Vision, Modeling, and Visualization (2015)
D. Bommes, T. Ritschel, and T. Schultz (Eds.)

Tongue S(t)imulator – A Comprehensive Parametrized Pose Model for Speech Therapy

L. Haraké[1], D. Bełtkiewicz[2] and G. Lochmann[1]

[1]Institut für Computervisualistik, Universität Koblenz-Landau, Germany
[2]Faculty of Philology, Pedagogical University of Cracow, Poland

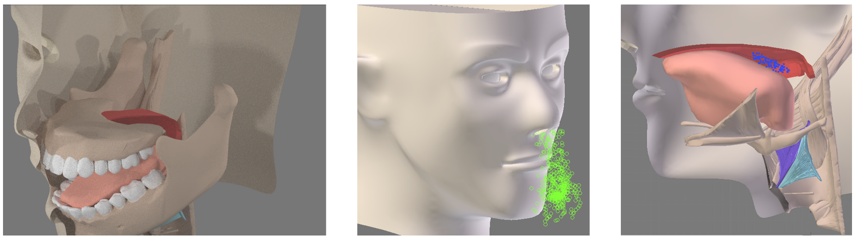


Figure 1: *Our 3D pose model serves as therapeutical tool to visualize not only mouth-interior articulators during phone production, but also as swallowing training.* ***Left:*** *Pronouncing the sound [a].* ***Middle:*** *Nasal expiration flow while pronouncing the phone [m].* ***Right:*** *Movements of involved body parts during swallowing process*

Abstract

Recent digital applications in speech therapy address patients to train auditive speech comprehension, reading or semantics, in a playful way. Virtual tutors consist of three-dimensional head models for assisting the patient with conversational exercises. However, speech therapists also have to give pronunciation instructions and motility training of the tongue very often, but only have two-dimensional drawings or their own mouths for demonstration. In this paper we propose a comprehensive application for speech therapy as a therapeutical tool, simulating the articulation of German phones including color-coded expiration flows and the deglutition process (swallowing). A three-dimensional visualization of anatomical models of pharyngo-laryngeal area can be used in an interactive way. For examining the benefits of our tool over common conventional therapy media, our approach considers iteratively the demands of speech therapists. A final expert interview was conducted to assess how the application could be involved in treatment and the application's limits.

Categories and Subject Descriptors (according to ACM CCS): I.6.8. [Simulation and Modeling]: Types of Simulation—Animation

1. Introduction

The motility and efficiency of human articulators is shaped spontaneously during the first stages of human life. Part of the articulators is mobile – tongue, lips, and soft palate with uvula (Figure 2). The motionless part includes teeth and hard palate. A well-balanced human development as well as the motility of speech organs undisturbed by external factors and not limited by anatomical defects make it possible for articulators to achieve their correct position, arrangement, and sequence of movements on the path of speech development. However, this process is often disrupted, resulting in reduced efficiency of mobile articulators and most commonly in incorrect tongue motility (the key, most mobile articulator) while pronouncing sounds, especially the ones which require achieving extreme positions.

In order to develop correct motility of speech organs, es-

DOI: 10.2312/vmv.20151266

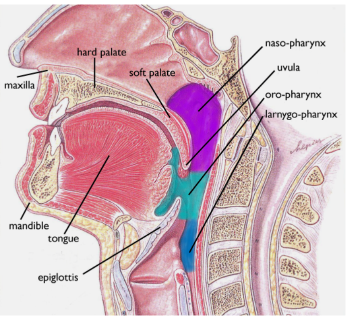


Figure 2: *Parts of the articulators in oro-pharyngo-laryngeal area used in our visualization [BSM14]*

pecially of the tongue, it is necessary to undergo regular training, and to visualize the arrangement and movement of articulators when pronouncing specific sounds. Interactive multimedia visualization can present the position of speech organs inside the oral cavity, which is not always possible for a speech therapist to show on his articulators, because the oral cavity is a closed space and it is impossible to demonstrate the movement nuances which occur inside. In addition, for presentation purposes a speech therapist often has to open his mouth as wide as possible, which is not a free arrangement of speech organs. Showing the correct positions on static sketches and illustrations prevents the demonstration of the dynamics of pronunciation and specific articulatory movements. Presentation with the traditional model also does not fully reflect the plasticity of speech organs. The presentation of how a particular sound should be pronounced (including the positions of all the areas of speech organs, air flow, and the recording of the model sound) demonstrates the full conditions of producing a sound in an interactive way. During speech acquisition, the sound model is adopted somehow intuitively through aural self-control.

