A Sparse Non-negative Matrix Factorization Framework for Identifying Functional Units of Tongue Behavior from MRI

Journal:	*Transactions on Medical Imaging*
Manuscript ID	Draft
Manuscript Type:	Full Paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Woo, Jonghye; Massachusetts General Hospital, Radiology; MGH Prince, Jerry; Johns Hopkins University, Electrical and Computer Engineering Stone, Maureen; University of Maryland Dental School, Neural and Pain Sciences, and Orthodontics Xing, Fangxu; Massachusetts General Hospital, Radiology Gomez, Arnold; Johns Hopkins University, Electrical and Computer Engineering Green, Jordan; MGH Institute of Health Professions, Communication Sciences and Disorders Hartnick, Christopher; Massachusetts Eye and Ear Infirmary, Harvard Medical School Brady, Thomas; Harvard Medical School, Radiology Reese, Timothy; Massachusetts General Hospital, Martinos Center for Biomedical Imaging Wedeen, Van; Massachusetts General Hospital and the Harvard Medical School, Athinoula A. Martinos Center for Biomedical Imaging El Fakhri, Georges; Massachusetts General Hospital and the Harvard Medical School, Radiology
Keywords:	Tracking (time series analysis) < General methodology, Magnetic resonance imaging (MRI) < Imaging modalities, Muscle < Object of interest, Machine learning < General methodology
Specialty/Area of Expertise:	

A Sparse Non-negative Matrix Factorization Framework for Identifying Functional Units of Tongue Behavior from MRI

Jonghye Woo, *Member, IEEE*, Jerry L. Prince, *Fellow, IEEE*, Maureen Stone, Fangxu Xing, Arnold Gomez, Jordan R. Green, Christopher J. Hartnick, Thomas J. Brady, Timothy G. Reese, Van J. Wedeen, Georges El Fakhri, *Fellow, IEEE*

Abstract*—Tongue motions in the course of speech or other lingual behaviors are synergies created by locally deforming regions, or functional units. Functional units are functional groupings of local structural elements within the tongue that compress, expand, and move in a cohesive and consistent manner. Identifying the functional units using tagged-Magnetic Resonance Imaging (MRI) provides an insight into the mechanisms of normal and pathological muscle coordination, potentially leading to improvement in surgical planning, treatment, or rehabilitation procedures. In this work, to mine this information, we propose a graph-regularized sparse non-negative matrix factorization and probabilistic graphical model framework by learning latent building blocks and the corresponding weighting map using a set of motion features from displacements extracted from tagged-MRI. Our tagged-MRI imaging and internal tissue motion tracking paradigm provide previously unavailable internal tongue motion patterns, thus illuminating the inner workings of the tongue during speech or other lingual behaviors. Spectral clustering using the weighting map is then performed to determine the coherent regions defined by the tongue motion that may involve multiple or undocumented regions. Two-dimensional image data are used to verify that the proposed approach clusters the different types of images accurately. Three-dimensional synthetic and *in vivo* tongue data carrying out simple non-speech/speech tasks are used to define subject/task-specic functional units of the tongue in localized regions.

***Index Terms*—Tongue Motion, Functional Units, Speech, Non-negative Matrix Factorization, MRI, Sparsity**

I. INTRODUCTION

Finding a suitable representation of high-dimensional and complex data for a variety of tasks, such as clustering and topic modeling, is a fundamental challenge in many areas such as computer vision, machine learning, data mining, and medical image analysis. Non-negative matrix factorization (NMF) [1], an unsupervised generative model, is a class of techniques to find a low-dimensional representation of a dataset suitable for a clustering interpretation [2]; it also has been used to transform seemingly disparate features to a common domain. NMF and its variants involving sparsity have received substantial attention since the seminal work by Lee and Seung [1] because of its ability to provide an interpretable and parts-based representation inspired by psychological and physiological observations about the human brain [23]. Specifically, NMF with a sparsity constraint focuses on data matrices whose elements are non-negative, allowing to model a data matrix as sparse linear combinations of basis vectors. In this work, we are interested in modeling the tongue's underlying behaviors using NMF, since non-negative properties of NMF are akin to the physiology of the tongue as reflected in the matrix decomposition process. That is, NMF does not allow negative combinations of basis vectors. This is consistent with the analysis of muscles, which either have positive activation or no activation, not negative activation.

The human tongue is a structurally and functionally complex muscular structure, comprising orthogonally oriented and inter-digitated muscles. The tongue is innervated by more than 13,000 hypoglossal motoneurons [3], [4]. The complexity and precision of the voluntary and involuntary movements of the tongue during the course of speaking, swallowing, and breathing are invoked by a complex set of neural excitations of tongue muscles. The tongue muscles interact with one another to carry out the oromotor behaviors, which are executed by deforming local functional units in this complex muscular array. Tongue motions are synergies created by locally deforming regions, or *functional units* [6], [7]. Functional units are regions of the tongue that exhibit homogeneous motion during the execution of the specific task identifying functional units and understanding the mechanisms of coupling among them can identify motor control strategy in both normal and adapted speech (e.g., tongue motion after tongue cancer surgery or brain injury such as Amyotrophic Lateral Sclerosis (ALS) or stroke). However, to date, the mechanisms of the muscle coordination and the relationship between tongue structure and function have remained poorly understood partly due to the greater complexity and variability of both muscular structures and their interactions.

Understanding the subject/task-specific functional organization of the tongue requires a map of the functional units of the human tongue for specific tasks using medical imaging. In particular, recent advances in medical imaging and associated image analysis techniques permit the non-invasive imaging of structural and functional components of the tongue. Magnetic

---

Part of this work was presented at the MICCAI 2014 and Acoustical Society of America in 2017. Jonghye Woo, Fangxu Xing, Thomas J. Brady, Timothy R. Reese, Van J. Wedeen, and Georges El Fakhri are with Department of Radiology, Massachusetts General Hospital and Harvard Medical School. Maureen Stone is with the University of Maryland Dental School. Jordan R. Green is with MGH Institute of Health Professions. Christopher J. Hartnick is with Massachusetts Eye and Ear Infirmary and Harvard Medical School. Arnold Gomez and Jerry L. Prince are with Department of Electrical and Computer Engineering at Johns Hopkins University.
E-mail: jwoo@mgh.harvard.edu

resonance imaging (MRI) technologies have shown that a large number of tongue muscles undergo highly complex deformations during speech and other lingual behaviors. For example, the ability to perform non-invasive imaging using MRI has allowed to image both tongue surface motion using cine-MRI [8], [9], [10] and internal tissue motion using tagged-MRI (tMRI) [5]. In addition, high-resolution MRI and diffusion MRI [25], [24] have provided the muscular and fiber anatomy, respectively. With advances in computational anatomy, a vocal tract atlas [11], [12], a representation of the tongue anatomy, has also been created and validated, which allows for investigating the relationship between tongue structure and function by providing a reference anatomical configuration to analyze similarities and variability of tongue motion.

In this work, we develop a computational approach to defining the subject-specific and data-driven functional units from tMRI and 4D (3D space with time) voxel-level tracking [13] by extending our previous approach [17]. We describe a refined algorithm including an advanced tracking algorithm and graph-regularized sparse NMF to determine spatio-temoprally varying functional units using simple non-speech/speech tasks and provide extensive validations on both synthetic and *in vivo* tongue data. The method integrates a regularization term that encourages the computation of distances on a manifold rather than the whole of Euclidean space in order to preserve the intrinsic geometry of the observed motion data. We assume a manifold of the data within an NMF approach, thereby capturing the intrinsic geometry and finding a low dimensional subspace of the motion features derived from tMRI. Since standard NMF assumes a standard Euclidean distance measure for its data, it fails to discover the intrinsic geometry of its data [23]. Thus, we improve the standard NMF by introducing a joint regularization scheme to determine the cohesive motion pattern of the tongue. Both quantitative and qualitative evaluation results demonstrate the validity of the proposed method and its superiority to conventional clustering algorithms.

