HHS Public Access

Author manuscript

Magn Reson Med. Author manuscript; available in PMC 2016 May 01.

Published in final edited form as:

Magn Reson Med. 2015 May ; 73(5): 1820–1832. doi:10.1002/mrm.25302.

High-Resolution Dynamic Speech Imaging with Joint Low-Rank and Sparsity Constraints

Maojing Fu[1,2], **Bo Zhao**[1,2], **Christopher Carignan**[3], **Ryan K. Shosted**[4], **Jamie L. Perry**[5], **David P. Kuehn**[6], **Zhi-Pei Liang**[1,2], and **Bradley P. Sutton**[2,7]

[1]Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois

[2]Beckman Institute of Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, Illinois

[3]Department of English, North Carolina State University, Raleigh, North Carolina

[4]Department of Linguistics, University of Illinois at Urbana-Champaign, Urbana, Illinois

[5]Department of Communication Sciences and Disorders, East Carolina University, Greenville, North Carolina

[6]Department of Speech and Hearing Science, University of Illinois at Urbana-Champaign, Urbana, Illinois

[7]Department of Bioengineering, University of Illinois at Urbana-Champaign, Urbana, Illinois

Abstract

Purpose—To enable dynamic speech imaging with high spatiotemporal resolution and full-vocal-tract spatial coverage, leveraging recent advances in sparse sampling.

Methods—An imaging method is developed to enable high-speed dynamic speech imaging exploiting low-rank and sparsity of the dynamic images of articulatory motion during speech. The proposed method includes: a) a novel data acquisition strategy that collects navigators with high temporal frame rate, and b) an image reconstruction method that derives temporal subspaces from navigators and reconstructs high-resolution images from sparsely sampled data with joint low-rank and sparsity constraints.

Results—The proposed method has been systematically evaluated and validated through several dynamic speech experiments. A nominal imaging speed of 102 frames per second (fps) was achieved for a single-slice imaging protocol with a spatial resolution of $2.2 \times 2.2 \times 6.5\ mm^3$. An eight-slice imaging protocol covering the entire vocal tract achieved a nominal imaging speed of 12.8 fps with the identical spatial resolution. The effectiveness of the proposed method and its practical utility was also demonstrated in a phonetic investigation.

Conclusion—High spatiotemporal resolution with full-vocal-tract spatial coverage can be achieved for dynamic speech imaging experiments with low-rank and sparsity constraints.

Correspondence to: Maojing Fu, 4259 Beckman Institute for Advanced Science and Technology, 405 N Mathews Ave, Urbana, IL 61801, USA, mfu2@illinois.edu.

Keywords

partial separability modeling; low-rank approximation; sparsity; spiral navigation; dynamic speech imaging

INTRODUCTION

Dynamic MRI has recently been recognized as a powerful tool for speech imaging, especially for monitoring dynamic changes in the oropharyngeal and nasopharyngeal regions and capturing articulatory gestures during the speech process. It also has the potential to reveal the soft-tissue structures (e.g., pharyngeal cavity and internal musculature) and articulatory dynamics across arbitrarily-oriented imaging planes. This capability has facilitated a broad spectrum of speech-related studies, such as examining the articulatory movement during speech production (1–3), the process of singing in diverse forms (4, 5), and the function of swallowing under normal and pathological conditions (6, 7). It is also useful for the analysis of a variety of physiological defects and disorders, such as cleft palate (8), velopharyngeal dysfunction (9), motor dysfunction (10) and sleep apnea (11). The work presented in this paper is to further enhance the capability of dynamic speech MRI with high temporal resolution, high spatial resolution and broad spatial coverage.

Ideally, an effective dynamic speech imaging method should have the following three properties: First, it should provide high temporal resolution to capture fast-varying speech dynamics. A common goal of many speech imaging applications is to accurately capture the speech process, in which the shapes and position of articulators experience rapid transitions over time. For instance, it has been shown that a frame rate of 30 to 50 fps was necessary to observe dynamic velar retraction and coarticulation effects during speech (12–14). Second, an imaging method should offer high spatial resolution to delineate fine features of the articulators, in particular fine scale articulation in the tongue tip and velopharyngeal port. For example, previous work has used an in-plane spatial resolution of 1.9 mm in order to effectively observe and model the tongue tip motion (15). Third, an imaging method should allow for increased coverage through multiple imaging planes to explore soft-tissue structures in the region of interest in the vocal tract. Some speech motions involve the interaction and coordination of articulators from different spatial locations, which require multiple imaging planes to visualize. Particularly, more than two imaging planes have been used to study English fricative sounds and Arabic pharyngeal sounds (16, 17).

Although trade-offs exist to satisfy each of the above-mentioned properties by compromising others, in general it remains challenging to simultaneously achieve adequate performance in all three properties in dynamic MRI. Significant efforts have been made in recent years to improve dynamic speech imaging methods. Combining a multishot spiral FLASH sequence with a field-inhomogeneity-corrected reconstruction method, Sutton et al. (18) resolved high-resolution spatial details of the velopharyngeal mechanism at 21.4 fps. Kim et al. (16) employed interleaved multislice data acquisitions with gridding reconstructions to visualize fricative production across three intersecting imaging planes at a frame rate of 6.1 fps. Scott et al. (19) applied an adaptive averaging technique to improve

the signal-to-noise ratio (SNR) for the visualization of velar activities on a mid-sagittal plane at 14.9 fps. Using radial trajectories, Li et al. (20) incorporated aggregated motion estimation into a nonlinear reconstruction method to visualize tongue movement at 30.3 fps for a single imaging plane. Despite these advances, higher spatiotemporal resolution and broader spatial coverage are needed for dynamic speech imaging applications.

Dynamic speech MRI, as a specific application of dynamic MRI, could benefit from many emerging techniques developed for faster MRI. Recent advances in fast pulse sequences (18,21) and parallel imaging (22,23) have proven to be effective in reducing acquisition time. In addition, model-based image reconstruction methods have also provided many opportunities to achieve sub-Nyquist sampling for dynamic MR experiments. These model-based methods are generally based on low-dimensional signal models. One category of methods uses the limited spatial-spectral support to relax the associated data acquisition requirements (24–27). Recently, compressed sensing (CS) based techniques have successfully used sparsity constraints to achieve high quality reconstructions despite sparse (**k**, *t*)-space sampling (28–35). Alternatively, sparse sampling methods based on the partial separability (PS) model (36–38) and its induced low-rank constraint (39–41) have also demonstrated effectiveness in various spatiotemporal imaging applications (42–44). Although the sampling methods were tailored based on the specific applications, the low-rank constraint has been combined with other constraints to yield improved reconstruction quality. The complementary advantages of low-rank and sparsity constraints have been previously discussed in (39, 40).

Expanding upon our early studies (45, 46), we propose a model-based imaging method for dynamic speech imaging. The method effectively integrates parallel imaging, spiral-navigator-based data acquisition and constrained image reconstruction. The method has been systematically evaluated through a number of dynamic speech imaging experiments, demonstrating its great potential in simultaneously achieving high spatial resolution, high temporal resolution and full-vocal-tract coverage. Also, its practical utility is demonstrated through a phonetic investigation of nasalization.

THEORY

Partial Separability (PS) Model

In a dynamic speech imaging experiment, the acquired (**k**, *t*)-space data, *d*(**k**, *t*), can be expressed as:

$$d(\mathbf{k},t)=\int I(\mathbf{r},t)e^{-i2\pi\mathbf{k}\cdot\mathbf{r}}d\mathbf{r}+\eta(\mathbf{k},t), \quad (1)$$

where *I*(**r**, *t*) is the desired image series and η(**k**, *t*) represents the measurement noise. The PS model expresses *I*(**r**,*t*) as (36):

$$I(\mathbf{r},t)=\sum_{l=1}^{L}\psi_l(\mathbf{r})\phi_l(t), \quad (2)$$

where L is the model order, $\{\psi_l(\mathbf{r})\}_{l=1}^{L}$ denotes a set of spatial basis functions, and $\{\phi_l(t)\}_{l=1}^{L}$ denotes a set of temporal basis functions. The PS model is based on the assumption that strong spatiotemporal correlation exists in $I(\mathbf{r}, t)$. In speech imaging experiments particularly, this assumption is often valid because a) spatial images corresponding to articulations of similar sounds are highly correlated, b) image voxels within certain bulk articulators (e.g. the tongue or velum) share similar temporal dynamics, and c) only a few driving muscles are involved in the production of a specific syllable. The PS model implies that a data matrix $\hat{\mathbf{I}}$ (referred to as a Casorati matrix in (36)) defined over any point set $\{I(\mathbf{r}_n, t_m)\}_{n,m=1}^{N,M}$,

$$\hat{\mathbf{I}} = \begin{bmatrix} I(\mathbf{r}_1, t_1) & \cdots & I(\mathbf{r}_1, t_M) \\ \vdots & \ddots & \vdots \\ I(\mathbf{r}_N, t_1) & \cdots & I(\mathbf{r}_N, t_M) \end{bmatrix}, \quad (3)$$

has a rank upper bounded by L (36,37), where N denotes the number of spatial encoding steps and M denotes the number of image frames. An equivalent data matrix $\hat{\mathbf{C}}$ can also be defined in the ($\mathbf{k}$, t)-space. Based on the low-rank property of $\hat{\mathbf{I}}$ (40), we can express it as: $\hat{\mathbf{I}} = \mathbf{UV}$, where $\mathbf{U} \in \mathbb{C}^{N \times L}$ denotes a matrix with its columns spanning the spatial subspace of $\hat{\mathbf{I}}$, and $\mathbf{V} \in \mathbb{C}^{L \times M}$ denotes a matrix with its rows spanning the temporal subspace of $\hat{\mathbf{I}}$.

Data Acquisition

We use the strategy proposed in (36,47) to sparsely sample two sets of ($\mathbf{k}$, t)-space data in an interleaved manner: a navigator data set and an imaging data set. For these two data sets, the navigator data is used to estimate $\mathbf{V}$, while the imaging data is used to estimate $\mathbf{U}$. Although each data set serves a different purpose, both data sets contribute to reproducing $\hat{\mathbf{C}}$.

