Intra-operative assessment of resection margins by three-dimensional ultrasound in patients with tongue squamous cell carcinoma

UNIVERSITY OF TWENTE.

By

N.M. Bekedam

A thesis submitted for the degree of *Master of Science*
Faculty Science and Technology
University of Twente
Enschede, The Netherlands

September, 2020

Graduation Committee

Technical supervisor (Chairman) Dr.ir. F. van der Heijden	Department of Robotics and Mechatronics TechMed Centre University of Twente Enschede, The Netherlands
Medical supervisor Prof.dr. L.E. Smeele	Department of Head and Neck Surgery & Oncology Netherlands Cancer Institute Antoni van Leeuwenhoek Amsterdam, The Netherlands
Process supervisor Dr. M. Groenier	Faculty Science and Technology University of Twente Enschede, The Netherlands
Technical supervisor (additional) Dr.ir. R.L.P. van Veen	Department of Head and Neck Surgery & Oncology Verwelius 3D Lab Netherlands Cancer Institute Antoni van Leeuwenhoek Amsterdam, The Netherlands
Technical supervisor (additional) Dr. M.J.A van Alphen	Department of Head and Neck Surgery & Oncology Verwelius 3D Lab Netherlands Cancer Institute Antoni van Leeuwenhoek Amsterdam, The Netherlands
Medical supervisor (additional) Dr. M.B. Karakullukçu	Department of Head and Neck Surgery & Oncology Verwelius 3D Lab Netherlands Cancer Institute Antoni van Leeuwenhoek Amsterdam, The Netherlands
External member Dr.ir. A.M. Leferink	Department of Applied Stem Cell Technologies TechMed Centre University of Twente Enschede, The Netherlands

Acknowledgements

I would like to thank my supervisors for all the support during my graduation. In addition, the help from all colleagues from the Antoni van Leeuwenhoek hospital to realize this research is much appreciated. I am very grateful for all the opportunities in the past year to learn and develop myself in academic, clinical, professional and personal ways. Lastly, a special thanks to family and friends who supported and stimulated me, so I could perform to the best of my abilities.

Abstract

Surgical excision is the most common treatment for tongue squamous cell carcinomas (TSCC). Surgeons aim to remove the tumor with a minimal resection margin of 5 mm to reduce the chance of recurrence. Currently, there is no intra-operative assessment available to determine if a 5 mm resection margin is achieved. To improve the surgical precision of TSCC resections, this research aims to provide surgical guidance during resections of TSCC using three-dimensional (3D) ultrasound (US) to minimize close resection margins. This research was divided into three parts which together investigated the feasibility of 3D US for intraoperative assessment of surgical resection margins of tongue squamous cell carcinomas.

The first part of this research provides a better understanding of data acquisition and reconstruction of 3D US. In a phantom study, the influence of the 1) reconstruction algorithm, 2) sweeping method, 3) US transducer frequency, 4) stabilization rails and 5) observer was investigated. The accuracy of the 3D US volumes was evaluated by the signal-to-noise ratio (SNR), contrast-to-noise ratio (CNR), derivative along a scan line and the Full Width at Half Maximum (FWHM) of the peak of the derivative of the pixel intensity. The results show that data acquisition was performed best using the highest US frequency possible, a single sweep method assisted by rails and performed by a single operator. This study could not identify a reconstruction algorithm performing better than others.

The second part of this research investigated as a proof of concept that deep learning is a feasible technique for fast automatic multi-class segmentation of tongue specimen and tumor in 3D freehand US volumes. The multi-class segmentation of tongue specimen and tumor was split into two binary segmentation problems by adopting the cascade strategy. Two identical UNet models were trained upon their own dataset (from a total dataset of 44 3D US volumes of 8 patients) and the influence of the loss function (Dice and binary cross-entropy (BCE)) and data augmentation was investigated. Evaluation based on the Dice similarity coefficient (DSC), showed 86% DSC (BCE loss with data augmentation applied) predicting the specimen and 18% DSC (Dice loss and data augmentation applied) predicting the tumor.

The third part of this research explored the correlation between resection margins assessed by 3D US and histopathology. This study included 8 patients of which the resection margins of TSCC were assessed intra-operatively by 3D US and post-operatively by histopathology. The correlation between the measurements by 3D US and histopathology was computed by the Pearson correlation coefficient. The results showed that the measurements of resection margins by 3D US and histopathology do not correlate statistically significant, meaning that 3D US could not provide correct intra-operative feedback to the surgeon.

