ACT: An Automatic Centroid Tracking tool for analyzing vocal tract actions in real-time magnetic resonance imaging speech production data

Miran Oh, and Yoonjeong Lee

Citation: The Journal of the Acoustical Society of America **144**, EL290 (2018); doi: 10.1121/1.5057367
View online: https://doi.org/10.1121/1.5057367
View Table of Contents: http://asa.scitation.org/toc/jas/144/4
Published by the Acoustical Society of America

Miran Oh and Yoonjeong Lee: JASA Express Letters https://doi.org/10.1121/1.5057367 Published Online 17 October 2018

ACT: An Automatic Centroid Tracking tool for analyzing vocal tract actions in real-time magnetic resonance imaging speech production data

Miran Oh[a] **and Yoonjeong Lee**[b]
Department of Linguistics, University of Southern California, Los Angeles, California 90089, USA
miranoh@usc.edu, yoonjeonglee@ucla.edu

Abstract: Real-time magnetic resonance imaging (MRI) speech production data have expanded the understanding of vocal tract actions. This letter presents an Automatic Centroid Tracking tool, ACT, which obtains both spatial and temporal information characterizing multi-directional articulatory movement. ACT auto-segments an articulatory object composed of connected pixels in a real-time MRI video, by finding its intensity centroids over time and returns kinematic profiles including direction and magnitude information of the object. This letter discusses the utility of ACT, which outperforms other similar object tracking techniques, by demonstrating its successful online tracking of vertical larynx movement. ACT can be deployed generally for dynamic image processing and analysis.

[CCC]
Date Received: July 10, 2018 **Date Accepted:** September 11, 2018

1. Introduction

Speech production data acquired using real-time Magnetic Resonance Imaging (rtMRI) have expanded our understanding of the articulatory dynamics of the vocal tract (Bresch and Narayanan, 2009; Lammert *et al.*, 2010, 2013; Tilsen *et al.*, 2016). Several techniques such as grid-based analysis (Proctor *et al.*, 2010; Lammert *et al.*, 2013; Kim *et al.*, 2014), region segmentation (Silva and Teixeira, 2015; Toutios and Narayanan, 2015; Labrunie *et al.*, 2018), and region-of-interest (ROI) analysis (Lammert *et al.*, 2010, 2013; Proctor *et al.*, 2011) have been used for processing and analyzing midsagittal behavior of the speech articulators. These techniques are well suited for quantifying articulatory movements of speech constriction actions and/or locating where such actions are manifested (e.g., place and manner of articulation/constriction location and degree).

However, the above methods may not be ideal for quantifying spatial information of (non-constriction) articulatory movements. For example, ROI techniques that use aggregating measures of pixel intensity in a fixed region, which is placed in reference to a passive articulatory structure, do not track the actual spatial change (e.g., in millimeters) traversed by a specific active articulatory structure. Thus, despite its adequateness in quantifying constriction-oriented articulatory actions, ROI analysis is not optimal for quantifying the direction and magnitude of actions of articulatory structures not engaged in forming a constriction. Movements of the velum and larynx, for example, are not characterized as having a linguistically specified constriction formation and release in the vocal tract (e.g., upward and downward movements of the larynx are achieved by raising or lowering the position of the larynx from one non-specific position to another non-specific position, not by moving the larynx toward a certain specified point or "place of articulation" in the vocal tract.). That is, speech-related laryngeal actions include upward and downward movements of the larynx (e.g., associated with tones or glottalic consonants); and the velum lowers during the articulation of nasals, and exactly where the lowered velum is positioned or reaches is not crucial for the creation of nasality (nasal airflow) if the action itself is of sufficient magnitude.[1] Moreover, while region segmentation and grid-based analyses can indeed provide adequate dynamic information, previous studies with such methods have not explicitly measured vertical larynx movement (Bresch and Narayanan, 2009;

[a]Author to whom correspondence should be addressed.
[b]Present address: Department of Head and Neck Surgery, UCLA School of Medicine, Los Angeles, CA, 90095, USA.

Toutios and Narayanan, 2015; Labrunie *et al.*, 2018). In addition, the methods used in those studies involve a substantial amount of manual outlining and correction, which makes the data processing and analysis inefficient and less reliable.

