3D Real-Time MRI of Vocal Tract Shaping

Yongwan Lim, Yinghua Zhu, Sajan Goud Lingala, Dani Byrd, Shrikanth Narayanan, Krishna Nayak

Ming Hsieh Department of Electrical Engineering,
University of Southern California
Los Angeles, CA, USA
Introduction – RT-MRI of Speech

Spiral 2D RT-MRI of speech

1. Lingala SG et al., JMRI. 2016;43:28–44
Introduction – RT-MRI of Speech

• Speech scientists seek a comprehensive understanding of human vocal tract shaping and its dynamics.

• RT-MRI techniques are being used to study dynamic function of articulators non-invasively\(^1,2\).

• However, most of RT-MRI techniques have been limited to one mid-sagittal or a few 2D imaging planes\(^1\).

1. Lingala SG et al., JMRI. 2016;43:28–44
Introduction – Vocal Tract Shaping During Speech

• Enormously complex in **geometry** and **temporal structuring**

• Cannot be fully understood from mid-sagittal constriction posture along the vocal tract.\(^3,4\)
  - E.g., grooving/doming/lateral shaping of tongue, asymmetries in tongue shape, resonate cavity volume

• **3D vocal tract shaping and its dynamics is essential.**

Goal of This Work

• To develop 3D RT-MRI technique of the full vocal tract at high temporal resolution during natural speech.

• To evaluate this technique using in-vivo vocal tract airway data and via comparison with interleaved multislice 2D RT-MRI.
Methods – Overview of 3D RT-MRI

Pseudo Golden Angle

Stack-of-spirals

Rotational angle

$k_z$ index

$\theta_1 = 0^\circ$  $\theta_2 = \theta_{GA}$  $\theta_3 = 2\theta_{GA}$  $\theta_N = (N-1)\theta_{GA}$  $\theta_{N+1} = 0^\circ$

12 TR  24 TR  ...  12*N TR
Methods – Overview of 3D RT-MRI

Pseudo Golden Angle

Stack-of-spirals

Rotational angle

Temporal window

\[ \theta_1 = 0^\circ \]
\[ \theta_2 = \theta_{GA} \]
\[ \theta_2 = 2\theta_{GA} \]
\[ \theta_N = (N-1)\theta_{GA} \]
\[ \theta_{N+1} = 0^\circ \]
Methods – Overview of 3D RT-MRI

- Pseudo Golden Angle
- Stack-of-spirals
- Undersampled by 13

Rotational angle

Undersampled in k_x-k_y plane

Temporal window

θ_1 = 0°, θ_2 = θ_{GA}, θ_2 = 2θ_{GA}, θ_N = (N-1)θ_{GA}, θ_{N+1} = 0°
Methods – Overview of 3D RT-MRI

- Pseudo Golden Angle
- Stack-of-spirals
- Undersampled by 13

Rotational angles:

\[
\begin{align*}
\theta_1 &= 0^\circ \\
\theta_2 &= \theta_{GA} \\
\theta_3 &= 2\theta_{GA} \\
\theta_N &= (N-1)\theta_{GA} \\
\theta_{N+1} &= 0^\circ
\end{align*}
\]

Temporal window:

Undersampled in \(k_x-k_y\) plane

IFT
Methods – Overview of 3D RT-MRI

- Pseudo Golden Angle
- Stack-of-spirals
- Undersampled by 13
3541: 3D Real-Time MRI of Vocal Tract Shaping

Methods – Overview of 3D RT-MRI

Pseudo Golden Angle

Stack-of-spirals

Undersampled by 13

Rotational angle

Undersampled in $k_x$-$k_y$ plane

Inverse FT along the $k_z$ direction
Pseudo Golden Angle

Stack-of-spirals

Undersampled by 13

Decompose 3D data into 2D slice data

Rotational angle

Temporal window

Undersampled in $k_x$-$k_y$ plane

Inverse FT along the $k_z$ direction

$\theta_1 = 0^\circ$, $\theta_2 = \theta_{GA}$, $\theta_2 = 2\theta_{GA}$, $\theta_N = (N-1)\theta_{GA}$, $\theta_{N+1} = 0^\circ$
Methods – Overview of 3D RT-MRI

Pseudo Golden Angle

Stack-of-spirals

Undersampled by 13

Decompose 3D data into 2D slice data

Slice-by-slice 2D Constrained Reconstruction

Rotational angle

Temporal window

Undersampled in $k_x$-$k_y$ plane

Inverse FT along the $k_z$ direction

Reconstructed 3D image series

$$\theta_1 = 0° \quad \theta_2 = \theta_{GA} \quad \theta_2 = 2\theta_{GA} \quad \theta_N = (N-1)\theta_{GA} \quad \theta_{N+1} = 0°$$
Methods – Some Details

• A minimum-phase slab excitation pulse\(^5\)
  • Slab thickness = 50 mm, FA = 5 degree, TBW = 16

• Reconstruction
  • Sparse-SENSE with spatiotemporal TV constraints\(^6,7\)
    \[
    \arg\min \frac{1}{2} \|FSm - y\|_2^2 + \lambda_s \left\| \sqrt{|D_xm|^2 + |D_ym|^2} \right\|_1 + \lambda_t \|D_tm\|_1
    \]