We present a fully automated approach to discovering the functional units using a graph-regularized sparse matrix factorization and probabilistic graphical model framework with the following main contributions:

- The most prominent contribution of this work is to use voxel level data-driven MRI (1.875 mm×1.875 mm×1.875 mm) methods incorporating internal tissue points to obtain subject-specific functional units of how tongue muscles coordinate to generate target observed motion.
- This work applies a graph-regularized sparse NMF and probabilistic graphical model to the voxel level motion data, allowing us to estimate the latent functional coherence, by learning simultaneously latent building blocks and the corresponding weighting map from a set of motion features.
- Our tMRI imaging and internal tissue motion tracking paradigm bring into light patterns of motion that have so far been intractable. The proposed approach is scalable to a variety of motion features derived from motion trajectories to characterize the coherent motion patterns. In this work, we consider the most representative features such as displacements and angle from tMRI.

The structure of this paper is as follows. Related work is reviewed in Section II. Section III shows the proposed approach to determining the functional units. Section IV presents the experimental results. Section V provides a discussion, and finally, Section VI concludes this paper.

II. RELATED WORK

In this section, we review recent work on NMF for clustering and functional units research that are closely related to our work.

NMF for Clustering. A multitude of NMF-based methods for unsupervised data clustering have been proposed over the last decade across various domains, ranging from video analysis to medical image analysis. In particular, the idea of using L_1 norm regularization (i.e., sparse NMF) for the purpose of clustering has been successfully employed [28]. The sparsity condition imposed on the weighting matrix (or coefficient matrix) indicates the clustering membership. For example, Shahnaz et al. [29] proposed an algorithm for document clustering. The matrix factorization was used to compute a low-rank approximation of a sparse matrix along with preservation of natural data property. Wang et al. [31] applied the NMF framework to gene-expression data to identify different cancer classes. Anderson et al. [32] presented an NMF-based clustering method to differential changes in default mode subnetworks in ADHD from multimodal data. In addition, Mo et al. [33] proposed a motion segmentation method using an NMF-based method. A key insight to use NMF for clustering purpose is that NMF is able to learn and discriminate localized traits of data with a better interpretation. Please refer to [19] for a detailed review on NMF for clustering purpose.

Speech Production. Research on determining functional units during speech has a long history (see e.g., Gick and Stavness [14] for a recent review). Determining functional units is considered as discovering a “missing link” between speech movements primitives and cortical regions associated with speech production [14]. In the context of lingual coarticulation, functional untis can be seen as “quasi-independent” motions [3]. A variety of different techniques have been used to address this problem. Stone et al. [15] showed that four midsagittal regions including anterior, dorsal, middle, and posterior functioned quasi-independently using ultrasound and microbeam data. Green and Wang attempted to characterize functionally independent articulators from microbeam data using covariance-based analysis [6]. More recently, Stone et al. [3] presented a method to determine functional segments using ultrasound and tMRI. That work examined compression and expansion between the anterior and posterior tongue and found that regions or muscles have functional segments strongly affected by phonemic constraints. Another key report is that of Ramanarayanan et al. [16], who used a convolutive NMF algorithm to determine tongue movement primitives from electromagnetic articulography (EMA). Our work is inspired by approaches discussed above and we use far richer

4D tMRI based tracking data and the augmented NMF framework with the addition of prior information on sparsity and intrinsic data geometry in defining functional units. Unlike other approaches that largely rely on the tracking of landmark points sparsely located on the fixed tongue surface such as the tip, blade, body, and dorsum, we aim to identify 3D cohesive functional units that involve multiple, and possibly undocumented, internal tongue regions.

III. PROPOSED FRAMEWORK

A. *Problem Statement*

Without loss of generality, let us first define the notations and definitions used in this work. Consider a set of P internal tongue tissue points tracked through tMRI, each with n scalar quantities (e.g., magnitude and angle of each track) tracked through F time frames. These quantities characterize each tissue point, which are used to group them into cohesive motion patterns, functional units. The location of the p-th tissue point at the f-th time frame can be expressed as (x_f^p, y_f^p, z_f^p). The tongue motion can then be represented by a $3F \times P$ spatio-temporal feature matrix $\mathbf{N} = [\mathbf{n}_1, ..., \mathbf{n}_P] \in \mathbb{R}^{3F \times P}$, where the p-th column is given by

$$\mathbf{n}_p = [x_1^p, \cdots, x_F^p, y_1^p, \cdots, y_F^p, z_1^p, \cdots, z_F^p]^T. \qquad (1)$$

We cast this problem of determining the functional units as a motion clustering problem as the functional units are considered to be regions of the tongue that exhibit homogeneous motion. However, different from generic motion clustering problems in computer vision, the tongue's function and physiology also need to be reflected and captured in our formulation. Thus, the goal is to determine a permutation of the columns to form $[\mathbf{N}_1 | \mathbf{N}_2 | \cdots | \mathbf{N}_c]$, where the submatrix $\mathbf{N}_i$ comprises point tracks associated with the i-th submotion—i.e., the i-th functional unit. The displacements and derived motion features for each underlying muscle are not completely independent; a subset of motion quantities from each muscle maps to one or more common latent dimensions, which can be interpreted via our model. These latent dimensions provide a sparse summary of the generative process behind the tongue's motion for a single or multiple muscles. In this work, sparse NMF is utilized to infer the latent structure of motion features derived from tMRI. The proposed method is described in more detail below; a flowchart is shown in Figure 1.

B. *MR Data Acquisition and Motion Tracking*

1) MR Data Acquisition: All MRI scanning is performed on a Siemens 3.0 T Tim Treo system (Siemens Medical Solutions, Malvern, PA) with 12-channel head and 4-channel neck coil. While subjects are speaking the same word repeatedly, the tMRI datasets are collected using Magnitude Imaged CSPAMM Reconstructed images [26]. The datasets have a 1 second duration, 26 time-frames with a temporal resolution of 36 ms for each phase with no delay from the tagging pulse, 6 mm thick slices (6 mm tag separation), and 1.875 mm in-plane resolution with no gap. The field-of-view is 24 cm.

