Prospects of EPG and OPG sensor fusion in pursuit of a 3D real-time representation of the oral cavity

Simon Preuß, Christiane Neuschaefer-Rube, Peter Birkholz

Clinic for Phoniatrics, Pedaudiology, and Communication Disorders
University Hospital Aachen and RWTH Aachen University
Pauwelsstraße 30, 52074 Aachen, Germany
sipreuss@ukaachen.de

Abstract: After giving a brief review of the underlying technologies electro-palatography and optical palatography, this paper presents the prototype of a device that combines electrical and optical palatographic measurements to reconstruct the tongue surface during articulation in real-time. The prototype consists of a mouth piece (pseudo-palate) and a control board, and improves on a previous design by integrating the entire measurement system into a single unit, increasing spatial resolution by a greater number of sensors, and reducing the number of wires necessary to contact the pseudo-palate for more comfortable use. Furthermore, a concept to approximate and visualize the tongue surface from the sensor readings using cubic spline interpolation is described. Future challenges in the next development stages are highlighted. The outlined system could be applied in, e.g., experimental phonetics, speech therapy, silent speech interfaces, and pronunciation training.

1 Introduction

A major challenge in speech therapy and pronunciation training in general (e.g., in foreign language acquisition) is inferring the actual position of a subject's tongue from the perceived audio signal. As Atal et. al. [1] have shown, there is no bijective mapping of articulatory positions to acoustic signals, i.e., the articulatory-to-acoustic transform is not invertible. But improvement of the subject's articulation still hinges on the instructor's ability to correctly deduce the position and shape of the tongue in faulty speech, as well as comprehensibly communicating the target position. To aid an instructor in this critical task, biofeedback techniques can help to eliminate false assumptions or misunderstandings and thus speed up and generally improve therapy or training, as shown, for example, in [12]. The most straight-forward approach in this context is a real-time visualization of the oral cavities of both the subject and the instructor.
Several techniques such as magnetic resonance imaging (MRI), sonography, electromagnetic articulography (EMA), cineradiography, and X-ray have been employed to visualize articulation [6]. But as each procedure has its very own major drawbacks (e.g., cost, radiation, necessary specialized environment etc.), so far none has emerged as both accurate and practical in day-to-day applications. However, among the most promising methods are two, which are closely related and ideally complement each other: electro-palatography (EPG) and optical palatography (OPG).

1.1 EPG

EPG is a well established technique to visualize the time-varying palato-lingual contact pattern that sees widespread use in phonetics and speech therapy across the world [10]. To record an

EPG, the subject wears a plastic plate that is tailored to fit its hard palate. On this artificial palate, a number of electrodes are mounted to detect lingual contact by registering a small AC voltage which is applied to the subject. Several systems were developed at the University of Reading, UK (*Reading* system), the University of Alabama, USA (*Kay Palatometer*), and the University of Tokyo, Japan (*Rion* system). More recently, the *SmartPalate* system was developed based on the *Kay Palatometer* and commercially distributed by *Complete Speech*. These systems all follow the same basic principle described above but differ in several aspects that are subsumed in Table 1.

Name	No. of palate electrodes	Sample rate [Hz]	Palate cost [€]	Status
Kay Palatometer	96	100	ca. 300	discontinued
Reading	62	100	ca. 220	available
Rion	63	N/A	N/A	discontinued
SmartPalate	122	100	ca. 230	available

Table 1 - Properties of several EPG systems (see [14], pricing according to manufacturer inquiry)

In addition to these characteristics, the artificial palates were designed with different layouts. The *Kay* palate, patented in 1978 [14], covered the teeth of the maxilla as well as the hard palate. It had two pairs of additional contacts, one on each buccal surface of the teeth, that made permanent contact with the cheeks for grounding purposes. Every single one of the 100 contacts was soldered to fine enameled copper wires, which exited the mouth behind the rear molars, were sealed in flexible tubing and soldered to two fifty pin connector plugs. This rather complex and expensive manufacturing process contributed significantly to the cease of production of the *Kay Palatometer* in 1998 [14].