In order to quantify non-constriction-based articulatory actions, this work presents a newly developed Automatic Centroid Tracking tool, ACT, deployed in MATLAB that finds time-varying pixel intensity *centroids* of a moving articulator in a rtMRI video [ACT is a MATLAB graphical user interface (GUI) available free of cost;[2] for details, see Sec. 3]. Intensity centroids are spatial positions in the image (cf. Tilsen *et al.*, 2016; Oh *et al.*, 2017), which are different from aggregating measures of pixel intensity used in ROI techniques (Lammert *et al.*, 2010, 2013). Mean pixel intensity is an abstract measure that cannot be directly converted to any geometrical units. On the other hand, an intensity centroid—the intensity-weighted average spatial position of an articulatory object[3]—represents the mean spatial location of tissue found in a defined region. The centroid value reflects the center position of the auto-segmented articulator in the region, and changes in the centroid value over time estimate the direction and magnitude of articulator motion in the selected region. Therefore, speech actions that are mainly characterized by directionality and magnitude, such as vertical larynx movement, can be better understood with centroid analysis given the spatiotemporal information it provides. The current tool, which is informed by the application of a centroid tracking method for oral constriction gestures introduced in Tilsen *et al.* (2016), is designed to track multi-directional movements of an articulatory object such as the larynx and velum. While the method used in Tilsen *et al.* (2016) calculates regions with reference to manual estimations of anatomical landmark labeling and contour drawing, our tool algorithmically segments an articulatory object, offering a more reliable basis for directional movement tracking. As an example of its application, we present the ACT tool's utility in characterizing the vertical aspect of larynx movement in rtMRI speech production data (Sec. 4) in which the locally defined region corresponds to the area of the image around the larynx. The vertical trajectory of the centroid can be interpreted as vertical larynx movement.

While there are various ways of quantifying the position and movement of the larynx in image data (Bresch and Narayanan, 2009; Honda *et al.*, 1999; Proctor *et al.*, 2013; Toutios and Narayanan, 2015; Labrunie *et al.*, 2018), using these methods can be computationally intensive and often demands a large amount of manual input by the user before automatic data processing becomes possible. For example, the method used in Honda *et al.* (1999) requires the user to manually trace visible outlines of the cricoid cartilage, and then the vertical larynx position can be identified by the rotation angle of the cricoid cartilage. In Proctor *et al.* (2013), the end points of a larynx outline are selected manually at frames that visually have the lowest and highest larynx positions to calculate vertical larynx displacement.

The study of Oh *et al.* (2017) directly compares the region segmentation method (Bresch and Narayanan, 2009; Toutios and Narayanan, 2015) with the current centroid tracking method. The results of the two methods are comparable (though greater displacement values of the same articulatory movement are observed with ACT), and the centroid analysis method performs much faster than the region segmentation method done with factor analysis. Moreover, the previous studies employ either static or non-dynamic methods that do not allow tracking of the larynx movement trajectories (Honda *et al.*, 1999; Proctor *et al.*, 2013). As such, quantifying larynx movement from dynamic vocal tract MRI has not been satisfactorily established. The present study proposes a novel method that automatically records time-varying changes in movement trajectories of the larynx from dynamic MRI data, further enabling research on dynamic relations (i.e., timing and coordination) between supralaryngeal and laryngeal articulators.

While the larynx as a highly multi-functional speech articulator seems to exhibit linguistically significant actions, giving rise to variations in voicing, pitch, tenseness, aspiration, etc., little has been known about the temporal and spatial dynamics of laryngeal movement in speech sounds (whether segmental or tonal). In large measure this is due to a dearth of appropriate imaging data and of concomitant efficient image processing methods, including direct methods of tracking and quantifying larynx movement. This paper introduces a tool that tracks objects' geometric center-points (i.e., centroids) in image sequences. The tool is highly automated, with a simple manual selection of the region in which the moving object of interest resides, and is time-efficient for processing time-series image data. In what follows, we introduce the rtMRI centroid tracking tool, exemplify how the ACT tool is used in investigating the

vertical aspect of larynx movement, and suggest broader applications in dynamic image processing and analysis.