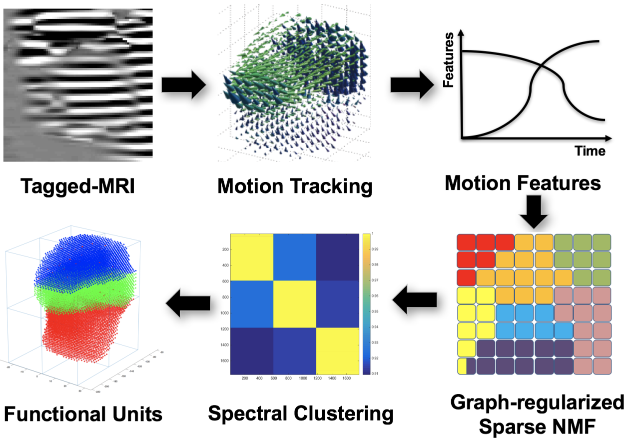

Fig. 1. Flowchart of the proposed method

2) Motion Estimation from Tagged-MRI: The phase vector incompressible registration algorithm (PVIRA) [35] is used to estimate deformation of the tongue from tMRI, yielding a sequence of dense 3D motion fields. Although the input is a set of sparsely acquired tMRI slices, PVIRA uses cubic B-spline to interpolate these 2D slices into denser 3D voxel locations. Then a harmonic phase (HARP) [20] filter is applied to produce HARP phase volumes from the interpolated result. Finally, PVIRA uses the iLogDemons method [21] on these phase volumes. Specifically, we denote the phase volumes as Φ_x, Θ_x, Φ_y, Θ_y, Φ_z, and Θ_z, where x, y, and z denote motion information from three cardinal directions usually contained in orthogonal axial, sagittal, and coronal tagged slices. The volume in the reference time frame is Φ and the volume in the deformed time frame is Θ. The motion update vector field is derived from these phase volumes. At each voxel in the image, the update vector is computed by

$$\delta\mathbf{v}(\mathbf{x}) = \frac{\mathbf{v}_0(\mathbf{x})}{\alpha_1(\mathbf{x}) + \alpha_2(\mathbf{x})/K}, \qquad (2)$$

Note that K is the normalization factor. $\mathbf{v}_0(\mathbf{x})$, $\alpha_1(\mathbf{x})$, and $\alpha_2(\mathbf{x})$ are defined by

$$\begin{aligned}
\mathbf{v}_0(\mathbf{x}) = {} & W(\Phi_x(\mathbf{x}) - \Theta_x(\mathbf{x}))(\nabla^*\Phi_x(\mathbf{x}) + \nabla^*\Theta_x(\mathbf{x})) \\
& + W(\Phi_y(\mathbf{x}) - \Theta_y(\mathbf{x}))(\nabla^*\Phi_y(\mathbf{x}) + \nabla^*\Theta_y(\mathbf{x})) \\
& + W(\Phi_z(\mathbf{x}) - \Theta_z(\mathbf{x}))(\nabla^*\Phi_z(\mathbf{x}) + \nabla^*\Theta_z(\mathbf{x})) \, , \\
\alpha_1(\mathbf{x}) = {} & ||\nabla^*\Phi_x(\mathbf{x}) + \nabla^*\Theta_x(\mathbf{x})||^2 + ||\nabla^*\Phi_y(\mathbf{x}) + \nabla^*\Theta_y(\mathbf{x})||^2 \\
& + ||\nabla^*\Phi_z(\mathbf{x}) + \nabla^*\Theta_z(\mathbf{x})||^2 \, , \\
\alpha_2(\mathbf{x}) = {} & W(\Phi_x(\mathbf{x}) - \Theta_x(\mathbf{x}))^2 + W(\Phi_y(\mathbf{x}) - \Theta_y(\mathbf{x}))^2 \\
& + W(\Phi_z(\mathbf{x}) - \Theta_z(\mathbf{x}))^2 .
\end{aligned} \qquad (3)$$

Wrapping of phase $W(\theta)$ is defined by

$$W(\theta) = \mathrm{mod}(\theta + \pi, 2\pi) - \pi \qquad (4)$$

and the "starred" gradient is defined by

$$\nabla^*\Phi(\mathbf{x}) := \begin{cases} \nabla\Phi(\mathbf{x}), & \text{if } |\nabla\Phi(\mathbf{x})| \le |\nabla W(\Phi(\mathbf{x}) + \pi)| \\ \nabla W(\Phi(\mathbf{x}) + \pi), & \text{otherwise.} \end{cases} \qquad (5)$$

After all iterations are complete, the forward and inverse deformation fields can be found by

$$\varphi(\mathbf{x}) = \exp(\mathbf{v}(\mathbf{x})) \text{ and } \varphi^{-1}(\mathbf{x}) = \exp(-\mathbf{v}(\mathbf{x})), \tag{6}$$

and they are both incompressible and diffeomorphic, making both Eulerian and Lagrangian computations available for the following continuum mechanics operations.

C. *Extraction of Motion Quantities*

The first step in our algorithm is to extract the motion features from PVIRA that characterize the cohesive motion patterns over time. We extract motion features including the magnitude and angle of the track similar to [34] described as

$$m_f^p = \sqrt{(x_{f+1}^p - x_f^p)^2 + (y_{f+1}^p - y_f^p)^2 + (z_{f+1}^p - z_f^p)^2} \tag{7}$$

$$cz_f^p = \frac{x_{f+1}^p - x_f^p}{\sqrt{(x_{f+1}^p - x_f^p)^2 + (y_{f+1}^p - y_f^p)^2}} + 1 \tag{8}$$

$$cx_f^p = \frac{y_{f+1}^p - y_f^p}{\sqrt{(y_{f+1}^p - y_f^p)^2 + (z_{f+1}^p - z_f^p)^2}} + 1 \tag{9}$$

$$cy_f^p = \frac{z_{f+1}^p - z_f^p}{\sqrt{(z_{f+1}^p - z_f^p)^2 + (x_{f+1}^p - x_f^p)^2}} + 1 \tag{10}$$

where m_f^p denotes the magnitude of the track and cz_f^p, cx_f^p, and cy_f^p represent the cosine of the angle projected in the z, x, and y axes plus one, respectively, which are in the range of 0 to 2 to satisfy the non-negative constraint in the NMF formulation. We then rescale all features into the range of 0 to 10 for each feature to be comparable.

For clustering, we gather all the motion features into a $5(F-1) \times P$ non-negative matrix $\mathbf{U} = [\mathbf{u}_1, ..., \mathbf{u}_n] \in \mathbb{R}_+^{m \times n}$, where the p-th column can be expressed as

$$\mathbf{u}_p = [m_1^p, \cdots, m_{F-1}^p, cz_1^p, \cdots, cz_{F-1}^p, cx_1^p, \cdots, cx_{F-1}^p, cy_1^p, \cdots, cy_{F-1}^p,]^T.$$

These features are always non-negative and can therefore be input to NMF.

Algorithm 1: Determination of the functional units

```
1. Extract motion features from displacement fields and
   construct U.
2. Apply graph-regularized sparse NMF to U to obtain
   V and W.
3. Compute affinity matrix A from W.
4. Apply spectral clustering to A and identify functional
   units.
```

D. *Graph-regularized Sparse NMF*

1) NMF: Given a non-negative data matrix $\mathbf{U}$ constructed from the motion quantities above and $k \leq \min(\mathrm{m}, \mathrm{n})$, let $\mathbf{V} = [v_{ik}] \in \mathbb{R}_+^{m \times k}$ be the *building blocks* and let $\mathbf{W} = [w_{kj}] \in \mathbb{R}_+^{k \times n}$ be the *weighting map*. The goal of NMF is to learn building blocks and corresponding weights such that the input $\mathbf{U}$ is approximated by a product of two non-negative matrices (i.e., $\mathbf{U} \approx \mathbf{VW}$). A typical way to define NMF is to use the Frobenius norm to measure the difference between $\mathbf{U}$ and $\mathbf{VW}$ [1] given by

$$\mathcal{E}(\mathbf{V}, \mathbf{W}) = \|\mathbf{U} - \mathbf{VW}\|_F^2 = \sum_{i,j} \left(u_{ij} - \sum_{k=1}^{K} v_{ik} w_{kj} \right)^2 \tag{11}$$

where $\|\cdot\|_F$ denotes the matrix Frobenius norm. The solution can be found through the multiplicative update rule [1]:

$$\mathbf{V} \leftarrow \mathbf{V}. * \mathbf{U}\mathbf{W}^T./\mathbf{V}\mathbf{W}\mathbf{W}^T \tag{12}$$

$$\mathbf{W} \leftarrow \mathbf{W}. * \mathbf{V}^T\mathbf{U}./\mathbf{V}^T\mathbf{V}\mathbf{W} \tag{13}$$