Due to the different function of these two data sets (36), separate considerations were imposed in terms of acquisition trajectory and sampling requirements. For the navigator data set, we propose a new spiral-trajectory-based technique to navigate the ($\mathbf{k}$, t)-space with high temporal resolution. Figure 1 illustrates this strategy with a simplified pulse sequence diagram and a corresponding ($\mathbf{k}$, t)-space sampling pattern. Unlike the previous techniques that navigate with Cartesian readouts (40–42), the proposed spiral navigator serves as a better trade-off between speed, SNR and the ability to capture temporal dynamics because a) the spiral trajectory provides high gradient efficiency compared with other trajectories, b) the spiral trajectory receives high SNR due to its natural oversampling at the center of k-space, and c) the spiral trajectory has the potential to cover broader k-space within certain time constraints or slew rate limits (48, 49). For the imaging data set, we use a Cartesian trajectory to acquire imaging data with high spatial resolution. The use of Cartesian imaging data not only simplifies the image reconstruction problem, but also maintains low sensitivity to magnetic susceptibility, which can be a challenge at a number of air-to-tissue interfaces in the oropharyngeal region.

Image Reconstruction

Given the acquired navigation data, the principal component analysis (PCA) or singular value decomposition (SVD) is performed to estimate the temporal subspace, i.e., matrix **V** (36, 47). Specifically, **V** is constructed from the L most significant right singular vectors of $\hat{\mathbf{C}}$ (36). With **V** estimated, we can determine **U** from the imaging data by solving a least-squares problem. By separating the estimation of **V** and **U** in two steps, the proposed method yields a convex and significantly simplified reconstruction problem (compared with methods that simultaneously determine **U** and **V**).

Also, it should be noted that direct determination of $\hat{\mathbf{U}}$ from least-squares fitting is usually ill-conditioned, especially when a higher L is used but limited samples are available (40). In this work, we employ a spatial-spectral sparsity constraint to regularize the ill-conditioned reconstruction problem. The sparsity constraint has been found effective in existing low-rank constrained reconstruction problems (40–42). For dynamic speech MRI, in particular, this constraint is appropriate because the (**x**, f)-domain is approximately sparse, given that subjects are using similar velopharyngeal motion for different speech samples during experiments. The combination of low-rank and sparsity constraint not only yields better conditioning, but also represents a broader range of non-periodic articulatory motion than using a sparsity constraint alone. It is worth noting that incorporating other sparsity constraints can be also mathematically straightforward (44, 50).

By jointly imposing the low-rank constraint and spatial-spectral sparsity constraint, the image reconstruction problem with sensitivity-encoded (**k**, t)-space data can be formulated as

$$\hat{\mathbf{U}}=\arg\min_{\mathbf{U}\in\mathbb{C}^{N\times L}}\sum_{q=1}^{Q}\|\mathbf{\Omega}(\mathbf{F}_s\mathbf{S}_q\mathbf{U}\mathbf{V})-\mathbf{d}_q\|_2^2+\lambda\|\mathrm{vec}(\mathbf{U}\mathbf{V}_f)\|_1, \quad (4)$$

where Q denotes the number of receiver coils, $\mathbf{\Omega}$ denotes a sparse sampling operator corresponding to the acquisition of the imaging data (and putting the data in vector form), $\mathbf{F}_s$ denotes a spatial Fourier transform matrix, $\mathbf{S}_q$ denotes the sensitivity map of the qth coil, $\mathbf{d}_q$ denotes the sparsely acquired imaging data samples from the qth receiver coil, λ denotes the regularization parameter, $\mathbf{V}_f$ is obtained by applying the temporal Fourier transform to each row of **V** and vec(.) denotes an operator that forms a vector by concatenating the columns of a matrix. A numerical algorithm based on half-quadratic regularization with continuation is applied to solve Eq. 4 (40).

METHODS

Validation Experiments

Several validation experiments were performed to demonstrate the effectiveness of the proposed method. The experiments were performed on a Siemens Trio scanner (Siemens Medical Solutions, Erlangen, Germany) with a field strength of 3 T, a gradient strength of 40 mTm^{-1} and a maximum slew rate of 176 $\mathrm{Tm}^{-1}\mathrm{s}^{-1}$. A 12-channel head receiver coil and a 4-channel neck receiver coil were jointly used to image the subject. A FLASH sequence was

developed to interleave a spiral navigation acquisition and a Cartesian imaging acquisition with a repetition time (TR) of 9.78 ms. The navigation and imaging data acquisition had an echo time (TE) of 0.85 ms and 2.3 ms, respectively. Other parameters were: acquisition matrix size = 128 × 128, FOV = 280 mm × 280 mm, spatial resolution = 2.2 mm × 2.2 mm and a slice thickness of 6.5 mm.

The experiment protocol enabled three choices of acquisition setups: a) acquisition of a single mid-sagittal imaging slice at a nominal frame rate of 102.2 fps, b) four slices at 25.5 fps or c) eight slices at 12.8 fps. When acquiring data that targets a model order of around 80, as was done in this work, the acquisition time for the single slice, four-slice and eight-slice experiments were 1 min 42 s, 6 min 49 s and 13 min 38 s, respectively.

The experiment protocol allowed flexible orientations for the imaging planes of interest. For single-slice acquisitions, an imaging plane was placed on the mid-line of the tongue to capture articulatory motion in the upper airway. For multi-slice acquisition, the imaging planes can be contiguously positioned or arbitrarily oriented. Specifically, eight contiguous planes centering at the mid-line of the tongue formed an imaging volume that covered the entire upper vocal tract. Reconstructions from this imaging volume can be resliced to synthesize oblique-coronal views of the vocal tract.

Accurate estimation of the receiver coil sensitivity profiles is important for high-quality image reconstruction. Prior to the acquisition of the dynamic imaging data, a pilot scan was performed to pre-determine the sensitivity profiles. The estimated sensitivity profiles were assumed to be time-invariant for the subsequent image reconstruction.

Five volunteers participated in the validation experiments. These subjects had an age range of 23 to 45 and three subjects were female. Among these five subjects, three were requested to produce /loo/-/lee/-/la/-/za/-/na/-/za/ recurrently at their own speaking paces; Another subject was asked to produce /za/-/na/-/za/ at controlled paces: fast (around 2 samples per second), medium (around 1 sample second) and slow paces (around 0.5 sample per second); The last subject was asked to read a passage from a publicly accessible data base of English dialects (51). The passage contains standard reading text that is devoid of explicit repetitions of words or phrases. This passage was chosen to demonstrate that the proposed method does not rely on repetitions in speech samples.

During acquisition, the voice of each subject was simultaneously recorded at a sampling rate of 22 kHz through a fiber-optic microphone with active noise cancellation (Dual Channel FOMRI, Optoacoustics, Or Yehuda, Israel). The head motion of each subject was minimized by fixing the positions of the patient's head in the receiver coil with foam pads. Informed consents were obtained for all subjects and the experiment was carried out in accordance with regulations of the Institutional Review Board at the University of Illinois at Urbana-Champaign.

Application to Nasalization Studies

The proposed method was also applied to analyze the level of nasalization during the production of a speech sample. Particular emphasis in these experiments was placed on

velopharyngeal activity. With regard to velopharyngeal activity, nasalization refers to the coupling between the nasal and oral cavities due to velar movement (52). The level of velopharyngeal coupling is influenced by the fast interaction between the velum and the velopharyngeal aperture. Previous work has indicated the potential of dynamic MRI to examine the spatiotemporal variation in the aperture and provide imagery basis for phonetic assessments of nasalization level (53,54). However, analysis of velopharyngeal activity at the current frame rate has not been possible.

French was chosen as the carrier language for the studies on nasalization. Unlike the English language, French contains nasal vowel sounds, oral vowel sounds and nasalization effects during speech. Therefore, it is an especially interesting language to study in relation to the coupling between the nasal and oral cavities to produce nasalized speech sounds, such as the alveolar nasal consonant (55–57). This experiment focused on a dialect of French, the northern metropolitan French, due to its prevalence. The carrier phrase was chosen as " Il retape X parfois ", where X denotes articulation of three nasal vowels: /ɑ̃/, /ɛ̃/ or /ɔ̃/. A healthy female subject was recruited for the experiment, during which she was requested to produce the phrase recurrently. No other requirement was imposed on the subject's speaking paces. This experiment was conducted in accordance with an approved protocol through the Institutional Review Board.

Other aspects of the experiment protocol were also optimized to better capture the dynamics of nasalization. First, the slice orientations of the four slices acquired were designed to cover multiple spatial locations to study velopharyngeal activities and their interaction with the rest of the vocal tract during the production of nasal vowels. Particularly, these four imaging planes were placed from anterior to posterior regions of the vocal tract, including a coronal imaging plane across the mid-section of the tongue, an oblique mid-sagittal imaging plane across the levator veli palatini muscle and two axial imaging planes across the mediopharynx and lower pharynx. These four imaging planes were acquired in a temporally interleaved fashion with the multi-slice experiment protocol described in detail above. Second, the slice orientations were carefully selected to better visualize activities of the velum and the relationship of the timing of its movement relative to the tongue tip / tongue blade, along with changes in the pharynx. Third, the production of the speech samples were recorded with a fiber-optic microphone. The audio recording and corresponding imaging data were synchronized and jointly studied in the ensuing linguistic analysis to reveal articulatory / acoustic relations.

RESULTS

Validation Experiments

Figure 2 shows reconstructions from a single-slice experiment, in which the subject was asked to produce /loo/-/lee/-/la/-/za/-/na/-/za/ sounds recurrently. Gestures of the tongue at multiple time points are illustrated and directions of tongue motion are indicated with arrows. Specifically, Figure 2a, b and c show tongue gestures at the onset of the /l/ sound in /lee/, /loo/ and /la/ syllables, respectively. Although the same /l/ sound is produced and the tongue tip is elevated in a similar fashion, the tongue body bends toward different regions of the vocal tract. By contrast, the tongue retracts to a resting position during the production

of /a/ in the /za/ syllable in Figure 2d. The proposed method captured the above motions in find spatial details.

Figure 3 shows reconstructions from a multi-slice experiment, in which an identical speech sample was produced as above. The oblique coronal reconstructions are resliced from 8 parallel mid-sagittal slices to enable 3D visualization of articulatory motion in arbitrary through-plane directions. Figure 3a captures tongue tip retraction from the alveolar ridge towards the lower jaw during the transition from /l/ to /ee/ in the /lee/ syllable. Figure 3b exhibits the formation of velopharyngeal opening as the velum moves towards the root of the tongue. The anterior velum surface is first curved to prepare for /n/ in the /na/ syllable, and later becomes straightened to produce /a/ in the /na/ syllable. Results in Figure 3 demonstrate the feasibility to capture 3D articulatory motion with full-vocal-tract coverage.

Figure 4 shows spatial images and temporal profiles from a single-slice experiment, in which the subject was asked to produce /za/-/na/-/za/ sounds at fast, medium and slow speaking paces. Three single-pixel-wide strips from the same reconstruction are plotted versus time in Figure 4a, b and c: two head-foot strips across the lips and mid-tongue, respectively, and an anterior-posterior strip across the velum. Contacts between bulk articulators (e.g., the lips, the tongue and the velum) with the associated boundaries are captured with fine spatial and temporal details. Although temporal periodicity was purposefully disrupted in the experiment, the proposed method remained robust to non-periodic articulatory motion.