For years, new technological solutions have been contributing to the development of logopedics and improvement of speech disorder prevention and speech therapy. In 2001, Józef Surowaniec argued in the therapeutic and educational context that “in the past decade, the development of information technology has been astronomical. Some even say there has been a technological gallop” [Sur01]. There are more and more cutting-edge tools, applications, and multibooks which contribute to the stimulation of language and improvement of linguistic communication. Among those tools are speech therapeutic stories for children presented in a multimedia version, which contain clear instructions on the articulatory arrangement and movement. Moreover, the multimedia form makes it possible to trigger a specific interaction between the protagonist of a speech therapeutic story (who is a kind of avatar guiding the child through the successive steps of speech improvement) and a child listener [Bel13]. New solutions are not an alternative to the traditional model of logopedics but an attractive improvement. We cannot forget about the key role of a speech therapist in leading and monitoring the progress of prevention and therapy, also with the use of new technologies.

In this paper, the Tongue S(t)imulator is an innovative tool that demonstrates how individual sounds should be articulated, with a particular focus on the tongue. The plasticity of the key mobile areas of speech organs makes it possible present a sound which a speech therapist wants to stimulate and consolidate in the patient. The presentation is based on clear animation accompanied by the model pronunciation pattern of a sound. It is also possible to implement articulatory motor (myofunctional) exercises aimed at improving the articulation of a given sound. It applies to the most motile articulator – the tongue – and to the key articulation areas in the oral cavity.

The Tongue S(t)imulator enables training during sessions with a speech therapist but also after providing the patient with appropriate instructions so that he can practice on his own everyday, which promotes regularity and determines faster and longer lasting effects.

2. Related work

A therapy meeting requires an appropriate medium for high-quality knowledge transfer and organization of the treatment plan. The variety of digital applications in a general logopedics context is enormous these days. It ranges from audiovisual tools providing exercises for children to products utilized by speech therapists for surgical interventions.

In terms of the edutainment concept, digital educational games permit an independent exercising for the patient beyond the therapy session. Often they embed video and audio clips for content transfer. The majority of those educational applications and games, such as *Revivo* or *SpeechCare* are used in therapy of language disorders. Programs particularly addressed to hearing-impaired patients, imply checking the patient’s articulation with the aid of audiovisual elements. The software *SpeechMaster* [KP06] provides a visual feedback on the correctness of articulation in Hungarian language. With an included speech recognition the phoneme-grapheme combination can be trained, where vocalization or sound volume are represented by playful elements. Profound considerations referring to audiovisual tools can be found in Kroeger et al., implemented in their software *Speech-Trainer* [KBHM10] to improve the client’s phonetic abilities. After choosing the target phoneme, its spectral extent is compared to that of the patient’s input speech. For that, the relevant features of the speech signal can be output acoustically. A two-dimensional sagittal view on vocal apparatus help to analyze the articulation process of the patient, speech data bases aid with an automatically recognition of phonetic mistakes.

Examples for non-pedagogical approaches can be found in three-dimensional, biomechanical models for medical simulations, taking real physical relations (like those of muscles) into account. The finite elements method is used as a fundamental means to approximate a complex model structure and to describe the deformation behavior of the tongue. Furthermore, the mechanics and strains of soft tissues can be represented by mass-spring systems. Perrier et al. employ a three-dimensional geometry for the speech organs which are derived from the *Visible Human Project* and adapted through MRI and CT scans [PPB*11]. Wilhelms-Tricario and Perkell introduced an approximated model of the vocal tract to simulate speech movements [WTP97], an overview on the subject is given by Wilhelms-Tricario's lecture "From Muscle Models to Tongue Models (And Back)" [WT06]. Biomechanical models can be used in surgical interventions and in predictive medicine, e.g. as a simulation of the upper respiratory tract during an examination of the voice box [RGC98]. The impact of a removal of one half of the tongue on its movement and speech production for instance, can similarly be simulated with the aid of a finite elements structure [BBPP07].

