ISCA Archive
http://www.isca-speech.org/archive

INTERSPEECH 2012
ISCA's 13th Annual Conference
Portland, OR, USA
September 9-13, 2012

Advances in combined electro-optical palatography

Peter Birkholz, Philippe Dächert, Christiane Neuschaefer-Rube

Clinic of Phoniatrics, Pedaudiology, and Communication Disorders,
University Hospital Aachen and RWTH Aachen University, Aachen, Germany
pbirkholz@ukaachen.de, philippe_daechert@yahoo.de, cneuschaefer@ukaachen.de

Abstract

This paper describes the development of a device that combines the electropalatographic measurement of tongue-palate contact with optical distance sensing to measure the mid-sagittal contour of the tongue and the position of the lips. The device consists of a thin acrylic pseudopalate that contains both contact sensors and optical reflective sensors. Application areas are, for example, experimental phonetics, speech therapy, and silent speech interfaces. With regard to the latter, the prototype of the system was applied to the recognition of vowels from the sensor signals. It was shown that a classifier using the combined input data from both the contact sensors and the optical sensors had a higher recognition rate than classifiers based on only one type of sensory input.

Index Terms: electropalatography, glossometry, silent speech interfaces

10.21437/Interspeech.2012-220

1. Introduction

Electropalatography (EPG) is a well-established and highly effective technique to measure the contact between the tongue and the palate with high temporal and spatial resolution. For this technique, the speaker wears a thin artificial palate (pseudopalate) with multiple electrodes distributed over its surface that detect contact with the tongue. Therefore, it is most valuable for articulatory feedback of phones with distinct tongue-palate contact like obstruents, laterals, and high front vowels. However, if the tongue is not touching the palate, there is no indication of its distance from the palate. Hence, the mid-sagittal shape of the tongue can usually not be reconstructed from EPG measurements. However, Chuang and Wang [1] showed that *reflective optical sensors* mounted onto a pseudopalate can be used to measure the tongue-palate distance and so to reconstruct the tongue contour in the oral cavity. This method was later advanced and modified by Fletcher *et al.* (e.g. [2]) and Wrench *et al.* [3, 4]. Because an optical sensor occupies more space on the pseudopalate than an EPG electrode, less measurement points are usually used for optical distance sensing than for contact sensing (e.g., four distance sensors along the mid-line of the palate in [1] and [2]). In this study, based on our previous experiments [5], we designed an *electro-optical* palatograph combining EPG electrodes and optical sensors in the same pseudopalate for a more detailed analysis of speech movements than with either sensor type alone. Eventually, our goal is to reconstruct the 3D shape of the oral cavity in real-time based on the combined sensor data. In this paper, we introduce the new device and analyze the benefit of combining both types of sensor data for the recognition of vowels.

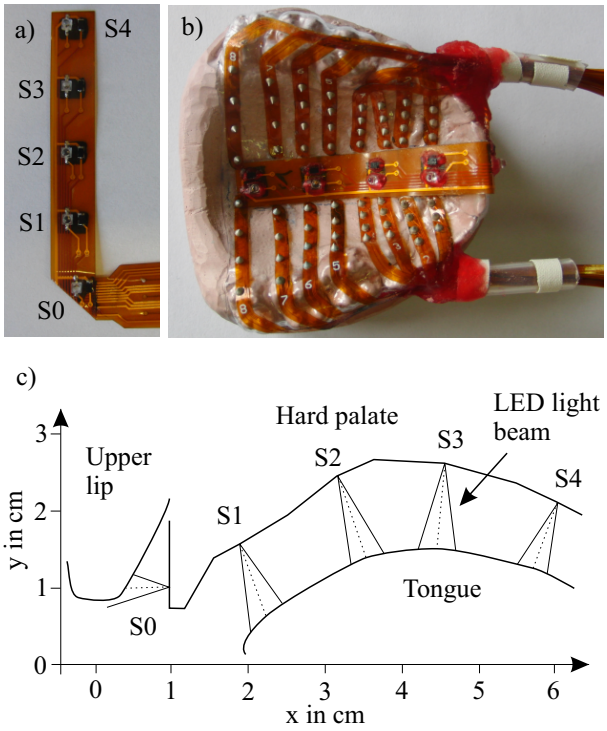

Figure 1: *a) Populated flexible circuit for the optical sensors. b) Arrangement of the flexible circuits for the contact sensors and optical sensors on the pseudopalate. c) Mid-sagittal section of the pseudopalate.*