2. rtMRI data acquisition

The data that will serve as a testbed to exemplify the ACT tool are rtMRI data of the midsagittal vocal tract and audio data simultaneously acquired using a custom rtMRI protocol as detailed in Toutios and Narayanan (2016) with an effective frame rate of 83.3 frames/s.

3. ACT for image processing and analysis

3.1 Overview

We introduce a MATLAB GUI tool that exploits a centroid tracking technique.[4] The proposed object tracking tool, ACT, is designed to automatically capture speech articulator action in rtMRI data. This tool is well suited for tracking moving objects in a specified image region, ideally in a gray scale video. ACT tracks the time-varying pixel intensity centroid of the object of interest in a manually selected ROI [currently rectangular, as in Fig. 1(a)]. Once the ROI is defined, the algorithm detects any objects that are composed of connected pixels brighter than a fixed intensity threshold within the given ROI. Then, the user selects the initial *seed*, which becomes a reference point to algorithmically find the first centroid of the object of interest among the auto-segmented objects [Fig. 1(b)]. Based on the information from the user-selected ROI and initial seed, the time-varying centroids of the object are automatically calculated for a given video file. Object segmentation and centroid tracking are processed simultaneously, thus speeding the overall tool application. The output trajectories of the centroids are useful for analyzing kinematics of articulatory objects.

3.2 Region placement

In this study, ACT is used to obtain the kinematic profiles of vertical larynx movement. A fixed rectangular ROI for the larynx is selected based on each speaker's cervical vertebra locations [e.g., between the midline of their 2nd cervical vertebra (C-2) and the bottom line of the 4th cervical vertebra (C-4); Fig. 1(a)]. In this implementation, the right side of the larynx region was always aligned with the rear pharyngeal wall, and the dimensions of the larynx ROI were the width of 3 to 4 pixels and the height of 11 to 16 pixels, which for each subject were tall enough to include the highest and lowest positions of the larynx and narrow enough to exclude most of the intruding tongue root inside the ROI. To clarify, within a specified ROI there may be other articulator objects in addition to the larynx object of interest (namely, tongue root and/or epiglottis); an initial *seed* selection informs which object to process in calculating subsequent centroids [indicated by a yellow asterisk [*] in Figs. 1(b)–1(c)]. Thus, in order to capture only the movement of the object of interest and to avoid spurious centroid weighting generated by any other intruding objects that may come into the ROI, the user needs to appropriately designate a region and a seed within the region. In cases when two articulators may make contact within the selected ROI, the centroid tracking could be erroneous, as the two distinct articulators are highly likely to be identified as a single object. To avoid such an error, a more restricted ROI, though compromising some small part of the object of interest (e.g., 1 to 2 pixel-wide horizontal portion of the larynx) over a few frames, might be appropriate. However, the effect of such a region placement on centroid tracking over a very short period of time is negligible, as such faulty fluctuations in the output trajectory are smoothed taking the neighboring values of the correctly tracked centroids into account (smoothing details described in Sec. 3.3).

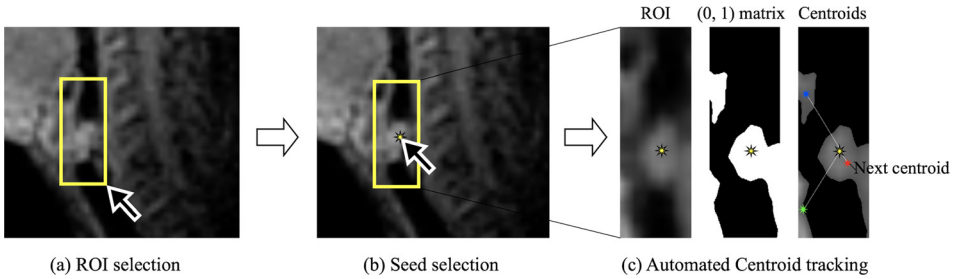

Fig. 1. (Color online) Example processing steps of object segmentation and centroid tracking in ACT (a user-selected seed is indicated by a yellow asterisk [*]).

