Towards a predictive model for functional loss after oral cancer treatment

M.J.A. van Alphen

TOWARDS A PREDICTIVE MODEL FOR FUNCTIONAL

LOSS AFTER ORAL CANCER TREATMENT

De promotiecommissie

Voorzitter en secretaris:
Prof. dr P.M.G. Apers   Universiteit Twente

Promotoren:
Prof. dr. ir. C.H. Slump   Universiteit Twente
prof. dr. A.J.M. Balm   Antoni van Leeuwenhoek

Assistent promotor:
Dr. ir. F. van der Heijden   Universiteit Twente

Referent:
Dr. R.J.J.H. van Son   Antoni van Leeuwenhoek

Leden:
Prof. dr. ir. P.H. Veltink   Universiteit Twente
Prof. dr. T. Ruers   Universiteit Twente
Prof. dr. L.E. Smeele   Antoni van Leeuwenhoek
Prof. dr. S.J. Bergé   Radboud UMC

TOWARDS A PREDICTIVE MODEL FOR FUNCTIONAL LOSS AFTER ORAL CANCER TREATMENT

PROEFSCHRIFT

ter verkrijging van
de graad van doctor aan de Universiteit Twente,
op gezag van de rector magnificus,
prof. dr. H. Brinksma,
volgens besluit van het College voor Promoties
in het openbaar te verdedigen
op donderdag 27 augustus 2015 om 14.45 uur

door

Maarten Jan Antony van Alphen
geboren op 16 juli 1987
te Nijmegen

Dit proefschrift is goedgekeurd door:

Promotoren: Prof. dr. ir. C.H. Slump
                      Prof. dr. A.J.M. Balm
Assistent promotor: Dr. ir. F. van der Heijden

Contents

Chapter 1

General introduction

A key concept in treatment planning for oral cavity and oropharyngeal cancer is functional inoperability, meaning a tumour can be resected radically, but the expected functional outcome will be unacceptable.[1] Expectations as to functional outcome, however, are highly subjective and thus unreliable. Therefore, a project was launched to develop a more reliable, evidence-based approach to predict patient-specific oral function post treatment: 'Virtual Therapy for Head & Neck Cancer – Prediction of Functional Loss', or 'Virtual Therapy' for short. This thesis is a part of this project, and describes an investigation of what will be needed to further improve existing tongue and lip models. This first chapter discusses anatomy, epidemiology, staging, and treatment options, as well as function loss issues, earlier research on functional inoperability, and the ultimate goal of the Virtual Therapy project, concluding with the aim and outline of this thesis.

1

Anatomy

The Virtual Therapy project focusses on the oral cavity and the oropharynx, as treatment of advanced tumours in these regions will inevitably affect vital functions like speech, mastication, and swallowing.[2–4] Making an accurate and objective expectation of the functional outcome in treatment planning is highly significant. Tongue and lip mobility, in particular, are essential in this respect.

The oral cavity is the first part of the digestive tract, confined anteriorly by the lips, and posteriorly by the junction of the hard and soft palate and the vallate papillae on the tongue (Figure 1.1). It includes the buccal mucosa, the upper and lower gums, the hard palate, the floor of mouth, and the mobile tongue anterior to the vallate papillae. The oral cavity plays a pivotal role in vital functions like transport of food towards the pharynx, swallowing, breathing, and communication by speech.[5]

The oropharynx is located behind the oral cavity (Figure 1.1), and includes the base of tongue, the vallecula, the tonsils, the posterior pharyngeal wall, the inferior surface of the soft palate, and the uvula.[5, 6]

The tongue is a complex muscular hydrostat, meaning it is a muscular organ, lacking skeletal support.[8] It is controlled by four paired extrinsic and intrinsic muscles. The extrinsic muscles connect to other structures, like the jaw or the hyoid bone, and insert into the tongue, whereas the intrinsic muscles are located entirely within the tongue (Figure 1.2). Basically, the extrinsic muscles control the position of the tongue and the intrinsic muscles control its shape.[6]

The four extrinsic tongue muscles are the genioglossus, the styloglossus, the palatoglossus, and the hyoglossus; the intrinsic tongue muscles are the transversus, the verticalis, the superior longitudinalis, and the inferior longitudinalis. All except the palatoglossus are innervated by the hypoglossal nerve, which is the twelfth cranial nerve.[9] The exact innervation of the palatoglossus remains uncertain: most likely it is a cranial root of cranial
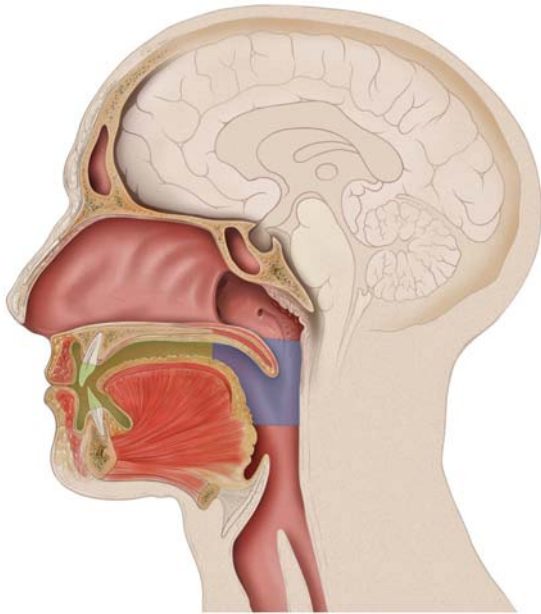
Figure 1.1 | Midsagittal plane of the head. Green area is defined as the oral cavity and blue as the oropharynx.
Adapted image, original downloaded from [7].
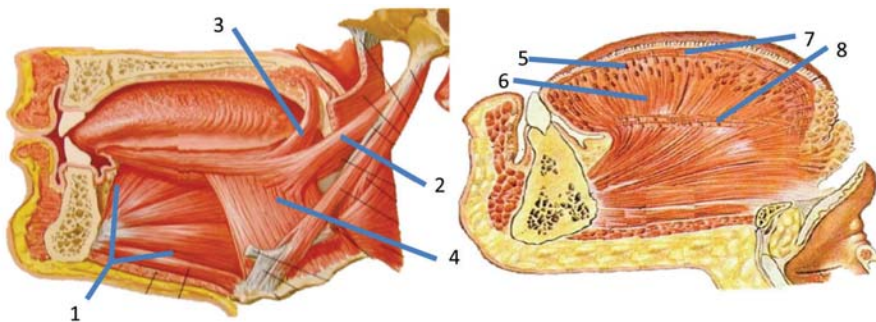
Figure 1.2 | Anatomical drawings showing the four extrinsic (left), and four intrinsic (right) tongue muscles: 1. genioglossus; 2. styloglossus; 3. palatoglossus; 4. hyoglossus; 5. transversus; 6. verticalis; 7. longitudinalis superior; 8. longitudinalis inferior.
Adapted image, original downloaded from [12, 13].

nerve XI (the accessory nerve) [10], but a vagal nerve branch (cranial nerve X) innervation has also been suggested [11].

Lip motion, and thereby the function, is controlled mainly by the complex facial architecture of overlapping and interdigitating muscles (Figure 1.3), which may show great anatomic variability both in and between individuals.[14] These muscles can be innervated both voluntarily and emotionally by different motoric branches of the facial nerve (cranial nerve VII).[15] As orofacial functions require the simultaneous contractions of several muscles, well-coordinated muscle innervation patterns are required. Also, similar functions may be performed through various alternative muscle contraction patterns, depending on personal preference.[16]

Important for mouth opening is the digastric muscle, which has a posterior and an anterior belly. The posterior belly originates from the mastoid process, the anterior belly from the symphysis menti. The two muscle bellies are interconnected by a tendon, which is attached to the body of the hyoid bone. They are innervated by different cranial nerves: the posterior belly by the digastric branch of the facial nerve, and the anterior belly by the mylohyoid nerve, which is a branch of the trigeminal nerve (cranial nerve V).[17] Liquidato et al. showed unilateral and bilateral anatomic variations for this muscle, and suggested that the unilateral variations, in particular, are of clinical importance, because they can cause asymmetry in the anterior part of the neck, the floor of mouth [18], or the temporomandibular joint [19], or induce imbalance of larynx motion.[20]

Epidemiology

Oral and oropharyngeal squamous cell carcinoma together rank sixth among the most common types of cancer [22, 23], with annual incidences of approximately 300,000 for lip and oral cavity cancers, and 142,000 for pharyngeal cancers (excluding the nasopharynx) worldwide.[24] And numbers are rising.[25] In the Netherlands, incidence figures for oral cavity and oropharyngeal cancer are 3.6 and 2.7 per 100,000, respectively.[26] Whereas in the US, the last decades have witnessed a decrease in laryngeal and oral cavity cancer, the age-adjusted incidence of oropharyngeal cancer has risen, particularly in the middle-aged (40-59 years of age).[27] A relation with human papillomavirus (HPV) has been presumed. In the Netherlands, however, the incidence of HPV-positive tumours remains at a rather low level.[28, 29]

In the US, around 39% of oral cavity or oropharyngeal cancers are located in the tongue. Other tumour sites include the floor of mouth (7%), the lips (8%), the gums and other parts of the mouth (17%), and the tonsils (29%).[30]

US figures for tongue, oral cavity, and pharyngeal cancer give a five-year overall survival of 62.7% over the period 2004 to 2010, with a much higher percentage for localized tumours than for regional or distant metastatic disease (Figure 1.4). Lip cancer comes out best with
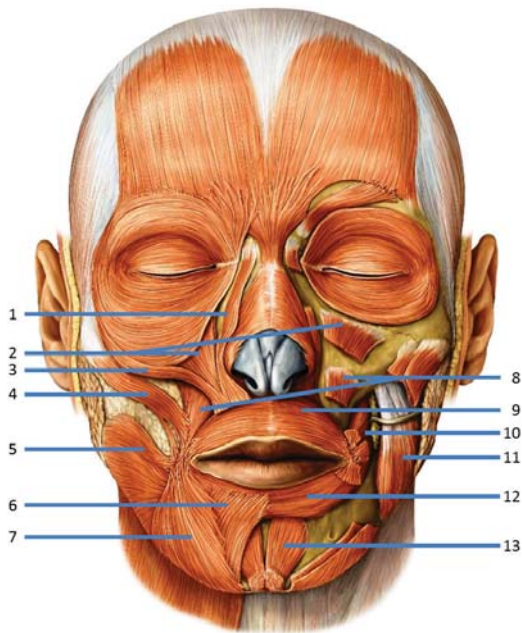
Figure 1.3 | Facial muscles influencing lip geometry: 1. Levator labii superioris alaeque nasi; 2. Levator labii superioris; 3. Zygomaticus minor; 4. Zygomaticus major; 5. Risorius; 6. Depressor labii inferioris; 7. Depressor anguli oris; 8. Levator anguli oris; 9. Orbicularis oris superior; 10. Buccinator; 11. Masseter; 12. Orbicularis oris inferior; 13. Mentalis. *Adapted image, original downloaded from [21].*

a five-year survival of over 89.5%.[30] Comparable figures were found in the Netherlands: 61% for oral cavity cancer and 91% for lip cancer over the period 2008 to 2012, whereas for oropharyngeal cancer a five-year survival rate of 47% was found.[26]

Well-known risk factors for developing oral and oropharyngeal cancer are alcohol and tobacco exposure.[31] Only 4% of patients are non-smokers and non-drinkers.[32, 33] Disease progression and survival figures in this group are similar to those in patients with drinking and/or smoking histories. A joint effect of smoking and drinking was found in several studies.[34–36] Additional risk factors are smokeless tobacco and betel quid chewing [37, 38], as well as working environment [39, 40], and in lip cancer ultraviolet light exposure.[22]

As mentioned above, another risk factor for oropharyngeal cancer is HPV [22, 24], [31, 41, 42], which mainly affects the younger age group. Tonsil cancer, in particular, is quite evidently correlated with HPV.[23, 43, 44] In Western countries, percentages between 50 and 70% are described for HPV-positive oropharyngeal cancers.[45–47] In the Netherlands, Rietbergen et al. and Henneman et al. found rates of 29% and 38%, respectively.[28, 29] Interestingly, HPV-positive cancers show better prognosis and treatment response.[48–52]
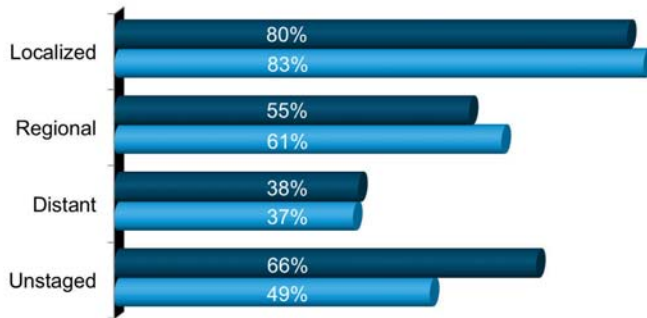
Figure 1.4 | Percentages of five-year relative survival for oral cavity and oropharynx cancer in the Netherlands (dark blue) and the USA (light blue).[26, 30]

1

As to genetics, first-degree relatives of head and neck cancer patients do appear to be at increased risk for developing the same types of cancer.[53] An underlying genetic disorder has been presumed [54], but not a single gene could be designated yet [55].

Finally, only 6% of head and neck cancers are seen in patients under 45 years of age.[56] However, both in the US and in the European Union a rising incidence in this young patient group has been observed.[57–63]

Table 1.1 | T and N classification of malignancies in the lip and oral cavity, and the oropharynx.[64]

	Lip and oral cavity	**Oropharynx**
T1	≤ 2 cm	≤ 2 cm
T2	> 2-4 cm	> 2-4 cm or more than 1 subsite
T3	> 4 cm	> 4 cm
T4a	Lip: through cortical bone, inferior alveolar nerve, floor of mouth, skin. Oral cavity: through cortical bone, deep/ extrinsic muscle of tongue, maxillary sinus, facial skin.	Through the larynx, deep/extrinsic muscle of tongue, medial pterygoid, hard palate, mandible.
T4b	Masticator space, pterygoid plates, skull base, internal carotid artery.	Through lateral pterygoid muscle, pterygoid plates, lateral nasopharynx, skull base, or encasement of carotid artery.
N1	Ipsilateral single ≤ 3 cm	Ipsilateral single ≤ 3 cm
N2	– Ipsilateral single > 3- 6 cm – Ipsilateral multiple ≤ 6 cm – Bilateral, contralateral ≤ 6 cm	– Ipsilateral single > 3- 6 cm – Ipsilateral multiple ≤ 6 cm – Bilateral, contralateral ≤ 6 cm
N3	> 6 cm	> 6 cm

13

Staging

Tumour staging is a strong prognostic tool in treatment planning. All cancers are staged according to the TNM system developed by the International Union Against Cancer (UICC), where the T-status represents the primary tumour status; N-status characterizes the status of the regional lymph nodes; and M-status describes the presence (M1) or absence (M0) of distant metastatic disease. The precise elaboration of this system for lip, oral, and oropharyngeal cancer is given in Table 1.1.[64]

On the basis of its TNM classification the disease can be staged on a scale from 0 (carcinoma in situ without any regional or distant spread) to IV C (distant metastatic disease). Stages I, II, and III represent a T1, T2, and T3 stage, respectively, without regional or distant metastasis as yet. Tumours are also designated as stage III, when T-status is not higher than T3, but the lymph node status is N1, and without distant metastasis. In stage IV A more positive lymph nodes are found ipsilateral or bilateral (N2), or the tumour has grown into its surrounding structures (T4a). In stage IV B the tumour shows further advancement into local structures (T4b), or lymph nodes are found with a diameter of 6 cm or more. In case of distant metastatic spread (M1), the disease is automatically staged IV C. The various stages are represented in Table 1.2.

Oral and oropharyngeal cancers often present in locally advanced stages of the disease and more than half are already stage IV at the time of diagnosis [65], which, of course, is an alarming fact given the predictive value of this staging system for survival [66].

Table 1.2 | Staging of oral cavity and oropharyngeal cancers, based on TNM classification.[64]

Stage	TNM classifications
I	T1 N0 M0
II	T2 N0 M0
III	T1-3 N1 M0 and T3 N0 M0
IV A	T1-3 N2 M0 and T4a N0-N2 M0
IV B	T4b N0-3 M0 and T1-4b N3 M0
IV C	T1-4b N0-3 M1

Treatment guidelines

Currently there are two curative treatment options for advanced oral cavity and oropharyngeal cancer; surgery (with or without adjuvant radiotherapy) and chemoradiotherapy. For oral cavity cancer, the primary choice of treatment is surgery, to be followed by adjuvant radiotherapy in case of a microscopically incomplete resection or the histological presence of bad prognostic indicators (perineural growth, (lymph-) angioinvasion, or sprouting growth pattern). Locally advanced oropharyngeal cancers are preferably treated with chemoradiation [67], mainly for functional reasons. Early-staged lesions, on the other hand,

can be treated with CO2-laser excisions, robotic surgery, photodynamic therapy (PDT), or radiotherapy, but only if function loss post treatment is expected to be limited.

Despite more advanced tumours are technically operable, thanks to increasing possibilities for reconstructive methods [67, 68], organ-sparing treatment with chemoradiation has been used more often as primary treatment, since the intial cure rates have become high.

Below are presented the Netherlands Cancer Institute guidelines for the most important areas of disease covered in this thesis. Guidelines for positive neck cases have not been included, because this thesis focusses on the functional consequences of treatment of the primary tumour in the oral cavity or oropharynx, even though a positive lymph node status will have definite impact on the treatment plan. These and other guidelines can be found on the website of the head and neck oncology and surgery multidisciplinary board at *www.hoofdhalskanker.info*.[69]

In lip cancer, T1 lesions are treated with surgery if <1 cm, and with radiotherapy if they are 1-2 cm in diameter, to minimize loss of function. T2 lesions are preferably treated with radiotherapy, and T3 and T4 lesions are surgically removed with reconstruction, an elective neck dissection, and adjuvant radiotherapy or chemoradiation. Inoperable tumours are treated with chemoradiation only.

In tongue cancer, the primary choice of treatment is surgery, regardless of T-status, with PDT as an alternative treatment option for superficial lesions with infiltration depths of <5 mm. Reconstruction is generally not indicated for T1 and T2 lesions. For T3 and T4 lesions, the tongue defect is generally reconstructed with a revascularized radial forearm flap with microvascular anastomoses. However, if speech and swallowing are expected to be unacceptably impaired by surgery, chemoradiation, or radiotherapy alone, is proposed. Anatomically inoperable tumours are always treated with chemoradiation.

As to cancers of the base of tongue or the tonsils, T1 and T2 lesions are generally treated surgically (robotic, CO2 laser, or PDT), with locoregional radiotherapy as an alternative treatment option. T3 lesions in these areas are preferably treated with radiotherapy or chemoradiation, depending on tumour volume; surgery is feasible, but may well lead to swallowing disorders and aspiration postoperatively. T4 lesions are generally treated with chemoradiation.

Function loss after treatment

Treatment of oral and oropharyngeal cancer can have serious impact on quality of life.[71, 72] A partial glossectomy, for instance, is bound to affect eating, drinking, swallowing, and speech, to a degree which will strongly correlate with T-status of the disease. Function loss in turn may lead to other problems like depression and deficient intake.[3, 22] Several studies have analysed consequences of surgery for speech [4, 73, 74], swallowing [75, 76], or both [77]. Langendijk et al. evaluated swallowing dysfunction after treatment with either
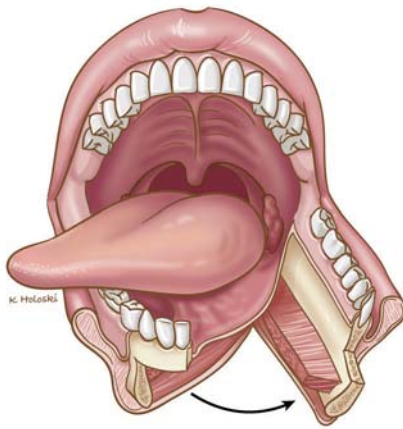
Figure 1.5 | Drawing of a commando resection. This procedure is performed when the mouth opening is too small for surgical exposure. By splitting of the mandibula a better approach is created.[70]

radiotherapy or chemoradiation [78]; Weber et al. studied multiple functional outcomes after surgery and/or chemoradiation [3]; and Dwivedi et al. reviewed literature on speech outcomes after any kind of treatment for oral cavity and oropharyngeal cancer [72]. All of these studies do confirm that factors like tumour size and location have predictive value for functional outcome, yet all they can do is show trends, rather than quantifying expected treatment effects in individual patients. So far, no randomised studies comparing oral functioning or quality of life, in patients, treated with either of the treatment options, have been described.[79]

Although speech, mastication, and swallowing seem straightforward activities for most people, the pertaining motions require complex coordinated contractions of multiple muscles in and around the oral cavity and the oropharynx. To perform the right motions in a controlled manner, proper timing, coordination, sensibility, and muscle strength are of vital importance. If any of these are compromised, patients may lose control and develop problems like dysphagia or pathologic speech, with potentially detrimental effects on their quality of life.[3, 77] There are options for surgical reconstruction [80, 81], but even then some degree of impaired function may remain [6].

In clinical treatment planning, functional outcome expectations are based on tumour location and planned resection volume.[82] Some locations are known to pose particular functional risks: the base of tongue is associated with dysphagia [6], and the genioglossus, the geniohyoid, and the mylohyoid muscles are linked to impaired speech and swallowing [83].

16

Oral function post-surgery was the focus of interest of Kreeft et al. in 2009 [82], who systematically reviewed relevant studies with populations of a least 20 patients with T2-T4 oral cavity and oropharyngeal cancers. Lip cancers were excluded, as was chemoradiation as primary treatment. The authors concluded that speech intelligibility had remained rather well after procedures in which the mobile tongue and the soft palate had not been part of the resection. The larger the resection volume had been, the greater impact was seen on speech intelligibility. Swallowing function, on the other hand, seemed affected by surgery more significantly, with a larger impact on quality of life. Aspiration was found mainly after oropharyngeal resections, with aspiration rates one year post surgery ranging from 12 to 50%. Besides resection volume, another contributing factor was found to be adjuvant radiotherapy.

Weber et al. in 2010 compared mouth opening effects of surgery with adjuvant radiotherapy on the one hand (n=82) and chemoradiation on the other (n=19), evaluating 101 patients with various types of head and neck cancer (oropharynx n=37; larynx n=29; hypopharynx n=16; other locations n=19) and stages (stage I n=2; stage II n=13; stage III n=25; stage IV n=59; stage 0 n=1; and unclassified n=1).[3] They found that 65% of oral and oropharyngeal cancer patients developed actual trismus (mouth opening <36 mm), whereas 73% reported any degree of mouth opening problems in a quality-of-life questionnaire. Other reported functional issues related to eating (65%), drinking (70%), xerostomia (92%), speech (68%), alteration of voice (62%), taste (60%), mastication (60%), swallowing (60%), choking on food (54%), and coughing while eating (52%).

Functional inoperability

Since the introduction of alternative curative treatment options, such as chemoradiation, surgical focus has shifted to preservation of function. Traditionally, tumours were considered inoperable if they invaded vital structures like the base of skull or the internal carotid artery; such tumours are now referred to as anatomically inoperable. Tumours which can be removed completely, but not without unacceptable loss of function are referred to as functional inoperable.[6]

As concluded by A.M. Kreeft in her PhD thesis, however, the assessment of functional operability is a highly subjective and variable process, based on personal experiences and preferences of individual members of the multidisciplinary tumour board.[6] Clear guidelines are lacking, whereas multiple patient-specific factors will influence treatment outcome. As described in previous sections, oral functions require complicated motions, initiated by many different muscles that must act in concordance. Complex anatomy makes it difficult to assess preoperatively exactly which individual muscles, nerve branches, or other structures will be part of the resection. Also, treatment-induced changes in anatomy can affect muscle synergy, and therefore muscle function. Apart from tumour location and volume, other

Figure 1.6 | Percentages of 67 Dutch head and neck oncologists that considered the functional result of the intervention unacceptable (red) and acceptable (green). Only statements are shown that are regarded as functional unacceptable by more than half of all respondents.[1]

patient-specific factors include age, profession, comorbidity, HPV status, and – last but not least – willingness to be treated. The choice of treatment also depends on the treating physicians' expertise in the fields of surgery, radiotherapy, and chemotherapy, as well as on availability of rehabilitation programs, dentistry, prosthetic support, and psychosocial support.

A survey amongst head and neck specialists in the Netherlands showed consensus on functional operability of a number of surgical procedures.[1] For instance, total glossectomy, with or without resection of the supraglottic larynx, was considered functionally inoperable by 93% of the respondents. Yet procedures like tonsil and base of tongue resections including removal of the vallecula and epiglottis proved controversial, with only 52% of respondents considering these procedures functionally inoperable. Figure 1.6 lists several procedures which more than half of the respondents considered functionally inoperable.

Case descriptions supported by magnetic resonance (MRI) images elicited disagreement mainly on T3 and T4 tongue-based cancers. A high level of consensus was found regarding the preservation of one hypoglossal nerve and lingual artery, which were assumed to be of vital importance for securing mobility of the remaining part of the tongue.[1]

Figure 1.7 chart showing percentages of respondents considering functional outcome as unacceptable (red) and acceptable (green):

Procedure	Unacceptable	Acceptable
Total soft palate resection	37	63
Total mandibulectomy	50	50
Total glossectomy	53	47
Resection of tonsil, base of tongue carcinoma with invasion in vallecula	53	47
Resection of tonsil, base of tongue carcinoma, dubious preservation of hypoglossal nerve and lingual artery	56	44
Total glossectomy + total laryngectomy	63	37
Resection of floor of mouth, segmental mandibula, subtotal glossectomy, sacrificing both hypoglossal nerves	64	36
Total glossectomy and supraglottic laryngectomy	85	15

Figure 1.7 | Percentages of respondents of the worldwide survey considering the functional outcome of the given surgical procedure as unacceptable (red) and acceptable (green). Composite of figures from Kreeft et al.[84]

1

A comparable survey was also conducted with 179 respondents worldwide.[84] Figure 1.7 shows the results for procedures comparable to the ones listed in Figure 1.6. In this study it was described that Dutch head and neck oncologists are more likely to consider a procedure functionally inoperable, with the clearest distinction with their counterparts from Northern America, who tend to opt for a surgical approach sooner than the European respondents would.

These surveys show disparity in opinions regarding functional inoperability. For some procedures, such as the surgical removal of both hypoglossal nerves, there is consensus on when to refrain from surgery, but many procedures prove controversial when it comes to functional operability. To solve this controversy and to allow for more evidence-based and objective treatment planning, it was proposed to develop a dynamic model of the oral cavity and oropharynx that could present in a virtual environment the expected functional outcome of the available treatment options in individual patients.[6]

Virtual Therapy project

As mentioned above, functional outcome predictions in clinical practice are based on personal experience of the members of the multidisciplinary tumour board, rather than on evidence provided by reliable tools.[6] Consequently, the optimal treatment plan for any
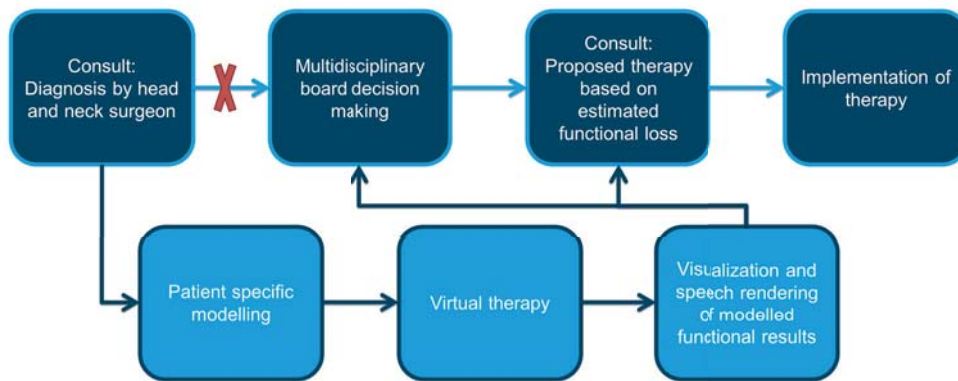
19

Figure 1.8 | Adapted clinical workflow to incorporate predicted functional loss in treatment planning and patient counselling. Upper row: current workflow. Lower row: additional steps in proposed workflow.

individual patient remains something of an educated guess, rather than being suggested by standardized, objective, and accurate data. This makes it difficult to reach consensus in the multidisciplinary meetings and to counsel the patient properly on the expected outcome. In some cases, patients and/or their physicians have even come to regret their choice of treatment, because it turned out they were insufficiently prepared for the functional result. The Virtual Therapy project aims to improve this situation by providing a new decision model incorporating a virtual tool that will be able to generate objective personalized outcome predictions, thereby leading to more constructive treatment planning an better patient counselling (Figure 1.8).

In the current workflow, a patient visits a head and neck specialist, who analyses clinical features to come up with a provisional diagnosis, and possibly differential diagnoses. The case is then presented to the multidisciplinary tumour board, where the best treatment strategy is decided upon.

The problem with this workflow, is the lack of patient-specific data. For one thing, the exact individual anatomy, including precise muscle locations and nerve branching patterns, is unknown, and so is dynamic information on individual motion performance: two people may perform the same motion by contracting different tongue muscles, and may therefore be affected by treatment in different ways. Another factor influencing tongue motion is aging, because of the structural and physiological changes it brings [85], as was also demonstrated in 2009 by Sanders et al., who found differences between newborns and adults in numbers of slow muscle fibres and distribution patterns of tongue muscles.[86] Furthermore, it can be hypothesized that the language or even the dialect one speaks may affect tongue motions, because of the specific muscle contractions involved in pronunciation. Finally, there is also inter-person variability of muscular innervation for lip motions, as has been demonstrated by

20

means of electromyography (EMG).[16] Such variability is likely to exist for tongue motions as well. All of these factors call for a tool to produce accurate, personalized functional outcome predictions.

To answer this call, the Virtual Therapy project proposes an adapted workflow incorporating the use of a biomechanical model that can predict movements of the tongue, lips, and other structures of the oral cavity and oropharynx post treatment, presenting in a virtual environment the individual treatment effects on mastication, swallowing, and speech. The resulting visualizations and audio data would be very helpful in deciding upon the best treatment plan and preparing the patient for the expected treatment result. By virtually performing post-treatment swallowing motions, for instance, the model could visualize any problems of aspiration that may arise. Patients could then be advised beforehand what types of food (liquids, solids) they will be able to eat after their treatment. Also, by means of articulatory speech synthesis patients could hear how their speech will change due to the harm done to their vocal tracts.

Aim and outline of thesis

The aim of this thesis is to formulate the technical requirements, consisting of modelling methods, and the incorporation of anatomical as well as clinical data, for building such a model. Investigations focus on the tongue and lips, as these structures predominantly determine oral function.

Crucial in any virtual-therapy model is patient-specificity. Some biomechanical models of the tongue [87, 88] and the lips [89] have already been developed, but these do not account for individual anatomic peculiarities, such as exact muscle fibre locations and nerve branching patterns. Accurate, detailed predictions of post-treatment function must be derived from patient-specific data as to anatomy, mechanical and dynamical properties of tissues, and neural input for tongue muscle activation in specific tasks.

