An analysis of vocal tract shaping in English sibilant fricatives using real-time magnetic resonance imaging

Erik Bresch¹, Daylen Riggs², Louis Goldstein², Dani Byrd³, Sungbok Lee², Shrikanth Narayanan¹,²

¹Department of Electrical Engineering, University of Southern California, Los Angeles, CA, USA ²Department of Linguistics, University of Southern California, Los Angeles, CA, USA

{bresch,daylenri,louisgol,dbyrd,sungbokl}@usc.edu, shri@ipi.usc.edu

Abstract

This study uses real-time MRI to investigate shaping aspects of two English sibilant fricatives. The purpose of this article is to 1) develop linguistically meaningful quantitative measurements based on vocal tract features that robustly capture the shaping aspects of the two fricatives, and 2) provide qualitative analyses of fricative shaping. Data was recorded in both midsagittal and coronal planes. The proposed three quantitative measures of this study provide robust results in categorizing shape. The qualitative analyses describe tongue shape in terms of grooving and doming and they support previous research.

Index Terms: real-time MRI, fricatives, sibilants, tongue shape analysis

1. Introduction

Real-time magnetic resonance imaging (RT-MRI) promises a new means for visualizing and quantifying the spatio-temporal articulatory details of speech production. This study uses RT-MRI imaging to study the shaping and dynamic aspects of two English sibilant fricatives /s/ and /S/. The RT-MRI data, which affords views of the entire moving vocal tract, and is accompanied by synchronized audio recordings, aims to build upon, and add to, the several excellent previous/ongoing studies that aim to fully capture the static and dynamic properties of fricatives. Acoustic studies are capable of illuminating the dynamical nature of fricatives, as changes in formant structure indicate changes in the vocal tract while producing speech sounds, (e.g., [1], [2]). However, acoustic studies are unable to characterize the exact shaping of the tongue (and perhaps other articulators) necessary for the production of fricatives. Other experimental methods for obtaining speech production information (e.g. EMA, static MRI, x-ray, etc.), lack information about change over time, or lack the spatial resolution necessary to properly characterize the complete shape of articulators during the production of fricatives. RT-MRI provides an ideal tool for measuring both the posture and temporal (dynamic) properties of fricatives and the effect of surrounding vowel context on the fricatives, as MRI provides us a (nearly) complete picture of the dynamics of the vocal tract. RT-MRI additionally provides valuable insights into the production of fricatives as it is able to examine the shaping of such fricatives in various planes. The current study employs this technique to obtain data from both midsagittal and coronal planes of the vocal tract.

Specifically, the aim of this paper two-fold. The primary objective is to investigate various derived measurements from MR images that are linguistically meaningful. These measurements are based on tongue shape and other properties of the vocal tract seen in the MR images, and they allow for analysis of the two fricatives under examination. The measurements described below allow for the explicit study of shaping differences between the tongue tip and tongue body gestures of /s/ and /S/. Sibilant fricatives are of particular interest because of the complexity displayed in their shape during production. This shaping of fricatives (e.g. the grooving of the tongue for sibilant fricatives) has been shown to be crucial in yielding their acoustic properties necessary for perception [3]. Thus, defining measurements that allow for shape to be investigated as a variable is crucial for the validation of any hypothesis addressing the role of shape in the production of fricatives. Three derived measures are shown to yield linguistically meaningful results. They are 1) tongue-palate area behind fricative constriction (midsagittal), 2) tongue-palate area deformation (midsagittal), and 3) fricative groove-depth (coronal). These measurements are discussed in the methods section.

Secondly, the study aims to describe various strategies used by speakers to produce sibilant fricatives. Based on previous research on the shaping of English fricatives (e.g., [4]) it is expected that /s/ will show grooving of the tongue and /S/ will show doming of the tongue. This study also asks what effect surrounding vowel context will have on the shaping of the fricatives. The fricatives under examination were preceded and followed by the vowels /i/ and /a/. The vowel /i/ has been described as showing a convex dome-like shape, whereas /a/ is considered to be flat [5]. The study thus seeks to determine what happens when two contrasting shapes (such as flat /a/ and grooved /s/) are adjacent. Two possible hypotheses are considered: target shaping in the fricative will be more pronounced when contrasting shapes are adjacent, or target shaping in the fricative will be more pronounced when comparable shapes are present.