Figure 5 shows reconstructions from a single-slice experiment, in which a subject was requested to read the passage that is devoid of explicit repetitions. Representative spatial images and temporal profiles along various directions are provided in Figure 5a, b, c and d, respectively. Furthermore, a reconstruction video is also included in the supplementary material of the paper. As can be seen, the proposed method well captured the spatiotemporal dynamics of each distinct syllable in the passage. Note that although the reading passage contains no repetition of words or phrases, the proposed method still provided high-quality reconstructions.

Application to Nasalization Studies

Figure 6 shows mid-sagittal images from the nasalization study on northern metropolitan French. The proposed method provided high spatial resolution to capture the interaction between the velum and the pharyngeal wall. Diverse gestures of the velum were observed and three are given as examples. Figure 6b shows the gesture of the velum as it retracts from the pharyngeal wall to prepare for the production of the nasal sound /ɑ̃/. Upon completion of the /ɑ̃/ sound, as illustrated with Figure 6c, the velum holds a relaxed position with a larger air passage. As a comparison, Figure 6d shows the closure of the velopharyngeal port at the production of the plosive sound /t/. The results in Figure 6 validate the effectiveness of the proposed method in capturing nasalization in in vivo experiments with high spatial resolution.

Figure 7 shows oblique coronal and axial images from a multi-slice in vivo experiment. The proposed method allowed vocal tract opening to be captured along four imaging planes.

Specifically, the vocal tract opening sizes are compared for three nasal vowels, the /ɑ̃/, /ɛ̃/ and /ɔ̃/ sounds. In the superior portion of the vocal tract, as illustrated with Figure 7a and b, the /ɑ̃/ sound has the largest opening size while the /ɔ̃/ sound has the smallest. Opposite observation, as illustrated in Figure 7c, is seen in the middle portion of the vocal tract, where /ɑ/ sound has the smallest opening size and /ɔ̃/ sound has the largest. Nearly identical opening sizes are observed in the inferior portion of the vocal tract, as illustrated in Figure 7d. The results in Figure 7 demonstrate that the proposed method can provide sufficient spatial coverage to capture variation of vocal tract opening sizes at different spatial locations.

Figure 8 shows an integrated analysis enabled by synchronizing the reconstructions with acoustic recordings to fully visualize the relationship between movement of the velum, nasalization and the spectrogram of the recorded speech sample (17,58,59). For ease of visualization of velar movement, we use the average pixel intensity (API) time series that is extracted from the high-spatiotemporal-resolution reconstruction. As illustrated in Figure 8a, an API series is calculated from a region of interest (ROI) where velopharyngeal closure takes place in order to summarize the motion with one measure, while keeping the ROI fixed in place. The API measures are temporally aligned with the audio and spectral signal for comparison. Since temporal variation in API is caused by the velum motion into or out of the ROI, variation of API captures subtle spatial motion in high temporal resolution. As illustrated in Figure 8b, an increase in API is observed during the production of /t/ and /p/ sounds, where velic positions are expected to be high, i.e., the velopharyngeal port is expected to be closed. The onsets and endings of these phonetic events are confirmed as changes in spectral patterns in Figure 8d. Production of the /ɑ̃/ sound is associated with a decrease in API in Figure 8b and manifests as a spectral pattern characteristic of vowels in Figure 8d. The results in Figure 8 demonstrate that the proposed method can accurately capture the fast articulatory transitions associated with speech motion and offers high-resolution spatial image series as added information to conventional phonetic analysis.

DISCUSSION

The proposed method provides broad spatial coverage to capture speech movements across multiple imaging planes. Sufficient spatial coverage is critical for phonetic studies that investigate the relations between multiple regions in the vocal tract. For example, traditional phonetic analysis has taken for granted that the superior region of the vocal tract undergoes a greater level of motion compared with the inferior region. However, imaging evidence has been lacking. In this work, this hypothesis is verified by applying the proposed method to examine movements at different levels of the vocal tract and the results are summarized in Figure 6. As can be seen, the opening sizes of these cavities vary significantly in the superior vocal tract but remain nearly identical in the inferior portion. The results confirm the above hypothesis and may suggest a pivot-like structure of muscle motion: the superior vocal tract may be more deformable while the inferior vocal tract may be more stable. Further investigation of the physiological basis and the clinical implications of this observation is necessary.

The proposed method offers high temporal resolution for some phonetic studies on nasalization. These studies usually involve controversy over the opening of the velopharyngeal port (61), asking questions such as: what sounds can tolerate velopharyngeal opening? To address this, a spectrogram of the acoustic signal and the API time series are jointly used. Note that this analysis is valid only when the temporal frame rate of the image series is comparable to the time scale of changes in spectral properties. An example of this analysis is provided in Figure 8, where the changes in spectral patterns match well with that in the API series. In particular, a slightly increased API value is observed for /t/ and /p/ sounds, with corresponding spectral pattern changes. Phonetically, this may reflect that the body of the velum is pulled towards the pharyngeal wall enabling air pressure to build up. Upon the completion of the /t/ and /p/ sounds, a decreased API value is observed. This may suggest that air pressure is released upon completion of these sounds. Through this example, the proposed method demonstrates great potential to assist phonetic analysis that requires high temporal frame rate.

Although the proposed method has demonstrated a number of merits in retrieving practical phonetic information, several aspects of the method may interact with the quality of reconstruction. First of all, rigorous quantification of resolution is important for the proposed method, although we empirically use the nominal frame rate to describe the imaging speed in this work. The resolution for conventional linear and shift-invariant reconstruction methods can often be measured via the point spread function. However, the proposed method is a nonlinear reconstruction method, for which defining a meaningful point spread function can be nontrivial. This is because the inherent non-linearity often renders the impulse response function spatiotemporal-location-dependent and imaging-object-dependent. Although there exist some empirical ways (60) to quantify the resolution for nonlinear reconstructions, in-depth investigation is still needed in future research.

As demonstrated in Figure 5, the proposed method works well for non-periodic speech imaging tasks, although the ($\mathbf{x}$, f)-sparsity constraint is used in the problem formulation. This is mainly due to the complementary roles of the low-rank and sparsity constraints (40): the low-rank model provides strong power to represent spatiotemporal dynamics (temporal periodicity is not required), while the ($\mathbf{x}$, f)-sparsity constraint effectively regularizes the ill-conditioning issues associated with the highly undersampled temporal subspace.

The proposed acquisition strategy employs spiral-trajectory-based navigation. For single-slice experiments, this trajectory captures high-frame-rate dynamics efficiently. For multi-slice experiments, however, our current approach uses a separate navigator acquisition for each slice. This is beneficial for non-parallel slice prescriptions (as shown in Figure 7). For parallel slice prescriptions, however, this approach may reduce the available temporal information. In this case, a single, 3D navigator may be preferable. For example, we have investigated a cone-trajectory-based navigation technique (46) to improve imaging speed for parallel, contiguous multi-slice experiments. Systematic analysis of this technique is under investigation and will be presented in future work.

The proposed method requires selection of the model order L. In this work, L is chosen based on visual inspection of image quality (e.g., SNR, reconstruction artifacts and temporal

blurring). However, it should be noted that the proposed method offers relatively robust performance over the selection of L due to the sparsity regularization (40). To demonstrate this, we reconstructed an in vivo data set (identical with the one used in Figure 4) using a range of L and the results are shown in Figure 9. As can be seen, reconstruction performance is robust for L ranging from 40 to 80, although a "too-low" (e.g., $L = 20$) or "too-high" (e.g., $L = 120$) model order can lead to sub-optimal performance. Despite the above observations, it is still important to explore quantitative metrics to enable automatic selection of L. For example, some information-theoretic metrics (62, 63) could be used, although they may lead to suboptimal reconstructions (64). Also, a resolution-based metric could be an interesting alternative (60), although in-depth investigation is still needed.

The proposed method also requires the experimenter to determine the regularization parameter λ. In this work, we assign different λ for different reconstructions based on the discrepancy principle (40), which has consistently yields good empirical results in the speech imaging and other dynamic imaging applications (40, 41). Alternative methods, such as SURE-based methods (65), could also be used for selecting regular-ization parameters, which may result in better performance.

The proposed method employs pilot scans to pre-determine the sensitivity maps. Similar to many dynamic imaging applications, we assume that the estimated sensitivity maps are time-invariant in reconstruction. However, it is possible to use the proposed method with time-varying sensitivity maps (66), which may lead to improved performance. Systematic exploration of this extension will be carried out in future research.

The proposed method may pose a computational burden for some research experiments and clinical applications due to the large-scale optimization problem involved. For instance, computation time for reconstruction of a single mid-sagittal slice for 10404 time frames at 102 fps from a 16-channel receiver coil was around 34 min on a 24-core SUN workstation without code optimization. It is worth noting that a decoupled computational structure (40) may potentially improve the computational efficiency. In addition, iterative image reconstruction algorithms on graphical processing units (GPUs) have already been optimized for structural MRI image reconstruction and achieved an acceleration factor of up to 150 folds (67). The proposed method may benefit from similar implementation leveraging the massively multi-core power of the GPUs, although considerations on computational efficiency and algorithmic optimization are beyond the scope of this work.

CONCLUSIONS

A new method has been proposed for dynamic speech imaging with high spatiotemporal resolution and broad spatial coverage. The proposed method is characterized by a) an acquisition strategy based on spiral navigators, and b) an image reconstruction method based on joint low-rank and sparsity constraints. The proposed method has been validated in speech imaging experiments, achieving a nominal imaging speed of 102 fps with a spatial resolution of $2.2 \times 2.2 \times 6.5\ mm^3$ for a single-slice imaging protocol, and a nominal imaging speed of 12.8 fps with the identical spatial resolution for an eight-slice protocol. The proposed method has demonstrated potential in assisting phonetic analysis on nasalization.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

This work was partially supported by NIH-1R03DC009676-01A1, NSF-1121780 and a Computational Science and Engineering fellowship (M. Fu) from the University of Illinois at Urbana-Champaign. The authors would also like to acknowledge Aaron Johnson for providing a reading passage for an in vivo experiment.