3.3 Object segmentation and centroid tracking algorithms

Given the user-indicated ROI and seed, the centroids of the object are automatically calculated for each frame over time using the following protocol. First, a binary (0, 1) matrix is calculated in order to capture the number of objects within the ROI. Based on each pixel's intensity values in an ROI, a binary matrix is generated by assigning 1 to pixels brighter than an intensity threshold and 0 otherwise. The intensity threshold is by default defined as a 95% confidence interval of the z-score ($z > 0.8225$). This binary matrix of zeros and ones is then used to get connected components (i.e., objects) in the ROI using the flood-fill algorithm (Yapa and Koichi, 2007). In other words, in identifying the number of objects in the ROI, any two adjacent pixels assigned a 1 constitute a part of a single connected component; therefore, an object is composed of a sequence of pixels assigned "1." Then, the intensity-weighted centroid of each connected component is calculated, and the centroid that has the closest Euclidean distance from the seed is tracked as the centroid of the first frame [Fig. 1(c)]. In subsequent frames, the closest centroid from the immediately preceding centroid is set as the current centroid of a given frame. The resulting output returns both horizontal and vertical centroid values (i.e., x and y coordinates) of each frame. Finally, in order to reduce noise and faulty intensity fluctuations, the trajectories of the vertical larynx position are smoothed by a LOESS smoothing (i.e., a locally weighted scatter plot smoothing method) using a quadratic polynomial regression model with a local span of data points [30 was used for the Hausa and International Phonetic Alphabet (IPA) datasets below and 50 for the Korean dataset] (Cleveland and Devlin, 1988).

3.4 User interface and data analysis

Once the video data are loaded in the MATLAB workspace, the user is prompted to define a local region and place a seed on a mean reference image, which displays the average pixel intensity values of an entire video. By way of example in Fig. 2, we use publicly available rtMRI data of phoneticians producing the sounds of the IPA (Toutios *et al.*, 2016)[5] to illustrate how ACT tracks moving objects in a sequence of images. Figure 2 shows a sample centroid tracking analysis using ACT. The example image used is the midsagittal view of a phonetician producing a voiced velar *implosive* in a VCV sequence (/ɑgɑ/). The utility of looking at the production of a glottalic consonant is that it critically involves vertical movement of the larynx (i.e., lowering in the case of an implosive).

Figure 2 shows how ACT records changes in the vertical larynx centroid values (in millimeters) over the course of 92 image frames. Based on the user-selected ROI, which is indicated by the yellow outlined rectangles in the left figure panel, ACT returns: a zoomed-in ROI image with the online tracking of the larynx object (a light gray pixelated object), its current centroid (red dot), and previous centroids (blue dots) at each frame. As shown in the zoomed-in ROI images, based on the algorithm explained above (Sec. 3.3) only the larynx object is tracked for the centroid weight calculation, while disregarding the centroids of the other articulatory objects invading the ROI. While executing ACT, the user can visually check how the larynx centroid at each frame is tracked (i.e., the frame-by-frame completion of the blue centroid

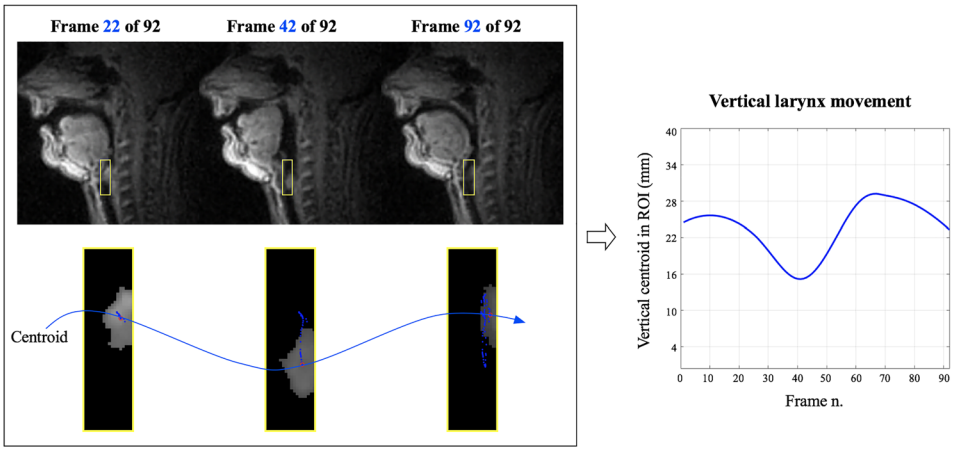

Fig. 2. (Color online) Sample visualization of ACT tracking of the vertical larynx movement during the production of /ɑgɑ/ (ROI size: 12 [width] × 40 [height] (in millimeters); rtMRI IPA data available online at http://sail.usc.edu/span/rtmri_ipa).

