# Electromagnetic articulography in the development of "serious games" for speech rehabilitation

**M. Brandon Haworth†, Elaine Kearney‡, Melanie Baljko†, Petros Faloutsos†, Yana Yunusova‡**
† Department of Electrical Engineering and Computer Science ‡ Department of Speech-Language Pathology
York University University of Toronto
†{brandon,mb,pfal}@cse.yorku.ca,‡{elaine.kearney,yana.yunusova}@mail.utoronto.ca

## Abstract

In this paper we discuss our current research project, which is a clinical application of biomechanical modeling for the purpose of motivating behaviour change (motor speech learning) among the clients of speech intervention services. We review and discuss previously developed computer-based speech therapy approaches and then discuss the challenge of deriving clinical-relevant intervention targets. Then we discuss issues that concern our data acquisition techniques, our data processing methods, our need for patient-specific modeling, and other issues that arise at the juncture of gamification and real-time visualization.

**Keywords:** Speech Rehabilitation, Visual Feedback, Computer Games, Computer-Based Speech Therapy (CBST)

## 1. Introduction

There has been a recent surge in the creation of game-based interventions for rehabilitation of motor disorders arising after a stroke or due to various neurological conditions (e.g., Parkinson disease, cerebral palsy etc). The games often use motion tracking technologies and enhanced visual feedback to provide engaging rehabilitation experiences. Development of visualization approaches for motor speech disorders has been challenging due to lack of technologies for tracking movements of the tongue, the primary articulator. With maturing technologies such as electromagnetic articulography, tracking the tongue in real time becomes possible.

Using a motor learning paradigm, our aim is to enhance the essential auditory and somatosensory feedback with visual information. Visual information plays an important role in speech development [1] as well as speech perception and intelligibility [2, 3]. Enhancing visual information might therapeutically benefit individuals with proprioceptive and tactile deficits such as individuals post stroke and those living with Parkinson's disease (PD) [4, 5]. Additionally, we hypothesize that these visualizations might provide more motivating and memorable learning experiences, leading to efficient learning.


PMHA-14, Aug 22-23, 2014, Vancouver, BC, CA.


## 2. State of The Art

The state of the art in clinical practices for speech rehabilitation is not fully benefiting from the state of the art in speech science. Speech scientists are now taking advantage of 3D electromagnetic articulography (EMA), a fairly recently developed sensor technology, to gather large volumes of real-time data about the movement of the tongue and other speech articulators from relatively large numbers of different speakers. Without EMA or other sensor technologies, the speech articulators cannot be easily studied, since most of the articulators are hidden from view and entail millisecond-duration movements. EMA represents a leap forward in accessibility over earlier technologies that have been employed by speech scientists, such as x-ray microbeam, in terms of logistic feasibility (such as cost and the need for highly specialized expertise, staff and infrastructure). EMA systems are on the cusp of becoming feasible for clinical use, but as of yet have not been adopted. Admittedly, EMA systems are expensive, but there is reason to conjecture that their price point could be significantly lowered, given favourable market forces. EMA can be seen as a disruptive technology to extant clinical practices, since oftentimes clinical interventions are not computer-based, let alone sensor-based. Extant clinical methods, however, offer simplicity and accessibility, both in terms of equipment and clinical training. Extant methods are driven largely by speech-language pathologist (SLP) expertise, which is gained via advanced training, and whose auditory perceptions of disordered speech are a main driver of the course of a given client's clinical intervention. Visual methods include, for example, the use of a mirror for reflecting the image of the patient's visible articulators (i.e., the jaw and lips). In general, these methods rely heavily on feedback from the SLP.

### 2.1. Computer-Based Speech Therapy (CBST)

Computer-Based Speech Therapy (CBST) systems provide a means of automatically acquiring, analyzing and providing feedback for particular speech parameters. When examining prior CBST systems, it can be helpful to distinguish between *product-oriented* and *process-oriented* approaches.

In the case of speech therapy, methods that focus on the use of various organs associated with the production of speech are *process-oriented* [6], whereas methods that focus on the final output of speech (i.e., the resultant sound) are said to be *product-oriented*. Historically, CBST systems tended to be product-oriented as opposed to process-oriented, since acoustical information is more readily available (e.g., only requiring a microphone) than data about the movement of the speech articulators (which are predicated on the development of appropriate sensor technologies or highly-accurate acoustic inversion techniques).