Several previous studies using MRI have shown shape as crucial parameter in the production of fricatives. Co-articulatory effects have been shown to be important in describing the shape of Swedish fricatives [6]. When examining English sibilant fricatives with MRI, various observations about tongue posture have been shown to be important. /S/ has been shown to be articulated with a raised tongue blade that is distributed across the alveolar ridge. /S/ also shows a sublingual cavity behind the constriction, whereas /s/ shows an absence of a sublingual cavity [7]. The concave shaping of the tongue has also been shown to be crucial for the production of English sibilant fricatives; /s/ consistently shows concavity behind the constriction region, whereas /S/ does not [4]. The area posterior to the constriction has also been shown to be important: the area functions derived for /s/ tend to be less smooth than those for /S/. This difference has been shown to be due to a slight raising of the tongue for /s/ postures. This study further investigates the importance of the area behind the constriction during fricatives, examining area and area deformation behind the tongue. These two variables are shown to be important to quantitatively characterize the shaping differences in the two fricatives.

2. Methods

Three native speakers of American English were used as subjects, two female (A2, S2), and one male (A1). Subjects had no known speech or hearing deficits.

2.1. Simuli

The fricatives /s/ and /S/ under analysis occur between either the vowels /i/ or /a/ in a carrier phrase (“Go pVCPVp okay”, V= /æ, ɪ/; C= /s, S/). There are four different surrounding vowel contexts: symmetrical: /i/i/ , /a/ a/; asymmetrical: /i/ a/, /a/ i/. As
there are two different fricatives, this yields eight different stimuli.

The sentences were grouped into fourteen blocks, seven blocks for each fricative. Thus each stimulus was read seven times by each subject. Each block containing four sentences was randomized, and the order of the blocks was randomized, with the constraint that blocks alternated according to fricative (two /s/ blocks were never consecutive).

2.2 Real-time MRI and synchronized audio acquisition

MR images were acquired on a GE Signa 1.5 Tesla scanner using a fast gradient echo pulse sequence with a 13-interleaf spiral readout [8] within the RTHawk framework [9]. A four-channel targeted phased-array receiver coil was employed. Images were formed from the data of two coils located in front of the subject’s face and neck through root sum of squares combining. The repetition time TR was 6.376ms. The MRI reconstruction was carried out using a standard gridding and sliding-window technique [10] with a window offset of 7 acquisitions. The resulting in a frame rate for processing and analysis was 22 frames per second. The slice thickness was 3mm.

Images were acquired in the midsagittal plane and in a coronal plane. The midsagittal field-of-view (FOV) was chosen to capture the entire vocal tract from glottis to lips. The image rotation was chosen so that the pharyngeal wall is approximately vertical. A sample midsagittal image sequence of the utterance “pa seep” is shown in Figure 1. The coronal scan plane was selected perpendicular to the midsagittal scan plane at the position of maximal doming of the hard palate. The FOV was identical in size with respect to the midsagittal value. A sample coronal image is shown in Figure 4.

Simultaneous synchronized speech audio was collected during the MRI scans. Subsequently, a noise cancellation procedure was applied to the audio signal to remove the MRI gradient noise [11]. Sample videos are available at http://sail.usc.edu/span/interSpeech2008/index.php.

2.3 Image analysis

In the midsagittal MR images the vocal tract contours were traced using the procedure described in [12]. Figure 2(a) shows an example midsagittal MR image of the fricative /s/ during its maximum constriction. Here the red lines delineate the outline of the vocal tract, and the green lines are the aperture lines which were computed using the methods described in [13].

In order to quantify the spatio-temporal shaping characteristics of the speech sounds of interest appropriate geometrical features have to be derived from each image. For the fricative sounds /s/ and /S/ one candidate feature is the midsagittal aperture at, and posterior to, the critical vocal tract constriction, which in these cases is formed using the tongue tip and the alveolar ridge/front palate. We hence proceed by defining a region of interest in the midsagittal profile of vocal tract bordered by the minimum opening between tongue and hard palate on the left (left blue line in Figure 2(a)) and a vertical line dropping from the hinge point of the velum (right blue line in Figure 2(a)). The selection of these boundaries is motivated by the relatively reliable detection of these anatomical landmarks. Figure 3(a) shows the time evolution of the size of this tongue-palate area for single tokens of the utterances “pee seep” and “pa sop.” During the interval of the fricative production, which was identified for all tokens using the spectrogram of the synchronized audio recording, we observe a local minimum in the time function for the /a/ context and a local maximum for the /i/ context. The same holds for the /S/ sound as shown in Figure 3(b).

