\ 0 \quad x_{n,q_{1}} \cdots \mathbf{x}_{m,q_{\ell}} \mathbf{0} \end{bmatrix}$$
$$= \mathbf{X} \begin{bmatrix} \overbrace{\mathbf{0}}^{\text{columns}} q_{1}, \dots, q_{\ell} \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}.$$
(5)

Similarly, the matrices $\mathbf{X}_{tongue}, \mathbf{X}_{lips}, \mathbf{X}_{velum} \in \mathbb{R}^{n \times 2p}$ focus on the tongue and lip contours. Summing such matrices produces a matrix that corresponds to a set of articulators. For instance, the matrix $\mathbf{X}_{jaw,tongue,lips} = \mathbf{X}_{jaw} + \mathbf{X}_{tongue} + \mathbf{X}_{lips}$ corresponds to the jaw, tongue, and lips.

C. Jaw factors

We first obtained the factors \mathbf{F}_{jaw} that capture the contribution of the jaw to vocal tract shaping (see the jaw factor in Fig. 3). We performed principal component analysis of the jaw (i.e., mandible and chin, cf. Fig. 1) contour vertices through eigendecomposition of the covariance matrix $\Sigma_{jaw} = \mathbf{X}_{jaw}^{\mathsf{T}} \mathbf{X}_{jaw}/(n-1)$ into an orthogonal matrix $\mathbf{Q}_{jaw} \in \mathbb{R}^{2p \times q_{jaw}}$ whose columns are the principal axes of \mathbf{X}_{jaw} and a diagonal matrix $\Lambda_{jaw} \in \mathbb{R}^{q_{jaw} \times q_{jaw}}$ whose diagonal entries are the variances of \mathbf{X}_{jaw} on the principal axes,

$$\Sigma_{jaw} = \mathbf{Q}_{jaw} \Lambda_{jaw} \mathbf{Q}_{jaw}^{-1}.$$
 (6)

The principal axes $(\mathbf{q}_{jaw})_{\cdot,1}(\mathbf{q}_{jaw})_{\cdot,2}, \ldots, (\mathbf{q}_{jaw})_{\cdot,q_{jaw}}$ and the variances $(\lambda_{jaw})_{1,1}, (\lambda_{jaw})_{2,2}, \ldots, (\lambda_{jaw})_{q_{jaw},q_{jaw}}$ on these axes capture the direction and variance of jaw motion. The jaw factors $(\mathbf{f}_{jaw})_{\cdot,1}, (\mathbf{f}_{jaw})_{\cdot,2}, \ldots, (\mathbf{f}_{jaw})_{\cdot,q_{jaw}}$ capture this jaw motion along with concomitant tongue and lip motion. The jaw factors are the columns of matrix $\mathbf{F}_{jaw} \in \mathbb{R}^{2p \times q_{jaw}}$,

$$\mathbf{F}_{jaw} = \Sigma_{jaw,tongue,lips} \mathbf{Q}_{jaw} \Lambda_{jaw}^{-1/2}.$$
(7)

Column $(\mathbf{f}_{jaw})_{\cdot,i}$ is the vector of covariances between the jaw, tongue, and lip contour vertices and the *z*-scored component scores $\mathbf{X}(\mathbf{q}_{jaw})_{\cdot,i}/\sqrt{(\lambda_{jaw})_{i,i}}$ for the *i*th jaw principal component. Thus, the factors capture motion of the tongue



FIG. 3. (Color online) Factors obtained for one participant with one jaw factor. four tongue factors, two lip factors, and one velum factor $(q_{jaw} = 1, q_{tongue} = 4,$ $q_{\rm lip} = 2$, $q_{\rm velum} = 1$). Each factor characterizes spatial variation in the position and shape of an articulator. The jaw factor captures jaw contour motion and concomitant tongue and lip motion. The tongue, lip, and velum factors capture tongue and lip contour motion that is not concomitant with jaw motion. The articulator contours in a given real-time magnetic resonance image are parameterized as the linear combinations of the factors. For a given participant, the coefficients of this linear combination change from image to image as the articulators move, while the factors remain constant.

and lips, which accompanies the motion of the jaw. Note that the matrix $\Lambda_{jaw}^{-1/2}$ is the inverse of the principal square root of Λ_{jaw} . Postmultiplying $\Sigma_{jaw,tongue,lips} Q_{jaw}$ by $\Lambda_{jaw}^{-1/2}$ z-scores the component scores XQ_{jaw} , whose covariances with the jaw, tongue, and lip contour vertices $X_{jaw,tongue,lips}$ are the entries of the jaw factors F_{jaw} .

The column space $\text{Col}(\mathbf{F}_{jaw})$ of \mathbf{F}_{jaw} is a q_{jaw} -dimensional subspace of \mathbb{R}^{2p} . Variance within $\text{Col}(\mathbf{F}_{jaw})$ captures jaw contour motion and concomitant tongue and lip motion [see Fig. 4(a)]. The projection $\hat{\mathbf{X}}$ of the data \mathbf{X} on the space $\text{Col}(\mathbf{F}_{jaw})$ is obtained through the Moore-Penrose pseudoinverse \mathbf{F}_{jaw}^+ of \mathbf{F}_{jaw} ,

$$\hat{\mathbf{X}} = \mathbf{X}\mathbf{F}_{jaw}\mathbf{F}_{jaw}^+. \tag{8}$$

The null space Null($\mathbf{F}_{jaw}^{intercal}$) is a $(2p - q_{jaw})$ -dimensional subspace of \mathbb{R}^{2p} . Variance within Null($\mathbf{F}_{jaw}^{intercal}$) captures velum motion along with the part of the tongue and lip motion that is independent of jaw motion. Section IV D describes the factors that characterize the variance of the tongue, lips, and velum within Null($\mathbf{F}_{jaw}^{intercal}$).

D. Tongue, lip, and velum factors

This section describes the procedure for obtaining the factors F_{tongue} , which capture the contribution of the

tongue to vocal tract shaping (see the tongue factors in Fig. 3). The projection \tilde{X} of the data matrix X on the space Null($F_{jaw}^{intercal}$) is the contour vertex motion that is independent of the jaw,

$$\mathbf{X} = \mathbf{X}(\mathbf{I}_p - \mathbf{F}_{jaw}\mathbf{F}_{jaw}^+) \tag{9}$$

$$= \mathbf{X} - \mathbf{X}.$$
 (10)

Specifically, $\tilde{\mathbf{X}}$ is independent of the jaw in the sense that it is statistically independent of $\hat{\mathbf{X}}$.

We performed principal component analysis of the tongue contour vertices through eigendecomposition of the covariance matrix $\tilde{\Sigma}_{tongue} = \tilde{\mathbf{X}}_{tongue}^{\mathsf{T}} \tilde{\mathbf{X}}_{tongue}/(n-1)$ into an orthogonal matrix \mathbf{Q}_{tongue} , whose columns are the principal axes of $\tilde{\mathbf{X}}_{tongue}$, and a diagonal matrix Λ_{tongue} , whose diagonal entries are the variances of $\tilde{\mathbf{X}}_{tongue}$ on the principal axes,

$$\tilde{\Sigma}_{\text{tongue}} = \mathbf{Q}_{\text{tongue}} \Lambda_{\text{tongue}} \mathbf{Q}_{\text{tongue}}^{-1}.$$
 (11)

The principal axes $(\mathbf{q}_{\text{tongue}})_{.,1}, (\mathbf{q}_{\text{tongue}})_{.,2}, ..., (\mathbf{q}_{\text{tongue}})_{.,q_{\text{tongue}}}$ and the variances $(\lambda_{\text{tongue}})_{1,1}, (\lambda_{\text{tongue}})_{2,2}, ..., (\lambda_{\text{tongue}})_{q_{\text{tongue}},q_{\text{tongue}}}$ on these axes capture the direction and variance of tongue motion, respectively.

Column $(\mathbf{f}_{tongue})_{,i}$ of the tongue factor matrix \mathbf{F}_{tongue} is the vector of covariances between the tongue contour vertices and



FIG. 4. (Color online) Percent variance explained (a) for the mandible and chin contours for each number of jaw factors, (b) for the tongue contour for different numbers of jaw and tongue factors, and (c) for the lip contours for different numbers of jaw and lip factors. Results averaged over participants.

the *z*-scored component scores $\mathbf{X}(\mathbf{q}_{\text{tongue}})_{.,i}/\sqrt{(\lambda_{\text{tongue}})_{i,i}}$ for the *i*th tongue principal component,

$$\mathbf{F}_{\text{tongue}} = \tilde{\Sigma}_{\text{tongue}} \mathbf{Q}_{\text{tongue}} \Lambda_{\text{tongue}}^{-1/2}.$$
 (12)

The column space $\text{Col}(\mathbf{F}_{\text{tongue}})$ of $\mathbf{F}_{\text{tongue}}$ is a q_{tongue} dimensional subspace of \mathbb{R}^{2p} . Variance within $\text{Col}(\mathbf{F}_{\text{tongue}})$ captures tongue contour motion that is not concomitant with jaw motion [see Figs. 4(b) and 4(c)]. The procedure for deriving lip and velum factors is the same as for the tongue factors except that \mathbf{X}_{lips} or $\mathbf{X}_{\text{velum}}$ is substituted for $\mathbf{X}_{\text{tongue}}$.