trajectory in the right figure panel). The resulting trajectory of vertical larynx movement that ACT generates can be saved for further analysis of inter-articulatory timing or articulatory-acoustic interface. For example, the output centroid values of the larynx (Fig. 2, right panel) can be temporally aligned with other types of time-series data such as the other co-existing articulatory signals and/or the corresponding acoustic signal.

4. Two testbeds

This section presents linguistic application examples from two sets of natural speech rtMRI data—in Seoul Korean and Hausa. These articulatory datasets serve as excellent testbeds for how the ACT tool enables a systematic linguistic analysis, as vertical larynx movement is critical to the realization of tone characterizing the contemporary Korean stop consonants and to the realization of the glottalic airstream mechanism used for the implosive and ejective consonants of Hausa.

4.1 Tense and lax stop consonants in Korean

In younger-generation speakers of Seoul Korean, the tense (aspirated /p^h/, fortis /p*/) and lax (lenis /p/, nasal /m/) stops are distinguished by high and low fundamental frequency ($f0$), respectively (Silva, 2006; Lee, 2018). The rtMRI data reported here (a subset of data presented in Lee, 2018) were obtained from three Seoul Korean speakers (2 females and 1 male) producing sentences with varying consonants (tense stops /p^h, p*/ vs lax stops /p, m/). The vertical position of the larynx at each image frame was obtained using ACT, and $f0$ during voiced intervals of the recorded speech was automatically tracked using Praat (Boersma and Weenink, 2018). Then, the $f0$ values and the corresponding vertical larynx centroid values at $f0$ peaks were analyzed to assess a correlation between the measures and to test whether the tense and lax consonants are distinguished by both $f0$ and the tracked vertical larynx movement.

Figure 3(a) presents results from one speaker (female) to exemplify ACT's utility. A strong positive correlation between the two measures is observed [$r = 0.769$ at $p < 0.05$; Fig. 3(a)]. Moreover, the tense consonants [red dots in Fig. 3(a)] are distinguished from lax consonants [blue dots in Fig. 3(a)] by their high $f0$ *and* by their larynx position values (all at $p < 0.05$). The other two speakers show similar patterns (figure not given), exhibiting the strong positive correlation between $f0$ and vertical larynx height and the similarity between the tonal patterns and vertical larynx movement (i.e., tense > lax). As such, ACT can contribute to an understanding of the articulatory mechanisms that express consonantally contrastive $f0$ organization.

4.2 Glottalic consonants in Hausa

Ejectives and implosives are reported to involve relatively rapid raising and lowering of the larynx, respectively (Maddieson and Ladefoged, 1996). Hausa has both ejectives and implosives in its consonant inventory and, as such, can serve as a testing ground for the use of ACT to elucidate the laryngeal mechanism in the production of these glottalic consonants. Although there have been previous attempts to quantify vertical larynx movement in the production of glottalic consonants, they use either "tutorial" production data from phoneticians (Shosted *et al.*, 2011) or non-speech examples, such

