Published in final edited form as:

Perspect ASHA Spec Interest Groups. 2019 December ; 4(6): 1644–1652.

doi:10.1044/2019_pers-19-00064.

Tongue Part Movement Trajectories for /r/ Using Ultrasound

Sarah Dugan^a, Sarah R. Li^b, Jack Masterson^b, Hannah Woeste^b, Neeraja Mahalingam^b,
Caroline Spencer^c, T. Douglas Mast^b, Michael A. Riley^a, Suzanne E. Boyce^c

^aDepartment of Psychology, University of Cincinnati, OH

^bDepartment of Biomedical Engineering, University of Cincinnati, OH

^cDepartment of Communication Sciences and Disorders, University of Cincinnati, OH

Abstract

Purpose: Because it shows the movement of different parts of the tongue in real time, ultrasound biofeedback therapy is a promising technology for speech research and remediation. One limitation is the difficulty of interpreting real-time ultrasound images of tongue motion. Our image processing system, TonguePART, tracks the tongue surface and allows for the acquisition of quantitative tongue part trajectories.

Method: TonguePART automatically identifies the tongue contour based on ultrasound image brightness and tracks motion of the tongue root, dorsum, and blade in real time. We present tongue part trajectory data from 2 children with residual sound errors on /r/ and 2 children with typical speech, focusing on /r/ (International Phonetic Alphabet ɹ) in the phonetic context /ɑr/. We compared the tongue trajectories to magnetic resonance images of sustained vowel /ɑ/ and /r/.

Results: Measured trajectories show larger overall displacement and greater differentiation of tongue part movements for children with typical speech during the production of /ɑr/, compared to children with residual speech sound disorders.

Conclusion: TonguePART is a fast, reliable method of tracking articulatory movement of tongue parts for syllables such as /ɑr/. It is extensible to other sounds and phonetic contexts. By tracking tongue parts, clinical researchers can investigate lingual coordination. TonguePART is suitable for real-time data collection and biofeedback. Ultrasound biofeedback therapy users may make more progress using simplified biofeedback of tongue movement.

Correspondence to Sarah Dugan: hamilsm@ucmail.uc.edu.

Disclosures

Financial: Sarah Dugan has no relevant financial interests to disclose. Sarah R. Li has no relevant financial interests to disclose. Jack Masterson has no relevant financial interests to disclose. Hannah Woeste has no relevant financial interests to disclose. Neeraja Mahalingam has no relevant financial interests to disclose. Caroline Spencer has no relevant financial interests to disclose. T. Douglas Mast has no relevant financial interests to disclose. Michael A. Riley has no relevant financial interests to disclose. Suzanne E. Boyce has no relevant financial interests to disclose.

NonFinancial: Sarah Dugan has no relevant nonfinancial interests to disclose. Sarah R. Li has no relevant nonfinancial interests to disclose. Jack Masterson has no relevant nonfinancial interests to disclose. Hannah Woeste has no relevant nonfinancial interests to disclose. Neeraja Mahalingam has no relevant nonfinancial interests to disclose. Caroline Spencer has no relevant nonfinancial interests to disclose. T. Douglas Mast has no relevant nonfinancial interests to disclose. Michael A. Riley has no relevant nonfinancial interests to disclose. Suzanne E. Boyce has no relevant nonfinancial interests to disclose.

Ultrasound biofeedback therapy (UBT) uses ultrasound imaging of the tongue in real time to assist with treatment of speech sound disorders. Primarily, the technique has been employed with children, but it has also been used with adults. In the United States, most research on UBT has focused on the population of children with residual speech sound disorders (RSSDs) whose difficulties involve articulatorily complex sounds, such as American English /r/, which involve concurrent movements of different parts of the tongue in different directions.¹ A typical UBT setup consists of an ultrasound scanner and ultrasound probe, sometimes with a method of recording video and audio data from the session. Existing motor-based speech sound treatment strategies (such as shaping, approximation, and using facilitative contexts) are commonly incorporated in UBT.² Ultrasound is used as a way to identify problems in the speech motor movements that produce sounds, as well as to see progress toward accurate movements.

One aspect of making accurate articulatory speech motor movements is tongue part differentiation or the ability to control the tongue tip/blade and tongue body independently (Gibbon, 1999; Zharkova, Hewlett, & Hardcastle, 2012). Tongue part differentiation during articulation has been related to the maturation of speech motor control (Cheng, Murdoch, Goozée, & Scott, 2007; Zharkova et al., 2012). Lack of tongue part differentiation (defined as independent movement of tongue front vs. back parts) is considered a characteristic of speech sound disorders (Gibbon, 1999; Gick et al., 2007). Part of the clinical utility of UBT is that ultrasound allows for a direct observation of the degree of differentiation of tongue gestures: The tongue dorsum and root can be simultaneously imaged, providing detail about tongue gestures during articulation that can only be indirectly recovered from the acoustic signal.