Chapter 2 describes the development of a dynamic three dimensional (3D) biomechanical model of the tongue. This model should be able to mimic tongue anatomy and simulate how treatment-induced changes in tissue characteristics will affect motion. Chapter 3 focusses on lip shape, evaluating if the use of surface EMG (sEMG) of the facial muscles in building a 3D lip shape model is feasible. In chapter 4 this is extrapolated to dynamics, investigating sEMG for modelling lip movements. Individual anatomic variation will be of key importance in a patient-specific virtual model. Therefore a new set-up is introduced in chapter 5 to measure tongue movements induced by stimulation of individual nerve branches. This set-up will provide insight into anatomic variability, which will help personalize and evaluate the biomechanical models. Chapter 6 proposes another measurement set-up, which will measure range of tongue motion in healthy subjects and patients. This will help develop an individualized grading system to identify risk groups on the basis of preoperative tongue mobility and tumour characteristics. Chapter 7, finally, presents short summaries per chapter and a general conclusion, as well as a discussion of future perspectives.

21

REFERENCES

[1] A. M. Kreeft, I. B. Tan, M. W. M. van den Brekel, F. J. Hilgers, and A. J. M. Balm, "The surgical dilemma of 'functional inoperability' in oral and oropharyngeal cancer: current consensus on operability with regard to functional results," *Clin. Otolaryngol.*, vol. 34, no. 2, pp. 140–146, 2009.

[2] S. M. Gore, A. K. Crombie, M. D. Batstone, and J. R. Clark, "Concurrent chemoradiotherapy compared with surgery and adjuvant radiotherapy for oral cavity squamous cell carcinoma," *Head Neck*, vol. 37, no. 4, pp. 518–523, 2015.

[3] C. Weber, S. Dommerich, H. W. Pau, and B. Kramp, "Limited mouth opening after primary therapy of head and neck cancer," *Oral Maxillofac. Surg.*, vol. 14, pp. 169–173, 2010.

[4] C. Chuanjun, Z. Zhiyuan, G. Shaopu, J. Xinquan, and Z. Zhihong, "Speech after partial glossectomy: A comparison between reconstruction and nonreconstruction patients," *J. Oral Maxillofac. Surg.*, vol. 60, pp. 404–407, 2002.

[5] A. Christopoulos, S. P. Moubayed, M. E. Nader, J. E. Ghannoum, T. Ayad, and A. D. Meyers, "Mouth Anatomy," 2013. [Online]. Available: http://emedicine.medscape.com/article/1899122-overview#a1. [Accessed: 11-Mar-2015].

[6] A. M. Kreeft, "Functional inoperability of oral and oropharyngeal cancer," University of Amsterdam, 2013.

[7] "Animal nutrition." [Online]. Available: http://biobook.nerinxhs.org/bb/systems/animal_nutrition.htm. [Accessed: 26-May-2015].

[8] W. M. Kier and K. K. Smith, "Tongues, tentacles and trunks: the biomechanics of movement in muscular-hydrostats," *Zool. J. Linn. Soc.*, vol. 83, no. 4, pp. 307–324, Apr. 1985.

[9] K. L. Moore and A. F. Dalley, "Head," in *Clinically oriented anatomy*, 5th ed., Philadelphia: Lippincott Williams & Wilkins, 2005, pp. 820–1052.

[10] A. M. R. Agur and A. F. Dalley, "Head," in *Grant's Atlas of Anatomy*, 11th ed., B. Sun, Ed. Philadelphia: Lippincott Williams & Wilkins, 2005, pp. 587–719.

[11] R. Drake, A. W. Vogl, and A. W. M. Mitchell, "Head and Neck," in *Gray's Anatomy for Students*, Second., W. Schmitt and R. Gruliow, Eds. Philadelphia: Churchill Livingstone, 2009, pp. 1087–1114.

[12] L. T. Gordon, "Oral cavity." [Online]. Available: http://www.unmc.edu/dissection/review18.cfm?d=-1&topic_id=18. [Accessed: 26-May-2015].

[13] E. Armstrong, "Journey of the Voice: Anatomy, Physiology and the Care of the Voice," 1998. [Online]. Available: http://www.yorku.ca/earmstro/journey/tongue.html. [Accessed: 26-May-2015].

[14] B. G. Lapatki, R. Oostenveld, J. P. Van Dijk, I. E. Jonas, M. J. Zwarts, and D. F. Stegeman, "Optimal placement of bipolar surface EMG electrodes in the face based on single motor unit analysis," *Psychophysiology*, vol. 47, pp. 299–314, 2010.

1

[15] B. G. Lapatki, D. F. Stegeman, and I. E. Jonas, "A surface EMG electrode for the simultaneous observation of multiple facial muscles," *J. Neurosci. Methods*, vol. 123, pp. 117–128, 2003.

[16] C. L. Isley and J. V Basmajian, "Electromyography of the human cheeks and lips.," *Anat. Rec.*, vol. 176, no. 2, pp. 143–147, 1973.

[17] H. Knipe and J. Jones, "Digastric muscle," 2015. [Online]. Available: http://radiopaedia. org/articles/digastric-muscle. [Accessed: 09-Feb-2015].

[18] T. Peker, H. Turgut, and A. Anil, "Bilateral anomaly of anterior bellies of digastric muscles," *Surg. Radiol. Anat.*, vol. 22, no. 2, pp. 119–121, 2000.

[19] J. C. Andreo, J. A. Caldas Navarro, and J. L. Toledo Filho, "Anatomical study on the variations of the anterior belly of the digastric muscle," *Rev. chil. anat*, vol. 15, no. 2, pp. 111–114, 1997.

[20] B. M. Liquidato, M. D. Barros, A. L. Alves, and C. S. B. Pereira, "Anatomical study of the digastric muscle: variations in the anterior belly," *Int J Morphol*, vol. 25, no. 4, pp. 797–800, 2007.

[21] M. Schünke, E. Schulte, and U. Schumacher, "Head," in *Thieme atlas of anatomy: head and neuroanatomy*, 1st ed., L. M. Ross, E. D. Lamperti, and E. Taub, Eds. New York: Thieme, 2010, p. 44.

[22] S. Warnakulasuriya, "Global epidemiology of oral and oropharyngeal cancer," *Oral Oncol.*, vol. 45, no. 4–5, pp. 309–316, 2009.

[23] A. Argiris, M. V Karamouzis, D. Raben, and R. L. Ferris, "Head and neck cancer.," *Lancet*, vol. 371, pp. 1695–1709, 2008.

[24] J. Ferlay, I. Soerjomataram, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D. M. Parkin, D. Forman, F. Bray, I. Soerjomataram I, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D. M. Parkin, D. Forman D, and F. Bray, "Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012," *Int. J. Cancer*, vol. 136, no. 5, pp. 359–386, 2015.

[25] A. K. Chaturvedi, W. F. Anderson, J. Lortet-Tieulent, M. Paula Curado, J. Ferlay, S. Franceschi, P. S. Rosenberg, F. Bray, and M. L. Gillison, "Worldwide trends in incidence rates for oral cavity and oropharyngeal cancers," *J. Clin. Oncol.*, vol. 31, pp. 4550–4559, 2013.

[26] IKNL, "Nederlandse Kankerregistratie." [Online]. Available: http://www.cijfersover kanker.nl/. [Accessed: 12-Mar-2015].

[27] K. K. Ang and E. M. Sturgis, "Human Papillomavirus as a Marker of the Natural History and Response to Therapy of Head and Neck Squamous Cell Carcinoma," *Seminars in Radiation Oncology*, vol. 22. pp. 128–142, 2012.

[28] M. M. Rietbergen, C. R. Leemans, E. Bloemena, D. A. M. Heideman, B. J. M. Braakhuis, A. T. Hesselink, B. I. Witte, R. J. B. De Jong, C. J. L. M. Meijer, P. J. F. Snijders, and R. H. Brakenhoff, "Increasing prevalence rates of HPV attributable oropharyngeal squamous cell carcinomas in the Netherlands as assessed by a validated test algorithm," *Int. J. Cancer*, vol. 132, pp. 1565–1571, 2013.

[29] R. Henneman, H. S. van Monsjou, C. V. M. Verhagen, M. L. F. van Velthuysen, N. T. ter Haar, E. M. Osse, M. I. Lopez-Yurda, A. J. M. Balm, and M. W. M. van den Brekel, "Incidence changes of human papillomavirus in oropharyngeal squamous cell carcinoma and effects on survival in the Netherlands Cancer Institute, 1980-2009 [Epub ahead of print]," *Anticancer Res.*, vol. 35, 2015.

[30] National Cancer Institute, "Seer Cancer Statistics Review 1975-2011." [Online]. Available: http://seer.cancer.gov/csr/1975_2011/browse_csr.php. [Accessed: 12-Mar-2015].

[31] H. S. Van Monsjou, A. J. M. Balm, M. W. M. Van Den Brekel, and V. B. Wreesmann, "Oropharyngeal squamous cell carcinoma: A unique disease on the rise?," *Oral Oncology*, vol. 46. pp. 780–785, 2010.

[32] F. Farshadpour, G. J. Hordijk, R. Koole, and P. J. Slootweg, "Non-smoking and non-drinking patients with head and neck squamous cell carcinoma: A distinct population," *Oral Dis.*, vol. 13, pp. 239–243, 2007.

[33] S. M. Wiseman, H. Swede, D. L. Stoler, G. R. Anderson, N. R. Rigual, W. L. Hicks, W. G. Douglas, D. Tan, and T. R. Loree, "Squamous cell carcinoma of the head and neck in nonsmokers and nondrinkers: an analysis of clinicopathologic characteristics and treatment outcomes.," *Ann. Surg. Oncol.*, vol. 10, pp. 551–557, 2003.

[34] M. Hashibe, P. Brennan, S. C. Chuang, S. Boccia, X. Castellsague, C. Chen, M. P. Curado, L. Dal Maso, A. W. Daudt, E. Fabianova, L. Fernandez, V. Wünsch-Filho, S. Franceschi, R. B. Hayes, R. Herrero, K. Kelsey, S. Koifman, C. La Vecchia, P. Lazarus, F. Levi, J. J. Lence, D. Mates, E. Matos, A. Menezes, M. D. McClean, J. Muscat, J. Eluf-Neto, A. F. Olshan, M. Purdue, P. Rudnai, S. M. Schwartz, E. Smith, E. M. Sturgis, N. Szeszenia-Dabrowska, R. Talamini, Q. Wei, D. M. Winn, O. Shangina, A. Pilarska, Z. F. Zhang, G. Ferro, J. Berthiller, and P. Boffetta, "Interaction between tobacco and alcohol use and the risk of head and neck cancer: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium.," *Cancer Epidemiol. Biomarkers Prev.*, vol. 18, pp. 541–550, 2009.

[35] F. Lewin, S. E. Norell, H. Johansson, P. Gustavsson, J. Wennerberg, A. Biörklund, and L. E. Rutqvist, "Smoking tobacco, oral snuff, and alcohol in the etiology of squamous cell carcinoma of the head and neck: A population-based case-referent study in Sweden," *Cancer*, vol. 82, pp. 1367–1375, 1998.

[36] J. L. Ferreira Antunes, T. N. Toporcov, M. G. H. Biazevic, A. F. Boing, C. Scully, and S. Petti, "Joint and Independent Effects of Alcohol Drinking and Tobacco Smoking on Oral Cancer: A Large Case-Control Study," *PLoS One*, vol. 8, no. 7, pp. 1–7, 2013.

[37] N. K. Proia, G. M. Paszkiewicz, M. A. S. Nasca, G. E. Franke, and J. L. Pauly, "Smoking and smokeless tobacco-associated human buccal cell mutations and their association with oral cancer - A review," *Cancer Epidemiology Biomarkers and Prevention*, vol. 15. pp. 1061–1077, 2006.

[38] S. Warnakulasuriya, "Areca nut use following migration and its consequences.," *Addict. Biol.*, vol. 7, pp. 127–132, 2002.

[39] M. L. Gillison, "Current topics in the epidemiology of oral cavity and oropharyngeal cancers," *Head and Neck*, vol. 29. pp. 779–792, 2007.

1

[40] O. Shangina, P. Brennan, N. Szeszenia-Dabrowska, D. Mates, E. Fabiánová, T. Fletcher, A. T'Mannetje, P. Boffetta, and D. Zaridze, "Occupational exposure and laryngeal and hypopharyngeal cancer risk in Central and Eastern Europe," *Am. J. Epidemiol.*, vol. 164, pp. 367–375, 2006.

[41] E. M. Sturgis and P. M. Cinciripini, "Trends in head and neck cancer incidence in relation to smoking prevalence: An emerging epidemic of human papillomavirus-associated cancers?," *Cancer*, vol. 110. pp. 1429–1435, 2007.

[42] G. D'Souza, A. R. Kreimer, R. Viscidi, M. Pawlita, C. Fakhry, W. M. Koch, W. H. Westra, and M. L. Gillison, "Case-control study of human papillomavirus and oropharyngeal cancer.," *N. Engl. J. Med.*, vol. 356, pp. 1944–1956, 2007.

[43] C. G. L. Hobbs, J. A. C. Sterne, M. Bailey, R. S. Heyderman, M. A. Birchall, and S. J. Thomas, "Human papillomavirus and head and neck cancer: a systematic review and meta-analysis.," *Clin. Otolaryngol.*, vol. 31, pp. 259–266, 2006.

[44] J. P. Klussmann, S. J. Weissenborn, U. Wieland, V. Dries, H. E. Eckel, H. J. Pfister, and P. G. Fuchs, "Human papillomavirus-positive tonsillar carcinomas: A different tumor entity?," *Med. Microbiol. Immunol.*, vol. 192, pp. 129–132, 2003.

[45] A. K. Chaturvedi, E. A. Engels, W. F. Anderson, and M. L. Gillison, "Incidence trends for human papillomavirus-related and -unrelated oral squamous cell carcinomas in the United States," *J. Clin. Oncol.*, vol. 26, pp. 612–619, 2008.

[46] L. Hammarstedt, D. Lindquist, H. Dahlstrand, M. Romanitan, L. Onelöv, J. Joneberg, N. Creson, J. Lindholm, W. Ye, T. Dalianis, and E. Munck-Wikland, "Human papillomavirus as a risk factor for the increase in incidence of tonsillar cancer," *Int. J. Cancer*, vol. 119, pp. 2620–2623, 2006.

[47] A. Näsman, C. Nordfors, S. Holzhauser, A. Vlastos, N. Tertipis, U. Hammar, L. Hammarstedt-Nordenvall, L. Marklund, E. Munck-Wikland, T. Ramqvist, M. Bottai, and T. Dalianis, "Incidence of human papillomavirus positive tonsillar and base of tongue carcinoma: A stabilisation of an epidemic of viral induced carcinoma?," *Eur. J. Cancer*, vol. 51, no. 1, pp. 55–61, 2015.

[48] L. Licitra, F. Perrone, P. Bossi, S. Suardi, L. Mariani, R. Artusi, M. Oggionni, C. Rossini, G. Cantù, M. Squadrelli, P. Quattrone, L. D. Locati, C. Bergamini, P. Olmi, M. A. Pierotti, and S. Pilotti, "High-risk human papillomavirus affects prognosis in patients with surgically treated oropharyngeal squamous cell carcinoma," *J. Clin. Oncol.*, vol. 24, pp. 5630–5636, 2006.

[49] K. K. Ang, J. Harris, R. Wheeler, R. Weber, D. I. Rosenthal, P. F. Nguyen-Tân, W. H. Westra, C. H. Chung, R. C. Jordan, C. Lu, H. Kim, R. Axelrod, C. C. Silverman, K. P. Redmond, and M. L. Gillison, "Human papillomavirus and survival of patients with oropharyngeal cancer.," *N. Engl. J. Med.*, vol. 363, no. 1, pp. 24–35, 2010.

[50] D. Lindquist, M. Romanitan, L. Hammarstedt, A. Näsman, H. Dahlstrand, J. Lindholm, L. Onelöv, T. Ramqvist, W. Ye, E. Munck-Wikland, and T. Dalianis, "Human papillomavirus is a favourable prognostic factor in tonsillar cancer and its oncogenic role is supported by the expression of E6 and E7," *Mol. Oncol.*, vol. 1, pp. 350–355, 2007.

[51] P. Lassen, J. G. Eriksen, S. Hamilton-Dutoit, T. Tramm, J. Alsner, and J. Overgaard, "Effect of HPV-associated p16INK4A expression on response to radiotherapy and survival in squamous cell carcinoma of the head and neck," *J. Clin. Oncol.*, vol. 27, pp. 1992–1998, 2009.

[52] C. Fakhry, W. H. Westra, S. Li, A. Cmelak, J. A. Ridge, H. Pinto, A. Forastiere, and M. L. Gillison, "Improved survival of patients with human papillomavirus-positive head and neck squamous cell carcinoma in a prospective clinical trial," *J. Natl. Cancer Inst.*, vol. 100, no. 4, pp. 261–269, 2008.

[53] E. Negri, P. Boffetta, J. Berthiller, X. Castellsague, M. P. Curado, L. D. Maso, A. W. Daudt, E. Fabianova, L. Fernandez, V. Wünsch-Filho, S. Franceschi, R. B. Hayes, R. Herrero, S. Koifman, P. Lazarus, J. J. Lence, F. Levi, D. Mates, E. Matos, A. Menezes, J. Muscat, J. Eluf-Neto, A. F. Olshan, P. Rudnai, O. Shangina, E. M. Sturgis, N. Szeszenia-Dabrowska, R. Talamini, Q. Wei, D. M. Winn, D. Zaridze, J. Lissowska, Z. F. Zhang, G. Ferro, P. Brennan, C. La Vecchia, and M. Hashibe, "Family history of cancer: Pooled analysis in the International Head and Neck Cancer Epidemiology consortium," *Int. J. Cancer*, vol. 124, pp. 394–401, 2009.

[54] V. Bongers, B. J. M. Braakhuis, H. Tobi, H. Lubsen, and G. B. Snow, "The relation between cancer incidence among relatives and the occurrence of multiple primary carcinomas following head and neck cancer," *Cancer Epidemiol. Biomarkers Prev.*, vol. 5, pp. 595–598, 1996.

[55] M. Lacko, B. J. M. Braakhuis, E. M. Sturgis, C. C. Boedeker, C. Suárez, A. Rinaldo, A. Ferlito, and R. P. Takes, "Genetic susceptibility to head and neck squamous cell carcinoma," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 89, no. 1, pp. 38–48, 2014.

[56] C. D. Llewellyn, N. W. Johnson, and K. A. Warnakulasuriya, "Risk factors for squamous cell carcinoma of the oral cavity in young people--a comprehensive literature review.," *Oral Oncol.*, vol. 37, no. 5, pp. 401–418, 2001.

[57] H. Møller, "Changing incidence of cancer of the tongue, oral cavity, and pharynx in Denmark.," *J. Oral Pathol. Med.*, vol. 18, no. 4, pp. 224–229, 1989.

[58] G. J. Macfarlane, P. Boyle, T. V. Evstifeeva, C. Robertson, and C. Scully, "Rising trends of oral cancer mortality among males worldwide: The return of an old public health problem," *Cancer Causes Control*, vol. 5, no. 3, pp. 259–265, 1994.

[59] C. H. Shiboski, S. C. Shiboski, and S. Silverman, "Trends in oral cancer rates in the United States, 1973-1996.," *Community Dent. Oral Epidemiol.*, vol. 28, no. 4, pp. 249–256, 2000.

[60] R. H. Depue, "Rising mortality from cancer of the tongue in young white males.," *N. Engl. J. Med.*, vol. 315, no. 10, p. 647, 1986.

[61] S. Davis and R. K. Severson, "Increasing incidence of cancer of the tongue in the United States among young adults.," *Lancet*, vol. 2, no. 8564, pp. 910–911, 1987.

[62] J. N. Myers, T. Elkins, D. Roberts, and R. M. Byers, "Squamous cell carcinoma of the tongue in young adults: increasing incidence and factors that predict treatment outcomes.," *Otolaryngol. Head. Neck Surg.*, vol. 122, pp. 44–51, 2000.

[63] L. J. Shemen, J. Klotz, D. Schottenfeld, and E. W. Strong, "Increase of tongue cancer in young men.," *JAMA*, vol. 252, no. 14, p. 1857, 1984.

[64] L. H. Sobin, M. Gospodarowicz, and C. Wittekind, "Head and Neck," in *TNM Classification of Malignant Tumours*, Seventh ed., Wiley-Blackwell, 2009, p. 336.

[65] Dutch Head and Neck Oncology Cooperative Group of the Association of Comprehensive Cancer Centres, "Oral Cavity and Oropharyngeal Cancer," 2004. [Online]. Available: http://www.oncoline.nl/mondholte-en-orofarynxcarcinoom. [Accessed: 12-Mar-2015].

[66] E. H. Dibble, A. C. L. Alvarez, M. T. Truong, G. Mercier, E. F. Cook, and R. M. Subramaniam, "18F-FDG metabolic tumor volume and total glycolytic activity of oral cavity and oropharyngeal squamous cell cancer: adding value to clinical staging.," *J. Nucl. Med.*, vol. 53, no. 5, pp. 709–15, 2012.

[67] J. P. Shah and Z. Gil, "Current concepts in management of oral cancer-surgery," *Oral Oncol.*, vol. 45, pp. 394–401, 2009.

[68] P. G. Cordeiro and J. J. Disa, "Challenges in midface reconstruction," *Semin. Surg. Oncol.*, vol. 19, no. 3, pp. 218–225, 2000.

[69] Head and Neck Center AVL, "Oral cavity and Oropharynx." [Online]. Available: http://www.hoofdhalskanker.info/wpavl/behandelrichtlijnen/mondholte-en-oropharynx/. [Accessed: 12-Feb-2015].

[70] Head and Neck cancer guide, "Glossectomy," 2015. [Online]. Available: http://www.headandneckcancerguide.org/adults/cancer-diagnosis-treatments/surgery-and-rehabilitation/cancer-removal-surgeries/glossectomy/. [Accessed: 19-Jun-2015].

[71] G. F. Funk, L. H. Karnell, and A. J. Christensen, "Long-term health-related quality of life in survivors of head and neck cancer.," *Arch. Otolaryngol. Head. Neck Surg.*, vol. 138, no. 2, pp. 123–33, 2012.

[72] R. C. Dwivedi, R. A. Kazi, N. Agrawal, C. M. Nutting, P. M. Clarke, C. J. Kerawala, P. H. Rhys-Evans, and K. J. Harrington, "Evaluation of speech outcomes following treatment of oral and oropharyngeal cancers," *Cancer Treat. Rev.*, vol. 35, no. 5, pp. 417–424, 2009.

[73] J. Sun, Y. Weng, J. Li, G. Wang, and Z. Zhang, "Analysis of Determinants on Speech Function After Glossectomy," *J. Oral Maxillofac. Surg.*, vol. 65, pp. 1944–1950, 2007.

[74] T. Bressmann, R. Sader, T. L. Whitehill, and N. Samman, "Consonant Intelligibility and Tongue Motility in Patients with Partial Glossectomy," *J. Oral Maxillofac. Surg.*, vol. 62, pp. 298–303, 2004.

[75] A. G. Schache, O. Lieger, P. Rogers, A. Kelly, L. Newman, and N. Kalavrezos, "Predictors of swallowing outcome in patients treated with surgery and radiotherapy for advanced oral and oropharyngeal cancer," *Oral Oncol.*, vol. 45, no. 9, pp. 803–808, 2009.

[76] N. Kalavrezos and R. Bhandari, "Current trends and future perspectives in the surgical management of oral cancer," *Oral Oncol.*, vol. 46, no. 6, pp. 429–432, 2010.

[77] A. C. Zuydam, D. Lowe, J. S. Brown, E. D. Vaughan, and S. N. Rogers, "Predictors of speech and swallowing function following primary surgery for oral and oropharyngeal cancer.," *Clin. Otolaryngol.*, vol. 30, no. 5, pp. 428–437, 2005.

[78] J. A. Langendijk, P. Doornaert, D. H. F. Rietveld, I. M. Verdonck-de Leeuw, C. René Leemans, and B. J. Slotman, "A predictive model for swallowing dysfunction after curative radiotherapy in head and neck cancer," *Radiother. Oncol.*, vol. 90, no. 2, pp. 189–195, 2009.

[79] A. H. Ackerstaff, I. B. Tan, C. R. N. Rasch, A. J. M. Balm, R. B. Keus, J. H. Schornagel, and F. J. M. Hilgers, "Quality-of-life assessment after supradose selective intra-arterial cisplatin and concomitant radiation (RADPLAT) for inoperable stage IV head and neck squamous cell carcinoma.," *Arch. Otolaryngol. Head. Neck Surg.*, vol. 128, no. 10, pp. 1185–1190, Oct. 2002.

[80] P. C. Neligan, P. J. Gullane, and R. W. Gilbert, "Functional reconstruction of the oral cavity," *World J. Surg.*, vol. 27, no. 7, pp. 856–862, 2003.

[81] W. C. Tsai, J. M. Yang, S. C. Liu, Y. H. Chu, W. S. Lai, Y. S. Lin, and J. C. Lee, "Management of different kinds of head and neck defects with the submental flap for reconstruction [Epub ahead of print]," *Eur. Arch. Oto-Rhino-Laryngology*, 2014.

[82] A. M. Kreeft, L. van der Molen, F. J. Hilgers, and A. J. Balm, "Speech and swallowing after surgical treatment of advanced oral and oropharyngeal carcinoma: a systematic review of the literature.," *Eur. Arch. Otorhinolaryngol.*, vol. 266, no. 11, pp. 1687–1698, 2009.

[83] M. M. Suarez-Cunqueiro, A. Schramm, R. Schoen, J. Seoane-Lestón, X. L. Otero-Cepeda, K. H. Bormann, H. Kokemueller, M. Metzger, P. Diz-Dios, and N. C. Gellrich, "Speech and swallowing impairment after treatment for oral and oropharyngeal cancer.," *Arch. Otolaryngol. Head. Neck Surg.*, vol. 134, no. 12, pp. 1299–1304, 2008.

[84] A. M. Kreeft, I. B. Tan, C. R. Leemans, and A. J. M. Balm, "The surgical dilemma in advanced oral and oropharyngeal cancer: how we do it.," *Clin. Otolaryngol.*, vol. 36, no. 3, pp. 260–266, Jun. 2011.

[85] J. W. Bennett, P. H. H. M. van Lieshout, and C. M. Steele, "Tongue control for speech and swallowing in healthy younger and older subjects.," *Int. J. Orofacial Myology*, vol. 33, pp. 5–18, 2007.

[86] I. Sanders, L. Mu, A. Amirali, H. Su, and S. Sobotka, "The human tongue slows down to speak: Muscle fibers of the human tongue," *Anat. Rec.*, vol. 296, no. 10, pp. 1615–1627, 2013.

[87] J. Dang, S. Fujita, E. Murano, and M. Stone, "Observation and Simulation of Large-scale Deformation of Tongue," *J. Acoust. Soc. Am.*, vol. 155, no. 2, pp. 853–870, 2004.

[88] I. Stavness, J. E. Lloyd, and S. Fels, "Automatic prediction of tongue muscle activations using a finite element model," *J. Biomech.*, vol. 45, no. 16, pp. 2841–2848, 2012.

[89] J. C. Lucero and K. G. Munhall, "A model of facial biomechanics for speech production," *J. Acoust. Soc. Am.*, vol. 106, no. 5, pp. 2834–2842, 1999.

Chapter 2

Towards virtual surgery in oral cancer to predict postoperative oral functions preoperatively

M.J.A. van Alphen

A.M. Kreeft

F. van der Heijden

L.E. Smeele

A.J.M. Balm

ABSTRACT

Our aim was to develop a dynamic virtual model of the oral cavity and oropharynx so that we could incorporate patient-specific factors into the prediction of functional loss after advanced resections for oral cancer. After a virtual resection, functional consequences can be assessed, and a more substantiated decision about treatment can be made. In this study we used a finite element model of the tongue, which can be implemented in the total virtual environment in the future. We analysed the movements and changes in volume, and the effects of changes in the material variables, to mimic scar tissue. The observed movements were in accordance with descriptions of in vivo movements. Affected movements caused by the mimicked scar tissue were also similar to expectations. Some changes in volume were measured, particularly in individual elements. We have taken the first steps in the development of a finite element model of the tongue. Now, refinement is necessary to make the model suitable for future use in virtual surgery.

INTRODUCTION

Resection, often combined with adjuvant radiotherapy, is the mainstay of the treatment of oral cancer. However, operations on the mouth might seriously interfere with speech, swallowing, and mastication, particularly in advanced disease, an alternative treatment to which is organ sparing. This consists of concurrent chemoradiotherapy, which has become more common during recent decades.[1] Chemoradiotherapy may lead to dysphagia, fibrosis and xerostomia, depending on the dose of radiation and site of the tumour.[2] Both regimens have comparable survival,[3] which gives the assessment of the functional consequences of each an important role in clinical decision-making.[4]

Nowadays the choice between resection and chemoradiotherapy depends mainly on the site and extension of the tumour, as the functional outcome is dependent on the amount of tissue that should be removed to achieve clear margins.[5] However, patient-related factors also have an important role in the functional consequences, and these are hard to take into account, which makes the predictions about functional outcome fallible. The choice between resection and organ-sparing chemoradiotherapy remains difficult and requires careful weighing of all patient-related and tumour-related factors. The term "functional inoperability" can be used if resection of the tumour would cause unacceptable loss of function. It is currently used in clinical decision-making, and indicates the irreversible losses of swallowing and speech postoperatively.[6]

In a web-based survey among head and neck surgeons worldwide, we clearly showed that opinions about functional inoperability vary significantly among individual physicians. Most surgeons based their decision between resection and chemoradiotherapy on the expected ability to swallow and speak postoperatively, and on the wishes and expectations of the patient.[6]

In this era of evidence-based medicine, clinical decisions should be based on integrated clinical expertise, the best available evidence, and the patient's values. The integration of these three elements increases the potential for successful outcomes.[7]

To achieve an objective way of estimating the operability and to involve the patient more in decision-making, the expected functional loss of the oral cavity should be assessed with a high degree of predictability. We therefore aimed to develop a realistic, dynamic, 3-dimensional model of the oral cavity on which to do virtual resections and visualize more accurately the functional impairments after treatment. To build such a predictive system the following items should be implemented: a patient-specific biomechanical/geometric model of the tongue and the lips, including the muscular and neural systems; an electromyographic (EMG) muscular innervation model of the tongue and the lips; a biomechanical model of scar tissue; a virtual resection module that adapts the models according to the planned

2

31

intervention; and an artificial speech system that simulates the corresponding pathological speech.

Our aim in this paper was to describe the first steps that we have taken to implement the first item, a biomechanical model of the tongue, which makes it possible to do virtual resections.

METHODS

Because the aim of this project was to make virtual operating possible, we should be able to use anatomical information such as the site of muscle fibres and the course of the hypoglossal nerve. The possibility of changing material properties to simulate scar tissue is also necessary. Because of those requirements a black box model, such as the principal component analysis models developed by Badin and Serrurier[8], is not sufficient. It is possible that the finite element method will fulfil those necessities, and we have chosen it to develop the tongue model.