Future research should focus on expanding the dataset and improving the data acquisition, by utilizing a high frequency US transducer and stabilization rails. In addition, remodeling of histopathological slices into 3D models and registering those 3D models towards the 3D US models could help the radiologist annotating more accurately. Also, research should investigate which hyperparameters in the deep learning models perform superior to obtain maximum DSC in predicting the specimen and tumor in 3D US volumes. Eventually after all these improvements, it is speculated that recalculating the correlation between the resection margins measured by 3D US and histopathology in tongue tumor specimens could be statistically significant.

Contents

List of Figures

List of Table

List of Acronyms

2D	Two-dimensional
3D	Three-Dimensional
AP	Anterior-Posterior
BCE	Binary Cross-Entropy
CN	Central Nerve
CNN	Convolutional Neural Network
CNR	Contrast to Noise Ratio
CT	Computed Tomography
DL	Deep Learning
DSC	Dice Similarity Coefficient
DW	Distance Weighted
EM	Electromagnetic
ENE	Extranodal extension
FWHM	Full Width at Half Maximum
HPV	Human Papilloma Virus
IS	Inferior-Superior
ML	Machine Learning
MRI	Magnetic Resonance Imaging
MSD	Mean Sum of Distances
NCI	Netherlands Cancer Institute
PBM	Pixel Based Methods
PNN	Pixel Nearest Neighbor
PV	Pixel Values
ReLU	Rectified Linear Unit
RL	Right-Left
ROI	Region of Interest
SD	Standard Deviation
SNR	Signal to Noise Ratio

TNM Tumor, Nodes, Metastasis

TSCC Tongue Squamous Cell Carcinoma

US Ultrasonography

VBM Voxel Based Methods

VNN Voxel Nearest Neighbor

VNN2 Voxel Nearest Neighbor2

1
General Introduction

In the Netherlands, the prevalence of tongue cancer was 2069 in 2018, and the incidence is 300 to 400 [1]. Tongue cancer is part of the larger subject oral cavity cancer. The dominant cancer of the oral cavity is squamous cell carcinoma (OSCC) and frequently associated with risk factors such as chronic smoking, alcohol use, the presence of human papilloma virus (HPV) [2–5] and sunlight exposure [6]. OSCC is frequently diagnosed in patients aged between 55 and 65 years old for men, and 50 and 75 years old for women [4, 7]. OSCC is treated by surgery, chemo- or radiotherapy or a combination of those. However, these treatments cause damage to the tongue and have negative effects on speech, swallowing and mastication [3, 8] and interferes with cosmetic appearance [4]. Complications can be minimized by prosthetic rehabilitation [9]. After treatment, patients should be encouraged to stop smoking and drinking alcohol as those are high risk factors for OSCC [4].

OSCC is mostly detected through physical examination by dental and general practitioners. Routine biopsy of precancerous lesions such as leukoplakia, erythroplakia and chronic traumatic ulcers, is performed to diagnose OSCC [6]. Additional endoscopy of the upper pulmonary tract is necessary, since oral cancer has a high risk in developing metastasis in the head and neck region and lungs. The severity of the disease is categorized following the Tumor, Nodes, Metastasis (TNM) staging system [6, 8, 9] of which the 8th edition of the AJCC Cancer Staging Manual is the most recent update, as shown in Table I and II [10].

Once OSCC is diagnosed, the best treatment will be discussed and planned by a multidisciplinary team. A preoperative planning is based on physical examination, computed tomography (CT) , magnetic resonance imaging (MRI) and ultrasonography (US) , which enable accurate evaluation of local spread, invasion of surrounding tissues and lymph node involvement [9, 11]. The shape and growth pattern of tongue cancer vary among patients [3]. Also, the feasibility of tumor-free resection margins and the postoperative quality of life are discussed by the multidisciplinary team.

Treatment of OSCC is based on patient's individual circumstances and TNM cancer staging. For resectable tumors (Tis/T1/T2), partial glossectomy is the most common treatment, if the general condition of the patient is sufficient [12]. More advanced tumors (T3/T4), with vessel and perineural invasion and lymph node involvement require additional postoperative radiotherapy and/or chemotherapy [6, 9]. Additional postoperative therapy is advised as well in case of positive or close resection margins, since increased chance of disease recurrence and poorer survival are known consequences of inadequate removal of the tumor [13].