King and Pareng present a parametrized tongue model to animate its movement during formation of English vocals and consonants. The surface of the model consists of B-Spline patches, the parametrization assigns possible tongue positions and deforms its geomtry [KP01]. Ilie et al. work with a parametrized model including virtual bones to control corresponding tongue segments at five control points. The virtual skeleton complies with the geometrical features of the tongue. A mathematical model describes the transformation by determining the bones' influence on its surrounding mesh points [INS12].

The virtual tutor *Baldi* [Mas06] assists deaf and hearing-impaired children with speech exercises and provides an three-dimensional, anatomically realistic, interior view on his oral cavity during articulation. To visualize a phone, the position of each participating body part is assigned to a specific target value.

Engwall et al. present a virtual speech training system for hearing-impaired, Swedish speaking children, called *ARTUR*. In [EBOK06] the authors study demands on an audio-visual system in general. They point out the importance of the recognition of a mistaken pronunciation, aiming to give to the patient an appropriate feedback. To improve the detection of false articulation, statistical methods in speech recognition are used as well as the extraction of visual information while video recording the client's face. Then an "articulatory inversion" reconstructs the user's articulation and presents it within a three-dimensional articulation model, providing an interior view on speech organs.

Survey: functionality	*Important*	*Unnecessary*	*No opinion*
To play preselected phones	9	0	3
To input phones via microphone	7	3	2
To move the tongue manually via mouse	8	2	2
To be able to involve muscles during 'free' transformation	1	5	6
To fade in and out particular body parts separately	11	0	1

Table 1: *Number of answers in questionnaire: Which functionality is of practical importance, which can be renounced? (Multiple answers possible)*

3. Assumptions and demands

We requested the demands of twelve therapists after the implementation of a first prototype.

Generally, from the experts' point of view, a tool should be used in a simple and intuitive way, without need for a long preparation time. Also only an accurate visualization of the physiological deformations during articulation process could have the benefits over commonly used simple two-dimensional image cards. The view on the three-dimensional model should be arbitrary, parts of the model like mandible, teeth or tongue should have an option to fade out. The therapists wanted to have a list with selectable phones, each of them playing an animation of its corresponding articulation movements. The animation itself should be able to stop and repeat by the therapist.

The therapists were familiarized with a first prototype, which they had to evaluate by questionnaires, verifying if implemented functionalities suffice their demands and if not, how they suggest to refine it. Thus they had to mark those functionalities, which would be necessary to employ in a final application. Table 1 shows the number of acceptances and denials. Those functions marked neither important nor unimportant, are numbered among 'no opinion'.

The majority of therapists found it reasonable to fade out specific body parts of the head model. The functionality to play in advance selected phones from a list, gave a likewise positive feedback, only three persons had no opinion about it. Most therapists argued that it is important to input phones via microphone. So, these three functions should be well preserved in a final application. Opposite statements were made referring to the possibility to move the tongue manually: in fact half of the interviewees gave a positive feedback, but only one person favored to move the tongue depending on its muscles. In contrast, the remaining speech therapists found it unnecessary to involve tongue muscles, or had no opinion of it.

Moreover, it was asked for the worthiness of anatomical correctness for a tongue pose in connection with playing animations of phones. Here, nine of twelve therapists are of the opinion that a motion should be presented anatomically accurate. Two replied that petty anatomical variances should be allowed, if they nevertheless yield a realistic demonstration. This result already emphasizes the aspect, that hitherto materials used in speech therapy do not suffice as sample for a patient.

The following questions on questionnaire served for estimation, if and how helpful therapists picture the first prototype in therapy use, and how an improvement would look like. On a scale from *1* (not helpful) to *5* (very supportive), six persons thought of the application at that time to be supporting (*4*) their therapy sessions. Even two therapists stated it very supportive. Patients with good spatial sense would be able to obtain a visual assistance for phone pronunciation. Thereby, target poses and exercises would be more comprehensible. Especially a three-dimensional perspective with a non-restricted view on articulation organs would be an enhancement compared to common two-dimensional images.

In spite of that predominantly positive feedback (not a single therapists did *not* state the prototype to be helpful), reasons against such an application were given: In case of patients with cognitive limitations it would not provide an understanding support. Also therapy sessions at the bedside or held as home visit (stroke patients) could not involve such a tool. It was demurred that younger patients could feel the visualization to be too complicated.

