CrossMark

ORIGINAL ARTICLE

Predicting 3D lip movement using facial sEMG: a first step towards estimating functional and aesthetic outcome of oral cancer surgery

Merijn Eskes[1,2] · **Maarten J. A. van Alphen**[1] · **Ludi E. Smeele**[1,3] ·
Dieta Brandsma[4,5] · **Alfons J. M. Balm**[1,3] · **Ferdinand van der Heijden**[1,2]

Abstract In oral cancer, loss of function due to surgery can be unacceptable, designating the tumour as functionally inoperable. Other curative treatments can then be considered. Currently, predictions of these functional consequences are subjective and unreliable. We want to create patient-specific models to improve and objectify these predictions. A first step was taken by controlling a 3D lip model with volunteer-specific sEMG activities. We focus on the lips first, because they are essential for speech, oral food transport, and facial mimicry. Besides, they are more accessible to measurements than intraoral organs. 3D lip movement and corresponding sEMG activities are measured in five healthy volunteers, who performed 19 instructions repeatedly, to create a quantitative lip model by establishing the relationship between sEMG activities of eight facial muscles bilaterally on the input side and the corresponding 3D lip displacements on the output side. The relationship between 3D lip movement and sEMG activities

was accommodated in a state-space model. A good relationship between sEMG activities and 3D lip movement was established with an average root mean square error of 2.43 mm for the first-order system and 2.46 mm for the second-order system. This information can be incorporated into biomechanical models to further personalise functional outcome assessment after treatment.

Keywords Oral cancer · Surface electromyography · Lips · State-space estimation · Kalman filter · Principal component analysis

1 Introduction

Oral cancer, including that of the lips, is the sixth most common cancer worldwide [15]. Surgery is still standard care [14] and can lead to deterioration of speech, swallowing and mastication with serious consequences on quality of life [6]. If surgical resection of a tumour results in an unacceptable loss of function, the tumour is designated as functionally inoperable, and other curative treatment options such as chemoradiotherapy or radiotherapy can serve as alternative treatments [7].

Accurate prediction of the functional consequences of surgery is an urgent need to make the right choice of treatment [19]. Functional prediction using virtual surgery is complex and involves several aspects of patient-specific anatomical geometry, biomechanical tissue properties, branching and distribution pattern of the nervous system and the muscle activation signals that control a particular function. Biomechanical modelling, including the muscular system, in the oral and oropharyngeal region, has been the subject of ongoing research [10, 16, 18, 23].

✉ Merijn Eskes
  m.eskes@nki.nl

1   Department of Head and Neck Oncology and Surgery, Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

2   MIRA Institute of Biomedical Engineering and Technical Medicine, University of Twente, Drienerlolaan 5, 7522 NB Enschede, The Netherlands

3   Department of Oral and Maxillofacial Surgery, Academic Medical Center, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands

4   Department of Neuro-Oncology, Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

5   Department of Neurology, Slotervaart Hospital, Louwesweg 6, 1066 EC Amsterdam, The Netherlands

 Springer

This paper focuses on the lips, since these are essential for speech, oral food transport and facial mimicry. To create a predictive model, a continuum of 3D lip shapes is needed ultimately to perform virtual surgery on a model. Former research on lip modelling, utilising surface electromyography (sEMG) of facial muscles, is more phonetic in nature and is mainly focused on the categorisation of facial expressions [4], categorisation of vowels [1] and words [2].

To our knowledge, only two studies have described lip shape modelling in combination with quantitative lip pose estimation using facial electromyography (EMG). Honda et al. [5] recorded lip motion in the 2D frontal projection of the face and sEMG signals from only one side. They used a direct linear mapping of EMG to the lip coordinates based on multiple regression analysis. A visual comparison between the measured and modelled lip shapes was made. Lucero and Munhall acquired intramuscular EMG data, using hooked-wire bipolar electrodes, of one side of the face, and simultaneously measured lip and face displacements on the other side [9]. The relationship between EMG activity and marker displacements was based on a facial finite element model and the connection between EMG feature and the steady-state force generated by the corresponding muscle was presumed to be linear. The quantitative evaluation was expressed in terms of cross-correlation between model-predicted and measured displacements of the individual markers. For markers on the lips, these cross-correlations were rather low (mean values: 0.0–0.91) with very low cross-correlation for protrusion. The instructions in these studies differed, Honda et al. used five Japanese vowels, and the subject in the study of Lucero and Munhall was asked to produce an English sentence. Both models were tested on one volunteer.

The goal of this study is twofold. First we want to demonstrate that sEMG signals contain enough information for controlling 3D dynamic models of facial expressions, particularly lip movements. The second goal is to establish the optimal processing configuration to extract information from facial sEMG data. To avoid the complexity and pitfalls of detailed biomechanical models, we first focus on an empirical model. If the results of this empirical model are promising, the premise is justified that sEMG signals are very useful to solve the ambiguity problems in inverse dynamic modelling [17]. In addition, our study also should reveal which sEMG processing configuration, e.g. sEMG feature type and time window, is most promising for sEMG-based inverse modelling. The ambiguity problem of inverse dynamic modelling stems from the fact that a desired movement can be accomplished in various ways [17]. The activation pattern that causes the desired movement is not unique. The addition of the sEMG could provide further information about a patient-specific activation pattern.

2 Methods

2.1 Volunteers and data acquisition

Data were obtained from five healthy volunteers ($k = 1, \ldots, 5$) consisting of two males and three females, age ranging from 21 to 30. The recording sites of the skin were cleansed with NuPrep abrasive gel and alcohol. The sEMG signals were recorded using a TMSi® Porti™ system (TMSi®, Oldenzaal, the Netherlands). The micro-sEMG sintered disc-shaped surface electrodes (1.5 mm diameter, Ag/AgCL, with shielded cables) were placed above eight muscles on both sides of the face ($n = 1, \ldots, 16$), as shown in Fig. 1. The locations were chosen based on human lip anatomy and a study of Lapatki et al. [8] showing the effects on lip shapes. Additionally, a common ground reference electrode was applied with a self-adhesive button