However, currently, there are challenges in the implementation of UBT for treating clinical conditions characterized by undifferentiated tongue gestures. Clients who use ultrasound technology are usually at least 8 years of age, due to the later development of some speech sounds (such as /r/) and also due to the cognitive and attentional demands of interpreting ultrasound images and translating what is perceived into an appropriate pattern of tongue motion. Most studies of RSSD treatment using UBT have shown improvement in speech sound production over the course of sessions, but the typical findings are that, (a) for most children, improvement requires a significant number of sessions, and (b) there are some children who show little or no response. While most children can be taught in a few sessions to answer questions about what their tongue is doing, they still have to absorb this knowledge and reorganize new habits of articulation. The visual display changes rapidly during speech, and the multiple regions of the tongue move quasi-independently. Interpreting these visual patterns in real time is cognitively demanding for the client and also for the clinician. Accordingly, progress toward producing target speech sounds may be protracted and uneven, which is frustrating for clinician and client alike.

¹In this article, we follow the typical clinical practice of using the Roman letter designation /r/ for the American English rhotic phoneme. The International Phonetic Association equivalent is /ɹ/.

²For examples of clinical protocols using ultrasound, see B. Bernhardt, Gick, Bacsfalvi, and Adler-Bock (2005) and the manual published through the Ultrax2020 Project (Cleland, Wrench, Lloyd, & Sugden, 2018).

A number of reasons have been suggested to explain lack of response to UBT. Among them are low motivation (M. B. Bernhardt et al., 2008; Heng, McCabe, Clarke, & Preston, 2016), poor phonological processing skills (Preston, Leece, & Maas, 2016; Preston, Maas, Whittle, Leece, & McCabe, 2016), an accompanying disorder (e.g., hearing impairment, attention-deficit/hyperactivity disorder, apraxia, dysarthria; Bacsfalvi & Bernhardt, 2011; Preston, Brick, & Landi, 2013), prosodic impairment with accompanying dysarthria (Preston, Leece, McNamara, & Maas, 2017), older age (Preston et al., 2014, 2017), poor integration of visual feedback (Preston, Maas, et al., 2016), poor visual-motor skills (Sjolie, Leece, & Preston, 2016), speech motor maturation related to young age (McAllister Byun, Hitchcock, & Swartz, 2014; Sjolie et al., 2016), and lack of stimulability for the speech sound (Preston, Leece, et al., 2016; Preston, Maas, et al., 2016; Preston et al., 2017; Sjolie et al., 2016). Because most UBT studies have a small sample size (around five to 10 participants), it is difficult to know the true proportion of clients who show nonresponse to treatment and to understand the distribution of reasons for little/no response to treatment. From the list of possible reasons why some children have low response to UBT, it can be surmised that there are significant cognitive and attentional demands associated with using the ultrasound image during speech therapy. These demands may intersect with the complexity of interpreting and responding to ultrasound biofeedback of tongue movement.

Thus, while UBT is considered a promising treatment option, currently, there are some limitations to using ultrasound during speech therapy. We propose that current clinical practice can be enhanced by a system that transforms the complex, rapid movements of the displayed tongue surface into a simpler visual display based on tracked motion of several tongue parts. This article describes a first step toward deploying such a system and is focused on development of a method to accurately track the movements of the tongue in real time using ultrasound.

Why Simplified Biofeedback About Tongue Part Movements May Improve Speech Sound Therapy Outcomes

A considerable number of studies in the field of non-speech motor control have shown that complex movements can be learned quickly when tasks are cast as “the control of perception” (Powers, 1973) or acting so as to bring about a particular perceptible outcome (Gibson, 1958), particularly through the use of interactive biofeedback to simplify and enhance perception of movement outcomes. For example, it is very difficult for most people to perform 4:3 rhythmic bimanual coordination patterns (in which the right hand completes four cycles of a circular movement while the left hand completes three cycles) when viewing their hands and thinking explicitly about controlling their movements. Yet, the pattern is easily achieved when subjects control the visual motion of two dots in a real-time, simplified biofeedback display, where the dots are mathematically transformed to move synchronously in a 1:1 pattern only when the person achieves the 4:3 hand pattern (Mechsner, Kerzel, Knoblich, & Prinz, 2001). In addition to simplifying perception of movement outcomes, such displays elicit an external focus of attention, which is known to enhance motor performance and learning by engaging automatic, implicit control processes rather than

explicit, conscious motor control, which is more effortful and error prone (Mechsner, 2004; Wulf, 2013; Wulf & Prinz, 2001; Wulf, Shea, & Lewthwaite, 2010).

The success of this approach has been demonstrated in many sensorimotor control tasks and also has been used to guide the design of interfaces for controlling complex human-machine systems (Pawlak & Vicente, 1996). It yields fast, robust learning, leading to high retention and transfer (Brenner & Smeets, 2011; Fernandez & Bootsma, 2008; Kovacs, Buchanan, & Shea, 2008, 2009; Shea, Wulf, Whitacre, & Park, 2001; Wang, Kennedy, Boyle, & Shea, 2013). For example, in Faugloire, Bardy, Merhi, and Stoffregen (2005), subjects had to produce either in-phase (0° relative phase) or antiphase (180°) rhythmic rotations of the lower leg and trunk about the ankle and hip, respectively. Rather than being instructed explicitly to perform those coordination patterns, participants were instructed to view a real-time, interactive feedback display that represented movements as a trace on a screen and move the body to draw either a positively or a negatively sloped line with the trace. That feedback was, unknown to subjects, an angle-angle diagram depicting ankle versus hip joint rotations, in which a positively sloped line corresponded to in-phase coordination and a negatively sloped line to antiphase. Subjects quickly learned to produce the ankle-hip coordination patterns by adopting an external focus on creating the desired feedback outcome (i.e., controlling what they saw in the feedback display). Those trained using this feedback produced better postural coordination faster than those who did not. Moreover, stroke patients who performed this task as a balance intervention showed long-lasting balance improvements (Varoqui, Froger, Pelissier, & Bardy, 2011).