#### *2.1.1. Product-Oriented Visual Feedback*

To be effective, product-oriented CBST systems depend on high-quality acoustic data acquisition and the selection of intervention goals that can be defined through acoustic parameters. These goals include timing, pitch and volume (or loudness).

One of the earliest of these systems provided feedback on pitch and loudness in two different scenarios [7]: one scenario entailed a game-like challenge, whereas the other was a straightforward feedback visualization. The game-like scenario was a basketball-like game with the goal to pass a ball through a wall and into a basket. The speaker could control the height of the ball, as the height was tied to the fundamental frequency (F0) of the acoustic signal that was produced by the speaker. Avoidance of the wall served to motivate the speaker to exert control over pitch. The other scenario entailed the display of an avatar, where the height of the laryngeal prominence was controlled by pitch and the size of the avatar's mouth was controlled by loudness.

Subsequent systems followed the template of this early example in that they provided visual feedback of acoustical parameters in the form of pleasing graphics or game-like challenges. SpeechViewer by IBM has a range of visualizations and scenarios for various acoustical speech parameters, as well as accompanying exercises (such as producing contrast between sounds) [8, 9]. These early systems illustrated the issue of setting reference standards during practice. In SpeechViewer, for instance, the speaker's best production would be used as the basis for setting a reference. The Indiana Speech Training Aid (ISTRA) used a speaker-dependent word recognition system [10]. The reference was formulated as a distance metric of the built-in speech recognition system. In another system, Box of Tricks, references were derived on the basis of a database of representative samples that were obtained from native speakers without speech disorders [11, 12].

#### *2.1.2. Process-Oriented Visual Feedback*

To be effective, process-oriented CBST systems depend on the acquisition of high-fidelity data about the function of speech articulators. Many process-oriented CBST systems employ visually-based anatomical models (using sagittal views or other types of cut-aways to expose the vocal tract) to provide information regarding the spatial properties of speech sounds and for user feedback. The relative role of the visual mode of feedback (on the basis of a third-person perspective of an anatomical model) in the construction of the neural representation of one's vocal tract is not known. However, the identification of the most effective mode of visual feedback is essential, given the reports of the conceptual difficulty that users have with using anatomical models (for instance, in the ARTUR system, children could not understand the representation of the palate [13]). This remains a key challenge for these systems — how to select an appropriate format for representation and feedback?

Prior systems have used a variety of representation paradigms. Anatomically correct, 3D talking head models with cut-aways, such as Baldi [14], have been developed. These systems give speakers a reference model during speech learning. The majority of existing models provide interactive references of idealized articulator behaviour during correct speech productions, but they do not provide feedback about the speaker's own articulators or production success. This approach was shown to be effective for improving the speech of children with hearing impairment, [14], however, it remains to be shown effective for motor speech impairment. Other than anatomical models, other process-oriented representations have been employed: for instance, the Box of Tricks system uses cartoon-like cross-sectional images (at least in its process-oriented mode), whereas the Speech Illumina Mentor (SIM) system employed a collection of magnetic resonance image (MRI) cross-sections and a speech recognition system to provide game-like feedback scenarios [15]. The efficacy of these approaches have not been reported.

Others systems, such as "the ARticulation TUtoR" (ARTUR) [16, 13, 17, 18] aim to provide user-specific feedback. ARTUR uses anatomical models in combination with acoustic-articulatory inversion, which is a technique for estimating speech articulator positions on the basis of acoustic signal. Acoustic-articulatory inversion is a challenging technique to perform and has a large error margin [19] as well as a non-trivial processing latency. EMA-based sensing techniques provide an alternative to acoustic-articulatory inversion in the provision of the speaker-specific feedback in the process-oriented mode.

## 3. Speech Targets

Another challenge in the development of the CBST system is the selection of appropriate speech intervention targets. Speech articulation therapy aims to improve speech intelligibility of a patient through training, structured around clinically viable goals. In the context of developing therapies, these clinical goals must first be defined on the basis of the nature of the speech 'signal' and then through the parameters that can be feasibly altered via motor learning. The selection of these goals/targets should also be theoretically

motivated and empirically validated. Because our approach is based on the acquisition of tongue motions, treatment targets of our envisioned therapy program are based on kinematic parameters of speech.