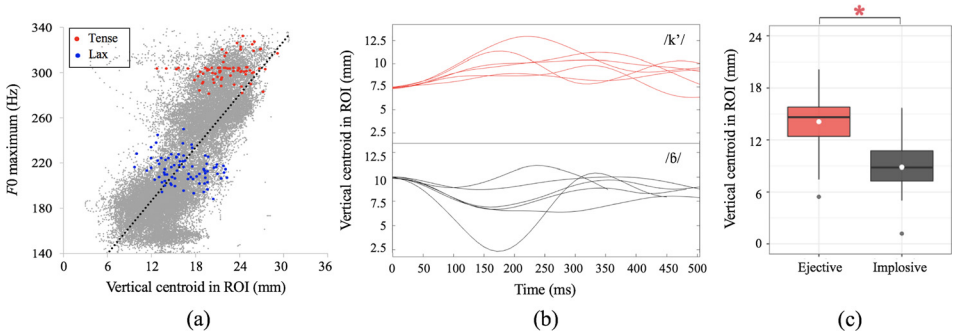

Fig. 3. (Color online) (a) Correlation between $f0$ and the corresponding vertical larynx (black dotted line: regression line, gray dots: values measured from all the voiced intervals) and $f0$ and the corresponding vertical larynx centroid values for tense (red dots) vs lax (blue dots) from a female speaker of Seoul Korean, (b) sample vertical larynx movement time functions in Hausa ejective /k'/ (red line) and implosive /ɓ/ (gray line), and (c) vertical larynx position at movement maximum in Hausa ejectives (/s', k', k$^{w'}$/) and implosives (/ɓ, ɗ/) (box heights: interquartile range [IQR], white dots: means; gray dots: outliers; horizontal lines: medians; vertical line heights: intervals between minimum and maximum values within 1.5 × IQR).

as beatboxing (Proctor *et al.*, 2013). Additionally, previous studies using an electromagnetic articulography (Shosted *et al.*, 2011) are limited to inferring larynx movement from external neck markers. The current work uses ACT and presents a sample rtMRI data analysis of glottalic consonant productions from a native Hausa speaker (female). Sentences with varying glottalic consonants such as ejectives (/s', k', $k^{w'}$/) and implosives (/ɓ, ɗ/) in a controlled vocalic and tonal frame were elicited.

Figure 3(b) shows vertical larynx movement patterns during multiple productions of ejective /k'/ (red line) and implosive /ɓ/ (dark gray line). (Trajectories are aligned by the vertical larynx *movement onset* time points, identified by an analysis algorithm that uses articulator trajectory velocity [see Tiede, 2010]). The upper panel of Fig. 3(b) shows that the larynx trajectories for ejective consonants (red line) have an upward movement. On the other hand, implosive consonants manifest the opposite patterns, i.e., larynx lowering. Figure 3(c) shows the box plots of the vertical larynx position at laryngeal movement maximum in ejectives (/s', k', $k^{w'}$/) and implosives (/ɓ, ɗ/). This indicates, as expected, that the vertical larynx position achieved for ejectives is significantly higher than that of implosives [ejectives > implosives: $t(43.93) = 7.33$, $p < 0.05$]. As such, the bidirectionality of the vertical larynx movement as well as the positional (height) differences between ejectives and implosives are sensibly, adequately, and informatively quantified by ACT.

5. Summary and future directions

The current work proposes a novel MATLAB-based application tool, ACT, that provides fast and efficient movement tracking of dynamic image data to facilitate articulatory image processing and analysis. The ACT image processing tool is shown to enable the spatiotemporal quantification of previously undocumented articulatory actions. The tool allows for the increased automating of speech movement analysis in real-time dynamic image data. ACT's intensity centroid tracking algorithm is especially advantageous for analyzing speech dynamics of non-constriction-based articulator actions, obtaining the articulatory object's temporal information (e.g., movement duration and speed) as well as spatial information (e.g., movement magnitude).

In addition to the larynx vertical actions, ACT is also well suited to quantifying other (non-constriction) articulatory movements, such as velum lowering for nasals, longitudinal laryngeal movement for voicing, pharyngeal expansion (e.g., ATR/RTR), and jaw movement. The ACT tool can be used with publicly available rtMRI data (Narayanan *et al.*, 2014; Toutios *et al.*, 2016; Sorensen *et al.*, 2017). Furthermore, ACT is not limited to analyzing speech but can be extended to quantifying non-linguistic oral behaviors such as yawning, coughing, swallowing (Zu *et al.*, 2013), and use of mouthpiece instruments (Iltis *et al.*, 2015). Last, the tool can be applicable to speech imaging data other than rtMRI, such as ultrasound (Whalen *et al.*, 2005) or x ray (Munhall *et al.*, 1995), as long as there is dynamic movement of distinguishable objects. The current version of the ACT tool has been developed with MATLAB, and its release and subsequent revisions will continue to make it accessible to a broad range of speech production and image processing laboratories.