Sutton et al. and Alicandri-Ciuflli et al. state that there is no consistency about the distance of a clear resection margin [13, 14]. Most literature considers >5 mm between invasive carcinoma and the surgical margin as a safe margin, 1-5 mm as close and less than 1 mm is a positive or involved resection margin [2, 9, 13, 15, 16]. These distances are substantiated by The Royal College of Pathologists [17].

After treatment, patients will be followed up frequently since any recurrences or second malignancies could occur. Wang et al. show that the recurrence rate of OSCC is 32,7% [18]. Additionally, another study shows that the 5 year survival rate is significantly higher in OSCC patients without recurrence (about 80-90%) than patients with recurrences (about 30%) and the overall survival rate is 60% [5]. Other important prognostic factors are tumor size, lymph node involvement [4, 6, 9] and the status of resection margin[13, 19]. Since none of the factors can influence the prognosis of OSCC alone, all prognostic factors should be taken into consideration when determining the prognosis of a patient [5].

The survival rates of early detected tumors (T1/T2) is >70% , which is higher than the survival rates for late discovered tumor (<43%) [2]. Also, Nóbrega et al. found that patients with tumors at stages T1 and T2 and absence of lymph node involvement at initial diagnosis have a higher survival rate [4]. A review of Jadhav et al. show that a disease-free period of three years was higher in patients with a tumor diameter <2 cm (84%) compared to patients with tumor diameters >2 cm (52%). Besides the diameter, an invasion depth of >5 mm corresponds to more cervical metastasis (64,7%) than tumors with a depth of invasion <5 mm (5,9%). This can be explained by the fact that lymphatic channels are present in deeper tissue and function as a pathway for cervical metastasis [5].

The involvement of lymph nodes is strongly correlated with metastasis, a lower survival rate and an increased risk of local recurrences. Especially, it is found that the macroscopic extension of the extracapsular spread of the lymph node is 1.5 times more likely to develop local recurrences than patients with microscopic extension [5]. In a study of Mourad et al. tumor thickness and invasion depth, assessed pre-operatively by MRI, seem to be important prognostic factors of metastasis in cervical lymph nodes [3].

Sutton et al. show that a narrow resection margin is related to poorer prognosis in terms of disease recurrence and survival [11, 19] in spite of radiotherapy to the primary site [13]. In addition, Jadhav et al. show a 5 year survival rate of 69% in patients with clear margins compared to 38% with involved margins [5]. On the other hand, a study of Weijers et al., excluding patients with positive margins as well as patients with epithelial dysplasia in the mucosal resection margins, found no significant difference in the development of local recurrence within patients of which the specimen contained tumor cells at <5 mm from the deep surgical margin compared to specimens containing tumor cells at >5 mm [19].

Table I. T Category for oral cancer in the 8th edition of the AJCC TNM staging system [10]

T Category	T Criteria
TX	Primary tumor cannot be assessed
Tis	Carcinoma in situ
T1	Tumor 2 cm, 5 mm depth of invasion (DOI) (DOI is depth of invasion and not tumor thickness)
T2	Tumor 2 cm, DOI >5 mm and 10 mm or tumor >2 cm but 4 cm, and 10 mm DOI
T3	Tumor >4 cm or any tumor >10 mm DOI
T4	Moderately advanced or very advanced local disease
T4a	Moderately advanced local disease: (lip) tumor invades through cortical bone or involves the inferior alveolar nerve, floor of mouth, or skin of face (ie, chin or nose); (oral cavity) tumor invades adjacent structures only (eg, through cortical bone of the mandible or maxilla, or involves the maxillary sinus or skin of the face); note that superficial erosion of bone/tooth socket (alone) by a gingival primary is not sufficient to classify a tumor as T4
T4b	Very advanced local disease; tumor invades masticator space, pterygoid plates, or skull base and/or encases the internal carotid artery

Table II. N Category for pathologic regional lymph nodes (pN) in the 8th edition of the AJCC TNM staging system [10]. ENE = extranodal extension.