Data from cine magnetic resonance images (MRIs) were used to define the geometry of the model of the tongue. Image sequences were captured in the sagittal plane using a 3-T MRI unit (Achieva, Philips, Best, Netherlands). A mildly T1 weighted turbo field echo technique with spectral presaturation with inversion recovery (SPIR) fat suppression was used with a repetition time of 4.5 ms and an echo time of 2.3 ms.
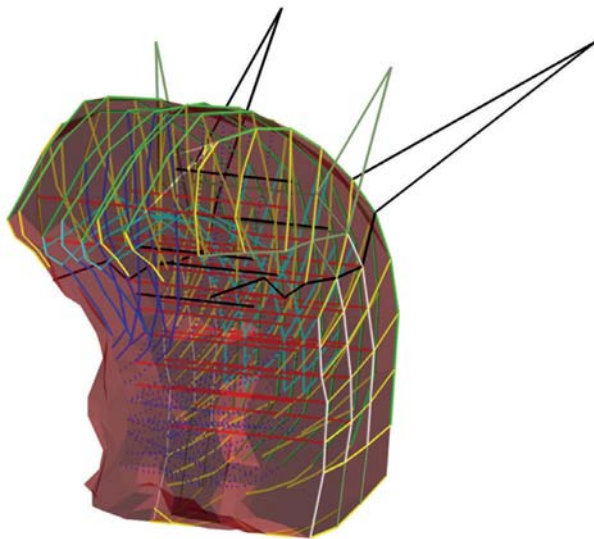

Figure 2.1 | Geometry of the finite element tongue model, including the tongue musculature. Colours indicate different muscles. Green = longitudinalis superior; red = transversus linguae; yellow= verticalis linguae; turquoise = longitudinalis inferior; blue = genioglossus; white = hyoglossus; black = styloglossus; and greyish green = palatoglossus.

The MRI were loaded into Matlab, which was also used for further development of the model. The tongue was segmented from the midsagittal slice, as well as four slices on one lateral side in equal steps of 4 mm from the midline. The opposite side was mirrored from the segmented side, assuming that the tongue is symmetrical, resulting in a total of nine slices, which enclose eight elements lateral to each other. For the meshing of the volume of the tongue we copied the model of Dang and Honda,[9] with six elements in the anteroposterior direction, and ten elements along the surface of the tongue. In total 480 3-dimensional solid quadrilateral elements were created. After the meshing process, the muscle fibres of four extrinsic and intrinsic muscles were placed between the vertices of the elements, according to the description given by Takemoto.[10] The muscles included are listed in Table 2.1. Figure 2.1 shows the musculature, as situated in the model.

Deformation and movement of the model caused by the forces generated by the activated muscles were calculated using Newton's second law:

$$\mathbf{Ma} + \mathbf{Cv} + \mathbf{Kd} = \mathbf{F} \qquad (2.1)$$

$\mathbf{M}$, $\mathbf{C}$, and $\mathbf{K}$ represent the mass, damping, and stiffness matrices, respectively, and $\mathbf{a}$, $\mathbf{v}$ and $\mathbf{d}$ are the acceleration, velocity, and displacement vectors. $\mathbf{F}$ is the force vector acting on the elements' nodes.

Activation of the muscles generates forces in the direction along the muscle fibres, and these activation patterns were set manually. In this way, the deformation caused by specific muscle contractions can be analysed carefully. At present it is not possible to obtain quantitative values about the movement of the tongue based on specific patterns of muscular activation. To retrieve a unique solution from equation (2.1), some displacements should be known. In the current model, the nodes on the jaw and the hyoid are fixed, meaning that displacements are kept at zero. The jaw and hyoid are segmented from the MRI.

The tongue is considered as a muscular hydrostat, a muscular organ that lacks skeletal support. One of the main biomechanical features of a muscular hydrostat is its incompressibility.[11] We therefore measured the volume to control if the model is indeed incompressible. The Poisson's ratio (v), a measure of the strain in the perpendicular directions as a result of an axial force in one direction, was set to 0.49. A ratio of 0.5 means perfect incompressible material, but it is not possible to use this value exactly, because of the mathematical limitations.[12]

To test if the effects of scar tissue could be mimicked, the stiffness was increased for several elements. Currently no reliable values are available for stiffness of scar tissue, and an 8-times increased stiffness was used.

Table 2.1 | Main actions of different muscles of the tongue based on qualitative observations in our model, and descriptions by Agur.[13] The m. genioglossus is divided into three part. Agur described it as one muscle, but did mention an independent action for the posterior part.

Muscle	Our model	Agur and Dalley
m. genioglossus anterior (GGA)	Depression of the apex and deepening the groove	Depresses tongue, posterior part pulls tongue anteriorly for protrusion
m. genioglossus middle part (GGM)	Protrusion of the apex and depression of the dorsum	
m. genioglossus posterior (GGP)	Protrusion	
m. hypoglossus (HG)	Depression of the dorsum	Depresses and retracts tongue
m. styloglossus (SG)	Retracts the tongue and lifts the dorsum	Retracts tongue and draws it up to create a trough for swallowing
m. palatoglossus (PG)	Elevation of the tongue	Elevates posterior part of the tongue
m. longitudinalis superior (SL)	Retraction of the tongue and elevation of the apex	Curls apex and sides of tongue superiorly and shortens tongue
m. longitudinalis inferior (IL)	Retraction of the tongue and pulls tongue tip down- and backward	Curls tip of tongue inferiorly and shortens tongue
m. transversus linguae (TRA)	Small protrusion of the tongue and elevation of the dorsum and narrowing the tongue body	Narrows and elongates the tongue
m. verticalis linguae (VER)	Downward movement of the tongue and slight protrusion	Flattens and broadens the tongue

RESULTS

A first, coarse, finite element tongue model was created, which is controlled manually. For some individual muscles the observed deformations are given in deformation is seen.

A combined contraction of m. transversus linguae, m. verticalis linguae, m. genioglossus posterior, and the middle part of m. genioglossus resulted in protrusion of the tongue (Figure 2.2). An anterior and caudal movement was seen for the movement of the tongue's tip.

A mean (SD) volume change of 103.1 (4.6)% was seen when individual muscles were activated, with a maximal volume increment of 112.3%, seen after activating the longitudinalis superior muscle, and the maximal volume decrement of 99.3% for activation of the genioglossus posterior. The volume changes/element can be larger, ranging from a mean (SD) decrement of 73.5 (32.5)% to 129.2 (29.7)%. At the start of the protrusion movement (activation of m. transversus linguae, m. verticalis linguae, m. genioglossus posterior, and m. genioglossus middle part) there was a small decrease in volume when the activation was held for a longer period and the muscle force was increased, there was a growth in volume, and a constant equilibrium of 108.8% of the original volume.
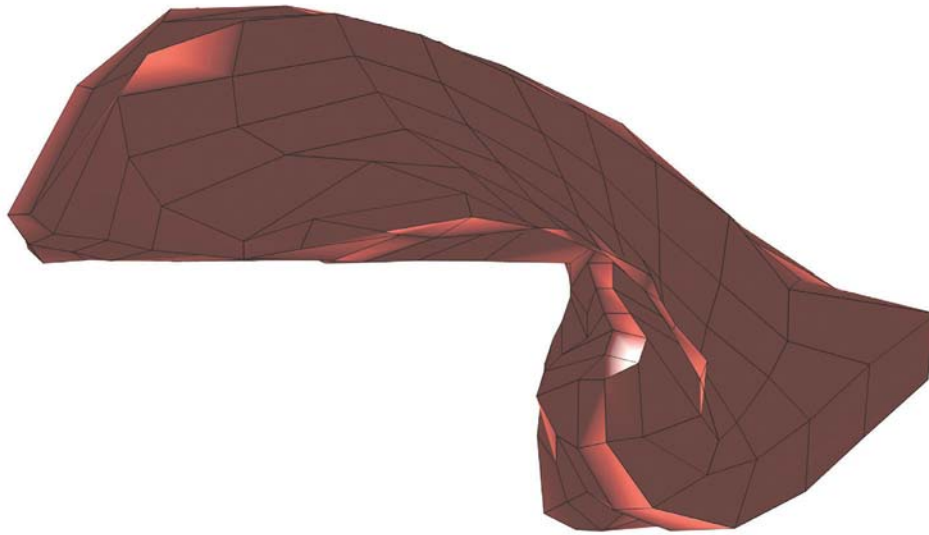
Figure 2.2 | Activation of the m. transversus linguae, m. verticalis linguae, m. genioglossus posterior, and m. genioglossus middle part, resulting in protrusion of the tongue.
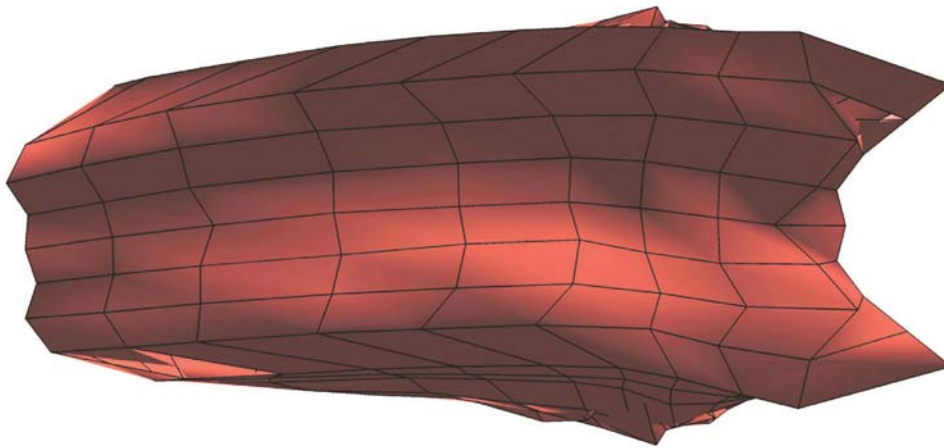
2

Figure 2.3 | Top view on resulting position after combined activation of m. transversus linguae, m. verticalis linguae, m. genioglossus posterior, and m. genioglossus middle part, with increased stiffness for elements on the left lateral cranial side, which resulted in deviation to the left.

After increasing the stiffness of several elements on the cranial left side the movement changed as expected. The resultant movement after combined activation of the m. genioglossus middle part, m. genioglossus posterior, m. transversus linguae, and m. verticalis linguae was a deviation of the tip of the tongue to the left side (Figure 2.3).

DISCUSSION

In the present study we have taken the first steps towards creation of a new model of the tongue. The choice was made for a finite element model that gives us the ability to detect causes of unexpected deformations, and enables us to adjust the inner workings of the model locally, based on virtual surgery.

The deformations as shown by the model are comparable to descriptions of in vivo movements (Table 2.1).[13] The combined activation of m. genioglossus posterior, m. genioglossus middle part, m. transverses linguae, and m. verticalis linguae is mentioned as the activation pattern for protrusion,[13] which was confirmed by our model. Because we found changes in volume, a volume constraint is desirable that prevents large volume changes but does not interfere with the movement of the tongue. Increasing the stiffness value for several elements shows deformations as they were expected beforehand. When exact tissue variables are known, our model should be able to mimic the changes in the movements of the tongue postoperatively.

When a virtual resection module of the tumour can be implemented, a patient-specific finite element model of the tongue will make an objective judgement of the expected possible losses in function. Our modelling method gives possibilities to adjust material properties in such a manner that scar and reconstructive tissue compartments can be matched.

Using diffusion tensor MRI sequences, the tracking of muscle fibres is an option to reach a higher specificity for the patient.[14] The same holds true for a possible method of imaging nerves.[15]

Implementing EMG signals for patterns of muscular activation during specific movements of the tongue would lead to further individualisation, and the number of elements should be increased to obtain a more accurate simulation and a smoother surface to the tongue. This would also improve future speech synthesis.[16]

Detailed information on tissue variables is necessary to simulate adequately the effects of resection. Finally, clinical validation studies will be necessary to assess the accuracy of the model quantitatively.

REFERENCES

[1]   A. J. M. Balm, J. H. Schornagel, and C. R. N. Rasch, "The role of simultaneous chemotherapy and radiotherapy in the treatment of locally metastasised tumours of the larynx, pharynx and oral cavity," Ned. Tijdschr. Geneeskd., vol. 149, no. 2, pp. 61–4, Jan. 2005.

[2]   B. Li, D. Li, D. H. Lau, D. G. Farwell, Q. Luu, D. M. Rocke, K. Newman, J. Courquin, J. A. Purdy, and A. M. Chen, "Clinical-dosimetric analysis of measures of dysphagia including gastrostomy-tube dependence among head and neck cancer patients treated definitively by intensity-modulated radiotherapy with concurrent chemotherapy.," Radiat. Oncol., vol. 4, no. 52, Jan. 2009.

[3]   N. Tomita, T. Kodaira, K. Furutani, H. Tachibana, Y. Hasegawa, A. Terada, K. Hanai, T. Ozawa, T. Nakamura, and N. Fuwa, "Long-term follow-up and a detailed prognostic analysis of patients with oropharyngeal cancer treated with radiotherapy.," J. Cancer Res. Clin. Oncol., vol. 136, no. 4, pp. 617–23, Apr. 2010.

[4]   A. H. Ackerstaff, I. B. Tan, C. R. N. Rasch, A. J. M. Balm, R. B. Keus, J. H. Schornagel, and F. J. M. Hilgers, "Quality-of-life assessment after supradose selective intra-arterial cisplatin and concomitant radiation (RADPLAT) for inoperable stage IV head and neck squamous cell carcinoma.," Arch. Otolaryngol. Head. Neck Surg., vol. 128, no. 10, pp. 1185–90, Oct. 2002.

[5]   A. M. Kreeft, L. van der Molen, F. J. Hilgers, and A. J. Balm, "Speech and swallowing after surgical treatment of advanced oral and oropharyngeal carcinoma: a systematic review of the literature.," Eur. Arch. Otorhinolaryngol., vol. 266, no. 11, pp. 1687–1698, 2009.

[6]   A. M. Kreeft, I. B. Tan, C. R. Leemans, and A. J. M. Balm, "The surgical dilemma in advanced oral and oropharyngeal cancer: how we do it.," Clin. Otolaryngol., vol. 36, no. 3, pp. 260–266, Jun. 2011.

[7]   D. Ilic, "The role of the internet on patient knowledge management, education, and decision-making.," Telemed. J. E. Health., vol. 16, no. 6, pp. 664–9, Jan. 2010.

[8]   P. Badin and A. Serrurier, "Three-dimensional modeling of speech organs : Articulatory data and models," Acoust. Soc. Japan, vol. 36, no. 5, pp. 421–426, 2006.

[9]   J. Dang and K. Honda, "Construction and control of a physiological articulatory model.," J. Acoust. Soc. Am., vol. 115, no. 2, pp. 853–870, 2004.

[10]   H. Takemoto, "Morphological analyses of the human tongue musculature for three-dimensional modeling.," J. speech, Lang. Hear. Res., vol. 44, no. 1, pp. 95–107, 2001.

[11]   W. M. Kier and K. K. Smith, "Tongues, tentacles and trunks: the biomechanics of movement in muscular-hydrostats," Zool. J. Linn. Soc., vol. 83, no. 4, pp. 307–324, Apr. 1985.

[12]   U. Hueck and H. L. Schreyer, "The use of orthogonal projections to handle constraints with applications to incompressible four-node quadrilateral elements," Int. J. Numer. Methods Eng., vol. 35, no. 8, pp. 1633–1661, Nov. 1992.

2

[13] A. M. R. Agur and A. F. Dalley, "Head," in Grant's Atlas of Anatomy, 11th ed., B. Sun, Ed. Philadelphia: Lippincott Williams & Wilkins, 2005, pp. 587–719.

[14] S. M. Mijailovich, B. Stojanovic, M. Kojic, A. Liang, V. J. Wedeen, and R. J. Gilbert, "Derivation of a finite-element model of lingual deformation during swallowing from the mechanics of mesoscale myofiber tracts obtained by MRI.," J. Appl. Physiol., vol. 109, no. 5, pp. 1500–14, Nov. 2010.

[15] S. Jambawalikar, J. Baum, T. Button, H. Li, V. Geronimo, and E. S. Gould, "Diffusion tensor imaging of peripheral nerves.," Skeletal Radiol., vol. 39, no. 11, pp. 1073–9, Nov. 2010.

[16] J. Beskow, "Talking Heads - Models and Applications for Multimodal Speech Synthesis," Institutionen för talöverföring och musikakustik, 2003.

Chapter 3

Predicting 3D lip poses using facial surface EMG

M.J.A. van Alphen

M. Eskes

A.J.M. Balm

L.E. Smeele

D. Brandsma

F. van der Heijden

ABSTRACT

The aim of this study is to proof that facial surface electromyography (sEMG) conveys sufficient information to predict 3D lip shapes. If this is possible, it implies that biomechanical models can also be controlled using sEMG recordings. 3D lip shapes are captured using a stereo camera setup, and simultaneously sEMG of facial muscles is recorded. Principal Component Analysis (PCA), and a modified form of General Regression Neural Network (GRNN) are used to link the sEMG measurements to 3D lip shapes. The measurements are performed on five subjects to test reproducibility. Different EMG features, window lengths, and configurations are evaluated to find the optimal settings for facial sEMG.

The error measures of the two methods were comparable. Lip shapes could be predicted with an accuracy of 2.76 mm using PCA and 2.78 mm using the modified GRNN. Slightly better results were seen using the Willison Amplitude feature and the shorter window lengths. Overall, we were able to predict lip shapes with satisfactory accuracy using sEMG.

INTRODUCTION

The treatment choice for oral cavity carcinoma is currently based on the subjective judgement of the treating physician and the multidisciplinary tumour board. Surgery, with or without adjuvant radiotherapy, is still the first treatment choice.[1] When the quality of life will be unacceptably diminished by the functional consequences of surgery, the tumour is designated as functional inoperable.[2] In these circumstances, organ sparing chemoradiation treatment serves as alternative curative treatment option. However, the functional inoperability of a particular tumour is unreliably and subjectively assessed by medical specialists.[3] To obtain more objective measures regarding the postoperative functional consequences, we proposed a virtual therapy tool, consisting of a patient specific biomechanical model of the oral cavity and oropharynx.[4] For further individualisation of such models, it was proposed to implement electromyography (EMG) signals as an estimate for muscular activations during specific tongue movements. This is difficult to achieve for the tongue region. Another part of the model of the oral cavity and oropharynx comprises the lips, which are of importance for speech function. The implementation of surface EMG (sEMG) in a lip model would be relatively easier, because sEMG measurements, as well as capturing 3D images of the lip shapes, are easier to accomplish. The lips are therefore a good proposition of the tongue model.

Important research areas regarding EMG measurement of facial muscles, are speech interfaces, mostly silent speech interfaces [5] and multi-modal speech synthesis.[6] Those studies are limited to one feature, mostly the root mean square (RMS). Often their models are also tested on a single subject. Their aim is the categorization of phonemes, words, or gestures based on facial and tongue EMG signals.[5,7–11] This is not appropriate for our purposes, where for function prediction 3D lip shapes should be estimated. Honda et al. [12] and Lucero et al. [13] published studies concerning the prediction of lip shapes. The model Honda et al. developed estimated 2D lip shapes based on video imaging. Because this model does not include jaw movements, and is restricted to 2D, it is not suitable for reaching our goal. Lucero et al. developed a 3D finite element model (FEM) of the face, including the lips.[13] Using this model, 3D locations on the lips can be estimated by controlling the facial muscles with hooked-wire EMG and the jaw movement with optical tracking. Nevertheless, their average correlation coefficients between measured and modelled position are relatively low (<0.78).

The objective of this paper is to proof that sEMG of the facial muscles and the m. digastricus convey sufficient information to predict 3D lip shapes including mouth opening. Therefore, two different methods are used, nonlinear regression by means of Principal Component Analysis (PCA) and by using a modified version of the Generalized Regression Neural Network (GRNN). This is relevant knowledge since it implies that a biomechanical model of
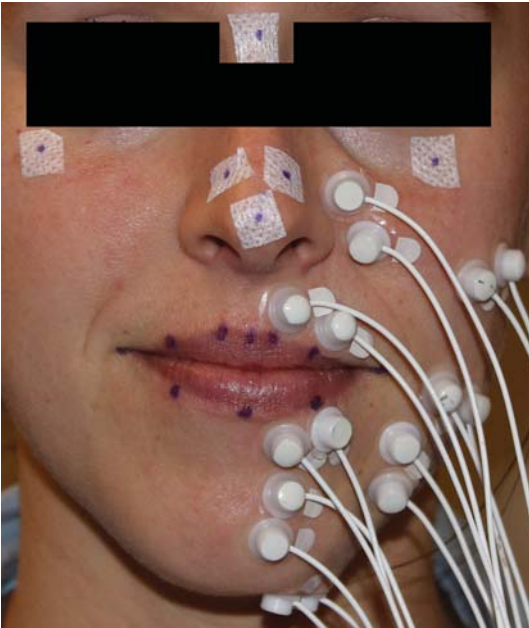
3

Figure 3.1 | Subject with surface EMG electrodes in bipolar configuration on the Zygomaticus major, Risorius, Orbicularis Oris Superior and Inferior, Mentalis, Depressor Anguli Oris, Levator Labii Superioris and the Digastricus, ten markers are locatied on the lips and six on the nose and infra-orbital.

the lips including the jaw, can be controlled by non-invasive sEMG. A biomechanical model could also be personalized by training with the sEMG measurements and the associated motion. As such this can be very helpful in solving the inverse dynamics problem [14], not only by reducing the solution space of an inverse dynamics algorithm, but also by finding the personalized muscle activation signals via a regression method trained by patients specific measurements. Furthermore in other areas, e.g. silent speech interfaces and multi modal speech synthesis, sEMG based lip modelling is relevant.[5,6,11]

MATERIALS AND METHODS

To test reproducibility, five healthy subjects were included, ages ranging from 24 to 25. To measure the 3D lip positions 10 markers were applied on the lips and six markers on the face (two infra-orbital and four on the nose), with a black fine line pen, see Figure 3.1. The markers on the face were used to compensate for head movements. The 3D locations were estimated using a stereo camera setup consisting of two consumer cameras (Casio® EX-FC100), which are placed in front of the subject. The camera setup is calibrated using a

calibration object with 27 nodes at known positions. The reconstruction error, quantified by its root mean square (RMS), averaged over 10x10x10 cm measurement space, and obtained via the leave-one-out method, is 0.63 mm. The facial muscles measured with sEMG were the m. zygomaticus (ZYG), m. risorius (RIS), m. orbicularis oris superior (OOS) and inferior (OOI), m. mentalis (MEN), m. depressor anguli oris (DAO), m. levator labii superioris (LLS). Furthermore the m. digastricus (DIG) was included to represent jaw opening in the model. The muscles were chosen based on human lip anatomy and a study of Lapatki et al.[15] The sEMG signals were measured using a Porti-system from TMSi (Oldenzaal, the Netherlands) with sintered disc shaped surface EMG electrodes (1.5 mm diameter, Ag/AgCl) with shielded cables, see Figure 3.1. A reference electrode was placed on the wrist. The EMG signals were recorded with a sample frequency of 2048 Hz. A band-pass filter was applied at 15 to 500 Hz, using a fourth order Butterworth filter. Since no significant 50 Hz power line interference was found a 50 Hz notch filter was omitted. Measurements were done with two configurations: unipolar and bipolar.

Figure 3.2 | The eight facial expressions being performed by the subjects: A. rest, B. smile, C. purse, D. growl, E. open, F. sad, G. blow, H. pout.

The subjects were asked to hold 12 poses: seven facial expressions; smile, pursing, growl, open, sad, blow, and pout (see Figure 3.2); and five vowels; /a/, /e/, /i/, /o/, /u/; for four seconds. Between each pose a rest position with closed mouth and relaxed muscles was adopted. The rest position was only measured to provide a reference for determining the magnitude of displacement of the marker positions of the other poses, and the displacements could be inferred from sEMGs. The facial expressions were selected for their correspondence to isolated contractions of facial muscles.[15] During each pose an image was acquired by both cameras. EMG measurements were performed simultaneously. This acquisition was repeated four times with a pause between each repetition. For each pose, the markers were manually selected in the images, resulting in a set of 40 image coordinates for the lips (10 markers, 2D positions in two cameras). These coordinates were reconstructed to a vector, $\mathbf{X}(p,r) \in \mathbb{R}^{30}$, that holds the 30 coordinates of the 10 markers

in 3D space. Here, $p = 0, \cdots, 12$ gives the pose number, where the rest pose is $p = 0$, and $r = 1, \cdots, 5$ is the repetition number. The markers on the face were used to register all poses to the same reference frame to correct for head motion. The manual selection of the image points induced an error. This error, and also the error induced by imperfect head motion compensation, was assessed by estimating the RMS of the differences between two runs of selection, the so-called intra-observer error. Suppose $x_m^{(1)}(p)$, $m = 1, \cdots, 30$, denotes the 3D coordinates of the marker position of pose p resulting from the first selection run and $x_m^{(2)}(p)$ from the second. Then e_{obs}, defined as the RMS of the Euclidean length of the differences, and calculated over the 10 markers and 13 poses, is:

$$e_{obs} = \sqrt{\frac{1}{130} \sum_{p=0}^{12} \sum_{m=1}^{30} \left(x_m^{(1)}(p) - x_m^{(2)}(p) \right)^2} \qquad (3.1)$$

The found RMS intra-observer error was $e_{obs} = 2.34$ mm, which implies that the reconstruction error caused by the calibration (0.63 mm) was negligible.

The magnitude of variations of the marker positions was assessed by means of the RMS of the Euclidean distance between the 10 markers of a pose and the corresponding markers in the rest position ($p = 0$). This was averaged over the 12 poses and the 5 repetitions:

$$d_{RMS} = \sqrt{\frac{1}{600} \sum_{r=1}^{5} \sum_{p=1}^{12} \sum_{m=1}^{30} \left(x_m(p,r) - x_m(0,r) \right)^2} \qquad (3.2)$$

Here, $x_m(p,r)$ denotes the mth element from the vector $\mathbf{X}(p,r)$. The magnitude of variation is corrected for the intra-observer (e_{obs}) by:

$$d_c = \sqrt{d_{RMS}^2 - e_{obs}^2} \qquad (3.3)$$

To estimate lip positions based on sEMG recordings we used two different nonlinear regression methods: using PCA and a modified version of GRNN. These methods will be described below. The accuracy of an estimated positions $\hat{\mathbf{X}}(p,r)$ from both regression methods was assessed by the RMS of the Euclidean length of the difference between marker positions of the estimated pose and the reference pose:

$$e_{RMS} = \sqrt{\frac{1}{600} \sum_{r=1}^{5} \sum_{p=1}^{12} \sum_{m=1}^{30} \left(\hat{x}_m(p,r) - x_m(p,r) \right)^2} \qquad (3.4)$$

where $\hat{x}_m(p,r)$ is the mth element of the vector $\hat{\mathbf{X}}(p,r)$. This RMS was corrected for the manual selection error of one run, thus $e_{obs}/\sqrt{2}$. The factor $1/\sqrt{2}$ appears since the e_{obs} is based on two manual selections runs, and in equation (3.4) we have only one selection run:

$$e_c = \sqrt{e_{RMS}^2 - \tfrac{1}{2}e_{obs}^2} \qquad (3.5)$$

To compare subjects with each other the ratio of the error to position deviation (e_r) is defined:

$$e_r = \frac{e_c}{d_c} \qquad (3.6)$$

To compare our results with literature [13], we calculated correlation coefficients between $\hat{x}_m(p,r)$ and $x_m(p,r)$. This was done per subject and per coordinate resulting in 30 correlation coefficients. These were averaged to retrieve the mean correlation coefficient of each subject.

Table 3.1 | Used EMG feature types and their formulas.

Feature	Formula
MAV	$g(j) = \dfrac{1}{N} \displaystyle\sum_{i=j-\frac{1}{2}(N-1)}^{j+\frac{1}{2}(N-1)} |s(i)|$
RMS	$g(j) = \sqrt{\dfrac{1}{N} \displaystyle\sum_{i=j-\frac{1}{2}(N-1)}^{j+\frac{1}{2}(N-1)} s(i)^2}$
WL	$g(j) = \displaystyle\sum_{i=j-\frac{1}{2}(N-1)}^{j+\frac{1}{2}(N-1)-1} |s(i+1) - s(i)|$
WAMP	$g(j) = \displaystyle\sum_{i=j-\frac{1}{2}(N-1)}^{j+\frac{1}{2}(N-1)-1} \left[v\big(|s(i+1) - s(i)|\big) \right]$ $v(s) = \begin{cases} 1, & \text{if } s \geq \text{ threshold} \\ 0, & \text{otherwise} \end{cases}$

In this study four EMG feature types are compared with each other. The features extracted from the EMG measurements were: Mean Absolute Value (MAV), Root Mean Square (RMS), Waveform Length (WL), Willison Amplitude (WAMP) with thresholds at 10 and 20 mV. In Table 3.1 the equations for the different features are given. These features were chosen because they performed well in a recent study.[16] The features were defined for

six different time windows: 50, 100, 150, 200, 250 and 300 ms, the time segments overlap so the features are determined for each time point i in the measured signals (except for half a time window at the beginning of the measurement and the ending). The features of the eight EMG channels were represented by 60 feature sets $\mathbf{g}(i,p,r)$, $p = 1,\cdots,12$ (the poses), and $r = 1,\cdots,5$ (the repetitions). From each feature set three seconds, i.e. 6145 samples, were used for further processing (so, $i = 1,\cdots,6145$). Therefore each feature set consisted of an equal number of samples. Because two different physical dimensions and units were present, all coordinates $\mathbf{X}(p,r)$ and EMG features $\mathbf{g}(i,p,r)$ were normalized with respect to their mean and variance, i.e. the mean was set to zero by subtraction, and the variance was set to one by scaling. The mean and variance were estimated per subject over all available time samples. Estimated coordinates were transformed back to the original domain by performing the inverse operation.

The PCA-based regression method was trained using the database consisting of the selected marker positions and associated feature vectors. For each lip position $\mathbf{X}(p,r)$, only one EMG feature vector was needed instead of the whole set of 6145 EMG feature vectors $\mathbf{g}(i,p,r)$. Therefore, the vectors in a set were first averaged over the time samples to yield a single 8D feature vector $\overline{\mathbf{g}}(p,r)$. To implement nonlinear regression, these vectors were augmented with the $\frac{1}{2}\times 8\times 9 = 36$ quadratic terms that can be formed from the 8 elements in $\overline{\mathbf{g}}(p,r)$. This gave a set of 44D data vectors $\overline{\mathbf{g}}_{aug}(p,r)$. The next step was to concatenate this vector with the 30 coordinates $\mathbf{X}(p,r)$ of the pose resulting in a 74D vector $\mathbf{z}(p,r)$:

$$\mathbf{z}(p,r) = \begin{bmatrix} \mathbf{X}(p,r) \\ \overline{\mathbf{g}}_{aug}(p,r) \end{bmatrix} \text{ with } \begin{array}{l} p = 1,\cdots,12 \text{ (poses)} \\ r = 1,\cdots,5 \text{ (repetitions)} \end{array} \qquad (3.7)$$

The training of the PCA model was combined with cross validation to prevent a possible bias in the performance evaluation. Training was based on the pooled data from four repetitions of 12 poses each. Validation was done on the remaining repetition. This process was iterated five times by rotating the five repetitions in the training pool and the validation set. Each training pool consisted of a 74×48 matrix $\mathbf{Z}$ the columns of which are vectors $\mathbf{z}(p,r)$.

