A 3D model of linguopalatal contact for VR biofeedback.

Chiara Bertini*, Paola Nicoli*, Niccolò Albertini*, Chiara Celata°
* Scuola Normale Superiore, Pisa
° Università degli studi di Urbino 'Carlo Bo'

In this paper we describe a 3D model of the linguopalatal contact issued from real multilevel data for the simulation in a virtual reality (VR) environment of the mechanisms underlying the production of lingual sounds. The outcome is an animation that can be experienced within a Unity 3D graphics engine, desktop or immersive environment.

Motivation: The model was developed within a project on speech motor disorders aimed at developing rehabilitation techniques based on VR visual biofeedback (Barone 2017). Modelling the spatiotemporal dynamics of linguopalatal contact is important in the context of many speech pathologies for both diagnosis and rehabilitation. Several biomechanical models of tongue movements exist that are based on mathematical models of muscular actions and their interactions (Moschos et al. 2011, Lloyd et al. 2012, Wrench & Balch 2015 among others). The model developed in this project takes an opposite kinematic perspective: the goal is to produce a patient-specific model starting from real articulatory data issued from specific experimental settings. As a matter of fact, our model exploits the information about the positioning of the tongue with respect to the palate, and this information is obtained from real data acquired by means of a digital ultrasound device for tongue imaging (UTI) and an electropalatograph (EPG) in synchronized combination (Spreafico et al. 2015, Celata et al. 2018). Furthermore, we do not model the movement of the tongue in general but, more specifically, the contact between the active (tongue) and passive (palate) articulator in the production of lingual sounds.

Data acquired: The 3D reconstruction of the palate has been obtained by acquiring and processing midsagittal and transversal ultrasound images of the speaker's palate; subsequently, the plaster cast of the palate and the artificial palate have been scanned. By superimposing the two spatial information, the 3D anatomy of the speaker's upper oral cavity has been reconstructed. An echogenic object of known size and shape (biteplane) has been used as reference for both the alignment of the virtual structures of the oral cavity and the analysis of the multilevel data obtained from different experimental sessions.
UTI data consisted in the discrete sampling over time of the mid-sagittal profile of the tongue during speech production. The Micro Speech Research Ultrasound system developed by Articulate Instruments Ltd was used, equipped with a micro-convex transducer (10mm; 5-8MHz; max FOV 150°); the software for the analysis was Articulate Assistant Advanced (AAA). At each ultrasound frame, a maximum of 42 discrete points, corresponding to the lines of sight of the ultrasound probe, are used to reconstruct the tongue midsagittal upper contour. The reading of the tongue profile data is therefore n <= 42 positions supplied as coordinates (in mm) in the x-y plane at each ultrasound frame.
EPG data consisted of binary information about presence/absence of contact (value 1 or 0) between the tongue and the 62 sensors arranged on the artificial palate worn by the speaker. This information is acquired by the WinEPG system by Articulate Instruments Ltd and analysed through the same software AAA.
Both UTI and EPG data were sampled at a frequency of 100 Hz. The data were then arranged in tabular form so as to have, for each row, the reference time point and the sequence of positions (for the UTI data) and contacts (for the EPG data).

Rigging and skinning of the model: The tongue model has been created by using chains of bones (i.e., chains of movement units for a 3D object during an animation), whose sum represents the skeleton of the virtual object (the tongue).
The skeleton was made up of 9 chains, each of which consisted of 42 bones. Each bone was positioned in the virtual space according to the spatial coordinates recorded by the articulatory instruments, and was directed towards the bone immediately in front of it. The central chain is the one responsible for receiving the positional information coming from the UTI data. When the acquired UTI data were < 42 (e.g. when the tongue is retracted and does not intersect the front rays), the missing data were filled through the Bézier interpolation algorithm. The central chain was therefore animated at each frame by the acquired UTI positional data and then transmitted this positional information to the side chains. The side chains corresponded each to a different sensor line of the artificial palate. When a sensor was contacted at a given time, the closest bone of the corresponding chain was detected and the position of the sensor was attributed to that bone. Since the same sensor was not always contacted by the same

area of the tongue, a function was created that evaluate which part of the tongue surface was most likely to contact a given sensor based on proximity.
The skeleton was first tested on a very simple polygonal skeleton (mesh), and subsequently incorporated into that of the definitive virtual model of the tongue. The skinning process was the definition of which vertices were influenced by which bone and with what weight. The automatic association provided by the software was manually corrected and validated.

