Acoustics'08 Paris
June 29-July 4, 2008
www.acoustics08-paris.org

A Method of Co-registering Multiple Magnetic Resonance Imaged Vocal Tract Volumes for Fricatives

Michael Proctor, Christine Shadle and Khalil Iskarous

Haskins Laboratories, 300 George St, New Haven, CT 06511, USA
proctor@haskins.yale.edu

In a study of fricative production, Magnetic Resonance Imaging was used to investigate three-dimensional tract geometries of fricatives produced in different vowel contexts. Scan sequences were designed to minimize the number of elicitations required, to reduce acquisition times, and to constrain the imaging of a given phone to a single session. A new method of assembling vocal tract volumes was developed to consolidate the data obtained using this approach.

Sagittal, axial and oblique-coronal imaging orientations were chosen to best resolve the tract in the midsagittal plane, the pharynx, and around the fricative constriction. Each fricative token was acquired using a 36 second imaging sequence. Three corresponding point clouds were created by sampling air-tissue boundaries in each image stack, and aligned using anatomical landmarks on the face and spinal column, creating a single, multiply-sampled volume. A composite tract model was developed, using the most suitable data source at each region of the tract.

Vocal tract models of eight English fricatives, produced in three vowel contexts, were constructed in this manner. The models have provided important insights into the articulation of the constriction and tongue grooving, as well as differences in the production of voiced/voiceless fricatives.

Introduction

The construction of accurate three-dimensional models of the vocal tract is an important analysis technique in the study of fricatives. A proper understanding of the mechanisms of production of fricatives requires detailed knowledge of the geometric configuration of the entire tract, from the glottis to the lips.

1 Modeling the Tract Using MRI

Although Magnetic Resonance Imaging (MRI) can provide detailed information about localized regions of the vocal tract [1], the creation of accurate holistic models from disparate sets of MRI images presents significant problems. In order to unify sets of images acquired from different scans and multiple regions of the tract, existing techniques have generally involved the imposition of an external coordinate system [2], or abstraction of the data using finite element or principle component analysis [2, 3].

Other major factors restricting the utility of MRI in the study of speech are the inherently long acquisition times, and the high numbers of scans typically required to image the entire tract. As a result of these difficulties, MRI studies of fricatives have usually involved restricted corpora, such as a single vocalic context [1], or else they have relied on restricted sets of imaging orientations [4].

The modeling technique we have been developing represents an attempt to overcome some of these limitations. By combining information sourced from multiple imaging orientations, an integrated tract model can be constructed. Three superimposed point clouds, each of which contributes superior resolution of a different region, collectively constitute an oversampled model of the tract from which detailed quantification of the tract geometry can be obtained.

Most importantly, this technique does not involve any normalization of the data, nor the imposition of an external system of reference – all alignment is performed with respect to the original coordination of the MRI data itself.

2 Image Acquisition

Vocal tract models were constructed from tokens of the English oral fricatives [f-v, θ-ð, s-z, ʃ-ʒ], each elicited in three maximally-distributed vocalic contexts [i-a-u]. Five mono-lingual native speakers of Standard American English, three women (W1-W3) and two men (M1, M2), were recruited as subjects. All were students of linguistics, aged between 21 and 26 years, who were paid for their participation. Non-naïve subjects were deliberately chosen so that stimuli could be presented in IPA, and the subjects could be instructed about the linguistic requirements of each task. Subjects were instructed to concentrate on maintaining a consistent vocalic context throughout, to sustain even frication throughout the production, and to ensure that they pronounced the vowel at the end of the fricative.

2.1 MRI Sequences

A Siemens Sonata 1.5T MRI Scanner was used to image the subjects' vocal tracts while they produced all fricative tokens over the course of two 90 minute sessions. Subjects lay supine in the scanner, sustaining each fricative in each vocalic context for 36 seconds. Prompts were presented in IPA, projected onto a screen which could be read by the subject from within the scanner bore. Some subjects interrupted the sustained frication in order to take one breath; they were instructed to do so with a minimum of oral movement to reduce image blur.

All fricatives elicited during the scanning session were monitored using a Phone-Or ceramic noise-cancelling microphone system integrated into the scanner so that the veracity of each fricative and vowel could be verified.

