Articulation analysis using real time spiral MRI with TRACER

Bo Xu[1], Sam Tilsen[2], Pascal Spincemaille[3], Madhur Srivastava[2], Peter Doerschuk[2], and Yi Wang[1]
[1]Biomedical Engineering, Cornell University, Ithaca, New York, United States, [2]Cornell University, Ithaca, New York, United States, [3]Weill Cornell Medical College, New York, New York, United States

TARGET AUDIENCE: Anyone interested in dynamic MRI

PURPOSE: To use real-time MRI to perform articulation analysis. The recently proposed fast dynamic imaging method TRACER[1] is used for data acquisition and reconstruction. Spiral sampling and constrained reconstruction help to achieve ultra-high temporal frame rate that facilitates analysis of speech articulation.

METHOD: ***Data acquisition*:** Variable density golden angle ordered spirals were used in a 2D gradient spoiled echo sequence for data sampling. Two subjects were scanned at 1.5T (GE EXCITE) using a 3-channel shoulder coil placed anterior to the mouth. Scan parameters were: TR/TE=8.1/1.9ms, FA=5˚, BW=±125kHz, FOV=28cm, spatial resolution= 2.2×2.2mm^2, sagittal slice thickness 24mm and acquired matrix size 128×128. The acquired slice is approximately mid-sagittal. ***Speech stimulation*:** Each subject was instructed to produce 16 vowel-consonant-vowel (VCV) response: [apa, api, ipa, ipi, ama, ami, ima, imi, ata, ati, ita, iti, ana, ani, ina, ini], first with stress on the initial vowel, then again with stress on the final vowel, resulting in 32 VCV responses per block. Each block was repeated 8 times. ***Reconstruction:*** Time resolved 2D images were reconstructed with a 8ms frame rate using TRACER. An initial frame was reconstructed from the fully sampled dataset. Each subsequent frame was reconstructed with one single spiral leaf and the previous frame as a constraint. ***Data analysis:*** Images from each block for a given subject were registered with the first block. To analyze patterns of speech articulation, a time-series of average pixel intensities were extracted for regions of interest which were defined on an anatomical and linguistic basis in order to represent articulatory gestures[2] (Fig.1). Because syllable stress is known to influence articulatory kinematics, we conducted a repeated-measures ANOVA on kinematic variables (movement velocities, ranges, and durations for consonantal closure and release gestures), with the factors consonant, preceding vowel, following vowel, and position of stress. The analysis was limited to one speaker (a native speaker of English) since the other acquired English as an adult and exhibited non-native articulation patterns.

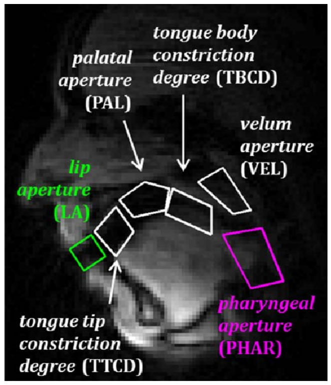

Fig. 1 Regions of interest used to extract tract variables.

RESULTS AND DISCUSSION: In Fig. 2 three representative images are shown at their corresponding time point for [ipa] utterances. Consonantal release gestures were greater in magnitude and speed when the consonant precedes a stressed syllable (F=37.91, p=0.002; F=31.9, p=0.002), which is consistent with previous studies[3]. Fig. 2 also shows that stressed vowels were associated with greater opening of the lips, and that vocalic movements were initiated approximately when the preceding consonantal constriction reaches a maximum[4] (cf. pharyngeal aperture during the [p]). This latter observation speaks to an important advantage of vocal tract rtMRI: due to superior coverage of posterior regions of the tract, rtMRI offers information on the dynamics of pharyngeal narrowing and velum aperture, whereas other technologies (e.g. ultrasound, electromagnetic articulometry, x-ray microbeam) cannot provide this information.

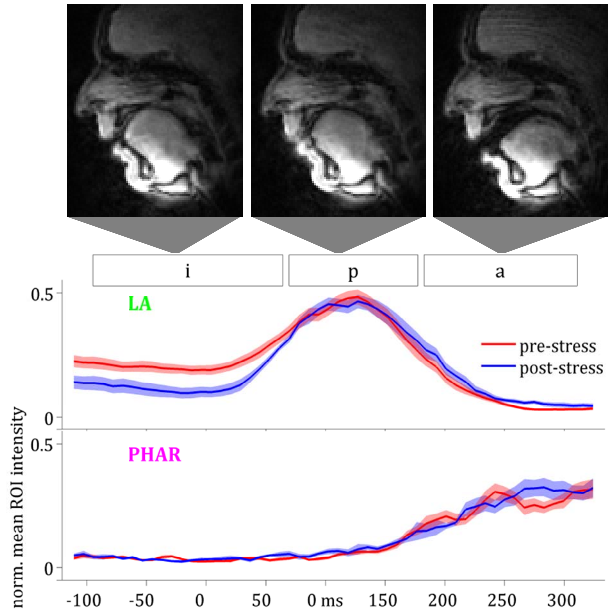

Fig. 2 Mean lip aperture and pharyngeal constriction (smoothed and normalized) in productions of [ipa] aligned to consonantal onset. Shading indicates the ±1 standard error at each time point.

CONCLUSION: Speech analysis with high temporal frame rate MR images obtained from TRACER is demonstrated here. Preliminary results have shown a great potential for investigation of speech motor control and development of clinical applications.

REFERENCES: [1] Xu et al. *MRM*, 2012, DOI: 10.1002/mrm.24253. [2] Saltzman et al. *Ecological Psychology*, 1989, 1(4), 333–382. [3] De Jong et al. *Journal of the Acoustical Society of America*, 1995, 97:1, 491-504. [4] Browman et al.1990, *Papers in laboratory phonology I: Between the grammar and physics of speech* (pp. 341-376)