The PCA model $\mathbf{Y}$ is a $74\times D$ matrix, containing the D most dominant eigenvectors $\mathbf{y}_{d'}$ resulting from $\mathbf{Z}\mathbf{Z}^T\mathbf{y}_d = \lambda_d\mathbf{y}_{d'}$ where λ_d are the eigenvalues. The eigenvectors, i.e. the principal components, are normalized such that $\mathbf{Y}\mathbf{Y}^T = \mathbf{I}$. The eigenvectors $\mathbf{y}_d$ were sorted in $\mathbf{Y} = [\mathbf{y}_1,\ldots,\mathbf{y}_D]$ in such a way that for the corresponding eigenvalues holds that $\lambda_d \geq \lambda_{d+1}$. Using $\mathbf{Y}$, any data vector $\mathbf{z}$ can be encoded in a lower D-dimensional coefficient vector $\mathbf{b}$:

$$\mathbf{b} = \mathbf{Y}^T\mathbf{z} \qquad (3.8)$$

46

Decoding of $\mathbf{z}$ from $\mathbf{b}$ uses the same model:

$$\mathbf{z} \approx \mathbf{Yb} \qquad (3.9)$$

To estimate the lip positions from the EMG data, equation (3.9) was adapted in such a way that only the EMG signals are reconstructed. We defined the submatrix $\mathbf{Y}_g$ of $\mathbf{Y}$ that only contained the EMG features; that is the lower 44 rows of $\mathbf{Y}$. We then have:

$$\overline{\mathbf{g}}_{aug} = \mathbf{Y}_g \mathbf{b} + \mathbf{v} \qquad (3.10)$$

where $\mathbf{v}$ contains the residuals that represent the approximation error in equation (3.9). Equation (3.10) can be regarded as a linear observation model of $\mathbf{b}$ with $\mathbf{Y}_g$ the observation matrix and $\mathbf{v}$ the observation noise. Least Squared Error (LSE) estimation of $\mathbf{b}$ is then straightforward:[17]

$$\hat{\mathbf{b}}_{LSE} = \left(\mathbf{Y}_g^T \mathbf{Y}_g \right)^{-1} \mathbf{Y}_g^T \overline{\mathbf{g}}_{aug} \qquad (3.11)$$

$\hat{\mathbf{b}}_{LSE}$ is the estimated coefficient vector. From that, the full vector $\mathbf{z}$, including the 30 coordinates $\mathbf{X}$, can be estimated by application of equation (3.9):

$$\hat{\mathbf{z}}_{LSE} = \mathbf{Y}\hat{\mathbf{b}}_{LSE} \qquad (3.12)$$

An extension of this estimation of $\mathbf{b}$ is the Minimum Mean Square Error (MMSE) estimation. This method exploits the prior knowledge that the PCA coefficients are uncorrelated with zero mean. The covariance matrix $\mathbf{C}_b$ of $\mathbf{b}$ is diagonal with diagonal elements λ_d. In case of uncorrelated residuals $\mathbf{v}$, its covariance matrix is proportional to the unity matrix $\mathbf{C}_v = \sigma_v^2 \mathbf{I}$. The unbiased MMSE estimate $\hat{\mathbf{b}}_{MMSE}$, based on the EMG features, is then given by:[17]

$$\hat{\mathbf{b}}_{MMSE} = (\mathbf{Y}_g^T \mathbf{Y}_g + \sigma_v^2 \mathbf{C}_b^{-1})^{-1} \mathbf{Y}_g^T \overline{\mathbf{g}}_{aug} \qquad (3.13)$$

Note that when σ_v is set to zero, $\hat{\mathbf{b}}_{MMSE}$ is equal to $\hat{\mathbf{b}}_{LSE}$. So, the LSE estimate is encompassed by the MMSE estimate, and there is no need to treat it separately.

The second regression method is an extension on the General Regression Neural Network (GRNN). GRNN is a nonlinear interpolation method that is based on Parzen kernel density models.[18] We combined the design and evaluation of the GRNN with cross validation in the same way as outlined above. First, we defined a linear index over the poses and repetitions: $c(r, p) \overset{def}{=} 12(r-1) + p$. Assume the vectors $\mathbf{X}_c$ and $\overline{\mathbf{g}}_c$ are associated, we have

3

a population of pairs available in a training pool $\left\{\left(\mathbf{X}_c, \overline{\mathbf{g}}_c\right) \mid c = 1, \cdots, 48\right\}$. Given a new EMG vector $\overline{\mathbf{g}}$, a GRNN estimates the associated vector $\hat{\mathbf{X}}$ by:

$$\hat{\mathbf{X}} = \sum_{c=1}^{48} w_c \mathbf{X}_c \quad \text{with:} \quad \mathbf{w}_c = \frac{s\left(\overline{\mathbf{g}}, \overline{\mathbf{g}}_c\right)}{s\left(\overline{\mathbf{g}}, \overline{\mathbf{g}}_1\right) + s\left(\overline{\mathbf{g}}, \overline{\mathbf{g}}_2\right) + \ldots + s(\overline{\mathbf{g}}, \overline{\mathbf{g}}_{48})} \tag{3.14}$$

where $s\left(\overline{\mathbf{g}}, \overline{\mathbf{g}}_c\right)$ is a similarity measure between $\overline{\mathbf{g}}$ and $\overline{\mathbf{g}}_c$ derived from a Parzen estimate of the underlying probability density. We replaced the Parzen kernel by the likelihood function $p(\mathbf{g} \mid c)$. We defined the 48 poses in a training pool each as an individual class. For each class, a feature set $\mathbf{g}(i, p, r)$ is available. This set was used to train the likelihood function $p(\mathbf{g} \mid c)$. Under the assumption of normal distributions for the likelihood function: $p(\mathbf{g} \mid c) = N\left(\mathbf{g} - \mu_c, \mathbf{C}_c\right)$, learning boils down to estimating the mean μ_c and the covariance matrix $\mathbf{C}_c$. That is, $\hat{\mu}_c = \overline{\mathbf{g}}_c$. The similarity measures associated with a new vector $\overline{\mathbf{g}}$ are defined as follows:

$$s(\overline{\mathbf{g}}, \overline{\mathbf{g}}_c) = N(\overline{\mathbf{g}} - \overline{\mathbf{g}}_c, \alpha^2 \mathbf{C}_c + \gamma \mathbf{I}) \tag{3.15}$$

The factor α^2 was added to improve the generalization capability. For each pose, only four repetitions in a training pool were available. Therefore, poses were not well populated in the 8-dimensional feature space. By spreading the Gaussians kernels with the factor α^2 the overlap between kernels increased. The term $\gamma \mathbf{I}$ was added for improving the numerical stability, but the choice of γ (around 10^{-6}) was not critical.

For each regression method, the best performing combination of feature type, window length, and configuration (that is, 60 different combinations) was determined using the cross validation technique mentioned above. In the PCA method, for each of such a combination, the parameter σ_v was evaluated over the range $\sigma_v = 0, 0.05, \cdots, 0.3$, and the PCA dimension D over the range $D = 1, \cdots, 48$. For the GRNN method the parameter α was analysed over the range $\alpha = 1, 2, \cdots, 10$.

RESULTS

In Table 3.2 the results of the five individual subjects and after averaging the errors over the five subjects, are presented for both methods. The averaged data showed the error, corrected for the intra-observer selection error, (e_c) was slightly lower using the PCA method compared to the GRNN method, 2.76 mm and 2.78 mm, respectively. The ratios of these error to position deviation (e_r) were 0,22 and 0,24. However, the GRNN method performed slightly better in four subjects.

Table 3.2 | The optimal settings and corresponding errors per subject and averaged over the subjects.

	e_c (mm)		e_r		Corr. Coef.		Config.[a]		Feat[b]		Window (ms)		D	σ_v	α
Method[c]	**P**	**G**	**P**	**G**	**P**	**G**	**P**	**G**	**P**	**G**	**P**	**G**	**P**		**G**
1	2.71	2.69	0.24	0.24	90	90	1	1	1	1	200	200	10	0.05	3
2	3.00	3.12	0.29	0.31	88	86	2	2	2	3	300	150	17	0.25	5
3	2.41	2.25	0.23	0.22	91	92	1	1	1	1	300	50	12	0.10	2
4	1.99	1.42	0.23	0.17	91	94	1	2	1	4	300	50	7	0.05	2
5	2.94	2.39	0.33	0.27	87	90	1	1	1	5	100	200	6	0.05	3
Average	2.76	2.78	0.22	0.24	89	89	1	2	1	4	300	50	9	0.05	2

a.   The monopolar configuration is represented by 1 and bipolar configuration by 2.

b.   The found optimal features are represented by numbers as follows 1. WAMP 10 mV; 2. WAMP 20 mV; 3. WL; 4. MAV; 5. RMS.

c.   The PCA method is represented by P; and the GRNN method by G.

The lowest averaged error was found using unipolar configuration, when positions are estimated using PCA. For GRNN bipolar configuration showed the lowest error on average, however, three subjects performed better using unipolar configuration. By changing the configuration, e_c become 3.78 mm and 3.01 mm, for PCA and GRNN respectively. The PCA method showed more consistency regarding the chosen feature. For all subjects the WAMP feature showed the best results, four of them with a threshold of 10 mV, which also showed the lowest error rates on average, and one with 20 mV. On average, the GRNN method performed somewhat better using the MAV feature. However, the optimal feature differed in the individual subjects, where two subjects performed better using the WAMP feature with threshold of 10 mV, and also the WL and RMS features were found as optimal settings for one subject each. In Table 3.3 the optimal settings per feature, averaged over the subjects, are given with the corresponding error measures, for both methods. Regarding the PCA method the WAMP with a threshold at 10 mV performs clearly better. The error measures between the different features using GRNN are much smaller, with a maximum difference in e_c of 0.4 mm between the features MAV and WAMP 20 mV. The methods showed differences regarding the window size which resulted in the lowest errors, with 300 ms for the PCA method and 50 ms using GRNN. However, comparing the different windows, the PCA showed a maximum difference in e_c of 0.01 mm, where for the GRNN method this error increases from 2.78 to 3.46 mm when a 300 ms window was used.

The optimal PCA dimension D ranged between subjects from 6 to 17, and showed an optimum in the averaged results at 9. By evaluating the results for all different dimensions, we found that the first four to five principal components have large influence on the prediction of the position. Using eight principal components the errors reach a plateau, after which only small changes in the errors are found. The explained variance by the PCA model using

nine principal components ranges from 92 to 96% between the different subjects. The optimal value for σ_v was never found to be zero, meaning the MMSE analysis performed better. On average a clear trend is found for higher e_c values when σ_v is increased. When σ_v is increased by 0.05, this error increases by 0.11 on average, ranging from 0.10 to 0.13. For four subjects the same trend was seen, and only for subject two a small decrease of the error was seen till a σ_v of 0.25.

The averaged results showed the lowest error with the parameter α at a value of two. Only subject two with a value of five deviated clearly, the error when α is two, is for this subject 0.50 mm larger, when the other settings are not adapted.

By comparing subjects based on the values in e_r it can be seen that the PCA method is more consistent, with values ranging from 0.23 to 0.33, whereas the modified GRNN showed values between 0.17 and 0.31. But the value e_r for the averaged data is comparable, 0.22 versus 0.24.

Table 3.3 | The optimal results, averaged over the subjects, obtained per feature using PCA and GRNN, with the corresponding configurations.

	RMS		MAV		WL		WAMP 10 mV		WAMP 20 mV	
Method[a]	P	G	P	G	P	G	P	G	P	G
e_c **(mm)**	3.31	3.10	3.33	3.04	3.28	3.32	2.76	3.12	3.27	3.44
e_r	0.34	0.31	0.34	0.31	0.33	0.34	0.28	0.32	0.33	0.35
Corr. Coef.	87	89	87	89	87	87	89	88	87	86
Config[b]	1	2	1	2	1	1	1	1	1	1
Window (ms)	50	50	50	50	50	50	300	150	200	50
D	7	NA	7	NA	9	NA	9	NA	16	NA
σ_v	0.05	NA	0.05	NA	0.05	NA	0.05	NA	0.1	NA
α	NA	2	NA	2	NA	2	NA	3	NA	2

a.   The PCA method is represented by P; and the GRNN method by G.
b.   The monopolar configuration is represented by 1 and bipolar configuration by 2.

DISCUSSION

This study has shown that static lip shapes can be estimated from sEMG measurements of facial muscles and the m. digastricus. The tested methods, PCA-based nonlinear regression, and a modified GRNN, gave comparable results with an accuracy of about 2.8 mm.

In the PCA based approach, MMSE performed better than LSE. The MMSE method uses the additional knowledge that the values in the coefficients **b** are uncorrelated with zero

means, and variances that are known from the PCA model.[17] Therefore, the growth of the coefficients is controlled, and a higher PCA dimension can be achieved leading to a more accurate estimations. The corrected RMS error of 2.76 mm and a correlation coefficient of 0.89 are promising. The modified GRNN showed almost identical results. Both models are applicable for different subjects, but PCA is the most consistent between subjects. However in four subjects the modified GRNN showed more accurate position estimates.

A disadvantage for using the modified GRNN could be the fact that it can be regarded as an interpolation method, using a lookup-table that is filled by the training set, and that is probed by the EMG feature vector of the unknown pose. Provided that the feature vector probes in the vicinity of feature vectors in the table, as was the case in this study, the method performs well. However, if the feature vector probes in a white area of the lookup-table, the results are less predictable. This can occur when a pose is adopted that is not present in the training set. In this respect, PCA might be advantageous, where it can be expected that its behaviour in the untrained regions, will be smoother. This is an important aspect, because if we want to predict functional loss, we should be able to predict functional movements of, among others, the lips based on activations of multiple muscles. But it is not possible to create a patient specific training set including all possible positions.

For the prediction of functional outcome post treatment, lip shapes of each individual subject should be mimicked as accurate as possible. As most models in literature don't give quantitative values it is difficult to compare our results. Others mentioned correlation coefficients, based on their lip shape prediction using a FEM of the face.[13] Intramuscular hooked-wire bipolar EMG measurements of facial muscles were used as input for their model. Mouth opening was simulated using an optical tracked marker. Five markers were placed on the lips. Vertical displacements and protrusion of these markers were estimated, and the method was assessed using cross-correlation coefficients. Mean correlation coefficients of 0.78 and 0.46 were found for the vertical direction and protrusion, respectively. Our methods performed better than this finite element approach, especially when one consider that in our model also mouth opening is predicted using sEMG signals measured from the DIG. However, it must be noted that a finite element could establish a one-to-one correspondence between anatomy and physiology on the one hand, and mathematical structures on the other hand. Our models solely describe phenomena, and hold the risk of losing the correlation with anatomical details. In view of the application for predicting functional loss this consideration is in favour of FEMs. Another potential advantage of the FEM approach is that they can include the (nonlinear) dynamics of the anatomy/physiology. For some poses, like the vowels /a/ and /e/, little persistent activation of the muscles is needed to hold these poses. Without considerations of the muscle activation (and the associated EMG patterns) that led to these poses, it is difficult to make a distinction between those poses.

3

Our study did not lead to very decisive conclusions about the best parameters of the EMG feature. A clear optimal feature type could not be determined. The PCA-based method performed slightly better using the WAMP feature, with a threshold at 10 mV, whereas the GRNN method did not show persistent behaviour in this respect. The window length did not show a clear preference when evaluating the PCA-based method. This is possibly caused by the fact that only static poses were evaluated with an EMG feature sequence averaged over three seconds. The GRNN method performed better with smaller windows. This might be caused by the fact that small windows have more fluctuations in the EMG features. This enlarges the covariance matrix and leads to a better kernel coverage in the 8-dimensional feature space. We could not find a preferred configuration. For most subjects, PCA performed better with unipolar measurements, whereas the modified GRNN had no clear preference.

The methods proposed in this research were tested on five subjects, with satisfying performances. This indicates that sEMG conveys sufficient information to estimate 3D positions of the lips. It must be noted that training sets are subject-specific and cannot be used for the estimation of lip poses of another subject as positioning of the EMG electrodes on the skin remains critical. The current models need to be extended for the development of a virtual therapy model, that can predict functional outcome. Dynamics should be incorporated as well as unilateral movements of the lips. Therefore video capturing of the lips and bilateral EMG measurements are necessary. If a regression based model including dynamics is developed, our recommendation for future research is to combine FEMs with nonlinear regression-based models. Since the regression techniques could be applied to model the person specific neural activations of simulated muscles, instead of lip positions, it could be the next step solving the inverse dynamics problem, and reaching personalized activation signals for a biomechanical model. The first step in developing such an activation model of the lips was taken.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

ETHICAL APPROVAL

The described project involved healthy human volunteers. The performed measurements were non-invasive and not stressful. There was no infringement of the (psychological) integrity of the volunteers. Therefore no ethical approval was required.

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 Societies).

Oral information and consent is obtained.

3

REFERENCES

[1]   S. M. Gore, A. K. Crombie, M. D. Batstone, and J. R. Clark, "Concurrent chemoradiotherapy compared with surgery and adjuvant radiotherapy for oral cavity squamous cell carcinoma," Head Neck, vol. 37, no. 4, pp. 518–523, 2015.

[2]   A. Kreeft, I. B. Tan, M. W. M. van den Brekel, F. J. Hilgers, and A. J. M. Balm, "The surgical dilemma of 'functional inoperability' in oral and oropharyngeal cancer: current consensus on operability with regard to functional results," Clin. Otolaryngol., vol. 34, no. 2, pp. 140–146, 2009.

[3]   A. M. Kreeft, L. van der Molen, F. J. Hilgers, and A. J. Balm, "Speech and swallowing after surgical treatment of advanced oral and oropharyngeal carcinoma: a systematic review of the literature.," Eur. Arch. Otorhinolaryngol., vol. 266, no. 11, pp. 1687–1698, 2009.

[4]   M. J. A. van Alphen, A. M. Kreeft, F. van der Heijden, L. E. Smeele, and A. J. M. Balm, "Towards virtual surgery in oral cancer to predict postoperative oral functions preoperatively," Br. J. Oral Maxillofac. Surg., vol. 51, no. 8, pp. 747–751, 2013.

[5]   B. Denby, T. Schultz, K. Honda, T. Hueber, J. M. Gilbert, and J. S. Brumberg, "Silent speech interfaces," Speech Commun., vol. 52, no. 4, pp. 270–287, 2010.

[6]   J. Beskow, "Talking Heads - Models and Applications for Multimodal Speech Synthesis," Institutionen för talöverföring och musikakustik, 2003.

[7]   S. P. Arjunan, H. Weghorn, D. K. Kumar, G. Naik, and W. C. Yau, "Recognition of human voice utterances from facial surface EMG without using audio signals," in Lecture Notes in Business Information Processing, 2008, vol. 12 LNBIP, pp. 366–378.

[8]   B. J. Betts, K. Binsted, and C. Jorgensen, "Small-vocabulary speech recognition using surface electromyography," Interact. Comput., vol. 18, no. 6, pp. 1242–1259, 2006.

[9]   M. Hamedi, S. H. Salleh, T. S. Tan, K. Ismail, J. Ali, C. Dee-Uam, C. Pavaganun, and P. P. Yupapin, "Human facial neural activities and gesture recognition for machine-interfacing applications.," Int. J. Nanomedicine, vol. 6, pp. 3461–3472, 2011.

[10]   T. Schultz and M. Wand, "Modeling coarticulation in EMG-based continuous speech recognition," Speech Commun., vol. 52, no. 4, pp. 341–353, 2010.

[11]   J. S. Brumberg, A. Nieto-Castanon, P. R. Kennedy, and F. H. Guenther, "Brain-computer interfaces for speech communication," Speech Commun., vol. 52, no. 4, pp. 367–379, 2010.

[12]   K. Honda, T. Kurita, Y. Kakita, and S. Maeda, "Physiology of the lips and modeling of lip gestures," J. Phon., pp. 243–254, 1995.

[13]   J. C. Lucero and K. G. Munhall, "A model of facial biomechanics for speech production," J. Acoust. Soc. Am., vol. 106, no. 5, pp. 2834–2842, 1999.

[14]   I. Stavness, J. E. Lloyd, and S. Fels, "Automatic prediction of tongue muscle activations using a finite element model," J. Biomech., vol. 45, no. 16, pp. 2841–2848, 2012.

[15] B. Lapatki, D. Stegeman, and M. Zwarts, "Selective contractions of individual facial muscle subcomponents monitored and trained with high-density surface EMG," in The Facial Palsies. Complementary Approaches, C. Beurskens, R. van Gelder, P. Heymans, J. Manni, and J. Nicolai, Eds. Utrecht, 2005, pp. 89–108.

[16] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, "Feature reduction and selection for EMG signal classification," Expert Syst. Appl., vol. 39, no. 8, pp. 7420–7431, Jun. 2012.

[17] F. van der Heijden, R. P. W. Duin, D. de Ridder, and D. M. J. Tax, Classification, Parameter Estimation and State Estimation. Chichester, UK: John Wiley & Sons, Ltd, 2004.

[18] D. F. Specht, "A general regression neural network.," IEEE Trans. Neural Netw., vol. 2, no. 6, pp. 568–576, 1991.

3

55

Chapter 4

On the feasibility of sEMG controlled models for lip motion

M. Eskes*

M.J.A. van Alphen*

L.E. Smeele

A.J.M. Balm

D. Brandsma

F. van der Heijden

* Contributed equally

ABSTRACT

If surgical resection of a tumour results in an unacceptable loss of function, the tumour is designated as functionally inoperable and alternative curative treatment options, like chemoradiotherapy or radiotherapy, can be considered. An accurate prediction of these functional consequences of surgery is required to make the right choice of treatment. In this study we focus on modelling the lips, since these are essential for speech, oral food transport, and facial mimicry. 3D lip movement and corresponding sEMG activities are measured in five healthy volunteers to create an accurate, 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. Each volunteer performed nineteen different instructions which they repeated five times. The relationship between 3D lip movement and sEMG activities were accommodated in a state space model. The state-space measurement model used a truncated Taylor series up to order two combined with PCA. We tested two state space models for modelling dynamics, a first and a second-order system. PCA coefficients were estimated with a discrete Kalman filter. We used some fudging parameters to fine-tune the system. Accuracy was assessed by calculation of a RMS error. A good relationship between sEMG activities and 3D lip movement was established with an average RMS error of 2.43 mm for the first-order system, and 2.46 mm for the second-order system.

INTRODUCTION

Oral cancer, including that of the lips, is the sixth most common cancer worldwide.[1] Surgery is still standard care [2], and can lead to deterioration of speech, swallowing, and mastication with serious consequences on quality of life [3]. If surgical resection of a tumour results in an unacceptable loss of function the tumour is designated as functionally inoperable, and alternative curative treatment options like chemoradiotherapy or radiotherapy can serve as alternative treatments.[4]

Accurate prediction of the functional consequences of surgery is an urgent need to make the right choice of treatment.[5] 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 neural 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.[6–9]

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, utilizing surface electromyography (sEMG) of facial muscles, is more phonetic in nature and is mainly focused on the categorization of facial expressions [10], categorization of vowels [11], and words [12].

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., recorded lip motion in the 2D frontal projection of the face, and sEMG signals from only one side.[13] 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 performed. Lucero & 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.[14] 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 lip, 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 used five Japanese vowels, and the subject in the study of Lucero et al. was asked to produce an English sentence. Both models were tested on one volunteer.

4

For the prediction of functional outcome of a treatment, inverse dynamic modelling is needed to find the muscular activation signals. However, inverse dynamic modelling suffers from ambiguity problems [15], and therefore the solution could easily differ from the specific muscular activities of a patient. The addition of the sEMG could provide further information about these patient-specific muscular activities. As a first step in exploiting this possible information source, the current study addresses the forward modelling to demonstrate that sEMG signals indeed provide this information for controlling dynamic models of facial expressions, and particularly lip movements. To avoid the complexity and pitfalls of detailed biomechanical models, we accommodated the dynamics of the system in an empirically obtained state space model. We aimed at creating an accurate model by establishing the relationship between bilateral sEMG activities of eight facial muscles, including the digastric muscle, as input, and the 3D lip displacements as the output.

MATERIALS AND METHODS

Volunteers and data acquisition

Data was obtained from five healthy volunteers ($k = 1,\cdots,5$) consisting of two males and three females, ages 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 ($n = 1,\cdots,16$), see Figure 4.1. The locations were chosen based on human lip anatomy and a study of Lapatki et al. showing the effects on lip shapes.[16] Additionally, a common ground reference electrode was applied with a self-adhesive button electrode on the left wrist. In Table 4.1 the measured muscles, their functions and electrode number, corresponding to the numbering in Figure 4.1, are given. Sixteen facial markers were defined using a skin marker. Ten markers covered the lip contour ($m = 1,\cdots,10$). The other six markers ($m_{OR} = 1,\cdots,6$) were located on the face (cheeks, nose and forehead; see Figure 4.1) and were used to compensate for head movement. The volunteers were positioned in front of a triple camera setup consisting of three cameras (Basler avA1000-100gc), which recorded the lip movement at 100 frames per second.

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.[18] 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.[16] Lastly, two asymmetric motions were performed from left to right to
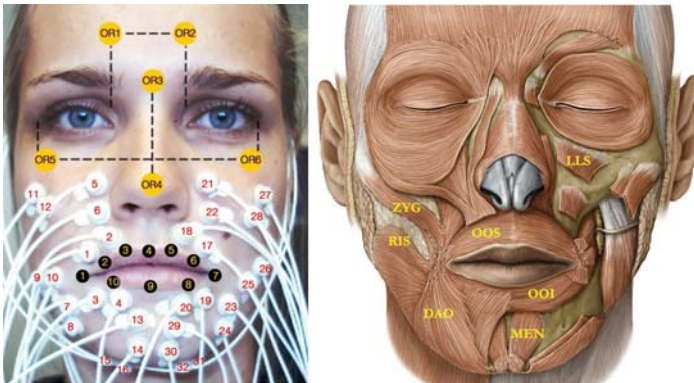
Figure 4.1 | Left: locations of electrodes, orientation markers, lip markers, and facial markers. Right: measured facial muscles, excluding the digastric muscle.[17]

left with closed lips, and with open lips, and one dynamic motion transfer between two expressions; purse lips to smiling to purse lips. Each volunteer was asked to repeat the nineteen instructions ($i = 1, \cdots, 19$) five times ($r = 1, \cdots, 5$). The instructions are shown Table 4.2.

Data processing and analysis

sEMG pre-processing

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 lowpass cutoff frequency of respectively 15 and 500 Hz, in accordance with van Boxtel et al.[19]

Table 4.1 | 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

61

Many different sEMG feature types have been proposed in 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 4.3.[20] Thresholds for the WAMP feature were set to 10 and 20 mV. With all 16 EMG channels stacked in a vector the result is denoted: $\mathbf{g}_f(t,i,r) \in \mathbb{R}^{16}$, with $f = 1,\cdots,5$ the feature type. Features were calculated over a sliding window. 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.

Table 4.2 | Instructions: visemes, facial expressions, and asymmetric movements.

#	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], [ɔ]	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

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., used 70 ms [13], whereas Vatikiotis et al., used different delays varying from 0 to 100 ms [21]. By minimizing 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.

Table 4.3 | sEMG features.

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

Video pre-processing

The facial markers were tracked in the three camera views, and the 2D coordinates were reconstructed to a set of 3D coordinates. The RMS error of the 3D localization 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 $\mathbf{X}(t,i,r)\in\mathbb{R}^{30}$.

The measurement model

State space estimation requires the availability of a measurement model that links the sEMG features $\mathbf{g}_f(t,i,r)$ 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 152D 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 $\underline{\mathbf{g}}(j)$ and corresponding marker positions $\mathbf{X}(j)$, with $j=1,\cdots,J$, is available. The exact construction

of this training set will be explained later. PCA was applied to the concatenation of these vectors:

$$\mathbf{z}(j) \overset{def}{=} \begin{bmatrix} \mathbf{X}(j) \\ \underline{\mathbf{g}}(j) \end{bmatrix} \tag{4.1}$$

The dimension of the vectors $\mathbf{z}(j)$ is 182. The set was normalized 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:

$$\mathbf{b} = \mathbf{Y}^T \mathbf{z} \quad \text{(encoding)}$$
$$\hat{\mathbf{z}} = \mathbf{Y}\mathbf{b} \quad \text{(decoding)} \tag{4.2}$$

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:

$$\hat{\underline{\mathbf{g}}} = \mathbf{Y_g}\mathbf{b} \tag{4.3}$$

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:

$$\underline{\mathbf{g}} = \mathbf{Y_g}\mathbf{b} + \mathbf{v} \tag{4.4}$$

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 $\mathbf{g}$ with quadratic terms, the measurement noise is not guaranteed to be uncorrelated, and the matrix $\mathbf{C_v}$ might be non-diagonal.

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:

$$\mathbf{b}(t+1) = \mathbf{F}\mathbf{b}(t) + \mathbf{w}(t) \qquad (4.5)$$

$\mathbf{b}(t)$ 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 $\mathrm{E}\left[\mathbf{b}(t+1)\mathbf{b}^T(t)\right] = \mathbf{F}\,\mathrm{E}\left[\mathbf{b}(t)\mathbf{b}^T(t)\right]$. In this equation, $\mathrm{E}[\ ]$ is the expectation operator, and therefore $\mathbf{F}$ can be estimated by $\hat{\mathbf{F}} = \overline{\mathbf{b}(t+1)\mathbf{b}^T(t)}\left(\overline{\mathbf{b}(t)\mathbf{b}^T(t)}\right)$. The covariance matrix $\mathbf{C_w}$ can be estimated from the training set using $\mathbf{w}(t) = \mathbf{b}(t+1) - \hat{\mathbf{F}}\mathbf{b}(t)$. 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:

$$\mathbf{x}(t) \overset{def}{=} \begin{bmatrix} \mathbf{b}(t-1) \\ \mathbf{b}(t) \end{bmatrix} \qquad (4.6)$$

with associated state equation:

$$\mathbf{x}(t+1) = \mathbf{F}\mathbf{x}(t) + \mathbf{w}(t) \qquad \text{with} \qquad \mathbf{F} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ \mathbf{F}_1 & \mathbf{F}_2 \end{bmatrix} \qquad (4.7)$$

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 (4.7) models D decoupled second-order autoregressive (AR) models, one for each coefficient $b_n(t)$ in $\mathbf{b}(t)$, i.e.,

$$b_n(t+1) = \alpha_n b_n(t) + \beta_n b_n(t-1) + w_n(t) \qquad \text{with } n = 1, \cdots, D \qquad (4.8)$$

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 characterized by their natural frequencies f_n and relative damping ζ_n given by:

$$f_n = \frac{\sqrt{1-\alpha_n-\beta_n}}{2\pi T} \quad \text{and} \quad \zeta_n = \frac{-\alpha_n-2}{2\sqrt{1-\alpha_n-\beta_n}} \qquad (4.9)$$

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.