E. Factor scores

According to Eq. (1), the data matrix \mathbf{X} is parameterized as the matrix product \mathbf{WF}^{T} . Sections IV C and IV D specify the factors \mathbf{F} . This section derives the factor scores \mathbf{W} from the factors \mathbf{F} and the data matrix \mathbf{X} ,

$$\mathbf{W} = \mathbf{X}(\mathbf{F}^{\mathsf{T}})^{+}.$$
 (13)

Superscript "+" denotes the Moore-Penrose pseudoinverse. In image *i* of the real-time MRI data-set, the vocal tract contours $\mathbf{x}_{i,.}$ is approximately equal to the linear combination $w_{i,1}\mathbf{f}_1^{\mathsf{T}} + w_{i,2}\mathbf{f}_2^{\mathsf{T}} + \cdots + w_{i,q}\mathbf{f}_q^{\mathsf{T}}$. The factor scores $w_{i,.}$ are the coefficients of the linear combination. Variance of the factor scores parameterizes the temporal variability of vocal tract shape.

V. FORWARD KINEMATIC MAP OF THE VOCAL TRACT

A. Estimation of the direct and differential kinematics

The forward kinematic map is the function that maps the shape and position of the articulators (here, parameterized by factor scores) to the corresponding constriction task variables at the phonetic places of articulation (Lammert *et al.*, 2013a). Although the function is nonlinear, in the neighborhood of a given point $\mathbf{w}_{j,\cdot}$ the function can be linearized to have the form of a linear system of equations. Specifically, the linearized forward kinematic map is the function **G**, where $\mathbf{w} \in \mathbb{R}^q$ is a vector of *q* factor scores and $\mathbf{z} \in \mathbb{R}^m$ is a vector of constriction task variables at the *m* phonetic places of articulation,

$$\mathbf{z} = \mathbf{G}(\mathbf{w}) \begin{bmatrix} 1\\ \mathbf{w} \end{bmatrix}$$
(14)

$$= \begin{bmatrix} \boldsymbol{\mu}(\mathbf{w}) & \mathbf{J}(\mathbf{w}) \end{bmatrix} \begin{bmatrix} 1 \\ \mathbf{w} \end{bmatrix}.$$
(15)

The first column $\mathbf{g}_{\cdot,1}$ of the matrix $\mathbf{G}(\mathbf{w}) \in \mathbb{R}^{m \times (q+1)}$ is the vector $\boldsymbol{\mu}(\mathbf{w}) \in \mathbb{R}^m$ of intercepts for each constriction task variable in the neighborhood of \mathbf{w} . The remaining columns $\mathbf{g}_{\cdot,2}(\mathbf{w}), \dots, \mathbf{g}_{\cdot,q+1}(\mathbf{w})$ define the Jacobian $\mathbf{J}(\mathbf{w})$ of the forward kinematic map in the neighborhood of \mathbf{w} .

$$\mathbf{J}(\mathbf{w}) = \begin{bmatrix} \frac{\partial \mathbf{z}_1}{\partial \mathbf{w}_1} & \cdots & \frac{\partial \mathbf{z}_1}{\partial \mathbf{w}_q} \\ \vdots & \ddots & \vdots \\ \frac{\partial \mathbf{z}_m}{\partial \mathbf{w}_1} & \cdots & \frac{\partial \mathbf{z}_m}{\partial \mathbf{w}_q} \end{bmatrix}.$$
 (16)

The Jacobian of the forward kinematic map indicates how much the constriction task variables change as the result of a small change in factor scores. The forward kinematic map is estimated using weighted least squares. The estimator $\hat{\mathbf{g}}_{i,\cdot}$ of row $\mathbf{g}_{i,\cdot}$ of **G** is defined locally at point **w** as the function that minimizes the weighted sum of squared errors (SSE)

$$SSE = (\mathbf{z}_{\cdot,i} - \hat{\mathbf{z}}_{\cdot,i})^{\mathsf{T}} \mathbf{C}(\mathbf{w}) (\mathbf{z}_{\cdot,i} - \hat{\mathbf{z}}_{\cdot,i})$$
(17)

for i = 1, 2, ..., m, where $\mathbf{z}_{.,i} \in \mathbb{R}^n$ is the vector of *n* constriction task variables measured at the *i*th constriction location; *n* is the number of images in the data-set; $\hat{\mathbf{z}}_{.,i} \in \mathbb{R}^n$ is the corresponding vector of *n* estimated constriction task variables at the *i*th constriction location; and entry $c_{k,k}$ of the diagonal weight matrix $\mathbf{C}(\mathbf{w})$ is defined by the tri-cubic kernel function *K* centered in the neighborhood of \mathbf{w} ,

$$c_{k,k}(\mathbf{w}) = K(\mathbf{w}, \mathbf{w}_{k,\cdot})$$

$$= \begin{cases} \left(1 - \left(\frac{\|\mathbf{w}_{k,\cdot} - \mathbf{w}\|}{h}\right)^3\right)^3, & \|\mathbf{w}_{k,\cdot} - \mathbf{w}\| < h \\ 0, & \text{otherwise.} \end{cases}$$
(18)

(19)

The parameter *h* is the radius of a spherical neighborhood centered on **w** containing exactly $\lfloor fn \rfloor$ data-points, where $f \in (0,1]$ and *n* is the number of data-points. The parameter *h* is found using the *k*-nearest neighbors algorithm, where $k = \lfloor fn \rfloor$. Equation (18) has that the forward kinematic map is computed within this neighborhood with points close to the center **w** contributing more to the forward map estimator than points at the edge of the neighborhood. The expression for the estimator of the forward kinematic map in the neighborhood of **w** is the following:

$$\hat{\mathbf{G}}(\mathbf{w}) = (\mathbf{W}^{\mathsf{T}} \mathbf{C}(\mathbf{w}) \mathbf{W})^{-1} \mathbf{W}^{\mathsf{T}} \mathbf{C}(\mathbf{w}) \mathbf{Z}.$$
(20)

This estimator is a modified version of the standard linear regression estimator $(\mathbf{W}^{\mathsf{T}}\mathbf{W})^{-1}\mathbf{W}^{\mathsf{T}}\mathbf{Z}$ in which data-points close to **w** are assigned greater weight by the matrix **C**. This results in a forward kinematic map whose form differs depending on where it is evaluated in the space of factor scores.

B. Cross-validation of the direct and differential kinematics

We evaluated the direct and differential kinematics with the median error and the 10th–90th percentile error range. Error of the direct and differential kinematics is an important parameter, as the forward kinematics underlies the technical performance of the articulator synergy biomarker. The median error and the 10th–90th percentile error range were computed using tenfold cross-validation. In each fold, the cross-validation assigned 90% of the data-point indices to the training set S and 10% to the test set T. No two folds had overlapping test sets.

The median_{G,k} for the forward kinematic map **G** at the *k*th phonetic place of articulation reflects deviation in the estimated constriction task variables $\hat{\mathbf{z}}_{\cdot,k} = \hat{\mathbf{G}}(\mathbf{w}_{j,\cdot}) \begin{bmatrix} 1 & \mathbf{w}_{j,\cdot}^{\mathsf{T}} \end{bmatrix}^{\mathsf{T}}$ from the observed constriction task variables $\mathbf{z}_{\cdot,k}$,

$$\text{median}_{\mathbf{G},k} = \text{median}(\sqrt{(z_{j,k} - \hat{z}_{j,k})^2}), \quad \text{for} j \in \mathcal{T}.$$
(21)

The median_{J,k} for the Jacobian matrix **J** of the forward kinematic map at the *k*th phonetic place of articulation reflects deviation in the estimated finite differences in constriction task variables $\Delta \hat{\mathbf{z}}_{\cdot,k} = \hat{\mathbf{J}}(\mathbf{w}_{j,\cdot})\Delta \mathbf{w}_{j,\cdot}$ from the observed finite differences in constriction task variables $\Delta \mathbf{z}_{\cdot,k}$,

median_{J,k} = median
$$(\sqrt{(\Delta z_{j,k} - \Delta \hat{z}_{j,k})^2}), \text{ for } j \in \mathcal{T},$$
(22)

where the finite difference $\Delta z_{j,k}$ is obtained by the central difference formula,

$$\Delta z_{j,k} = (z_{j+1,k} - z_{j-1,k})/2.$$
(23)

The evaluation was performed for scale parameter f in the range of 0.2–0.9 (i.e., for neighborhoods containing 20%–90% of training data-points). For a given f-value and phonetic place of articulation, the tenfold cross-validation produced ten median_{G,k} values and ten median_{J,k} values. The reported median_{G,k} and median_{J,k} values are the medians of these ten

values. The results reported in this section were obtained with one jaw factor, four tongue factors, and two lip factors.

