INTERSPEECH 2019: Show & Tell Contribution
September 15–19, 2019, Graz, Austria

Using Ultrasound Imaging to Create Augmented Visual Biofeedback for Articulatory Practice

Colin T. Annand, Maurice Lamb, Sarah Dugan, Sarah R. Li, Hannah M. Woeste, T. Douglas Mast, Michael A. Riley, Jack A. Masterson, Neeraja Mahalingam, Kathryn J. Eary, Caroline Spencer, Suzanne Boyce, Stephanie Jackson, Anoosha Baxi, Reneé Seward

University of Cincinnati - USA

annandct@mail.uc.edu, lambmi@mail.uc.edu, hamilsm@ucmail.uc.edu, lisr@mail.uc.edu, woestehm@mail.uc.edu, masttd@ucmail.uc.edu, rileym@ucmail.uc.edu, masterjk@mail.uc.edu, mahalina@mail.uc.edu, earykn@mail.uc.edu, spenceco@mail.uc.edu, boycese@ucmail.uc.edu, jacksi@mail.uc.edu, baxiah@mail.uc.edu, sewardr@ucmail.uc.edu

Abstract

Ultrasound images of the tongue surface can be used to provide real-time visual feedback for clinical practitioners and speakers adjusting pronunciation patterns. However, rapid and complex movements of the tongue can be difficult to interpret and directly relate to desired changes. We are developing a method for simplified visual feedback controlled by efficient, real-time tracking of tongue contours in ultrasound images. Our feedback and control paradigm are briefly discussed, and video of a potential game-like biofeedback stimulus is demonstrated.

Index Terms: Ultrasound, Articulation, Feedback, Gamification, Speech sound errors.

1. Introduction

Instructive feedback based on ultrasound imaging of articulators can be an important tool in clinical speech therapy, both by providing insights to clinicians and by providing visual feedback to speakers to support communication and implementation of clinical instructions. However, due to complexity and rapidity of tongue movement, this visual feedback may only opaquely indicate to a speaker how to adjust the articulatory trajectory of an undesired movement pattern. Augmented visual feedback, with representative movement of tongue parts, may improve clinical outcomes. We demonstrate the concept of a practice game in which real-time tongue movement controls in-game elements. This augmented biofeedback game is intended to guide speakers towards desired pronunciation and acquisition of the American English rhotic /ɹ/ sound, which is often difficult for children or second-language speakers to acquire [1,2]. This proceeding briefly describes the ultrasound imaging algorithm which provides the realtime data and focuses on the biofeedback-controlled game design concept, control and clinical applicability.

2. Methodology

We have developed a real-time tracking program TonguePART (Tongue Profiles with Automatic Rapid Tracking) that provides tongue contour data from ultrasound images at 36 frames per second. The tongue contour is manually identifiable in a single frame by an input of one calibration point and two threshold brightness values, after which the algorithm automatically tracks the tongue contour based on local brightness maxima. The tongue contour is separated into three equal-length regions corresponding to the root, dorsum, and blade regions of the tongue (see Figure 1). Within the target tongue regions, three reference points are automatically defined and used to calculate the average position of each region in each ultrasound frame. Reference point position and regional displacement are calculated and transmitted to the feedback program using a TCP/IP connection to provide a real-time control signal. Because the target /ɹ/ productions may have quasi-independent motion of the root, dorsum and blade tongue regions, motion of these three regions may potentially provide a clinically relevant control signal for the biofeedback display [3,4,5].