4

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

$$\mathbf{C_w} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{22} \end{bmatrix} \tag{4.10}$$

$\mathbf{C}_{22}$ 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.

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 $2D$. In practice, D, being the result of the PCA, is much smaller than the dimension of the measurement vector, $\underline{\mathbf{g}}(t)$, which is 152. Therefore, the Kalman filter was used in the following form:

$$\begin{aligned} \left. \begin{aligned} \hat{\mathbf{x}}(t\,|\,t-1) &= \mathbf{F}\hat{\mathbf{x}}(t-1\,|\,t-1) \\ \mathbf{C}(t\,|\,t-1) &= \mathbf{FC}(t-1\,|\,t-1)\mathbf{F}^T + \mathbf{C}_w \end{aligned} \right\} &\text{ prediction} \\[2ex] \left. \begin{aligned} \mathbf{C}(t\,|\,t) &= \left(\mathbf{C}^{-1}(t\,|\,t-1) + \mathbf{H}^T\mathbf{C}_v^{-1}\mathbf{H} \right)^{-1} \\ \hat{\mathbf{x}}(t\,|\,t) &= \mathbf{C}(t\,|\,t)\left(\mathbf{C}^{-1}(t\,|\,t-1)\hat{\mathbf{x}}(t\,|\,t-1) + \mathbf{H}^T\mathbf{C}_v^{-1}\underline{\mathbf{g}}(t) \right) \end{aligned} \right\} &\text{ updating} \end{aligned} \tag{4.11}$$

This is computationally more efficient than the typical form. In equation (4.11), $\mathbf{H}$ 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.

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 regularization 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 regularization parameter c_w. That is, the submatrix $\mathbf{C}_{22}$, was replaced by the matrix $(1-c_w)\mathbf{C}_{22} + c_w diag(\overline{\mathbf{C}_{22}(\ell,\ell)})$. Here, $diag(\overline{\mathbf{C}_{22}(\ell,\ell)})$ 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 gave 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 optimized using training data.

We performed cross-validation for training and testing. The procedure is depicted in Figure 4.2. It was applied per volunteer, per feature type, and per window size. Data from the various instructions were pooled by concatenating the data: $\mathbf{g}_f(t,r) = \begin{bmatrix} \mathbf{g}_f(t,1,r) & \cdots & \mathbf{g}_f(t,19,r) \end{bmatrix}$. The data from four repetitions were pooled to get the training data: $\mathbf{g}_f(t) = \begin{bmatrix} \mathbf{g}_f(t,1) & \cdots & \mathbf{g}_f(t,4) \end{bmatrix}$. 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 minimization of the RMS error by varying these parameters one by one and applying successive parabolic optimization.
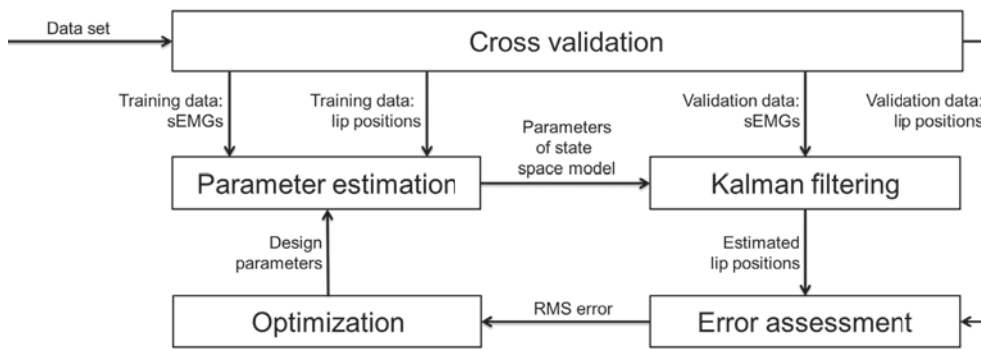

Figure 4.2 | The optimization and evaluation of design parameters using cross validation.

67

RESULTS

The best results for the static, first-order, and second-order state space models are summarized in Table 4.4 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. Four subjects showed the best results when the WAMP feature was used. The threshold 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 a threshold of 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.

Table 4.4 | 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 (20mV)	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

The RMS errors for the different features for the first and second order system models are presented in Table 4.5. It can be seen that for both tested threshold values, the WAMP feature performs comparable. Regarding the other features, the WL performs slightly worse compared to the WAMP, and RMS and MAV showed the poorest results.

Table 4.5 | Optimal settings averaged over the volunteers obtained per feature and system order.

	RMS		MAV		WL		WAMP 10 mV		WAMP 20 mV	
System order	1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd
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

The influence of the different parameters and window on the error in the second-order system can be seen in Figure 4.3. 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 damping correction c_d had the minimum error at a factor of 0.7. The constant c_w showed little influence but had an optimum on average at 0.2. The window lengths show a similar trend in all subjects, with the best results for medium length windows. For the first-order system comparable values were found.

DISCUSSION

For the prediction of the functional and aesthetic consequences of therapy 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 muscular activation signals, needed for the biomechanical models, is difficult.[22] As an alternative, 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 patient-specific information. As far as we know, we are the first who expresses 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. 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- and second-order system is neglectable. A higher-order system has more parameters which has 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
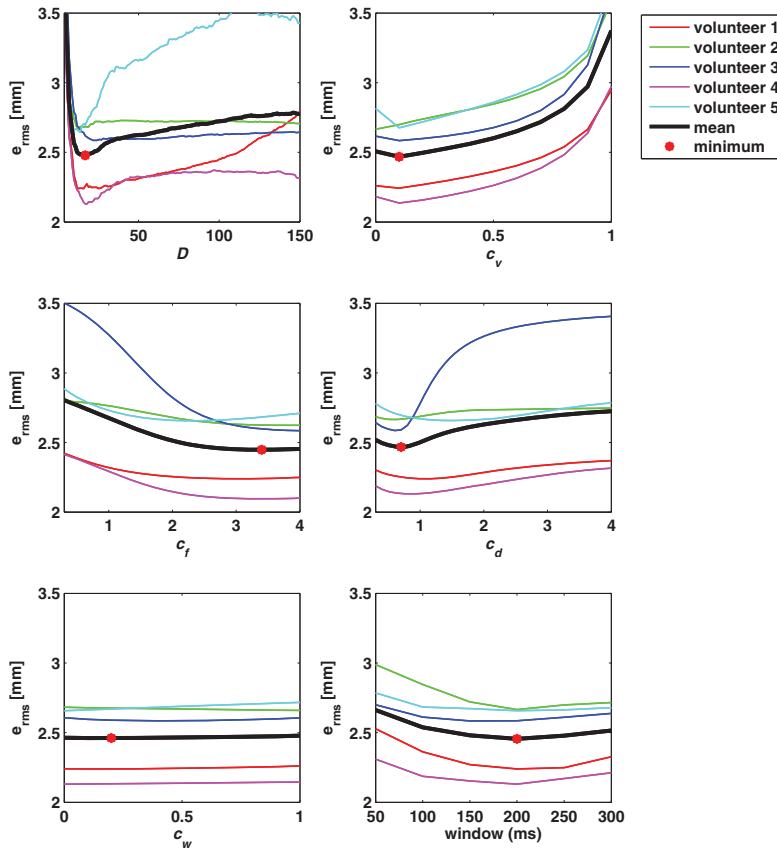
4

Figure 4.3 | The dependence of the RMS error on the various parameters and window in the second-order system.

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 outweighed the errors induced by the deviations in the estimated parameters.

The fudging parameters were used to optimize the model per volunteer, hence make it volunteer specific. Only the regularization 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 optimization. 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. Lastly 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 feature with threshold 10 and 20 mV were most promising. Perhaps that thresholds in-between these values could perform better. One can also think of optimizing the WAMP threshold per muscle channel for optimal results. The widely used RMS feature performed worse. This was also found by Phinyomark.[20]

Table 4.6 | Differences in experiment setup in related studies.

	Current study	Honda et al.	Lucero & Munhall
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)	A sentence
Volunteers	5	1	1
Error assessment	RMS distance	Visual	Visual and cross-correlations

Because of different error assessments, the differences of our model with those found in literature, will be discussed qualitatively. The main differences are stated in Table 4.6. 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 & Munhall controlled jaw movement using an optical tracked marker instead of EMG signals. Non-surprisingly the modelled facial tissue followed this movement well. We added EMG measurements of the digastric muscle to make the prediction of jaw movement possible. Honda and Lucero & Munhall measured EMG signals on one side of the face, disregarding asymmetry in facial

4

morphology and lip movement as described by Campbell.[23] The use of hooked-wire or needle electrodes is attractive to overcome the problem of crosstalk, but for clinical applications this will be impractical because they are invasive and thus patient-unfriendly. Therefore, in generating future personalized 3D models controlled by EMG signals, sEMG will have our preference, being easy applicable and patient-friendly. Furthermore, we tested our model on five subjects, indicating it is general applicable.

A limitation of the proposed setup 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.[24] 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 cocontraction. Also volunteers can use different muscle activation patterns to perform the same subject instruction. We saw similar results when facial expressions were performed as described by Schumann et al. [25] Most volunteers were able to selectively activate the LLS, whereas most volunteers had difficulty in pulling their lip corners down. Purse lips, pout lips, and voluntary smiling all induced multiple muscle activations, showing the difficulty in selective muscle activation of facial muscles.

Finally, the ultimate goal is to achieve objective functional predictions in oral cancer treatment. Therefore the integration of patient-specific anatomical information and tissue parameters are required for simulating treatment effects. The current model does not account for those aspects. A biomechanical model would be able to incorporate this physical relationship, and adjust them according to the performed treatment. Therefore a biomechanical model could be able to simulate functional complications. For patient-specific function prediction, the activation signals of the musculature in a biomechanical model should be similar to the actual neural control by the patient. EMG provides an indication of the neural control, but dissimilarity between EMG and the activation signals will occur. An EMG feature of a muscle channel cannot be transferred directly as an activation signal for a muscle group in a biomechanical model, because of crosstalk, misplacement of electrodes, a non-linear relationship, and other artefacts. We showed that using a statistical model the relationship between 3D lip motion and EMG can be described. Inverse modelling may also be possible with our approach. This can be used to detect misplacements of electrodes, in that case the EMG signal of the misplaced electrode will show unexpected results for specific movements.

Also the EMG signal can be smoothened and artefacts can be removed, due to dimension reduction of PCA which will cause the model not to react on outliers. This will be a useful tool to solve existing problems regarding inverse modelling in biomechanical models, because the estimated EMG signals will reduce the solution space drastically and make the activation signals patient specific. In future studies, it can be evaluated if the relationship between activation signals and sEMG can be described more clearly, using the mutual information theory for instance.

CONCLUSION

This study presented a next step towards the personalization of the functional outcome assessment after treatment of oral cancer. The two dynamic modelling methods proved a continuum of 3D lip positions can be predicted based on person-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. In future studies this method can be used to solve problems concerning inverse modelling in biomechanical models, by reduction of the solution space and including patient-specific information.

CONFLICT OF INTEREST

ACKNOWLEDGEMENTS

FUNDING

ETHICAL APPROVAL

Although this research involved healthy human volunteers, no ethical approval was required. Because it did not infringe the (psychological) integrity of the subjects. Measurements were non-invasive and not stressful. The study was performed within the Dutch legislation

4

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 Societies). Oral information and consent is obtained.

REFERENCES

[1]   R. Siegel, D. Naishadham, and A. Jemal, "Cancer Statistics , 2012," *CA Cancer J Clin*, vol. 62, no. 1, pp. 10–29, 2012.

[2]   J. P. Shah and Z. Gil, "Current concepts in management of oral cancer-surgery," *Oral Oncol.*, vol. 45, pp. 394–401, 2009.

[3]   A. Kreeft, I. B. Tan, M. W. M. van den Brekel, F. J. Hilgers, and A. J. M. Balm, "The surgical dilemma of 'functional inoperability' in oral and oropharyngeal cancer: current consensus on operability with regard to functional results," *Clin. Otolaryngol.*, vol. 34, no. 2, pp. 140–146, 2009.

[4]   A. M. Kreeft, I. B. Tan, C. R. Leemans, and A. J. M. Balm, "The surgical dilemma in advanced oral and oropharyngeal cancer: how we do it.," *Clin. Otolaryngol.*, vol. 36, no. 3, pp. 260–266, Jun. 2011.

[5]   M. J. A. van Alphen, A. M. Kreeft, F. van der Heijden, L. E. Smeele, and A. J. M. Balm, "Towards virtual surgery in oral cancer to predict postoperative oral functions preoperatively," *Br. J. Oral Maxillofac. Surg.*, vol. 51, no. 8, pp. 747–751, 2013.

[6]   I. Stavness, B. Gick, D. Derrick, and S. Fels, "Biomechanical modeling of English /r/ variants.," *J. Acoust. Soc. Am.*, vol. 131, no. 5, pp. EL355–360, May 2012.

[7]   I. Stavness, M. A. Nazari, P. Perrier, D. Demolin, and Y. Payan, "A Biomechanical Modeling Study of the Effects of the Orbicularis Oris Muscle and Jaw Posture on Lip Shape," *J. Speech, Lang. Hear. Res.*, vol. 56, no. 3, pp. 878–890, 2013.

[8]   J. P. V. Pelteret and B. D. Reddy, "Development of a computational biomechanical model of the human upper-airway soft-tissues toward simulating obstructive sleep apnea," *Clin. Anat.*, vol. 27, no. 2, pp. 182–200, 2014.

[9]   F. Vogt, J. E. Lloyd, and P. Perrier, "Efficient 3D Finite Element Modeling of a Muscle-Activated Tongue," no. Figure 1, pp. 19–28, 2006.

[10]   M. Hamedi, S. H. Salleh, and T. T. Swee, "Surface electromyography-based facial expression recognition in Bi-polar configuration," *J. Comput. Sci.*, vol. 7, no. 9, p. 1407, 2011.

[11]   S. Arjunan, D. Kumar, W. Yau, and H. Weghorn, "Unvoiced speech control based on vowels detected by facial surface electromyogram," in *e-Society 2006, IADIS Proceedings*, 2006, pp. 381 – 388.

[12]   B. J. Betts, K. Binsted, and C. Jorgensen, "Small-vocabulary speech recognition using surface electromyography," *Interact. Comput.*, vol. 18, no. 6, pp. 1242–1259, 2006.

[13]   K. Honda, T. Kurita, Y. Kakita, and S. Maeda, "Physiology of the lips and modeling of lip gestures," *J. Phon.*, pp. 243–254, 1995.

[14]   J. C. Lucero and K. G. Munhall, "A model of facial biomechanics for speech production," *J. Acoust. Soc. Am.*, vol. 106, no. 5, pp. 2834–2842, 1999.

[15]   I. Stavness, J. E. Lloyd, and S. Fels, "Automatic prediction of tongue muscle activations using a finite element model," *J. Biomech.*, vol. 45, no. 16, pp. 2841–2848, 2012.

4

[16] B. Lapatki, D. Stegeman, and M. Zwarts, "Selective contractions of individual facial muscle subcomponents monitored and trained with high-density surface EMG," in *The Facial Palsies. Complementary Approaches*, C. Beurskens, R. van Gelder, P. Heymans, J. Manni, and J. Nicolai, Eds. Utrecht, 2005, pp. 89–108.

[17] M. Schünke, E. Schulte, and U. Schumacher, "Head," in *Thieme atlas of anatomy: head and neuroanatomy*, 1st ed., L. M. Ross, E. D. Lamperti, and E. Taub, Eds. New York: Thieme, 2010, p. 44.

[18] N. van Son, T. M. I. Huiskamp, A. J. Bosman, and G. F. Smoorenburg, "Viseme classifications of Dutch consonants and vowels," *J. Acoust. Soc. Am.*, vol. 96, no. 3, pp. 1341–1355, 1994.

[19] A. van Boxtel, "Optimal signal bandwidth for the recording of surface EMG activity of facial, jaw, oral, and neck muscles.," *Psychophysiology*, vol. 38, no. 1, pp. 22–34, 2001.

[20] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, "Feature reduction and selection for EMG signal classification," *Expert Syst. Appl.*, vol. 39, no. 8, pp. 7420–7431, Jun. 2012.

[21] E. Vatikiotis-Bateson, K. G. Munhall, M. Hirayama, Y. V. Lee, and D. Terzopoulos, "The Dynamics of Audiovisual Behavior in Speech," in *Speechreading by Humans and Machines*, vol. 150, D. G. Stork and M. E. Hennecke, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1996, pp. 221–232.

[22] X. Wu, J. Dang, and I. Stavness, "Iterative method to estimate muscle activation with a physiological articulatory model," *Acoust. Sci. Technol.*, vol. 35, no. 4, pp. 201–212, 2014.

[23] R. Campbell, "Asymmetries in moving faces.," *Br. J. Psychol.*, vol. 73, no. 1, pp. 95–103, 1982.

[24] G. Kamen and D. Gabriel, "EMG instrumentation," in *Essentials of Electromyography*, 1st ed., L. D. Robertson, E. H. Mustain, K. Maurer, E. Evans, S. Calderwood, and J. Sexton, Eds. Champaign, Illinois: Human kinetics, 2009, p. 66.

[25] N. P. Schumann, K. Bongers, O. Guntinas-Lichius, and H. C. Scholle, "Facial muscle activation patterns in healthy male humans: a multi-channel surface EMG study.," *J. Neurosci. Methods*, vol. 187, no. 1, pp. 120–8, Mar. 2010.

Chapter 5

In vivo intraoperative hypoglossal nerve stimulation for quantitative tongue motion analysis

M.J.A. van Alphen

M. Eskes

L.E. Smeele

A.J.M. Balm

F. van der Heijden

ABSTRACT

This is the first study quantitatively measuring tongue motion in 3D after in vivo intraoperative neurostimulation of the hypoglossal nerve and its branches during a neck dissection procedure. Firstly this study is performed to show whether this setup is suitable for innervating different muscles or muscle groups with an identifiable corresponding motion pattern by stimulating the main stem and the visible branches and by performing an intra-patient comparison over the captured 3D trajectories. Secondly an inter-patient comparison is performed to analyse if similar branches lead to comparable movements of the tongue tip. Our results showed the measurement setup works, and that we are able to capture distinguishable trajectories for different branches. The inter-patient comparison showed a poor match in trajectories for similar branches, which could be an indication of the presence of several anatomical variations between individuals.

INTRODUCTION

Choices between treatment options for tongue carcinoma might be difficult, particularly if decisions on whether or not to operate are dependent on a critical judgment of functional inoperability of the disease, i.e. unacceptable function loss after surgery.[1] In advanced tongue cancer, the choice between surgery and organ sparing chemoradiation is currently based on the personal expertise of the treating physician. It is made on subjective expectations regarding functional loss post-surgery, i.e. expected disturbance of speech, swallowing, and mastication. Functional outcomes post-surgery are patient specific and differ significantly among individuals. The development of a patient specific biomechanical model of the oral cavity is proposed, allowing the prediction of functional consequences after treatment of tongue carcinoma.[2–4] These models should enable personalized simulation of the effects of the surgery on important functions for the quality of life like speech and swallowing [5], and the subsequent animation and visualization thereof.

Although several biomechanical models of the tongue are currently available [3, 6–10], they are still generic and not patient specific. Knowledge regarding inter-patient variability of the branching of the hypoglossal nerve and distribution patterns is lacking. This understanding is needed for patient specific prediction of function loss and basic knowledge on muscle activation patterns is a prerequisite. A tongue movement can be defined as a coordinated sequence of muscle activations contributing to, for instance, efficient swallowing and speech. Currently, the most popular method to determine these activation patterns is inverse dynamic modelling: finding the activation pattern is formulated as an optimization problem in which a model-predicted motion is fitted to an observed motion by adjusting hypothesized activation signals. This method harbours the risk of a low specificity, since the solution space is not unique and the optimization process can easily get stuck in a local optimum. Another problem is the lack of quantitative validation of patient specific motions after activation of individual muscle groups.

Individualized activation patterns can be improved by additional physiological measurements that represent the muscular activations. The hypoglossal nerve is pivotal for motoric innervation of the tongue muscles. Its anatomy is dense and complex [11]: the nerve bifurcates multiple times into finer branches distributed over the tongue muscles. Knowledge of this branching and distribution pattern is crucial for model based prediction of functional loss. Firstly, if a branch is sacrificed during surgery, the associated muscle or muscle group lose completely or partly (depending on possible existing anastomosis) their ability to contract voluntarily. Secondly, knowledge of the distribution pattern might lead to more parsimonious activation models with a more restricted solution space of the inverse dynamic modelling. The branching and distribution pattern of the hypoglossal nerve is only globally known. In vivo imaging techniques enabling patient specific visualization and 3D localization of neural structures inside the tongue are not available, as yet. A first step to

5

unravel the innervation of the various tongue muscles might therefore be the application of intra-operative neurostimulation.

In this study we present a unique and controlled in vivo setup that enables tracking of 3D tongue motions that are innervated by stimulation of branches of the hypoglossal nerve which are exposed during a neck dissection. The first aim of this study is to show that our in vivo measurement setup is suitable for innervating different muscles or muscle groups with an identifiable corresponding motion pattern. This is performed by an intra-patient comparison of the 3D trajectories of the tongue tip. Secondly we compare those motion patterns between individual patients to see if corresponding branches innervate similar muscles.

MATERIALS AND METHODS

A modified or radical neck dissection is a surgical procedure with exposure of the hypoglossal nerve and its branches. This procedure is ideal for an experimental setup that allows neural stimulation of selected branches and simultaneous recording of the resulting motion of the tongue. See Figure 5.1 and Figure 5.2.

We developed a method to optically track well-defined locations on the tongue in 3D. As the tongue has no anatomical easy recognizable positions, markers are drawn on the tongue using a skin marker (Figure 5.2). A triple-camera setup enables 3D tracking of the markers. The three cameras, placed in a case with four LED lights, are attached to an arm with hinges, so it can be positioned to visualize the tongue properly. The colour cameras (Basler avA1000-100gc) have a frame rate of 100 frames per second (fps) and a resolution of 1024x1024 pixels. A preliminary experiment, with voluntary movements, showed that rapid tongue movements have frequency components up to about 15 Hz (measured 40 dB below maximum) [12]. Thus, 100 fps is able to track tongue motion with negligible aliasing errors. In order to improve the view on the tongue, a lip retractor was placed (see Figure 5.2). If necessary, also a jaw spreader was used to improve the intraoral view. A surgical drape separated the operation field from the camera setup (Figure 5.2).

Usually, calibration is done with multiple images of a chess board [13], but due to limited space and time in the operation room, this method is not practicable in the current setup. Instead, a wired frame of a $10\times10\times10$ cm cube, with 27 markers on well-defined locations was placed in front of the camera's. Using this wired frame, the system could be calibrated using only one image per camera, just before the surgery started. Validated by leave-one-out method applied to the 27 markers, the accuracy of the 3D estimated positions, quantified by its root mean square, was 0.7 mm. The $10\times10\times10$ cm measurement space appeared to be large enough to cover the area of the tongue movement.

In order to validate this experimental setup, ten patients were included (6 males and 4 females; age 51-89 years) without any present or past oral disease. All signed informed
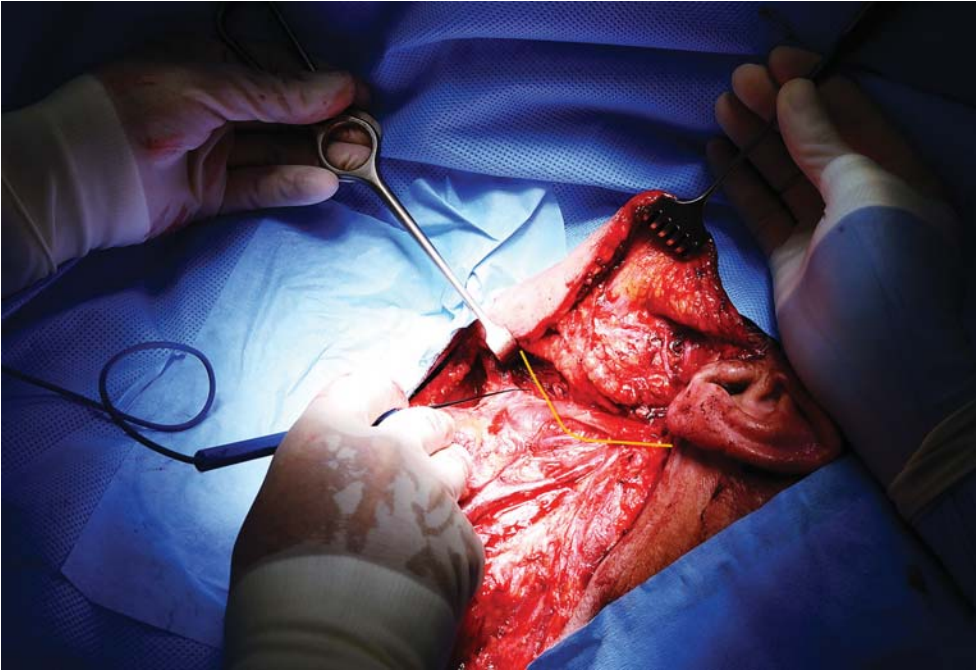
Figure 5.1 | Peroperative view of a neck dissection and superficial parotidectomy displaying the first branches of the hypoglossal nerve. The left hook is retracting the posterior border of the mylohyoid muscle to create a better view on the hypoglossal nerve branches. The yellow line, demarcates the course of the hypoglossal nerve.

5

consent. Patients were intubated nasally to prevent interference of the intubation tube with the tongue movements. This study was approved by the Ethical committee and board of directors of the Netherlands Cancer Institute.

After calibrating and positioning the cameras, the surgeon stimulated (4 Hz with 0.8 mA) the main stem and next the visible branches of the hypoglossal nerve using a NIM®-nerve stimulator (Medtronic) for several seconds. Simultaneously recordings were performed with the triple-camera setup. The surgeons drew the nerve branches schematically on a paper in a sterile environment and numbered them in the order they were stimulated. These drawings were used for inter-patient comparison. The number of stimulated branches was dependent on the local extension of surgery.

The captured videos were analysed offline. Tracking in 3D was based on a Kalman filter. A PCA based outlier detection algorithm allowed the tracking of markers even if a marker was occluded in one or more cameras.[12] For each patient (indexed by p with $p = 1, \cdots, P$), and for each branch (indexed by k with $k = 1, \cdots, K_p$; $k = 1$ is the main stem) the marker on the tip of the tongue was used for analysis, resulting in K_p different measurement sequences $\mathbf{X}_{p,k}$. That is, $\mathbf{X}_{p,k} = \begin{bmatrix} \mathbf{x}_{p,k,1}^T & \mathbf{x}_{p,k,2}^T & \cdots \end{bmatrix}^T$ where $\mathbf{x}_{p,k,m}$ is a 3D point at time frame m. K_p can

Figure 5.2 | Surgical drape separates the non-sterile environment from the operation field. The triple-camera setup is positioned in front of the mouth opening. The tongue markers are made visible by a lip retractor.

differ between patients, because it depends on the amount of branches the surgeon could find during surgery. In each sequence $\mathbf{X}_{p,k}$, a number of periods of a motion pattern was visible due to the periodic stimulation. Those patterns should also be periodic as the motion was caused by a muscle or muscle group innervated by the same branch. The periods are enumerated by $n = 1, \cdots, N_{p,k}$. A motion pattern of a single period was represented by a trajectory consisting of a series of 3D coordinates $\mathbf{Z}_{p,k,n} = \left[\mathbf{z}_{p,k,n,1}^{T} \cdots \mathbf{z}_{p,k,n,M}^{T} \right]^{T}$, where $\mathbf{z}_{p,k,n,m}$ represented the 3D coordinates in frame m. Each trajectory holds $M = 25$ frames, because the cameras captured videos at 100 fps, and we stimulated at 4Hz.

Since the neural stimulation was not synchronized with the cameras the time frame m in which the stimulation started, and the motion pattern was initiated, could differ between the K_p different measurement sequences. For a proper comparison of the trajectories of different branches, each sequence should be synchronised by time-shifting it such that the start of the stimulation pulses occur in all sequences at the same time frame. The first branch

(main stem) was used as a reference. For each other branch, $k = 2, \cdots, K_{p'}$ the sequence $\mathbf{X}_{p,k}(s)$ was time-shifted with shifts ranging from $s = -4$ to $s = 4$, yielding nine shifted versions $\mathbf{X}_{p,k}(s)$. Next, the inter-intra distances, $J_p(1,k)$, a distance measure between two datasets, in this case $\mathbf{X}_{p,1}$ and $\mathbf{X}_{p,k}(s)$ (see below), were determined. The time-shifted version $\mathbf{X}_{p,k}(s_{\min})$, with $s_{\min}$ the lag that gave the minimum distance amongst the nine shifts was used in further analysis.

For the intra-patient comparison, also the inter-intra distance was used as a distance measure. For each patient, the distance was determined between all branches that showed motion at the tongue tip. To calculate the inter-intra distance between two sequences $\mathbf{X}_{p,k}$ and $\mathbf{X}_{p,\ell}$, the corresponding trajectories within each sequence, i.e. $\mathbf{Z}_{p,k,n}$ and $\mathbf{Z}_{p,\ell,n'}$ were regarded as random samples in a $3M$-dimensional space. For both sequences, the averages, $\hat{\boldsymbol{\mu}}_k$ and $\hat{\boldsymbol{\mu}}_\ell$, over the $N_{p,k}$ and $N_{p,\ell}$ trajectories $\mathbf{Z}_{p,k,n}$ and $\mathbf{Z}_{p,\ell,n}$ were calculated. Next, all the trajectories were projected on the $1D$ subspace spanned by the difference $\mathbf{m} = \hat{\boldsymbol{\mu}}_k - \hat{\boldsymbol{\mu}}_\ell$.

$$y_{p,k,n} = \mathbf{Z}_{p,k,n} \frac{\mathbf{m}}{\|\mathbf{m}\|} \quad \text{and} \quad y_{p,\ell,n} = \mathbf{Z}_{p,\ell,n} \frac{\mathbf{m}}{\|\mathbf{m}\|} \tag{5.1}$$

Then the within scatter (S_w) was determined,

5

$$S_w = \frac{N_k \hat{\sigma}_k^2 + N_\ell \hat{\sigma}_\ell^2}{N_k + N_\ell} \tag{5.2}$$

where $\hat{\sigma}_k^2$ and $\hat{\sigma}_\ell^2$ are the variances of $y_{p,k,n}$ and $y_{p,\ell,n'}$ respectively. The inter distance was defined as the squared distance $\|\mathbf{m}\|^2$ between the two averages of the two datasets, and, finally, the inter-intra distance was defined as the ratio between inter distance and within scatter:

$$J_p(k,\ell) = \frac{\|\mathbf{m}\|^2}{S_w} \tag{5.3}$$

Before calculating this distance measure, first offset neutralization was applied to each trajectory by translation of the trajectory such that the first point of the trajectory becomes the origin, $\mathbf{x}_{p,k,m} := \mathbf{x}_{p,k,m} - \mathbf{x}_{p,k,1}$. By this, the coordinates of the first point of a trajectory is always zero, and therefore this point could be deleted from the trajectory reducing the number of points in a trajectory to $M = 24$, and the dimension of $\mathbf{Z}_{p,k,n}$ becomes $3 \times 24 = 72$. The inter-intra distances were calculated for all combinations of pairs (k, ℓ) with $k \neq \ell$ of a patient.