Apparently, many statements of the therapists referred to fundamental problems, independent of the first prototype. We did not expect those opinions to change after implementing the final application. Also it was not our intention to tackle such aspects in our work. Other Suggestions, concerning our models and functionality, contained:

- A strict differentiation between hard and soft palate resp. uvula as being important places of articulation
- Only accurately modeled and moving lips and uvula are able to illustrate sounds that are produced with one or both lips
- The resting position of the tongue should be involved into animations. It is an important exercise of the myofunctional therapy and a requisite for correct speech and swallowing
- The possibility to adjust the animation speed for watching the movements in slow motion
- To play not just the animation of one phone, but rather to play the phone's coarticulated surrounding phones or syllables in a word. It is argued that in spontaneous speech, the phones do never occur separately, but always in connection with other phones, influencing it in its movement and pronunciation
- When entering a phone via microphone, a correctly or falsely recognized phone would give additional feedback on how to pronounce it right
- To map the deglutition process physiologically, because the infantile swallowing movements differ noticeably from those of adults. In the treatment of difficulty in swallowing its exercising is essential.

4. Tongue S(t)imulator approach

Through questionnaires we made the observation that practically very few or rather no computer tools are used in assistance during quotidian speech therapy sessions, although therapists consider them actually to be reasonable [Har14]. Thus we developed a therapeutic tool to impart both articulation movements of German language and deglutition, integrated in a three-dimensional model of oro-pharyngolaryngeal body parts. Our concept is comprehensive, given that it is applicable to many aspects in speech therapy without changing multi-modalities. The model should give the possibility to zoom in and to rotate it around its axis, so that during articulation and swallowing animations, the therapist would have a sight from every perspective upon it. Furthermore the movements of speech organs require an anatomically accurate visualization.

As graphical environment we use the modeling tool *Blender*, providing already a model import and a rendering window for a three-dimensional view, easy to adapt to the user's needs and to append new functionality. Thus we are able to afford a consistent work and visualization mode to a speech therapist, so that there is no need for him to change exercise materials. The user interface itself concentrates on the essential control elements and simple mouse interaction, assuming that usability plays a subordinated role in this work and the visualization is going to be foregrounded.

Parts of our virtual model are taken from the *Database Center for Life Science* of University of Tokyo. The head is modeled by hand and anatomically simplified; mainly lips and mandible are modeled in detail as they are essential actors in articulation process. To ensure anatomical correctness, all models were concurrently compared to medical images of human vocal tract. Each virtual body part is transformed manually to animate articulation and deglutition process to assure the overall interaction between all models.

4.1. Pose Model

Several animations can be defined for one model, where an animation follows basically Blender's shape keys concept. A shape key represents one shape transformation of particular mesh vertices at a particular time, its value generally ranges between 0 and 1 and dictates the influence on object transformation per frame relative to the basis key (initial pose).

The overall animation of the object therefore corresponds to a spatial and temporal interpolation between single shape keys (blending). Blender offers more than one interpolation mode, by default a Bézier interpolation is applied between the frames.

In order to permit smooth transformations on specific subregions of the tongue, we used a symmetrically and constantly subdivided polygon mesh (triangulated quads) that defines tongue tip, body and root. Also the tongue margins got the ability to be moved, given that they play an important role in the production of most of German phones. The use of vertex groups on those tongue regions is able to control specific tongue vertices (Figure 3(a)), additionally the tongue itself can be transformed to all necessary poses without affecting its geometry (Figure 3(b)). An important example pose is the resting tongue as being an requisite for a correct speech and swallowing (Figure 4.1).

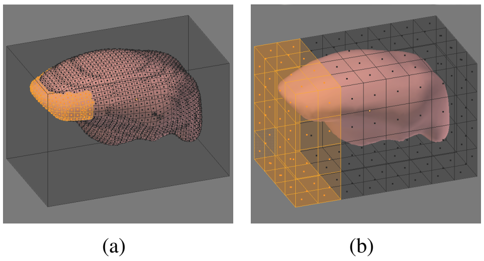

(a) (b)

Figure 3: *Two vertex groups get combined to obtain an adequate transformation of the tongue model. (a) Vertex group on polygon mesh of tongue, (b) Vertex group on mesh deform modifier*

To demonstrate an animation of the tongue, a therapist can either select a phone from a list (in a certain speed or with sound output) or transform the tongue manually in realtime via interface sliders. The list includes those German phones, that mostly make difficulties from a therapeutical point of view, such as strongly and softly spoken consonants. Moreover, a phone can be produced correctly in more than one manner, so that most of those phones are posed through two different transformations: e.g. phone [d] can be accomplished both with tongue tip and with tongue body. For a direct access, all phones are internally stored as a dictionary, and include a characterization of their phonation mode (soft, exploding, constant) and location (mouth, nose). That way, new phones, either of german or another language, can be easily added.