However, in order to better discriminate between the shaping difference of /s/ and /S/ it is also desired to devise a shape feature that is largely independent of the morphology of the subject’s hard palate. To be more specific, we would like a measure of how parallel the palate and the tongue contours are in the region of interest in order to be able to deduce information on the nature of the airway channel. We hence propose the use of a tongue-palate area deformation measure, which is illustrated in Figure 2(b). Here, the aperture function is shown (green) in addition to the location of the boundaries of the tongue-palate area of interest (vertical blue lines). As the deformation measure we use the variation (black shaded area) of the aperture about its mean value (magenta line) in the region of interest. This scalar deformation value will be near zero if the tongue is largely parallel to the hard palate, irrespective of the actual shape, and it will be large for non-parallel configurations. Figures 3(c) and 3(d) show sample time functions for the deformation feature for /s/ and /S/,
behind tongue front constriction (based on [4]), since it was an-
other was motivated to capture the nature of shaping in the region
combination of contour points for a given frame.
complished through an exhaustive search over all triangle com-
tact by finding the tangent (blue) onto the surface which
bending tongue groove depth. It is derived from the coronal tongue sur-
For the coronal images, the tongue contour was traced using a
other sections of the vocal tract contours.
Hard palate contour location, if left unconstrained during the
signal to noise ratio leading to faint MRI contrast. Hence the
Since bone has a very low hydrogen content it produces a weak
larger image noise. This likely due to the fact that it consists
ure 1 the region of the hard palate is generally subject to rather
proposed feature extraction process. As can be seen in Fig-
maximum
derived measurements discussed above reveal quantitative support
for this finding. These are further discussed below.
Observational differences found in this study also point to
wide inter-speaker variability. One such parameter of inter-
speaker variability is seen in the symmetry of groove/dome forma-
tion in the coronal plane. Subject S1 generally had very sym-
metrical tongue shapes during the formation of the groove of
/t/ and /S/ showed doming, whereas /s/ showed grooving and the /a/
was expected to be flat (or sometimes, slightly concave).
The expectations are largely seen in the speakers’ produc-
tions in this study. Speaker A2 behaved the most canonically
with respect to the above expectations. /i/ and /S/ both showed
doming, whereas /s/ showed grooving and the /a/ vowel was flat.
Speakers A1 and S1 also behaved as expected; in fact, speaker
A1 showed a slight grooving (concavity) of /a/ when this vowel
was adjacent to both fricatives.
One hypothesis that is evaluated in the current study deals
with the effect of the surrounding vowels on the shaping of the
fricative: whether contrasting tongue shapes in the surround-
mg vowels will cause the shaping of the fricatives to be more
pronounced, or whether tongue shape follows “inertia,” and the
shaping of the fricative is more pronounced with comparable
shapes of adjacent vowels. The trends seen in the data support
the latter hypothesis. For example, the grooving of /s/ is more
profound in the /a/a/ context than in the /i/i/ context. The de-
were adjacent to both fricatives.
Differences in the grooving and doming of the two sibilant frica-
tives and shape of the surrounding vowel were examined. Ta-
ble 1 describes tongue shaping aspects based on both the coro-
nal and midsagittal planes. Here, “D,” “G” and “F” stand for
“domed,” “grooved,” and “flat,” respectively. The sequence in
each cell is the sequence of the segments: the preceding vowel,
the fricative, and the following vowel.
3. Results
3.1. Descriptive analysis
Differences in the grooving and doming of the two sibilant frica-
tives and shape of the surrounding vowel were examined. Ta-
ble 1 describes tongue shaping aspects based on both the coro-
nal and midsagittal planes. Here, “D,” “G” and “F” stand for
domed,” “grooved,” and “flat,” respectively. The sequence in
each cell is the sequence of the segments: the preceding vowel,
the fricative, and the following vowel.
Table 1: Qualitative analysis of tongue shaping in the three sub-
jects (“D,” “G” and “F” stand for “domed,” “grooved,” and
“flat,” respectively).