Fig. 1 *Left* locations of electrodes, orientation markers and lip markers. *Right* measured facial muscles (excluding the digastric muscle) [16] © Springer, with permission of Springer
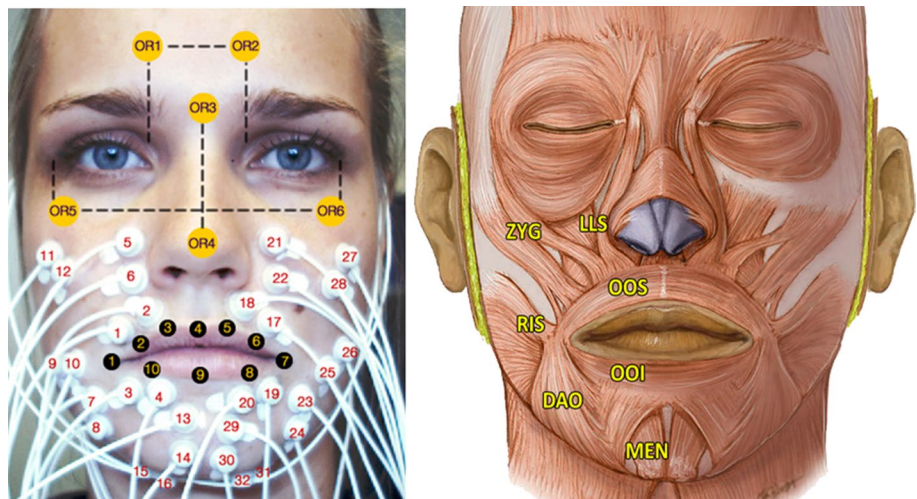
Table 1 Muscle, muscle function and corresponding electrode number

Muscle	Function	Electrodes number (right/left)
Zygomaticus major (ZYG)	Elevates the corners of the mouth in lateral direction	11, 12/27,28
Risorius (RIS)	Retracts angle of mouth	9, 10/25, 26
Orbicularis oris superior (OOS)	Narrows orifice of mouth, purses lips and puckers lip edges	1, 2/17, 18
Orbicularis oris inferior (OOI)	Narrows orifice of mouth, purses lips and puckers lip edges	3, 4/19, 20
Mentalis (MEN)	Draws up the skin of the chin and causes the lower lip to protrude	13, 14/29, 30
Depressor anguli oris (DAO)	Draws the corners of the mouth downwards and laterally	7, 8/23, 24
Levator labii superioris (LLS)	Elevates and everts upper lip	5, 6/21, 22
Digastricus (DIG)	Depresses mandible, opening mouth and/or elevates larynx	15, 16/31, 32

electrode on the left wrist. In Table 1, the measured muscles, their functions and electrode number, corresponding to the numbering in Fig. 1, are given. Sixteen facial markers were defined using a skin marker. Ten markers covered the lip contour ($m = 1, \ldots, 10$). The other six markers ($m_{OR} = 1, \ldots, 6$) were located on the face (cheeks, nose and forehead; see Fig. 1) and were used to compensate for head movement. The volunteers were positioned in front of a triple-camera set-up consisting of three cameras (Basler avA1000-100gc), which recorded the lip movement at 100 frames per second.

2.2 Instructions to volunteers

A study of van Son et al., showed that Dutch (experienced) lip readers were able to recognise five consonantal and five vowel visemes [21]. Visemes are groups of speech sounds that are visually indistinguishable. These Dutch viseme instructions were used in this study. Besides these visemes, six facial expressions that maximised independent contraction of the measured muscles were included. These selected expressions were based on the work of Lapatki et al. [8]. Lastly, two asymmetric motions were performed from left to right to left with closed lips, and with open lips, and one dynamic motion transfer between two expressions; purse lips to voluntary smiling to purse lips. Each volunteer was asked to repeat the 19 instructions ($i = 1, \ldots, 19$) five times ($r = 1, \ldots, 5$). The instructions are shown Table 2.

2.3 Data processing and analysis

2.3.1 sEMG preprocessing

The sEMG signals $s_n(t, i, r)$ were recorded in bipolar configuration with a sample frequency of 2048 Hz. Here, t is the time index, i is the instruction and r is the repetition number. All recorded signals were band-pass filtered with a high- and low-pass fourth-order Butterworth filter with cut-off frequencies of, respectively, 15 and 500 Hz, in accordance with van Boxtel [20].

Many different sEMG feature types have been proposed in the literature. Based on the results of Phinyomark et al., who examined 37 feature types, and our results of a preliminary experiment, we chose to investigate four feature types given in Table 3 [11]. Thresholds for the WAMP feature ($x_{\lim}$) were set to 10 and 20 mV. With all 16 sEMG channels stacked in a vector the result is denoted: $\mathbf{g}_f(t, i, r) \in \mathbb{R}^{16}$, with $f = 1, \ldots, 5$ the feature type. Features were calculated over a sliding window with maximum overlap. The different window lengths examined were: 50, 100, 150, 200, 250 and 300 ms.

The videos were recorded concurrently with the sEMG. To synchronise the recorded sEMG signals with the video recordings a synchronisation pulse was fed to the TMSi® Porti™ system when the cameras started their recordings. Thereafter the sEMG signals were cut and resampled to 100 Hz, equivalent to the frame rate of the cameras.

There is a small time delay between a sEMG activity and the corresponding muscle activation. It is difficult to define a default value for this delay. Honda et al. [5] used 70 ms, whereas Vatikiotis et al. [22] used different delays varying from 0 to 100 ms. By minimising the estimation errors of the lip marker positions, we empirically determined a mean muscle activation delay of 30 ms, which we compensated in all records.

2.3.2 Video preprocessing

The facial markers were tracked in the images of the three cameras, and the 2D coordinates were reconstructed to a set of 3D coordinates. The root mean square (RMS) error of the 3D localisation of markers, obtained via the leave-one-out method, was 0.73 mm. The resulting 3D positions of the ten markers on the lip, corrected for head movement, are denoted by $\mathbf{X}(t, i, r) \in \mathbb{R}^{30}$.