The median error of the forward kinematic map was smaller than the 2.4 mm in-plane spatial resolution of the realtime MRI pulse sequence when 20%–90% training datapoints were in the neighborhood (i.e., for all $f \in [0.2,0.9]$; see Fig. 5). The median error was smaller than the standard deviation of the observed constriction task variables for all participants and for all neighborhood sizes.

The median error of the Jacobian matrix was smaller than the 2.4 mm in-plane spatial resolution of the real-time MRI pulse sequence when 20%–90% training data-points were in the neighborhood (i.e., for all $f \in [0.2,0.9]$; see Fig. 5). For many participants, the median error approached the standard deviation of the frame-to-frame finite differences in constriction task variable, especially for the velar and pharyngeal places of articulation. The reason for this is that error for the Jacobian matrix was small, but the frame-to-frame differences in constriction task variables varied over a small range to begin with.

In sum, the error of the forward kinematic map is small enough to reliably quantify speech behavior for the purpose of the present study. Whether the error of the Jacobian matrix is small enough to reliably quantify speech behavior was quantified through the bias and precision of the articulator synergy biomarker, a topic taken up in Sec. VI.

VI. ARTICULATOR SYNERGY BIOMARKERS

A. Biomarker definition

In the neighborhood of $\mathbf{w}(t)$, the Jacobian $\mathbf{J}(\mathbf{w}(t))$ of the forward kinematic map provides the following relation



FIG. 5. (a) Median error (solid line) and 10th–90th percentile error range (shaded) of the forward kinematic map estimator of constriction task variables. (b) Median error (line) and 10th–90th percentile error range (shaded) of the Jacobian matrix estimator of frame-to-frame finite differences in constriction task variables. Data-points are the errors computed over all ten folds of cross-validation. Neighborhood size is given as percentage of training data-points. The standard deviation of observed (frame-to-frame finite differences in) constriction task variables is indicated as a dashed line whenever the standard deviation is small enough to fit within the *y*-axis limits.

between change $\dot{\mathbf{z}}(t)$ in constriction task variables and change $\dot{\mathbf{w}}(t)$ in articulator shape and position:

$$\int_0^T \dot{\mathbf{z}}(t) \, \mathrm{d}t = \int_0^T \mathbf{J}(\mathbf{w}(t)) \dot{\mathbf{w}}(t) \, \mathrm{d}t = \sum_{k=1}^q \int_0^T \mathbf{J}(\mathbf{w}(t)) \mathbf{P}_k \dot{\mathbf{w}}(t) \, \mathrm{d}t.$$
(24)

Time 0 is the temporal onset of a constriction, time *T* is the temporal offset of the subsequent release (see Fig. 6), *q* is the number of factors, and the $q \times q$ diagonal projection matrix \mathbf{P}_k has the (k,k)-entry equal to unity and all other entries equal to zero, breaking the integral down into the contributions of each factor score. Term *k* of the outer summation is the theoretical contribution of factor $\mathbf{f}_{.k}$ to elapsed change in constriction task variables \mathbf{z} over the time-course of a constriction.

Since real-time MRI provides a discretized sequence of images, the constriction task variables z and factor scores w are discrete-time signals. The discrete-time version of Eq. (24) is the following:

$$\sum_{j=0}^{N} \Delta \mathbf{z}_{j,\cdot} = \sum_{j=0}^{N} \mathbf{J}(\mathbf{w}_{j,\cdot}) \Delta \mathbf{w}_{j,\cdot} = \sum_{k=1}^{q} \sum_{j=0}^{N} \mathbf{J}(\mathbf{w}_{j,\cdot}) \mathbf{P}_{k} \Delta \mathbf{w}_{j,\cdot}.$$
(25)

Sample 0 is the temporal onset of a constriction, and sample *N* is the temporal offset of the subsequent release. As in the continuous-time Eq. (24), term *k* of the outer summation is the contribution of factor $\mathbf{f}_{.k}$ to elapsed change in constriction task variables \mathbf{z} over the time-course of a constriction.

The discrete-time signal $\lambda_{\mathcal{U},\ell}[n]$ is the cumulative sum of contributions of the articulator whose factor indices are in the set \mathcal{U} to narrowing or widening the ℓ th constriction task variable z_{ℓ} ,



FIG. 6. (Color online) Quantitative readout of the contributions of the jaw (dark bar) and tongue (light bar) to a palatal approximation for [i] during the transition from [a] to [i] and back to [a] in the sequence [dia]. Time runs left to right. Vertical length of the bars indicates the total elapsed change in constriction task variable at the palatal place. The breakdown into dark and light parts indicates the contribution of the jaw and tongue to the constriction, respectively. From the onset of movement to the time of maximum constriction, the jaw and tongue produce a narrowing at the palatal place, and constriction task variable decreases to a minimum. After achieving maximum constriction, the jaw and tongue widen the constriction task variable at the palatal place, and constriction task variable increases. See Sec. II A for operational definitions of movement onset and movement offset.

The set \mathcal{U} contains the indices of factors corresponding to the target articulator. For example, when the numbers of factors are $q_{jaw} = 1$, $q_{tongue} = 4$, $q_{lip} = 2$, and $q_{velum} = 1$, then $\mathcal{U} = \{1\}$ is the jaw; $\mathcal{U} = \{2, 3, 4, 5\}$ is the tongue; $\mathcal{U} = \{6, 7\}$ is the lips; and $\mathcal{U} = \{8\}$ is the velum. If N + 1 is the number of real-time magnetic resonance (MR) images acquired during a single utterance, then the integer *n* starts at 0 (i.e., the temporal onset of a constriction) and increases to *N* (i.e., the temporal offset of the subsequent release). As *n* increases from the onset 0 of a constriction to the offset *N* of the subsequent release, the signal $\lambda_{\mathcal{U},\ell}[n]$ dips to a minimum at the time-point of greatest constriction and then rises back up during the release (cf. Fig. 6).

Let \mathcal{J} be the set of jaw factor indices and \mathcal{E} be the set of lip or tongue factor indices. Specifically, the set \mathcal{E} contains lip factor indices for the bilabial place of articulation and tongue factor indices for the coronal, palatal, velar, and pharyngeal places of articulation. We define the articulator synergy biomarker ν_{ℓ} for place of articulation ℓ as the range of $\lambda_{\mathcal{J},\ell}[n]$ divided by the range of $\lambda_{\mathcal{J},\ell}[n] + \lambda_{\mathcal{E},\ell}[n]$ over all samples $n \in \{0,1,2,\ldots,N\}$. Range is computed as the difference between the 90th percentile P_{90} and 10th percentile P_{10} . Thus, the articulator synergy biomarker ν_{ℓ} is the following percentage:

$$\nu_{\ell} = 100 \times \frac{P_{90}(\lambda_{\mathcal{J},\ell}[n]) - P_{10}(\lambda_{\mathcal{J},\ell}[n])}{P_{90}(\lambda_{\mathcal{J},\ell}[n] + \lambda_{\mathcal{E},\ell}[n]) - P_{10}(\lambda_{\mathcal{J},\ell}[n] + \lambda_{\mathcal{E},\ell}[n])}.$$
(27)

The articulator synergy biomarker ν_{ℓ} is the percent contribution of the jaw to narrowing and widening the vocal tract for a constriction. The quantity 1 - ν_{ℓ} is the percent contribution of the lips (for the bilabial place) or tongue (for the coronal, palatal, velar, and pharyngeal places) to a constriction. Through Eq. (26), these quantities are based on the kinematic relations between factor scores and constriction task variables (i.e., the Jacobian of the forward kinematic map), and how the factor scores and the constriction task variables evolve in time.

B. Bias

This section evaluates the bias of the articulator synergy biomarker. Bias is the difference between the expected value of a measurement and its true value. The true value of the articulator synergy biomarker is unknown for any given MRI scan of a particular subject. For this reason, the present study designed a computer simulation method that generated synthetic vocal tract movements for which the true value of the biomarker could be controlled. By varying the true biomarker value over the range 0%–100% and comparing to the measured biomarker value, the present study estimated the bias of the articulator synergy biomarker.

The synthetic data-set was generated through simulation using the $n \times m$ matrix **Z** of constriction task variables at *m* places of articulation in the *n* observed MR images (Sec. III), the $2p \times q$ matrix of factors **F** (Sec. IV, parameters: $q_{jaw} = 1$, $q_{tongue} = 4$, $q_{lips} = 2$), and the forward kinematic map (Sec. V, parameter f = 1.0). The synthetic data-set consisted of vocal tract contours from 40 utterances (10 repetitions each of [apa], [ata], [aia], [aka]). Each utterance started from the open vocal tract posture characteristic of an initial vowel [a]. The utterance consisted of two movements. The first movement was the oral constriction for [p], [t], [k], or [i]. The second movement was the pharyngeal constriction that returned the vocal tract to the open posture characteristic of the vowel [a] (Wood, 1979).