The *Reading* system's pseudo-palate [11] is affixed by Adams clasps around the teeth. The 62 contacts are soldered to individual copper wires that exit the mouth around the back of the rear molars, as with the *Kay* palate, are sealed in flexible tubing and soldered to a double-sided edge connector card. This palate design is still in use in combination with the *WinEPG* measurement hardware by *Articulate Instruments* [14].

In 1977, *Rion Co Ltd*, Japan, filed a flexible circuit based palate design for patent [14]. Based on this original design, *Rion* developed the SP-02 flexible artificial palate, that could be fitted directly without the need for a plaster mold. Due to failure to comply with materials-related safety regulations however, it was discontinued.

The currently available *SmartPalate* system by *Complete Speech* (formerly *Logometrix*) also uses a flexible circuit board that is glued to a plastic mouthpiece. The three-lobed circuit board itself was filed for patent in 2001 [14] and includes one pair of lip contacts to register labial opening. Due to its shape, it can be easily trimmed to fit smaller palates at the expense of the number of available contact electrodes by cutting away posterior and lateral area [14].

1.2 OPG

In contrast to the EPG, which registers palato-lingual contact, Glossometry or, more commonly, optical palatography (OPG) uses optical distance sensors mounted on an artificial palate to determine the distance between the hard palate and the tongue at multiple sampling points. The basic design was first introduced in 1978 by Chuang and Wang [5], refined by Fletcher et al. in the 1980s [7, 8] and further expanded upon in 2011 and 2012 by Birkholz et al. [3, 4]. Chuang and Wang and Fletcher et al. used four paired infrared light emitting diodes (LEDs) and phototransistor transducers mounted along the midsagittal line of an artificial pseudo-palate. Birkholz et al. added a fifth pair, mounted facially on the maxillary incisors, that was directed towards the lips and used to measure labial aperture.

Even though OPG was able to accurately track the midsagittal tongue position, it has not been widely adopted due to a number of limitations. Two major drawbacks of the earlier systems, namely the large sensor size, that made the pseudo-palates cumbersome to speak with, and the time and material cost to manufacture them, have since been amended by [4].

1.3 Synergetic potential of EPG and OPG

While EPG is only capable of registering palato-lingual contact, but across the entire hard palate, OPG can measure the distance between palate and tongue, but typically only along the midsagittal line. EPG is thus suited to study sounds with ample palato-lingual contact, i.e., the lingual obstruents /t, d, k, ɡ, s, z, ʃ, ʒ, tʃ, dʒ , tr, dr/, close vowels and diphthongs such as /iː, ɪ, eɪ, ɛ/, the nasals /n, ŋ/, the palatal approximant /j/ and the lateral approximant /l/ [10]. With OPG, the articulation of diphthongs and the vowels /aː, æ, eː, ɛ, iː, ɪ, oː, ɔ, uː, ʊ/ was successfully studied [8]. If both techniques were employed in conjunction, it would enable a wide range of sounds to be analyzed. Additionally, the fused information might also permit further discrimination of sounds that could not be unambiguously identified by either method on its own.

In a first survey of this concept, Birkholz et al. [2, 3] designed a prototype system combining a *Reading* EPG palate with OPG sensors according to [4]. Building upon this design, we present in this paper recent advances of this concept with an increased number of optical and contact sensors, and highlight further development steps en route to a fully integrated measurement system.

2 The Opto-EPG prototype

The proposed Opto-EPG measurement system comprises a concept study of a pseudo-palate, a control board, and a simple measurement software to display electrical and optical sensor readings simultaneously.

2.1 The pseudo-palate

The concept study of the new Opto-EPG pseudo-palate, which abandons the *Reading* design used in [2, 3], is realized on a rigid FR4 printed circuit board (PCB). The final device will use a flexible PCB that is affixed on a plastic plate individually molded to fit a subject's hard palate. The top layer, facing the tongue in the final system, contains 128 copper contact electrodes (Ø1 mm) and six paired VSMY2850 infrared SMD LED and TEMT7100 SMD phototransistors (see Figure 1a). On the bottom layer, four ADG731 32:1 SMD analog multiplexers are mounted, whose input pins are connected to the contact electrodes (see Figure 1b). The multiplexers' supply voltage, their three common control and four individual output pins, the 12 tracks needed to control the optical sensors (six LED control signals, six phototransistor signals), and common ground for all components are wired to a pin header connector.

