Lucic, I. (2016). Virtual Talking Head [Review]. In J. Levis, H. Le, I. Lucic, E. Simpson, & S. Vo (Eds). *Proceedings of the 7th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Dallas, TX, October 2015 (pp. 284-288). Ames, IA: Iowa State University.

SOFTWARE REVIEW

Virtual Talking Head

Ivana Lucic, Iowa State University

INTRODUCTION

Animation could play a potential role in improving the human learning process, especially in promoting deep understanding of the subject matter (Ahmah Zamzuri, 2013). Using more than one modality in learning enables the creation of referential connections, which facilitates learning (Zhu, Fung & Wang, 2012). The speech mapping concept has become more influential in the ways that visual modeling is done since the early 1990s. It is viewed as a useful framework for pronunciation training that provides a visual display of a spectrum of sounds produced by a particular speaker. It enables better understanding of the link between speech production and speech perception (Abry & Badin, 1996).

The three-dimensional virtual talking head was developed as a virtual anthropomorphic robot based on physical modelling of the articulatory, aerodynamic and acoustic phenomena involved in the audio-visual production of speech (Badin, Bailly & Boe, 1998). This specific model was developed in Grenoble, France at a CNRS (Centre National de la Recherche Scientifique) research unit. Even though this model is not publicly available, concepts and frameworks provided by these researchers could be used for developing a similar model to aid second language speakers in their learning of pronunciation. CNRS created the Virtual Talking Head to manipulate audio-visual speech stimuli in order to fulfill two main tasks:

(1) Evaluating and improving the learner's perception of the target language sounds,

(2) Helping the learner produce the corresponding articulations by acquiring the internalization of the relations between articulatory gestures and resulting sounds (Badin et al., 1998).

An L2 learner can be considered phonologically deaf in regard to particular sound categories, which means they are not able to distinguish speech sounds that do not belong to the phonological inventory of their L1, or they are not similar enough to the existing sound map. If there is a perceptive issue with sound recognition, production as well will most likely be problematic (Badin et al., 1998). In addition, existing research related to psychology and neuroscience shows that speech production and speech perception occur in separate paths in human brains (Skoyles, 2010). This means that examining and analyzing both speech production and speech perception as complementary skills could lead to improvement of the process of pronunciation teaching in language classrooms (Badin et al., 1998). In order for the learner to grasp proper production of perceptively

acquired sounds "they must shape their vocal tract and dynamically coordinate articulators to produce these specific acoustic targets by means of maneuvers that may be new to him/her" (Badin et al., 1998, p.1). This is similar to any type of muscle exercising: the vocal tract consists of many muscles that need to learn how to move in a different manner, and the more practice the learner provides for it, the better the acquisition and production of sounds.

HOW IT WORKS

To further explain the importance of both production and perception form pronunciation improvements, Badin et al. (1998) quotes LeBel who said that three of the "grands moyens [big means]" in the domain of phonetic correction are directly related to perception and production:

> (1) Auditory discrimination (one can pronounce well only what one can perceive well),
>
> (2) Articulatory and acoustic composition (the learning process will be more efficient if the learner knows which articulator he/she should pay attention to in order to correct a specific problem),
>
> (3) Combinatory phonetics (various coarticulation effects can be used to induce the right articulatory gestures for a given phoneme).

The space of articulator's positions, the geometric and the acoustic/auditory space, and the relations between them are implemented in a virtual talking head, which is an anthropomorphic model of speech production.

In order for the virtual talking head concept to function accurately, it was necessary to obtain complementary data from various experimental setups for various reference subjects. The subjects involved had to produce the same speech material in the same (controlled) conditions. In this way, the framework provided complete and accurate representations of the different mechanisms involved at different levels in the speech production chain and at constructing a comprehensive model (Badin et al., 1998). Some of the most important methods used to reach a valuable level of accuracy are:

> "(1) Cineradiography that produces limited but extremely valuable sets of midsagittal vocal tract contours,
>
> (2) Pneumotachometry that provides air flow at the lips and intraoral pressure, video labiometry that furnishes a geometric description of lips from front and profile views,
>
> (3) Electromagnetic articulometry that delivers the X/Y coordinates in the midsagittal plane of a few points attached the tongue or to the jaw,
>
> (4) Magnetic Resonance Imaging that results in full 3D geometric descriptions of sustained articulations" (Engwall, 2003).

The experimental setups consisted of multiple tests that needed to be performed on participating subjects, including MRI (magnetic resonance imaging), EPG (electropalatography) and EMA (electromagnetic articulography). "The shape and parameters are determined through statistical analysis of static MRI data, the parameter activation is based on the combination of MRI and EPG, and the timing of the movements is determined from EMA data" (Engwall, 2003, p. 312). Even though such creation of the model ensures reliability in recreation of virtual model sounds, the conclusion Engwall (2003) came to is that the static MRI data needed to be complemented with real-time data, in order to generate a model that is fully representative of running speech. This makes the model somewhat inconvenient. Nevertheless, the three-dimensionality of the framework provides learners with a more accurate representation of the inner processes, as it enables them to visualize how the vocal tract works when producing specific sounds.