2.3.3 The measurement model

State-space estimation requires the availability of a measurement model that links the sEMG features

Table 2 Instructions: visemes (1–10), facial expressions (11–17) and asymmetric movements (18–19)

#	Viseme	Phonemes	Description/Location	Instruction
1	1	[p], [b], [m]	Bilabial consonants	APA
2	2	[f], [v], [ʋ]	Labiodental consonants	AVA
3	3	[s], [z], [ʃ]	Non-labial front fricatives	AZA
4	4a	[t], [d], [n], [j], [l]	Other nonlabial front consonants	ATA
5	4b	[k], [χ], [ʀ], [ŋ]	Other nonlabial back consonants	AKA
6	5a	[i], [I], [e], [ɛ]	Close and half-close front vowels (unrounded)	KIEK
7	5b	[ɛi], [a], [o]	Half-open and open vowels (unrounded)	KAK
8	6	[u], [y], [œ], [ɔ]	Short back vowels (rounded)	KOEK
9	7	[Ø], [o]	Long back vowels (rounded)	KOOK
10	8	[au], [œy]	Closing and rounding diphthongs	KAUK

11   12   13   14   15   16

17   18   19

Table 3 sEMG features

f		Feature	Formula
1		RMS	$\sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}$
2		MAV	$\frac{1}{N}\sum_{i=1}^{N} |x_i|$
3		WL	$\sum_{i=1}^{N-1} |x_{i+1} - x_i|$
4	$x_{\lim} = 10$ mV	WAMP	$\sum_{i=1}^{N-1} \left[f(|x_n - x_{n+1}|) \right]$
5	$x_{\lim} = 20$ mV		with $f(x) = \begin{cases} 1 & \text{if } x \geq x_{\lim} \\ 0 & \text{otherwise} \end{cases}$

$\mathbf{g}_f(t, i, r)$ to marker positions $\mathbf{X}(t, i, r)$. The relationship between these quantities is nonlinear, whereas a linear model was preferred. To arrive at a linear approximation, a truncated Taylor series in $\mathbf{g}_f(t, i, r)$ up to order two was used. For this, the 16D feature vector $\mathbf{g}_f(t, i, r)$ was augmented with all the 136 quadratic terms and cross products of its elements yielding a $152 \times D$ vector $\underline{\mathbf{g}}_f(t, i, r)$.

To establish the measurement model, first a principal component analysis (PCA) was applied. Suppose that a training set consisting of J observed sEMG

features $\mathbf{g}(j)$ and corresponding marker positions $\mathbf{X}(j)$, with $j = 1, \ldots, J$, is available. The exact construction of this training set will be explained later. PCA was applied to the concatenation of these vectors:

$$\mathbf{z}(j) \stackrel{\text{def}}{=} \begin{bmatrix} \mathbf{X}(j) \\ \underline{\mathbf{g}}(j) \end{bmatrix} \qquad (1)$$

The dimension of the vectors $\mathbf{z}(j)$ is 182. The set was normalised with respect to mean and variance of each element before applying the PCA, because the sEMG features and the 3D coordinates present two different physical dimensions. The PCA resulted in a $182 \times D$-dimensional orthogonal matrix $\mathbf{Y}$ containing the first D principal components of the set. Encoding of a vector $\mathbf{z}$ in a D-dimensional coefficient vector $\mathbf{b}$, and subsequent decoding, occurs according to:

$$\begin{aligned} \mathbf{b} &= \mathbf{Y}^{\mathrm{T}}\mathbf{z} \ (\text{encoding}) \\ \hat{\mathbf{z}} &= \mathbf{Y}\mathbf{b} \ (\text{decoding}) \end{aligned} \qquad (2)$$

The mean of the coefficient vector $\mathbf{b}$ is zero, and the covariance matrix $\mathbf{C_b}$ is diagonal with the elements sorted in descending order.

To arrive at a (pseudo-) linear measurement model, we constructed the matrix $\mathbf{Y_g}$ from $\mathbf{Y}$ by leaving out the first 30 rows corresponding to the positions $\mathbf{X}$. We then have:

$$\hat{\underline{\mathbf{g}}} = \mathbf{Y_g}\mathbf{b} \qquad (3)$$

Suppose that the residuals of $\hat{\underline{\mathbf{g}}}$ are given by $\mathbf{v}$, such that $\underline{\mathbf{g}} = \hat{\underline{\mathbf{g}}} + \mathbf{v}$, then:

$$\underline{\mathbf{g}} = \mathbf{Y_g}\mathbf{b} + \mathbf{v} \qquad (4)$$

This can be regarded as a linear measurement model of $\mathbf{b}$ with $\underline{\mathbf{g}}$ the measurement vector, $\mathbf{Y_g}$ the measurement matrix and $\mathbf{v}$ the measurement noise. The covariance matrix $\mathbf{C_v}$ of $\mathbf{v}$ is a 152×152-dimensional matrix which can easily be estimated from the training set. Due to augmentation of $\underline{\mathbf{g}}$ with quadratic terms, the measurement noise is not guaranteed to be uncorrelated, and the matrix $\mathbf{C_v}$ might be non-diagonal.

2.3.4 State-space modelling

For dynamic modelling, two state-space models were implemented, a first-order and a second-order system. In the first-order system, a time series $\mathbf{b}(t)$ was modelled dynamically with:

$$\mathbf{b}(t + 1) = \mathbf{F}\mathbf{b}(t) + \mathbf{w}(t) \qquad (5)$$

$\mathbf{b}(t)$ is the D-dimensional state vector, and $\mathbf{F}$ is the $D \times D$ system matrix. The process noise $\mathbf{w}(t)$ was assumed to be zero mean and uncorrelated in time. Its covariance matrix is $\mathbf{C_w}$. The system matrix $\mathbf{F}$ was estimated from the training set using $\mathrm{E}\big[\mathbf{b}(t+1)\mathbf{b}^T(t)\big] = \mathbf{F}\,\mathrm{E}\big[\mathbf{b}(t)\mathbf{b}^T(t)\big]$.

Here, $\mathrm{E}[]$ is the expectation operator; hence, $\hat{\mathbf{F}} = \overline{\mathbf{b}(t+1)\mathbf{b}^T(t)}\left(\overline{\mathbf{b}(t)\mathbf{b}^T(t)}\right)^{-1}$. The covariance matrix $\mathbf{C_w}$ can be estimated from the training set using $\mathbf{w}(t) = \mathbf{b}(t+1) - \mathbf{F}\mathbf{b}(t)$. Preliminary experiments showed that both $\mathbf{F}$ and $\mathbf{C_w}$ are diagonal. This was expected as the PCA decorrelated the coefficients $\mathbf{b}(t)$. In addition, the system matrix $\mathbf{F}$ appeared to approximate the identity matrix $\mathbf{I}$. This was also expected as the sampling period, 10 ms, is rather small compared to the expected time constant of lip motions.