To play the animation of an selected phone, its associated speed (resp. number of frames) then gets assigned and those key-frames get adjusted. To illustrate airflow and presence/absence of voice of phones like [p] or [g], those phones are animated with the appended vocal [a]. Tongue tip and uvula obtain their own shape keys to show their rapid movements during production of phones [r] and [ʀ]. The speed of each animation is additionally controllable through an timeline interface element.

Our approach offers the possibility to exemplary output sound while playing a phone animation at normal speed and combined with the vocal [a]. In terms of co-articulation, the audio samples explain the sound initiation according to its medial (*a_'consonant'_a*) or initial (*'consonant'_a*) position in syllables. Movements of the body parts are animated synchronous to the sound effect, hence the animation curve matches the audio curve.

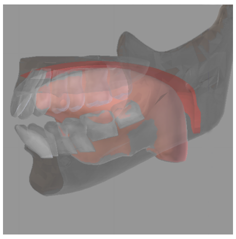


Figure 4: *Exercising the tongue's resting position is essential for correct speech and swallowing*

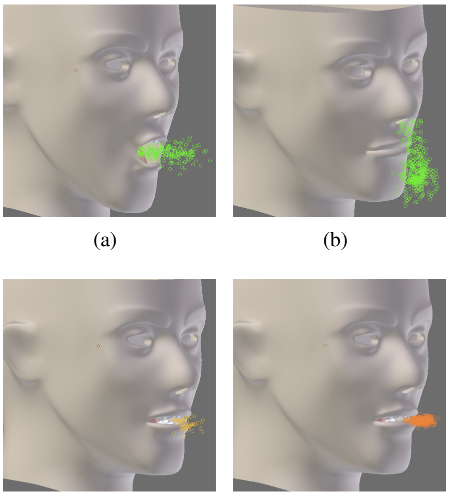
4.2. Expiration flow

Parts of the head model can be faded in and out, by activating and deactivating visibility layers. It enables the therapist to look closer at some object without being obscured by other objects. So if maxilla is shown, upper teeth can be fade out to see during articulation, if the tongue tip abuts on upper teeth-ridge or if tongue body abuts on palate.

The phonation airflow serves as supplementary visualization at normal speed for *how intensive* the phone should be produced (Figure 5). For this, we use a particle system with Newtonian physics, in which particles move between a start and a target object. A third control object defines the property of particle behavior. The particle system considers three kinds of expiration flow, differentiating the particles in their amount and lifetimes and hence influencing directly their visualization. For each expiration type, the start and end frames need to be set for those particles, the values for that are manually adjusted. For a clear separation the particles are additionally color-coded: a constant flow holds a green color, moving slowly with an increased amount and at the same time taking up a flat broad area, in the direction of the target object. The expiration type holds a few yellow particles, which have a short lifetime and therefore appear shortly. The strong, almost exploding type holds many orange colored particles, agglomerating on a small area and moving fast, but with a short lifetime.

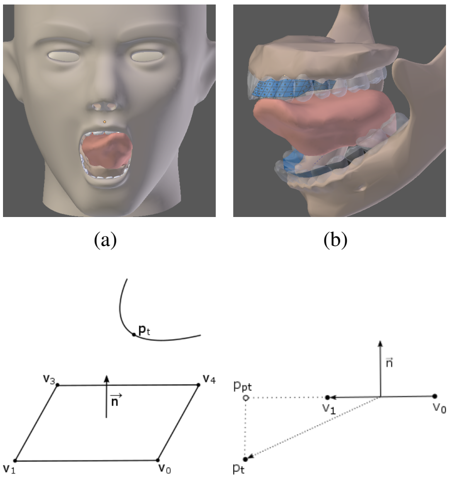
4.3. 'Free' transformation

Instead of playing predefined animations, the therapist can as well transform the tongue 'freely' to a particular target



(a) (b) (c) (d)

Figure 5: *Constant oral, constant nasal, soft and strong expiration airflow by comparison. (a) Constant for phone [o], (b) Nasal for phone [m], (c) Soft for phone [b], (d) Strong for phone [t]*



(a) (b) (c) (d)