To visualize the clustering of trajectories per branch, multidimensional scaling (MDS) mapping was performed. MDS is a nonlinear mapping from a high dimensional space (in our case $3M$-dimensional) to a 2D or 3D space in such a way that distances between vectors in both spaces are preserved as much as possible.[14] We applied MDS per patient to all trajectories of all sequences, i.e. all stimulated branches. The Euclidian distance between any pair of such trajectories was the input of the MDS. Per patient, this resulted in a 2D map of vectors visualizing the intra distance of trajectories and the inter distances between branches.

The position of a patient's head, relative to the calibration cube, varied from patient to patient. Therefore, to compare the inter-patient variability, the coordinate systems of two different patients should be registered first, and the trajectories should be aligned accordingly. For the alignment of the origin we shifted all trajectories such that the first point of the trajectory became the new origin (as was done before).

For the assessment and alignment of the orientation, we used the second order moment matrices:

$$\mathbf{M}_{p,k} = \frac{1}{N_{p,k}} \sum_{n=1}^{N_{p,k}} \mathbf{Z}_{p,k,n} \mathbf{Z}_{p,k,n}^{T} \qquad\qquad \textbf{(5.4)}$$

and especially the one belonging to the main stems $(k = 1)$ was used: $\mathbf{M}_{p,1}$. It was assumed that the orientation of the trajectories of the main stem would be similar between patients since all muscles innervated by the hypoglossal nerve were activated. The orientation of each patient was normalized by rotating the coordinate system such that eigenvectors, ordered descending according to the corresponding eigenvalues, became the new x-, y-, and z-direction. If needed, the direction of an eigenvector was flipped to assure a right handed system with comparable trajectories.

To enable comparison between the lengths of the trajectories, also a scale parameter for each patient was calculated based on the trace of the second order matrix:

$$\alpha_p = \sqrt{trace(\mathbf{M}_{p,1})} \qquad\qquad \textbf{(5.5)}$$

Since the lengths of the trajectories might be influenced by a patient dependent factor that is common for all stimulated branches, we normalized the trajectories by scaling them with the patient specific factor $1/\alpha_p$.

For the test on inter-patient variability, we considered each pair (p,q) of patients. If inter-patient variability is small, and in the ideal case, for each branch k from patient p, there will be a corresponding branch $\ell = L(k)$ from patient q. Since the surgeon could have overlooked a branch, this association would not always be available.

Two criteria were used to express the correspondence between trajectories: differences in orientations and differences in length. After normalization with respect to origin, orientation, and scale, the second order matrix $\mathbf{M}_{p,k}$ was calculated for each patient p and each branch k. Then the first principal direction from each $\mathbf{M}_{p,k}$ was calculated (being the eigenvector with largest eigenvalue). Finally, the angle $\varphi_{p,q,k,\ell}$ between principal directions from patient/branch (p,k) and patient/branch (q,ℓ) was used an indicator of correspondence. A second indicator was the relative difference in length between trajectories defined as $d_{p,k,q,\ell} = (\lambda_k - \lambda_\ell)/\sqrt{\lambda_k \lambda_\ell}$. The hypothesis is that small inter-patient variability leads to small angles and small differences in length between corresponding branches $\ell = L(k)$ relative to angles and differences in length between non-corresponding branches $\ell \neq L(k)$.

RESULTS

In ten cases, we observed well distinguishable trajectories in two or more different measurements. With these data, the intra- and inter-patient analysis was performed. In Table 5.1 the results for intra-patient comparison are showed. One patient underwent a bilateral neck dissection with an intraoperative stimulation of both hypoglossal nerves. Each side is analysed and treated as a separate patient (left patient 3 and right patient 4). The number of stimulated branches ranged from five to eight. In patient 9, all branches that were stimulated (seven in total) showed motion. The other patients had one to three branches which did not initiate motion. In six out of nine cases, where the first cranially curving branch was stimulated, no motion was seen. In three of those patients also the second cranially curving nerve did not induce movement. In patient 2, five branches were stimulated, of which three without motion at the tongue tip. Therefore, for the intra-patient comparison of this patient, only one branch could be compared with the motion pattern of the main stem. For the branches with a distinguishable motion, the results regarding the inter-intra distance (J) and maximum displacements of the averaged trajectories are summarized per patient in Table 5.1. The maximum displacements of the averaged trajectories was calculated for the different branches in the individual patients. The median of these maximal displacement over all branches and all patients was 2.8 mm. The minimum was 0.3 mm and the maximum 12.4 mm. Within patients the median of the maxima ranged from 1.0 to 6.9 mm.

A median inter-intra distance (J) between different branches within patients of 26.9 was found, which differed from 4.1 (patient 8) to 67.2 (patient 9) in individual patients. In patient 8, two branches gave resembling motion patterns with an inter-intra distance of 0.4. The maximal displacements were 1.3 and 1.2 mm, and the branches were adjacent of each other. This was the only occasion with such a small distance. For the other patients, the minimum inter-intra distance values ranged from 1.7 to 11.6, with a median of 4.3. The maximum distances ranged from 8.5 to 701.7, with a median of 145.8.

5

Table 5.1 | Number of stimulated branches, branches that showed motion (K_p) at the tongue tip, median, minimum and maximum inter-intra distance (J), and the median, minimum and maximum deflection at the tongue tip (mm) of the ten patients and the median values over the 10 patients. The median inter-intra distance and deflection over all patients is determined based on all inter-intra distances and trajectories.

Patient	Stim. branches	K_p	Median J	Min. J	Max. J	Median deflection (mm)	Min. deflection (mm)	Max. deflection (mm)
1	7	4	19.2	5.3	31.1	2.0	1.8	2.3
2	5	2	8.5	8.5	8.5	2.8	1.6	3.9
3	5	5	56.2	3.3	701.7	5.0	1.5	7.3
4	7	6	27.3	1.7	85.8	1.0	0.5	2.1
5	8	6	22.7	11.6	35.1	2.9	1.1	7.4
6	7	6	35.5	3.0	205.8	4.7	2.2	7.9
7	7	5	12.7	2.7	235.3	5.4	1.9	12.4
8	6	5	4.1	0.6	24.2	1.3	0.3	5.7
9	7	7	67.2	6.7	529.6	6.9	1.4	9.2
10	6	5	38.8	8.1	232.2	1.7	1.2	4.8
Median	7	6	26.9	4.3	145.8	2.8	1.5	6.5

The MDS-plots (Figure 5.3), representing the Euclidean distances between different trajectories of a patient, showed clear clusters of trajectories that resulted from the same branch. Patient 9 had the maximum median inter-intra distance of 107.7 amongst all other patients. Each branch forms a clear cluster without overlap from trajectories of other branches. Patient 8 had the minimum median inter-intra distance of 4.1. Each branch tends to cluster, but they also show some dispersion causing an overlap between clusters.

In Figure 5.4, the results of the inter-patient comparison are shown. The angles between branches that would match according to the surgeons drawings were not smaller than the angles between branches that did not match. The same is true for the relative length difference. When pairs of patients are compared this trend is seen in all combinations, except two. The comparison of patient 2, where only one branch could be compared, and patient 9 gave the lowest angle difference for comparable branches of 17.9°, where non matching branches showed an angle difference of 70.0°. The relative length differences were 2.0 and 0.5 for comparable and non-comparable branches respectively. And the comparison of patient 1 and 5 showed a 21.9° angle and 3.2 length differences for matching and 66.0° and 2.9 for non-matching branches.

DISCUSSION

This is the first study quantitatively measuring tongue motion after intraoperative neurostimulation of the hypoglossal nerve during a neck dissection procedure. After stimulation of different branches we can detect corresponding motion patterns, even if the
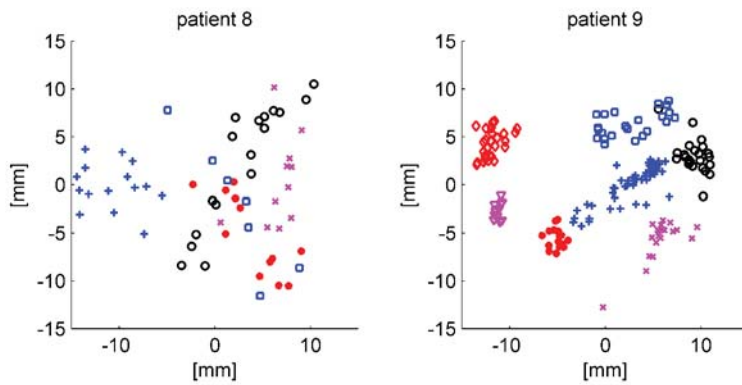
Figure 5.3 | MDS plots of patient 8 and 9. Patient 8, five branches showing motion at the tongue tip, showed the lowest median of the inter-intra distances and patient 9, seven branches showing motion, the highest. The different markers represent different stimulated branches and each individual maker represents a single trajectory.
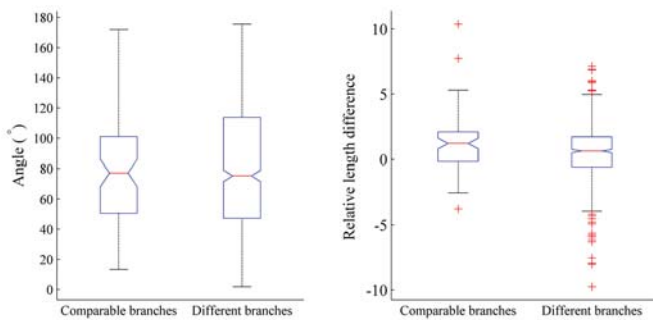

Figure 5.4 | Left: box plots showing the inter-patient angle (°) differences between trajectories of comparable and different branches. Right: box plots showing the inter-patient in the relative length difference between trajectories of comparable and different branches.

covered distance is small. The smallest covered distance of a measurement was 0.3 mm. The results of the intra-patient comparison showed that a good distinction could be established between motion patterns after stimulation of different hypoglossal nerve branches patterns (2-7 branches per patient with induced motion). The MDS-plots of nine out of ten patients showed clear clusters of the different trajectories which were well distinctive. Only the MDS-plot of patient 8 showed clusters, representing trajectories of different branches that overlapped for some trajectories. However, one has to take into account that a $2D$ visualization of $72D$ data (24 time samples times 3 coordinate values) could contribute

to this phenomenon. The median inter-intra distance of patient 8 is 4.1, implying that the clusters are still well separated.

In contrast, in the inter-patient study we were not able to discriminate similar trajectories, i.e. muscle groups, for comparable branches of individuals. This concerns not only the orientation but also the length of the trajectories. This observation could be explained by an inaccurate drawing and numbering of the branches or even by missing a neural branch during the intraoperative neurostimulation experiment. Also spread of current to nearby neural branches or muscles might contribute to a more diverse pattern. Nevertheless, the impact of these aforementioned negative factors cannot be so significant, since different orientations were found for each branch, meaning that the surgeon had a good overview over the region during the stimulations. Moreover, the used current was low (0.8 mA). It seems therefore rather plausible that anatomical variability is of greater importance in causing non-similar trajectories in the inter-patient experiments.

There are certain limitations to this study. First of all, the positioning of a patient's head is strongly dependent on the locoregional anatomical situation in combination with the required surgical view. Therefore the head is deviated sideways, and the tongue may slightly drop to one side due to gravity. Since this effect varies among patients, the inter-patient comparison is compromised. Also the intraoperative visualization of nerve branches is strongly dependent on the individual anatomical conditions and extent of neck surgery, and leads to different numbers of visible nerve branches. A good example of this issue is represented by the case of the bilateral neck dissection (listed as patients 3 and 4), where on the right side seven branches were exposed and on the left side only five. Another issue regarding our measurement setup is the fact that we only capture movements of the anterior part of the tongue surface. In eight patients one or two cranially curving branches were stimulated which did not show any movement at the locations of the placed markers. In six cases this was the first cranially curving branch, and in three cases the first two. It is assumed stimulation involves a muscle in the floor of mouth since a clear motion could be observed at that location, which was not propagated to the tongue tip.

A limitation regarding the intra-patient analysis is the calculation of the inter-intra distance between measurements. Currently the $72D$ data of each trajectory is projected on a $1D$ subspace, given by $\mathbf{m}$. Also the within scatter (S_w) is based on variances of the $1D$ representations of all the measured trajectories of the two compared branches. Because of the reduction in dimensionality, the trajectories of the branches can show overlap in $1D$, where in the $72D$ space probably less, or no, overlap would be present. Therefore, the inter-intra distance measure will underestimate the distinguishability of the trajectories of different branches, but if it performs well we know that two branches can really be separated properly. An alternative distance measure could be the Mahalanobis distance [14]. However, as the number of trajectories is much lower (< 20) than the $72D$

space, this could give an unrealistic good distinction between the trajectories of two different branches.

In future research we will focus on the transfer of the motion data to a finite element tongue model. The advantage of such a finite element model is the ability to include gravity as a parameter, which could compensate for the orientation of the patient's head. Such a model should be made patient specific, in terms of geometry and anatomy, before inverse modelling can be applied on our data, due to the relative small deflections of the tongue after stimulation of a nerve branch. Slight differences in, for instance in muscle locations, will lead to slightly different deflections in the model. For those small deflections this will result in a relative large error measure in the solution space, causing the inverse algorithm to find other results, e.g. by adding other muscles to compensate for the non-appropriate deflection or even the innervation of non-related muscles. As an extension on the current setup, simultaneous needle electromyography (EMG) measurements could be performed of different tongue muscles to analyse in more detail the distinct muscle innervations.

This setup gives opportunities for (a) quantitative validation of individualized biomechanical models; (b) experiments to study inter-patient variability concerning branching and distribution patterns; and (c) to come to a strategy to reduce the solution space for inverse dynamic modelling.

ACKNOWLEDGEMENTS

5

REFERENCES

[1]   A. M. Kreeft, I. B. Tan, C. R. Leemans, and A. J. M. Balm, "The surgical dilemma in advanced oral and oropharyngeal cancer: how we do it.," Clin. Otolaryngol., vol. 36, no. 3, pp. 260–266, Jun. 2011.

[2]   S. Fujita, J. Dang, N. Suzuki, and K. Honda, "A Computational Tongue Model and its Clinical Application," Oral Sci. Int., vol. 4, no. 2, pp. 97–109, 2007.

[3]   M. J. A. van Alphen, A. M. Kreeft, F. van der Heijden, L. E. Smeele, and A. J. M. Balm, "Towards virtual surgery in oral cancer to predict postoperative oral functions preoperatively," Br. J. Oral Maxillofac. Surg., vol. 51, no. 8, pp. 747–751, 2013.

[4]   S. Buchaillard, M. Brix, P. Perrier, and Y. Payan, "Simulations of the consequences of tongue surgery on tongue mobility: Implications for speech production in post-surgery conditions," Int. J. Med. Robot. Comput. Assist. Surg., vol. 3, no. 3, pp. 252–261, 2007.

[5]   A. C. Zuydam, D. Lowe, J. S. Brown, E. D. Vaughan, and S. N. Rogers, "Predictors of speech and swallowing function following primary surgery for oral and oropharyngeal cancer.," Clin. Otolaryngol., vol. 30, no. 5, pp. 428–437, 2005.

[6]   J. Dang and K. Honda, "Construction and control of a physiological articulatory model.," J. Acoust. Soc. Am., vol. 115, no. 2, pp. 853–870, 2004.

[7]   S. Buchaillard, P. Perrier, and Y. Payan, "A biomechanical model of cardinal vowel production: muscle activations and the impact of gravity on tongue positioning.," J. Acoust. Soc. Am., vol. 126, no. 4, pp. 2033–2051, 2009.

[8]   I. Stavness, J. E. Lloyd, and S. Fels, "Automatic prediction of tongue muscle activations using a finite element model," J. Biomech., vol. 45, no. 16, pp. 2841–2848, 2012.

[9]   Y. K. Wang, M. P. Nash, A. J. Pullan, J. A. Kieser, and O. Röhrle, "Model-based identification of motion sensor placement for tracking retraction and elongation of the tongue," Biomech. Model. Mechanobiol., vol. 12, no. 2, pp. 383–399, 2013.

[10] J. P. V. Pelteret and B. D. Reddy, "Development of a computational biomechanical model of the human upper-airway soft-tissues toward simulating obstructive sleep apnea," Clin. Anat., vol. 27, no. 2, pp. 182–200, 2014.

[11] L. Mu and I. Sanders, "Human tongue neuroanatomy: Nerve supply and motor endplates," Clin. Anat., vol. 23, no. May, pp. 777–791, 2010.

[12] T. A. G. Hageman, "Tracking of the tongue in three dimensions using a visual recording system," University of Twente, 2013.

[13] Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations," in Proceedings of the Seventh IEEE International Conference on Computer Vision, 1999, vol. 1, pp. 666–673.

[14] F. van der Heijden, R. P. W. Duin, D. de Ridder, and D. M. J. Tax, Classification, Parameter Estimation and State Estimation. Chichester, UK: John Wiley & Sons, Ltd, 2004.

Chapter 6

A new accurate 3D measurement tool to assess the range of motion of the tongue in oral cancer patients: a standardized model

S. van Dijk

M.J.A. van Alphen

I. Jacobi

L.E. Smeele

F. van der Heijden

A.J.M. Balm

ABSTRACT

Introduction Function loss as speech and swallowing problems in oral cancer treatment can be severe, due to reduction of lingual mobility. Until now, no accurate measurement tool for tongue mobility exists and preoperative prediction of function loss is based on expert opinion instead of evidence based insight.

Purpose To assess the reliability of a triple camera setup for the measurement of tongue range of motion (ROM) in healthy adults and its feasibility in patients with partial glossectomy.

Method A triple camera setup was used and 3D coordinates of the tongue in five standardized tongue positions were achieved in 15 healthy volunteers. Maximum distances between the tip of the tongue and the maxillary midline were calculated. Each participant was recorded twice and each movie was analysed three times by two separate raters. Intrarater, interrater and test-retest reliability were the main outcome measures. Secondly, feasibility of the method was tested in 10 patients treated for oral tongue carcinoma.

Results Intrarater, interrater and test-retest reliability all showed high correlation coefficients of >0.9 in both study groups. All healthy subjects showed perfect symmetrical tongue ROM. In patients, significant differences in lateral tongue movements were found, due to restricted tongue mobility after surgery.

Conclusions This triple camera setup is a reliable measurement tool to assess three-dimensional information of tongue ROM. It constitutes an accurate tool for objective grading of reduced tongue mobility after partial glossectomy.

INTRODUCTION

Oral cancer is the sixth most common cancer worldwide and the global incidence is around 275,000 per year.[1] In The Netherlands 1056 patients were diagnosed with oral cancer in 2012 with a five years cancer specific survival (adjusted for age related death; period 1989-2010) of 59%.[2]

Oral tongue carcinoma accounts for one third of all oral carcinomas and most frequently arises at the lateral border of the tongue. Surgical resection is the preferred treatment for stage T1 – T3 lesions, whereas larger lesions will be treated with chemoradiation if functionally irresectable.[3] Depending on the stage of the disease surgical removal may have serious impact on the mobility of the tongue. Reduction of this mobility has been proven to be one of the main causes of impaired function as speech and swallowing deteriorate and problems with oral food transport.[4–7] Furthermore, rehabilitation (e.g. logopedic training) is predominantly based on the improvement of the lingual mobility to regain effective oral function.[8] Although lingual movement and its range of motion (ROM) is of great importance, until now no accurate and reproducible measuring system exists. As a result preoperative prediction of reduced mobility and function loss is still matter of expert opinion and not based on scientific evidence.[9] If, however, we would be able to grade the reduced mobility and function loss, we would be able to: (a) assess each individual patient on functional inoperability; (b) inform the patient better on the expected function loss; (c) adjust rehabilitation therapy to the specific needs of the patient to improve swallowing and speech function.

The constant problem in the assessment of tongue mobility is the fact that the tongue is concealed in the oral cavity and therefore hard to follow during movements. Different methods have been tried to capture tongue movements during action. Cine MRI, video fluoroscopy and ultrasound are used methods but all have their limitations: poor temporal resolution, exposure to radiation, and velocity measurement instead of information about the actual ROM of the tongue, respectively.[10–12]

This study aims at the development of an objective and accurate grading scale to assess tongue ROM after partial glossectomy. This as a first step towards prediction of expected function loss for different stages of oral tongue carcinoma. We investigated the reliability of a triple camera setup for the measurement of tongue ROM with respect to intra-, interrater and test-retest reliability. Secondly, feasibility of the measuring system was tested in ten patients surgically treated for oral tongue carcinoma.

6

MATERIALS AND METHODS

Participants

First part of this study consists of 15 healthy volunteers who were recruited on institute premises. Participants had to be 18 years or older and were eligible if they were healthy and had no history of tongue carcinoma, no neurovascular diseases or other conditions that could influence tongue mobility.

Secondly, to assess feasibility of the method, we examined tongue ROM of 10 patients who had undergone a partial glossectomy. Surgery had to be at least six months ago and patients did not receive radiotherapy at the tongue.

Movements

All participants were asked to perform five movements: reach with tip of the tongue to the left side (I), reach with tip of the tongue to the right side (II), protrude the tongue (III), elevation: reach to the tip of the nose (IV) and depression: reach to the tip of the chin (V). Participants were asked to reach as far as they were able in all directions. Movements were performed twice: first as described above and secondly with a plastic widener in place to retract the lips. The widener was included in the first part of this study in the search to retain the most controlled and reliable measurement condition. Because of no additional value, this device was abandoned in the second part of this study when tongue ROM in patients was assessed.

Camera system and Measurements

Tongue movements were recorded with a standardised camera setup consisting of three cameras in a fixed position. This setup made it possible to reconstruct 3D coordinates based on known positions in the three video frames. The chosen cameras are the avA1000-100gc. Its specifications are: resolution 1024x1024 pixels, frame rate of 101 fps, sensor HAI-1050 ½ inch, connections gigE 12-pinconnector, size of 40.7x62x62 mm and weight 300g. A fixed focal 8.5 mm lens is mounted on these cameras in order to get a view of 25 cm in height (roughly the head size) when placing them 30 cm from the subject. Cameras were placed in 15 cm distance of each other. Lateral ones were placed in 20 degrees rotation towards the middle (Figure 6.1). The calibration parameters of the cameras were obtained with a calibration object consisting of 27 beads that are arranged on a 3x3x3 orthogonal grid.

For 3D information on tongue position two reference points were applied on the surface of the tongue and four additional markers on the face, used for correction of head movements, see Figure 6.2. Recorded movies were analysed offline. First the frames in which the maximal position of the tongue in the five different directions was achieved were selected. Next, in these frames, the four references markers and the tongue tip were selected manually.
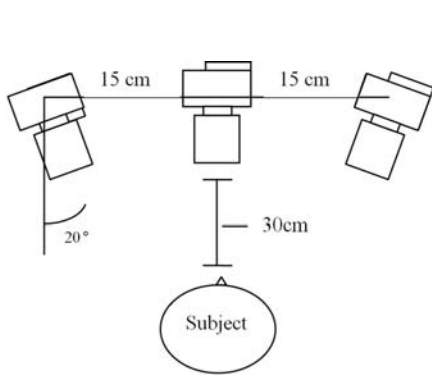
Figure 6.1 | Camera position setup.

Figure 6.2 | Subject with four face markers, and two markers on the tongue tip.

The 2D image coordinates of each marker visible in the frames of the three cameras were reconstructed to a 3D position. The Root Mean Square (RMS) error of this 3D localization was 0.73 mm estimated with a leave-one-out method. From the 3D positions of the markers, the distances between the interdental papilla of the maxillary midline and the tip of the tongue was calculated.

To determine intrarater variability, the five tongue distances of all participants were measured three times in a row by the same author (SvD). For interrater variability the second author (MvA) also measured all tongue movements of eight participants, three times in a row. The second author was blinded for the measurements made by the first author. Test-retest reliability was assessed by recording tongue ROM of all participants for a second time on another day. Because healthy adults were tested, symmetrical movements were expected. Thus, for the method to be valid, left and right distances should be equal and therefore differences in left-right distances were calculated.

The Medical Ethic Committee of the Netherlands Cancer Institute (NKI-AvL) did approve the study protocol and informed consent was received of all participants and patients.

Statistical analysis

All statistical analyses were performed using IBM SPSS statistics version 20.0. As a degree of reliability of the method, Intra Class Correlation coefficient (ICC) is computed and intrarater, interrater and test-retest ICCs are given. An ICC of >0.75 is considered as an excellent strength of agreement, a value of 0.4-0.75 between fair and good agreement and an ICC value of <0.4 as poor agreement. Differences of tongue movements between participants, as well as the influence of other factors (the presence of a widener and differences between days, gender or raters) on tongue movements were analysed with regression analysis. Wald chi-square p-values ≤0.05 are considered as a significant difference between groups.

6

RESULTS

We included 15 healthy participants and 10 surgically treated oral tongue cancer patients. All participants were able to perform all movements as instructed and were recorded successfully.

The group of healthy participants consisted of eight men (53%) and seven women (47%) with a mean age of 27 years (SD 2.4). Mean distances between the tip of the tongue and the interdental papilla of the maxillary incisors without mouth widener were: 36.1 mm (protrusion), 49.0 mm (left), 48.6 mm (right), 19.4 mm (elevation) and 62.3 mm (depression). With mouth widener distances were 36.3 mm, 44.5 mm, 44.5 mm, 17.1 mm and 55.3 mm for protrusion, left side, right side, elevation and depression respectively (Table 6.1).

Table 6.1 | Tongue range of motion in healthy participants in mm (SD).[a]

Side	Without widener	With widener	P-value[b]
Protrusion	36.1 (7.1)	36.3 (6.5)	0.137
Left	49.0 (6.9)	44.5 (5.6)	< 0.001
Right	48.6 (7.3)	44.5 (5.5)	< 0.001
Elevation	19.4 (5.2)	17.1 (5.0)	0.056
Depression	62.3 (10.9)	55.3 (7.3)	< 0.001

a.   Distance between tip of the tongue and maxillary incisor midline
b.   Significance value with paired T-Test between measurements with-without widener

Reliability

Intrarater reliability of the first author showed an ICC of 0.99 (95% CI: 0.99-0.99). The ICC of the second author was 0.99 (95% CI: 0.98-0.99). Measurements between the two separate raters showed an ICC of 0.95 (95% CI: 0.93-0.97) as a degree of interrater reliability. When measurements of day 1 and day 2 are compared an ICC of 0.93 (95% CI: 0.89-0.93) was found as a degree of test-retest reliability (Table 6.2).

Mouth Widener

All participants performed the five movements with and without mouth widener. There was a significant difference between the overall distances with and without the device in situ (p <0.001). When sides were analysed separately the difference was found for the two lateral movements 'left' and 'right', both p <0.001 and for 'depression' of the tongue (p <0.001). Elevation and protrusion were not influenced by the presence of the widener; p = 0.06 and p = 0.743 respectively. These results are summarized in Table 6.1 and Table 6.3.

Table 6.2 | Intra Class Correlations of measurement of tongue ROM with a 3D camera setup.

		ICC	95% Confidence Interval	
			Lower bound	Upper bound
Intrarater[a]	**1st author**	0.99	0.986	0.991
	2nd author	0.99	0.977	0.993
Interrater[b]		0.95	0.927	0.969
Test-retest[c]		0.93	0.889	0.952

a.   ICC3(A,k)
b.   ICC2(A,1)
c.   ICC3(A,k)

Table 6.3 | Influence of different factors on tongue range of motion.

		β	SE	Wald χ^2	df	p-value
Widener		3.32	0.65	26.40	1	<0.001*
Side				787.81	4	<0.001*
	Forward vs lateral	-10.49	1.12	88.00	1	<0.001*
	Left vs right	-0.39	0.61	0.409	1	0.522
Gender		1.29	1.96	0.435	1	0.509

* Significant influence or difference between measurements of range of tongue motion
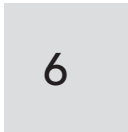
Symmetry of movement

Distances of the forward movement differed significantly of both lateral movements left and right (p <0.001). Distances toward left and right showed no significant difference (p = 0.486) (Table 6.3).

Gender

The healthy participants group consisted of 15 persons of which eight were men and seven were women. There was no significant difference between the range of tongue motion in men and women, ß = 1.29, χ^2(1) = 0.44, p = 0.51 (Table 6.3).

Inter-participant variability

Mean distances of tongue ROM of all 15 participants are described in Table 6.1. Measurements of participant number 2 were most near the mean values of the total group. Therefore, participant 2 was set up as comparison participant in regression analysis. This analysis showed that there was a non-significant difference with only one other participant (no. 2 and no. 3 with ß = 0.11, χ^2(1) = 0.20, p = 0.652). All other participants significantly differed in tongue ROM of subject 2 with P < 0.001.

6

Patients

To determine the feasibility of the method we included ten partial glossectomy patients in the study. The patient group consisted of five men (50%) and five women (50%) with a mean age at the time of resection of 65 years (SD 10.0). All patients underwent lateral tongue resection with primary closure. Five patients had stage I disease (50%), one patient stage II (10%) and four patients stage III (40%). Two patients underwent re-excision because of incomplete resection and one other patient, four months later, because of leukoplakia. Only one patient underwent selective neck dissection and received postoperative radiotherapy on the ipsilateral neck. Mean time between operation and assessment of tongue ROM was 29 months (SD 18.1).

Mean distances between the tip of the tongue and the interdental papilla of the maxillary incisors were: 35.9 mm (protrusion), 42.3 mm (left), 47.3 mm (right), 17.8 mm (elevation) and 52.1 mm (depression) (Table 6.4). These measurements did not differ significantly from the overall measurement of the healthy subjects (t (1134) = 1.86; p = 0.06, 95% CI: -0.1-4.96).

In healthy individuals, lateral movements did not differ significantly. The lateral movements of patients, divided in resection side versus contralateral side, were significantly different (p < 0.001) with distances of 47.8 mm and 42.2 mm respectively (Table 6.4).

When measurements per patient were analysed it is seen that, in general, the lateral movement towards the resected side is not altered after partial glossectomy.

Table 6.4 | Mean distances of tongue range of motion in partial glossectomy patients.[a]

Side	Mean distance in mm (SD)		Mean distance in mm (SD)	
Protrusion	35.9 (9.0)			
Left	42.3 (4.6)	Contralateral side[b]	42.2 (5.8)	p<0.001[c]
Right	47.3 (8.0)	Side of resection	47.5 (7.0)	
Elevation	17.8 (5.3)			
Depression	52.1 (13.2)			

a. Distance between tip of the tongue and maxillary incisor midline
b. 60% of the patients underwent glossectomy at the right side
c. p-value between resection side vs. contralateral side

DISCUSSION

This study shows a new and reliable technique for the assessment of the range of motion of the tongue with excellent intra-, interrater and test-retest reliability. Furthermore, it shows that there is a high degree of symmetry in the range of tongue motion in healthy adults. The used 3D camera setup seems a promising method to accurately measure tongue ROM. The detected altered tongue movements in patients treated for oral tongue carcinoma, is

the first step toward a tongue mobility grading system, followed by predicting functional consequences after tongue surgery.