2. The electro-optical palatograph

The pseudopalate prototype developed in this study is shown in Fig. 1b. It consists of a 0.5 mm sheet of acrylic plastic, which was thermoformed on a plaster model of the hard palate and carries the EPG electrodes and opti-

cal sensors. The EPG electrodes were integrated as for the Articulate palate [6]. In this design, the electrodes are laid out on precast flexible circuit strips. Using these strips considerably reduces the time and cost needed to manufacture a palate compared to the traditional designs. Therefore, we adopted this method for the integration of the optical sensors. We designed a flexible circuit to carry five reflective optical sensors, as shown in Fig. 1a. This circuit was placed along the midline of the palate and bent around the upper incisors. Figure 1c illustrates the resulting position and orientation of the sensors S0–S4 in the mid-sagittal plane. While S1–S4 were directed towards the tongue to measure the tongue-palate distance, S0 was arranged to measure the light reflected by the upper lip, which varies for different degrees of lip protrusion and lip aperture [5]. After fixing the flexible circuits for all sensors, a cover layer of acrylic plastic was formed over the base layer to seal the circuits. Finally, the electrodes and optical sensors were exposed. In the following we provide some more detail on the optical sensors and the measurement system.

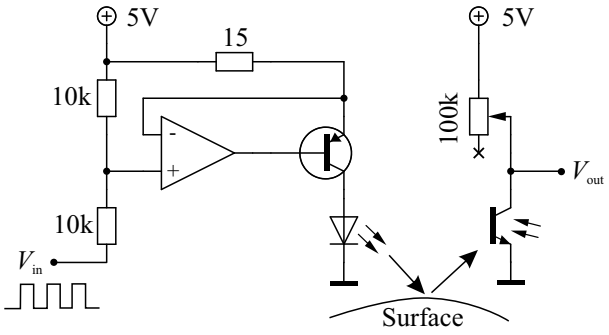

Figure 2: *Reflective light-sensing circuit.*

2.1. Reflective optical sensors

A reflective optical distance sensor consists of an LED and a phototransistor placed next to each other. The phototransistor detects the the light that is reflected when a surface in front of the sensor is illuminated by the LED. The greater the distance between the sensor and the surface, the less light is received by the phototransistor. This relation is used to infer the the distance from the output of the phototransistor. The choice of a suitable sensor type in terms of the used LED, phototransistor, and their arrangement, is a major pre-condition for effective distance sensing in the oral cavity. The sensor should be as small and flat as possible, have a measuring range of at least 25 mm, and be as insensitive as possible to coating with saliva. Because commercially available integrated distance sensors were usually not designed for these requirements, we arranged and tested different combinations of *discrete* LEDs and phototransistors. We tested the following three combinations: (A) The LED VSMY2850 and the phototransistor VEMT2020; (B) The LED VSMY2850 and the phototransistor TEMT7100; (C) The LED VSMY1850 and the phototransistor TEMT7100 (all components by Vishay Semiconductors). While the LED VSMY1850 and the phototransistor TEMT7100 have a flat and tiny 0805 package with a height of only 0.85 mm, the LED VSMY2850 and phototransistor VEMT2020 include a lens and have a footprint of 2.3 x 2.3 mm^2 and a height of 2.8 mm. The distance between the optical centers of the LED and the phototransistor were between 3.0 and 3.5 mm for all three tested sensors. Figure 2 shows the circuit for driving the LEDs with a current of 165 mA and measuring the phototransistor output values. The voltage V_{out} was digitized using a 10-bit ADC with a reference voltage of 5 V.