References

1. Ventura SMR, Freitas DRS, Tavares JMR. Toward dynamic magnetic resonance imaging of the vocal tract during speech production. J Voice. 2011; 25:511–518. [PubMed: 20471801]
2. Ettema SL, Kuehn DP, Perlman AL, Alperin N. Magnetic resonance imaging of the levator veli palatini muscle during speech. Cleft Palate Craniofac J. 2002; 39:130–144. [PubMed: 11879068]
3. Echternach M, Markl M, Richter B. Dynamic real-time magnetic resonance imaging for the analysis of voice physiology. Curr Opin Otolaryngol Head Neck Surg. 2012; 20:450–457. [PubMed: 23086261]
4. Sundberg J. Articulatory configuration and pitch in a classically trained soprano singer. J Voice. 2009; 23:546–551. [PubMed: 18504111]
5. Proctor M, Bresch E, Byrd D, Nayak K, Narayanan S. Paralinguistic mechanisms of production in human beatboxing: A real-time magnetic resonance imaging study. J Acoust Soc Am. 2013; 133:1043–1054. [PubMed: 23363120]
6. Vijay Kumar K, Shankar V, Santosham R. Assessment of swallowing and its disorders: A dynamic MRI study. Eur J Radiol. 2012; 82:215–219. [PubMed: 23068561]
7. Zhang S, Olthoff A, Frahm J. Real-time magnetic resonance imaging of normal swallowing. J Magn Reson Imaging. 2012; 35:1372–1379. [PubMed: 22271426]
8. Shinagawa H, Ono T, Honda E-I, Masaki S, Shimada Y, Fujimoto I, Sasaki T, Iriki A, Ohyama K. Dynamic analysis of articulatory movement using magnetic resonance imaging movies: methods and implications in cleft lip and palate. Cleft Palate Craniofac J. 2005; 42:225–230. [PubMed: 15865454]
9. Atik B, Bekerecioglu M, Tan O, Etlik O, Davran R, Arslan H. "Evaluation of dynamic magnetic resonance imaging in assessing velopharyngeal insufficiency during phonation. J Craniofac Surg. 2008; 19:566–572. [PubMed: 18520366]
10. Wein BB, Drobnitzky M, Klajman S, Angerstein W. Evaluation of functional positions of tongue and soft palate with MR imaging: initial clinical results. J Magn Reson Imaging. 1991; 1:381–383. [PubMed: 1802152]
11. Suto Y, Matsuo T, Kato T, Hori I, Inoue Y, Ogawa S, Suzuki T, Yamada M, Ohta Y. Evaluation of the pharyngeal airway in patients with sleep apnea: value of ultrafast MR imaging. Am J Roentgenol. 1993; 160:311–314. [PubMed: 8424340]
12. Bae Y, Kuehn DP, Conway CA, Sutton BP. Real-time magnetic resonance imaging of velopharyngeal activities with simultaneous speech recordings. Cleft Palate Craniofac J. 2011; 48:695–707. [PubMed: 21214321]
13. Niebergall A, Zhang S, Kunay E, Keydana G, Job M, Uecker M, Frahm J. Real-time MRI of speaking at a resolution of 33 ms: Undersampled radial FLASH with nonlinear inverse reconstruction. Magn Reson Med. 2013; 69:477–485. [PubMed: 22498911]
14. Uecker M, Zhang S, Voit D, Karaus A, Merboldt K-D, Frahm J. Real-time MRI at a resolution of 20 ms. NMR Biomed. 2010; 23:986–994. [PubMed: 20799371]
15. Stone M, Davis EP, Douglas AS, NessAiver M, Gullapalli R, Levine WS, Lundberg A. Modeling the motion of the internal tongue from tagged cine-MRI images. J Acoust Soc Am. 2001; 109:2974–2982. [PubMed: 11425139]
16. Kim Y-C, Proctor MI, Narayanan SS, Nayak KS. Improved imaging of lingual articulation using real-time multislice MRI. J Magn Reson Imag. 2012; 35:943–948.

17. Shosted R, Fu M, Benmamoun A, Liang Z-P, Sutton BP. Using partially separable functions to image spatiotemporal aspects of Arabic pharyngealization. J Acoust Soc Am. 2012; 132:2091.
18. Sutton BP, Conway CA, Bae Y, Seethamraju R, Kuehn DP. Faster dynamic imaging of speech with field inhomogeneity corrected spiral fast low angle shot (FLASH) at 3 T. J Magn Reson Imaging. 2010; 32:1228–1237. [PubMed: 21031529]
19. Scott AD, Boubertakh R, Birch MJ, Miquel ME. Adaptive averaging applied to dynamic imaging of the soft palate. Magn Reson Med. 2013; 70:865–874. [PubMed: 23023822]
20. Li H, Haltmeier M, Zhang S, Frahm J, Munk A. Aggregated motion estimation for image reconstruction in real-time MRI. arXiv preprint. 2013; 1304:5054. arXiv.
21. Uecker M, Zhang S, Voit D, Merboldt K-D, Frahm J. Real-time MRI: recent advances using radial FLASH. Imag Med. 2012; 4:461–476.
22. Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P. SENSE: sensitivity encoding for fast MRI. Magn Reson Med. 1999; 42:952–962. [PubMed: 10542355]
23. Griswold MA, Jakob PM, Heidemann RM, Nittka M, Jellus V, Wang J, Kiefer B, Haase A. Generalized autocalibrating partially parallel acquisitions (GRAPPA). Magn Reson Med. 2002; 47:1202–1210. [PubMed: 12111967]
24. Tsao J, Boesiger P, Pruessmann KP. k-t BLAST and k-t SENSE: Dynamic MRI with high frame rate exploiting spatiotemporal correlations. Magn Reson Med. 2003; 50:1031–1042. [PubMed: 14587014]
25. Xu D, King KF, Liang Z-P. Improving k-t SENSE by adaptive regularization. Magn Reson Med. 2007; 57:918–930. [PubMed: 17457871]
26. Aggarwal N, Bresler Y. Patient-adapted reconstruction and acquisition dynamic imaging method (PARADIGM) for MRI. Inverse Probl. 2008; 24:045015.
27. Mistretta C, Wieben O, Velikina J, Block W, Perry J, Wu Y, Johnson K. Highly constrained backprojection for time-resolved MRI. Magn Reson Med. 2006; 55:30–40. [PubMed: 16342275]
28. Lustig, M.; Santos, JM.; Donoho, DL.; Pauly, JM. k-t SPARSE: High framerate dynamic MRI exploiting spatio-temporal sparsity; Seattle, USA. In Proceedings of the 14th Annual Meeting of ISMRM; 2006. p. 2420
29. Gamper U, Boesiger P, Kozerke S. Compressed sensing in dynamic MRI. Magn Reson Med. 2008; 59:365–373. [PubMed: 18228595]
30. Trzasko JD, Haider CR, Borisch EA, Campeau NG, Glockner JF, Riederer SJ, Mand-uca A. Sparse-CAPR: Highly accelerated 4D CE-MRA with parallel imaging and nonconvex compressive sensing. Magn Reson Med. 2011; 66:1019–1032. [PubMed: 21608028]
31. Jung H, Sung K, Nayak KS, Kim E, Ye J. k-t FOCUSS: A general compressed sensing framework for high resolution dynamic MRI. Magn Reson Med. 2009; 61:103–116. [PubMed: 19097216]
32. Bresch E, Kim Y-C, Nayak K, Byrd D, Narayanan S. Seeing speech: Capturing vocal tract shaping using real-time magnetic resonance imaging. IEEE Signal Proc Mag. 2008; 25:123–132.
33. Huang F, Lin W, Duensing GR, Reykowski A. k-t sparse GROWL: Sequential combination of partially parallel imaging and compressed sensing in k-t space using flexible virtual coil. Magn Reson Med. 2012; 68:772–782. [PubMed: 22162191]
34. Liang D, DiBella EVR, Chen RR, Ying L. k-t ISD: Dynamic cardiac MR imaging using compressed sensing with iterative support detection. Magn Reson Med. 2012; 68:41–53. [PubMed: 22113706]
35. Usman M, Prieto C, Schaeffter T, Batchelor PG. k-t group sparse: A method for accelerating dynamic MRI. Magn Reson Med. 2011; 66:1163–1176. [PubMed: 21394781]
36. Liang, Z-P. Proceedings of IEEE International Symposium on Biomedical Imaging. Washington D.C., USA: 2007. Spatiotemporal imaging with partially separable functions; p. 988-991.
37. Haldar, JP.; Liang, Z-P. Proceedings of IEEE International Symposium on Biomedical Imaging. Rotterdam, The Netherlands: 2010. Spatiotemporal imaging with partially separable functions: a matrix recovery approach; p. 716-719.
38. Zhao, B.; Haldar, JP.; Brinegar, C.; Liang, Z-P. Proceedings of IEEE International Symposium on Biomedical Imaging. Rotterdam, The Netherlands: 2010. Low rank matrix recovery for real-time cardiac MRI; p. 996-999.