The multiplexers are in a 48-lead TQFP package and have a specified height of no more than 1.2 mm, which contributes to the overall thickness of the pseudo-palate. However, the number of tracks necessary to address all 128 contact sensors decreases significantly from 129 (128 output signals and common ground) to 10 (supply voltage, ground, four control signals and four output signals). As all wires must exit through the mouth opening, keeping their number low is imperative to retain a natural feel while speaking and outweighs the possible slight discomfort

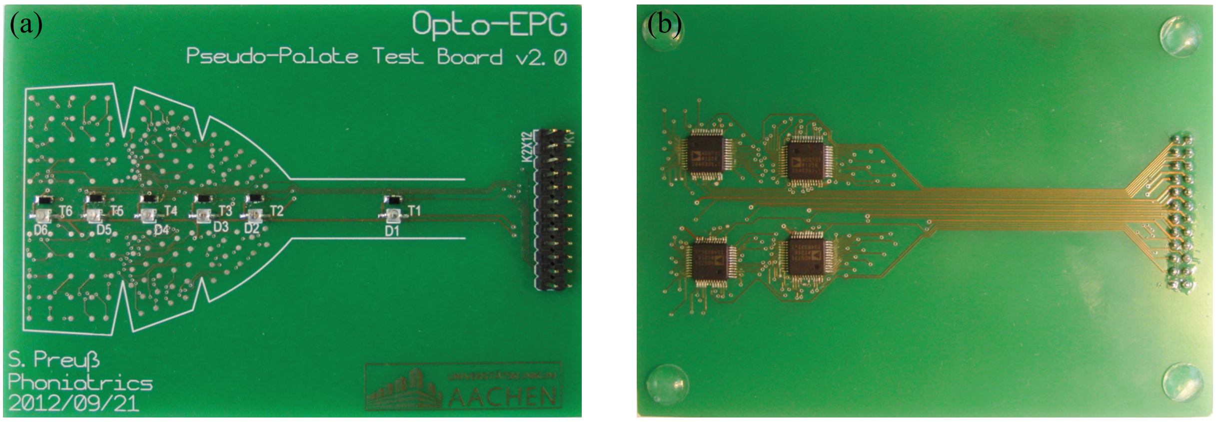

Figure 1 - FR4 PCB of the Opto-EPG pseudo-palate concept study containing 128 electrical and six optical sensors and four multiplexer units. (a) Sensor layer, directed towards the oral cavity in the final system. (b) Bottom layer, wedged between oral cavity and artificial palate plate in the final system

caused by the rather small additional thickness of the system.

The contact electrodes have been symmetrically arranged with respect to the midsagittal line and are more densely spaced in the alveolar and post-alveolar part. This results in a higher spatial resolution there (2 mm), than in the palatal and velar region (3 mm). The *Kay* and *Reading* palates also provided variable spacing to improve identification of gaps in alveolar tongue contact (required for production of /s/ and /z/), but with a generally lower resolution of less than 3 mm in the alveolar region and greater than 3 mm in the palatal and velar region [14]. All EPG signal tracks, the supply voltage and all multiplexer control tracks are 0.15 mm wide, while the ground track is 0.4 mm wide. The LEDs D1-D6 are placed at a distance of 3.2 mm (center to center) to the phototransistors T1-T6 along the midsagittal line, with 10 mm spacing between units 2-6 and 26.2 mm between unit 2 and unit 1, the labial aperture sensor. Even though all signal tracks should be kept as narrow as possible in the confined space, the LED control tracks are 0.4 mm wide to avoid heating due to the comparatively high currents here (ca. 166 mA).