In a parallel argument, it has been suggested that UBT helps children with RSSD learn to produce speech sounds because it externalizes the focus of attention on their movement (McAllister Byun & Hitchcock, 2012). However, although children watch their tongue movements on an external ultrasound monitor when engaged in UBT, the complexity of the ultrasound image may still present challenges, and children are typically provided instructions that nonetheless direct them to pay explicit attention to controlling their tongue rather than on controlling a simplified feedback outcome. These features may be limiting the potential of UBT as it is currently implemented.

In our work, we are translating recent innovations in biomedical image processing and motor learning theory into a new type of speech biofeedback system that will employ a simplified, interactive display. Our ultimate goal is to use this image processing system to drive an interactive, “gamified” display that transforms independent movements of the tongue parts into simplified, real-time visual feedback guiding the speaker to achieve the accurate tongue coordination pattern. In such an interface, displacement of the tip/blade, dorsum, and root of the tongue can be used to control the movement of an object in a three-dimensional game space. Such a simplified, interactive, gamified system may improve speech motor learning while enhancing speaker motivation and adherence to therapy. En route to this goal, in this article, we demonstrate how we can automatically discretize the tongue contour into regions of interest for each tongue part and track the movement of the respective parts. Accurate, real-time tongue part tracking is a prerequisite for the development and implementation of the enhanced feedback system we envision.

Method

Image Processing System: TonguePART

Our custom image processing program, Tongue- PART (Tongue Profiles with Automatic Rapid Tracking), presently is written and implemented in MathWorks, 2018b, Versions R2016a–R2018b). The program processes real-time grayscale B-mode ultrasound images captured as video sequences and is adaptable to work with any clinical ultrasound scanner. There are three processes comprising the TonguePART: tongue surface identification, tongue part delineation, and recording of trajectories for each tongue part. When employing the TonguePART for non– real-time use, such as quantifying tongue part trajectories for later analysis, identification and isolation of sound sequences can be performed as a preprocessing step.

Preprocessing for Off-Line Usage: Identification of Sound Sequences

Sound sequences (such as a syllable or a single sound) are identified manually from spectrographic data, using the acoustic record from a video-recorded data collection session. The corresponding ultrasound image frames are then extracted using the TonguePART. At present, sound sequences are identified by a researcher, but in our eventual system, this step will be performed automatically.

Tongue Surface Identification

Each image frame is cropped to isolate the ultrasound image, adjusted with a time gain compensation curve to darken regions inferior to the tongue surface, and low-pass filtered to smooth speckle. Then, the brightness maxima of each image column are located. Next, an initial point on the tongue surface is found as the local maximum nearest to a user-defined calibration point. Additional points along the tongue surface are found by first estimating positions via second-order Taylor series approximations along the anterior and posterior directions, then finding local maxima within 60×1 -pixel search windows centered on each estimated position. The process is repeated until the maxima fall below brightness thresholds defined by the user for the anterior and posterior directions.

Tongue Part Delineation

Our approach automatically divides the tongue into three regions, corresponding to three equal portions of the horizontal span of the identified tongue surface as represented on the ultrasound image. These regions broadly correspond to the tongue root, dorsum, and blade, but the division of regions depends on the tongue span visible on the ultrasound image. Ultrasound, in general, does not capture the extreme anterior and posterior regions of the tongue due to shadows associated with air and mandible below the tongue tip and the hyoid bone below the posterior part of the root, near the base of the tongue (Stone, 2005). However, using our ultrasound transducer (Siemens C6–2, with a 90° field of view), most of the tongue surface is visible. Figure 1 shows an ultrasound image frame from our Siemens ultrasound system where the root (red), dorsum (green), and blade (blue) have been automatically delineated using the TonguePART. Three reference points for displacement tracking are then defined along the tongue surface in each region, evenly spaced at 1/3, 1/2,

and 2/3 of the horizontal distance spanned by that region. Movements of these points are then tracked over multiple image frames with near real-time latency (less than 0.1 s per tracked frame).

Trajectories

In order to measure trajectory data, displacements of each tongue part are recorded from frame to frame. Tongue surface points from the previous frame are used as initial estimates for the following frame, and new points are determined from local brightness maxima within corresponding search windows. Average displacements of the three reference points in each region are normalized to the length of the speaker's tongue, represented as the distance between the middle point assigned to the blade and the middle point assigned to the root in the initial frame.

Tongue part displacements are tracked along the vertical direction for the tongue blade and dorsum, with positive displacement corresponding to an upward motion. The tongue root is tracked in a diagonal direction, with positive displacement corresponding to the root moving upward and back, as the tongue root can expand transversely with limited backward movement in the pharyngeal region (Buchaillard, Perrier, & Payan, 2009). Thus, for each tongue part, positive displacement corresponds to local narrowing of the vocal tract.