39. Lingala SG, Hu Y, DiBella E, Jacob M. Accelerated dynamic MRI exploiting sparsity and low-rank structure: k-t SLR. IEEE Trans Med Imaging. 2011; 30(5):1042–1054. [PubMed: 21292593]
40. Zhao B, Haldar JP, Christodoulou AG, Liang Z-P. Image reconstruction from highly undersampled (k, t)-space data with joint partial separability and sparsity constraints. IEEE Trans Med Imaging. 2012; 31:1809–1820. [PubMed: 22695345]
41. Christodoulou A, Zhang H, Zhao B, Hitchens T, Ho C, Liang Z-P. High-resolution cardiovascular MRI by integrating parallel imaging with low-rank and sparse modeling. IEEE Trans Biomed Eng. 2013
42. Brinegar C, Schmitter SS, Mistry NN, Johnson GA, Liang Z-P. Improving temporal resolution of pulmonary perfusion imaging in rats using the partially separable functions model. Magn Reson Med. 2010; 64:1162–1170. [PubMed: 20564601]
43. Christodoulou, AG.; Zhao, B.; Zhang, H.; Ho, C.; Liang, Z-P. Proceedings of IEEE International Symposium on Biomedical Imaging. Chicago, USA: 2011. Four-dimensional MR cardiovascular imaging: Method and applications; p. 3732-3735.
44. Zhao, B.; Lu, W.; Liang, Z-P. Highly accelerated parameter mapping with joint partial separability and sparsity constraints; Melbourne, Australia. In Proceedings of the 20th Annual Meeting of ISMRM; 2012. p. 2233
45. Fu, M.; Christodoulou, AG.; Naber, AT.; Kuehn, DP.; Liang, Z-P.; Sutton, BP. High-frame-rate multi-slice speech imaging with sparse sampling of (k, t)-space; Melbourne, Australia. In Proceedings of the 20th Annual Meeting of ISMRM; 2012. p. 12
46. Fu, M.; Zhao, B.; Holtrop, JL.; Kuehn, DP.; Liang, Z-P.; Sutton, BP. “High-frame-rate full-vocal-tract imaging based on the partial separability model and volumetric navigation; Salt Lake City, USA. In Proceedings of the 21st Annual Meeting of ISMRM; 2013. p. 4269
47. Sen Gupta, A.; Liang, Z-P. Dynamic imaging by temporal modeling with principal component analysis; Glasgow, Scotland. In Proceedings of the 9th Annual Meeting of ISMRM; 2001. p. 102001
48. Adalsteinsson E, Irarrazabal P, Topp S, Meyer C, Macovski A, Spielman DM. Volumetric spectroscopic imaging with spiral-based k-space trajectories. Magn Reson Med. 2005; 39:889–898. [PubMed: 9621912]
49. Blasco, MS.; Krishnan, S.; Moratal, D.; Ramamurthy, S.; Brummer, ME. High spatial frequencies are more dynamic than low spatial frequencies in cardiac motion; Hawaii, USA. In Proceedings of the 17th Annual Meeting of ISMRM; 2009. p. 1425-1434.
50. Zhao, B.; Haldar, JP.; Christodoulou, AG.; Liang, Z-P. Proceedings of IEEE International Symposium on Biomedical Imaging. Chicago, USA: 2011. Further development of image reconstruction from highly undersampled (k, t)-space data with joint partial separability and sparsity constraints; p. 1593-1596.
51. The Rainbow Passage. [Access: March 23, 2014] Feb 1. http://www.dialectsarchive.com/the-rainbow-passage. Published
52. Kuehn DP. A cineradiography investigation of velar movement variables in two normals. Cleft Palate J. 1976; 13:88–103. [PubMed: 1062249]
53. Byrd D, Tobin S, Bresch E, Narayanan S. Timing effects of syllable structure and stress on nasals: a real-time MRI examination. J Phon. 2009; 37:97–110. [PubMed: 20046892]
54. Perry, JL.; Sutton, BP.; Kuehn, DP.; Gamage, JK. Using MRI for assessing velopharyngeal structures and function. Cleft Palate Craniofac J. 2013. http://dx.doi.org/10.1597/12-083
55. Fougeron, C.; Smith, C. Handbook of the International Phonetic Association. Cambridge: Cambridge University Press; 1999.
56. Cohn, AC. Ph.D. dissertation. Los Angeles: University of California; 1990. Phonetic and phonological rules of nasalization.
57. Carignan, C.; Shosted, R.; Fu, M.; Liang, Z-P.; Sutton, B. The role of the pharynx and tongue in enhancement of vowel nasalization: A real-time MRI investigation of French nasal vowels; Lyon, France. In Proceedings of the 14th Annual Conference of the International Speech Communication Association; 2013.
58. Proctor, MI.; Lammert, AC.; Katsamanis, A.; Goldstein, LM.; Hagedorn, C.; Narayanan, SS. Direct estimation of articulatory kinematics from real-time magnetic resonance image sequences;

Florence, Italy. In Proceedings of the 12th Annual Conference of the International Speech Communication Association; 2011. p. 281-284.

59. Shosted, R.; Sutton, B.; Benmamoun, A. Using magnetic resonance to image the pharynx during Arabic speech: Static and dynamic aspects; Portland, USA. In Proceedings of the 13th Annual Conference of the International Speech Communication Association; 2012. p. 2182-2185.
60. Wech T, Stab D, Budich JC, Fischer A, Tran-Gia J, Hahn D, Kostler H. Resolution evaluation of MR images reconstructed by iterative thresholding algorithms for compressed sensing. Med Phys. 2012; 39:4328–4338. [PubMed: 22830766]
61. Ohala JJ, Ohala M. The phonetics of nasal phonology: theorems and data. Nasals, Nasalization, the Velum. 1993; 5:225–249.
62. Stoica P, Yngve S. Model-order selection: a review of information criterion rules. IEEE Signal Process Mag. 2004; 21:36–47.
63. Bumham KP, Anderson DR. Model selection and multimodel inference: a practical information-theoretic approach. Springer. 2002
64. Haldar JP. Constrained imaging: denoising and sparse sampling. Doctoral dissertation, University of Illinois at Urbana-Champaign. 2011
65. Ramani S, Liu Z, Rosen J, Nielsen J, Fessler JA. Regularization parameter selection for nonlinear iterative image restoration and MRI reconstruction using GCV and SURE-based methods. IEEE Trans Image Process. 2012; 21:3659–3672. [PubMed: 22531764]
66. Kellman P, Epstein FH, McVeigh ER. Adaptive sensitivity encoding incorporating temporal filtering (TESENSE). Magn Reson Med. 2001; 45:846–852. [PubMed: 11323811]
67. Gai J, Obeid N, Holtrop JL, Wu X-L, Lam F, Fu M, Haldar JP, Hwu W, Liang Z-P, Sutton BP. More IMPATIENT: A gridding-accelerated toeplitz-based strategy for non-Cartesian high-resolution 3D MRI on GPUs. J Parallel Distrib Comput. 2013; 73:686–697. [PubMed: 23682203]

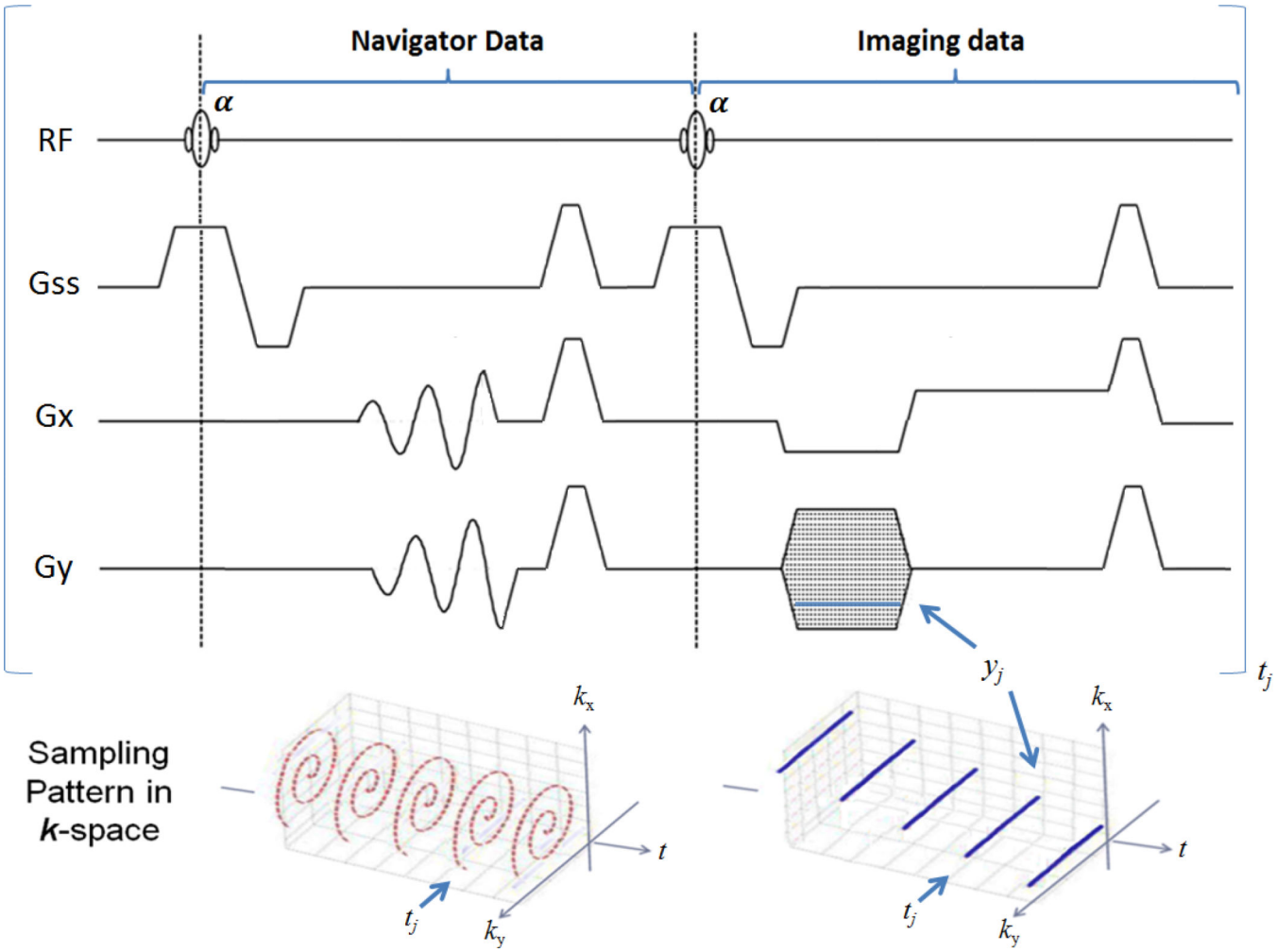

Figure 1.

A simplified pulse sequence diagram for the proposed PS model-based data acquisition strategy with illustration of (**k**, *t*)-space sampling patterns. The navigator data set is acquired using a spiral trajectory. The imaging data set is acquired using a Cartesian trajectory with random phase encoding.