In the second-order system, the state vector was defined as:

$$\mathbf{x}(t) \stackrel{\text{def}}{=} \begin{bmatrix} \mathbf{b}(t-1) \\ \mathbf{b}(t) \end{bmatrix} \qquad (6)$$

with associated state equation:

$$\mathbf{x}(t+1) = \mathbf{F}\mathbf{x}(t) + \mathbf{w}(t) \quad \text{with} \quad \mathbf{F} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ \mathbf{F_1} & \mathbf{F_2} \end{bmatrix} \qquad (7)$$

Preliminary experiments showed that the submatrices $\mathbf{F_1}$ and $\mathbf{F_2}$ are diagonal which again is in line with the uncorrelatedness of the coefficients $\mathbf{b}(t)$. Equation (7) models D decoupled second-order autoregressive (AR) models, one for each coefficient $b_n(t)$ in $\mathbf{b}(t)$, i.e.

$$\begin{aligned} b_n(t+1) &= \alpha_n b_n(t) + \beta_n b_n(t-1) + w_n(t) \\ &\text{with} \quad n = 1, \ldots, D \end{aligned} \qquad (8)$$

where α_n is a diagonal element from $\mathbf{F_2}$ and β_n a diagonal element from $\mathbf{F_1}$. The AR models represent second-order differential equations in the continuous time that are characterised by their natural frequencies f_n and relative damping ζ_n given by:

$$f_n = \frac{\sqrt{1 - \alpha_n - \beta_n}}{2\pi T} \quad \text{and} \quad \zeta_n = \frac{-\alpha_n - 2}{2\sqrt{1 - \alpha_n - \beta_n}} \qquad (9)$$

where T is the sampling period. The natural frequency determines the bandwidth of the corresponding coefficient. The damping determines the spectrum of the signal around the natural frequency. We used these parameters to fine-tune the state-space model during training.

The process noise $\mathbf{w}(t)$ has zero elements in the first D elements. Thus, the covariance matrix is built as follows:

$$\mathbf{C_w} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C_{22}} \end{bmatrix} \qquad (10)$$

$\mathbf{C_{22}}$ is a diagonal matrix as the coefficients of a PCA are uncorrelated. Preliminary results showed that this was indeed the case. To determine the influence of dynamic modelling we also performed static modelling by enforcing the Kalman filter, which is described below, to use only measurements, and to ignore the predictions. This was effectuated by setting the standard deviation of the process noise to almost infinity.

Fig. 2 Optimisation and evaluation of design parameters using cross-validation
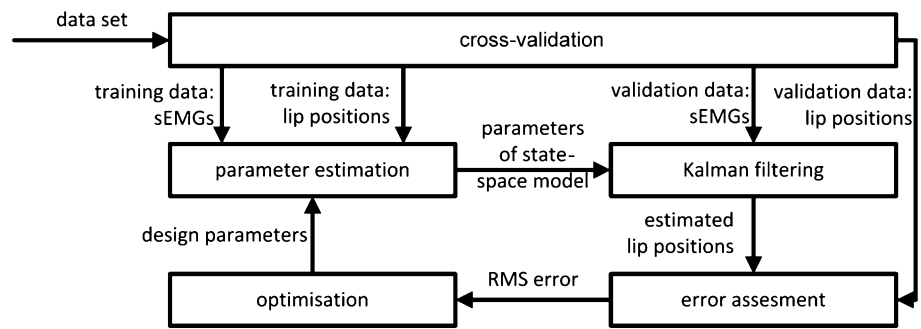
2.3.5 Estimation

The estimation of the coefficients of the PCA was done with a discrete Kalman filter. The dimension of the state vector is in the first-order system D and in the second-order system $2 \times D$. In practice, D, being the result of the PCA, is much smaller than the dimension of the measurement vector, $\mathbf{g}(t)$, which is 152. Therefore, the Kalman filter was used in the following form:

$$\left. \begin{array}{l} \hat{\mathbf{x}}(t|t-1) = \mathbf{F}\hat{\mathbf{x}}(t-1|t-1) \\ \mathbf{C}(t|t-1) = \mathbf{F}\mathbf{C}(t-1|t-1)\mathbf{F}^T + \mathbf{C}_w \end{array} \right\} \text{prediction}$$

$$\left. \begin{array}{l} \mathbf{C}(t|t) = \left(\mathbf{C}^{-1}(t|t-1) + \mathbf{H}^T \mathbf{C}_{\mathbf{v}}^{-1} \mathbf{H} \right)^{-1} \\ \hat{\mathbf{x}}(t|t) = \mathbf{C}(t|t) \left(\mathbf{C}^{-1}(t|t-1)\hat{\mathbf{x}}(t|t-1) + \mathbf{H}^T \mathbf{C}_{\mathbf{v}}^{-1} \underline{\mathbf{g}}(t) \right) \end{array} \right\} \text{updating}$$

(11)

This is computationally more efficient than the typical form. In Eq. (11), $\mathbf{H}$ is the measurement matrix, which equals $\mathbf{Y_g}$ in the first-order system and $\begin{bmatrix} \mathbf{0} & \mathbf{Y_g} \end{bmatrix}$ in the second-order system.

2.3.6 Training and testing

The algorithm needs training data to find the PCA components $\mathbf{Y}$, the covariance matrices $\mathbf{C_v}$ and $\mathbf{C_w}$ and in case of the first-order system the system matrix $\mathbf{F}$ and in case of the second-order system, its submatrices $\mathbf{F}_1$ and $\mathbf{F}_2$. The dimension D of the PCA is a design parameter. Additional design parameters were introduced to fine-tune the models. These were as follows:

- The measurement noise covariance matrix $\mathbf{C_v}$ was corrected with a regularisation parameter c_v. Instead of $\mathbf{C_v}$, the matrix $(1 - c_v)\mathbf{C_v} + c_v\overline{\mathbf{C_v}(\ell, \ell)}\,\mathbf{I}$, with $\overline{\mathbf{C_v}(\ell, \ell)}$ the average of the diagonal elements, was used.
- The process noise covariance matrix $\mathbf{C_w}$ was corrected with a regularisation parameter c_w. That is, the submatrix $\mathbf{C}_{22}$ was replaced by the matrix $(1 - c_w)\mathbf{C}_{22} + c_w\text{diag}(\overline{\mathbf{C}_{22}(\ell, \ell)})$. Here, $\text{diag}(\overline{\mathbf{C}_{22}(\ell, \ell)})$ is the diagonal matrix that is built with a smoothed version of the diagonal elements of $\mathbf{C}_{22}$.

- The diagonal matrices $\mathbf{F}_1$ and $\mathbf{F}_2$, which holds the second-order AR parameters α_n and β_n, respectively, were corrected by application of a proportionality constant to the corresponding natural frequencies and damping by constants c_f and c_d. So, instead of f_n and ζ_n, the parameters $c_f f_n$ and $c_d \zeta_n$ were used.

This resulted into three design parameters, D, c_v and c_w, for the first-order system and five design parameters, D, c_v, c_w, c_f and c_d, for the second-order system. These parameters were optimised using training data.