<table>
<thead>
<tr>
<th>subject</th>
<th>fricative</th>
<th>/a/</th>
<th>/a/a/</th>
<th>/a/S/</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>/s/</td>
<td>DDR</td>
<td>FDF</td>
<td>DDR</td>
</tr>
<tr>
<td>A1</td>
<td>/S/</td>
<td>DDR</td>
<td>DFF</td>
<td>DDR</td>
</tr>
<tr>
<td>A2</td>
<td>/s/</td>
<td>DDR</td>
<td>DFF</td>
<td>DDR</td>
</tr>
<tr>
<td>A2</td>
<td>/S/</td>
<td>DDR</td>
<td>DDF</td>
<td>DDR</td>
</tr>
<tr>
<td>S1</td>
<td>/s/</td>
<td>DDR</td>
<td>DFG</td>
<td>DDR</td>
</tr>
<tr>
<td>S1</td>
<td>/S/</td>
<td>DDR</td>
<td>FDF</td>
<td>DDR</td>
</tr>
</tbody>
</table>

Based on previous studies of sibilant fricative shaping (e.g., [4]), /s/ is expected to show concave grooving of the front
tongue, whereas /S/ is expected to show convex doming. The
vowels are also expected to show specific shapes: /i/ has a
domed tongue shape, and /a/ is expected to be flat (or sometimes,
slightly concave).
The expectations are largely seen in the speakers’ produc-
tions in this study. Speaker A2 behaved the most canonically
with respect to the above expectations. /i/ and /S/ both showed
doming, whereas /s/ showed grooving and the /a/ vowel was flat.
Speakers A1 and S1 also behaved as expected; in fact, speaker
A1 showed a slight grooving (concavity) of /a/ when this vowel
was adjacent to both fricatives.
One hypothesis that is evaluated in the current study deals
with the effect of the surrounding vowels on the shaping of the
fricative: whether contrasting tongue shapes in the surround-
mg vowels will cause the shaping of the fricatives to be more
pronounced, or whether tongue shape follows “inertia,” and the
shaping of the fricative is more pronounced with comparable
shapes of adjacent vowels. The trends seen in the data support
the latter hypothesis. For example, the grooving of /s/ is more
profound in the /a/a/ context than in the /i/i/ context. The de-
were adjacent to both fricatives.

Another dimension of much inter-speaker variability was the
tongue front apicality or laminality in producing the fricatives
studied. English sibilant fricatives have been described as pro-
duced either apically or laminally. The three speakers examined
in this study showed wide variability with respect to this aspect.
Subject A2 showed an apical /s/ and a laminal /S/. Subject S2
had apical articulation for both /s/ and /S/, whereas speaker A1
had a laminal articulation for both of the fricatives.

It should be noted that while the choice of the scan plane
was adjacent to both fricatives.

Observational differences found in this study also point to
wide inter-speaker variability. One such parameter of inter-
speaker variability is seen in the symmetry of groove/dome forma-
tion in the coronal plane. Subject S1 generally had very sym-
metrical tongue shapes during the formation of the groove of
/s/ and the dome of /S/. This symmetry was not generally seen
for speaker A1. A1 had an asymmetrical tongue shape during
groove/dome formation; one side of the tongue, to the observer’s
left of the observed groove in /s/, was often higher than the other
side. This asymmetry was likewise observed for /S/: one side of
the tongue was higher than the other during the formation of the
dome.

Another dimension of much inter-speaker variability was the
tongue front apicality or laminality in producing the fricatives
studied. English sibilant fricatives have been described as pro-
duced either apically or laminally. The three speakers examined
in this study showed wide variability with respect to this aspect.
Subject A2 showed an apical /s/ and a laminal /S/. Subject S2
had apical articulation for both /s/ and /S/, whereas speaker A1
had a laminal articulation for both of the fricatives.

It should be noted that while the choice of the scan plane
was adjacent to both fricatives.

Generally, speakers A2 and S1 showed a narrow constric-
tion region for /s/ and a wide constriction region for /S/. Co-
ardiculatory effects of the surrounding vowels also played an im-
portant role on the shaping of the fricatives. The constriction for
/S/ is made further back before the back vowel /a/ than the front

Figure 3: Midsagittal features sample time functions.