USEFULNESS IN TEACHING

The main tasks of a teacher who uses the virtual talking head framework is to both evaluate and improve the learner's ability to perceive the vowels and consonants of the target language. Elaborating teaching strategies is another way the virtual talking head could positively influence the learner – teachers have a chance to help them find and understand the right articulatory gestures to produce what they learned to perceive (Badin et al.,1998). The virtual talking head framework can be used for generating of appropriate stimuli in order to evaluate the learner's ability to discriminate sounds in the target language, and to progressively improve the relationship between their productive and perceptive skills by helping them build the auditory map of the target language for their vocal tract practicing, starting with mapping of the L1 phonological inventory.

One of the complicated parts to implementing the virtual talking head framework into classroom environment is the necessity for teacher training. Teachers need to be educated in the area of acoustic and articulatory phonetics in order to skillfully approach learner training. It is then teacher's responsibility to combine their knowledge of the articulatory-acoustic relations to successfully guide learners during the acquisition of the appropriate articulatory gestures. The teacher can experiment with the virtual talking head in order to find the most successful facilitating strategies (Badin et al., 1998). It would be useful for instructors to include diagnostic testing of the learner target group to better understand their needs, and find room for improvement. Then, the classroom can be oriented to address specific needs of the learner group, and teacher can set reachable goals for a set amount of time to ensure productiveness. Even though such setup would require more time and resources, it would ensure good quality pronunciation practice for the learners, and a fruitful research environment for the teacher. This would facilitate more knowledge gain of how virtual talking heads function, in order to develop strategies with supporting evidence to guarantee future learner advancement. Developing a widely-available web-implemented interface would also be very beneficial in the realm of pronunciation learning (I-Chen Lin et al., 1999).

The virtual talking head offers a valuable input of audiovisual nature which, apart from facilitating motivation, helps provide for a multitude of learner types. Badin (2008)

claims that the flexibility in the features and capabilities of such model can lead to promising applications in the domain of speech therapy for speech impaired children, perception and production rehabilitation of hearing impaired children, and pronunciation training for second language learners.

CONCLUSIONS

The virtual talking head framework seems to have the characteristics necessary to become a part of essential instructional material in L2 pronunciation learning. There are many positive examples in research that show such visualization has significant contribution within the teaching of pronunciation. The virtual talking head as a pronunciation assistant, provides the practice in working memory structure via both visual and verbal channels, which minimizes the issues of limited capacity (Ahmah Zamzuri, 2013). The model presented is partially inconvenient, due to its requirement for real-time data. That makes it a time-consuming practice which requires more resources and teacher training. Nevertheless, the virtual talking head has a multitude of benefits. It could help instructors evaluate and improve learner's ability to perceive the sounds of the target language. Learners could identify and visualize their own pronunciation difficulties and, at the same time, improve the relationship between speech production and speech perception. The three-dimensional platform of the virtual talking head could result in the necessary positive impact on second language acquisition of pronunciation.

REFERENCES

Abry, C. & Badin, P. (1996). Speech mapping as a framework for an integrated approach to the sensory-motor foundation of language. *1st ETRW on Speech Production Modeling*, May 1996, Autrans, 175-184.

Ahmah Zamzuri, M. A., & Segaran, K. (2013). 3D Talking-Head Mobile App: A conceptual framework for English pronunciation learning among non-native speakers. *English Language Teaching, 6*(8), 66-76.

Badin, P., Bailly G., & Boë L. J. (1998). Towards the use of a Virtual Talking Head and of Speech Mapping tools for pronunciation training, *Proceedings of the ESCA Tutorial and Research Workshop on Speech Technology in Language Learning, 98*, 167-170.

Badin, P., Elisei, F., Bailly, G., & Tarabalka, Y. (2008). An audiovisual talking head foraugmented speech generation: Models and animations based on a real speaker'sarticulatory data. In *Articulated Motion and Deformable Objects*, Proceedings, 5098, 132–143.

Engwall, O. (2003). Combining MRI, EMA & EPG measurements in a three-dimensional tongue model. *Speech Communication, 41*, 303-329.

Lin, I. C., Hung, C. S., Yang, T. J., & Ming, O. Y. (1999). A speech driven talking head system based on a single face image. *Computer Graphics and Applications, Proceedings*. Seventh Pacific Conference Seoul, 43-49.

Skoyles, J. R. (2010). Mapping of heard speech into articulation information and speech acquisition. *Proceedings of the National Academy of Sciences of the United States of America, 107*(18), E73. http://doi.org/10.1073/pnas.1003007107

Zhu, Y., Fung, A. S. L., & Wang, H. (2012). Memorization effects of pronunciation and strokeorder animation in digital flashcards. *CALICO Journal, 29*(3), 563–577. Retrieved from http://www.jstor.org/stable/calicojournal.29.3.563