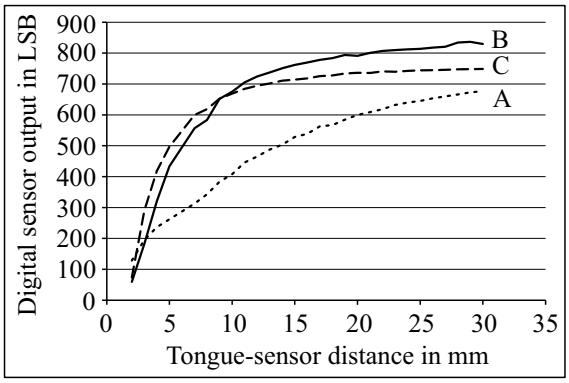

Figure 3: *Distance-sensing functions for the three tested optical sensors.*

For each of the three sensors, the digital sensor output was measured in-vitro as a function of the distance to the tongue. To keep the tongue surface in well-defined distances to the sensors, we designed special spacers made from plexiglas tubes (26 mm inner diameter; lengths between 2 mm and 30 mm in steps of 1 mm) with a wide-meshed strong net spanned over the opening for the tongue. Figure 3 shows the measured distance-sensing functions. With respect to the insensitivity against noise and other perturbations, a sensor is best suited when the slope of the curve is high. While all three curves have a high slope at small distances, it substantially differs for greater distances. The average slopes between 20 and 25 mm are 9.6, 4.3, and 1.6 LSB/mm for the sensors A, B, and C, respectively. Hence, the slope is highest, when the LED and phototransistor both have a lens, smallest, when both have a flat surface, and in between for the combination of the LED with the lens and the flat phototransistor. For our prototype palate we opted for sensor B, which is a compromise with respect to both the slope and the physical size. After populating the flexible circuit with five sensors of type B as shown in Fig. 1a and sealing the margins and electrical contacts of the components with

modeling resin, we measured the distance-sensing function for each of them (Fig. 4) using the spacers described above in steps of 5 mm. These functions were used to obtain the distances from the measured sensor output values in the experiments.

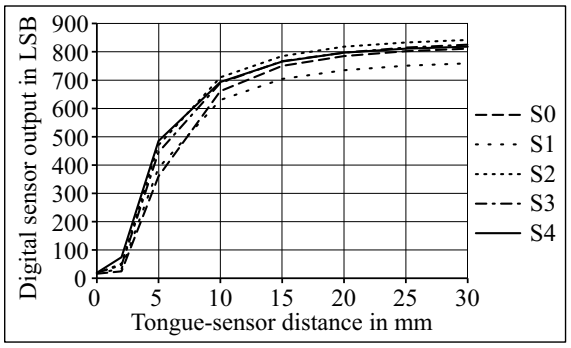

Figure 4: *Distance-sensing function of the five optical sensors after they were soldered on the flexible circuit board and sealed with modeling resin at the margins.*

2.2. Measurement system

The measurement system was identical to that in [5]. Both the EPG contact data and the output of the optical sensors were measured at a rate of 100 Hz. EPG data were measured using the WinEPG system by Articulate Instruments, and the optical sensor data were measured with a custom-made electronic unit, where each sensor was driven with the circuitry shown in Fig. 2. The optical sensors were switched in sequence to avoid optical crosstalk. The separate data streams from the WinEPG system and the optical sensor unit were synchronized and combined on a laptop computer with a custom-made software to display and analyze the data.

3. Evaluation

To analyze the performance of the new prototype, we recorded a corpus with the logatoms [bVbVbVbV] with V $\in$ {aː,eː,iː,oː,uː,ɛː,øː,yː}, [aCaCaCaCaCa] with C $\in$ {p,t,k,f,s,ʃ,ç,x,l}, and a read passage in a German book [7, p. 163] with a duration of 148 s. Figure 5 illustrates the average EPG patterns and tongue contours for selected vowels and consonants of the logatom corpus. The EPG patterns and most of the tongue contours conform with previous knowledge about articulation. However, there are two peculiarities. First, the constriction for [ʃ] in the mid-sagittal display seems too wide. We assume that the most constricted region for [ʃ] was actually about halfway between the sensors S1 and S2 and was therefore poorly captured with the current arrangement of the optical sensors. Second, the tongue contours for [aː] and [uː] seem distorted in the anterior part, probably as a result of inaccurate sensor calibration.