We performed cross-validation for training and testing. The procedure is depicted in Fig. 2. It was applied per volunteer, per feature type and per window size. Data from the various instructions were pooled by concatenating the data: $\underline{\mathbf{g}}_f(t, r) = \begin{bmatrix} \underline{\mathbf{g}}_f(t, 1, r) & \cdots & \underline{\mathbf{g}}_f(t, 19, r) \end{bmatrix}$. The data from four repetitions were pooled to get the training data: $\underline{\mathbf{g}}_f(t) = \begin{bmatrix} \underline{\mathbf{g}}_f(t, 1) & \cdots & \underline{\mathbf{g}}_f(t, 4) \end{bmatrix}$. Testing was performed on the fifth repetition. Cross-validation took place by rotating the repetitions. The final evaluation criterion was defined as the RMS of the error calculated over all marker coordinates and all repetitions. The design parameters were obtained by minimisation of the RMS error by varying these parameters one by one and applying successive parabolic optimisation. The one-sided paired Wilcoxon test was used to test for significant differences between the static and the two dynamic systems. The one-sided test was justified because the static model is in fact included in the dynamic model as a special case, and as such the optimised performance of the dynamic model cannot be less than the static model.

3 Results

The best results for the static, first-order and second-order state-space models are summarised in Table 4 for the individual subjects and on average. The first-order system for state-space modelling performed best on average, with a RMS error of 2.43 mm on average. The first-order and the second-order system showed

Table 4 RMS error, optimal feature and window, found per volunteer and on average for the static, first-order and second-order dynamic model

Volunteer	1	2	3	4	5	Average
Static						
RMS error (SD) (mm)	2.34 (0.21)	2.55 (0.21)	3.02 (0.12)	2.32 (0.15)	2.92 (0.14)	2.70 (0.19)
Feature	WAMP (20 mV)	WAMP (20 mV)	WAMP (20 mV)	WAMP (10 mV)	WAMP (10 mV)	WAMP (10 mV)
Window (ms)	200	200	300	250	300	250
First-order system						
RMS error (SD) (mm)	2.10 (0.17)	2.29 (0.19)	2.64 (0.17)	2.10 (0.19)	2.66 (0.19)	2.43 (0.18)
Feature	WAMP (20 mV)	WL	WAMP (20 mV)	WAMP (10 mV)	WAMP (10 mV)	WAMP (10 mV)
Window (ms)	200	200	200	250	250	200
Second-order system						
RMS error (SD) (mm)	2.02 (0.19)	2.42 (0.18)	2.58 (0.18)	2.13 (0.21)	2.66 (0.21)	2.46 (0.18)
Feature	WAMP (20 mV)	WL	WAMP (20 mV)	WAMP (10 mV)	WAMP (10 mV)	WAMP (10 mV)
Window (ms)	200	250	150	200	200	200

Table 5 Optimal settings averaged over the volunteers obtained per feature and system order

	RMS		MAV		WL		WAMP10		WAMP20	
System order	First	Second	First	Second	First	Second	First	Second	First	Second
RMS error (mm)	2.74	2.70	2.67	2.64	2.50	2.50	2.43	2.46	2.45	2.46
Window (ms)	200	200	250	200	200	200	200	200	200	200

statistically significant better results than the static system ($p = 0.03$). No significant difference was found between the two dynamic systems. Four subjects showed the best results when the WAMP feature was used. The optimal threshold $x_{\lim}$ differed between the subjects. A window length of 200 ms performed best on average. The average RMS error was 2.46 mm for the second-order state-space model, also using the WAMP feature with $x_{\lim} = 10$ mV and a window length of 200 ms. As expected, static modelling showed poorer results, but performed also best when the WAMP feature was used.

The RMS errors for the different features for the first-order and second-order system are presented in Table 5. It can be seen that for both $x_{\lim} = 10$ mV and $x_{\lim} = 20$ mV the WAMP feature performance was comparable. Regarding the other features, the WL performed slightly worse compared to the WAMP. RMS and MAV showed the poorest results.

The influence of the different parameters and window length on the error in the second-order system can be seen in Fig. 3. Each graph shows the influence of one parameter on the RMS error while the others are set to values which lead to the optimal results on average. The dimension of the PCA, D, shows a plateau after 20 components. Optimal values of 0.1 and 3.4 were found for c_v and c_f, respectively. The regularisation parameter c_d had the minimum error at a factor of 0.7. The constant
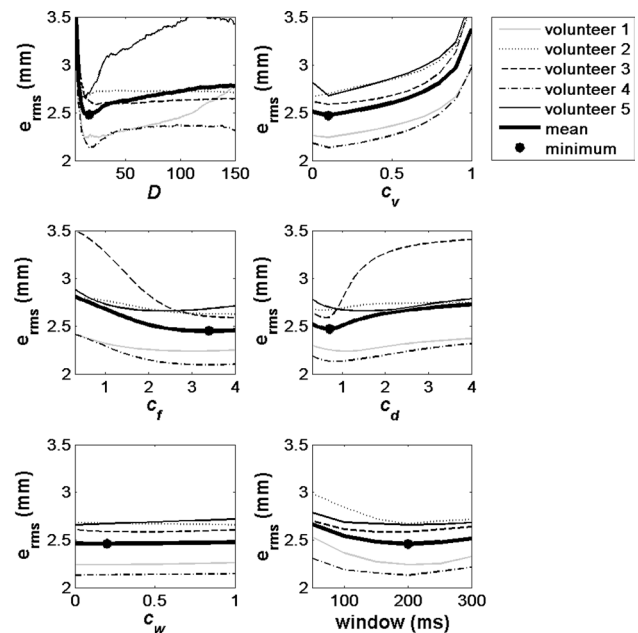

Fig. 3 Dependence of the RMS error on the various parameters and window length in the second-order system

c_w showed little influence but had an optimum on average at 0.2. The window lengths showed a similar trend in all subjects, with the best results for medium length windows. For the first-order system, comparable values were found.