Trajectory data capture changes in position from frame to frame for each tongue region. In a trajectory, each sample represents the tongue part's normalized displacement, defined as the average displacement of the three reference points for each region, relative to that region's respective position at the beginning of the sequence for each /*ar*/ production.

Measurements

To demonstrate the type of data that can be achieved using the TonguePART image processing system, two children with RSSD with /*r*/ misarticulations and two typically speaking children were recorded producing /*r*/. This sequence was chosen for two reasons: (a) Both /*α*/ and /*r*/ involve the tongue root and thus would test the ability of our system to track that tongue part, and (b) most children with /*r*/ misarticulations can produce an /*α*/. Note that misarticulations of /*r*/ are typically transcribed as /*w*/ in word-initial position but may be transcribed as schwa, rounded schwa, /*α*/, or /*ʊ*/ in nucleus or rime positions.

Children with and without RSSD were recruited from the University of Cincinnati clinic. Demographic characteristics of children are shown in Table 1, along with results of a stimulability probe based on the study of Miccio (2002). Fifteen to 18 productions of /*ar*/ were recorded for each speaker.

For ultrasound imaging, a custom head stabilization unit was employed, following Stone's (2005) recommendations for head and transducer support systems (see Figure 2). The subject's forehead and point of chin were kept in a fixed position using the stabilization unit, which can be raised or lowered to accommodate subject height.

Data for the ultrasound image processing system were captured from a Siemens Acuson X300 Premium Edition Diagnostic Ultrasound System, using a C6–2 curved linear array

transducer with an image depth of 8 cm, 4.0 MHz, and a frame rate of 36 fps. Ultrasound images were recorded as digital video at 60 fps with 1024×768 resolution, along with an audio sampled at 48 kHz.

In order to provide a detailed articulatory picture of structures around the tongue, children were also recorded producing sustained /α/ and /r/ in a magnetic resonance imaging scanner (Phillips 3T system) at the Cincinnati Children's Hospital Medical Center. Children were recorded sustaining these sounds in supine position for 13 s. These midsagittal magnetic resonance (MR) images provide vocal tract information that cannot be seen with ultrasound.

Data Analysis

Data from the ultrasound recordings were processed by our image processing program, TonguePART, which calculated trajectories of identified tongue parts. To ensure that only reliably tracked trajectories were analyzed, the mean-square value of all frame-to-frame differences in normalized displacement was computed for each production. Productions were excluded from further analysis if this mean-square value fell below a threshold of 0.001, which was observed to occur only in the presence of large, artifactual frame-to-frame jumps in displacement due to tracking errors. This exclusion criterion resulted in a total of nine to 18 analyzed productions for each subject. Tongue part trajectories were interpolated to a standard length of 39 samples, equal to the number of frames in the longest measured production of the syllable /αr/ (39 frames). The mean and standard deviation of each tongue part's displacement were calculated at each time point from these interpolated trajectories.

To investigate tongue part differentiation, trajectories for the three possible pairwise combinations of tongue parts (root vs. dorsum, dorsum vs. blade, and root vs. blade) were cross-correlated. Trajectories from all analyzed productions were concatenated for each tongue part from each speaker. The Pearson *r* correlation coefficient was then computed between the concatenated trajectories (root vs. dorsum, root vs. blade, and dorsum vs. blade) of the same speaker. Similarly, Stone, Epstein, and Iskarous (2004) used Pearson correlation coefficients to demonstrate functional independence of different tongue regions at two points in time, where correlation coefficients of .7 were considered strong and indicative of low independence. Here, we have instead assessed the related concept of tongue part differentiation, based on correlations between concatenated trajectories of different tongue parts.

Results

Figure 3 shows average trajectories of blade, root, and dorsum between /α/ and /r/ for syllable /αr/. Corresponding MR images of /α/ and /r/ display the configuration of the tongue for each isolated phoneme.

Three accurate /r/ tongue shapes are presented in Figure 3, provided by Children 2, 3, and 4. The /r/ tongue shape is shown in the bottom MR image for each child. American English /r/ can be produced by diverse tongue shapes, but many accounts collapse these articulatory variations into a single opposition between “bunched” and “retroflex.” Following these categories, Child 2 has a tongue shape for /r/ that is classically considered bunched, with a

lower tongue blade. Children 3 and 4 have shapes that are more retroflex, where the tongue tip and blade are positioned higher than the dorsum and root. Child 1 has an error tongue shape for /r/, with a high dorsum, lack of constriction from the posterior tongue root, and low tongue blade. This “humped” error shape has been noted in other children with residual errors on /r/ (Boyce, 2015).

The tongue part trajectories for Child 1 exhibit little overall displacement from phoneme to phoneme and large standard deviations. His tongue parts did not appear to move independently of each other; for /ar/, there was a very minimal increase in displacement for all tongue parts. This is consistent with the MR images of his /a/ versus his /r/; the tongue tip and blade move slightly forward for /r/, but there is very little movement of the tongue root or dorsum. In terms of acoustic percept, his /r/ error presents as a derhoticized distortion (transcribed /ə/; Shriberg, Flipsen, Karlsson, & McSweeny, 2001). His lack of tongue root movement contributes to this derhoticized quality (Boyce, Hamilton, & Rivera-Campos, 2016). Child 2, who also had a history of residual sound errors, was able to produce a perceptually accurate /r/ after his course of therapy, which is displayed in his /r/ MR image. However, he was unable to extend his success producing the phoneme /r/ to the /ar/ syllabic context. His /a/ is distorted in this context, sounding more like /r/ than /a/, which is evident in the minimal displacement seen in his tongue part trajectories measured by ultrasound. However, he was able to produce an accurate /a/ in isolation, as shown in his MR image.