The white outline in Figure 1a is drawn according to statistical palate dimensions data also used in [2]. Notches are included to allow the circuit to follow the arched palate contour once it is realized on a flexible PCB.

In summary, the advances in comparison to [2, 3] are a greater number of contact sensors (128 vs. 62) and optical sensor units (6 vs. 5) to improve spatial resolution, and the introduction of analog multiplexers to reduce the number of wires that exit the mouth thus improving the device's wearing comfort and lowering speech obstruction.

2.2 The control board

All signals from the pseudo-palate are connected to a second board by a ribbon cable. The multiplexer control signals are directly wired to an ATmega16 microcontroller unit (MCU), while their analog output signals are first passed through a simple comparator based 1-bit A/D converter before connecting to the MCU. Six precision current sources, controlled by the MCU, drive the LEDs, and the incoming phototransistor signals are passed through a detector circuit as in [2]. The control board furthermore includes a 9 V block battery, the necessary voltage regulation to generate the 5 V supply voltage for all active components, a connector to program the MCU and a serial RS232 interface for communication with a PC. This design significantly improves on [2, 3] by incorporating OPG and EPG in a single compact measurement unit.

2.3 The measurement software

After initializing its serial peripheral interface (SPI), RS232 and A/D converter functions, the MCU enters a loop in which it continually acquires the sensor data and transmits it to a connected PC via RS232 (see Figure 2a) with a frame rate of 100 Hz. A simple graphical user interface (GUI) receives, displays and logs the data to an ASCII-File (see Figure 2b) on the PC. The visualization is still done by treating the EPG and OPG parts of the system separately. The EPG display does not represent the actual 3D position of the contact sensors, a fact that will be addressed in Sec. 3.1. There is also no function implemented yet, which maps the raw A/D converted values of the OPG to an absolute distance.