Table 6 Differences in experimental set-up in related studies

	Current study	Honda et al. [5]	Lucero and Munhall [9]
Dimension	3D	2D	3D
Measurement	Bilateral	Unilateral	Unilateral
Muscles	16	6	7
Lip markers	10	7	5
EMG	Surface	Surface	Hooked-wire
Jaw movement	Digastric muscle	Omitted	Manual
Model	PCA MMSE + Kalman	Multiple regression	FEM
Instructions	Visemes and facial expressions (19)	Japanese vowels (5)	English sentence
Volunteers	5	1	1
Error assessment	RMS distance	Visual	Visual and cross-correlations

4 Discussion

For the prediction of the functional and aesthetic consequences of treatment in oral cancer, dynamic models of the lips are required. Biomechanical modelling is physics based and as such the most direct method to predict these consequences. However, finding the patient-specific muscle activation signals needed for the biomechanical models is difficult [24]. sEMG signals may contain information to help in finding these patient-specific activations signals. To find the optimal sEMG processing configuration and to prove that sEMG signals contain sufficient information to do so, the current study describes an empirically derived model that is able to estimate the dynamics of lip displacements with an average RMS error of 2.43 mm. This empirical model is sEMG driven, which incorporates volunteer-specific information. As far as we know, we are the first who expressed distance errors of lip motion prediction based on sEMG features.

The approach used here, incorporated the dynamics of the system by means of a state-space model. To test whether dynamical modelling was superior to static modelling, we implemented both. As expected, incorporation of dynamics improved the model. In comparison with the static system RMS errors decreased in every volunteer, with an average of 0.27 mm for the first-order system and 0.24 mm for the second-order system. An advantage of a dynamic system is that bandwidth can be sacrificed to improve the signal-to-noise ratio. Apparently, in the current application such a sacrifice pays off, but not drastically.

The difference between the first-order and second-order system is negligible. A higher-order system has more parameters which have to be estimated, making the filter more sensitive for deviations in those parameters to the optimum settings. An optimal equilibrium has to be found between modelling accurate dynamic behaviour for which higher-order systems are beneficial, and confining the impact of errors in the estimated parameters for which a lower-order system is preferred. In this study, the advantages of a second-order system over a first-order system did not outweigh the errors induced by the deviations in the estimated parameters.

The fudging parameters were used to optimise the model per volunteer and hence make it volunteer-specific. Only the regularisation parameter c_w for the process noise covariance matrix $\mathbf{C_w}$ did not have much influence. For each parameter, a similar trend was seen regarding the optimal values, but the level of influence differed per volunteer. The optimal values found in this study can be used to set the limits for future volunteers, thereby decreasing computational time of the parabolic optimisation. The dimension of the PCA reached a plateau at 20, four dimensions more than the original 16 dimensions, suggesting that the cross products of the sEMG provided additional information. Preliminary experiments indicated that leaving out all the nonlinear cross products seriously deteriorated the results. Hence, the nonlinearity of the system is substantial. Finally, window length was optimal at medium lengths. A possible explanation is that short window lengths are prone to noise, whereas longer window lengths smooth the signals too much.

The different features also had a noticeable influence on the RMS error. The WAMP features with $x_{lim} = 10$ and $x_{lim} = 20$ mV were most promising. Perhaps that thresholds in-between these values could perform better. One can also think of optimising x_{lim} per muscle channel for optimal results. The widely used RMS feature performed worse. This was also found by Phinyomark et al. [11].

Because of different error assessments, the differences of our model compared to studies in the literature will be discussed qualitatively. The main differences are stated in Table 6. The current model showed results for a more extensive set of instructions, including asymmetric movements. Furthermore, more 3D lip markers and more muscles were included. Our model predicts 3D movement of the lips based on measurements on both sides of the face and therefore is more realistic. Honda et al. used a linear statistical approach, which is inadequate for modelling nonlinear soft tissue changes. To allow the model to cope with nonlinear behaviour, we calculated the

cross products of the sEMG signal features to add nonlinearity. The model of Honda et al. did not include the factor of jaw movement, making the estimation of vertical movement prone to errors. Lucero and Munhall controlled jaw movement by tracking an optical marker instead of EMG signals. Non-surprisingly, the modelled facial tissue followed this movement well. We added sEMG measurements of the digastric muscle to make the prediction of jaw movement possible. Both Honda et al. and Lucero and Munhall measured EMG signals on one side of the face, disregarding asymmetry in facial morphology and lip movement as described by Campbell [3]. The use of hooked-wire, as used by Lucero and Munhall, or needle electrodes is attractive to overcome the problem of crosstalk, but for clinical applications this will be impractical because they are invasive and consequently patient-unfriendly. Therefore, in generating future personalised 3D models controlled by EMG signals, sEMG will have our preference, being easy applicable and patient-friendly. Furthermore, we tested our model in five subjects, indicating it is general applicable.

A limitation of the proposed set-up is the high number of required sEMG electrodes. This makes the current approach time-consuming which is inconvenient in future clinical practice. Monopolar derivations can be used to halve the number of electrodes; however, this configuration does not reduce unwanted noise from the recordings by using the differential amplifier design [12]. Another approach to lessen the number of electrodes is to identify less influential muscle channels for the estimation of motion, and include only those channels that affect motion prediction most.

Another difficulty is the variability in muscle anatomy, and overlying soft tissue, which makes standardisation of the measurements difficult. Additionally, physiological orofacial functions usually require simultaneous contraction of various muscles. These muscles therefore lack training in isolated contractions, resulting in relatively high co-contraction of muscles. Also volunteers can use different muscle activation patterns to perform the same instruction. We saw similar results when facial expressions were performed as described by Schumann et al. [13]. Most volunteers were able to selectively activate the LLS, whereas most volunteers had difficulty in pulling their lip corners down (DAO). Purse lips, pout lips and voluntary smiling all induced multiple muscle activations, showing the difficulty in selective muscle activation of facial muscles.

The two main pillars of our study were to demonstrate that sEMG signals contain sufficient information to control 3D dynamic models of lip movements and to determine the best sEMG processing configuration for this purpose. These two steps are necessary for our ultimate goal to enable inverse biomechanical modelling of the lips, oral cavity and tongue, in order to retrieve patient-specific muscle activation signals inducing oral functions. These activation signals are needed to enable prediction of functional consequences after surgery.

Besides patient-specific activation signals, patient-specific anatomical information and tissue parameters are required for simulating treatment effects. The current model does not account for these aspects yet. However, a biomechanical model should incorporate this physical relationship and patient-specific parameters before mimicking performed treatments. The simulated activations controlling the biomechanical model should be similar to the actual muscular control of the patient. sEMG is an instrument to provide the information for these simulated activations. Unfortunately, sEMG is a rough estimate, because of a nonlinear relationship, crosstalk, misplacement of electrodes and other artefacts. We showed that the relationship between 3D lip motion and sEMG can be accurately described by a statistical model. So it can be expected that with our approach the ambiguity problem of inverse modelling can be solved. Our next studies will focus on the relationship between activation signals and sEMG in biomechanical models.