Correlation coefficients between trajectories for the three delineated tongue parts are shown in Figure 4. Large negative correlations indicate tongue parts that are moving in opposing directions (i.e., differentiation), large positive correlations reflect tongue parts moving together in the same direction (undifferentiated displacements), and negligible or small correlations indicate a lack of coordination between the compared tongue parts. Both typically developing (TD) children exhibited large negative correlation of dorsum and blade, as well as of root and blade, indicating substantial tongue part differentiation. This differentiated movement is also evident from the trajectories shown in Figure 3, where both Children 3 and 4 from the TD group showed a large positive displacement in the tongue blade from /a/ to /r/, while the tongue root and dorsum moved in the opposite direction. In contrast, both children from the RSSD group showed lower magnitude correlation coefficients between the root versus dorsum and root versus blade trajectories compared to the TD group, indicating a relative lack of differentiation and thus lack of independent movement in these tongue parts.

Discussion

TonguePART was successful at tracking tongue part movements. This is important because the ability to track tongue part trajectories using ultrasound yields noninvasive, real-time data on lingual coordination during speech.

Besides being crucial for eventual use in a simplified, gamified biofeedback system, tongue part trajectories tracked with the TonguePART can be used to characterize speech motor movements in typical and disordered speech. We found that tongue part trajectories recorded during the production of two phonemes in a syllable (“are”—/ar/) exhibited differences in

articulation between TD children and children with RSSD. The two children with RSSD on /r/ showed much less differentiation of tongue parts during /ɑr/ articulation, as evidenced by predominantly positive correlations of tongue part displacements. The two children with typical speech, however, showed greater differentiation of tongue parts. In particular, both children with typical speech displaced the tongue blade positively, with minimal displacement of the root.

Those results support the findings of Stavness, Gick, Derrick, and Fels (2012), in which the stability of the root when articulating /ɑ/ to /r/ was noted in simulations using a biomechanical tongue–jaw–hyoid model. It is likely that, when making a more bunched /r/, the trajectories of the blade, root, and dorsum would appear different and that these displacements may be smaller, possibly showing less differentiation. Although Child 2, who has RSSD on /r/, was able to produce an accurate bunched /r/ in isolation, he is not a good candidate to investigate trajectories of bunched tongue shapes. He was unable to produce the sound accurately with the /ɑ/ vowel, leading to trajectories with minimal tongue part displacement for /ɑr/ that look qualitatively similar to those from Child 1. Although we interpret negative correlations as evidence of greater differentiation of tongue parts, which are associated with accurate /r/ productions in our data, it would be premature to state that differentiation, as measured by correlation of tongue part movements, is a diagnostic marker for accurate /r/. It may be that speakers producing an accurate /ɑr/ using a bunched /r/ would have less tongue part differentiation than a speaker using a retroflex /r/, based on the fact that the tongue blade is displaced minimally between /ɑ/ and bunched /r/. Further research should investigate whether error tongue part movement displacement patterns cluster together in a manner distinct from accurate productions.

Greater movement variability is a characteristic of younger children (Smith & Zelaznik, 2004; Walsh & Smith, 2002; Zharkova, Hewlett, & Hardcastle, 2011) or children who have speech errors (Howell, Anderson, Bartrip, & Bailey, 2009; Nijland et al., 2002). In our small sample, our data suggest that children with RSSD exhibit more variation in tongue part displacement trajectories (larger standard deviations) across multiple productions of /ɑr/, compared to children with typical speech. Variation in trajectories is illustrated by the larger standard deviations of tongue part displacement at different points in time, represented in Figure 3 as bands of color around the average trajectories for tongue parts for each child. Children with RSSD in our sample were slightly younger than those in the TD group, which may have influenced the variation in tongue part trajectories.

Ultrasound provides a complement to previous studies of tongue dynamics that allows for safe, real-time imaging of the entire tongue surface and quantification of tongue part trajectories during speech articulation. As noted, imaging and quantification of tongue part trajectories is useful to researchers by providing a way to characterize accurate and error productions. It is also useful to clinicians by providing a way to visualize how the tongue is moving relative to a desired target pattern of movement associated with accurate production of a sound, as well as a means of quantifying improvement over the course of therapy. As described in the introduction, clinical use of UBT is currently limited by the complexity associated with children having to comprehend the ultrasound image and translate that understanding into an appropriate motor control solution. While a simplified system may not

benefit all potential users, clinical biofeedback interventions stand to benefit from a tongue tracking system such as the TonguePART that can support a simplified, interactive biofeedback interface that utilizes principles from perception–action theory and motor learning to enhance the acquisition of accurate speech motor patterns.