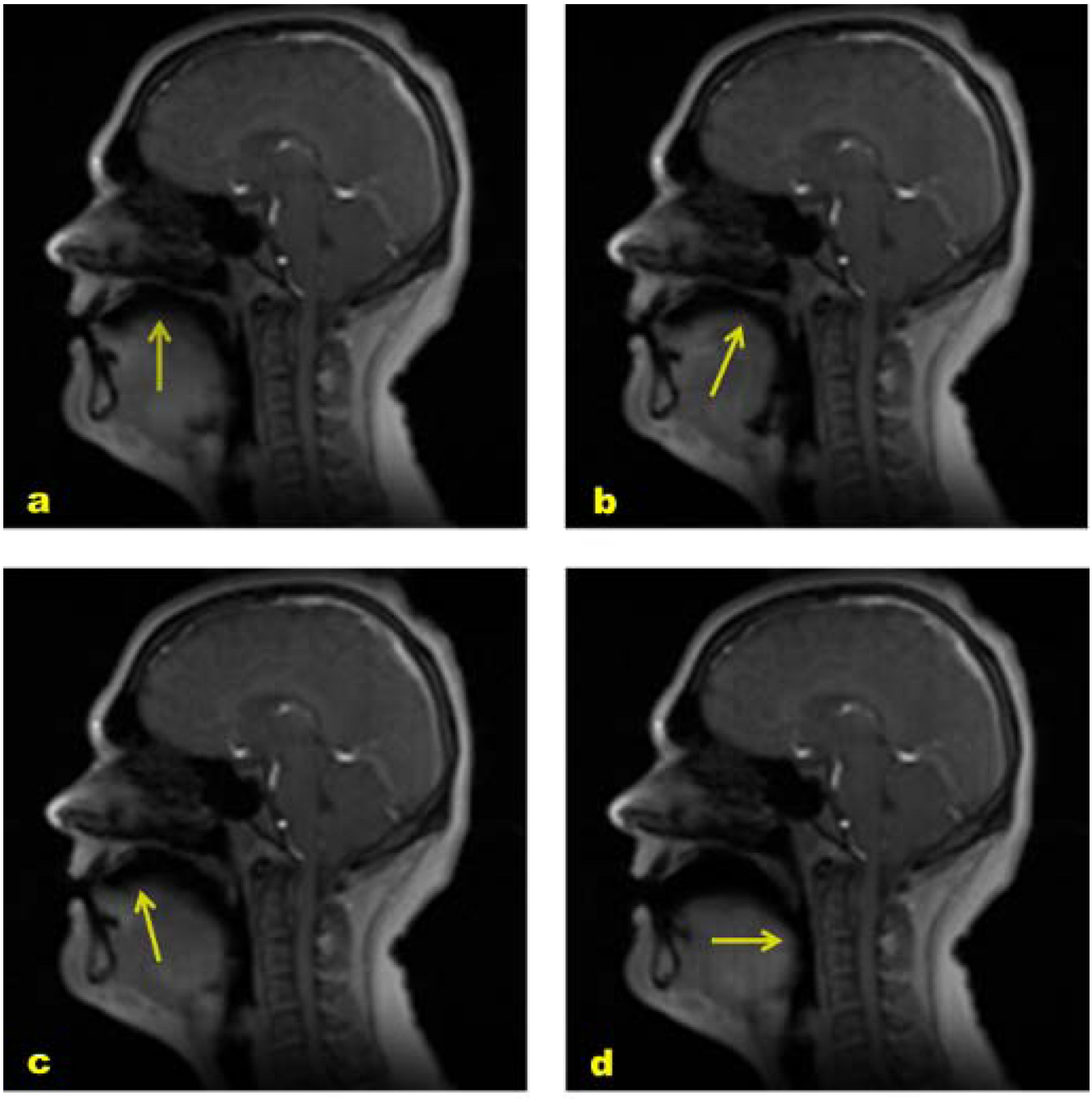

Figure 2.

Mid-sagittal reconstructions of the upper vocal tract during the production of /loo/-/lee/-/la/-/za/-/na/-/za/ syllables. The directions of the movement of the mid-tongue are indicated with arrows during (a) /l/ of the /lee/ syllable; (b) /l/ of the /loo/ syllable; (c) /l/ of the /la/ syllable; (d) /a/ of the /za/ syllable.

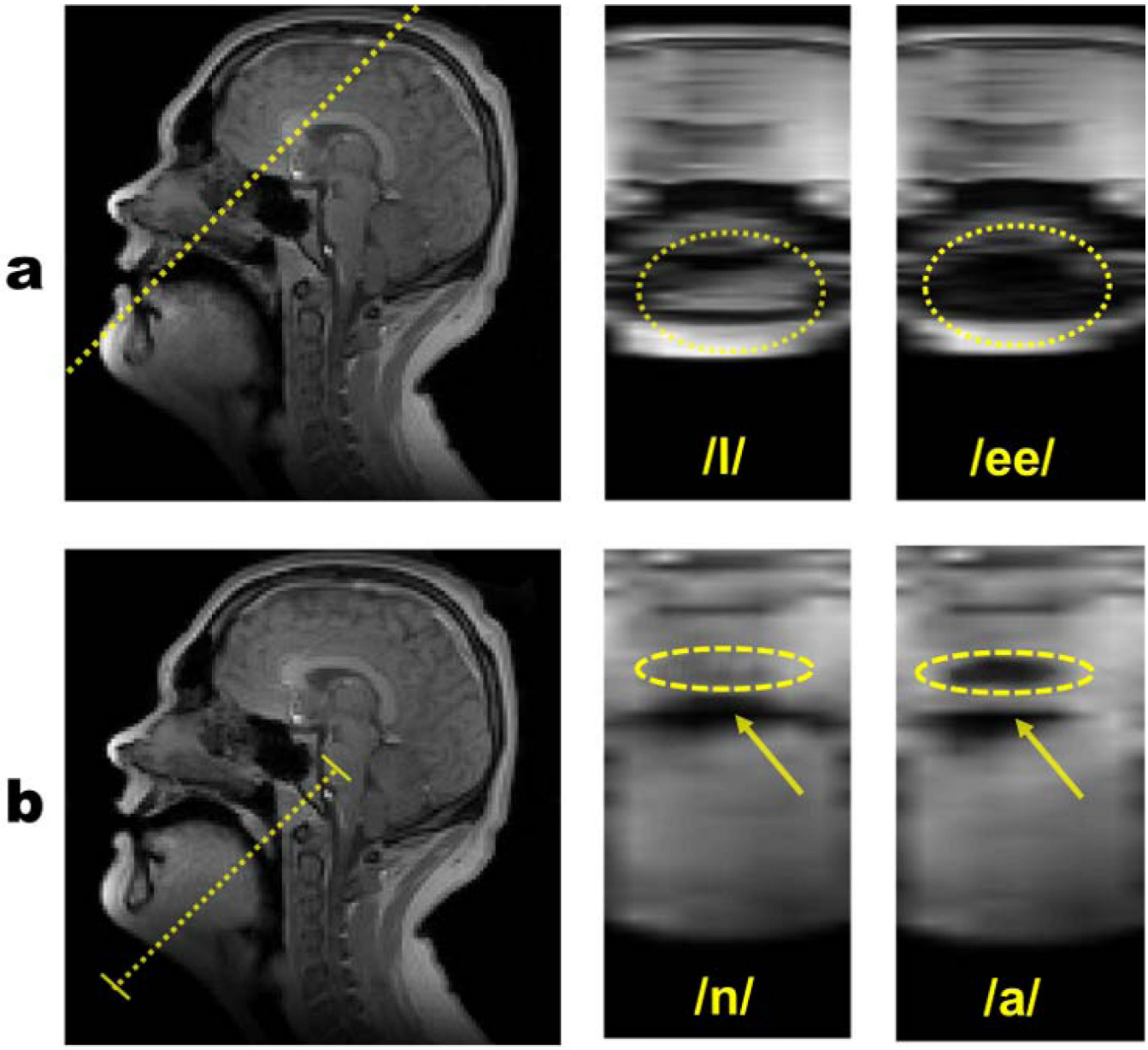

Figure 3.

Multi-slice mid-sagittal reconstructions covering the entire vocal tract. The left column shows the positions and orientations of the resliced oblique-coronal planes. The middle and right columns show resliced images. 3D articulatory motion is observed during the production of /loo/-/lee/-/la/-/za/-/na/-/za/ syllables: (a) an oblique plane across the lower incisor teeth and the alveolar ridge; (b) an oblique plane across the body of the lower jaw and the velopharyngeal closure point.

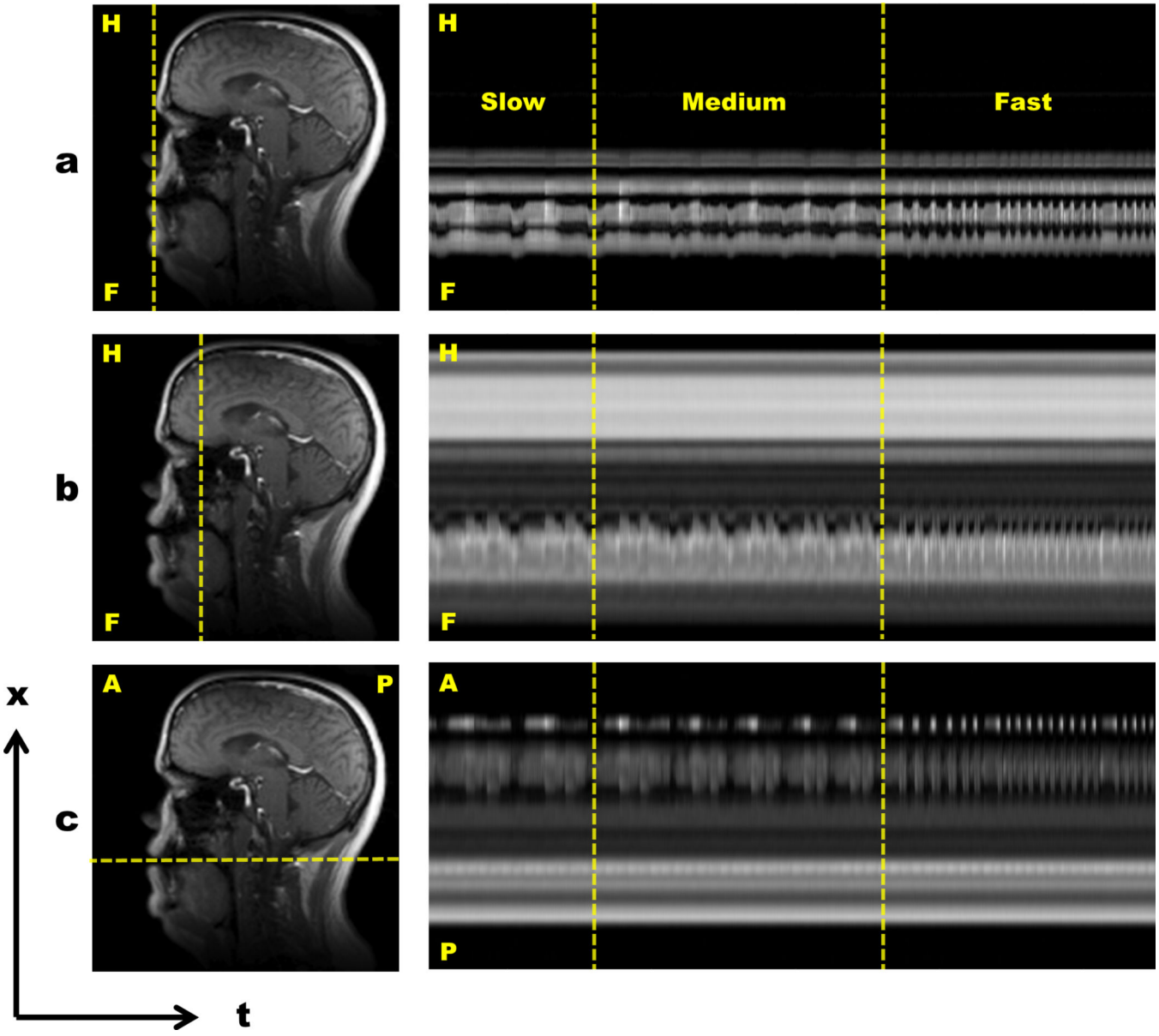

Figure 4.