Following task dynamics (Saltzman and Munhall, 1989), change in the vector \mathbf{z} of constriction task variables evolves according to the following equation for time *t* in the interval [$t_0, t_0 + T$], where t_0 is the start time and *T* is the duration:

$$\ddot{\mathbf{z}} = -\mathbf{K}(t)[\mathbf{z} - \mathbf{z}_0(t)] - \mathbf{B}(t)\dot{\mathbf{z}}.$$
(28)

The vector \mathbf{z}_0 contains *m* constriction task variable targets, where *m* is the number of places of articulation. The matrices **K** and **B** are diagonal matrices of *m* stiffness and damping coefficients, respectively. The parameters \mathbf{z}_0 , **K**, and **B** are constant for time *t* in $[t_0, t_0 + T/2)$ and for time *t* in $[t_0 + T/2, t_0 + T]$, with the parameters changing at the midpoint t_0 + T/2 separating the two movements. When the constriction is for task variable z_i , then the parameters take on the following values, where ω is the natural frequency of the gesture: target $z_{0i} = 0 \text{ mm}$ for [p],[t], [k] and $z_{0i} = 2.38 \text{ mm}$ for [i],[**a**]; stiffness $k_{ii} = \omega^2$; damping $b_{ii} = 2\omega$; and k_{jj} = $b_{jj} = 0$ for $j \neq i$. The natural frequency $\omega = 10 \text{ Hz}$ and simulation duration T = 1 s were set arbitrarily, as the relative usage of the jaw, tongue, and lips does not depend on the timescale of the simulation.

Given model parameters **K**, **B**, \mathbf{z}_0 , and initial conditions $z(t_0)$, $\dot{z}(t_0)$, the solution $\mathbf{z}(t)$ to Eq. (28) is unique for time t in $[t_0, t_0 + T]$. The unique solution $\mathbf{z}(t)$ maps to many paths $\mathbf{w}(t)$ of factor scores. The forward kinematic map **G**, its Jacobian $\mathbf{J}(\mathbf{w})$, and the set of weights $v_{11}, v_{22}, \dots, v_{qq}$ on the q factors determine one particular path $\mathbf{w}(t)$ out of the many

possible paths as the solution to the following equation for time *t* in $[t_0, t_0 + T)$:

$$\begin{split} \ddot{\mathbf{w}} &= \mathbf{J}^+(\mathbf{w})(-\mathbf{B}\mathbf{J}(\mathbf{w})\dot{\mathbf{w}} - \mathbf{K}(\mathbf{G}(\mathbf{w}) - \mathbf{z}_0)) \\ &- \mathbf{J}^+(\mathbf{w})\dot{\mathbf{J}}(\mathbf{w}, \dot{\mathbf{w}})\mathbf{w}. \end{split} \tag{29}$$

This follows from the change of variables $\mathbf{z} = \mathbf{G}(\mathbf{w}), \dot{\mathbf{z}} = \mathbf{J}(\mathbf{w})\dot{\mathbf{w}}, \ddot{\mathbf{z}} = \mathbf{J}(\mathbf{w})\ddot{\mathbf{w}} + \dot{\mathbf{J}}(\mathbf{w}, \dot{\mathbf{w}})\dot{\mathbf{w}}$, and the weighted Jacobian pseudoinverse $\mathbf{J}^+ = \mathbf{V}^{-1}\mathbf{J}^{\mathsf{T}}(\mathbf{J}\mathbf{V}^{-1}\mathbf{J}^{\mathsf{T}})^{-1}$, where \mathbf{V} is the diagonal matrix of weights $v_{11}, v_{22}, \dots, v_{qq}$ for the q factors (Saltzman and Munhall, 1989).

Due to its deterministic nature, the dynamical system generates vocal tract contours with a covariance structure that does not closely resemble real data. In order to demonstrate that the factors **F** could be recovered from synthetic vocal tract contours with a covariance structure similar to that of actual segmentations of real-time MR images, we estimated factors $\hat{\mathbf{F}}$ from a synthetic data-set of vocal tract contours $\mathbf{X} = \Xi \mathbf{F}$, where Ξ is a matrix whose rows are samples from the multivariate normal distribution with mean and covariance estimated from observed factor scores.

Numerical simulation of the dynamical system generated paths $\mathbf{w}(t)$ of factor scores that solved Eq. (29) for each of the 40 utterances. True biomarker values were computed using the simulated paths $\mathbf{w}(t)$ and the Jacobian matrix **J** used for the simulation according to Eqs. (26) and (27). Measured biomarker values were extracted from the synthetic vocal tract contours $\mathbf{x}(t) = \mathbf{F}\mathbf{w}(t)$ using the same procedure as for real data (i.e., without knowledge of the parameters used to generate them).

The simulation was repeated 15 times, each with a different value of the jaw weight v_{11} in the range of $10^{-3}-10^2$. This generated a range of true biomarker values from 0% to 100%. Figure 7(a) demonstrates an inverse relation between jaw weight and biomarker value. The measured biomarker values closely matched the true biomarker values over the range from 0% to 100% [cf. Figs. 7(b) and 7(c)]. A twosided paired-sample *t*-test detected no systematic bias (p = 0.80). The 95% limits of agreement are -1.03% and



FIG. 7. (Color online) (a) Relationship between measured biomarker values and theoretical jaw weight parameter values. The jaw weight parameter controls jaw usage in the dynamical systems simulation of vocal tract movement. Large biomarker values close to 100% correspond to small jaw weight parameters (indicating great jaw usage). Small biomarker values close to 0% correspond to large jaw weight parameters (indicating little jaw usage). (b) Relationship between the true and measured biomarker values. True values were obtained from synthetic data generated in a dynamical systems simulation. (c) Bland-Altman diagram graphs the difference between true and measured biomarker values (y axis) against the average of the true and measured biomarker values differed from the true value by -1.03%-0.68%. Measurement bias (-0.17%) was not significantly different from zero.

0.68%, meaning that most errors are contained in the interval [-1.03%, 0.68%].

C. Precision

This section evaluates the precision of the articulator synergy biomarker. Precision is the agreement between replicate measurements of a vocal tract constriction by the same participant for the same constriction task (Kessler et al., 2015; Sullivan et al., 2015). The same-day test-retest repeatability experiment evaluated the repeatability of the articulator synergy biomarker under variable conditions of MRI operator variability, image analysis variability, and shortterm physiological variability (Töger et al., 2017). MRI operator variability includes subject positioning within the scanner bore and scan plane localization. Image analysis variability includes variability in the manual step of time-point annotation and the manual initialization of the segmentation algorithm. Short-term physiological variability includes same-day scan-to-scan variability in speech motor control. Precision is an important parameter as it establishes a limit on effect size and group differences that the method can resolve.

Study participants repeated the MRI experiment for a total of two MRI scans. Agreement between scan 1 and scan 2 was quantified using the intraclass correlation coefficient (ICC). The ICC is a quantitative measure of test-retest repeatability for articulator synergy biomarkers. The ICC is the ratio of inter-participant variability to total variability. The greater the inter-participant variability compared to total variability, the greater the reliability because random error is smaller relative to the variance of the articulator synergy biomarker between experiment participants. On the basis of a recent review (LeBreton and Senter, 2008), ICC values were categorized as poor (0.00-0.30), weak (0.31-0.50), moderate (0.51-0.70), strong (0.71-0.90), and very strong (0.91-1.00). The ICC was computed using a linear mixed effects model fitted with the package lme4 (Bates et al., 2015) in R (R Development Core Team, 2008). Consider the sample of n = 8 participants, each with k = 20 repeated measurements of articulator synergy (ten from xcan 1, ten from scan 2). The articulator synergy biomarker ν_{ii} for replicate measurement *j* and participant *i* was

$$\nu_{ij} = m + p_i + e_{ij},\tag{30}$$

coronal

6 tongue factors 1

0.75

0.5

0.25

8

0

where *m* was the group mean, p_i was the random intercept for participant *i*, and e_{ij} was the intra-participant error. The random effects p_i and e_{ij} were independently and identically distributed with mean 0 and the inter-participant variance σ_p^2 and intra-participant variance σ_e^2 to be estimated from the data using the restricted maximum likelihood procedure. The ICC is the proportion of variance in the articulator synergy biomarker value due to biological variation among participants, compared to the total variance of the articulator synergy biomarker,

$$ICC(\nu) = \frac{\hat{\sigma}_p^2}{\hat{\sigma}_p^2 + \hat{\sigma}_e^2}.$$
(31)

If the ICC is close to one, variance in the articulator synergy biomarker mostly reflects variability among participants. If the ICC is close to zero, then variability among participants accounts for little of the total variance in the articulator synergy biomarker.