2) Sparsity Constraint: In this work, since we aim to identify the simplest muscle coordinations given many different combinations based on the current wisdom on phonological theories [18], we impose a sparsity constraint on the weighting map $\mathbf{W}$. The sparsity constraint allows us to encode the high-dimensional tongue motion data using a small number of active components, thereby making the weighting map simple and easy to interpret. In particular, the weighting map obtained this way will represent optimized tongue behavior that could generate the observed motion. In the NMF framework, it has been reported that a fractional regularizer using the $L_{1/2}$ norm outperformed the L_1 norm regularizer and gave sparser solutions [38]. Thus, we incorporate the $L_{1/2}$ sparsity constraint into the NMF framework, which can be expressed as

$$\mathcal{E}(\mathbf{V}, \mathbf{W}) = \frac{1}{2} \|\mathbf{U} - \mathbf{VW}\|_F^2 + \eta \|\mathbf{W}\|_{1/2}, \tag{14}$$

where the parameter $\eta \geqslant 0$ controls the sparseness of $\mathbf{W}$ and $\|\mathbf{W}\|_{1/2}$ is defined as

$$\|\mathbf{W}\|_{1/2} = \left(\sum_{i=1}^{k} \sum_{j=1}^{n} w_{ij}^{1/2} \right)^2 . \tag{15}$$

3) Manifold Regularization: Despite the high-dimensional configuration space of human motions, many human motions lie on low-dimensional manifolds that are non-Euclidean [37]. NMF with the $L_{1/2}$ norm sparsity constraint, however, produces a weighting map based on a Euclidean structure in the high-dimensional data space. Thus, the intrinsic and geometric relation between motion features may not be reflected accurately. To address this, we incorporate a manifold regularization that respects the intrinsic geometric structure as in [23], [39], [40]. The manifold regularization favors the local geometric structure while serving as a smoothness operator by reducing the interference of noise. Our final objective function incorporating both the manifold regularization and the sparsity constraint is then given by

$$\mathcal{E}(\mathbf{V}, \mathbf{W}) = \frac{1}{2}\|\mathbf{U} - \mathbf{V}\mathbf{W}\|_F^2 + \frac{1}{2}\lambda \mathrm{Tr}(\mathbf{W}\mathbf{L}\mathbf{W}^T) + \eta\, \|\mathbf{W}\|_{1/2} \tag{16}$$

where λ is a balancing parameter of the manifold regularization, Tr$(\cdot)$ denotes the trace of a matrix, $\mathbf{Q}$ is a heat kernel weighting, $\mathbf{D}$ is a diagonal matrix where $\mathbf{D}_{jj} = \sum_l \mathbf{Q}_{jl}$, and $\mathbf{L} = \mathbf{D} - \mathbf{Q}$, which is the graph Laplacian.

4) Minimization: The objective function in Eq. (16) is not convex in both $\mathbf{V}$ and $\mathbf{W}$ and therefore we use a multiplicative iterative method akin to that used in [40]. Let $\mathbf{\Psi} = [\psi_{mk}]$ and $\mathbf{\Phi} = [\phi_{kn}]$ be Lagrange multipliers subject to $v_{mk} \geq 0$ and $w_{kn} \geq 0$, respectively. By using the definition of the Frobenius norm, $\|\mathbf{U}\|_F = (\mathrm{Tr}(\mathbf{U}^T\mathbf{U}))^{1/2}$, and matrix calculus, the Lagrangian $\mathcal{L}$ is expressed as

$$\begin{aligned}\mathcal{L} = &\frac{1}{2}\mathrm{Tr}(\mathbf{U}\mathbf{U}^T) - \mathrm{Tr}(\mathbf{U}\mathbf{W}^T\mathbf{V}^T) + \frac{1}{2}\mathrm{Tr}(\mathbf{V}\mathbf{W}\mathbf{W}^T\mathbf{V}^T)\\ &+\frac{\lambda}{2}\mathrm{Tr}(\mathbf{W}\mathbf{L}\mathbf{W}^T) + \mathrm{Tr}(\mathbf{\Psi}\mathbf{V}^T) + \mathrm{Tr}(\mathbf{\Phi}\mathbf{W}^T) + \eta\,\|\mathbf{W}\|_{1/2}\,.\end{aligned} \tag{17}$$

The partial derivatives of $\mathcal{L}$ with respect to $\mathbf{V}$ and $\mathbf{W}$ are given by

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \mathbf{V}} &= -\mathbf{U}\mathbf{W}^T + \mathbf{V}\mathbf{W}\mathbf{W}^T + \mathbf{\Psi}\\ \frac{\partial \mathcal{L}}{\partial \mathbf{W}} &= -\mathbf{V}^T\mathbf{U} + \mathbf{V}^T\mathbf{V}\mathbf{W} + \lambda\mathbf{W}\mathbf{L} + \frac{\eta}{2}\mathbf{W}^{-1/2} + \mathbf{\Phi}.\end{aligned} \tag{18}$$

Finally, the update rule is found by using Karush-Kuhn-Tucker conditions—i.e., $\mathbf{\Psi}_{mk}\mathbf{V}_{mk} = 0$ and $\mathbf{\Phi}_{kn}\mathbf{W}_{kn} = 0$:

$$\begin{aligned}&\mathbf{V} \leftarrow \mathbf{V}.*\mathbf{U}\mathbf{W}^T./\mathbf{V}\mathbf{W}\mathbf{W}^T\\ &\mathbf{W} \leftarrow \mathbf{W}.*(\mathbf{V}^T\mathbf{U} + \lambda\mathbf{W}\mathbf{Q})./(\mathbf{V}^T\mathbf{V}\mathbf{W} + \frac{\eta}{2}\mathbf{W}^{-1/2}\\ &+\lambda\mathbf{W}\mathbf{D}).\end{aligned} \tag{19}$$

E. Spectral Clustering

The non-negative weighting map that is simple and sparse obtained in Eq. (19) provides a good measure of regional tissue point similarity. To obtain the final clustering results from the weighting map, spectral clustering is used to determine the cohesive motion patterns as spectral clustering outperforms traditional clustering algorithms such as the K-means algorithm [41].

Once $\mathbf{W}$ is determined from Eq. (19), an affinity matrix $\mathbf{A}$ is first constructed:

$$\mathbf{A}(i,j) = \exp\left(-\frac{\|w(i) - w(j)\|_2}{\sigma}\right), \tag{20}$$

where $w(i)$ is the i-th column vector of $\mathbf{W}$ and σ denotes the scale (we set σ = 0.01 in this work). The column vectors of $\mathbf{W}$ form nodes in the graph, and the similarity $\mathbf{A}$ computed between column vectors of $\mathbf{W}$ form the edge weights. On the affinity matrix, we apply a spectral clustering technique using a normalized cut algorithm [42]. From a graph cut perspective, our method can be seen as identifying subgraphs representing motions that exhibit distinct characteristics.