Strip plots of the production of /za/-/na/-/za/ syllables at fast, medium and slow speaking paces. These strip plot demonstrate the temporal dynamics along reference lines that are taken as: (a) a vertical line across the upper and lower lip; (b) a vertical line across the roof of the mouth and the mid-tongue; (c) a horizontal line across the bottom of the upper lip and the upper pharyngeal wall.

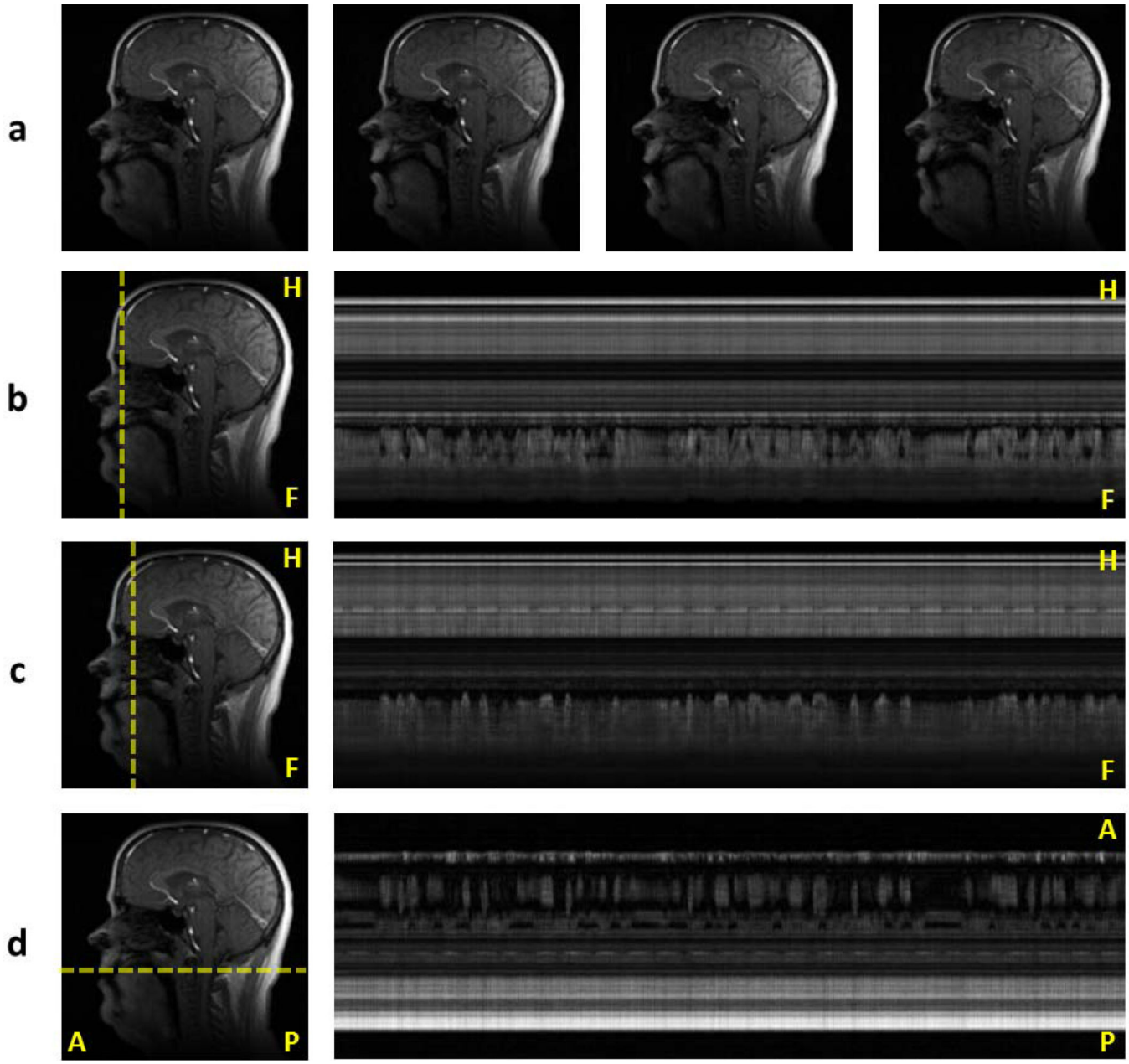

Figure 5.

Mid-sagittal reconstructions and strip plots of the production of a reading passage that contains no repetitions of words or phrases: (a) representative articulatory motion at four different time instances; (b) temporal profile taken along a vertical strip across the tongue tip; (c) temporal profile taken along a vertical strip across the mid-tongue; (d) temporal profile taken along a horizontal strip across the velum.

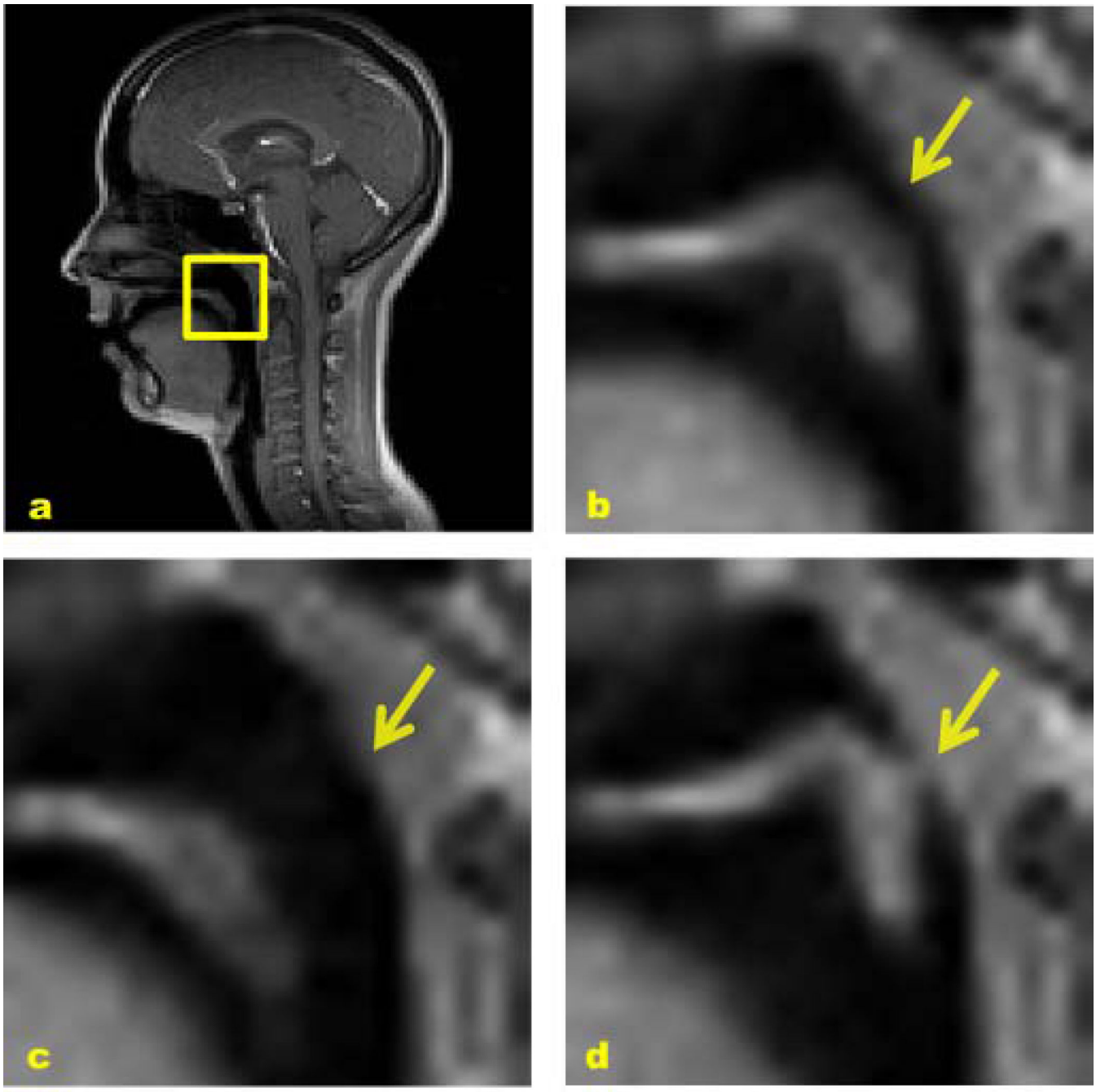

Figure 6.

Velar movement within a 18 pixel × 18 pixel region of interest as illustrated in (a). (b) An air passage is formed between the velum body and the pharyngeal wall during the production of the nasal vowel /ɑ̃/. (c) The relaxed velum creates maximum velopharyngeal opening during the breathing period. (d) The velum seals the velopharyngeal port during the production of the plosive /*t*/.