The search for speech targets/goals as elements of central motor control has a rich history in speech science. The extensive debate over the supremacy of the auditory-acoustic over the articulatory (or vice versa) information in coding speech targets [20, 21] has resulted in the prevalence of integrated models [22, 23, 24, 25]. In these integrated models, auditory as well as somatosensory information is represented inside the motor commands associated with various speech segments. Somatosensory information could be coded as target regions [26] or as various movement trajectory characteristics including its overall shape, velocity profile or curvature [27, 28, 29, 30]. All of these elements can be visualized through digital means for rehabilitative purposes.

As a first step, we consider the *articulatory working space* (AWS) as a key representative feature of the tongue's motion during speech. AWS characterizes the volume that is recruited by the tongue, represented via a point-parameterized technique, during the production of speech. AWS is known to be sensitive to disease-related changes in the speech production of individuals with motor speech disorders. Specifically, individuals with Parkinson's disease (PD) show a consistent reduction in articulatory movements across all oral articulators (jaw, lips and tongue) [31, 32], which is reflected in the reduction of individuals' AWS [33]. Thus, it is our position that AWS, as calculated over an entire utterance, provides a suitable clinical target for sentence-level motor practice in PD. There remain, however, several key empirical questions: first is the degree to which AWS can be changed through motor learning, and second is the degree to which motor learning can be enhanced via instructive visualizations of the AWS.

## 4. Experience Design

Our clinical application is predicated on the hypothesis that providing augmented visual feedback, in the form of computer games, will facilitate the expansion of the AWS in patients with PD and improve their speech intelligibility.

The promise of so-called 'gamification' approaches, which have proliferated so wildly in the last decade, has been to elicit psychological, and then subsequently, behavioural outcomes, via the mechanism of motivation as elicited by various affordances [34]. The efficacy of the myriad of affordances in eliciting behavioural outcomes have been empirically evaluated and validated, and include provisions such as points, leader-boards, levels, story/narrative, progress indications, and feedback. Evidence of the efficacy of feedback for speech motor learning is also compelling [35]. Thus, the premise of our design is positioned at a convergence point between the gamification and the motor learning research literature.

The basic template of our interactive system is as follows: (i) the system provides a visual representation of the clinical speech target to the interactant, (ii) the system cues or otherwise elicits speech production from the interactant, (iii) the interactant is presented with a visualization of his or her speech production, which provides feedback that instructively relates the elicited production to the clinical target. This 'interaction template' is embedded within a broader, possibly multi-session context, which also includes phases for training-familiarization-learning, for repetition, and for the provision of summative and/or integrative feedback. As might be clear from the outset, there are many engineering and design issues that must be addressed in computationally instantiating this system, which we discuss below. First, however, we present an overview of the system architecture.

### 4.1. System Architecture

The core processes of our therapy require a means to drive games with tongue motions. This requirement is realized through the integration of two enabling technologies. First, the Wave Speech Research System (Northern Digital Inc., Canada) that provides articulator motion tracking. Second, the Unity3D game development tool provides a means of rapidly prototyping visually rich interactive scenarios.

Our high-level system architecture models two interacting client-server relationships, one between the recording system and a middle-layer and another between this middle layer and the Unity3D engine. The middle-layer runs on the visualization system and serves as the primary link between data and visualization. This effectively creates a flow of data from tongue to game, making the tongue into a flexible game controller. An early prototype of this system was presented in [36]. Figure 1 shows an overview of the system, which has the following components.

#### *4.1.1. Clinical/Experimental User Interface (UI)*

This module allows the clinician to control the parameters of a session. Here they can select speech stimuli, design motor exercises (blocks, repetitions, feedback frequency), set therapy targets/goals and choose visualizations for feedback.

#### *4.1.2. Middle Layer*

- **Unity Plug-in**: The main interface between the middle-layer and the Unity game engine.
- **Wave Filter**: A series of filters that transform and clean the raw motion data into usable control signals.
- **Dispatcher**: A module that is responsible for distributing high level data between the middle-layer components and the external network client (i.e. the clinician).
- **Data Layer**: A module that implements the main database operations for storing traces, sessions, and patient–related information.