F. Model Selection

To achieve the best clustering quality, we need to determine the optimal number k of clusters, which is a challenging task [45]. In this work, we use a consensus matrix whose entries indicate probabilities that samples i and j belong to the same cluster by repeating NMF 10 times. To compute the dispersion coefficient of a consensus matrix C, the dispersion coefficient ρ is defined as

$$\rho = \frac{1}{n^2}\sum_{i=1}^{m}\sum_{j=1}^{n} 4(\tilde{c}_{ij} - 0.5)^2, \tag{21}$$

where $\tilde{c}_{ij}$, m, and n denote each entry of the matrix, the row and column size of the matrix, respectively. In the ideal case, for a consensus matrix whose entries are all 0s or 1s, we have $\rho = 1$; for a scattered consensus matrix, we have $0 < \rho < 1$. The optimal number of clusters is determined as the one with the maximal ρ_k [46]. We set $2 \leq k \leq 6$ in this work.

IV. EXPERIMENTAL RESULTS

We describe the qualitative and quantitative evaluation to validate the proposed approach. We both show synthetic and real *in vivo* data to demonstrate the accuracy of our approach.

A. Experiments Using 2D Data

We first used two 2D datasets to demonstrate the clustering performance of the proposed method. The first dataset is the COIL20 image library, which contains 20 classes (32×32 gray scale images of 20 objects). The second dataset is the CMU PIE face database, which has 68 classes (32×32 gray scale face images of 68 persons). In order to compare the performance of the different algorithms, we used a K-means clustering method (K-means), a normalized cut method (N-Cut) [42], standard NMF with K-means clustering (NMF-K), graph-regularized NMF with K-means clustering (G-NMF-K) [23], graph-regularized NMF with spectral clustering (G-NMF-S), graph-regularized sparse NMF with K-means clustering (GS-NMF-K), and our method (GS-NMF-S). Two metrics, the Normalized Mutual Information (NMI) and the accuracy (AC), were used to measure the clustering performance as used in [23]. Table 1 lists the NMI and AC values, demonstrating that the proposed method outperformed other methods. We also compared the $L_{1/2}$ and L_1 norms experimentally, and the $L_{1/2}$ norm had slightly better results.

B. Experiments Using Synthetic Tongue Motion Data

Since there is no ground truth in our *in vivo* tongue motion data, we evaluated the performance of the proposed method using four synthetic tongue motion datasets using a composite Lagrangian displacement field of individual muscle groups based on a tongue atlas [11]. Each muscle group was defined by a mask volume with a value of 1 inside the muscle group, and zero elsewhere. Since the masks were known, it also provided ground truth labels to assess the accuracy of the output of the clustering method. The first and second datasets used genioglossus (GG) and superior longitudinal (SL) muscles with and without interdigitated regions. The GG

TABLE I
CLUSTERING PERFORMANCE: NMI AND AC

NMI (%)	K-means	N-Cut	NMF-K	G-NMF-K	GS-NMF-K	G-NMF-S	Our method
COIL20 (K=20)	73.80%	76.56%	74.36%	87.59%	90.11%	90.24%	**90.63**%
PIE (K=68)	54.40%	77.13%	69.82%	89.93%	89.95%	90.95%	**91.74**%
AC (%)	K-means	N-Cut	NMF-K	G-NMF-K	GS-NMF-K	G-NMF-S	Ours
COIL20 (K=20)	60.48%	66.52%	66.73%	72.22%	83.75%	84.58%	**85.00**%
PIE (K=68)	23.91%	65.91%	66.21%	79.3%	79.93%	80.60%	**84.31**%

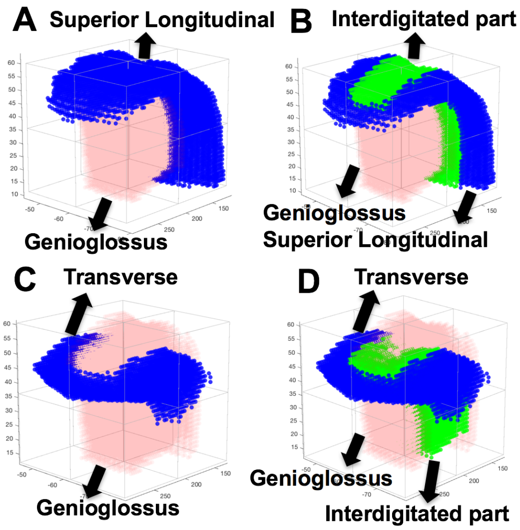

Fig. 2. Illustration of synthetic tongue motion simulation results: (a) translation plus rotation without interdigitated regions (2 clusters), (b) rotations with interdigitated regions (3 clusters), (c) translation plus rotation without interdigitated regions (2 clusters), and (d) rotations with interdigitated regions (3 clusters). It is noted that our approach identified each label accurately as visually assessed.

muscle was translated while the SL muscle was rotated -0.1 radians about the x direction in Fig. 2(a) in the course of 11 time frames. The GG muscle was rotated -0.1 radians about the x direction while the SL muscle was rotated 0.1 radians about the x direction in Fig. 2(b) in the course of 11 time frames. The third and fourth datasets were generated by applying the same composite Lagrangian displacement field to the GG and transverse muscles with and without interdigitated regions as in Fig. 2(c) and (d), respectively. Fig. 2 showed the final clustering results using our method. We attempted to cluster each dataset into two (first and third datasets) and three (second and fourth datasets) distinct motions, respectively, where we obtained 100% clustering accuracy for all datasets when evaluated against the ground truth labels.

C. Experiments Using In Vivo Tongue Motion Data

TABLE II
CHARACTERISTICS OF *in vivo* TONGUE MOTION DATA

Task	Protrusion	/s/-/u/	/i/-/s/
Time frames	1-13	10-17	15-22
Number of clusters	2-4	2-4	2-4

We also tested our method using a simple non-speech protrusion task and speech tasks: "a souk" and "a geese". Ten subjects performed "a souk" and "a geese" tasks and one subject performed the protrusion task. We used the features including the magnitude and angle of each track as our input to the NMF framework. Table II lists the characteristics of our *in vivo* tongue motion data including time frames analyzed and the number of clusters based on the dispersion coefficient.

First, the functional units have been extracted using our method for two clusters (Fig. 3(b)), three clusters (Fig. 3(c)), and four clusters (Fig. 3(d)), respectively, based on the dispersion coefficient as in Fig. 4 and visual assessment. The outer tongue layer expands forward and upward (but not backward), and the region near the jaw has little motion as shown in Fig. 3(a). Fig. 3(b) is a good representation of forward protrusion (red) vs. small motion (blue). In addition, as the number of clusters increases as shown in Fig. 3(c) and (d), subdivision of large regions in small motion (blue, Fig. 3(b)) into small functional units was observed.

Second, the functional units during $/s/$ to $/u/$ from "a souk" were determined using our method for two clusters (Fig. 5(b)), three clusters (Fig. 5(c)), and four clusters (Fig. 5(d)), respectively, based on the dispersion coefficient shown in Fig. 6 and visual assessment. These motions are characterized by forward to upward/backward motion of the tongue tip, upward motion of the tongue body, and forward motion of the posterior tongue as in Fig. 5(a). Fig. 5(b) shows two clusters including the tip plus bottom of the tongue (red) versus the tongue body. Three clusters as in Fig. 5(c) show a good representation of the tip, body and posterior of the tongue and four clusters as in Fig. 5(d) further subdivided the tongue tip and bottom.