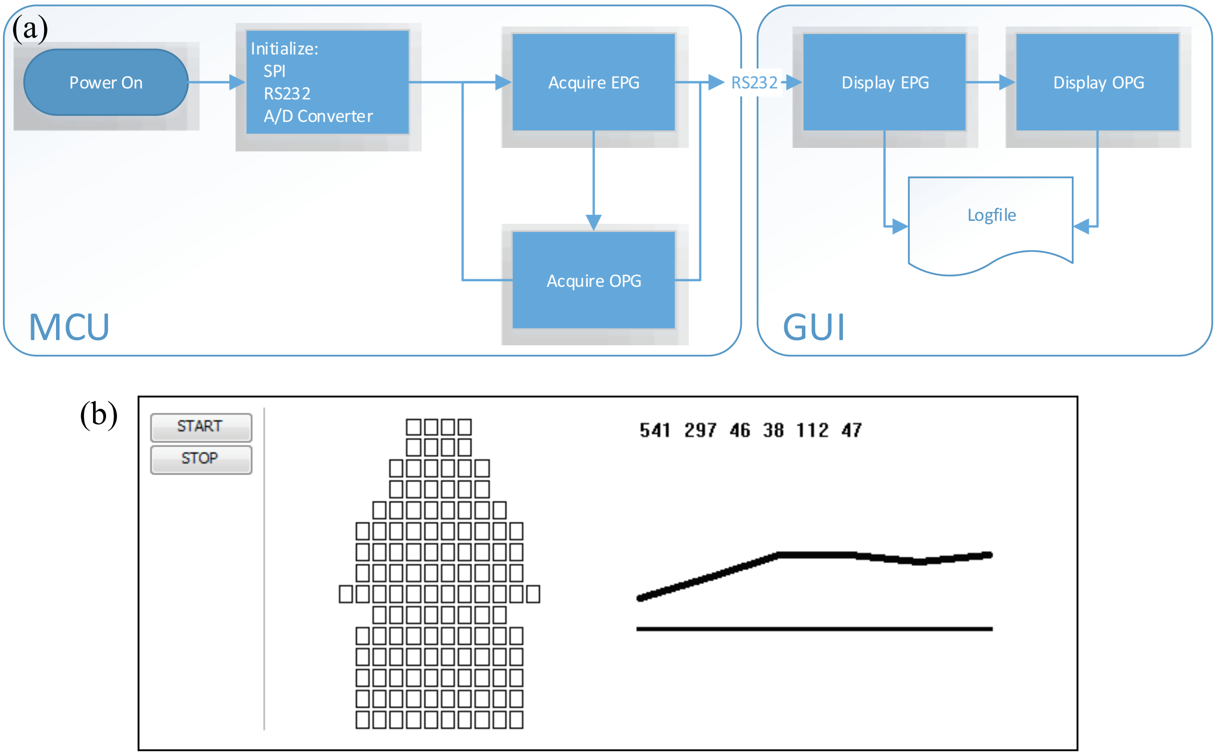

Figure 2 - Block diagram and GUI screenshot of the measurement software. (a) Overview of the data flow and functional blocks in the MCU software and the GUI. (b) Simple graphical user interface showing the EPG pattern on the left and the OPG measurement with a thin baseline on the right

2.4 Proof of concept

The described prototype is capable to read the contact sensor voltages and the phototransistor currents quasi-simultaneously (i.e., in sufficiently quick succession), with a frame rate of 100 Hz, where a single data frame is a complete reading of all 128 contact sensors and 6 optical units. It is therefore suited to perform EPG and OPG with a sufficiently high temporal and spatial resolution to analyze speech movements. The next iteration of this design will be realized on a flexible circuit board. This board will then be affixed on an artificial palate, that is individually molded to an impression of a subject’s maxilla.

3 3D reconstruction of the oral cavity

The aim of the presented work is to reconstruct the oral cavity in 3D from the information about the tongue shape sampled at discrete points by the sensors. But while the visualization in the simple exploratory GUI shows an equidistant grid of contact points, this is indeed not the case on the actual pseudo-palate (see Section 2.1 and Figure 1a). Once the sensors are mounted on a flexible PCB and affixed on an arched plastic palate mold, the contact sensors will also not always be in exactly the same place between subjects. To map the OPG readings to a defined point along the midsagittal plane, the optical axes of the optical sensor units need to be determined as well. It is therefore necessary to accurately localize every sensor in a 3D coordinate system.

3.1 3D sensor localization

In order to localize the sensors, the compound system needs to be visually captured by a 3D scanner after affixing the standardized flexible circuit on the individual artificial palate of a subject. The most important criterion this scanner needs to satisfy is a high spatial resolution.
Though currently no standard test protocols exist to evaluate the performance of 3D imaging systems, Guidi et al. [9] have published an extensive analysis of seven commercial scanners. Their study shows, that the relatively low-cost scanner *NextEngine* by *NextEngine Inc.* adequately meets our requirements. The overall production cost of the Opto-EPG system can thus still be kept low without compromising its spatial accuracy.

3.2 Reconstruction of the tongue shape

To show how we intend to exploit the complementary properties of EPG and OPG summarized in Sec. 1.3, we will use the example of the vowel /eː/. By extending the results in [2] to the greater number of sensors, we expect the EPG pattern of this sound to be similar to Figure 3a, and the OPG measurement to Figure 3b. The optical sensor unit 1 is used to measure labial aperture and will for now be neglected.
The points of palatal contact on the tongue surface and the tongue position in the midsagittal plane are thus known across the entire palate. The tongue shape is now reconstructed by interpolating between these known points using cubic spline interpolation in both coronal and sagittal direction. This interpolation technique ensures that the resulting shape is sufficiently smooth, a constraint imposed by the real-world tongue. At the points of palato-lingual contact, however, the tongue tissue is no longer smooth but slightly warped by the applied pressure and follows the palate contour. To represent this behavior a boundary condition is enforced after interpolation, that no point on the tongue surface can be higher than the palate height at that given position. The reconstruction scheme is illustrated for a single coronal section in Figure 3c. After 2D interpolation, the resulting shape is filtered with a 2D low-pass filter for further smoothing. Figure 3d shows a MATLAB plot of the final reconstruction.
Spline interpolation is a well established technique and, on a sufficiently powerful computer, the calculations can be done in real-time. To implement the described reconstruction scheme, we intend to use non-uniform rational B-splines (NURBSs) as they are in wide-spread use in computer graphics and have proven efficient for modeling 3D shapes by interpolating between a number of control points [13]. The evaluation of this concept of reconstruction will be possible once the next three important steps have been completed: mounting the sensors on a flexible circuit board, affixing it on an artificial palate, and determining the 3D positions of the sensors.