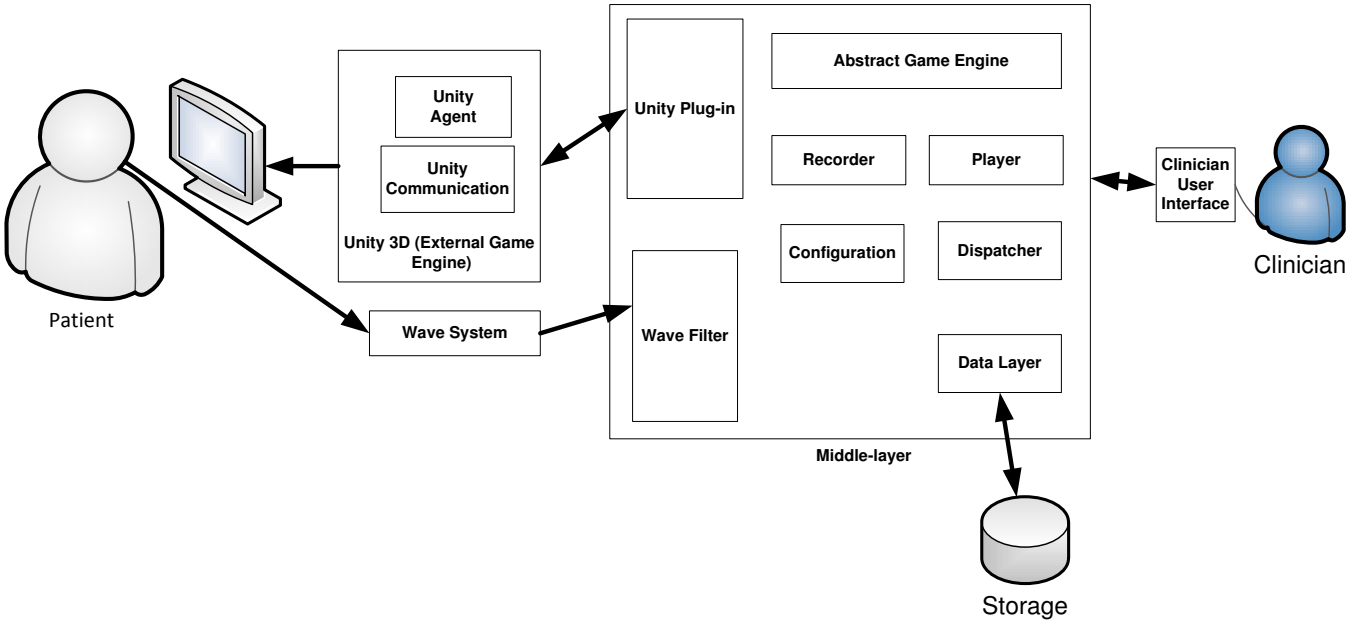



**Figure 1:** An overview of our game system framework.

- **Abstract Game Engine**: A layer that serves to make the framework game engine independent from the other middle-layer modules.
- **Configuration**: A module that configures each session on the basis of the parameters set by the clinician.
- **Recorder**: A module that records each training session and stores its relevant data in the main database.
- **Player**: A module that affords replay of any previously recorded session that is stored in the main database.

*4.1.3. Unity (External Game Engine and Development Environment)*

- **Unity Communication**: An abstraction of the communication with our middle layer.
- **Unity Agent**: A module that translates between motion data in the abstract game engine's format to Unity3D data types.

## 5. Data Acquisition and Modeling Issues

There are a number of technical issues and subtleties associated with the output of most motion tracking systems. Our system is not an exception. In this section, we describe our motion capture device and how our data processing pipeline addresses such issues.

### 5.1. Data Acquisition

Articulatory movements were captured using a 3D electromagnetic articulograph, the Wave Speech Research System. The accuracy of the Wave system has previously been reported as $< 0.5mm$ [37]. The Wave system can track the position and orientation of up to six sensors. Our sensor array is composed of a 6DOF sensor positioned on the head as well as a pair of jaw and a pair of tongue sensors. The tongue sensors are located on the tongue blade (1 cm away from the tip) and on the tongue dorsum (3 cm away from the tongue blade sensor). Data for each sensor can be captured up to a maximum frequency of $400Hz$. Since the frequency range of the tongue's motion associated with normal speech production is lower than $15Hz$ [38], a sampling frequency of $400Hz$ is sufficient for our purposes. Our current experiments, and therefore this section, focus primarily on positional data.

*5.1.1. Processing Data for Clinical Target Identification*

During early experiments we noticed that the sampling frequency of our Wave system is not uniform. The sampling period can vary by a few milliseconds from its nominal value. To address this issue, the first stage of our pipeline interpolates the recorded samples with a cubic spline, from which it then draws an equal number of uniformly spaced samples.