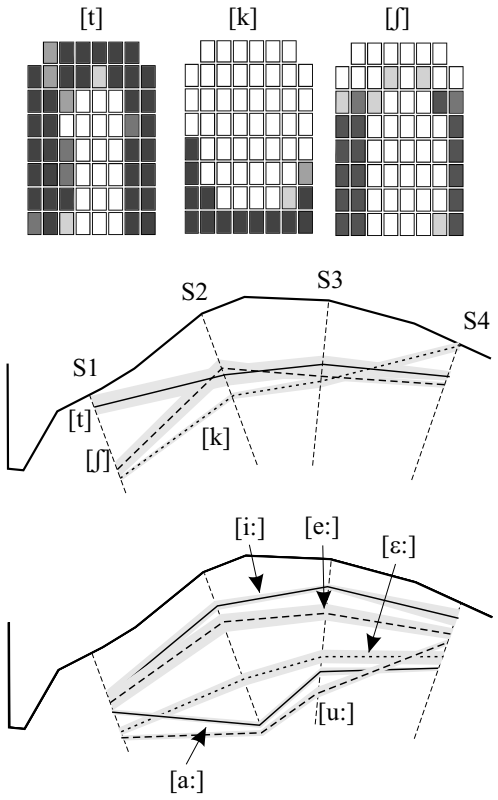

Figure 5: *Average EPG patterns and tongue contours of the consonants [t,k,ʃ] in the context of the vowel [aː], and tongue contours of the German vowels [iː, eː, ɛː, aː, uː]. The gray regions indicate the $\pm 2\sigma$ ranges of the corresponding contours.*

The read text in the corpus was used to analyze the benefit of combining input data from the EPG electrodes and the optical sensors for the classification of vowels. Therefore, we compared the performance of three classifiers: one using only the EPG data as input, one using only the optical sensor data as input, and one using both types of sensor data. As classifiers we used feed-forward neural networks with one layer of input neurons representing the sensor data and one layer of output neurons representing the vowel classes. To keep the experiment simple, we refrained from using more sophisticated classifiers and considered only the long (tense) vowels for classification at this stage. For each long vowel in the text, the sensor data frame at the acoustic midpoint of the vowel was extracted as sample for the corresponding vowel. Table 1 shows the number of samples per vowel. The vowels [ɛː] and [øː] appeared less than four times and were therefore not considered. Each of the six vowels was represented by one output neuron in the networks. The activation of an output neuron was trained to be 1 when the input data represented a sample of the corresponding vowel, and otherwise 0. To prevent overfitting of the neural networks despite the limited number

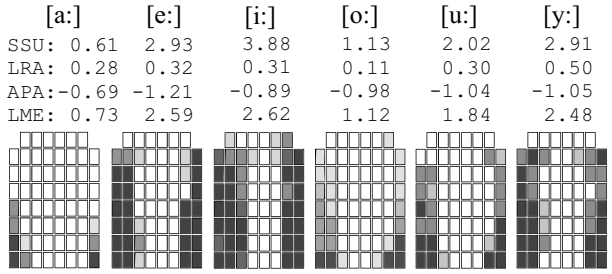

Figure 6: *Average EPG patterns of the six selected vowels and the corresponding EPG index values [8].*

of training data, the number of input neurons (and hence the number of connections/weights in the network) had to be suitably low. Hence, it was not advisable to represent each EPG contact as an individual input neuron. Instead, each EPG patterns was reduced to a vector of only four "EPG indices" based on a 2D cosine transform of the patterns according to [8]. Figure 6 shows the average EPG patterns for the vowels in the read text and the corresponding index values. Each index was represented by one input neuron. In addition, one input neuron was used to represent the distance (in cm) measured by each of the five reflective sensors. For the networks trained with only one type of sensory input, the input neurons for the other type were omitted. The networks were trained using the tool *JavaNNS* [9] with the Backpropagation learning method, a learning rate of $\eta = 0.01$, and random initial weights. The performance was assessed by four-fold cross-validation. The mean recognition rates are shown in Tab. 1.