Figure 6: *Collision detection during transformation. (a) The tongue body is moved upwards in direction of front upper teeth and the tongue tip is moved left, (b) Collision models of upper and lower teeth (blue), (c) A face of the collision model and a vertex p_t of the tongue are tested against collision, (d) p_{pt} is the projected vertex on the face's plane*

position via interface sliders. These access the corresponding shape keys of the tongue model, setting the tongue to a specific pose. The tongue root here represents the transformation's origin. One slider opens the mouth and the joining body parts models, while the tongue gets situated at the lower lingual teeth. Two other sliders let the tongue tip and the body be transformed separately upwards, downwards, left and right. The tongue body can additionally moved in or out along its x-axis, and be slimmed or expanded along local z-axis (Figure 6(a)). Those basic tongue movements can already be used to illustrate myofunctional exercises, without the in reality much more elastic tongue.

Every shifting-step of a slider can cause the tongue to hit the upper or lower front teeth. To avoid those tongue poses, we implemented a simple collision detection between the tongue tip vertices and the surfaces of the collision models. The latter consist of a simple geometry and adapt to the upper and lower front teeth (Figure 6(b)). For this purpose the the k-th polygonal face of a collision model is represented as tupel in the form of:

$$face_k = (i, \vec{n}, [v_0, v_1, v_2, v_3]),$$

where i represents the face index in the overall model mesh, $\vec{n}$ describes the face's normal and v_0 to v_3 the vertices on the face in 3D world coordinates. The vertices of the polygonal mesh of the tongue tip are as well managed in a list, so that each of those vertices p_t and each face normal $\vec{n}$ of the simplified teeth models can be tested against collision (Figure 6(c)). Firstly with the scalar product ($\circ$) between $\vec{n}$ and a tongue vertex p_t we check, if p_t lies in the upper or lower half-space of $face_k$.

If $((p_t - v_0) \circ \vec{n}) > 0$,

the tongue lies in the upper half-space of the face. Otherwise, a case distinction is necessary as a collision is not inevitably existent: p_t can lie on the face's plane, in the lower half-space of the face, where a collision had already occurred, or still on the face itself. For that, p_t is projected onto the plane of the face element with

$$p_{pt} = (sc \cdot (v_1 - v_0)) + v_0,$$

where p_{pt} is the projected tongue vertex and sc the scalar from

$$sc = (v_1 - v_0) \circ (p_t - v_0).$$

Figure 6(d) illustrates the projection. It is obvious that no collision could have been occurred, if $sc < 0$. Otherwise, the tongue vertex still does not lie neccessarily on the face, if

$|p_{pt} - v_0| \leq |v_1 - v_0|$, so a second projection including the second face vertex v_3 is made with $|p_{pt} - v_0| \leq |v_3 - v_0|$.

If a collision could be detected, the interface sliders are set back to their original values.

4.4. Deglutition

The human deglutition can also be simulated via interface control elements. The models of the participating body parts

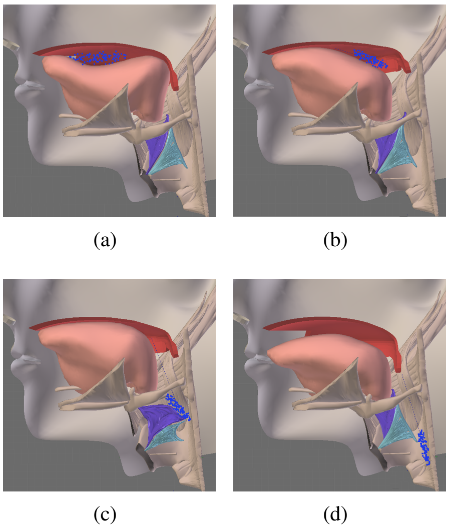


Figure 7: *The deglutition movement in four phases. The bolus is blue colored, epiglottis is purple colored. (a) Bolus in tongue bowl, (b) Soft palate is raised, (c) Swallowing reflex, (d) Bolus transportation*

obtain each five shape keys, representing the five stages of swallowing process. Figure 7 summarizes four of five stages. The chewed mass of food (bolus) is depicted as object with its own particle system, its position and transformations are predefined for animation. The movements of the other models are controlled via shape keys at defined frames. A coloration helps to distinguish involved body parts.

