Humans rely on the vocal tract and upper airway for many important functions, namely speech, swallowing, and breathing. This is an anatomic region of interest to basic scientists and clinicians. Imaging technologies such as X-ray, computed tomography (CT), ultrasound, and electromagnetic articulometry (EMA) have all been applied but have significant limitations. X-ray and CT involves exposure to ionizing radiation. Ultrasound imaging (US) is non-invasive and allows visualization of soft tissue and interfaces between air and tissue. In contrast, magnetic resonance imaging (MRI) involves no ionizing radiation, is inherently 3D, and provides excellent visualization of soft tissue and interfaces between air and tissue boundaries (see Figure 1). MRI is, however, notoriously slow, and our group has developed methods for acquiring images in real-time (roughly 24 frames per second), which is the first time imaging was possible in real-time (roughly 24 frames per second), which is the first time imaging was possible. The systems we use also provide interactive control over the imaging slice and scan parameters, allowing for faster examinations with fewer prescription errors. RT-MRI allows for monitoring of the vocal tract movements during speech, swallowing, and breathing tasks.

The majority of our work has been done as part of the Speech Production and Articulation Knowledge Group (SPAN, http://sail.usc.edu/span), which seeks to achieve a broader understanding of speech production and vocal tract shaping in humans. This interdisciplinary effort involves researchers from linguistics, otolaryngology, computer science, and electrical and biomedical engineering. We collect RT-MRI data at the Los Angeles County Hospital on the USC Health Science Campus. RT-MRI speech experiments, we also record synchronized and noise-cancelled audio, which relies on adaptive signal processing. Image segmentation is performed either manually or automatically to extract the vocal tract shape and location of articulators in each frame (see Figure 2). Figure 3 shows one example of RT-MRI being used in a study of “nasal” sounds (sounds that involve the velum lowering so that the air stream goes through the nose). RT-MRI allows for monitoring of vocal tract movements during speech, swallowing, and breathing tasks.

We are developing RT-MRI acquisition and reconstruction techniques that allow for retrospective selection of the temporal resolution and field-of-view (FOV). This is useful, because humans have natural variations in speaking rate and variables in the speed of different articulators. For example, the motion of articulators involved in production of vowel sounds or during pauses before words is slow. The motion of the tongue in production of stop consonants or diphthongs is very rapid.

The scheme we are currently exploring utilizes a “golden-ratio” sampling pattern in the spatial frequency domain (also known as k-space, the domain in which MRI data is acquired). Figure 4 compares imaging FOV as a function of temporal resolution for conventional and golden-ratio acquisitions. The golden-ratio method provides consistent and larger imaging FOV than the conventional approach. In addition, we are investigating approaches for automatic temporal resolution selection for each image region. Our ultimate goal is to improve overall image quality compared to the conventional approach.

We are finding ways to incorporate compressed sensing (CS) to speed up the acquisition of upper airway MRI. Resonance offsets that are present at interfaces between air and tissue pose some difficulty for conventional CS-MRI approaches. In our initial work in this area, we added a phase constraint step, and demonstrated reduced image artifact, and acceleration of 3D vocal tract imaging by 5. When CS is combined with parallel imaging using a multiple receive coil array, we could achieve 8-fold acceleration. The end result enabled acquisition of 1.33 mm isotropic resolution 3D imaging of the vocal tract in 7 seconds, which is suitable for sustained sound production. Examples are shown in Figure 5. We are currently exploring applications of this same approach to dynamic 2D and 3D upper airway MRI.

Synchronization using Audio Signals

One of our ultimate goals is 3D RT-MRI. As a partial solution, we are developing a methodology that constructs 3D dynamic movies from RT-MRI 2D data of several parallel sagittal slices acquired during repetitions of the same utterance. We then align and time warp the videos based on the synchronized audio recordings, using a general class of method called dynamic time warping.

Figure 6 shows temporal alignment of speech signals recorded at different slice imaging and three captured frames from reformatted videos of two parallel coronal slices. In addition to this approach, we are exploring highly accelerated 3D imaging approaches and novel receiver coils that may ultimately enable RT-MRI in 3D.