N Category	N Criteria
NX	Regional lymph nodes cannot be assessed
Nis	No regional lymph node metastasis
N1	Metastasis in a single ipsilateral lymph node, 3 cm or less in greatest dimension and ENE-negative
N2	Metastasis in a single ipsilateral lymph node, 3 cm or less in greatest dimension and ENE-positive; or more than 3 cm but not more than 6 cm in greatest dimension and ENE-negative; or metastases in multiple ipsilateral lymph nodes,none more than 6 cm in greatest dimension and ENE-negative; or metastasis in bilateral or contralateral lymph nodes,none more than 6 cm in greatest dimension, ENE-negative
N2a	Metastasis in a single ipsilateral or contralateral lymph node 3 cm or less in greatest dimension and ENE-positive; or metastasis in a single ipsilateral lymph node more than 3 cm but not more than 6 cm in greatest dimension and ENE-negative
N2b	Metastasis in multiple ipsilateral lymph nodes, none more than 6 cm in greatest dimension and ENE-negative
N2c	Metastasis in bilateral or contralateral lymph nodes, none more than 6 cm in greatest dimension and ENE-negative
N3	Metastasis in a lymph node more than 6 cm in greatest dimension and ENE-negative; or metastasis in a single ipsilateral lymph node more than 3 cm in greatest dimension and ENE-positive; or metastasis in multiple ipsilateral, contralateral,or bilateral lymph nodes, with any ENE-positive
N3a	Metastasis in a lymph node more than 6 cm in greatest dimension and ENE-negative
N3b	Metastasis in a single ipsilateral node more than 3 cm in greatest dimension and ENE-positive; or metastasis in multiple ipsilateral, contralateral, or bilateral lymph nodes, with any ENE-positive

1.1 Problem Statement

Surgeons aim to remove the tongue tumor with a minimal resection margin of 5 mm, while preserving vital structures in the oral cavity [11, 19]. However, in the current clinical setting there is no intra-operative feedback providing any assessment of the resection margins [12]. Currently, these resection margins can only be confirmed post-operatively by histopathological assessment [13]. Therefore, real-time assessment of resection margins of tongue tumors is highly preferred to improve accurate resections [12, 15]. It is thought that this will decrease the functional disability of the tongue, as well as the need for secondary postoperative chemotherapy, radiotherapy and/or surgery. When developing such a real-time assessment tool, it is important to take into account the difference in margins during surgery and pathological assessment. It is known that the margin assessed pathologically is less than the margin aimed for during surgical resection because of tissue shrinkage caused by fixation, pathological processing [12, 13, 19] and intrinsic muscle contraction [20].

Miyawaki et al. studied intra-operative frozen section histological analysis of the resection margins of resected OSCC specimens [11]. Mentioned advantages of this technique are 1) readily anatomical orientation, 2) direct macroscopically observing the resection margin in cross-sectional plane, 3) possibility of reliable sampling and 4) reflecting the in-situ position from the specimen when resection margin is close or involved [11]. However, this method only assesses the resection margin in one plane of the specimen. An aggressive growth and invasion pattern of the tumor could result in positive margins elsewhere other than the cross-sectional plane. Secondly, a pathologist is required, at the operating theater, to perform the frozen section analysis, which is not feasible in most hospitals.

Additionally, frozen sections may suffer from sample errors [12].

Other studies evaluated the specimens using MRI. Since high soft tissue capability and definement of true extent, loco-regional involvement and tumor depth, MRI had been frequently used to assess carcinomas of the tongue [3]. Most studies specifically aimed at evaluation of resection margins and invasion depth used 1.5T MRI. Steens et al. tried to improve this and aimed to evaluate the feasibility and validity of ex-vivo 7T MR for evaluation of resection margins in tongue squamous cell carcinoma's (TSCC). He states that in tumors larger than >3 mm they expect to predict whether the resection margin is too small. However, the total time of preparation and MR examination was too long for clinical application. They implied to compare the MR with US, as it provides promising results about the analysis of resection margins of TSCC [12].

Brouwer de Koning et al. studied the correlation between MRI and US measurements of the greatest dimension and tumor thickness of OSCC. As a result, measuring the tumor thickness using US is more accurately for pre-operative tumor staging [21]. This technique is more applicable in the operating theater if implemented for assessment of resection margins. And compared to the MR study of Steens et al., US provides sufficient resolution for tissue determination in less scanning time [12]. Therefore, US seems to be a feasible technique to assess resection margins intra-operatively.