Third, the functional units during $/i/$ to $/s/$ from "a geese" were determined using our method for two clusters (Fig. 7(b)), three clusters (Fig. 7(c)), and four clusters (Fig. 7(d)), respectively, based on the dispersion coefficient shown in Fig. 8 and visual assessment. These motions are characterized by an upward motion of the tongue tip, upward/backward motion of the tongue body, and forward motion of the posterior tongue as in Fig. 7(a). Two clusters as in Fig. 7(b) show a division between the tip plus bottom of the tongue (red) and the tongue body. Three clusters as in Fig. 7(c) are a good representation of the tip, body and posterior of the tongue and four clusters as in Fig. 7(d) subdivided the posterior of the tongue further.

V. DISCUSSION

In this work, we proposed a novel approach to characterizing multiple functional degrees of freedom of the tongue, which is critical to understand the tongue's role in speech and other lingual behaviors. This is because the functioning

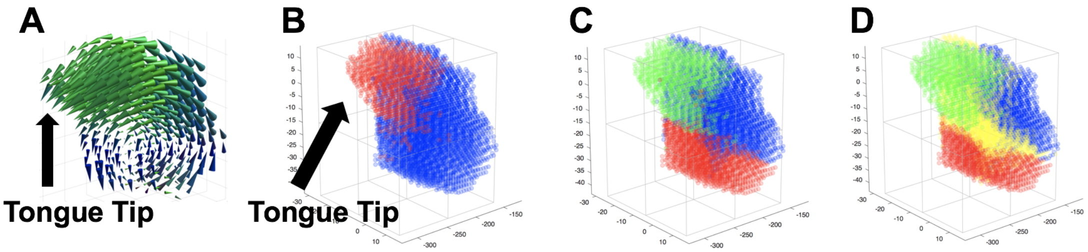

Fig. 3. Illustration of functional units during the tongue protrusion task, showing (a) 3D Lagrangian displacement field, (b) functional units (2 clusters), (c) functional units (3 clusters), and (d) functional units (4 clusters). It is noted that the colored clustering results are plotted in the tongue shape of the first time frame showing the neutral tongue position.

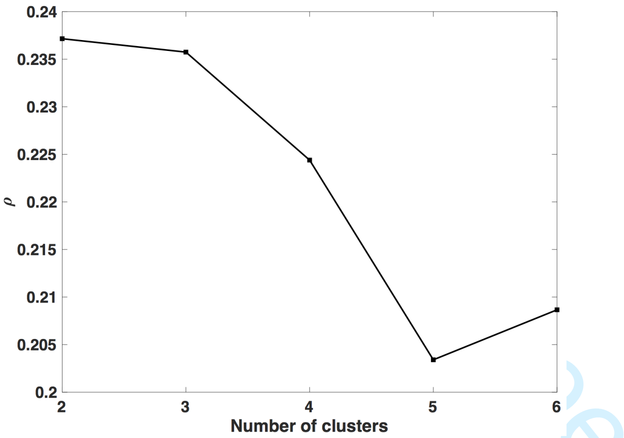

Fig. 4. Plot of the dispersion coefficient versus the different number of clusters for the tongue protrusion task.

of the tongue in speech or other lingual behaviors entails successful orchestration of the complex system of 3D tongue muscular structures over time. Determining functional units from healthy controls plays an important role in understanding motor control strategy, which in turn could elucidate adapted motor control strategy when analyzing patient data such as tongue cancer patients. It has been a long-sought problem that many researchers attempted using various techniques.

Inspired by recent advances in MR motion tracking and data mining schemes including sparse NMF and manifold learning, we presented a novel method for determining functional units from tMRI, which opens new vistas to study speech production. Unsupervised data clustering using NMF is the task of identifying semantically meaningful clusters using a low-dimensional representation from a dataset. Unlike previous algorithms, this proposed work aimed at identifying the internal, coherent manifold structure of high-dimensional 4D motion data. Two constraints in addition to the standard NMF were employed to reflect the physiological properties of 4D tongue motion during speech. Firstly, the sparsity constraint was introduced to capture the simplest and the most optimized weighting map. Sparsity has been one of important properties for phonological theories [18], and our work attempted to decode this phenomenon within a sparse NMF framework. Secondly, the manifold regularization was added to capture the intrinsic and geometric relationship between motion features. It also allows preserving the geometric structure between motion features, which is particularly important when dealing with tongue motions that lie on low-dimensional non-Euclidean manifolds. Our method performed better than K-means, N-Cut, NMF-K, G-NMF-K, GS-NMF-K, and G-NMF-S using 2D data.

As for the input features in our framework, we used instantaneous velocity information derived from the point tracks. More features could be investigated such as those reflect mechanical properties including principal strains, curvature, minimum-jerk, two-thirds power law, and isochrony [48] or motion descriptors combining those individual features.

The selection of the number of clusters is often performed manually as there is no definite model selection method available. In this work, we built a “consensus matrix” from multiple runs for each k and assessed the presence of block structure. As an alternative, one can compare reconstruction errors for different number k or examine the stability (i.e., agreement between results) from multiple randomly initialized runs for each k. Since there is no ground truth in our tongue data, we have used both visual assessment and the model selection approach in which the model selection approach provided an upper limit of the number of clusters.

There are a few directions to improve the current work. First, we used a data-driven approach to determine the functional units, which was visually assessed due to the lack of ground truth. This could be improved by further studies using model-based approaches via biomechanical stimulations [43] or electromyography [44] to co-validate our findings. For biomechanical simulations, subject-specific anatomy and the associated weighting map could be input and inverse simulation can then be used to verify the validity of the obtained weighting map. Second, we used magnitude and angle of each track as our input features. In order to equal the weight of each input feature, we normalized the feature values in the same range. In our future work, we will further study automatic relevance determination methods to model the interactions among these features to yield the best clustering outcome. Finally, the identified functional units as shown in our experimental results may involve multiple regions that correspond to sub-muscles or multiple muscles. Therefore, we will further study the identified functional units in the context of the muscular anatomy from individual high-resolution MRI, diffusion MRI, or a high-resolution atlas [11].

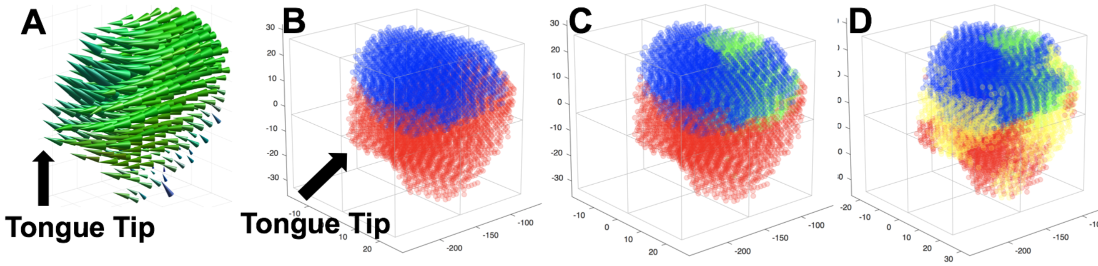

Fig. 5. Illustration of functional units during $/s/$ to $/u/$ from "a souk" showing (a) 3D Lagrangian displacement field, (b) functional units (2 clusters), (c) functional units (3 clusters), and (d) functional units (4 clusters). It is noted that the colored clustering results are plotted in the tongue shape of the first time frame showing the neutral tongue position.