In this study, an arbitrary study group of 15 participants was chosen. Furthermore, the study group consisted of participants with a mean age of 27 years which is lower than the average patient with oral cavity carcinoma.[13] Because this study aimed to test the reliability of a new method to assess tongue ROM, this difference seems not to be of importance. Also, tongue function as mean swallowing pressure, does not differ between younger and older age groups.[14]

Regression analysis in the participant group showed significant differences between tongue ROM in all directions of all but two subjects, so no standard or minimal ROM can be suggested from this study. Nevertheless, we showed that there definitely is a measurable difference in tongue ROM in patients with oral tongue carcinoma compared to healthy subjects. Especially the asymmetry of lateral movement in patients seems to be of importance. We hypothesize that patients with a smaller tongue ROM before treatment are more likely to develop larger functional problems compared to patients with a larger tongue ROM with the same tumour (or resection) volume and location. This might be one of the core concepts why patients with the same tumour characteristics evolve different speech and swallowing problems.[6]

To our knowledge only a few studies describe movements of the tongue in relation to function. Matsui et al. found that objective (as measured by a qualified speech pathologist) and naïve (as experienced by the patient) speech intelligibility are correlated with tongue tip elevation.[6] In this study tongue tip elevation was measured on a 3-point elevation score: 0: no elevation possible; 1: can elevate but not contact the palate; 2: can reach the palate. Less tongue mobility was significantly correlated with lower speech functioning scores. The same correlation was found in the study of Bressmann et al. where nine different tongue movements were tested on a 3-point scale (1:normal, 2:mild impairment, 3:marked impairment).[5] In both studies no absolute measurements of the tongue movements were performed and therefore it gives only a broad idea about the tongue mobility and subtle differences cannot be measured. Lazarus et al. developed a tongue ROM-score (combination of protrusion, lateralization and elevation score) with the use of the Therabite measuring disc.[7] They found statistical significant correlation between ROM-scores and eating and oral transport scores, with lower ROM indicating larger functional deficits. Although the measuring disc was used and more detailed information was assessed, in the analysis the patients were divided in subgroups of 'Normal', 'Mild-moderately impaired', 'Severely impaired' and 'Totally impaired' movement. As our study shows, no 'normal' ROM seems appropriate because a large variation in distances (in all directions) between healthy subjects exists. Also no additional data on the test-retest or rater reliability of the measuring disc are delivered.

6

To assess feasibility of the 3D camera setting in measuring tongue ROM we analysed five different tongue movements: elevation, depression, protrusion and both lateral directions. These movements were chosen based on international literature. Matsui et al. used elevation of the tongue's tip and dorsum as a mobility test and Bressmann et al. used nine different movements for a combined mobility score and no sub analysis of the different movements are described.[5,6] In a third study, of Konstantinović and Dimić, tongue's mobility was also measured, but correlation with functional aspects was not investigated.[4] They used a 3-points scale and six different movements were tested. This study showed that lateral movements significantly differed between patient groups with different reconstructions and therefore we included lateral movements next to protrusion. Elevation and depression of the tongue body are of great functional importance as it reflects the 'finger function' of the tongue that facilitates manipulation and shaping of the food bolus and clearance of food particles from the labial and buccal sulcus; as described in Chepeha et al.[15] Future research should focus on correlation of altered ROM and functional consequences. We measured the maximal possible movement of the tongue in all directions, because reliability of the method was the first incentive of the study. Question is if, or to what extent, these maximal possible positions contribute to the swallow and speech function. Future research should focus on this aspect. Nevertheless, detailed information on tongue positions can accurately be acquired with the triple camera setup.

In the search to retain the most controlled and thus reliable measurement condition we added a mouth widener in the test array. Results show that without the mouth widener in place, good and reliable measurements of the range of tongue motion can be made as well. Also, the presence of a mouth widener significantly restricted tongue ROM for the movements 'left', 'right' and depression and therefore should be excluded in future experiments.

This study shows that our 3D camera setup is a new reliable method to assess the range of motion of the tongue. Future research should focus on confirmation of our findings in larger patients groups treated for oral tongue carcinoma to develop an accurate grading system of altered ROM in patients with oral tongue carcinoma. Correlation of this grading system with functional outcomes as speech and swallowing deficits will assist head and neck surgeons in informing patients better than they can nowadays with respect to expected function loss.

CONCLUSION

This 3D camera setup is a reliable method for measuring tongue ROM in healthy subjects as well as in patients who underwent partial glossectomy. Our results show an excellent intrarater, interrater as well as test-retest reliability. Regression analysis confirmed these

findings and also showed symmetry of movement (left and right distances as being equal) as a core concept of tongue ROM in healthy adults.

Capturing tongue movement is challenging and tongue ROM had never been evaluated in detail as it has been in this study. With this 3D camera setup a first step in resolving this problem has been made. Future experiments should confirm the reliability of this measuring system in assessing the range of motion of the tongue in patients treated for different stages of oral tongue carcinoma.

Gender

The healthy participants group consisted of 15 persons of which eight were men and seven were women. There was no significant difference between the range of tongue motion in men and women, ß = 1.29, $\chi^2(1)$ = 0.44, p = 0.51 (Table 6.3).

Inter-participant variability

Mean distances of tongue ROM of all 15 participants are described in Table 6.1. Measurements of participant number 2 were most near the mean values of the total group. Therefore, participant 2 was set up as comparison participant in regression analysis. This analysis showed that there was a non-significant difference with only one other participant (no. 2 and no. 3 with ß = 0.11, $\chi^2(1)$ = 0.20, p = 0.652). All other participants significantly differed in tongue ROM of subject 2 with P < 0.001.

Patients

To determine the feasibility of the method we included ten partial glossectomy patients in the study. The patient group consisted of five men (50%) and five women (50%) with a mean age at the time of resection of 65 years (SD 10.0). All patients underwent lateral tongue resection with primary closure. Five patients had stage I disease (50%), one patient stage II (10%) and four patients stage III (40%). Two patients underwent re-excision because of incomplete resection and one other patient, four months later, because of leukoplakia. Only one patient underwent selective neck dissection and received postoperative radiotherapy on the ipsilateral neck. Mean time between operation and assessment of tongue ROM was 29 months (SD 18.1).

Mean distances between the tip of the tongue and the interdental papilla of the maxillary incisors were: 35.9 mm (protrusion), 42.3 mm (left), 47.3 mm (right), 17.8 mm (elevation) and 52.1 mm (depression) (Table 6.4). These measurements did not differ significantly from the overall measurement of the healthy subjects (t (1134) = 1.86; p = 0.06, 95% CI: -0.1-4.96).

In healthy individuals, lateral movements did not differ significantly. The lateral movements of patients, divided in resection side versus contralateral side, were significantly different (p < 0.001) with distances of 47.8 mm and 42.2 mm respectively (Table 6.4).

6

When measurements per patient were analysed it is seen that, in general, the lateral movement towards the resected side is not altered after partial glossectomy.

ACKNOWLEDGEMENTS

REFERENCES

[1]   S. Warnakulasuriya, "Global epidemiology of oral and oropharyngeal cancer," Oral Oncol., vol. 45, no. 4–5, pp. 309–316, 2009.

[2]   "www.cijfersoverkanker.nl," Assessed July 12, 2014.

[3]   D. G. Sessions, G. J. Spector, J. Lenox, B. Haughey, C. Chao, and J. Marks, "Analysis of treatment results for oral tongue cancer.," Laryngoscope, vol. 112, no. 4, pp. 616–25, Apr. 2002.

[4]   V. Konstantinović and N. Dimić, "Articulatory function and tongue mobility after surgery followed by radiotherapy for tongue and floor of the mouth cancer patients," Br. J. Plast. Surg., vol. 51, no. 1998, pp. 589–593, 1998.

[5]   T. Bressmann, R. Sader, T. L. Whitehill, and N. Samman, "Consonant Intelligibility and Tongue Motility in Patients with Partial Glossectomy," J. Oral Maxillofac. Surg., vol. 62, pp. 298–303, 2004.

[6]   Y. Matsui, K. Ohno, Y. Yamashita, and K. Takahashi, "Factors influencing postoperative speech function of tongue cancer patients following reconstruction with fasciocutaneous/myocutaneous flaps--a multicenter study.," Int. J. Oral Maxillofac. Surg., vol. 36, no. 7, pp. 601–9, Jul. 2007.

[7]   C. L. Lazarus, H. Husaini, A. S. Jacobson, J. K. Mojica, D. Buchbinder, D. Okay, and M. L. Urken, "Development of a new lingual range-of-motion assessment scale: normative data in surgically treated oral cancer patients.," Dysphagia, vol. 29, no. 4, pp. 489–99, Aug. 2014.

[8]   H. Husaini, G. P. Krisciunas, S. Langmore, J. K. Mojica, M. L. Urken, A. S. Jacobson, and C. L. Lazarus, "A survey of variables used by speech-language pathologists to assess function and predict functional recovery in oral cancer patients.," Dysphagia, vol. 29, no. 3, pp. 376–86, Jun. 2014.

[9]   A. M. Kreeft, I. B. Tan, C. R. Leemans, and A. J. M. Balm, "The surgical dilemma in advanced oral and oropharyngeal cancer: how we do it.," Clin. Otolaryngol., vol. 36, no. 3, pp. 260–266, Jun. 2011.

[10]   A. M. Kreeft, C. R. N. Rasch, S. H. Muller, F. A. Pameijer, E. Hallo, and A. J. M. Balm, "Cine MRI of swallowing in patients with advanced oral or oropharyngeal carcinoma: a feasibility study.," Eur. Arch. Otorhinolaryngol., vol. 269, no. 6, pp. 1703–11, Jun. 2012.

[11]   O. Rastadmehr, T. Bressmann, R. Smyth, and J. C. J. Irish, "Increased midsagittal tongue velocity as indication of articulatory compensation in patients with lateral partial glossectomies," Head Neck, no. June, pp. 718–727, 2008.

[12]   C. L. Lazarus, J. A. Logemann, A. W. Rademaker, T. R. H. Lurie, C. R. Larson, and B. B. Mittal, "Swallowing and Tongue Function and Oropharyngeal Cancer," J. Speech. Lang. Hear. Res., no. August, pp. 1011–1024, 2000.

[13]   "www.oncoline.nl/mondholte-en-orofarynxcarcinoom," Assessed July 17, 2014.

[14]   S. R. Youmans, G. L. Youmans, and J. A. G. Stierwalt, "Differences in tongue strength across age and gender: is there a diminished strength reserve?," Dysphagia, vol. 24, no. 1, pp. 57–65, Mar. 2009.

6

[15] D. B. Chepeha, T. N. Teknos, J. Shargorodsky, A. G. Sacco, T. Lyden, M. E. Prince, C. R. Bradford, and G. T. Wolf, "Rectangle tongue template for reconstruction of the hemiglossectomy defect.," Arch. Otolaryngol. Head. Neck Surg., vol. 134, pp. 993–998, 2008.

Chapter 7

Summary, conclusions, and future perspectives

SUMMARY PER CHAPTER

Chapter 1

Functional inoperability means that surgical removal of a tumour will lead to unacceptable function loss. The concept of functional inoperability poses clinical dilemmas in treatment planning for oral cavity and oropharynx cancer, as there is often little consensus on whether to treat locally advanced tumours with surgery or with an alternative, organ-sparing treatment, like radiotherapy or chemoradiation. Decisions in this matter are bound to be subjective, as they derive from the treating physician's personal expertise, with complex anatomy and patient variability clouding expectations of functional outcome. This is a growing problem, as incidence rates for oral and oropharyngeal cancer are rising, especially in younger patients.

The current treatment guidelines are mainly based on TNM classifications, which do not account for patient-specific factors. A personalized model of the oral cavity and oropharynx that allows for the performance of virtual therapy, would enable the multidisciplinary board to predict postoperative function more accurately, and plan treatment on more objective grounds, like visualized post-treatment effects on mastication and swallowing, and synthesized post-treatment speech. This would also improve patient counselling, as patients could base their opinions and wishes on realistic expectations.

To develop such model, the project 'Virtual Therapy for Head & Neck Cancer – Prediction of Functional Loss' was started in 2009, in which project this thesis is embedded. Aim of this thesis has been to formulate the specific technical requirements for building a patient-specific model, with modelling methods, and methods for incorporation of both anatomic and clinical information.

Chapter 2

To accurately predict individual functional treatment outcome in oral cancer patients, we would need a personalized, dynamic virtual model of their oral cavity and oropharynx that could present the functional consequences of surgical resection both visually and by means of sound. The most logical approach for creating such a model would be a finite-element based biomechanical model.

Chapter 2 described a biomechanical model of the tongue that could be implemented in a future total virtual environment of the oral cavity and oropharynx. The three-dimensional (3D) geometry of the model was segmented manually from magnetic resonance images, captured in a sagittal plane by means of 3-tesla MRI. Four extrinsic and four intrinsic muscles were positioned according to descriptions in literature. The simulated motions of a 'healthy' tongue were found to be in accordance with literature. As expected, the combined contractions of the transversus, the verticalis, and the posterior and middle part of the

7

genioglossus indeed resulted in a protruding motion. Fibrosis was imitated by increasing the stiffness of several elements on the left lateral side of the tongue model. This resulted in a bending movement towards the left, again as expected.

This chapter showed the possibilities of finite-element modelling, and discussed future steps needed to create a model for virtual surgery and accurate function predictions. Those future steps include implementing diffusion tensor MRI images for personalized representations of muscle fibres, and EMG measurements to individualize muscle activation patterns in specific functional tasks. Further refinement of the model is necessary in terms of element size and tissue variables to evaluate its quantitative accuracy.

Chapter 3

As mentioned in chapter 2, we propose EMG measurements for further individualization of biomechanical models. Because EMG of the tongue is difficult to perform, we started with the lips, which are essential for speech and facial expressions. We opted for the patient-friendly method of surface EMG (sEMG). To our knowledge, no studies so far have predicted 3D lip shapes accurately using surface electrodes. Chapter 3 evaluated whether sEMG conveys sufficient information to build a 3D static lip shape model. To eliminate the complexity of a full biomechanical model, two regression methods were used: one based on principal component analysis (PCA) and another on a modified generalized regression neural network (GRNN). Various EMG features, windows, and configurations were tested to find the optimal settings for facial sEMG. Five healthy subjects were included to evaluate generally applicability. We found that both PCA and modified GRNN are able to predict 3D lip shapes with satisfactory accuracy. However, PCA proved more consistent, and could better predict lip shapes not present in a training set. No clear optimal settings were found, except for the WAMP feature and unipolar configuration, which proved slightly more accurate.

Chapter 4

Dynamics were the focus of chapter 4, which investigated whether 3D lip movements could be predicted by means of sEMG. We measured 3D lip movements and corresponding sEMG activities in five healthy volunteers to create an accurate, 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 five volunteers carried out nineteen different instructions, which were each repeated five times. The relationship between 3D lip positions and movement and sEMG activities were accommodated in a state-space model. The measurement model applied in the state estimator used PCA to establish the relationship between sEMG activity and 3D lip positions. A truncated Taylor series up to order two was used to incorporate possible nonlinearities in this relationship. We tested two state-space models for modelling dynamics: a first-order

and a second-order discrete Kalman filter. Fudging parameters were used to fine-tune the system. A root mean square (RMS) error was calculated to assess accuracy, showing good correlations of sEMG activities and 3D lip movements with an average RMS error of 2.43 mm for the first-order system, and 2.46 mm for the second-order system. This model could be used in future research to improve and personalize inverse modelling algorithms.

Chapter 5

Inverse modelling poses several problems in biomechanical models. Especially in the tongue it is difficult to determine the precise patient-specific activation signals of the multiple muscles. This is caused by their large solution space and the possible presence of several mathematical solutions, which could all result in a motion of the model similar to the input motion, for instance a captured 3D course of the tongue. Furthermore, no possibility exists to validate quantitatively an individualized biomechanical tongue model as yet.

Precise anatomy is essential in predicting function loss, yet little is known about the anatomic variability in distribution pattern of the hypoglossal nerve in the tongue. Therefore, during (modified) radical neck dissection in ten patients, we stimulated the main branch and the first branches of the hypoglossal nerve, thus initiating motions of the tongue tip. Analysing video footage captured with a triple-camera setup, we were then able to reconstruct the individual 3D apical tongue trajectories. An intra-patient comparison showed the different trajectories were well distinguishable, meaning different muscles were innervated after stimulation of a different nerve branch. In an inter-patient comparison trajectories captured in the different patients were compared using drawings made by the surgeons during surgery. We found variations in the trajectories which were expected to be similar based on these drawings. This indicates the presence of functional anatomic variations.

In future research this setup can be used for the evaluation of individualized biomechanical models, and to reduce the solution space in inverse dynamic modelling in order to find the patient-specific activation signals.

Chapter 6

Treatment for oral cancer can seriously compromise speech and swallowing, due to reduced tongue mobility. However, no accurate measurement tool for tongue mobility has been developed as yet. This chapter proposed a new measurement setup for acquiring quantitative data on tongue mobility. Its reliability was assessed by measuring the tongue range of motion (ROM) in fifteen healthy adults, as well as in ten patients who had undergone partial glossectomies. Intrarater, interrater, and test-retest reliability were evaluated to make sure the setup gave consistent results and could be used in clinical practice by multiple persons without the introduction of a selection bias. A triple-camera system was used, so that 3D coordinates could be reconstructed by selecting image coordinates in frames captured

7

by each of the three cameras. Two markers were placed on the tip of the tongue, and four on the face. The markers on the face were used to secure the right position in between measurements. The participants were asked to move the tongue to five standardized tongue positions: protruded, left, right, upward, and down. At each position the face markers, the tip of the tongue, and the maxillary midline were selected. The maximum distances between the tip of the tongue and the maxillary midline were calculated. The healthy subjects were measured at two different time points, to allow for a test-retest comparison. Furthermore, each recording was analysed three times by two different raters to define intrarater and interrater reliability. The ten patients were all measured once.

All tests, intrarater, interrater, and test-retest alike, showed high correlation coefficients of >0.9. All test persons showed perfect symmetrical tongue ROM. In patients, significant differences in lateral tongue movements were found, with restricted tongue mobility post-surgery mainly affecting motions towards the contralateral side. The triple-camera setup proved a reliable way of gathering 3D tongue ROM data and can be used for objective grading of reduced tongue mobility after partial glossectomy. Future studies will need to confirm these findings in larger cohorts, including patients treated for various stages of oral tongue cancer. With these data we could develop a grading system that could link ROM to functional outcomes with respect to speech and swallowing.

CONCLUSIONS

This thesis has sprung from the project 'Virtual Therapy for Head & Neck Cancer – Prediction of Functional Loss', which aims to solve the clinical dilemma of functional inoperability in oral cavity and oropharyngeal cancer. Due to complex anatomy and anatomic variability, consensus on operability of individual patients is hard to reach. Preoperative expectations of functional outcome are subjective and therefore uncertain, frustrating proper patient counselling. To solve this dilemma, a virtual, patient-specific model has been proposed. This thesis has aimed to formulate the technical requirements for building such model, with focus on modelling methods, and possibilities for the incorporation of both anatomic and clinical information. Structures of major interest are the tongue and lips.

A biomechanical model based on finite element method is a useful tool to simulate movement of the tongue. The anatomical structures, like muscles and nerves, can be replicated in such a model, giving the possibility to personalize such a model. When material properties are adjusted, for instance to mimic fibrosis or other treatment effects, the model can simulate the effects on mobility and show what the resulting motion would be. This makes the model suitable for performing virtual therapy, and predicting the functional consequences.

The lips form another important part of the oral cavity. Adequate speech synthesis will require a proper 3D lip model covering both shape and movements. Lip shape is mainly

defined by the individual architecture of facial muscles. In a model based on principal component analysis (PCA), we achieved a promising 3D estimation of 2.76 mm using unilateral EMG measurements of eight facial muscles, including the digastric muscle. This model proved generally applicable, which is important for future clinical applications. The feature used most often to analyse EMG measurements, RMS, proved less adequate than WAMP in predicting lip shapes by means of facial sEMG. Future investigations will have to decide which feature is to be preferred.

Extending the model with dynamics, we were able to predict lip motions, too, with promising precision, even showing a slightly lower average error rate. Again the WAMP feature performed best, and therefore seems quite suitable for analysing the facial musculature. This model, too, proved generally applicable with its capacity to tune fudging parameters, showing optimal results in the same region for each subject. Consequently, we have demonstrated that 3D lip movements in speech as well as in facial expressions can be predicted accurately with generally applicable models based on sEMG of the facial and the digastric muscles.

With a unique setup we measured 3D tongue motions in a number of patients during surgery, selectively exciting the various nerve branches. The differences found in branching pattern as drawn by the surgeons, as well as the trajectories induced by comparable nerve branches, make the conclusion plausible that many variations exist in the nervous distribution of the hypoglossal nerve. The measured 3D trajectories could help quantitatively evaluate biomechanical models and improve inverse-modelling algorithms.

Quantitative measurements of 3D tongue ROM with a triple-camera setup showed high intrarater, interrater, and test-retest reliability. This high accuracy should allow for the development of an objective grading system for reduced tongue mobility. Since patients with smaller ROM are more likely to develop functional complications after treatment, such grading system could help decide which patients are prone to suffer major functional harm from therapeutic interventions.

FUTURE PERSPECTIVES

This section discusses various areas for future research, starting with aspects of anatomy and physiology, then moving on to inverse modelling and some perspectives on biomechanical modelling, and concluding with the potential benefit of clinical measurements.

Anatomy
The current biomechanical tongue model as described in chapter two is a generic one, and cannot predict individual function post treatment. Personalization of this model would require a (semi) automatic segmentation program for tongue images. Lee et al. created
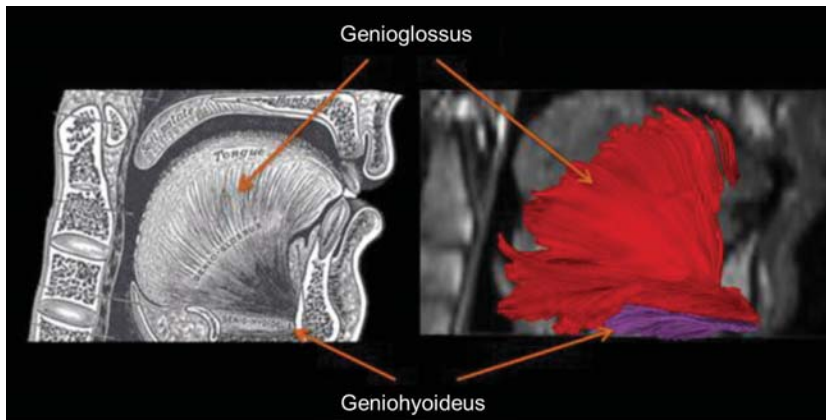
7

Figure 7.1 | DTI tractography for mapping of tongue muscles. Left: anatomic illustration. Right: muscular fibres detected with DTI tractography of the genioglossus and geniohyoideus. *Courtesy: Drs. J. Oudeman and dr. ir. A.J. Nederveen, Academic Medical Center, Amsterdam.*

a program that uses dynamic MRI images and a random walker algorithm, starting from manually marked regions that separate the tongue from its surrounding tissues.[1] The results seem promising, showing a 92% overlap with manual segmentation, which has been defined as the golden standard. The algorithm showed a 97% overlap in repeated runs. Furthermore, the segmented tongue volume showed less variability than when segmented manually.

Besides distinguishing the tongue from its surroundings, however, we also need to mark out the other relevant anatomic structures in the oral cavity and oropharynx. The algorithm should therefore be extended so that it could accurately segment multiple structures from the oral cavity and oropharynx without too much hassle. Perhaps a coupled multi-shape model might be incorporated, as has already been implemented for segmenting brain structures and organs in the pelvic floor.[2] This method is based on mutual shape variability, defining regions of structures in terms of locations and shape characteristics of other structures.

The involved musculature could be personalized by introducing the technique of diffusion tensor imaging (DTI), as proposed in chapter 2. DTI is a sophisticated MRI technique that shows how water diffuses in tissues. Originally the technique was developed to visualize white matter in the brain, but of late this technique has proven very useful in imaging individual muscle fibres: since water diffuses only along muscle fibres, and not in a perpendicular fashion, this imaging technique will show the actual directions of muscle fibres.[3–5] A pilot study applied DTI for imaging the tongue muscles (see Figure 7.1), and Ye et al. were able to distinguish the crossing fibres of the transversus, verticalis, and genioglossus
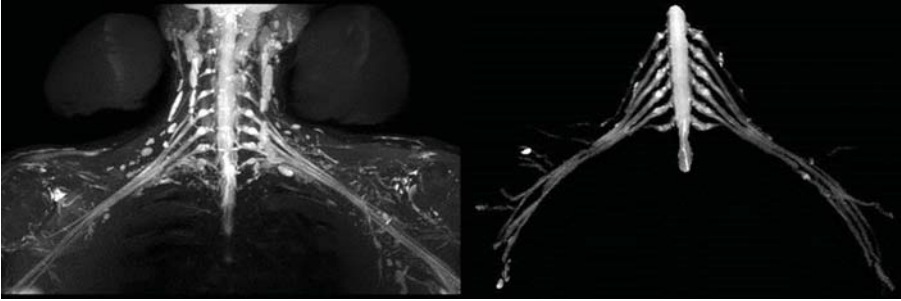
112

Figure 7.2 | Visualization of nerves using D-prep MRN. Left: maximum intensity projection of a D-prep MR neurography, showing the nerves at the level of the cervical vertebrae. Right: segmentation of the nerves as seen in the left image.
Courtesy: Drs. J. Oudeman and dr. ir. A.J. Nederveen, Academic Medical Center, Amsterdam.

muscles.[6] Further refinement will be necessary if DTI is to detect all individual muscle groups. Preliminary investigations demonstrate that this idea is feasible. Once further refined, DTI could help personalize musculature in the biomechanical tongue model, in relation to the extension of the tumour.

Another important aspect is the individual anatomy of nerve branching pattern. As discussed in chapter 5, distribution patterns vary. To further investigate this anatomic variation, 3D mapping of the hypoglossal nerve in multiple subjects is needed. A new MRI sequence called diffusion-prepared magnetic resonance neurography (D-prep MRN), can specifically visualize nerve sheaths by suppressing fat, vessel, and muscle signals (Figure 7.2).[7] Initial results are promising; not only large nerves in the head and neck region were visualized, but smaller structures were seen, too, with diameters comparable to that of the hypoglossal nerve (1.6 mm).[8] If accurate, D-prep MRN could be used to add personalized 3D maps of the hypoglossal nerve into the biomechanical model.

To validate the accuracy of D-prep mapping, a post-mortem study has been planned, performing D-prep neurography on post-mortem tongue specimens by means of 7-tesla MRI, and then serially dissecting the specimens in thin slices of 7μm, which are stained using hematoxylin and eosin (H&E) and S-100 immunohistochemic, to visualize the neural structures. By tracking the nerves through all slices we will be able to draw up a 3D map of the tongue branches, and compare this map with the one obtained from D-prep neurography. If proven accurate, D-prep mapping would be a way of investigating individual neural distribution patterns more closely in vivo.
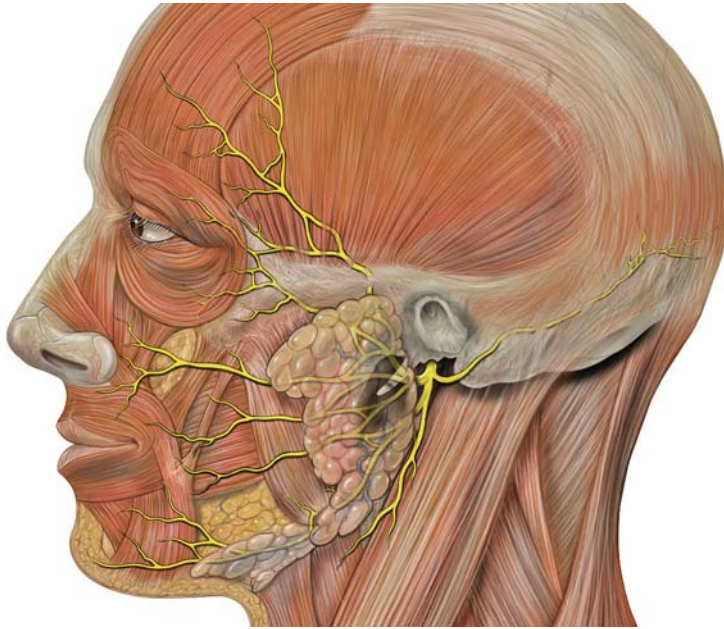
7

Figure 7.3 | Facial nerve branches (yellow) run below the parotid gland. After removal of the superficial lobe, the nerve can be stimulated.
Image created by Patrick Lynch, downloaded from [9].

Physiology

As stated before, the lips are an essential part of the oral cavity. The lip model described in chapters 3 and 4 is trained through voluntary muscle contractions of the facial musculature as a whole. How individual muscles or small muscle groups contribute to lip movements remains a subject for further investigation.

To find out more about the variability of facial nerve branching in relation to lip motion, we will need to perform further experiments with intraoperative nerve stimulation during superficial parotidectomies. During this procedure 3D lip motion can be measured as well as muscle activity using sEMG, without interference of voluntary or emotional compensation. The lip model can be trained using this data. Additional preoperative measurements can be used as an evaluation set, to see whether the statistic PCA model can accurately predict voluntary lip motions if trained through individual muscle or small muscle group contractions.

As mentioned in chapter 5, the current setup for in-vivo stimulation of the hypoglossal nerve could be extended with EMG measurements. We can measure the 3D trajectories of the tongue apex using marker tracking to examine which motion patterns relate to which nerve branches. EMG measurements during those intraoperative hypoglossal nerve stimulations

will give more insight which muscles or muscle groups are involved, allowing for a more reliable comparison between patients.

For individual muscles, needle EMG could be used. However, exact positioning of the needles remains difficult, due to the complex tongue muscle architecture. High-density EMG could serve as an alternative, recording muscle activity with a grid of small electrodes fixed onto a flexible substrate. These electrodes will all record the same action potentials, but from slightly different locations. Analysis of the EMG channels will yield information on depth, size, and location of the motor units, as well as on muscle fibre directions [10], telling us which muscles are innervated by which individual nerve branches. (See Figure 7.4 for a schematic view of a high-density surface EMG recording.) Positioning such a grid takes less precision and is easier done, but to our knowledge no high-density EMG grids are available as yet that could be fixed onto the tongue to measure its larger surface area.