Scripting and animation: The software used to virtually create the oral cavity and animate the tongue was 3ds Max 2019. A script was implemented to automate the process of acquiring data from files and creating the animation. The script is executable within the 3ds Max 2019 program via .mcr file so that it can be called up directly from the user interface.

Management of the mandibular movement: To get an estimate of the angle of mandibular rotation, light sensors were positioned at specific points of the speaker's face. Using the Intel® RealSense ™ D400 camera for facial recognition, the movements of the sensors during the production of the speech stimuli were recorded. The resolution of the camera allowed a good estimate of the angle of rotation of the jaw. These measures were inserted as correction values directly in the virtual display interface.

Visualization in a virtual environment: The digital elements necessary for the creation of the interface were the tongue, the palate and the jaw. The palate was derived from the acquisition of a real cast by laser scanning and the production of a model for Unity 3D. For the mandible model, the mesh available in Artisynth (Lloyd et al. 2012) was used. Once extracted, the polygonal object underwent a modeling process on 3DS Max. For the tongue, the Artisynth model was initially used but subsequently a new mesh was modelled according to the needs of the rigging calibration with the data obtained through the articulatory instruments. After defining the geometric structure, the display was optimized according to standard procedures.
The last phase focused on the development of the application using Unity 3D in order to obtain an immersive 3D visualization allowing interactivity and customization of parameters and animations (see figure below).

Corpus: The speech corpus used for the development of the prototype is very small and will have to be enlarged. The testing and validation were carried out on a set of 12 bisyllabic pseudo-words produced by a female speaker. These included all Italian vowels and, for the consonants, alveolar and velar stops, the lateral approximant and the alveolar trill.

Future perspectives: We will discuss the multiple possibilities for further development of the current prototype, from the introduction of an automatic system based on neural networks for the correction of the experimental noise, to gamification options to facilitate the use of the app by children. We will also show possible uses in speech therapy and speech teaching.

Barone V. (2017-2020). *Disturbi motori nel parlato e biofeedback visivo: Simulare i movimenti articolatori in 3D*. Progetto finanziato da Fondazione Pisa presso Scuola Normale Superiore di Pisa.

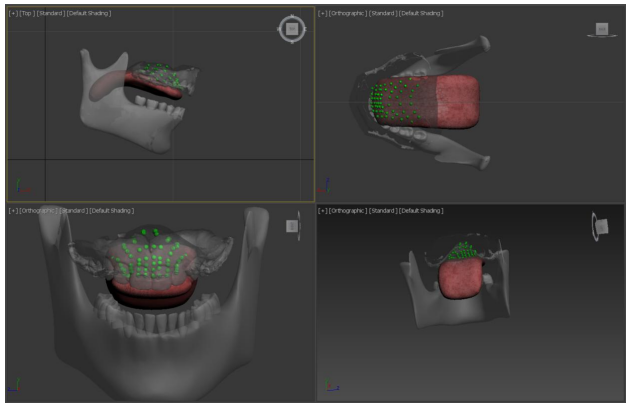
Celata / Vietti / Spreafico (2018). An articulatory account of rhotic variation in Tuscan Italian: synchronized UTI and EPG data. In *Romance Phonetics & Phonology*, Oxford University Press.

Lloyd / Stavness / Fels (2012). ArtiSynth: A fast interactive biomechanical modeling toolkit combining multibody and finite element simulation. In *Soft Tissue Biomechanical Modeling for Computer Assisted Surgery*, 355-394.

Moschos / Nikolaidis/ Pitas / Lyroudia (2011). A virtual anatomical 3D head, oral cavity and teeth model for dental and medical applications. In *Man-Machine Interactions* 2, 197-206.

Spreafico / Celata / Vietti / Bertini / Ricci (2015). An EPG + UTI study of Italian /r/. In *Proceedings of 18th International Congress of Phonetic Sciences (ICPhS)*, Glasgow, 2015.

Wrench / Balch (2015) Towards a 3D Tongue model for parameterising ultrasound data. In *Proceedings of 18th International Congress of Phonetic Sciences (ICPhS)*, Glasgow, 2015.



An instant of the animation.