Acknowledgments

This research was supported by NIH Grant Nos. DC003172 and DC007124. The manuscript preparation and revision were supported in part by NIH Grant No. DC01797 and NSF Grant No. IIS-1704167. We thank Dani Byrd, Louis Goldstein, Shrikanth Narayanan, Krishna Nayak, and Asterios Toutios for their support and constructive input on this project. We also thank Yongwan Lim, Tanner Sorensen, Asterios Toutios, and Colin Vaz for help with data collection. We thank the three anonymous reviewers for their constructive feedback and suggestions.

References and links

[1]Changes in velic aperture (also relevant for nasality) are appropriately measured with ROI analysis.
[2]Download from https://github.com/miranoh/ACT.
[3]An articulatory object found in MR video is an auto-segmented active articulator defined by the area of a group of connected tissues brighter than a fixed pixel intensity threshold.
[4]A poster of the preliminary version of the tool was presented in Oh *et al.* (2017).
[5]http://sail.usc.edu/span/rtmri_ipa.

Boersma, P., and Weenink, D. (**2018**). "Praat: Doing phonetics by computer (Computer program)," Version 6.0.37. http://www.praat.org/ (Last viewed February 3, 2018).
Bresch, E., and Narayanan, S. S. (**2009**). "Region segmentation in the frequency domain applied to upper airway real-time magnetic resonance images," IEEE Trans. Med. Imaging **28**(3), 323–338.

Cleveland, W. S., and Devlin, S. J. (**1988**). "Locally weighted regression: An approach to regression analysis by local fitting," J. Am. Stat. Assoc. **83**(403), 596–610.

Honda, K., Hirai, H., Masaki, S., and Shimada, Y. (**1999**). "Role of vertical larynx movement and cervical lordosis in F0 control," Lang. Speech **42**(4), 401–411.

Iltis, P. W., Frahm, J., Voit, D., Joseph, A. A., Schoonderwaldt, E., and Altenmüller, E. (**2015**). "High-speed real-time magnetic resonance imaging of fast tongue movements in elite horn players," Quantitative Imaging Med. Surg. **5**(3), 374–381.

Kim, J., Kumar, N., Lee, S., and Narayanan, S. S. (**2014**). "Enhanced airway-tissue boundary segmentation for real-time magnetic resonance imaging data," in *Proceedings of the International Seminar on Speech Production*, pp. 222–225.

Labrunie, M., Badin, P., Voit, D., Joseph, A. A., Frahm, J., Lamalle, L., Vilain, C., and Boë, L. J. (**2018**). "Automatic segmentation of speech articulators from real-time midsagittal MRI based on supervised learning," Speech Commun. **99**, 27–46.

Lammert, A. C., Proctor, M. I., and Narayanan, S. S. (**2010**). "Data-driven analysis of realtime vocal tract MRI using correlated image regions," in *INTERSPEECH*, Makuhari, Chiba, Japan, pp. 1572–1575.

Lammert, A. C., Ramanarayanan, V., Proctor, M. I., and Narayanan, S. S. (**2013**). "Vocal tract cross-distance estimation from real-time MRI using region-of-interest analysis," in *INTERSPEECH*, Lyon, France, pp. 959–962.

Lee, Y. (**2018**). "The prosodic substrate of consonant and tone dynamics," Doctoral dissertation, University of Southern California.

Maddieson, I., and Ladefoged, P. (**1996**). *The Sounds of the World's Languages* (Blackwell Publishing, Malden, MA).

Munhall, K. G., Vatikiotis-Bateson, E., and Tohkura, Y. I. (**1995**). "X-ray film database for speech research," J. Acoust. Soc. Am. **98**(2), 1222–1224.

Narayanan, S., Toutios, A., Ramanarayanan, V., Lammert, A., Kim, J., Lee, S., Nayak, K., Kim, Y. C., Zhu, Y., Goldstein, L., Byrd, D., Bresch, E., Ghosh, P., Katsamanis, A., and Proctor, M. (**2014**). "Real-time magnetic resonance imaging and electromagnetic articulography database for speech production research (TC)," J. Acoust. Soc. Am. **136**(3), 1307–1311.