Pushing an equivalent button offers the possibility to speak a word into a connected microphone. If it is recognized, its letters are stored in a selective list in order to preview their animations individually or consecutively. The corresponding phone of a letter comprises 15 frames, according to that, the animation end depends on the word's length.

5. Evaluation

We carried out an expert interview with one speech therapist to evaluate our Tongue S(t)imulator approach. It followed that the presentation of the articulation process with aid of animations is not only a dynamic method, but also a more comprehensible visualization in comparison to simple two-dimensional black and white drawings. The fade out of three-dimensional models offers an alternative to show therapy exercises on one's own articulators, which impossibly provides a close view on movement nuances.

From a logopedic point of view, the visualized expiration flows seem to be a good option to demonstrate different articulation intensities in their fineness. Not only can phones be distinguished between their lips position ([b] and [p]), but also nasal phones can be described in such a way. In contrast, the parameters to 'freely' move the tongue are not beneficial in myofunctional therapy; organs behind tongue do not matter, if the tongue is only moved towards the mouth's corners or circularly contacting the lips.

The functionality to input words via microphone was viewed critically: On one hand it should be clarified, if our speech recognition is able to differentiate between potentially false and correctly spoken words. If not, the application may recognize vaguely spoken words or accept words with swapped phones. Also, correct spoken words may not be recognized at all. On the other hand, this function supports a therapy with morphemes, gearing to quotidian exercises. The visualization then would be a profitable side effect. Referring to this, the timeline for controlling the movements temporally (e.g. reducing speed) seems to be reasonable.

The use of our application in therapy depends on the severity of the patient's impairment. The client does not only need to be able to attend the therapy session, but also to necessitate the visual perception. This is not the case when treating bedridden patients with neurological disruptions. In treatment of pathological deglutition, the application assists exercices strengthening the tongue: taking the tongue pose for phone [k], or practicing the pose of tongue tip for phone [t]. The simulation of deglutition process brings out wave-like movements of tongue anatomically correct, as the tip touches at first hard palate and the tongue body then presses against soft palate, which is not able to demonstrate on one's own mouth.

6. Conclusion

The therapist confirmed that our Tongue S(t)imlator could be supportive as a part of the therapy lesson and supplementary to the treatment. The animation of phones and the possibility to fade out model objects for an unrestricted sight on oral cavity as advantage over a demonstration on their own articulation organs or commonly used materials. Our approach uses anatomical precise models with accurate lip poses (Figure 8). It was confirmed, that the lips were modeled in such a way that they are able to differentiate visually between bilabial phones like [p] and [b]; especially the demonstration of the phone [s] needed distinguishable, rounded lips. Kröger et al. present a distinctive model of oral cavity, however just sagittal view on it. The approaches of Massaro and Engwall et al. use a three-dimensional model, but do not image all articulation organs thoroughly, some models are visualized anatomically approximated.

Restrictions were made in co-articulation. Normally, when speaking a word, a characteristic of a phone in the middle of that word, is already performed in the beginning. While pronouncing the German word 'Glück' for example, the lips are already rounded at initial letter, to be able to form

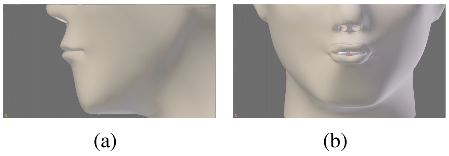


Figure 8: *Accurately modeled lips provide an differentiation between (a) phone [p] and (b) [ʃ]*

the letter 'ü' later on. If our speech recognition recognized a spoken word and the user decides to play all phones of that word consecutively, each phone is pronounced individually. Yet we did not investigate, if other interpolation methods are able to approximate co-articulation, this is part of speech synthesis. For *Baldi*, Massaro uses a weighting function for a speech unit to determine the influence of an organ's pose on the pose of a neighboring unit.

Our approach offers a visualization of expiration flows in form of a particle system during generating a sound. This demonstration received a positive feedback from speech therapists, although it has not been a primary demand. Adjusting the speed of an animation and the timeline control element have been demands for our final application and indeed accepted as advantageous ancillary functions. According to that, these functions seem to be requisites for our application to control animations directly.

Our sound output functionality likewise received a positive feedback in the expert interview. It addresses the patient's auditive perception when initiating a phone within a syllable. However, the use of such a function needs to be reconsidered, when the therapist auditions the phone by himself. Then, only the simultaneous three-dimensional visualization is able to support a syllable training.