We also noticed in our experiments that the data indicating the position of a sensor exhibits frequencies above the range that is commonly observed in normal speech production ($> 15Hz$ [38]). The second stage of our processing pipeline removes this high frequency noise by applying a bi-directional, fifth order, low pass Butterworth filter with an empirically determined cutoff frequency of 13.5 Hz.

During long speech utterances, such as passages, we noted that spikes occurred in the data stream which were in viola-

tion of physical or biomechanics constraints. For instance, sampled positions would fall outside of the possible range as if the tongue were penetrating the hard palate. An investigation of this phenomena is still underway. Our conjecture is that the occurrence of these spike artifacts is correlated with the duration of the speech data. In the meantime, while our investigation continues, we added a third stage in the pipeline to handle this issue. The third stage in our pipeline classifies the passage data with a $2\sigma$ error ellipsoid. Intuitively, this means that all data within the ellipsoid are classified as speech data, whereas data outside the ellipsoid are considered non-speech. This ellipsoid is defined by the squared Mahalanobis distance, or generalized squared interpoint distance [39], and a constant threshold $K$:

$$(x - m_x)^T \Sigma_x^{-1} (x - m_x) \leq K \quad (1)$$

where $x$ is a data point, $m_x$ is the mean of $x$, and $\Sigma_x$ is the covariance matrix. Setting $K = 7.84$ corresponds to $95\%CI$ or $2\sigma$ [40]. All points with distance equal to or less than $K$ are on the surface or within the ellipsoid respectively.

Once processed, the data can then be examined to define speaker-specific therapy goals - in this case AWS expansion volume. In order to do this, we acquire approximately an hour worth of articulatory data captured during utterances of various speech elicitation stimuli. We examine the data off-line to make estimates of where the participant is with respect to their AWS and where they should be moved during learning.

### 5.2. Real-time Data Processing

In order to provide our clinical targets as visual feedback, real-time processing in the form of computational geometry and other mathematics typical to computer graphics takes place in the game engine.

#### *5.2.1. Head Correction*

All sensor data is head corrected in real-time so that the speaker may sit and move comfortably as they speak without changing the relative position of articulator sensors (tongue, jaw, etc). This transformation is predicated on the position and rotation information of the 6 Degree of Freedom (DOF) head reference sensor. First, a snapshot of the initial head position and orientation are taken just after sensor setup. At each frame during the session, an inverse translation and quaternion rotation are derived from the current head position and applied to each of the tracked 5 DOF articulator sensors. This effectively places the sensors in head space.

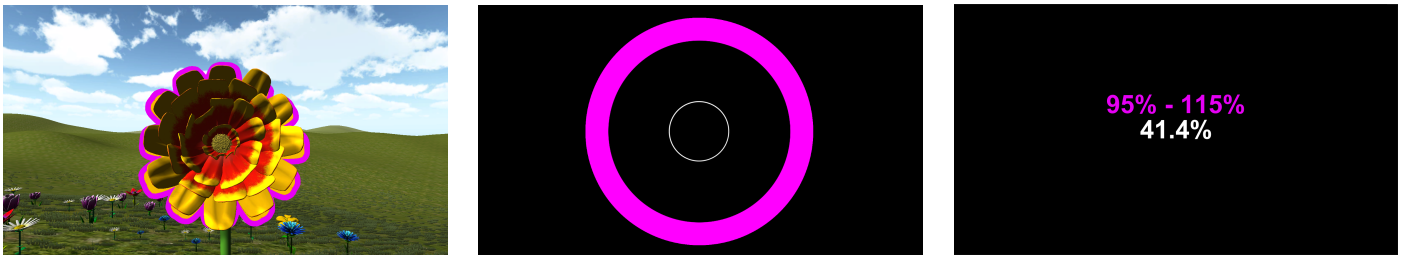
#### *5.2.2. Modeling of Speaker AWS*

In the current stage of our research project, the clinical target for the speaker is to achieve an AWS of a particular volume. The desired behavioural outcome, in terms of motor speech, is *expansion* of the AWS volume through the increase in the range of speech movements. To derive the speaker's progress with respect to our clinical target parameters, we use an accumulation of the tetrahedrons from a 3D Delaunay tetrahedralization to establish an instantaneous speaker-specific AWS model, as defined using a convex hull. This approach is effective because a Delaunay triangulation in three dimensions generates space-filling tetrahedrons whose combined free surface forms a convex hull. To do this in real time we make use of a modified version of the MIConvexHull Library [41], which implements a highly optimized version of the QuickHull algorithm. The *target* AWS is derived on a speaker-specific basis and is a proportional enlargement of the baseline AWS as defined by prior analysis (our current empirical work has demonstrated that the enlargement ratio should be tailored to each client).