Ultimately, our enhanced UBT interface will allow the user to control an interactive graphical object in real time through tongue part displacements (processed, detected, and recorded via the TonguePART) and thus enhance motor learning by casting the acquisition of tongue part coordination as a matter of creating a simplified perceptual outcome in the feedback display. We are developing such a system in our ongoing work, using the TonguePART combined with the Unity programming environment (used for creating games and virtual reality applications). In Unity, the movement of a game object (such as a car) can be controlled by assigning TonguePART’s record of tongue part movements to its screen coordinates. As the tongue moves, the TonguePART captures displacements of tongue parts in real time, which will be fed directly to Unity to specify the coordinates of the game object via either interprocess communication or a Transmission Control Protocol/ Internet Protocol connection. Combined tongue part movements of the blade, dorsum, and root will control the game object’s movement in three dimensions. For example, as the tongue dorsum moves up or down, a car may deviate slightly in its trajectory as seen in the game interface. We believe there will be additional benefits to “gamifying” the interactive biofeedback system to enhance motivation and treatment adherence. The client would not see the actual ultrasound image, though it could be available to a clinician on a second screen or an optional simultaneous recording for later inspection.

Conclusion

The TonguePART system is a unique contribution to the growing field of ultrasound speech data processing. By tracking the tongue in real time, we aim to embed our system in clinical practice and provide a fast, robust, and accurate tool with which clinical researchers can gather data on the coordination of tongue parts during speech. In addition, we aim for the TonguePART to provide the basis for innovative new biofeedback tools to promote effective and fast learning of tongue part coordination to achieve accurate speech sound production.

Acknowledgments

This work was supported by the University of Cincinnati Creating Our Third Century funding support and National Institute on Deafness and Other Communication Disorders Grants R01DC-013668 (awarded to Suzanne Boyce and Doug Whalen) and R01DC017301 (awarded to Suzanne Boyce, Michael A. Riley, and T. Douglas Mast). We gratefully acknowledge the contributions of Olivia Wnek and all the participants in this study. We thank Siemens Medical Solutions USA, Inc., for making their Acuson ultrasound scanner available for this project.

References

- Bacsfalvi P, & Bernhardt BM (2011). Long-term outcomes of speech therapy for seven adolescents with visual feedback technologies: Ultrasound and electropalatography. *Clinical Linguistics & Phonetics*, 25(11–12), 1034–1043. 10.3109/02699206.2011.618236 [PubMed: 22106893]
- Bernhardt B, Gick B, Bacsfalvi P, & Adler-Bock M (2005). Ultrasound in speech therapy with adolescents and adults. *Clinical Linguistics & Phonetics*, 19(6–7), 605–617. 10.1080/02699200500114028 [PubMed: 16206487]