A 2D True-FISP scan sequence (T_r=200ms, T_e=3.3ms, Flip Angle=70°) was chosen as the best compromise between image resolution and scan time. Each token was repeated three times, and imaged in one of three orientations: sagittally (from ear to ear), axially (upper-trachea to nasal cavity), and obliquely, 45° to axial planes, providing cross-sectional imaging of the tract in the alveolar region (lips to velum). In each orientation, slices of 4mm thickness were used, spaced at 4.8mm intervals. The number of slices acquired for each subject and each orientation was varied in order to sample the entire tract over the region of interest.

3 Model Construction

3.1 Segmentation

Initial processing of images was performed using Able 3D-Doctor software [5]. Image stacks were assembled from DICOM files in each orientation: sagittal, axial and oblique. Contrasts were adjusted to enhance air-tissue boundaries. Tracts were auto-segmented where possible, then adjusted manually, resulting in a set of parallel boundaries representing the intersection of the tract with the center of each imaging plane (**Fig. 1**).

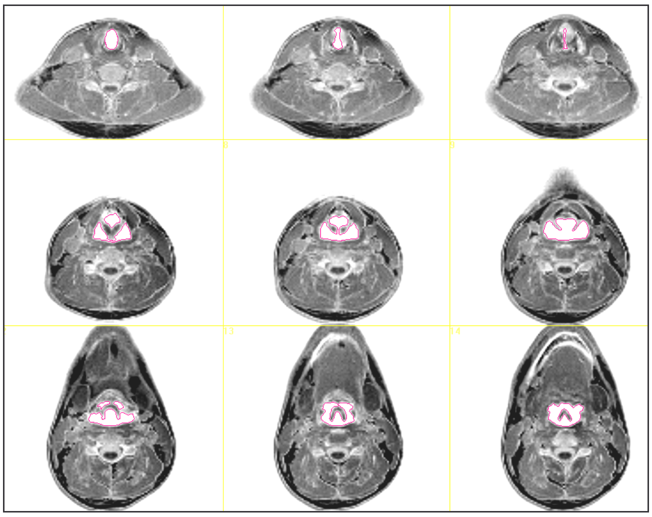


Fig.1 Segmentation of Axial MRI Images: Subj. M1 [asa] (3 tracheal, 3 epiglottal, 3 pharyngeal slices, from complete set of 30 slices covering upper trachea to nasal cavity)

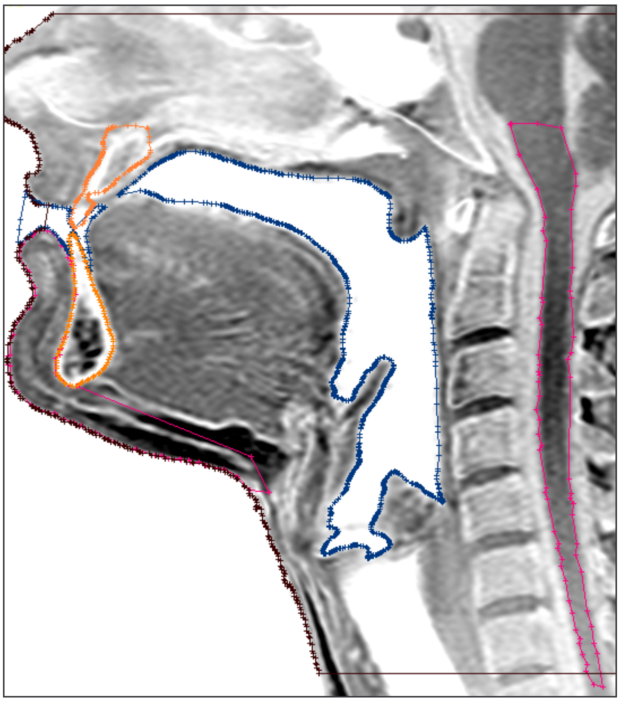


Fig.2 Segmentation of Sagittal Images: Tract, Dentition and Anatomical Landmarks: Subject M1 [asa]

Dental boundaries were identified in sagittal images where soft tissue enclosed the dentition, and superimposed on other sagittal images where necessary. Subject's faces, spinal cords and chins were segmented separately in each image stack to provide anatomical landmarks (**Fig. 2**).