## 6. Visual Feedback and Control

At this time, we are still experimenting with and empirically evaluating different visualization paradigms for both the clinical target and for the feedback. For example, we are currently focusing on three visualization variants: (i) a blooming flower (fig. 2a); (ii) a 2D circle (fig. 2b); and (iii) numerical feedback (fig. 2c). These visualizations provide increasing levels of abstract feedback on AWS expansion progress.

While our current experimental setup provides visualizations of characteristics derived from the interactant's tongue motions, such as the AWS that was utilized in the visualization scenarios, we have also shown, in prototype visualizations, that we can directly visualize the interactant's tongue motions within the volume of a playable game space as in Figure 3. For these type of visualization scenarios, there is a need to ensure that the system always visualizes the speech motions within the playable volume. To address this requirement, we have undertaken several data collection studies and have established that a profile of a speaker's range-of-motion can be effectively derived from the AWS that formed during the utterance of a common speech passage [42]. The system applies a transformation from speech movement space to game space. The transformation is derived on the basis of the component ranges, the centroid of the speaker-specific AWS, and the space constraints of the game. When scaled to a normalized space (i.e., [-1,1] in all three dimensions) the speaker-specific AWS can be used to formulate a relative control device, like an analog control stick, in three dimensions. Indeed, this characterization of an individual's overall speech space affords many control designs (for instance, it would be possible to discretize this space and to formulate regions as "buttons"). This flexibility allows for a rich set of motor speech targets and exercise designs within the framework.




(a) (b) (c)

**Figure 2:** An overview of our experimental visualizations (a) blooming flower (b) circle (c) percentage. Each visualization is an animated representation of expansion controlled by the speaker's instantaneous AWS volume. Target expansion is visualized as a bandwidth of acceptable range using high-contrast colours.

(a) (b)

**Figure 3:** An overview of various prototype tongue movement visualizations (a) pollinating bee (b) exercise ball. The pollinating bee scenario provides feedback on 3D tongue movements. The exercise ball visualizes one dimensional tongue movement (tongue elevation, or vertical movement).

## 7. Conclusion

We have presented a prototype system that provides visual feedback, in the form of interactive computer games, to patients undergoing speech therapy interventions. A key aspect of the system is its ability to track the motion of the patient's tongue in real-time, and compute suitable abstractions, such as the AWS, that can be connected to elements in a game. Our next step is to evaluate the clinical efficacy of the system with a user study at the UHN – Toronto Rehabilitation Institute.

## 8. Acknowledgments

The authors gratefully acknowledge funding and infrastructure support from multiple sources, including the NSERC Discovery Grant Program, the ASHA Foundation (New Investigator Award), the University Health Network – Toronto Rehabilitation Institute, the Parkinson Society of Canada Pilot Project Grant Program, and the Centre for Innovation in Information Visualization and Data-Driven Design (CIV-DDD).

## References

[1] P. K. Kuhl and A. N. Meltzoff, "The bimodal perception of speech in infancy." American Association for the Advancement of Science, 1982.

[2] H. McGurk and J. MacDonald, "Hearing lips and seeing voices," 1976.

[3] K. W. Grant and P.-F. Seitz, "The use of visible speech cues for improving auditory detection of spoken sentences," *The Journal of the Acoustical Society of America*, vol. 108, no. 3, pp. 1197–1208, 2000.

[4] R. Marchese, M. Diverio, F. Zucchi, C. Lentino, and G. Abbruzzese, "The role of sensory cues in the rehabilitation of parkinsonian patients: a comparison of two physical therapy protocols," *Movement Disorders*, vol. 15, no. 5, pp. 879–883, 2000.

[5] E. Seiss, P. Praamstra, C. Hesse, and H. Rickards, "Proprioceptive sensory function in parkinson's disease and huntington's disease: evidence from proprioception-related eeg

potentials," *Experimental Brain Research*, vol. 148, no. 3, pp. 308–319, 2003.

[6] D.-J. Povel and N. Arends, "The visual speech apparatus: Theoretical and practical aspects," *Speech communication*, vol. 10, no. 1, pp. 59–80, 1991.