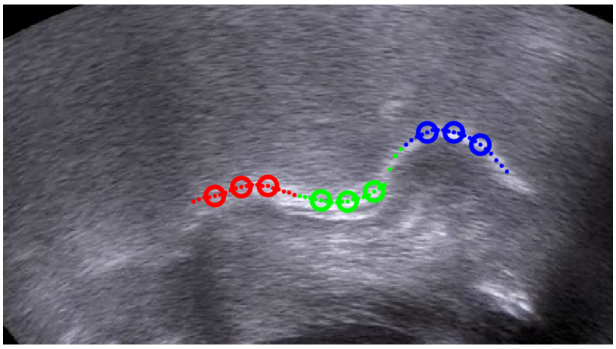

Figure 1: *Midsagittal ultrasound tongue image at the midpoint of an accurate /ɹ/ production. Average position of root (red), dorsum (green), and blade (blue) regions (identified by three circles) can be mapped to game parameters, turning the tongue contour into a controller.*

2.1. Augmented Visual Biofeedback

Our potential biofeedback format is a 'basketball shooting' paradigm, in which the real-time tongue tracking information will provide a control signal moving the basketball towards or away from a basketball hoop (Figure 2). A prototype basketball game animation was developed in Unity3D (Unity Technologies, San Francisco, CA, USA), which allows for viewing onscreen or in a virtual reality headset display. In our initial concept, the goal of the game is to make a basket, where the ball motion will be determined based on conformance to target tongue trajectories to be determined from measurements of accurate /ɹ/ production. In principle, the parameters of the application can be manipulated so that an accurate production will make a basket, whereas misarticulated productions will miss the basket. This creates a paradigm of real-time visual feedback, intended to be intuitive to a speaker.

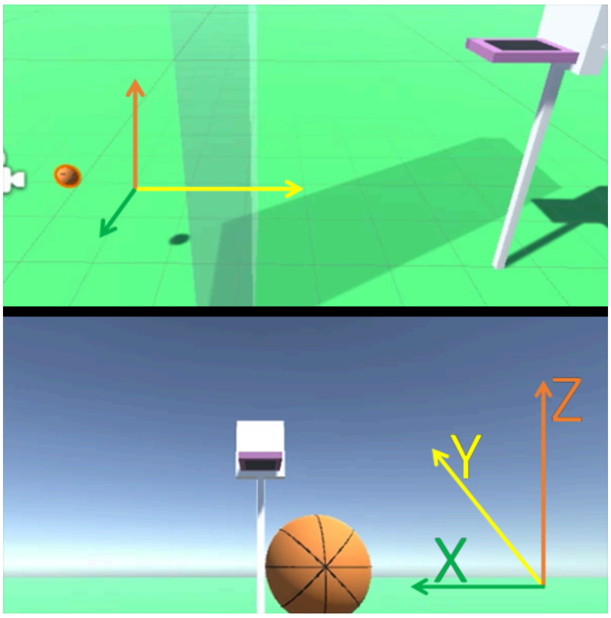

Figure 2: *(Top) Side view of play area. Semi-transparent block is used for some control schemes in which the ball can be released from user control at a specified distance from the basketball hoop. (Bottom) User view; any of the tongue regions can be assigned to control movement along the X, Y, or Z axes.*

The basketball motion and target goal are meant to provide external visual feedback, building on research in motor learning which suggests that feedback structured by an external focus of attention can enhance performance [6]. Providing relative motion information feedback allows a learner to self-organize new coordinated motion patterns through an implicit adjustment process [7]. This approach to exploring new trajectories may be useful for individuals learning sound production in acquired languages or with residual speech sound disorders. Our preliminary evidence characterizes /ɹ/ productions of the latter population as having relatively small tongue displacements compared to age matched counterparts [8]. Gradually increasing the tongue displacement needed to shoot a basket could guide RSSD speakers to slowly adjust their tongue trajectory towards a more desired movement pattern.

2.2. Demonstration and Video

The Show and Tell demonstration will feature the basketball video game concept displayed on a screen or in a virtual reality headset. Frame-by-frame data from sample /ɹ/ productions from speakers with and without RSSD will be fed into the game and produce different basketball shot trajectories in real time as the speaker attempts to produce the /ɹ/ sound. Adjusting the task parameters will change the trajectory of the basketball shot, exemplifying how a clinician could make adjustments for specific speakers, e.g. by adjusting the system parameters to require greater articulatory displacement to make the basket. This type of gamification in research may create greater motivation for practicing vocalizations, particularly for children who may otherwise not understand instructions or adjustments based on ultrasound images. A simplified goal of making a basket can be more tangible than adjustment by instructive cues that the speaker may not be able to perceptually differentiate.