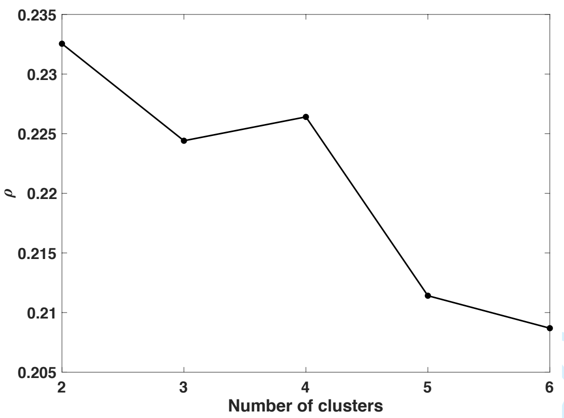

Fig. 6. Plot of the dispersion coefficient versus the different number of clusters for the task of $/s/$ to $/u/$ from "a souk".

VI. CONCLUSION

We have presented a new algorithm to determine local functional units that link muscle activity to surface tongue geometry during non-speech and speech tasks. Our work applied a graph-regularized sparse NMF method that incorporates joint sparse and manifold regularizations to the motion tracking data from tMRI. Both synthetic and *in vivo* tongue data were used to verify the performance of the proposed method, demonstrating that the proposed method was able to accurately cluster the tongue motion. Our results suggest that it is feasible to identify the functional units using a set of motion features including magnitude and angle of each track, and this proposed method has great potential in the improvement of diagnosis, treatment, and rehabilitation in patients with speech-related disorders.

ACKNOWLEDGMENT

This research was supported in part by NIH R21DC016047, R00DC012575, R01DC014717, and R01CA133015.

REFERENCES

[1] D. D. Lee, H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," Nature, 401(6755), pp. 788–791, 1999

[2] G. Trigeorgis, K. Bousmalis, S. Zafeiriou, B. W. Schuller, "A deep matrix factorization method for learning attribute representations," IEEE Transactions on Pattern Analysis and Machine Intelligence. 2016 Apr 15.

[3] M. Stone, M. A. Epstein, K. Iskarous, "Functional segments in tongue movement," Clinical Linguistics & Phonetics 18(6-8), pp. 507–521, 2004

[4] J. R. O'Kusky and M. G. Norman, "Sudden infant death syndrome: increased number of synapses in the hypoglossal nucleus," Journal of Neuropathology and Experimental Neurology, 54, pp. 627–34, 1995

[5] V. Parthasarathy, J. L. Prince, M. Stone, E. Z. Murano, M. NessAiver, "Measuring tongue motion from tagged cine-MRI using harmonic phase (HARP) processing," Journal of the Acoustical Society of America, 121(1), pp. 491–504, January, 2007

[6] J. R. Green, "Tongue-surface movement patterns during speech and swallowing," The Journal of the Acoustical Society of America 113(5), pp. 2820–2833, 2003

[7] J. A. S. Kelso, "Synergies: Atoms of Brain and Behavior," In Progress in Motor Control, pp. 83–91, Springer, 2009

[8] M. Stone, E. P. Davis, A. S. Douglas, M. N. Aiver, R. Gullapalli, W. S. Levine, A. J. Lundberg, "Tongue-surface movement patterns during speech and swallowing," Journal of Speech, Language, and Hearing Research, 113(5), pp. 2820–2833, 2001

[9] E. Bresch, Y. C. Kim, K. Nayak, D. Byrd, S. Narayanan, "Seeing speech: Capturing vocal tract shaping using real-time magnetic resonance imaging," IEEE Signal Processing Magazine, 25(3), pp. 123–132, 2008

[10] M. Fu, Z. Bo, C. Carignan, R. K. Shosted, J. L. Perry, D. P. Kuehn, Z.P. Liang, and B. P. Sutton, "Highresolution dynamic speech imaging with joint lowrank and sparsity constraints," Magnetic Resonance in Medicine, 73(5), pp. 1820-32, 2015

[11] J. Woo, J. Lee, E. Murano, F. Xing, A. Meena, M. Stone and J. Prince, "A high-resolution atlas and statistical model of the vocal tract from structural MRI," Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, pp. 1–14, 2015

[12] M. Stone, J. Woo, J. Lee, T. Poole, A. Seagraves, M. Chung, E. Kim, E. Z. Murano, J. L. Prince, S. S. Blemker, "Structure and variability in human tongue muscle anatomy," Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, pp. 1–9, 2016

[13] F. Xing, J. Woo, E. Z. Murano, J. Lee, M. Stone, J. L. Prince, "3D tongue motion from tagged and cine MR images," International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), LNCS 8151, pp. 41–48, 2013

[14] B. Gick and I. Stavness, "Modularizing speech," Frontiers in psychology, 4, pp. 1–3, 2013.

[15] M. Stone, "A three-dimensional model of tongue movement based on ultrasound and x-ray microbeam data," Journal of the Acoustical Society of America, 87, pp.2207-2217, 1990

[16] V. Ramanarayanan, L. Goldstein, and S. Narayanan, "Spatio-temporal articulatory movement primitives during speech production: extraction, interpretation, and validation," Journal of the Acoustical Society of America, 134(2), pp. 1378–1394, 2013

[17] J. Woo, F. Xing, J. Lee, M. Stone, J. L. Prince, "Determining Functional Units of Tongue Motion via Graph-regularized Sparse Non-negative Matrix Factorization," International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), 17(Pt 2), pp. 146–53, 2014

[18] C. Browman and L. Goldstein, "Dynamics and articulatory phonology," in Mind as Motion: Explorations in the Dynamics of Cognition, edited by R. F. Port and T.van Gelder (MIT Press, Cambridge, MA), pp. 175-194.

[19] T. Li and C. H. Q. Ding, "Nonnegative Matrix Factorizations for Clustering: A Survey," In: Data Clustering: Algorithms and Applications. Chapman and Hall/CRC, pp. 149-176, 2013

[20] N. F. Osman, W. S. Kerwin, E. R. McVeigh, J. L. Prince, "Cardiac motion tracking using CINE harmonic phase (HARP) magnetic resonance imaging," Magnetic resonance in medicine, 42(6), pp. 1048-1060, 1999

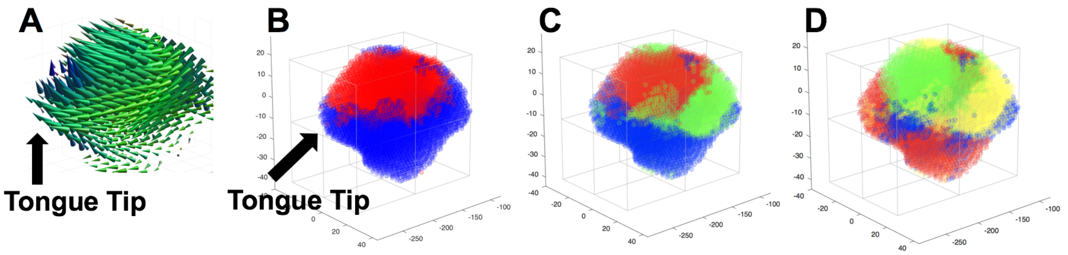

Fig. 7. Illustration of functional units during $/i/$ to $/s/$ from “a geese”, showing (a) 3D Lagrangian displacement field, (b) functional units (2 clusters), (c) functional units (3 clusters), and (d) functional units (4 clusters). It is noted that the colored clustering results are plotted in the tongue shape of the first time frame showing the neutral tongue position.