[7] R. S. Nickerson and K. N. Stevens, "Teaching speech to the deaf: Can a computer help?" *Audio and Electroacoustics, IEEE Transactions on*, vol. 21, no. 5, pp. 445–455, 1973.

[8] H. Crepy, B. Denoix, F. Destombes, G. Rouquie, and J.-P. Tubach, "Speech processing on a personal computer to help deaf children," in *IFIP Congress*, 1983, pp. 669–671.

[9] F. R. Adams, H. Crepy, D. Jameson, and J. Thatcher, "Ibm products for persons with disabilities," in *Global Telecommunications Conference and Exhibition'Communications Technology for the 1990s and Beyond'(GLOBECOM).* IEEE, 1989, pp. 980–984.

[10] D. Kewley-Port, C. Watson, M. Elbert, D. Maki, and D. Reed, "The indiana speech training aid (istra) ii: Training curriculum and selected case studies," *Clinical Linguistics & Phonetics*, vol. 5, no. 1, pp. 13–38, 1991.

[11] K. Vicsi, P. Roach, A.-M. Öster, Z. Kacic, P. Barczikay, and I. Sinka, "Speco – a multimedia multilingual teaching and training system for speech handicapped children," in *Sixth European Conference on Speech Communication and Technology, EUROSPEECH 1999, Budapest, Hungary, September 5-9*, 1999, pp. 859–862.

[12] K. Vicsi, P. Roach, A. Öster, Z. Kacic, P. Barczikay, A. Tantos, F. Csatári, Z. Bakcsi, and A. Sfakianaki, "A multimedia, multilingual teaching and training system for children with speech disorders," *International Journal of speech technology*, vol. 3, no. 3-4, pp. 289–300, 2000.

[13] O. Engwall, O. Bälter, A.-M. Öster, and H. Kjellström, "Designing the user interface of the computer-based speech training system artur based on early user tests," *Behaviour & Information Technology*, vol. 25, no. 4, pp. 353–365, 2006.

[14] D. W. Massaro and J. Light, "Using visible speech to train perception and production of speech for individuals with hearing loss," *Journal of Speech, Language, and Hearing Research*, vol. 47, no. 2, pp. 304–320, 2004.

[15] A. J. Soleymani, M. J. McCutcheon, and M. Southwood, "Design of speech illumina mentor (sim) for teaching speech to the hearing impaired," in *Biomedical Engineering Conference, 1997., Proceedings of the 1997 Sixteenth Southern*. IEEE, 1997, pp. 425–428.

[16] O. Engwall, P. Wik, J. Beskow, and G. Granström, "Design strategies for a virtual language tutor," in *Proc. of ICSLP-04*, vol. 3, 2004, pp. 1693–1696.

[17] E. Eriksson, O. Bälter, O. Engwall, A.-M. Öster, and H. Kjellström, "Design recommendations for a computer-based speech training system based on end-user interviews," in *in proceedings of SPECOM*, 2005.

[18] O. Bälter, O. Engwall, A.-M. Öster, and H. Kjellström, "Wizard-of-oz test of artur: a computer-based speech training system with articulation correction," in *Proceedings of the 7th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, 2005, pp. 36–43.

[19] S. Panchapagesan and A. Alwan, "A study of acoustic-to-articulatory inversion of speech by analysis-by-synthesis using chain matrices and the maeda articulatory model," *The Journal of the Acoustical Society of America*, vol. 129, no. 4, pp. 2144–2162, 2011.

[20] J. S. Perkell, M. L. Matthies, M. A. Svirsky, and M. I. Jordan, "Trading relations between tongue-body raising and lip rounding in production of the vowel/u: A pilot "motor equivalence"study," *The Journal of the Acoustical Society of America*, vol. 93, no. 5, pp. 2948–2961, 1993.

[21] S. M. Nasir and D. J. Ostry, "Somatosensory precision in speech production," *Current Biology*, vol. 16, no. 19, pp. 1918–1923, 2006.

[22] F. H. Guenther, S. S. Ghosh, and J. A. Tourville, "Neural modeling and imaging of the cortical interactions underlying syllable production," *Brain and language*, vol. 96, no. 3, pp. 280–301, 2006.

[23] J. F. Houde and S. S. Nagarajan, "Speech production as state feedback control," *Frontiers in human neuroscience*, vol. 5, 2011.