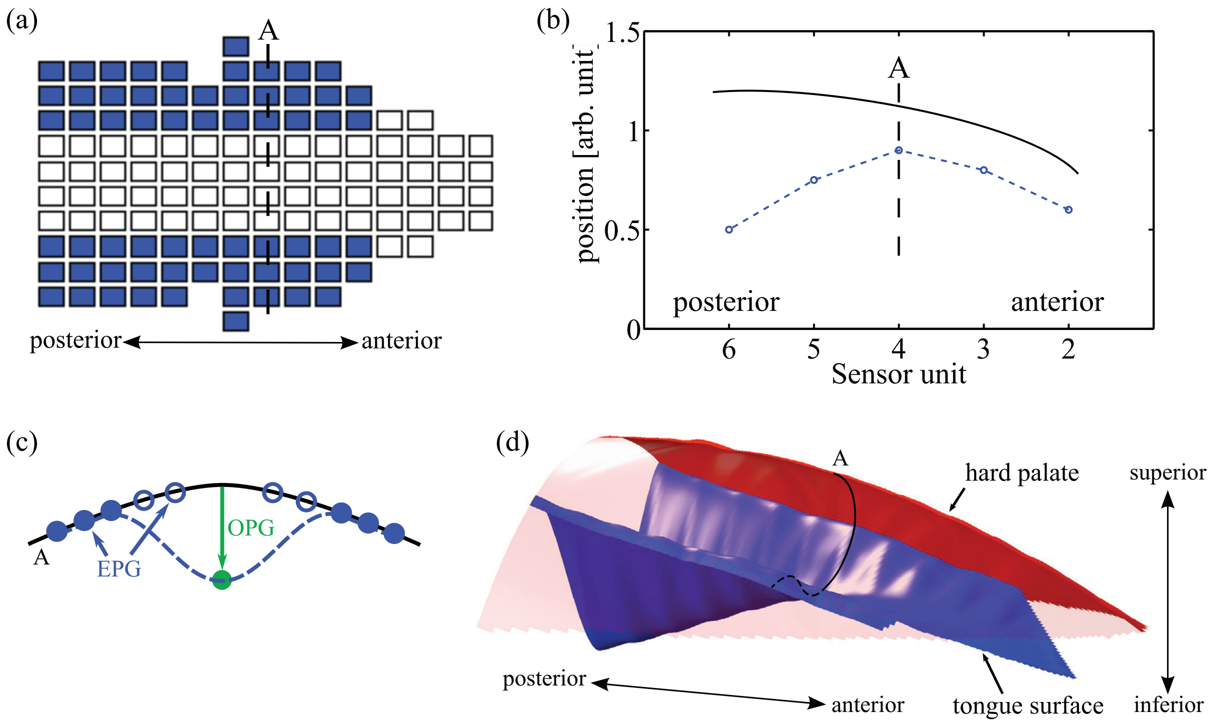

Figure 3 - Concept of reconstruction using the example of the sound /eː/ by extending the data in [2] to the greater number of sensors and cubic spline interpolation between sampled points. (a) EPG pattern: points of palato-lingual contact shown in blue. (b) OPG measurement: exemplary palate contour in black, tongue position in blue. (c) Exemplary reconstructed tongue contour in a single coronal section A: palate roof is black, EPG measurements are blue with a filled circle representing registered contact, OPG measured distance is green, resulting interpolated curve is dashed. (d) 3D reconstruction of the tongue surface (blue) with exemplary palate (red).