The repeatability of the articulator synergy biomarker was evaluated for different parameterizations. Parameters include: imaging parameters of the scanner pulse sequence, image reconstruction parameters, the number of jaw, tongue, and lip factors, and the neighborhood size for the forward kinematic map estimator. We investigated how varying the number of factors and the neighborhood size affected repeatability, keeping the pulse sequence and reconstruction algorithm constant. The reason that the present study investigated parameters of the statistical analysis and did not investigate parameters of image acquisition and reconstruction was that the statistical analysis directly related to the novel proposal of the present study, namely, estimating the forward kinematic map from MRI. Image acquisition and reconstruction was only indirectly related to this aim. Nevertheless, repeatability will depend on imaging parameters, and if different imaging parameters are used for future studies on the articulator synergy biomarker, the repeatability of the biomarker should be evaluated before drawing scientific conclusions.

Figure 8 shows the repeatability of the articulator synergy biomarker for different numbers of jaw, tongue, and lip factors. Regardless of the neighborhood size used for the forward kinematic map estimator, the error of the estimator was well in the subvoxel range (cf. Sec. V B). For this reason, neighborhood size was fixed at 70% of training data-points.

1

0.75

0.5

0.25

8

0

pharyngeal

6 tongue factors

velar

6 tongue factors

1

0.75

0.5

0.25

8

0



palatal

6 tongue factors

2 3 lip factors

bilabial

0.75

0.5

0.25

0

1

0.75

O 0.5

0.25

0

1 jaw factor

iaw factors

3 jaw factors



FIG. 9. (Color online) Sample distribution of the articulator synergy biomarker for bilabial stop [p], coronal stop [t], palatal approximation [i], velar stop [k], and pharyngeal approximation [a]. The biomarker indicates the percent of a constriction that was produced by the jaw. A value of 0% indicates that lip or tongue motion produced the entire constriction, whereas a value of 100% indicates that jaw motion produced the entire constriction. Sample distribution by participant shown with a different color for each participant.

The images of participant F3 were excluded from the repeatability analysis due to poor image quality in the second scan.

Repeatability of the articulator synergy biomarkers ranged from poor to strong over the wide range of factor analysis parameterizations tested. The bilabial stop had moderate to strong repeatability (range: 0.6–0.71, median: 0.68). The coronal stop had poor to moderate repeatability (range: 0.21–0.52, median: 0.36). The palatal approximation had moderate repeatability (range: 0.54–0.65, median: 0.58). The velar stop had weak to moderate repeatability (range: 0.31–0.6, median: 0.44). The pharyngeal approximation had poor to weak repeatability (range: 0.22–0.48, median: 0.36).

The range of ICC values obtained for the coronal stop, velar stop, and pharyngeal approximation included some ICC values in the poor to weak range. The reason for this may be different for the different constriction tasks. For the velar stop, the total variance of the biomarker is small (interquartile range: 7.3%-24%). Even if the intra-participant variance is small to begin with, the intra-participant variance makes up a substantial part of the total variance (see the his-tograms in Fig. 9). In contrast, for the coronal stop and pharyngeal approximation, the total variance of the biomarker is large (coronal stop inter-quartile range: 17%-37%), suggesting that intra-participant variance is substantial (see the histograms in Fig. 9).

Although the velar stop and pharyngeal approximation may not involve a very great contribution of the jaw, the number of jaw factors nevertheless affects ICC (see panels "velar" and "pharyngeal" of Fig. 8). This is due to the fact that using more jaw factors increases the variance of the tongue that is explained by jaw factors [cf. Fig. 4(b)], and some of this variance may reflect the performance of the velar stop and pharyngeal approximation.

VII. TESTING THE TASK-SPECIFICITY OF ARTICULATOR SYNERGIES

The present study tested the task-dependence of articulator synergies by determining whether the relative contribution of the jaw, tongue, and lips differs by constriction task using a linear mixed effects model fitted with the package lme4 (Bates *et al.*, 2015) in *R* (R Development Core Team, 2008). Specifically, the present study tested the null hypotheses that there are no pairwise differences in articulator synergy biomarker values between constriction tasks.

Consider the sample of n = 800 articulator synergy biomarker values (eight participants × five constriction tasks × ten repeated measurements of the articulator synergy biomarker × two scans). Let $y_{i,j,k,\ell}$ be the biomarker value for constriction task *i*, participant *j*, and replicate measurement *k* from scan ℓ (i.e., scan 1 or scan 2). The linear mixed effects model of $y_{i,j,k,\ell}$ is

$$y_{i,j,k,\ell} = m + b_i + p_j + c_\ell + q_{i,j} + r_{j,\ell} + e_{i,j,k,\ell},$$
(32)

where *m* is the baseline mean, b_i is the fixed effect for constriction task, p_i is the random intercept for participant *i*, c_ℓ is the fixed effect for scan number, $q_{i,j}$ is the by-participant random slope for constriction task, $r_{j,\ell}$ is the by-participant random slope for scan number, and $e_{i,j,k,\ell}$ is the intra-participant error. Multiple comparisons are corrected for using Tukey's range test with the package multcomp (Hothorn *et al.*, 2008) in *R* (R Development Core Team, 2008). This section reports adjusted *p*-values. See Table II for results.

The coronal stop had 32% larger biomarker values than the bilabial stop (z = 6.8, $p = 4.3 \times 10^{-11}$), 18% larger biomarker values than the palatal approximation (z = 4, $p = 4.5 \times 10^{-4}$), 39% larger biomarker values than the velar stop (z = 8.5, $p = 1.1 \times 10^{-16}$), and 28% larger biomarker values than the pharyngeal approximation (z = 6.6, $p = 3.9 \times 10^{-10}$). In addition to having 18% smaller biomarker values than the coronal stop (see immediately

TABLE II. Results for statistical tests of the null hypothesis that the contrast is zero. Rows indicate separate tests. *p*-values corrected for multiple comparisons with Tukey's range test (adjusted *p*-values reported).

Contrast	Estimate (%)	Ζ	р
Bilabial stop-coronal stop	-32	-6.8	4.3×10^{-11}
Bilabial stop-palatal approximation	-14	-2.1	0.21
Bilabial stop-velar stop	6.4	1.2	0.7
Bilabial stop-pharyngeal approximation	-4.9	-0.86	0.89
Coronal stop-palatal approximation	18	4	$4.5 imes 10^{-4}$
Coronal stop-velar stop	39	8.5	1.1×10^{-16}
Coronal stop-pharyngeal approximation	28	6.6	3.9×10^{-10}
Palatal approximation-velar stop	20	5.2	$1.8 imes 10^{-6}$
Palatal approximation-pharyngeal	9.1	3.5	0.0042
approximation			
Pharyngeal approximation-velar stop	11	3.4	0.0049

above), the palatal approximation had 20% larger biomarker values than the velar stop (z = 5.2, $p = 1.8 \times 10^{-6}$) and 9.1% larger biomarker values than the pharyngeal approximation (z = 3.5, p = 0.0042). The velar stop had 11% smaller biomarker values than the pharyngeal approximation (z = 3.4, p = 0.0049).

We infer that the jaw contributed significantly more to the coronal stop than to the bilabial stop, palatal approximation, velar stop, and pharyngeal approximation, the jaw contributed significantly more to the palatal approximation than to the velar stop and pharyngeal approximation, and the jaw contributed significantly less to the pharyngeal approximation than to the velar stop. We reject the null hypothesis that the articulator synergy biomarker does not differ by constriction task. Synergies differ in terms of inter-articulator coordination depending on the constriction task (see Fig. 10 for summary).

The sample distribution of the articulator synergy biomarker for the velar stop had small dispersion about a distinct peak at 10% (median: 12%, inter-quartile range: 7.3%–24%; see histograms in Fig. 9). The sample distribution of the articulator synergy biomarker for the coronal stop had a distinct peak at 60% (median: 56%, inter-quartile range: 44%–64%; see histograms in Fig. 9). The distinctly peaked sample distributions of the articulator synergy biomarkers for the coronal stop and velar stop likely contributed to the statistically significant bilabial stop-velar stop, coronal stop-palatal approximation, coronal stop-velar stop, palatal approximation-velar stop, palatal approximation-pharyngeal approximation, velar stop-pharyngeal approximation, and coronal stop-pharyngeal approximation contrasts.

In order to determine whether the results reported above depended on the choice of a particular number of jaw, tongue, and lip factors or on the choice of a particular neighborhood size for the forward kinematic map estimator, we repeated the statistical analysis with different parameter values (number of jaw factors: 1,2,3; number of tongue factors:



FIG. 10. (Color online) Synergies differ in terms of inter-articulator coordination depending on the constriction task. Constriction tasks are ordered from top to bottom in terms of jaw usage. Vertical lines indicate a statistically significant contrast. Compare with numeric results in Table II.