- Bernhardt MB, Bacsfalvi P, Adler-Bock M, Shimizu R, Cheney A, Giesbrecht N, ... Radanov B (2008). Ultrasound as visual feedback in speech habilitation: Exploring consultative use in rural British Columbia, Canada. *Clinical Linguistics & Phonetics*, 22(2), 149–162. 10.1080/02699200701801225 [PubMed: 18253873]
- Boyce SE (2015). The articulatory phonetics of /r/ for residual speech errors. *Seminars in Speech and Language*, 36(4), 257–270. 10.1055/s-0035-1562909 [PubMed: 26458201]
- Boyce SE, Hamilton SM, & Rivera-Campos A (2016). Acquiring rhoticity across languages: An ultrasound study of differentiating tongue movements. *Clinical Linguistics & Phonetics*, 30(3–5), 174–201. 10.3109/02699206.2015.1127999 [PubMed: 26913954]
- Brenner E, & Smeets JBJ (2011). Quickly ‘learning’ to move optimally. *Experimental Brain Research*, 213(1), 153–161. 10.1007/s00221-011-2786-9 [PubMed: 21748334]
- Buchallard S, Perrier P, & Payan Y (2009). A biomechanical model of cardinal vowel production: Muscle activations and the impact of gravity on tongue positioning. *The Journal of the Acoustical Society of America*, 126(4), 2033–2051. 10.1121/1.3204306 [PubMed: 19813813]
- Cheng HY, Murdoch BE, Goozée JV, & Scott D (2007). Electropalatographic assessment of tongue-to-palate contact patterns and variability in children, adolescents, and adults. *Journal of Speech, Language, and Hearing Research*, 50(2), 375–392. 10.1044/1092-4388(2007/027)
- Cleland J, Wrench A, Lloyd S, & Sugden E (2018). ULTRAX2020 : Ultrasound technology for optimising the treatment of speech disorders: Clinicians’ resource manual. Glasgow, United Kingdom: University of Strathclyde 10.15129/63372
- Faugloire E, Bardy BG, Merhi O, & Stoffregen TA (2005). Exploring coordination dynamics of the postural system with real-time visual feedback. *Neuroscience Letters*, 374(2), 136–141. 10.1016/j.neulet.2004.10.043 [PubMed: 15644280]
- Fernandez L, & Bootsma RJ (2008). Non-linear gaining in precision aiming: Making Fitts’ task a bit easier. *Acta Psychologica*, 129(2), 217–227. 10.1016/j.actpsy.2008.06.001
- Gibbon FE (1999). Undifferentiated lingual gestures in children with articulation/phonological disorders. *Journal of Speech, Language, and Hearing Research*, 42(2), 382–397. 10.1044/jslhr.4202.382
- Gibson JJ (1958). Visually controlled locomotion and visual orientation in animals. *British Journal of Psychology*, 49, 182–194. 10.1111/j.2044-8295.1958.tb00656.x [PubMed: 13572790]
- Gick B, Bacsfalvi P, Bernhardt BM, Oh S, Stolar S, & Wilson I (2007). A motor differentiation model for liquid substitutions in children’s speech. *Proceedings of Meetings on Acoustics*, 1(1), 1–9. 10.1121/1.2951481
- Heng Q, McCabe P, Clarke J, & Preston JL (2016). Using ultrasound visual feedback to remediate velar fronting in preschool children: A pilot study. *Clinical Linguistics & Phonetics*, 30(3–5), 382–397. 10.3109/02699206.2015.1120345 [PubMed: 26810299]
- Howell P, Anderson AJ, Bartrip J, & Bailey E (2009). Comparison of acoustic and kinematic approaches to measuring utterance-level speech variability. *Journal of Speech, Language, and Hearing Research*, 52(4), 1088–1096. 10.1044/1092-4388(2009/07-0167)
- Kovacs AJ, Buchanan JJ, & Shea CH (2008). Perceptual influences on Fitts’ law. *Experimental Brain Research*, 190(1), 99–103. 10.1007/s00221-008-1497-3 [PubMed: 18648783]
- Kovacs AJ, Buchanan JJ, & Shea CH (2009). Bimanual 1:1 with 90 degrees continuous relative phase: Difficult or easy! *Experimental Brain Research*, 193(1), 129–136. 10.1007/s00221-008-1676-2 [PubMed: 19093104]
- MathWorks. (2018b). Retrieved from <https://www.mathworks.com/content/dam/mathworks/mathworks-dot-com/support/sysreq/files/system-requirements-release-2018b-windows.pdf>
- McAllister Byun T, & Hitchcock ER (2012). Investigating the use of traditional and spectral biofeedback approaches to intervention for /r/ misarticulation. *American Journal of Speech-Language Pathology*, 21(3), 207–221. 10.1044/1058-0360(2012/11-0083) [PubMed: 22442281]
- McAllister Byun T, Hitchcock ER, & Swartz MT (2014). Retroflex versus bunched in treatment for rhotic misarticulation: Evidence from ultrasound biofeedback intervention. *Journal of Speech, Language, and Hearing Research*, 57(6), 2116–2130. 10.1044/2014_JSLHR-S-14-0034
- Mechner F (2004). A psychological approach to human voluntary movements. *Journal of Motor Behavior*, 36(4), 355–370. 10.1080/00222895.2004.11007993 [PubMed: 15695214]

- Mechsner F, Kerzel D, Knoblich G, & Prinz W (2001). Perceptual basis of bimanual coordination. *Nature*, 414(6859), 69–73. 10.1038/35102060 [PubMed: 11689944]
- Miccio AW (2002). Clinical problem solving: Assessment of phonological disorders. *American Journal of Speech-Language Pathology*, 11(3), 221–229. 10.1044/1058-0360(2002/023)
- Nijland L, Maassen B, Van der Meulen S, Gabrieëls F, Kraaimaat FW, & Schreuder R (2002). Coarticulation patterns in children with developmental apraxia of speech. *Clinical Linguistics & Phonetics*, 16, 461–483. 10.1080/02699200210159103 [PubMed: 12469451]
- Pawlak WS, & Vicente KJ (1996). Inducing effective operator control through ecological interface design. *International Journal of Human-Computer Studies*, 44(5), 653–688. 10.1006/ijhc.1996.0028
- Powers WT (1973). *Behavior: The control of perception*. New York, NY: Hawthorne.
- Preston JL, Brick N, & Landi N (2013). Ultrasound biofeedback treatment for persisting childhood apraxia of speech. *American Journal of Speech-Language Pathology*, 22(4), 627–643. 10.1044/1058-0360(2013/12-0139) [PubMed: 23813207]
- Preston JL, Leece MC, & Maas E (2016). Intensive treatment with ultrasound visual feedback for speech sound errors in childhood apraxia. *Frontiers in Human Neuroscience*, 10, 1–9. 10.3389/fnhum.2016.00440 [PubMed: 26858619]
- Preston JL, Leece MC, McNamara K, & Maas E (2017). Variable practice to enhance speech learning in ultrasound biofeedback treatment for childhood apraxia of speech: A single case experimental study. *American Journal of Speech-Language Pathology*, 26(3), 840–852. 10.1044/2017_AJSLP-16-0155 [PubMed: 28715554]
- Preston JL, Maas E, Whittle J, Leece MC, & McCabe P (2016). Limited acquisition and generalisation of rhotics with ultrasound visual feedback in childhood apraxia. *Clinical Linguistics & Phonetics*, 30(3–5), 363–381. 10.3109/02699206.2015.1052563 [PubMed: 26237652]
- Preston JL, McCabe P, Rivera-Campos A, Whittle JL, Landry E, & Maas E (2014). Ultrasound visual feedback treatment and practice variability for residual speech sound errors. *Journal of Speech, Language, and Hearing Research*, 57(6), 2102–2115. 10.1044/2014_JSLHR-S-14-0031
- Shea CH, Wulf G, Whitacre CA, & Park J-H (2001). Surfing the implicit wave. *The Quarterly Journal of Experimental Psychology: Section A*, 54(3), 841–862. 10.1080/713755993
- Shriberg LD, Flipsen P Jr., Karlsson HB, & McSweeney JL (2001). Acoustic phenotypes for speech-genetics studies: An acoustic marker for residual /s/ distortions. *Clinical Linguistics & Phonetics*, 15(8), 631–650. 10.1080/02699200110069429
- Sjolie GM, Leece MC, & Preston JL (2016). Acquisition, retention, and generalization of rhotics with and without ultrasound visual feedback. *Journal of Communication Disorders*, 64, 62–77. 10.1016/j.jcomdis.2016.10.003 [PubMed: 27973322]
- Smith A, & Zelaznik HN (2004). Development of functional synergies for speech motor coordination in childhood and adolescence. *Developmental Psychobiology*, 45(1), 22–33. 10.1002/dev.20009 [PubMed: 15229873]
- Stavness I, Gick B, Derrick D, & Fels S (2012). Biomechanical modeling of English /r/ variants. *The Journal of the Acoustical Society of America*, 131(5), 355–360. 10.1121/1.3695407
- Stone M (2005). A guide to analysing tongue motion from ultrasound images. *Clinical Linguistics & Phonetics*, 19(6–7), 455–501. 10.1080/02699200500113558 [PubMed: 16206478]
- Stone M, Epstein MA, & Iskarous K (2004). Functional segments in tongue movement. *Clinical Linguistics & Phonetics*, 18(6–8), 507–521. 10.1080/02699200410003583 [PubMed: 15573487]
- Varoqui D, Froger J, Pelissier JY, & Bardy BG (2011). Effect of coordination biofeedback on (re)learning preferred postural patterns in post-stroke patients. *Motor Control*, 15(2), 187–205. [PubMed: 21628724]
- Walsh B, & Smith A (2002). Articulatory movements in adolescents: Evidence for protracted development of speech motor control processes. *Journal of Speech, Language, and Hearing Research*, 45(6), 1119–1133. 10.1044/1092-4388(2002/090)
- Wang CY, Kennedy DM, Boyle JB, & Shea CH (2013). A guide to performing difficult bimanual coordination tasks: Just follow the yellow brick road. *Experimental Brain Research*, 230(1), 31–40. 10.1007/s00221-013-3628-8 [PubMed: 23811738]