3.2 Alignment

Stacks of boundaries created in this manner in 3D-Doctor were imported into Matlab as sets of x-y-z coordinates, and reassembled as point clouds representing tissue surfaces sampled in three different orientations. The set of points corresponding to one of these imaging orientations – the sagittal acquisition – is shown in **Figure 3**.

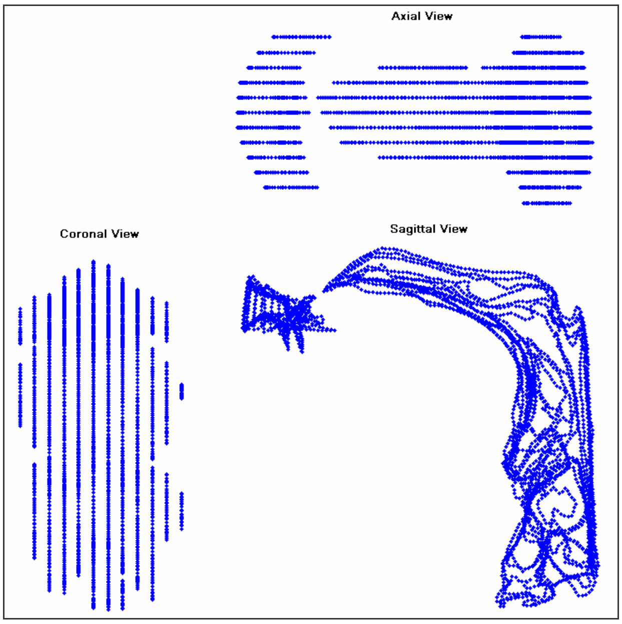


Fig.3 Sagittally-sampled tract: Subject M1 [asa]

Anatomical landmarks were resampled, and centers of gravity were calculated. Sagittal and oblique point clouds were automatically rotated and translated into the axial orientation such that the centers of gravity were coincident, then manually adjusted where necessary to align all three clouds as closely as possible (**Fig. 4**).

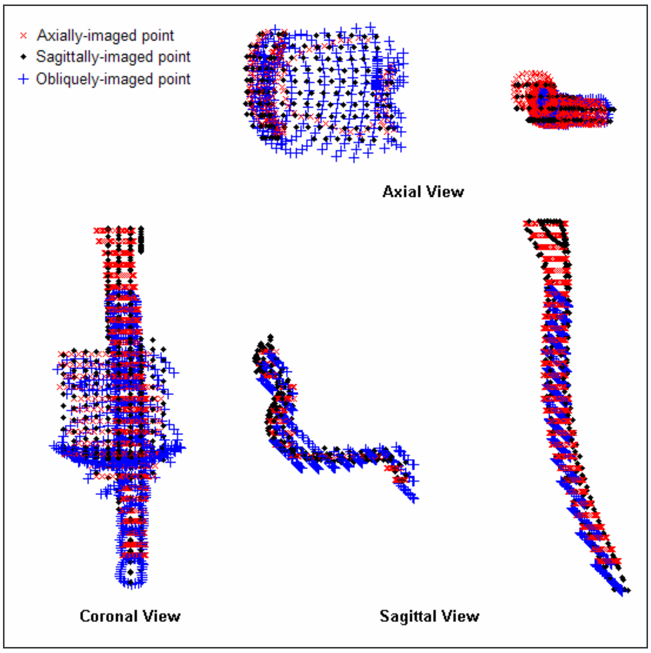


Fig.4 Alignment of anatomical landmarks: spinal column and facial surface around chin (Subject M2 [aʃa])

The spatial alignment between point-clouds was further checked by comparing the three facial surfaces sampled using different imaging orientations (**Fig. 5**).