• Coil maps: ESPIRiT\(^8\) using time-averaged 3D data

### Methods – Imaging Parameters

- **MRI Protocols**
  - GE Signa 1.5T scanner

<table>
<thead>
<tr>
<th></th>
<th>2D RT-MRI</th>
<th>3D RT-MRI</th>
</tr>
</thead>
<tbody>
<tr>
<td># of slices</td>
<td>3 orthogonal (interleaved)</td>
<td>12 (# of kz encode)</td>
</tr>
<tr>
<td>TR (ms)</td>
<td>6</td>
<td>5.05</td>
</tr>
<tr>
<td>TE (ms)</td>
<td>0.8</td>
<td>0.68</td>
</tr>
<tr>
<td>FA (degree)</td>
<td>15°</td>
<td>5°</td>
</tr>
<tr>
<td>FOV (mm³)</td>
<td>200 x 200 x 6</td>
<td>200 x 200 x 70 mm³</td>
</tr>
<tr>
<td>Spatial resolution (mm³)</td>
<td>2.4 x 2.4 x 6 mm³</td>
<td>2.4 x 2.4 x 5.8 mm³</td>
</tr>
<tr>
<td>Temporal resolution (ms / frame)</td>
<td>18 ms (1 spirals / frame)</td>
<td>61 ms (1 spirals in kₓ-kᵧ + 12 kₓ / frame)</td>
</tr>
<tr>
<td>Acceleration</td>
<td>13x</td>
<td></td>
</tr>
</tbody>
</table>

Results – 2D Multislice vs 3D RT-MRI

2D three-slice

<table>
<thead>
<tr>
<th>Mid-sagittal</th>
<th>Axial</th>
<th>Coronal</th>
</tr>
</thead>
<tbody>
<tr>
<td>lips</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3D

<table>
<thead>
<tr>
<th>Mid-sagittal</th>
<th>Axial</th>
<th>Coronal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Results – 2D Multislice vs 3D RT-MRI

Image vs. time profile along cut 1

Image vs. time profile along cut 2

2D Multislice
18 ms/frame

3D
61 ms/frame

alveolar ridge

lip

tongue

pharyngeal wall

0s 16s 0s 16s
Results – 3D RT-MRI with Sync. Audio

Reformatted image planes

Stimuli
Type ‘a slab,’ Abigail.
Type ‘pass lab,’ Abigail
Type ‘a Sal,’ Abigail
Type ‘a say lab,’ Abigail
Type ‘a sap lab,’ Abigail
Results – Consonants /s/ and /l/

[l]: tongue sides low, allowing lateral airflow
[s]: tongue sides up and braced and the tongue surface grooved for central airflow
Linguistic Study – Contrasting /s/ and /l/

Mean ROI Intensity

Time [s]

Mean ROI Intensity [au]

Time [s]

Mean ROI Intensity [au]

Time [s]

Mean ROI Intensity [au]

Time [s]

Mean ROI Intensity [au]

Time [s]

Mean ROI Intensity [au]

Time [s]

Mean ROI Intensity [au]

Time [s]

Mean ROI Intensity [au]

Time [s]

Mean ROI Intensity [au]

Time [s]

Mean ROI Intensity [au]

Time [s]

Mean ROI Intensity [au]

Time [s]

Mean ROI Intensity [au]

Time [s]

Mean ROI Intensity [au]

Time [s]

Mean ROI Intensity [au]

Time [s]
Linguistic Study – Syllable-Final and Initial /l/

Little onset timing difference between TT up, TD back, TS down

TS down lower and longer
Results – Vocal Tract Area Function

2D static MRI (multi-slice)


3D static MRI

Skordilis Z et al, ICASSP, 2017
Results – Vocal Tract Area Function

2D static MRI (multi-slice)

a. Location of planes used to calculate area functions
b. Area function for Portuguese vowel [8]
c. Mid-sagittal slice with showing the Portuguese vowel [8]

3D static MRI


3D RT-MRI

Skordilis Z et al, ICASSP, 2017

3541: 3D Real-Time MRI of Vocal Tract Shaping

/loo/-/lee/-/la/-/za/-/na/-/za/
Conclusion

• Demonstrate feasibility of 3D RT MRI of the full vocal tract, with spatiotemporal resolution adequate to visualize lingual movements and vocal tact shaping occurring during natural speech.

• Use a minimum-phase slab excitation, pseudo GA stack-of-spirals gradient echo, and spatio-temporal finite difference constrained reconstruction.

• Achieve $2.4 \times 2.4 \times 5.8 \text{ mm}^3$ spatial resolution, 61 ms temporal resolution, and a $200 \times 200 \times 70 \text{ mm}^3$ FOV, with 13-fold acceleration.
Acknowledgements

- USC Magnetic Resonance Engineering Laboratory (MREL)
- USC Speech Production and Articulation kNowledge (SPAN) group

- Funding
  - NSF grant 1514544
  - NIH grant R01-DC007124