Doubts concerning sound input via microphone seem to be justified, as the function depends on the quality of speech recognition. A more refined software would be able to analyze spectral frequency ranges of spoken phones, or to integrate particular vocabularies to compare with, but probably would not be able to eliminate interferences or spoken mistakes. Kröger et al. describe a method where the spectrum of speech signal is compared to that of the corresponding correct articulation. Engwall et al. use statistical techniques. Nevertheless, it was no essential requirement in our approach. It was rather our purpose to select and play phones of a recognized word and to give the patient a feedback about his spoken language and a monitoring of his progress.

References

[BBPP07] BUCHAILLARD S., BRIX M., PERRIER P., PAYAN Y.: Simulations of the consequences of tongue surgery on tongue mobility: Implications for speech production in post-surgery conditions. *The international journal of medical robotics and computer assisted surgery (MRCAS) 3*, 3 (2007), 252–261. 3

[Bel13] BELTKIEWICZ D.: A new horizon in logopaedics: Speech therapeutic story – innovative use of a story in the therapy of children speech impediments. In *Procedia - Social and Behavioral Sciences* (2013), vol. 109, pp. 149–216. 2

[BSM14] BARTOLOME G., SCHRÖTER-MORASCH H. (Eds.): *Schluckstörungen. Diagnostik und Rehabilitation*, 5 ed. Elsevier, 2014. 2

[EBOK06] ENGWALL O., BÄLTER O., ÖSTER A.-M., KJELLSTRÖM H.: Designing the user interface of the computerbased speech training system artur based on early user tests. *Journal of Behaviour & Information Technology 25*, 4 (2006), 353–365. 3

[Har14] HARAKÉ L.: *Entwicklung einer interaktiven 3D-Visualisierung der oro-pharyngo-laryngealen Region für die Sprechtherapie*. Master thesis, Universität Koblenz-Landau, 2014. 4

[INS12] ILIE M. D., NEGRESCU C., STANOMIR D.: An efficient parametric model for real-time 3d tongue skeletal animation. *9th International Conference on Communications (COMM)* (2012), 129–132. 3

[KBHM10] KRÖGER B. J., BIRKHOLZ P., HOFFMANN R., MENG H.: Audiovisual tools for phonetic and articulatory visualization in computer-aided pronunciation training. In *Proceedings of the Second International Conference on Development of Multimodal Interfaces: Active Listening and Synchrony* (Berlin, Heidelberg, 2010), COST'09, Springer, pp. 337–345. 2

[KP01] KING S. A., PARENT R. E.: A parametric tongue model for animated speech. *The Journal of Visualization and Computer Animation1 12*, 3 (2001), 107–115. 3

[KP06] KOCSOR A., PACZOLAY D.: Speech technologies in a computer-aided speech therapy system. In *Proceedings of the 10th International Conference on Computers Helping People with Special Needs* (Berlin, Heidelberg, 2006), ICCHP'06, Springer, pp. 615–622. 2

[Mas06] MASSARO D. W.: The psychology and technology of talking heads: Applications in language learning. In *Proceedings of International Workshop on Natural, Intelligent and Effective Interaction in Multimodal Dialogue System* (2006), pp. 183–214. 3

[PPB*11] PERRIER P., PAYAN Y., BUCHAILLARD S., NAZARI M. A., CHABANAS M.: Biomechanical models to study speech. *Faits de Langues 37* (2011), 155–171. 3

[RGC98] RODRIGUES M., GILLIES D., CHARTERS P.: Modelling and simulation of the tongue during laryngoscopy. *Computer Networks and ISDN Systems 30*, 20-21 (1998), 2037–2045. 3

[Sur01] SUROWANIEC J.: Informacja internetowa w edukacji logopedycznej (internet information in logopaedic education). In *Literatura i sztuka a wychowanie (Literature and Art vs Education)* (2001), Kida J., (Ed.), Wyższej Szkoły Pedagogicznej, pp. 209–216. 2

[WT06] WILHELMS-TRICARICO R.: From muscle models to tongue models (and back), November 2006. Presentation at Haskins Laboratories. 3

[WTP97] WILHELMS-TRICARICO R., PERKELL J.: Biomechanical and physiologically based speech modeling. In *Progress in Speech Synthesis*, van Santen J., Olive J., Sproat R., Hirschberg J., (Eds.). Springer New York, 1997, pp. 221–234. 3