In addition to the aforementioned application during stimulation of the hypoglossal nerve, a high-density grid could also be used during swallowing and speech to individualize muscle activation patterns during functional movements. If we could combine the use of a high-density grid with DTI and D-prep MRN, we could further personalize the tongue model as to anatomy and physiology, allowing for realistic predictions of treatment outcome with respect to swallowing and speech on the basis of individual muscle activation pattern, without the learning effects of retrained activation patterns.
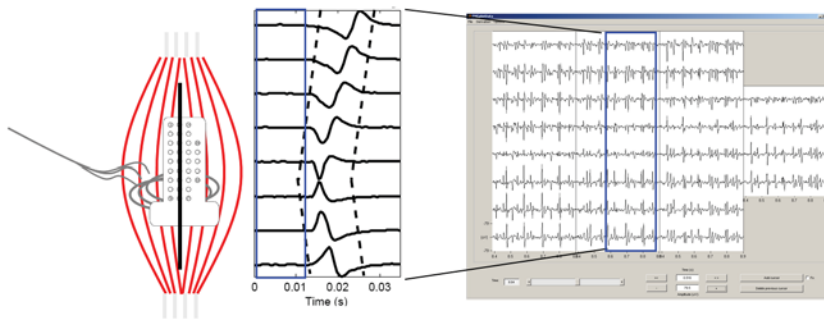

Figure 7.4 | Schematic overview of high-density surface EMG recording of muscle activity.
Left: schematic drawing of a muscle with electrode array on top of it. Grey lines with circles indicate the nerve endings and motor endplates. Right: recorded signals from the electrode array. Middle: a single motor unit action potential at different electrodes. The change in propagation direction that occurs at the innervation zone is clearly visible.
Courtesy: Prof. dr. ir. H.J. Hermens, Biomedical Signals and Systems, University of Twente.

7

115

Inverse dynamic modelling

An inverse-modelling algorithm could calculate muscle activation patterns from captured motion patterns. To this purpose, 3D marker trajectories are to be registered to corresponding locations on the biomechanical tongue model. At present inverse modelling is not yet sufficiently developed to draw conclusions regarding muscles that were innervated for functional movements. To calculate the muscle activation signals an error function is defined with various error terms with a weighing factor per term. The main term would be the Euclidean distance between the marker in the model and the measured marker location. Other terms may include changes in tongue volume, high excitation values for individual muscles, the number of innervated muscles, and changes in excitation value at succeeding time points for one muscle. Restrictions could also be implemented, like a manually set maximum excitation change. It is difficult to decide which terms should be incorporated into the error function, and what their corresponding weighing factors should be. Alteration of the error function would lead to different outcomes. Even if a proper error function could be defined, we would still have to deal with the fact that similar motions can be initiated through different muscle activation patterns.

EMG measurements of individual tongue muscles could solve this problem by showing which muscles are innervated at what activation levels. A trained statistical model as described in chapter 4 could also be used. Switching the roles of EMG signals and position coordinates, we could estimate the EMG signals from the measured marker positions. This would reveal patient-specific signalling, and provide us with an additional error term that expresses the degree of correspondence between estimated and measured EMG signals. Mutual information would be useful as a measure of correspondence.

With inverse dynamic modelling we hope to be able to calculate which muscular activation signals in our biomechanical model result in the observed motions of the lips/tongue. Another way of doing so would be to reduce the solution space significantly by introducing an error term that describes the correspondence between hypothesized muscular activation signals and measured sEMG signals. Again, mutual information would be useful for this purpose.

Individual geometry and muscular architecture will influence the calculated activation patterns in our model. Yet tissue properties, too, will need to be accounted for, as aspects like stiffness and damping are bound to bear their effects on tongue mobility as represented in the biomechanical model and thus the inverse-modelling algorithm.

Inverse dynamic modelling is not only valuable for our present research goals, but also for future clinical applications. For instance, it could be used to analyse whether a patient will be able to develop compensatory functional movements.[11] If so, this patient could be trained to compensate for their function loss. In this respect, inverse dynamic modelling is essential in building a tool that can accurately predict functional treatment outcome.

Biomechanical modelling

As described above, the current lip models are statistical, so-called black-box models that estimate lip movements from sEMG signals of facial muscles. However, to accurately predict individual function loss we need to account for individual anatomy, too. The existing black-box models should therefore be extended into full biomechanical models, with minimal loss of accuracy. EMG measurements can be used to backtrack muscle activation patterns in the biomechanical models. Subsequently, to eliminate errors in the estimation of those activations due to crosstalk and misplacement of the electrodes, the current models can be used in combination with inverse dynamic modelling, optimizing individual muscle actuation input.

The amplitude of EMG signals can vary between subjects due to electrode positioning in relation to the motor unit, but subject–specific factors, too, will be of influence, like thickness of the subcutaneous fat layer. We could use the statistical model to determine which channels have greatest influence in certain motion patterns. Transferring this information to the biomechanical model and combining it with inverse modelling, we could then define the correct muscles and muscles strengths on an individual basis.

To effectively simulate functionalities, the biomechanical model should cover the whole oral cavity and oropharynx, rather than just the lips and tongue. Speech synthesis, for instance, requires knowledge of the full vocal tract, starting at the vocal cords. If our model has to synthesize pathologic speech, it should incorporate all structures, also those not located within the treated area.

As discussed above, tissue properties, too, must be accounted for in calculating movement. Treatment effects on tissue characteristics are bound to affect function, as discussed in chapter 2. Therefore, fibrosis and other effects of on tissue parameters should be studied closely. Some studies have estimated tissue parameters by means of finite-element modelling. The properties found vary considerably, with a shear modulus ranging from 1.17 to 8.33 kPa.[12–14] A finite-element model will not be easily personalized; especially tissues affected by tumour or treatment will generate a wide variety of solutions. Therefore in-vivo methods of measuring tissue properties are preferred.

7

One suitable method may be magnetic resonance elastography, a technique that can measure viscoelastic properties of soft tissues in vivo. In 2011 Cheng et al. applied this technique to measure the viscoelastic properties of the tongue.[15] They found a relatively small range of 2.21 to 3.09 kPa for the elastic component, and 0.77 to 0.97 kPa for the viscous component in the tongue. This small range indicates the technique is robust, and there is little variability between subjects. However, the number of subjects in this study was rather low (n=7), and only a single age group was included: mean age 25.4 years with a standard deviation of 2.6 years. A similar study with a larger cohort, including a larger diversity of age, race, and gender, would probably convey more relevant information concerning the

variability of material properties in the tongue and other soft tissues. Additionally, a study cohort of patients, treated for oral cavity or oropharynx cancer, will show how treatment affects tissue properties and to what extent those effects vary in patients.

Clinical measurements

Chapter 6 introduced a new setup to quantitatively measure 3D tongue ROM. So far, only healthy subjects and partial glossectomy patients have been evaluated to confirm the reliability of this setup. Future research should focus on larger groups of treated patients, using speech analysis and quality-of-life questionnaires to correlate pre-existing restricted ROM to functional complications of treatment.

By measuring functional parameters and ROM pre- and post-surgery we could establish which areas of the tongue are associated with major loss of function if resected. With this information, we could develop a grading system that links preoperative ROM and function to risk of functional complications, assuming that patients with larger ROM before treatment can afford to lose more tissue before suffering unacceptable function loss than patients with limited tongue ROM and pre-existing functional issues. Such grading system would be a more objective and unbiased tool to decide which patients can be treated with surgery and which patients will be better off with a different treatment plan.

Finally, with measurements pre- and posttreatment we could validate the individual biomechanical models, for instance by comparing virtually calculated ROMs with actual ROMs. Also the capability of the model to adapt the virtual anatomy, including musculature and neural distribution, according to the performed treatment can be evaluated by measuring pre- and postoperative ROM.

REFERENCES

[1] J. Lee, J. Woo, F. Xing, E. Z. Murano, M. Stone, and J. L. Prince, "Semi-automatic segmentation for 3D motion analysis of the tongue with dynamic MRI," *Comput. Med. Imaging Graph.*, vol. 38, no. 8, pp. 714–724, 2014.

[2] A. Tsai, W. Wells, C. Tempany, E. Grimson, and A. Willsky, "Mutual information in coupled multi-shape model for medical image segmentation," *Med. Image Anal.*, vol. 8, no. 4, pp. 429–445, 2004.

[3] M. Froeling, J. Oudeman, S. van den Berg, K. Nicolay, M. Maas, G. J. Strijkers, M. R. Drost, and A. J. Nederveen, "Reproducibility of diffusion tensor imaging in human forearm muscles at 3.0 T in a clinical setting," *Magn. Reson. Med.*, vol. 64, no. 4, pp. 1182–1190, Oct. 2010.

[4] F. M. Zijta, M. Froeling, M. P. van der Paardt, M. M. E. Lakeman, S. Bipat, a D. M. van Swijndregt, G. J. Strijkers, a J. Nederveen, and J. Stoker, "Feasibility of diffusion tensor imaging (DTI) with fibre tractography of the normal female pelvic floor," *Eur. Radiol.*, vol. 21, no. 6, pp. 1243–1249, Jun. 2011.

[5] M. Froeling, A. J. Nederveen, D. F. R. Heijtel, A. Lataster, C. Bos, K. Nicolay, M. Maas, M. R. Drost, and G. J. Strijkers, "Diffusion-tensor MRI reveals the complex muscle architecture of the human forearm," *J. Magn. Reson. Imaging*, vol. 36, no. 1, pp. 237–248, 2012.

[6] C. Ye, A. Carass, E. Murano, M. Stone, and J. L. Prince, "A bayesian approach to distinguishing interdigitated muscles in the tongue from limited diffusion weighted imaging," *Bayesian Graph Model. Biomed Imaging*, vol. 8677, pp. 13–24, 2014.

[7] M. Yoneyama, T. Takahara, T. C. Kwee, M. Nakamura, and T. Tabuchi, "Rapid High Resolution MR Neurography with a Diffusion-weighted Pre-pulse," *Magn. Reson. Med. Sci.*, vol. 12, no. 2, pp. 111–119, 2013.

[8] T. Buckle, B. Verbist, F. Balm, M. Malessy, M. van Buchem, and F. Van Leeuwen, "Feasibility of nerve specific MRI for surgical planning and virtual surgery in the head and neck area," *J. Nucl. Med.*, vol. 55, no. 1, p. 1442, May 2014.

[9] B. di Muzio and F. Gaillard, "Facial nerve," 2015. [Online]. Available: http://radiopaedia.org/articles/facial-nerve. [Accessed: 27-May-2015].

[10] D. F. Stegeman, B. U. Kleine, B. G. Lapatki, and J. P. Van Dijk, "High-density Surface EMG: Techniques and Applications at a Motor Unit Level," *Biocybern. Biomed. Eng.*, vol. 32, no. 3, pp. 3–27, 2012.

[11] A. G. Hannam, I. K. Stavness, J. E. Lloyd, S. S. Fels, A. J. Miller, and D. a Curtis, "A comparison of simulated jaw dynamics in models of segmental mandibular resection versus resection with alloplastic reconstruction.," *J. Prosthet. Dent.*, vol. 104, no. 3, pp. 191–198, Sep. 2010.

[12] C. X. C. Xu, M. J. Brennick, and D. M. Wootton, "Image-based three-dimensional finite element modeling approach for upper airway mechanics," in *IEEE Engineering in Medicine and Biology 27th Annual Conference*, 2005, vol. 3, pp. 2587–2590.

7

[13] Z. S. Liu, X. Y. Luo, H. P. Lee, and C. Lu, "Snoring source identification and snoring noise prediction," *J. Biomech.*, vol. 40, no. 4, pp. 861–870, 2007.

[14] A. Malhotra, Y. Huang, R. B. Fogel, G. Pillar, J. K. Edwards, R. Kikinis, S. H. Loring, and D. P. White, "The male predisposition to pharyngeal collapse: Importance of airway length," *Am. J. Respir. Crit. Care Med.*, vol. 166, no. 10, pp. 1388–1395, 2002.

[15] S. Cheng, S. C. Gandevia, M. Green, R. Sinkus, and L. E. Bilston, "Viscoelastic properties of the tongue and soft palate using MR elastography," *J. Biomech.*, vol. 44, no. 3, pp. 450–454, 2011.

Chapter 8

Summary

Samenvatting

Curriculum vitae

Dankwoord

SUMMARY

Oral and oropharyngeal squamous cell carcinoma together rank sixth among the most common types of cancer worldwide, and incidence is rising. The most important treatment options are surgery and chemoradiotherapy. In current clinical practice, the primary choice of treatment for oral cavity cancer is surgery, given that the tumour is not designated as inoperable. If inoperability is assumed, the alternative of organ-sparing chemoradiotherapy is proposed. Two forms of inoperability exist: 'anatomical' and 'functional'. In case of anatomic inoperability, a radical resection of the tumour is not possible due to invasion of vital structures. Generally, using preoperative imaging this can be determined properly, and will not lead to discussions. Functional inoperability means a tumour can be resected radically, but the serious consequences for functions such as speech, mastication, and swallowing, are considered as unacceptable. However, the estimation of functional inoperability is subjective and unreliable. Patient-specific factors, like anatomical variability of the musculature and nerve distribution, can have major impact on the postoperative functions. Currently, knowledge regarding those factors is insufficient to reach an individualized estimation of the functional outcome. To achieve an objective and accurate estimation of the patient-specific functional consequences, the project 'Virtual Therapy for Head & Neck Cancer – Prediction of Functional Loss´ was launched. This estimation can contribute to the discussions regarding the inoperability of tumours during a multidisciplinary tumour board meeting. Furthermore, it will be useful for counselling the patient with respect to its functional consequences of the proposed surgical procedure.

This thesis is a part of this project, and aims on formulating the technical requirements for a patient-specific model, able to predict postoperative functions. The optimal modelling method, as well as the possibility to incorporate anatomical and clinical data are investigated. This thesis focusses on modelling the tongue and lips, which have large influence on the functions of the oral cavity and oropharynx.

Firstly, the optimal modelling method to incorporate patient-specific factors and simulate consequences of the surgery are described. We showed that using a biomechanical model of the tongue, based on the finite element method, motions initiated by individual muscles or muscle groups are comparable with literature. Also the effect of changed tissue parameters can be imitated.

One of the aspects necessary for accurate function prediction is the person-specific muscle activation. Electromyography (EMG) can be used to obtain an indication of the several activation levels. To evaluate if patient-friendly surface EMG (sEMG) holds sufficient information to incorporate in a biomechanical model, we investigated if accurate estimations could be made of positions and motions of the lips, using sEMG. Using statistical models, we showed that lip movements can be estimated with an accuracy of around 2.5 mm.

8

Next to activation signals, also the anatomical features should be personalized, especially because of the possible variations in the neural distribution. We developed a set-up that enables us to attain more insight in these variations and, in future research, could help evaluate biomechanical models quantitatively. For this set-up the branches of the nervus hypoglossus are stimulated during a radical neck dissection, and simultaneously the corresponding trajectories of the tongue are captured three dimensional (3D). The first measurements showed that we are able to innervate different muscles or muscle groups, and, in addition, the results indicate that several anatomical variations exist.

No methods to acquire quantitative data of tongue mobility has been developed as yet. Such a method can be used to measure the postoperative motion restriction in an objective and accurate manner. Therefore, we introduced a 3D camera system, which can measure the tongue range of motion using markers on the tip of the tongue. We demonstrated the measurement system is reliable and shows reproducible results. Furthermore it was shown that the range of motion in patients, treated for a tongue carcinoma, is diminished to the contralateral side.

In conclusion we can say that a biomechanical model of the oral cavity and oropharynx gives us opportunities to perform patient-specific functional predictions. sEMG can be used to estimate muscle activities and the corresponding motions. In vivo stimulations of the nervus hypoglossus give unique possibilities to investigate anatomical variations and can be used to evaluate biomechanical models. Finally, the range of motion of the tongue can be measured accurate, which could facilitate a grading system for risk levels.

Future research will focus on further personalization of biomechanical models, by means of a better representation of the patient-specific musculature and nerve distribution, as well as the geometry. This personalization also includes optimizing the tissue characteristics of the model. A more extensive clinical investigation regarding the tongue mobility and the corresponding functions will be give important insights. Further exploration of the physiology will be done by measuring sEMG during intra-operative stimulations of the nervus hypoglossus as well as the nervus facialis. Furthermore EMG will also play a major role in optimizing inverse modelling algorithms, such that those can contribute to an improved estimation of the postoperative consequences.

124

SAMENVATTING

Mond- en keelholte tumoren vormen gezamenlijk de zesde meest voorkomende vorm van kanker wereldwijd, en vertonen een stijgende incidentie. De belangrijkste behandelopties zijn chirurgie en chemoradiatie. In de huidige klinische praktijk is chirurgie de eerste behandel keuze voor mondholtetumoren, mits de tumor niet beschouwd wordt als inoperabel. In dat geval wordt het alternatief van orgaansparende chemoradiatie voorgesteld. Er bestaan twee vormen van inoperabiliteit: 'anatomische' en 'functionele'. In het geval van anatomische inoperabiliteit is een radicale resectie van de tumor niet mogelijk vanwege ingroei in vitale structuren. Meestal kan dit door preoperatieve beeldvorming goed worden vastgesteld, en leidt dit doorgaans niet tot discussies. Functionele inoperabiliteit betekent dat de tumor wel chirurgisch radicaal te verwijderen is, maar dat de ernstige beperkingen voor spraak-, kauw-, en slikfuncties als niet acceptabel moeten worden beschouwd. De inschatting van functionele inoperabiliteit is echter subjectief en onbetrouwbaar. Patiënt specifieke factoren, zoals variaties in de anatomie van spieren en zenuwen, kunnen een grote invloed hebben op de functies na een chirurgische ingreep. Kennis daarover is momenteel ontoereikend voor een individuele inschatting van de functionele uitkomst. Om een objectieve en accurate beoordeling van de patiënt specifieke functionele gevolgen te realiseren is het project 'Virtual Therapy for Head & Neck Cancer – Prediction of Functional Loss' opgezet. Deze beoordeling kan bijdragen aan feitelijke discussies over de inoperabiliteit van tumoren tijdens multidisciplinair overleg. Bovendien kan dit een bijdrage leveren aan een verbeterde voorlichting van de patiënt over de functionele gevolgen van de voorgestelde chirurgische ingreep.

Dit proefschrift is een onderdeel van dit project en heeft als doel inzicht te verkrijgen in de specifieke technische vereisten voor een patiënt-specifiek model dat in staat moet zijn om postoperatieve functies te voorspellen. Hiervoor wordt zowel gekeken naar de optimale methode om te modelleren, als de mogelijkheden om anatomische en klinische data te incorporeren in een model. Dit proefschrift behandeld de modellering van tong en lippen, die mede bepalend zijn voor de functionaliteit van mond- en keelholte.

Allereerst is aandacht besteed aan de keuze voor het optimale model om patiënt specifieke eigenschappen te incorporeren en gevolgen van een ingreep te voorspellen. We hebben aangetoond dat middels een biomechanisch model van de tong, gebaseerd op de eindige elementen methode, de door activatie van individuele spieren of spiergroepen geïnitieerde gesimuleerde bewegingen overeenkomen met beschrijvingen in de literatuur. Daarnaast kan het effect van veranderde weefseleigenschappen worden nagebootst.

Een van de aspecten die nodig zijn voor een accurate functie voorspelling is een persoon specifieke spier activatie. Hiervoor kan elektromyografie (EMG) gebruikt worden om een indicatie te geven van de verschillende activatie levels. Om te evalueren of patiënt

vriendelijke oppervlakte EMG (sEMG) voldoende informatie kan bevatten om te integreren in een biomechanisch model, is bestudeerd of accurate voorspellingen mogelijk zijn van de posities en bewegingen van de lippen op basis van sEMG. Met statistische modellen hebben we aangetoond dat lip bewegingen met een nauwkeurigheid van ongeveer 2,5 mm voorspeld kunnen worden.

Naast de activatie patronen moeten ook de anatomische kenmerken gepersonaliseerd worden, vooral ook vanwege de mogelijke variaties in het vertakkingspatroon van de zenuwen. We hebben een setup ontwikkeld die het mogelijk maakt om meer inzicht te krijgen in deze variaties en die het in de toekomst mogelijk maakt biomechanische modellen kwantitatief te evalueren. Hiervoor worden tijdens een halsklierdissectie de aftakkingen van de nervus hypoglossus gestimuleerd en worden tegelijkertijd de bewegingspatronen van de tong drie dimensionaal (3D) vastgelegd. Uit de eerste metingen is gebleken dat we in staat zijn verschillende spieren en spiergroepen te innerveren, en daarnaast lijken de resultaten erop te wijzen dat er anatomische variaties bestaan.

Tot op heden waren er nog geen methoden beschikbaar om kwantitatieve data te verkrijgen over de beweeglijkheid van de tong, die gebruikt kunnen worden voor het objectief en accuraat vastleggen van postoperatieve bewegingsbeperking. Daarom hebben we een 3D camera systeem geïntroduceerd dat het bereik van de tong aan de hand van markers op de tongpunt in verschillende richtingen meet. We hebben aangetoond dat de meetopstelling betrouwbare en reproduceerbare resultaten laat zien. Daarnaast is ook aangetoond dat het bereik van de tong bij patiënten, behandeld voor een tong carcinoom, is verminderd naar de contralaterale zijde.

Concluderend kan gezegd worden dat een biomechanisch model van de mond- en keelholte de mogelijkheid biedt om uiteindelijk patiënt specifieke functionele voorspellingen te kunnen doen. sEMG kan gebruikt worden om een inschatting te maken van de spier activiteiten en de bijbehorende bewegingen. De in vivo stimulatie van de nervus hypoglossus geeft unieke mogelijkheden om de anatomische variaties te onderzoeken en kan gebruikt worden voor de evaluatie van biomechanische modellen. Tenslotte kan het bereik van de tong in verschillende richtingen betrouwbaar gemeten worden, wat mogelijkheden geeft om tot een gradering van risico levels te komen.

Toekomstig onderzoek zal zich onder andere richten op het verder personaliseren van de biomechanische modellen, door zowel de patiënt specifieke spier en zenuw locaties als de geometrie beter te representeren. Hieronder valt ook het optimaliseren van weefseleigenschappen in het model. Ook een uitgebreidere klinische studie naar de beweeglijkheid van de tong en corresponderende functies, zullen hierbij een belangrijke rol spelen. De fysiologie zal nader bestudeerd worden door tijdens intra-operatieve zenuwstimulatie van zowel de nervus hypoglossus als de nervus facialis, gelijktijdig sEMG te meten. Daarnaast zal EMG ook een belangrijke rol gaan spelen in het optimaliseren

van inverse modelering algoritmen, zodat deze kunnen bijdragen aan een verbeterde inschatting van de gevolgen na behandeling.

8

CURRICULUM VITAE

Maarten van Alphen werd geboren op 16 juli 1987 te Nijmegen. Hij groeide op in Deurne en behaalde daar in 2005 zijn diploma op het St.-Willibrord Gymnasium. Hierna is hij begonnen met de studie *Technische Geneeskunde* aan de Universiteit Twente te Enschede. In 2008 begon hij met de master specialisatie *Robotics and Imaging*.

Als onderdeel van deze master heeft hij vier korte klinische onderzoeksstages gelopen op verschillende afdelingen en in meerdere ziekenhuizen. Het uitgevoerde onderzoek tijdens deze stages omvatte uiteenlopende onderzoeksgebieden zoals de ontwikkelingen van modellen, medische beeldvorming, en beeldbewerking.

Reeds tijdens de eerste stage, die hij ook liep op de afdeling *hoofd-halsoncologie en -chirurgie* in het Antoni van Leeuwenhoek, werd zijn interesse gewekt voor het idee van virtuele therapie. In 2011 heeft hij op deze afdeling ook zijn afstudeeronderzoek uitgevoerd. De focus was toen op de ontwikkeling van een biomechanisch tong model. Aansluitend aan het afronden van de opleiding *Technical Medicine* in december 2011, vervolgde hij zijn onderzoek in een promotietraject op ditzelfde project. De resultaten van het daarin uitgevoerde onderzoek zijn beschreven in dit proefschrift.

Maarten van Alphen was born on the 16th of July 1987 in Nijmegen. He grew up in Deurne, where he graduated at the St.-Willibrord Gymnasium in 2005. Thereafter he started with the study *Technical Medicine* at the University of Twente, in Enschede. In 2008 he began his master-specialization *Robotics and Imaging*.

As part of this master, he performed four clinical-research internships at different departments in several hospitals. The research topics during those internships contained subjects as the development of models, medical imaging, and image processing.

During his first internship, also conducted at the department of *head and neck oncology and surgery* of the Netherlands Cancer Institute, he got interested in the idea of virtual therapy. In 2011 he performed his graduation research at the same department, with the focus on the development of a biomechanical tongue model. After graduating *Technical Medicine* in December 2011, he continued his research as a PhD student on the same project. The results of this research are presented in this thesis.

8

DANKWOORD

Een proefschrift is niet het werk van één persoon. Ik wil hier kort enkele mensen bedanken zonder wie het schrijven van dit proefschrift niet mogelijk, of veel moeilijker, was geweest. Allereerst wil ik mijn promotor prof. dr. A.J.M Balm, bedanken. Hij kwam met het idee een virtuele dubbelganger van een patiënt te maken voor de inschatting van functie gevolgen. Hij heeft velen kunnen overtuigen van dit idee, waardoor het project *Virtual Therapy* nu loopt. Daarnaast wil ik hem bedanken voor de constructieve samenwerking.

Uiteraard wil ik ook mijn andere promotor, prof. dr. ir. C.H. Slump, bedanken voor het mogelijk maken van dit promotietraject, en mijn co-promotor, dr. ir. F. van der Heijden, voor de intensieve ondersteuning op technologisch vlak, de veelvuldig sparringsmomenten, en de talloze autoritten richting Amsterdam.

Daarnaast waren alle onderstaande personen van groot belang. Mijn dank gaat uit naar ieder van hen.

Dr. A.M. Kreeft heeft veel werk verzet om het probleem 'functionele inoperabiliteit' op de kaart te krijgen, en heeft mij geholpen tijdens mijn eerste stappen binnen dit project. Ook de andere co-auteurs, Prof. dr. L.E. Smeele, dr. D. Brandsma, dr. I. Jacobi, waren altijd bereid om mee te denken vanuit hun eigen specialisme.

Dr. R.J.J.H. van Son heeft vele meetings bijgewoond en ons onderzoekers bijgestaan met zijn kennis op het gebied van taal en alles wat daarmee samenhangt. Tevens was hij bereid als referent te fungeren voor dit proefschrift.

Meerdere studenten Technische Geneeskunde hebben aan gedeelten van het project gewerkt. Zij waren van grote waarde voor mijn onderzoek vanwege het verrichte werk maar ook hun frisse blik. Bovendien brachten zij de nodige gezelligheid met zich mee.

Alle onderzoekers van de 3de etage van het O-gebouw en de andere 'TG' promovendi in het AVL wil ik bedanken voor de gezellige periode die ik daar heb mogen meemaken. En vooral ook mijn kamergenoten, Sheima en Marnix, die mij óók hebben geduld op de momenten van frustratie.

De hoofd-halschirurgen uit het Antoni van Leeuwenhoek hebben hun medewerking verleend, zodat de metingen tijdens de OK verricht konden worden. Daarnaast zorgden zij met hun prettige houding richting alle medewerkers voor een fijne en motiverende werksfeer.

Twee personen die ik zeker niet wil vergeten zijn Marion van Zuilen en Henny Buis. Ik kon hen altijd om hulp vragen als er iets geregeld moest worden voor mij of een van de studenten. Ook aan de Universiteit Twente had ik zo'n aanspreekpunt, Sandra Westhoff. Zij heeft mij veel geholpen met alles wat daar geregeld moest worden.

Dankzij de bijdrage van de Maurits en Anna de Kock stichting konden we de benodigde apparatuur financieren. De technische dienst in het AVL, met name Michiel van der Meer,

8

Michiel Sinaasappel, en Ton Vlasveld, hebben er hard aan gewerkt om van deze apparatuur een bruikbare opstellingen te maken. Deze opstellingen worden nog steeds gebruikt in lopende studies, en zijn daarmee van groot belang voor het totale project.

Bij Jos Oudeman (AMC) en Tessa Buckle (LUMC) kon ik altijd terecht voor vragen over complexe MRI sequenties en de mogelijkheden hiermee.

Wetenschappelijk onderzoek, en zeker klinisch onderzoek, valt of staat bij de bereidwilligheid van patiënten en vrijwilligers om mee te doen aan studies. Ik heb vooral veel respect voor de patiënten die, in een voor hen moeilijke en onzekere periode, bereid waren tijd vrij te maken en metingen te ondergaan, zodat later anderen hier van kunnen profiteren.

Maja Keizers is bereid geweest om grote delen van dit proefschrift door te nemen en tekstueel te verbeteren. Het is de leesbaarheid en duidelijkheid van deze teksten zeker ten goede gekomen dat zij zich hier voor heeft willen inspannen.

Verder kan een promotie niet plaatsvinden zonder commissie. Ik wil prof. dr. ir. P.H. Veltink, prof. dr. T. Ruers, prof. dr. L.E. Smeele, en prof. dr. S.J. Bergé danken dat zij de tijd vrij hebben willen maken om mijn proefschrift te beoordelen en plaats te nemen in de commissie.

De oplettende lezer had misschien al door dat ik twee co-auteurs nog niet heb genoemd. Merijn Eskes en Simone van Dijk, twee andere promovendi binnen het Virtual Therapy project, waarmee ik zodoende intensief heb samengewerkt. Ik ben blij dat zij mij op 27 augustus wederom willen bijstaan als paranimf.

Tenslotte wil ik nog mijn vrienden en familie bedanken voor het tonen van hun interesse in mijn onderzoek, maar vooral de leuke en gezellige momenten die ik met jullie heb meegemaakt. In het bijzonder ook mijn ouders voor hun steun, en dat ik altijd welkom was om de drukke stad te ontvluchten om in het rustige Deurne bij te komen. En tot slot wil ik Sjoukje natuurlijk bedanken, die er altijd voor me was als luisterend oor maar vooral ook voor de mooie momenten die we hebben beleefd en waar er hopelijk nog veel van zullen volgen.

Stellingen

behorende bij proefschrift

Towards a predictive model for functional loss after oral cancer treatment

1. Met een virtueel biomechanisch model van de mond- en keelholte kan functionele inoperabiliteit objectief ingeschat worden. *Dit proefschrift*

2. Het gebruik van patiënt vriendelijk oppervlakte EMG maakt nauwkeurige voorspelling van lipbewegingen mogelijk. *Dit proefschrift*

3. Selectieve stimulatie van de nervus hypoglossus en synchrone opname van 3D bewegingen van de tong levert unieke mogelijkheden op voor het verbeteren van inverse dynamica algoritmen. *Dit proefschrift*

4. Analyse van drie dimensionale tongbewegingen na selectieve peroperatieve stimulatie van de nervus hypoglossus tijdens een halsklierdissectie maakt onderzoek naar anatomische variaties mogelijk. *Dit proefschrift*

5. Bij het ontwikkelen van medische technologieën moet men beginnen vanuit de klinische praktijk in plaats vanuit de technologische mogelijkheden.

6. '*Meten is weten*' is van belang, maar '*leren door proberen*´ is minstens zo belangrijk.

7. Voor vernieuwende ideeën moet men niet beperkt worden door te veel kennis over het onderwerp.

8. De kunst van onderzoek doen op de grens van de technische en medische wereld, is medici de mogelijkheden laten inzien en technici de beperkingen.

9. Essentially, all models are wrong, but some are useful. *George E.P. Box*

10. De rol van de Technisch Geneeskundige in het complexe medische technisch handelen wordt steeds belangrijker.

Maarten van Alphen
Enschede, 27 augustus 2015