4 Summary and conclusion

In this paper we proposed a concept for a real-time system that measures tongue-palate contact as well as the midsagittal contour of the tongue. Compared to [2, 3], the spatial resolution was improved by increasing the number of contact sensors from 62 to 128 and the number of optical sensor units from 4 to 5. By introducing analog multiplexers, the number of wires leading out of the mouth was greatly reduced for more wearing comfort. Furthermore, the measurement systems for the contact sensors and the distance sensors were integrated into a single unit. It was also shown, how in theory the 3D tongue surface can be approximated from this information by cubic spline interpolation between measured samples, if the 3D positions of the sensors are known. The next design steps and challenges were emphasized. We are confident that the concepts described herein will allow real-world measurement of articulated sounds in the next design stage and a thorough evaluation of the reconstruction algorithm.

5 Acknowledgments

This work was funded by the German Research Foundation (DFG), grant BI 1639/1-1.

References

[1] ATAL, B. S., J. J. CHANG, M. V. MATHEWS and J. W. TUKEY: *Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer-sorting technique*. The Journal of the Acoustical Society of America, 63(5):1535–1555, 1978.

[2] BIRKHOLZ, P., P. DÄCHERT and C. NEUSCHAEFER-RUBE: *Advances in combined electro-optical palatography*. In *Proc. of Interspeech 2012*, Portland, Oregon, USA, 2012.

[3] BIRKHOLZ, P. and C. NEUSCHAEFER-RUBE: *Combined optical distance sensing and electropalatography to measure articulation*. In *Proc. of Interspeech 2011*, pp. 285–288, Florence, Italy, 2011.

[4] BIRKHOLZ, P. and C. NEUSCHAEFER-RUBE: *A new artificial palate design for the optical measurement of tongue and lip movements*. In WOLFF, M. (ed.): *Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung 2012*, pp. 89–95. TUD-Press, Dresden, Germany, 2012.

[5] CHUANG, C.-K. and W. S.-Y. WANG: *Use of optical distance sensing to track tongue motion*. J Speech Hear Res, 21(3):482–496, 1978.

[6] EARNEST, M. M. and L. MAX: *En route to the three-dimensional registration and analysis of speech movements: instrumental techniques for the study of articulatory kinematics*. Contemporary Issues in Communication Science and Disorders, 30:2–25, 2003.

[7] FLEGE, J. E., S. G. FLETCHER, M. J. MCCUTCHEON and S. C. SMITH: *The physiological specification of American English Vowels*. Language and Speech, 29(4):361–388, 1986.

[8] FLETCHER, S. G., M. J. MCCUTCHEON, S. C. SMITH and W. H. SMITH: *Glossometric measurement in vowel production and modification*. Clinical Linguistics & Phonetics, 3(4):359–375, 1989.

[9] GUIDI, G., M. RUSSO, G. MAGRASSI and M. BORDEGONI: *Performance Evaluation of Triangulation Based Range Sensors*. Sensors, 10(8):7192–7215, 2010.

[10] HARDCASTLE, W., W. JONES, C. KNIGHT, A. TRUDGEON and G. CALDER: *New developments in electropalatography: A state-of-the-art report*. Clinical Linguistics & Phonetics, 3(1):1–38, 1989.

[11] HARDCASTLE, W. J., F. E. GIBBON and W. JONES: *Visual display of tongue-palate contact: Electropalatography in the assessment and remediation of speech disorders*. International Journal of Language & Communication Disorders, 26(1):41–74, 1991.

[12] KATZ, W. F., S. V. BHARADWAJ and B. CARSTENS: *Electromagnetic Articulography Treatment for an Adult With Broca's Aphasia and Apraxia of Speech*. J Speech Lang Hear Res, 42(6):1355–1366, 1999.

[13] PIEGL, L.: *On NURBS: a survey*. Computer Graphics and Applications, IEEE, 11(1):55 –71, jan. 1991.

[14] WRENCH, A. A.: *Advances in EPG palate design*. International Journal of Speech-Language Pathology, 9(1):3–12, 2007.