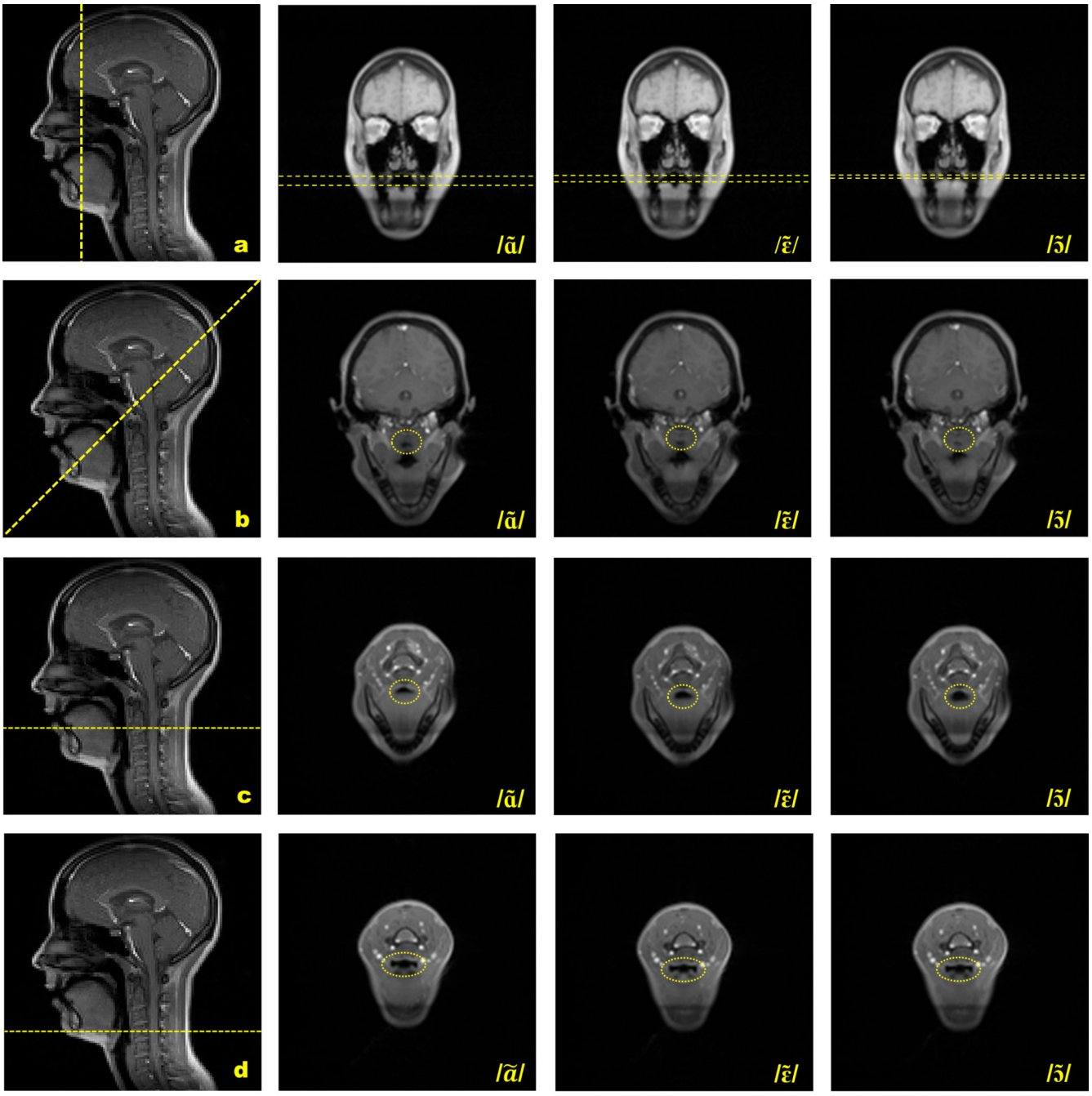

Figure 7.

Comparison of oblique coronal reconstructions for nasal vowels /ɑ̃/, /ɛ̃/ and /ɔ̃/. (a) /ɑ̃/ has the largest distance between the median portion of the tongue and the palate. (b) /ɑ̃/ has largest velopharyngeal opening size. (c) /ɑ̃/ has the smallest opening between the root of the tongue and the pharynx. (d) Three vowels have nearly identical opening between the epiglottis and the pharyngeal wall.

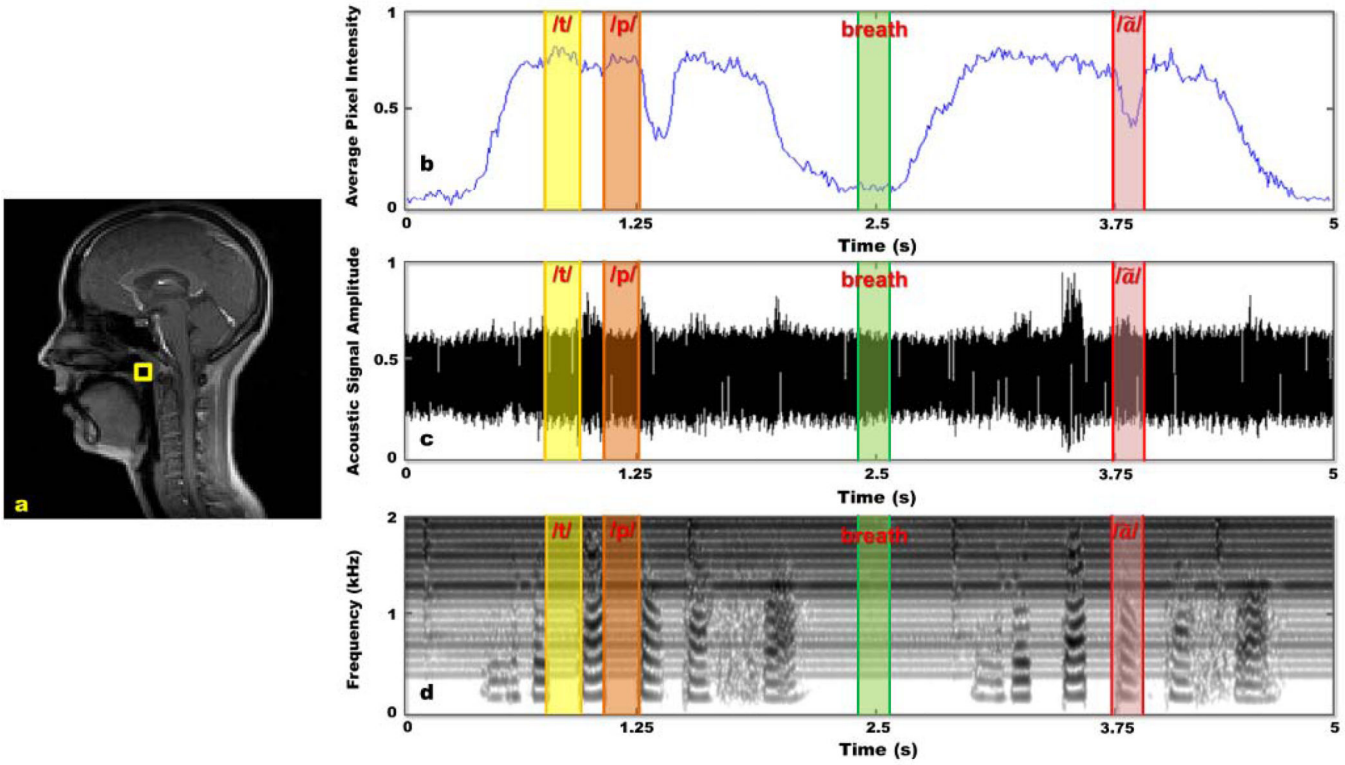

Figure 8.

Linguistic analysis integrating imaging information with acoustic properties. Colored rectangles represent temporal windows corresponding to the production of different sounds. (a) Illustration of a 5 pixel × 5 pixel region of interest where the average image intensity (API) in (b) is calculated. (b) Temporal evolution of API in the square region of (a). (c) The recorded acoustic signal. (d) The spectrogram of the acoustic signal calculated with an window width of 10 ms.

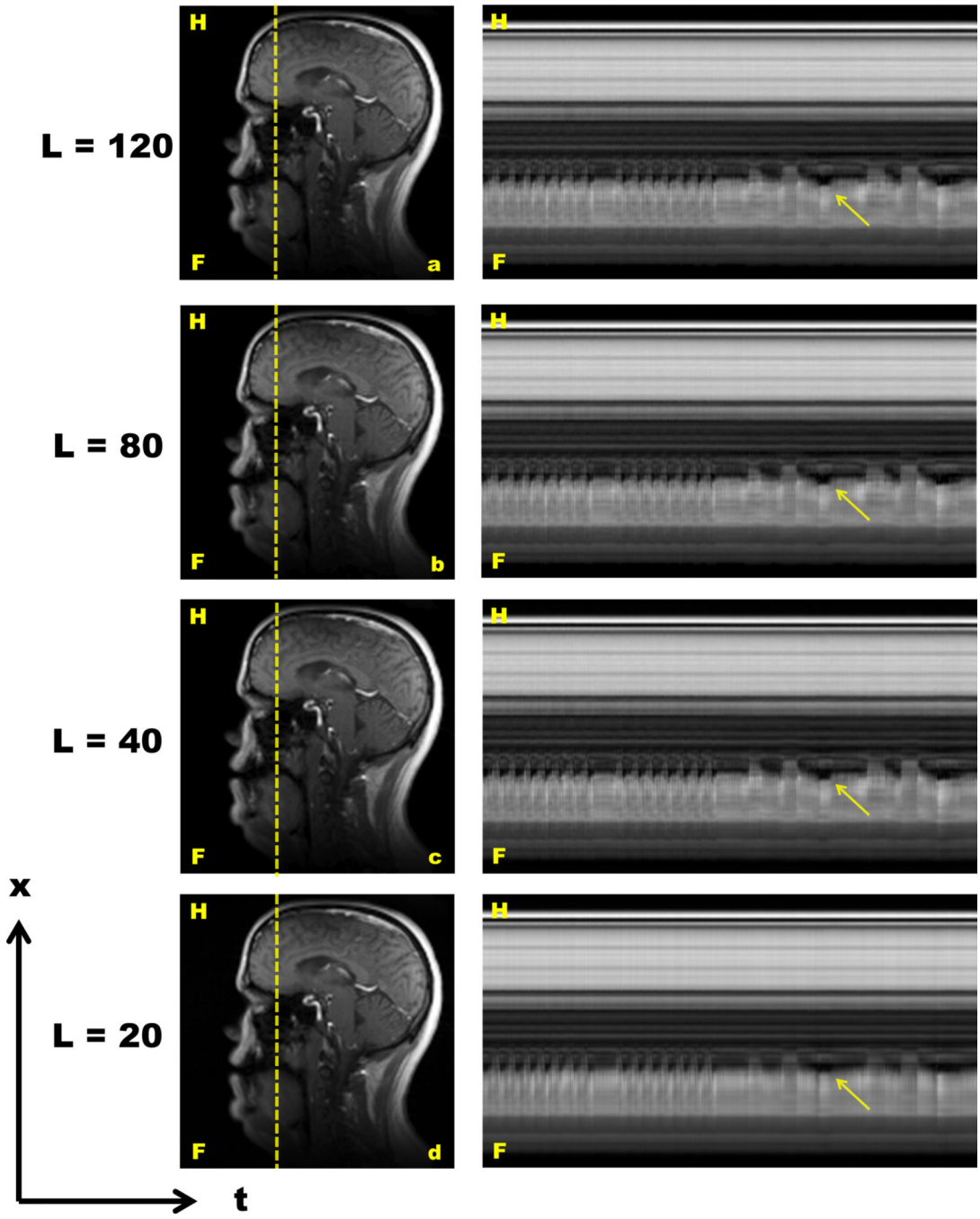

Figure 9.

Reconstruction of an experiment data set using different model orders. The first column shows representative mid-sagittal reconstructions with model orders: a) $L = 20$, b) $L = 40$, c) $L = 80$ and d) $L = 120$. The second column shows corresponding strip plots taken from a vertical line across the roof of the mouth and the mid-tongue. Variations of spatiotemporal dynamics on the strip plots are indicated by arrows.