- Wulf G (2013). Attentional focus and motor learning: A review of 15 years. *International Review of Sport and Exercise Psychology*, 6(1), 77–104. 10.1080/1750984X.2012.723728
- Wulf G, & Prinz W (2001). Directing attention to movement effects enhances learning: A review. *Psychonomic Bulletin and Review*, 8(4), 648–660. 10.3758/BF03196201 [PubMed: 11848583]
- Wulf G, Shea C, & Lewthwaite R (2010). Motor skill learning and performance: A review of influential factors. *Medical Education*, 44(1), 75–84. 10.1111/j.1365-2923.2009.03421.x [PubMed: 20078758]
- Zharkova N, Hewlett N, & Hardcastle WJ (2011). Coarticulation as an indicator of speech motor control development in children: An ultrasound study. *Motor Control*, 15(1), 118–140. 10.1123/mcj.15.1.118 [PubMed: 21339517]
- Zharkova N, Hewlett N, & Hardcastle WJ (2012). An ultrasound study of lingual coarticulation in /sV/ syllables produced by adults and typically developing children. *Journal of the International Phonetic Association*, 42(2), 193–208. 10.1017/S0025100312000060

Figure 1. Midsagittal ultrasound image of the tongue surface during /r/ articulation, with tongue parts identified automatically using the TonguePART (right side of image is anterior portion of vocal tract). Root region is red, dorsum is green, and blade is blue.

Figure 2.

Images of a head-stabilizing unit used for data collection. Panel A shows the custom three-dimensional printed cradle for the ultrasound probe, Panel B shows head positioning on top of the ultrasound probe with head restricted by the forehead strap, and Panel C shows the user seated in the head-stabilizing unit. Images shown with permission of the subject in the images. The subject is not a patient.

Figure 3.

Average tongue part trajectories for syllable /ɑr/ (with bands indicating 1 *SD*) from children with residual speech sound disorders (Child 1, Child 2) and children who have typically developing speech (Child 3, Child 4). Root displacement over time is represented in red, dorsum in green, and blade in blue. For articulatory reference, magnetic resonance images of each child sustaining vowel /ɑ/ and vowel /r/ are shown beneath the child's average trajectory of /ɑr/.

Figure 4.

Correlation of tongue part movements during /ar/ production for two children with residual speech sound disorder (RSSD; Children 1 and 2) and for two typically developing children (TD; Children 3 and 4). Negative correlations indicate tongue parts moving in opposing directions.

Table 1

Demographic and Therapy Data for Child Participants

	Group	Sex	Age	Stimulability
Child 1	RSSD	M	9;7	None
Child 2	RSSD	M	8;8	Only in isolation
Child 3	Typically developing	F	14;0	All accurate
Child 4	Typically developing	F	10;7	All accurate