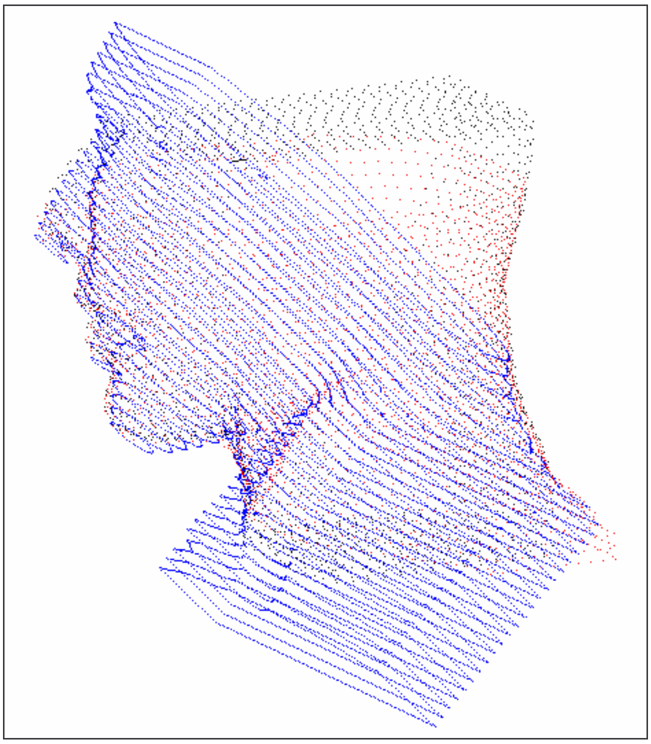


Fig.5 Alignment of facial surfaces: Subject M2 [aʃa]

From the resulting alignments, two transformation matrices were calculated, one representing the mapping of sagittally-imaged points onto the axial system, and another representing the oblique-to-axial mapping. With the spatial relationships between the three imaging orientations established, tract boundaries were imported, resampled and transformed using the landmark mappings. The result of this process was a superimposition of three sets of points representing the same vocal tract, sampled in three different orientations (**Fig.6**).

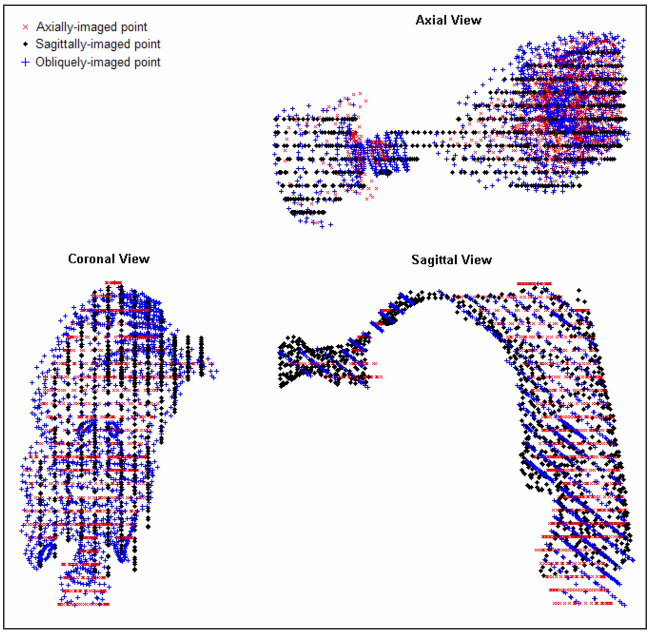


Fig.6 Multiply-sampled tract: Subject M1 [aʃa]

A thrice-sampled composite tract model was assembled in this manner for each fricative produced in each vocalic context by each subject – 120 models in total.

4 Model Rationalization

Because the construction of a composite tract using this technique involves superimposition of three fully-sampled models, the resulting tracts contain many more data points than are ultimately necessary for many purposes. As a result, some methods of data reduction are also required to rationalize the model for specific uses.

4.1 Piecemeal Tract Models

The most straightforward means of extracting a model useful for fine quantification of tract geometry is a piecemeal approach in which different regions of the tract are primarily represented with data drawn entirely from one of the underlying models.

Because the axial images provide the most accurate representation of the pharyngeal region, these data points alone are initially used in the part of the model from the glottis to the upper pharynx. Oblique sections are used in the region anterior to the mid-palate, potentially extending as far as the oral cavity. In other regions, the tract is represented by a series of points sampled where planes perpendicular to the tract intersect the sagittal tract model.

After applying the spatial transformations derived from the alignment of anatomical landmarks, axial and oblique point clouds are superimposed on the midsagittal slice (**Fig. 6**).