4,6,8; number of lip factors: 2,3; neighborhood size: 20%–90% in 10% steps). The articulator synergy biomarker was greater for the coronal stop than for the bilabial stop in 96/96 cases (100%), palatal approximation in 32/96 cases (33%), velar stop in 88/96 cases (92%), and pharyngeal approximation in 83/96 cases (86%). The articulator synergy biomarker was greater for the palatal approximation than for the velar stop in 71/96 cases (74%) and pharyngeal approximation in 65/96 cases (68%). The articulator synergy biomarker was greater for the pharyngeal approximation than for the velar stop in 25/96 cases (26%). Overall, these results support the inference that the jaw contributed significantly more to the coronal stop than to the bilabial stop and velar stop, and the jaw contributed significantly more to the palatal approximation than to the velar stop. However, the coronal stop-palatal approximation, coronal stop-pharyngeal approximation, palatal approximation-pharyngeal approximation, and velar stop-pharyngeal approximation contrasts should be interpreted with caution since the effect size is smaller than for the more robust effects, and the significance of these contrasts depends on parameterization.

In sum, the present study shows that the jaw contributes least to the velar stop for [k], more to pharyngeal approximation for [a], still more to palatal approximation for [i], and most to the coronal stop for [t]. Additionally, the jaw contributes more to the coronal stop for [t] than to the bilabial stop for [p] (see Fig. 10).

VIII. INTER- AND INTRA-PARTICIPANT VARIABILITY

Section VII shows an effect of constriction task on the articulator synergy biomarker values. This section investigates inter- and intra-participant variability in this effect. Inter-participant variability is evaluated by testing the significance of by-participant random slopes for constriction task. The linear mixed effects model of Sec. VII [cf. Eq. (32)] is compared to a reduced model that does not have byparticipant random slopes for constriction task. The likelihood ratio test indicates that the by-participant random slopes for constriction task are a significant source of variance $[\chi^2(14) = 560; p = 1.7 \times 10^{-109}]$. This indicates variability by participant in the effect for constriction task. By-participant variability in the effect for constriction task is further characterized by determining whether individual participants display similar effects for constriction task (i.e., same pattern of average values) and similar variances across constriction tasks. For each participant, the Mann-Whitney U test identifies all pairs of constriction tasks that differ in average biomarker value, and the Fligner-Killeen test identifies all pairs of constriction tasks that differ in biomarker variance (see Fig. 11). The study performed (number of participants) \times [(number of constriction tasks)² - (number of constriction tasks)]/2 = 80 Mann-Whitney U tests and the same number of Fligner-Killeen tests for a total of 160 statistical tests. Statistical tests are considered significant at the Bonferroni-corrected significance level $\alpha = 0.05/160$.

The results of the Mann-Whitney tests demonstrate that the pattern of average values across constriction tasks is consistent with the effect for constriction task discovered by the



FIG. 11. (Color online) Sample distribution of the articulator synergy biomarker (y axis) by participant (panel) and constriction task (color, x axis). Kernel density estimate (shaded) and 95% confidence interval for the mean (whiskers) provided for each distribution. Brackets below each panel indicate pairs of constriction tasks that significantly differ in average value (Mann-Whitney test) or variance (Fligner-Killeen test). *p*-values corrected for multiple comparisons with the Bonferroni method. Adjusted *p*-values used to determine significance.

linear mixed effects model (coronal stop > palatal approximation > pharyngeal approximation > velar stop; and coronal stop > bilabial stop; cf. Sec. VII) with three exceptions: velar stop-pharyngeal approximation constrast of participant M1; palatal approximation-velar stop and palatal approximation-pharyngeal approximation constrasts of M4. Although the participants individually display consistent patterns of average values, the dispersion of the data is large and the sample size within each participant is small. For this reason, the Mann-Whitney U test does not reject all the null hypotheses.

The results of the Fligner-Killeen tests demonstrate that participants can display high biomarker variance for some constriction tasks but low variance at others. However, fewer Fligner-Killeen tests reject the null hypothesis than do the Mann-Whitney U tests. This indicates that fewer variances differ by constriction task than do the average values.

IX. DISCUSSION

A. Task specificity of articulator synergies

The present study shows that the jaw contributes least to the velar stop for [k], more to pharyngeal approximation for [a], still more to palatal approximation for [i], and most to the coronal stop for [t]. Additionally, the jaw contributes more to the coronal stop for [t] than to the bilabial stop for [p]. This supports the hypothesis that different articulator synergies have different patterns of interarticulator coordination. The effect of constriction type on the articulator synergy biomarker demonstrates that inter-articulator coordination differs depending on the constriction task. Jaw usage varies by place of articulation (i.e., constriction location), active articulator (i.e., end-effector), and manner of articulation (i.e., target constriction degree; Vatikiotis-Bateson and Ostry, 1995). Both of these sources of variance presumably combine to produce the effect of constriction type reported in this paper. This introduces a confound in interpreting the effect for constriction task, and thus is a limitation of the present study. Future research should focus on how place and manner of articulation interact to determine articulator synergy biomarker values.

If synergies organize the articulators on a temporary basis for achieving motor goals such as vocal tract constrictions, as proposed in theories of motor control (Saltzman and Kelso, 1987; Turvey, 1977), theories of phonological organization (Browman and Goldstein, 1989; Ohala et al., 1986), and robotic systems (Herbort *et al.*, 2010), then the pattern of inter-articulator coordination varies over time as the vocal tract deploys different synergies. The articulator synergy biomarker provides the means to characterize this timevarying pattern of inter-articulator coordination in terms of the percent contribution of each articulator to changing the constriction task variable at the synergy's place of articulation. This complements the finding that articulator synergies are task-dependent in terms of inter-articulator coupling (Lancia and Rosenbaum, 2018), inter-articulator correlation (Jackson and Singampalli, 2009), and response to mechanical perturbation of articulator positions (Kelso et al., 1984).

B. Technical performance

The proliferation of vocal tract imaging databases (Narayanan et al., 2014; Sorensen et al., 2017) and the increasingly complex computational methods for studying the morphological (Lammert et al., 2013b) and functional (Dawson et al., 2016) complexities captured therein underscore the importance of evaluating the technical performance of quantitative imaging biomarkers of speech. The articulator synergy biomarker does not have systematic bias. However, precision is weak or poor at the coronal stop, velar stop, and pharyngeal approximation. Since the intraclass correlation coefficient is the ratio of inter-participant variability to total variability, which is the sum of intra- and interparticipant variability [cf. Eq. (31)], low precision may be due to large intra-participant variance, small interparticipant variance, or some combination of the two. For the velar stop, low precision is due to small inter-participant variance (cf. Fig. 9). For the coronal stop and pharyngeal approximation, low precision is due to large intra-participant variance. Intra-participant variance is unavoidable in voluntary movement due to short-term physiological variability that arises from the way the brain regulates noise in the motor system (Harris and Wolpert, 1998; van Beers, 2009; Wu et al., 2014). Although we do not discount other technical sources of variance such as MRI operator variability and image analysis variability, here we emphasize short-term physiological variability as the fundamental obstacle to achieving high precision in biomarkers of voluntary movement.

C. Parametric estimation for task dynamics

The forward kinematics relates articulator movements to the changes in constriction task variables that these movements produce. In the task dynamics model of speech production (Saltzman and Munhall, 1989), the forward kinematics is specified by the forward kinematic map and its Jacobian matrix. The study estimated these parameters from real-time MR images of speech production and evaluated the estimator by cross-validation. Error was well below the spatial resolution of the scanner.

The inverse kinematics relates changes in constriction task variables to the articulator movements that produce them. In the task dynamics model of speech production (Saltzman and Munhall, 1989), the percent contribution of each articulator in a synergy is determined by assigning weights to the articulators. In contrast to studies that manually assigned weights to the articulators based on theoretical considerations (for example, see Simko and Cummins, 2010, for an assignment of weights based on articulator mass), the present study is the first to obtain a quantitative readout of these weights from speech production data. Analysis of synthetic data in Sec. **VIB** suggested that the articulator synergy biomarker is a monotonic function of the jaw weight parameter [Fig. 7(a)]. The function will depend on the number of articulator degrees of freedom, the coordinate system for the articulator degrees of freedom, the constriction task, and the weights of other articulators. Further work is necessary to characterize these sources of variance, but the present study suggests that the articulator synergy biomarker can be mapped to articulator weights, and thus jaw weight parameters can be estimated from real-time MR images of speech production.