5 Conclusion

This study presented a next step towards the personalisation of the functional outcome assessment after treatment of oral cancer. The two dynamic modelling methods proved that a continuum of 3D lip positions can be predicted based on volunteer-specific sEMG features. The discrete Kalman filter with a first-order state estimation performs slightly better than a second-order system, with a mean RMS error of 2.43 mm. The optimal sEMG processing configuration was found to be the WAMP feature with $x_{lim} = 10$ mV and a window length of 200 ms. In future studies, this method may be used to solve the problems concerning inverse modelling in biomechanical models, by reduction in the solution space and including patient-specific information.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval Although this research involved healthy human volunteers, no ethical approval was required as the study did not infringe the (psychological) integrity of the volunteers. Measurements were non-invasive and not stressful. The study was performed within the Dutch legislation regarding the Agreement on Medical Treatment Act, Personal Data Protection Act, and the Code of Conduct for Responsible Use of the Federa (Dutch Federation of Biomedical Scientific Socie-

ties). Verbal information and consent is obtained. Furthermore, verbal consent is obtained for publishing the photographs in Fig. 1 and Table 2.

References

1. Arjunan SP, Kumar DK, Yau WC, Weghorn H (2006) Unvoiced speech control based on vowels detected by facial surface electromyogram. In: Isaías P, McPherson M, Bannister F (eds) Proceedings of the IADIS international conference on e-Society. IADIS Press, Dublin, pp 381–388
2. Betts BJ, Binsted K, Jorgensen C (2006) Small-vocabulary speech recognition using surface electromyography. Interact Comput 18:1242–1259. doi:10.1016/j.intcom.2006.08.012
3. Campbell R (1982) Asymmetries in moving faces. Br J Psychol 73:95–103. doi:10.1111/j.2044-8295.1982.tb01794.x
4. Hamedi M, Salleh S-H, Swee TT, Kamarulafizam (2011) Surface electromyography-based facial expression recognition in Bi-polar configuration. J Comput Sci 7:1407–1415. doi:10.3844/jcssp.2011.1407.1415
5. Honda K, Kurita T, Kakita Y, Maeda S (1995) Physiology of the lips and modeling of lip gestures. J Phon 23:243–254. doi:10.1016/S0095-4470(95)80046-8
6. Kreeft AM, Tan IB, van den Brekel MWM et al (2009) The surgical dilemma of "functional inoperability" in oral and oropharyngeal cancer: current consensus on operability with regard to functional results. Clin Otolaryngol 34:140–146. doi:10.1111/j.1749-4486.2009.01884.x
7. Kreeft AM, Tan IB, Leemans CR, Balm AJM (2011) The surgical dilemma in advanced oral and oropharyngeal cancer: how we do it. Clin Otolaryngol 36:260–266. doi:10.1111/j.1749-4486.2011.02299.x
8. Lapatki BG, Stegeman DF, Zwarts MJ (2005) Selective contractions of individual facial muscle subcomponents monitored and trained with high-density surface EMG. In: Beurskens CHG, van Gelder RS, Heymans PG et al (eds) Facial palsies complement approaches. Lemma, Utrecht, pp 89–108
9. Lucero JC, Munhall KG (1999) A model of facial biomechanics for speech production. J Acoust Soc Am 106:2834–2842. doi:10.1121/1.428108
10. Pelteret J-PV, Reddy BD (2014) Development of a computational biomechanical model of the human upper-airway soft-tissues toward simulating obstructive sleep apnea. Clin Anat 27:182–200. doi:10.1002/ca.22313
11. Phinyomark A, Phukpattaranont P, Limsakul C (2012) Feature reduction and selection for EMG signal classification. Expert Syst Appl 39:7420–7431. doi:10.1016/j.eswa.2012.01.102
12. Robertson LD, Mustain EH, Maurer K et al (2009) EMG instrumentation. In: Kamen G, Gabriel D (eds) Essentials of electromyography, 1st edn. Human kinetics, Champaign, Illinois, p 66
13. Schumann NP, Bongers K, Guntinas-Lichius O, Scholle HC (2010) Facial muscle activation patterns in healthy male humans: a multi-channel surface EMG study. J Neurosci Methods 187:120–128. doi:10.1016/j.jneumeth.2009.12.019
14. Shah JP, Gil Z (2009) Current concepts in management of oral cancer-surgery. Oral Oncol 45:394–401. doi:10.1016/j.oraloncology.2008.05.017
15. Siegel R, Naishadham D, Jemal A (2012) Cancer statistics, 2012. CA Cancer J Clin 62:10–29. doi:10.3322/caac.20138
16. Prendergast PM (2013) Anatomy of the face and neck. In: Shiffman MA, Di Giuseppe A (eds) Cosmetic surgery. Art and techniques. Springer, Berlin, Heidelberg, p 36
17. Stavness I, Lloyd JE, Fels SS (2012) Automatic prediction of tongue muscle activations using a finite element model. J Biomech 45:2841–2848. doi:10.1016/j.jbiomech.2012.08.031
18. Stavness I, Nazari MA, Perrier P et al (2013) A biomechanical modeling study of the effects of the orbicularis oris muscle and jaw posture on lip shape. J Speech Lang Hear Res 56:878–890. doi:10.1044/1092-4388(2012/12-0200)
19. Van Alphen MJA, Kreeft AM, van der Heijden F et al (2013) Towards virtual surgery in oral cancer to predict postoperative oral functions preoperatively. Br J Oral Maxillofac Surg 51:747–751. doi:10.1016/j.bjoms.2013.06.012
20. Van Boxtel A (2001) Optimal signal bandwidth for the recording of surface EMG activity of facial, jaw, oral, and neck muscles. Psychophysiology 38:22–34. doi:10.1111/1469-8986.3810022
21. Van Son N, Huiskamp TMI, Bosman AJ, Smoorenburg GF (1994) Viseme classifications of Dutch consonants and vowels. J Acoust Soc Am 96:1341–1355. doi:10.1121/1.411324
22. Vatikiotis-Bateson E, Munhall KG, Hirayama M et al (1996) The dynamics of audiovisual behavior in speech. In: Stork DG, Hennecke ME (eds) Speechreading by humans mach. Springer, Berlin, Heidelberg, pp 221–232
23. Vogt F, Lloyd JE, Perrier P (2006) Efficient 3D finite element modeling of a muscle-activated tongue. In: Harders M, Székely G (eds) Biomedical Simulation. Springer, Berlin, Heidelberg, pp 19–28
24. Wu X, Dang J, Stavness I (2014) Iterative method to estimate muscle activation with a physiological articulatory model. Acoust Sci Technol 35:201–212. doi:10.1250/ast.35.201

Merijn Eskes, M.Sc received his M.Sc in 'Technical Medicine', with specialisation in 'Medical Sensing and Stimulation' in 2013 at the University of Twente. His master's thesis focused on the development of a 3D lip model controlled by surface electromyographic signals. He is currently pursuing his Ph.D. in a collaborative project between the Netherlands Cancer Institute and the University of Twente. His research interests are in the area of head and neck surgery and oncology, biomedical signal processing and analysis and image processing with applications in personalised medical treatment planning.