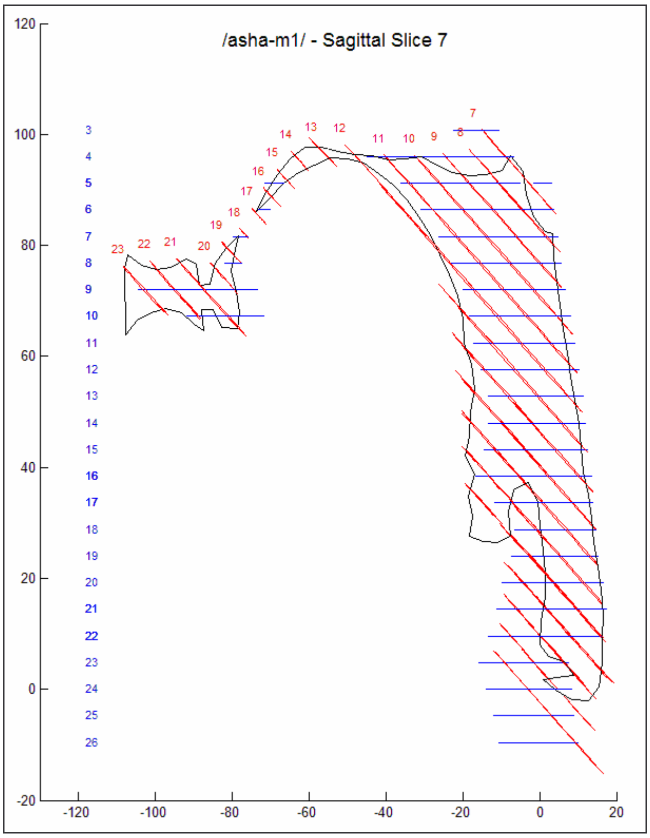
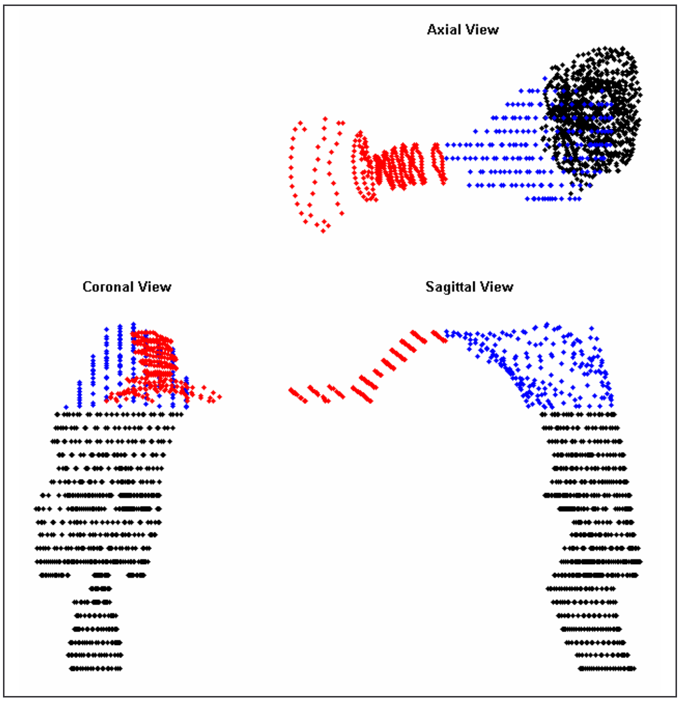


Fig.6: Construction of Piecemeal Tract Model: Axial and oblique-coronal slices (numbered) superimposed on the reference mid-sagittal plane (Subject M1: [aʃa])

The subset of axial slices which will be used to represent the pharynx is nominated, as well as the range of oblique slices to be used around the alveolar region. Between these regions, and forward of the oblique slices, a tract center is calculated at regular intervals, through which planes normal to the tract interval centerline are projected onto the sagittally-sampled point cloud (**Fig. 7**).



.**Fig.7: Construction of Piecemeal Tract Model**: Compilation of tract-normal sections (Subject M1: [aʃa])