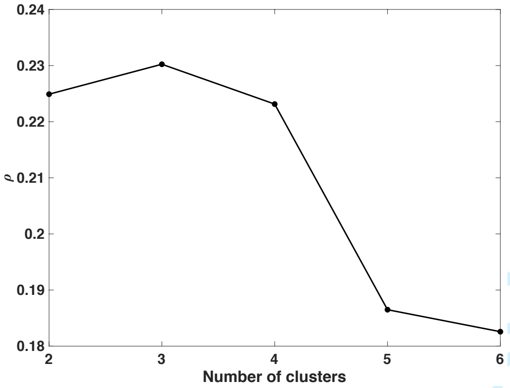

Fig. 8. Plot of the dispersion coefficient versus the different number of clusters for the task of $/i/$ to $/s/$ from “a geese”.

[21] T. Mansi, X. Pennec, M. Sermesant, H. Delingette, and N. Ayache, “iLogDemons: A demons-based registration algorithm for tracking incompressible elastic biological tissues,” International journal of computer vision, 92(1), pp. 92–111, 2011

[22] F. Xing, C. Ye, J. Woo, M. Stone, J. L. Prince, “Relating speech production to tongue muscle compressions using tagged and high-resolution magnetic resonance imaging,” SPIE Medical Imaging, Florida, FL, 2015

[23] D. Cai, X. He, J. Han, T. S. Huang, “Graph regularized nonnegative matrix factorization for data representation,” IEEE Transactions on Pattern Analysis and Machine Intelligence 33(8), pp. 1548–1560, 2011

[24] H. Shinagawa, E. Z. Murano, J. Zhuo, B. Landman, R. Gullapalli, and J. L. Prince, M. Stone, “Tongue muscle fiber tracking during rest and tongue protrusion with oral appliances: A preliminary study with diffusion tensor imaging,” Acoustical science and technology, 29(4), pp. 291–294, 2008

[25] T. A. Gaige, T. Benner, R. Wang, V. J. Wedeen, R. J. Gilbert, R.J., “Three dimensional myoarchitecture of the human tongue determined in vivo by diffusion tensor imaging with tractography,” Journal of Magnetic Resonance Imaging, 26(3), pp. 654–661, 2007

[26] M. NessAiver and J. L. Prince, “Magnitude image CSPAMM reconstruction (MICSR),” Magnetic Resonance in Medicine, 50(2), pp. 331–342, 2003

[27] J. Woo, M. Stone, Y. Suo, E. Murano, J. L. Prince, “Tissue-Point Motion Tracking in the Tongue From Cine MRI and Tagged MRI,” Journal of Speech, Language, and Hearing Research, 57(2), pp. 626–636, 2013

[28] J. Kim and H. Park, “Sparse nonnegative matrix factorization for clustering,”Technical Report GT-CSE-08-01, Georgia Institute of Technology, 2008

[29] F. Shahnaz, M. W. Berry, V. P. Pauca, and R. J. Plemmons, “Document clustering using nonnegative matrix factorization,” Information Processing and Management, 42(2), pp. 373–386, 2006

[30] Y. Gao, G. Church, “Improving molecular cancer class discovery through sparse non-negative matrix factorization,” Bioinformatics, 21(21), pp. 3970–3975, 2005

[31] J. Wang, X. Wang, X. Gao, “Non-negative matrix factorization by maximizing correntropy for cancer clustering,” BMC Bioinformatics, 14(1), pp. 1–11, 2013

[32] A. Anderson, P. K. Douglas, W. T., Kerr, V. S. Haynes, A. L. Yuille, A.L., J. Xie, Y. N. Wu, J. A. Brown, M. S. Cohen, “Non-negative matrix factorization of multimodal MRI, fMRI and phenotypic data reveals differential changes in default mode subnetworks in ADHD,” NeuroImage, 102(1), pp. 207–219, 2014

[33] Q. Mo, and B. A. Draper, “Semi-nonnegative matrix factorization for motion segmentation with missing data,” European Conference on Computer Vision, pp. 402–415, Berlin, Heidelberg, 2012

[34] A. Cheriyadat and R. J. Radke, “Non-negative matrix factorization of partial track data for motion segmentation,” IEEE 12th International Conference in Computer Vision, pp. 865–872, 2009

[35] F. Xing, J. Woo, A. Gomez, D. L. Pham, P. V. Bayly, M. Stone, and J. L. Prince, “Phase Vector Incompressible Registration Algorithm (PVIRA) for Motion Estimation from Tagged Magnetic Resonance Images,” IEEE Trans on Medical Imaging, 36(10), pp. 2116–2128, Oct, 2017

[36] F. Xing, J. Woo, E. Z. Murano, J. Lee, M. Stone, J. L. Prince, “3D tongue motion from tagged and cine MR images,” 16th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), Nagoya, Japan, pp. 865–872, September, 2013

[37] A. Elgammal and C. S. Lee, “The role of manifold learning in human motion analysis,” In B. Rosenhahn, R. Klette, D. Metaxas, (eds.) Human Motion. Computational Imaging and Vision, vol. 36, pp. 25–56. Springer, Netherlands, 2008

[38] Y. Qian, S. Jia, J. Zhou, J., A. Robles-Kelly, “Hyperspectral unmixing via sparsity-constrained nonnegative matrix factorization,” IEEE Transactions on Geoscience and Remote Sensing, pp. 4282–4297, 2011

[39] S. Yang, C. Hou, C. Zhang, Y. Wu, “Hyperspectral unmixing via sparsity-constrained nonnegative matrix factorization,” IEEE Transactions on Geoscience and Remote Sensing, pp. 4282–4297, 2011

[40] X. Lu, H. Wu, Y. Yuan, P. Yan, X. Li, “Manifold regularized sparse NMF for hyperspectral unmixing,” IEEE Transactions on Geoscience and Remote Sensing, 51(5), pp. 2815-2826, 2013

[41] U. Von Luxburg, “A tutorial on spectral clustering,” Statistics and Computing, 17(4), pp. 395–416, 2007

[42] J. Shi and J. Malik, “Normalized cuts and image segmentation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(8), pp. 888–905, 2000

[43] I. Stavness, J. E. Lloyd and S. Fels, “Automatic prediction of tongue muscle activations using a finite element model,” Journal of Biomechanics, 45(16), pp. 2841–2848, 2012

[44] L. Pittman and E. F. Bailey, “Genioglossus and intrinsic electromyographic activities in impeded and unimpeded protrusion tasks,” Journal of Neurophysiology, 101(1), pp. 276–282, 2009

[45] Z. Zhang, Y., T. Li,, C. Ding, R., X.-W., and X. S. Zhang, “Binary matrix factorization for analyzing gene expression data. Data Mining and Knowledge Discovery,” 20(1), pp. 28–52, 2010

[46] K. Da, J. Choo, and H. Park, “Nonnegative matrix factorization for interactive topic modeling and document clustering,” In Partitional Clustering Algorithms, pp. 215–243, 2015.

[47] C. Boutsidis and E. Gallopoulos, “SVD based initialization: A head start for nonnegative matrix factorization,” Pattern Recognition, 41(4) pp. 1350-1362, 2008.

[48] P. Viviani, T. Flash, “Minimum-jerk, two-thirds power law, and isochrony: converging approaches to movement planning,” Journal of Experimental Psychology: Human Perception and Performance, Feb;21(1):32, 1995.

[49] J. Woo, F. Xing, M. Stone, J. Green, T. Reese, T. Brady, V. Wedeen, J. Prince, and G. El Fakhri,, “Speech Map: A Statistical Multimodal Atlas of 4D Tongue Motion During Speech from Tagged and Cine MR Images,” Journal of Computer Methods in Biomechanics and Biomedical Engineering (CMBBE): Special Issue on Imaging and Visualization, 2017