2.3. Clinical Application

Previous biofeedback interventions for rhotic production have identified that optimal outcomes are reached when individuals can explore different tongue configurations and movement patterns [9]. The design goal attempts to elicit differentiated tongue part movements to produce an accurate /ɹ/ production without a focus on internal proprioception. There are two generalized types of /ɹ/ productions considered bunched and retroflex, however, even within these trajectories there can be significant variation [10]. The biofeedback concept described here would allow the shifting of control parameters to accommodate individual variations while keeping the same task parameters required for clinical intervention [11].

3. Conclusions

This paradigm attempts to enhance adjustment of articulation by supplanting cues towards complex and variant tongue movement through realtime exploration towards an externally focused goal. Future steps include implementing this concept in a real-time biofeedback system within the clinical setting.

4. Acknowledgements

This work was funded by University of Cincinnati Creating our Third Century Funding Support and by NIH/NIDCD grants 1R01DC013668-01A1 and 1R01DC017301-01.

5. References

[1] Boyce, S., Hamilton, S., Rivera Campos, A. (2016). Acquiring rhoticity across languages: An ultrasound study of differentiating tongue movements. *Clin. Ling. Phon.,* 30, 174-201.

[2] Culton, G. L. (1986). Speech disorders among college freshmen: a 13 year survey. *J. Speech Hearing Disorders.*, 51(1), 3-7.

[3] Delattre, P. C., & Freeman, D. C. (1968). A dialect study of American r's by x-ray motion picture. *Linguistics*, 44, 29-68.

[4] Zhou, X., Espy-Wilson, C. Y., Boyce, S., Tiede, M., Holland, C., & Choe, A. (2008). A magnetic resonance imaging-based articulatory and acoustic study of "retroflex" and "bunched" American English /r/. *J. Acoust. Soc. Am.*, 123, 4466-4481.

[5] Zharkova, N., Gibbon, F. E., Hardcastle, W. J. (2015). Quantifying lingual coarticulation using ultrasound imaging data collected with and without head stabilisation. *Clin. Ling. Phon.*, 29, 249-265.

[6] Wulf, G., Höß, M., & Prinz, W. (1998). Instructions for motor learning: Differential effects of internal versus external focus of attention. *Journal of Motor Behavior*, *30*(2), 169–179.

[7] Fowler, C. A., & Turvey, M. T. (1978). Skill acquisition: An event approach with special reference to searching for the optimum of a function of several variables. In G. E. Stelmach (Ed.), *Information Processing in Motor Control and Learning* (pp. 1–40). New York, NY: Academic Press.

[8] Li, S., Woeste, H., Dugan, S., Mast, T. D., Riley, M., Annand, C., Masterson, J., Mahalingam, N., Spencer, C., Boyce, S. (2019). Differentiating normal vs. misarticulated tongue trajectories from ultrasound for fast automatic articulatory biofeedback. Proceedings of the International Congress of Phonetic Sciences. (ICPhS), Melbourne, Australia

[9] Byun, T. M., Hitchcock, E. R., & Shwartz, M. T. (2014). Retroflex Versus Nunched in Treatment for Rhotic Misarticulation: Evidence From Ultrasound Biofeedback Intervention. *Journal of Speech, Language and Hearing Research, 69* (6)

[10] Alwan, A., Narayana, S., & Haker, K. (1997). Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part II. The rhotics. *The Journal of the Acoustical Society of America, 101*(2)

[11] Boyce, S. E. (2015). The Articulatory Phonetics of /r/ forResidual Speech Errors. *Seminars in Speech and Language*, *36*(4)