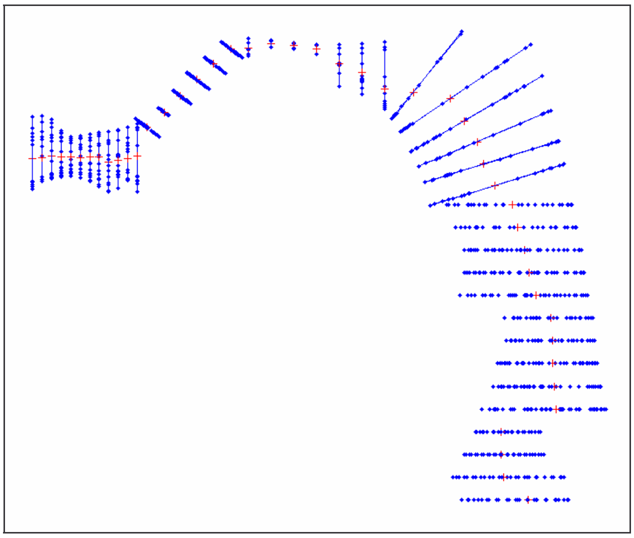
The model is thus assembled by compiling data drawn from companion sets of imaging data, and arranging these data points within a common coordinate space. Because this approach does not utilize any rescaling or interpolation of points, the tract resolution provided by the original images is inherent in the model.

There are four main sources of error in the construction of this composite model: the resolution of the original MR images, errors arising from the image segmentation, interpolation errors which arise in regions where the tract is not modeled using original data points, and errors in the identification and alignment of anatomical landmarks.

While the first three sources of error are common to any type of MRI-derived model, the potential for interpolation error in this approach is considerably smaller than in other methods. The legitimacy of models derived using this approach is largely determined by the accuracy with which the landmarks are resolved.

5 Extraction of Area Functions

Piecemeal tract models constructed in this manner are particularly well suited to the calculation of Area Functions, because the sectional areas are calculated directly from images which were acquired using planes normal to the axis of the tract, except in the intermediate regions, where data is sampled from the sagittal stack (**Fig. 8**).



.**Fig.8: Extraction of Area Functions** from Piecemeal Tract Model. Subject M1 [aʃa]: Sagittal view.

6 Results and Applications

Tract models produced using this method are especially useful for the study of fricatives, where accurate knowledge of the 3d-geometry of the tract is critical [6]. By drawing on data from imaging orientations especially selected to target the region of the constriction, the tract models assembled in this manner provide superior resolution of the region of interest to this study.

Vocal tract models constructed in this manner have already informed our knowledge of the importance of vowel context on production of fricatives, the mechanisms of tongue grooving in sibilants, the morphology of the sublingual cavity, and differences in pharyngeal volume between voiced and voiceless fricatives [7].

Although the methodology described here was specifically developed to allow for the incorporation of oblique-coronal imaging planes into a tract model, this approach to the construction of hybrid tract models could be used to assemble tracts from any matching sets of MRI images, provided they contain sufficient information about a common set of anatomical landmarks to allow for accurate cross-image alignment.

7 Conclusion

A novel technique for the construction of models of the vocal tract from multiply-imaged MRI data has been presented. The composite models produced using this approach integrate data from three different imaging orientations, each of which offers superior resolution of different parts of tract. At the macro level, the combination of three sets of points can provide a more accurate representation of the gross morphology of the tract through oversampling and minimization of the errors introduced by any one imaging orientation. At smaller scales, a subset of the data points most relevant to the area of tract under consideration can be utilized.

Acknowledgments

The work presented here is part of an ongoing project supported by NIH grant NIDCD R01 DC 006705.

References

[1] S. Narayanan, A. Alwan, K. Haker, "An articulatory study of fricative consonants using magnetic resonance imaging." *J. Acoust. Soc. Am.* 98, 1325-1347 (1995)

[2] S. Maeda, "Un modele articulatoire de la langue avec des composantes lineaires." *Proc. 10th Journees d'Etude sur la Parole (Grenoble)* 152-163 (1979)

[3] D. Beautemps, P. Badin, G. Bailly, "Linear degrees of freedom in speech production: Analysis of cineradio- and labio-film data and articulatory-acoustic modeling." *J. Acoust. Soc. Am.* 109, 2165-2180 (2001)

[4] B. H. Story, I. R. Titze, E. Hoffman, "Vocal tract area functions from magnetic resonance imaging". *J. Acoust. Soc. Am.* 100, 537-554 (1996)

[5] 3D-Doctor version 4.0, (Able Soft Corp: 5 Appletree Lane, Lexington, MA 02420 http://www.ablesw.com)

[6] C. H. Shadle, "The effect of geometry on source mechanisms of fricative consonants." *J. Phon*. 19, 409-424 (1991)

[7] C. Shadle, M. I. Proctor, K. Iskarous. "MRI Study of the Effect of Vowel Context on English Fricatives", Proc. ASA (2008)