Table 1: *Vowel recognition rates of the neural networks with input neurons for only the contact sensors, only the optical sensors, and both sensor types.*

Vowel	#Items	Recognition rate %		
		Contact sensors	Optical sensors	All sensors
[aː]	30	96.7	100.0	100.0
[eː]	11	18.2	9.1	36.4
[iː]	29	86.2	89.7	79.3
[oː]	15	46.7	40.0	53.3
[uː]	8	0.0	0.0	25.0
[yː]	4	0.0	0.0	0.0
All	97	64.9	64.9	69.1

Hence, based on only the contact sensors or only the optical sensors as input, a recognition rate of 64.9% was achieved. When the combined data were used as input, the recognition rate increased to 69.1%. While the recognition rate was generally high for the frequent vowels [aː] and [iː], it was lowest for the infrequent vowels [uː] and [yː].

4. Discussion and conclusions

The device presented in this paper was designed for the combined measurement of tongue-palate contact and distance. In this form, it provides feedback about the essential aspects of both vowel production and consonant production. The major improvements compared to our previous prototype regard the choice of the optical sensors and the manufacturing of the palate based on flexible circuits. The evaluation of the prototype suggested that the determination of the tongue contour for consonants would benefit from using *more* than four sensors along the midline of the palate, and that the method for the calibration of the optical sensors needs refinement.

With regard to the recognition of vowels from the sensor data, it was shown that the combination of both types of sensors improves the recognition rate. However, the improvement was lower than we actually expected. One reason is probably that for this speaker and the limited set of vowels, the EPG indices alone already allow a quite good discrimination of the vowels. Furthermore, more sophisticated and tuned classifiers for real-world speech recognition would probably increase the overall recognition rate considerably.

5. Acknowledgements

We would like to thank Martin Humperdinck, Alexander Füglein, Ulrike Fritz, André Barloi, Udo Höhn, Sebastian Scharrmann, René Bohne, Alex Röll, Doris Mücke, Martine Grice, Phil Hoole and Alan Wrench for their contributions to this research.

6. References

[1] C.-K. Chuang and W. S. Wang, "Use of optical distance sensing to track tongue motion," *Journal of Speech and Hearing Research*, vol. 21, pp. 482–496, 1978.

[2] S. G. Fletcher, M. J. McCutcheon, S. C. Smith, and W. H. Smith, "Glossometric measurements in vowel production and modification," *Clinical Linguistics and Phonetics*, vol. 3, no. 4, pp. 359–375, 1989.

[3] A. A. Wrench, A. D. McIntosh, and W. J. Hardcastle, "Optopalatograph: Development of a device for measuring tongue movement in 3d," in *EUROSPEECH '97*, Rhodes, Greece, 1997, pp. 1055–1058.

[4] A. A. Wrench, A. D. McIntosh, C. Watson, and W. J. Hardcastle, "Optopalatograph: Real-time feedback of tongue movement in 3d," in *5th International Conference on Spoken Language Processing (ICSLP 1998)*, Sydney, Australia, 1998.

[5] P. Birkholz and C. Neuschaefer-Rube, "Combined optical distance sensing and electropalatography to measure articulation," in *Interspeech 2011*, Florence, Italy, 2011, pp. 285–288.

[6] A. A. Wrench, "Advances in EPG palate design," *Advances in Speech-Language Pathology*, vol. 9, no. 1, pp. 3–12, 2007.

[7] S. Singh, *Fermats letzter Satz: Die abenteuerliche Geschichte eines mathematischen Rätsels*. Carl Hanser Verlag, München, 1998.

[8] N. Nguyen, "EPG bidimensional data reduction," *European Journal of Disorders of Communication*, vol. 30, no. 2, pp. 175–182, 1995.

[9] "Javanns," http://www-ra.informatik.uni-tuebingen.de/software/JavaNNS/welcome.html.