Figure 4: Coronal sample image, tongue contour (red), groove
tangent (blue), groove depth feature (green) during the fricative production in “pa seep.”

respectively, for the /a/a/ and /i/i/ context, and we observe a local
non-parallelity maximum during the interval of the fricative
production. We assume that at the maximum point the constric-
tion target was achieved the preparation of the production of the
following vowel will set it.

Our subsequent investigations will utilize the averaged max-
imum deformation over all tokens of a particular type, as well
as the averaged area size at the time of maximum deformation.
However, at this point we also want to point out a limitation of
the proposed feature extraction process. As can be seen in Fig-
ure 1 the region of the hard palate is generally subject to rather
larger image noise. This likely due to the fact that it consists
of bone which is covered with a rather than layer of soft-tissue.
Since bone has a very low hydrogen content it produces a weak
signal to noise ratio leading to faint MRI contrast. Hence the
hard palate contour location, if left unconstrained during the
automatic tracing process, can be adversely affected more than
other sections of the vocal tract contours.

2.4. Coronal plane images
For the coronal images, the tongue contour was traced using a
semi-automatic method [13]. Contours were initialized and cor-
rected manually. Figure 4 shows an example coronal image with
the traced tongue contour (red).

As the analysis feature of interest we chose the coronal
tongue groove depth. It is derived from the coronal tongue sur-
faccle contour by finding the tangent (blue) onto the surface which
forms the triangle with a maximum height (green). This is ac-
complished through an exhaustive search over all triangle com-
bination of contour points for a given frame.

It should be noted that while the choice of the scan plane
was motivated to capture the nature of shaping in the region
behind tongue front constriction (based on [4]), since it was an-
anchored to an anatomical landmark, the images are not expected
to align to any key shaping landmark such as maximal grooving
for /s/ or doming for /S/, but instead provide an indication of the
general shaping in that region.

3. Results
3.1. Descriptive analysis
Differences in the grooving and doming of the two sibilant frica-
tives and shape of the surrounding vowel were examined. Ta-
ble 1 describes tongue shaping aspects based on both the coro-
nal and midsagittal planes. Here, “D,” “G” and “F” stand for
“domed,” “grooved,” and “flat,” respectively. The sequence in
each cell is the sequence of the segments: the preceding vowel,
the fricative, and the following vowel.
vowel /i/. It is also seen that the grooving of /s/ is more profound when this consonant is between /a/ than /i/.

3.2. Quantitative analysis

Using the measures discussed above, clear differences between the two English fricatives /s/ and /S/ can be seen. Figures 5, 6, and 7 show the measurements for subjects A1, A2, and S1.

The tongue-palate area at the time of maximum constriction shows robust differences between /s/ and /S/ for all three subjects. The fricative /s/ has a greater tongue-palate area at time of maximum constriction across all vowel contexts.

The differences in the measures were most robust for speaker A2. For tongue-palate area deformation and for tongue-palate area, clear differences are seen for the two fricatives. The maximum tongue-palate area deformation for speaker A2 was much greater for /s/ than it is for /S/ across all vowel contexts.

The grooving for /s/ was consistently seen in all subjects; the groove depth measured from the coronal slice was significantly greater under /a/ than /i/. It is also seen that the grooving of /s/ is more profound when this consonant is between /a/ than /i/.

4. Discussion

The measurements of MR images of the vocal tract discussed here provide useful techniques for studying the differences between the two English fricatives /s/ and /S/. The measurements of area and variance (midsagittal) and groove depth (coronal) are able to provide a way to distinguish between the two fricatives, which is not always a transparent and simple task.

Deriving concrete measurements of the articulatory properties of fricatives is necessary for further studies examining the shaping properties of fricatives. The real-time MRI techniques described here provide for fruitful analyses of the co-articulation effects between vowels and consonants. Examining the concavity of the fricatives with respect to the convexity of /i/ and the flatness of /a/ allows for several linguistic hypotheses to be addressed. The level of explicit control of tongue shape during fricative production is one such research question that will be considered using the data set discussed above.

Another possible avenue of research involves examining the variability across speakers. In our data, A1 uses a different part of the tongue tip to produce the sibilant fricatives than the other two subjects. Palatal morphology or other physiological differences could play a role. The MR data and measures developed here provide for clear ways of answering this question and others.

5. Acknowledgements

This work was supported by NIH grant R01 DC007124-01. The authors would like to thank the USC Imaging Science Center.

6. References