Maarten J. A. van Alphen, Ph.D. received his, M.Sc degree in 'Technical Medicine', with specialisation in 'Robotics and Imaging' in 2011, from the University of Twente. In 2015, he received his Ph.D. degree from the University of Twente. The Ph.D. thesis was entitled 'Towards a predictive model for functional loss after oral cancer treatment', and the research was accommodated in the Netherlands Cancer Institute and the chair of Robotics and Mechatronics—Medical Applications at the

University of Twente. He is currently a postdoctoral researcher at the Netherlands Cancer Institute. His research interests are in the area of technological innovations in head and neck surgery and oncology, with applications in biomechanical modelling, image processing, navigation, and treatment planning.

Prof. Ludi E. Smeele MD Ph.D. received his Dental, Ph.D. and Medical degrees, in, respectively, 1984, 1988 and 1994. His Ph.D. thesis was entitled 'Connections between the middle ear and temporomandibular joint–Ontogeny and phylogeny in some mammals'. He started his training in head and neck surgical oncology in 1995. His professional experience started at the Free University in Amsterdam with a fellowship in surgical oncology. He continued his career with a Head and Neck fellowship in the Toronto General Hospital in Toronto, Canada. He followed an observership at the division of Head and Neck Oncology in the Memorial Sloan Kettering Cancer Center, New York, USA. He became a Fellow of the European Board of Oro-Maxillofacial Surgery in 1998. Smeele is a member of various international head and neck societies and presented basic and clinical research papers at many international meetings. He was member of advisory boards and research programme committees for several international meetings. His research over the last 5 years focused on the dysphagia and trismus intervention in patients treated with concurrent chemoradiation for advanced head and neck cancer. With the Department of Paediatric Oncology of the Amsterdam Medical Center, he participates in the project on growth inhibition by surgery and brachytherapy. Since 2009 he is involved in the development of a predictive model for functional inoperability with emphasis on the classification of tongue deformities and hypoglossal innervation. Smeele was chairman of the Netherlands Head and Neck Cooperative Group from 2011 to 2012. He is director of the Residency Program of the Department of Oral and Maxillofacial Surgery at the Amsterdam Medical Center. He is staff member at the Department of Head and Neck Surgical Oncology in the Netherlands Cancer institute and at the Department of Oral and Maxillofacial Surgery in the Amsterdam Medical Center.

Dieta Brandsma, MD, Ph.D. received her Medical Biology and Medical degree in 1995 and 1998 at the University Utrecht, the Netherlands. She specialised in Clinical Neurology at the University Medical Center in Utrecht. In 2005 she received her Ph.D. degree with her thesis entitled 'Adhesion in leptomeningeal metastases—towards an early diagnosis and treatment'. After she finished her specialisation in Clinical Neurology in 2007, she continued her career as a neurologist at the Netherlands Cancer Institute—Antoni van Leeuwenhoek Hospital and Slotervaart Hospital in Amsterdam. Her scientific interest is focused on improving the CSF diagnostics for leptomeningeal metastases, new treatments for brain metastases and glioma (clinical studies), molecular markers of meningioma and the clinical neurophysiological aspects of the functional inoperability and virtual surgery of head and neck cancers. Dr.

Dieta Brandsma is a member of various international/national neuro-oncology societies. She (co-) authored more than 30 peer-reviewed papers and is a reviewer for several neuro-oncological journals.

Prof. Alfons J. M. Balm MD Ph.D. FACS FRCS received his Medical and Ph.D. degrees, in 1976 and 1982, respectively, from the Free University in Amsterdam, the Netherlands. His Ph.D. thesis was entitled 'Mononuclear phagocyte function in head and neck cancer'. After his specialisation in ENT surgery in 1982, he continued his career as an Assistant Professor at the Department of Head and Neck Surgery and Oncology, the Free University Hospital, Amsterdam. In 1984, he worked on the production of monoclonal antibodies as a Visiting Research Scientist at the University of Michigan, USA. In 1989, he moved to the Netherlands Cancer Institute and was appointed as Chairman of the Multidisciplinary Head and Neck Cooperative Group of the Netherlands Cancer Institute and the Academic Medical Center. He is a Fellow of the American College of Surgeons and Royal College of Surgeons of England. In 2003, he accepted the extraordinary Chair of Head and Neck Oncology and Surgery at the Academic Medical Center, University of Amsterdam. In the same year, he became director of the Head and Neck Oncology Residency Program in het NKI/AVL. Since 2005 he is appointed as Principal Investigator of the Academic Medical Center in Amsterdam. To date, 17 graduate students completed their Ph.D. theses under his supervision. Twenty-six Head and Neck fellowships were supervised by him. His scientific interest is focused on both clinical and translational research topics, including prognostic/predictive markers after chemoradiation, rehabilitation after total laryngectomy, functional inoperability and virtual surgery and development of new treatment strategies. Balm is a member of various international/national head and neck societies and presented basic and clinical research papers at many international meetings. He (co-) authored more than 225 peer-reviewed papers on a diversity of clinical scientific and translational research issues and serves as a reviewer for many clinical scientific journals.

Associate Prof. Ferdinand van der Heijden, Ph.D. received the M.Sc and Ph.D degree from the University of Twente in 1981 and 1992, respectively. He is currently Associate Professor affiliated with the chair of Robotics and Mechatronics—Medical Applications at the University of Twente. His research is accommodated in the MIRA Institute on Biomedical Technology and Technical Medicine. Since 2014, he is also associated with the Netherlands Cancer Institute-Antoni van Leeuwenhoek Hospital. His research interests are in the area of image processing and pattern analysis with applications in surgical navigation and medical image analysis and modelling. He is the author of two successful textbooks, 'Image Based Measurement Systems' and 'Classification, Parameter Estimation and State Estimation' (of which a second edition is in preparation).