D. Decomposing the tongue into multiple articulators

For coronal stop [t], palatal approximation [i], velar stop [k], and pharyngeal approximation [a] constriction tasks, the articulator synergy indicates the relative contribution of the jaw and tongue. For example, a biomarker value of 60% indicates a jaw contribution of 60%. The remaining 40% is understood to come from the tongue. The present study considers the contribution of the tongue in aggregate and does not decompose its contribution into subparts such as tongue body and tongue tip. An extension of the articulator synergy biomarker would be to consider not simply a binary distinction between jaw and tongue, but a ternary distinction among jaw, tongue body, and tongue tip or even a quaternary distinction among jaw, tongue root, tongue dorsum, and tongue tip. This section provides a preliminary indication of how this is possible within the framework presented in the present study.

The method by which Sec. IV obtained jaw factors offers a recipe for obtaining factors that are associated with the motion of a subset of data-points. First, we obtain the jaw factors [Fig. 12(b)] as in Sec. IV. The null space of the



FIG. 12. (Color online) (a) Operational definition of jaw, tongue body, and tongue tip contours. (b)–(e) Mean vocal tract contour (black) with jaw, tongue body (\times 2), and tongue tip factors overlaid (red contour: +2 standard deviation; blue: -2 standard deviation).

transposed jaw factors captures the part of tongue and lip motion that is independent of jaw motion.

Second, we project the data matrix on the null space of the transposed jaw factors as in Sec. IV. This is the contour motion that is independent of the jaw. Rather than subject the whole tongue contour to principal component analysis as in Sec. IV, we perform principal component analysis only on the tongue body contour vertices. The tongue body factors [Figs. 12(c) and 12(d)] are the vectors of covariances between the z-scored tongue body principal component scores and the tongue body and tongue tip contour vertices. The null space of the transposed set of jaw and tongue body factors captures the part of tongue tip motion that is independent of jaw and tongue body motion.

Third, we project the data matrix on the null space of the transposed jaw and tongue body factors. This is the contour motion that is independent of the jaw and tongue body. We perform principal component analysis on the tongue tip contour vertices. The tongue tip factors [Fig. 12(e)] are the vectors of covariances between the *z*-scored tongue tip principal component scores and the tongue tip contour vertices. The null space of the transposed set of jaw, tongue body, and tongue tip factors captures residual variance that is independent of jaw, tongue body, and tongue tip motion.

Whereas Sec. IV extracts all tongue factors from the null space of the transposed jaw factors, the procedure described above extracts tongue body factors from the null space of transposed jaw factors and then extracts tongue tip factors from the null space of the transposed jaw and tongue body factors. This procedure offers a systematic way to decompose the tongue into subparts that form a kinematic chain (Craig, 2005). Future work that pursues this approach will provide greater detail in the analysis of jaw-tongue coordination in synergies.

E. Forward kinematics and the nervous system

The nervous system has an internal model of the forward kinematics (Guenther, 2016; Shadmehr *et al.*, 2010). This model encodes the expected result of motor commands in terms of expected sensory consequences (proprioceptive and auditory consequences in the case of speech). Both the forward kinematics of the vocal tract and the nervous system's internal model of the forward kinematics are important components of a computational model of motor control (for a theoretical model of speech motor control that cleanly distinguishes these components; see Ramanarayanan *et al.*, 2016; cf. "forward kinematics" and "model of forward kinematics" blocks; see also Houde and Nagarajan, 2011; Todorov and Jordan, 2002). Although the present study estimates the forward kinematics, it does not characterize the nervous system's internal model of the forward kinematics. While both relate the articulator degrees of freedom to task variables, the coordinate system that represents the articulator degrees of freedom in the nervous system may differ from the coordinate system of the present study. That is, the nervous system may represent the articulator degrees of freedom differently than with jaw, tongue, and lip factors. If the coordinate system for the nervous control of vocal tract movement were known, insight into motor variability, motor equivalence, and redundancy could be gleaned from analysis of the forward kinematic map using the uncontrolled manifold approach (Scholz and Schöner, 1999). However, the results of such studies are inconclusive without knowledge of the coordinate system used in the nervous system (Sternad et al., 2010). Knowledge of the coordinate system could potentially be obtained through detailed modeling of the innervation of head and neck muscles. Indeed, physiological knowledge guides the choice of coordinate system for analysis of the uncontrolled manifold in human upper limb movement (cf. argument of Scholz and Schöner, 2014, Sec. 7.1, point 3). Biomechanical models offer tools for investigating the coordinate system available to the nervous system for producing vocal tract movements (Lloyd et al., 2012). Szabados and Perrier (2016) used a biomechanical model to investigate motor equivalence using the uncontrolled manifold approach. Future work should investigate whether MRI can be used to model the nervous system's internal model of the forward kinematics of the vocal tract.

X. CONCLUSIONS

The present study shows that the jaw contributes least to the velar stop for [k], more to pharyngeal approximation for [α], still more to palatal approximation for [i], and most to the coronal stop for [t]. Additionally, the jaw contributes more to the coronal stop for [t] than to the bilabial stop for [p]. This supports the idea that synergies organize the articulators on a temporary basis for achieving motor goals such as vocal tract constrictions, and the pattern of interarticulator coordination varies over time as the vocal tract deploys different synergies.

The following are four potential threads of research that build on the results of the present study. First, the present study estimated parameters of vocal tract *kinematics*, not parameters of vocal tract *dynamics*. That is, the parameters have to do with motion, not with the forces that produce the motion. In the task dynamics model of speech production, dynamical parameters include gestural parameters (e.g., stiffness, damping, mass) and parameters of inter-gestural coordination (e.g., coupling, blending). Future studies may attempt to estimate these parameters from MRI. Second, the present study showed that the articulator synergy biomarker varied by constriction task. Future studies may evaluate whether the articulator synergy biomarker depends on linguistic context (e.g., phonetic, lexical, syntactic), speech conditions (e.g., speech in noise vs clean speech, variable speech rate), sociolinguistic factors, sex, anatomy, and age, especially during childhood development. Third, while the present study focused on jaw-tongue and jaw-lip synergies, the proposed method could be used to study other synergies such as those between the dorsum and tip of the tongue or between the tongue and velum. Future research studies should investigate these synergies. Fourth, a forward kinematic map from articulators to acoustic formant frequencies is used by the Directions into Velocities of Articulators (DIVA) model of speech production (Guenther, 1995). By using real-time MRI with simultaneously recorded speech audio, the proposed framework could be extended to estimate this forward kinematic map from data (McGowan and Berger, 2009).

XI. REPRODUCIBILITY AND REPLICATION

The scripts required to reproduce the present study are available online.¹ The MRI data-set is available online for free use by the research community (see Töger *et al.*, 2017).² A replication of the present study was performed using the USC Speech and Vocal Tract Morphology MRI Database (Sorensen *et al.*, 2017), and is available as supplementary material.³

ACKNOWLEDGMENTS

The authors acknowledge funding through National Institutes of Health (NIH) Grant Nos. R01DC007124 and T32DC009975, and National Science Foundation (NSF) Grant No. 1514544. The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or NSF.

- ¹https://github.com/TannerSorensen/task_spec_synergies (Last viewed March 7, 2019).
- ²http://sail.usc.edu/span/test-retest/ (Last viewed March 7, 2019).
- ³See supplementary material at https://doi.org/10.1121/1.5093538 for the results of the replication study.
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). "Fitting linear mixed-effects models using lme4," J. Stat. Software 67(1), 1–48.
- Bresch, E., and Narayanan, S. (2009). "Region segmentation in the frequency domain applied to upper airway real-time magnetic resonance images," IEEE Trans. Med. Imaging 28(3), 323–338.
- Browman, C. P., and Goldstein, L. (**1989**). "Articulatory gestures as phonological units," Phonology **6**(2), 201–251.
- Craig, J. J. (2005). *Introduction to Robotics: Mechanics and Control*, 3rd ed. (Pearson/Prentice Hall, Upper Saddle River, NJ).
- Dawson, K. M., Tiede, M. K., and Whalen, D. H. (2016). "Methods for quantifying tongue shape and complexity using ultrasound imaging," Clin. Ling. Phonetics 30(3–5), 328–344.
- Guenther, F. H. (1995). "Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production," Psychol. Rev. 102(3), 594–621.
- Guenther, F. H. (2016). Neural Control of Speech, 1st ed. (MIT Press, Cambridge, MA).
- Harris, C. M., and Wolpert, D. M. (1998). "Signal-dependent noise determines motor planning," Nature 394, 780–784.