[24] J. A. Jones and K. Munhall, "Remapping auditory-motor representations in voice production," *Current Biology*, vol. 15, no. 19, pp. 1768–1772, 2005.

[25] P. Perrier, L. Ma, and Y. Payan, "Modeling the production of vcv sequences via the inversion of a biomechanical model of the tongue," *arXiv preprint physics/0610170*, 2006.

[26] F. H. Guenther, "Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production." *Psychological review*, vol. 102, no. 3, p. 594, 1995.

[27] S. G. Adams, G. Weismer, and R. D. Kent, "Speaking rate and speech movement velocity profiles," *Journal of Speech, Language, and Hearing Research*, vol. 36, no. 1, pp. 41–54, 1993.

[28] S. M. Tasko and J. R. Westbury, "Speed–curvature relations for speech-related articulatory movement," *Journal of Phonetics*, vol. 32, no. 1, pp. 65–80, 2004.

[29] S. Tremblay, D. M. Shiller, and D. J. Ostry, "Somatosensory basis of speech production," *Nature*, vol. 423, no. 6942, pp. 866–869, 2003.

[30] J. H. Abbs and V. L. Gracco, "Sensorimotor actions in the control of multi-movement speech gestures," *Trends in Neurosciences*, vol. 6, pp. 391–395, 1983.

[31] Y. Yunusova, G. Weismer, J. R. Westbury, and M. J. Lindstrom, "Articulatory movements during vowels in speakers with dysarthria and healthy controls," *Journal of Speech, Language, and Hearing Research*, vol. 51, no. 3, pp. 596–611, 2008.

[32] K. Forrest, G. Weismer, and G. S. Turner, "Kinematic, acoustic, and perceptual analyses of connected speech produced by parkinsonian and normal geriatric adults," *The Journal of the Acoustical Society of America*, vol. 85, no. 6, pp. 2608–2622, 1989.

[33] G. Weismer, Y. Yunusova, and K. Bunton, "Measures to evaluate the effects of dbs on speech production," *Journal of Neurolinguistics*, vol. 25, no. 2, pp. 74–94, 2012.

[34] J. Hamari, J. Koivisto, and H. Sarsa, "Does gamification work?–a literature review of empirical studies on gamification," in *System Sciences (HICSS), 2014 47th Hawaii International Conference on*. IEEE, 2014, pp. 3025–3034.

[35] C. L. Ludlow, J. Hoit, R. Kent, L. O. Ramig, R. Shrivastav, E. Strand, K. Yorkston, and C. Sapienza, "Translating principles of neural plasticity into research on speech motor

control recovery and rehabilitation," *Journal of Speech, Language, and Hearing Research*, vol. 51, no. 1, pp. S240–S258, February 2008.

[36] M. Shtern, M. Haworth, Y. Yunusova, M. Baljko, and P. Faloutsos, "A game system for speech rehabilitation," in *Motion in Games*, ser. Lecture Notes in Computer Science, M. Kallmann and K. Bekris, Eds. Springer Berlin Heidelberg, 2012, vol. 7660, pp. 43–54.

[37] J. J. Berry, "Accuracy of the ndi wave speech research system," *Journal of Speech, Language, and Hearing Research*, vol. 54, no. 5, pp. 1295–1301, 2011.

[38] V. L. Gracco, "Analysis of speech movements: practical considerations and clinical application," *Haskins Laboratories status report on speech research SR-109/110*, pp. 45–58, 1992.

[39] R. Gnanadesikan and J. R. Kettenring, "Robust estimates, residuals, and outlier detection with multiresponse data," *Biometrics*, vol. 28, no. 1, pp. pp. 81–124, 1972. [Online]. Available: http://www.jstor.org/stable/2528963

[40] M. I. Ribeiro, "Gaussian probability density functions: Properties and error characterization," *Instituto Superior Tcnico, Lisboa, Portugal, Technical Report*, 2004.

[41] D. Sehnal and M. Campbell, *MIConvexHull Library, Version "1.0.10.1021"*. [Online]. Available: https://miconvexhull.codeplex.com/

[42] M. B. Haworth, E. Kearney, Y. Yunusova, P. Faloutsos, and M. Baljko, "Rehabilitative speech computer game calibration using empirical characterizations of articulatory working space (aws)," 2014, poster presented at 17th Biennial Motor Speech Conference, February 27 - March 2.