Oh, M., Toutios, A., Byrd, D., Goldstein, L., and Narayanan, S. S. (**2017**). "Tracking larynx movement in real-time MRI data," J. Acoust. Soc. Am. **142**(4), 2579.

Proctor, M., Bone, D., Katsamanis, A., and Narayanan, S. S. (**2010**). "Rapid semi-automatic segmentation of real-time magnetic resonance images for parametric vocal tract analysis," in *INTERSPEECH*, Makuhari, Chiba, Japan, pp. 1576–1579.

Proctor, M., Bresch, E., Byrd, D., Nayak, K., and Narayanan, S. S. (**2013**). "Paralinguistic mechanisms of production in human 'beatboxing': A real-time magnetic resonance imaging study," J. Acoust. Soc. Am. **133**(2), 1043–1054.

Proctor, M., Lammert, A., Katsamanis, A., Goldstein, L., Hagedorn, C., and Narayanan, S. S. (**2011**). "Direct estimation of articulatory kinematics from real-time magnetic resonance image sequences," in *INTERSPEECH*, Florence, Italy, pp. 281–284.

Shosted, R. K., Carignan, C., and Rong, P. (**2011**). "Estimating vertical larynx position using EMA," in *9th International Seminar on Speech Production*, Montreal, Canada, pp. 139–146.

Silva, D. J. (**2006**). "Acoustic evidence for the emergence of tonal contrast in contemporary Korean," Phonology **23**, 287–308.

Silva, S., and Teixeira, A. (**2015**). "Unsupervised segmentation of the vocal tract from real-time MRI sequences," Comp. Speech Lang. **33**(1), 25–46.

Sorensen, T., Skordilis, Z., Toutios, A., Kim, Y. C., Zhu, Y., Kim, J., Lammert, A., Ramanarayanan, V., Goldstein, L., Byrd, D., Nayak, K., and Narayanan, S. S. (**2017**). "Database of volumetric and real-time vocal tract MRI for speech science," in *INTERSPEECH*, Stockholm, Sweden, pp. 645–649.

Tiede, M. (**2010**). "MVIEW: Multi-channel visualization application for displaying dynamic sensor movements," Haskins Laboratories.

Tilsen, S., Spincemaille, P., Xu, B., Doerschuk, P., Luh, W. M., Feldman, E., and Wang, Y. (**2016**). "Anticipatory posturing of the vocal tract reveals dissociation of speech movement plans from linguistic units," PLoS One **11**(1), e0146813.

Toutios, A., Lingala, S. G., Vaz, C., Kim, J., Esling, J., Keating, P. A., Gordon, M., Byrd, D., Goldstein, L., Nayak, K., and Narayanan, S. S. (**2016**). "Illustrating the production of the International Phonetic Alphabet sounds using fast real-time magnetic resonance imaging," in *INTERSPEECH*, San Francisco, CA, pp. 2428–2432.

Toutios, A., and Narayanan, S. S. (**2015**). "Factor analysis of vocal tract outlines derived from real-time magnetic resonance imaging data," in *International Congress of Phonetic Sciences*, Glasgow, United Kingdom.

Toutios, A., and Narayanan, S. S. (**2016**). "Advances in real-time magnetic resonance imaging of the vocal tract for speech science and technology research," APSIPA Trans. Signal Inf. Process. **5**, e6 (1-12).

Whalen, D. H., Iskarous, K., Tiede, M. K., Ostry, D. J., Lehnert-LeHouillier, H., Vatikiotis-Bateson, E., and Hailey, D. S. (**2005**). "The Haskins optically corrected ultrasound system (HOCUS)," J. Speech, Lang., Hear. Res. **48**(3), 543–553.

Yapa, R. D., and Koichi, H. (**2007**). "A connected component labeling algorithm for grayscale images and application of the algorithm on mammograms," in *Proceedings of the 2007 ACM Symposium on Applied Computing*, pp. 146–152.

Zu, Y., Narayanan, S. S., Kim, Y. C., Nayak, K., Bronson-Lowe, C., Villegas, B., Ouyoung, M., and Sinha, U. K. (**2013**). "Evaluation of swallow function after tongue cancer treatment using real-time magnetic resonance imaging: A pilot study," JAMA Otol.–Head. Neck Surg. **139**(12), 1312–1319.