- Herbort, O., Butz, M. V., and Pedersen, G. (2010). "The SURE_REACH model for motor learning and control of a redundant arm: From modeling human behavior to applications in robotics," in *From Motor Learning to Interaction Learning in Robots*, edited by O. Sigaud and J. Peters, in Studies in Computational Intelligence, 1st ed. (Springer, Berlin), Vol. 264, Chap. 5, pp. 85–106.
- Hothorn, T., Bretz, F., and Westfall, P. (2008). "Simultaneous inference in general parametric models," Biom. J. 50(3), 346–363.
- Houde, J. F., and Nagarajan, S. S. (2011). "Speech production as state feedback control," Front. Hum. Neurosci. 5, 1–14.
- Jackson, P. J., and Singampalli, V. D. (2009). "Statistical identification of articulation constraints in the production of speech," Speech Commun. 51(8), 695–710.
- Kelso, J. S., Tuller, B., Vatikiotis-Bateson, E., and Fowler, C. A. (1984). "Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures," J. Exp. Psychol. Hum. Percept. Perform. 10(6), 812–832.
- Kessler, L. G., Barnhart, H. X., Buckler, A. J., Choudhury, K. R., Kondratovich, M. V., Toledano, A., Guimaraes, A. R., Filice, R., Zhang, Z., Sullivan, D. C., and QIBA Terminology Working Group (2015). "The emerging science of quantitative imaging biomarkers: Terminology and definitions for scientific studies and regulatory submissions," Stat. Methods Med. Res. 24(1), 9–26.
- Lammert, A., Goldstein, L., Narayanan, S., and Iskarous, K. (2013a). "Statistical methods for estimation of direct and differential kinematics of the vocal tract," Speech Commun. 55(1), 147–161.
- Lammert, A., Proctor, M., and Narayanan, S. S. (2013b). "Morphological variation in the adult hard palate and posterior pharyngeal wall," J. Speech Lang. Hear. Res. 56(2), 521–530.
- Lancia, L., and Rosenbaum, B. (2018). "Coupling relations underlying the production of speech articulator movements and their invariance to speech rate," Biol. Cybern. 112, 253–276.
- Latash, M. L. (2008). Synergy (Oxford University Press, Oxford, UK).
- LeBreton, J. M., and Senter, J. L. (2008). "Answers to 20 questions about interrater reliability and interrater agreement," Organ. Res. Methods 11(4), 815–852.
- Lingala, S. G., Zhu, Y., Kim, Y.-C., Toutios, A., Narayanan, S., and Nayak, K. S. (2017). "A fast and flexible MRI system for the study of dynamic vocal tract shaping," Magn. Reson. Med. 77(1), 112–125.
- Lloyd, J. E., Stavness, I., and Fels, S. (2012). "Artisynth: A fast interactive biomechanical modeling toolkit combining multibody and finite element simulation," in *Soft Tissue Biomechanical Modeling for Computer Assisted Surgery* (Springer, Berlin), pp. 355–394.
- Maeda, S. (1990). "Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model," in *Speech Production and Speech Modelling* (Springer, Berlin), pp. 131–149.
- McGowan, R. S., and Berger, M. A. (2009). "Acoustic-articulatory mapping in vowels by locally weighted regression," J. Acoust. Soc. Am. 126(4), 2011–2032.
- Narayanan, S., Toutios, A., Ramanarayanan, V., Lammert, A., Kim, J., Lee, S., Nayak, K., Kim, Y.-C., Zhu, Y., Goldstein, L., Byrd, D., Bresch, E., Ghosh, P., Katsamanis, A., and Proctor, M. (2014). "Real-time magnetic resonance imaging and electromagnetic articulography database for speech production research (TC)," J. Acoust. Soc. Am. 136(3), 1307–1311.
- Ohala, J. J., Browman, C. P., and Goldstein, L. M. (1986). "Towards an articulatory phonology," Phonology 3, 219–252.
- Ramanarayanan, V., Parrell, B., Goldstein, L., Nagarajan, S., and Houde, J. (2016). "A new model of speech motor control based on task dynamics and state feedback," in *Interspeech 2016*, pp. 3564–3568.
- R Development Core Team (2008). "R: A language and environment for statistical computing," R Foundation for Statistical Computing, Vienna, Austria, available at www.R-project.org (Last viewed March 7, 2019).
- Saltzman, E., and Kelso, J. A. S. (1987). "Skilled actions: A task-dynamic approach," Psychol. Rev. 94(1), 84–106.
- Saltzman, E. L., and Munhall, K. G. (1989). "A dynamical approach to gestural patterning in speech production," Ecol. Psychol. 1(4), 333–382.
- Scholz, J. P., and Schöner, G. (1999). "The uncontrolled manifold concept: Identifying control variables for a functional task," Exp. Brain Res. 126(3), 289–306.
- Scholz, J. P., and Schöner, G. (2014). "Use of the uncontrolled manifold (UCM) approach to understand motor variability, motor equivalence, and self-motion," in *Progress in Motor Control*, edited by M. Levin, Advances in Experimental Medicine and Biology (Springer Verlag, Berlin, Heidelberg), Vol. 826, pp. 91–100.

- Scott, A. D., Wylezinska, M., Birch, M. J., and Miquel, M. E. (2014). "Speech MRI: Morphology and function," Physica Medica: Eur. J. Med. Phys. 30(6), 604–618.
- Shadmehr, R., Smith, M. A., and Krakauer, J. W. (2010). "Error correction, sensory prediction, and adaptation in motor control," Annu. Rev. Neurosci. 33, 89–108.
- Simko, J., and Cummins, F. (2010). "Embodied task dynamics," Psychol. Rev. 117(4), 1229–1246.
- Sorensen, T., Skordilis, Z., Toutios, A., Kim, Y.-C., Zhu, Y., Kim, J., Lammert, A., Ramanarayanan, V., Goldstein, L., Byrd, D., Nayak, K., and Narayanan, S. (2017). "Database of volumetric and real-time vocal tract MRI for speech science," in *Interspeech 2017*, pp. 645–649.
- Sorensen, T., Toutios, A., Goldstein, L., and Narayanan, S. S. (2016). "Characterizing vocal tract dynamics across speakers using real-time MRI," in *Interspeech 2016*, pp. 465–469.
- Sternad, D., Park, S.-W., Müller, H., and Hogan, N. (2010). "Coordinate dependence of variability analysis," PLoS Comput. Biol. 6(4), e1000751.
- Sullivan, D. C., Obuchowski, N. A., Kessler, L. G., Raunig, D. L., Gatsonis, C., Huang, E. P., Kondratovich, M., McShane, L. M., Reeves, A. P., Barboriak, D. P., Guimaraes, A. R., and Wahl, R. L. (2015). "Metrology standards for quantitative imaging biomarkers," Radiology 277(3), 813–825.
- Szabados, A., and Perrier, P. (2016). "Uncontrolled manifolds in vowel production: Assessment with a biomechanical model of the tongue," in *Interspeech 2016*, pp. 3579–3583.
- Todorov, E., and Jordan, M. I. (2002). "Optimal feedback control as a theory of motor coordination," Nat. Neurosci. 5(11), 1226–1235.
- Töger, J., Sorensen, T., Somandepalli, K., Toutios, A., Lingala, S. G., Narayanan, S., and Nayak, K. (2017). "Test–retest repeatability of human

speech biomarkers from static and real-time dynamic magnetic resonance imaging," J. Acoust. Soc. Am. **141**(5), 3323–3336.

- Toutios, A., and Narayanan, S. S. (2015). "Factor analysis of vocal tract outlines derived from real-time magnetic resonance imaging data," in *International Congress of Phonetic Sciences (ICPhS)*, available at www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/ Papers/ICPHS0514.pdf (Last viewed March 7, 2019).
- Toutios, A., and Narayanan, S. S. (2016). "Advances in real-time magnetic resonance imaging of the vocal tract for speech science and technology research," APSIPA Trans. Signal Inf. Process. 5, e6.
- Turvey, M. T. (1977). "Preliminaries to a theory of action with reference to vision," in *Perceiving, Acting and Knowing: Towards an Ecological Psychology*, edited by R. Shaw and J. Bransford (Routledge Taylor and Francis Group, London), pp. 211–266, available at www.haskins.yale.edu/ sr/sr041/SR041_01.pdf (Last viewed March 7, 2019).
- Uecker, M., Ong, F., Tamir, J. I., Bahri, D., Virtue, P., Cheng, J. Y., Zhang, T., and Lustig, M. (2015). "Berkeley advanced reconstruction toolbox," in *Proceedings of the 23rd Annual Meeting of the International Society for Magnetic Resonance in Medicine*, International Society for Magnetic Resonance in Medicine.
- van Beers, R. J. (2009). "Motor learning is optimally tuned to the properties of motor noise," Neuron 63(3), 406–417.
- Vatikiotis-Bateson, E., and Ostry, D. J. (1995). "An analysis of the dimensionality of jaw motion in speech," J. Phonetics 23(1–2), 101–117.
- Wood, S. (**1979**). "A radiographic analysis of constriction location for vowels," J. Phonetics **7**, 25–43.
- Wu, H. G., Miyamoto, Y. R., Castro, L. N. G., Ölveczky, B. P., and Smith, M. A. (2014). "Temporal structure of motor variability is dynamically regulated and predicts motor learning ability," Nat. Neurosci. 17(2), 312–321.