A different study of Brouwer de Koning et al., shows that US is feasible for intra-operative assessment of deep resection margins of TSCC. Advantages of US are 1) easy to implement in surgical workflow, 2) no specific training of the operator, 3) not time-consuming, 4) available in almost every operating complex and 5) not expensive [22]. However, scans of the specimens by US were made in only 2 axis. Only those slices could be examined and compared to histopathological analysis. Therefore, as recommended, three-dimensional (3D) scanning of the whole tumor volume by US would provide information about the deep resection margins in all slices surrounding the tongue tumor. A second recommendation is to create an operator independent setup. This setup will be more reproducible and ease the proceedings of the surgeon. Finally, advise is given to solve the orientation of the resected specimen to the resection field before implementing this technique [22]. This challenging problem of orientation of the specimen to the resection field is noticed by Hinni et al. as well [23].

1.1.1 Aim of the study

This study is part of a larger project focusing on the overall improvement of surgical precision of the TSCC resections. This involves pre-operative planning by different imaging modalities and intraoperative surgical guidance. This research can be seen as successor of Brouwer de Koning and will aims to provide surgical guidance during resections of TSCC using 3D US to minimize involved resection margins and preserve maximum functionality of the tongue.

The first objective of this research is to determine how to create a 3D model of a resected tongue specimen from US images. This objective includes the reconstruction method and variables which come across during the study. Secondly, the objective is to segment specimen and tumor from the 3D US volume fast and automatically. Based on these segments, the resection margin could be computed for the entire specimen. The final objective is to determine whether the computed resection margin in 3D US correlates to histopathology.

2 Background

2.1 Clinical Background

The tongue is located in the center of the oral cavity and partially within the oropharynx [24]. The tongue enables taste of food [25] and plays a critical role in speech, swallowing and breathing [24, 26–29].

It consists of three parts: the base, body and blade, of which the base is attached to the mandible and hyoid bone. The sulcus terminalis divides the tongue in an oral (anterior) and pharyngeal (posterior) part or the base, in which the tongue can easily be explained. At the end of this pharyngeal part is located the vallecula, which is the transition of smooth mucosa between the tongue base to the epiglottis [24]. The body of the tongue extents from the sulcus terminalis to the frenulum linguae. The part anterior of the frenulum linguae is the blade [29]. The tongue has two symmetrical muscular halves separated by the fibrofatty lingual septum, except for the blade [24].

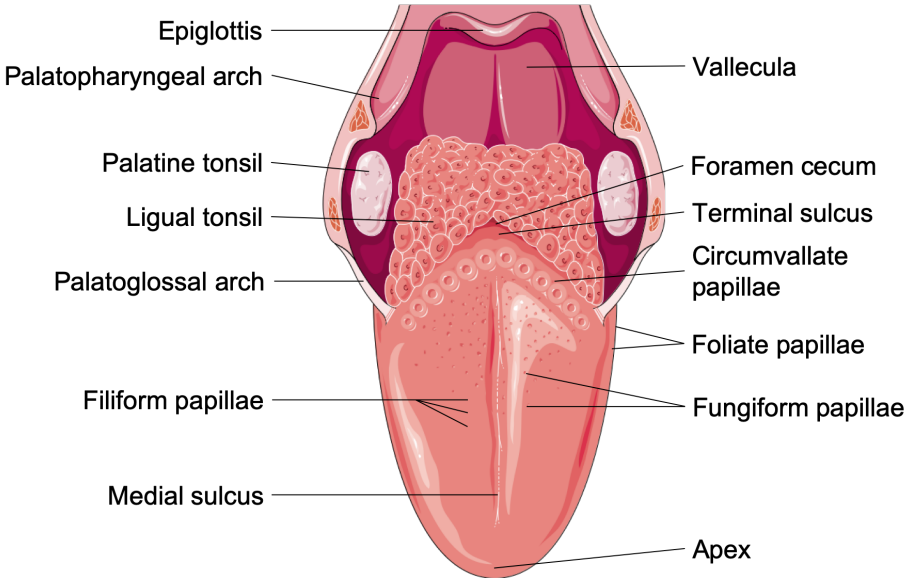


Figure 1. The major anatomical structures related to the tongue.

There are four types of papillae on the tongue, as shown in Fig. 1, of which three (fungiform, circumvallate and foliate) contain taste buds, which enable taste, while filiform papillae plays a role in eating, controlling the food [30] and providing information about temperature, texture and pain. The

filiform papillae is the most presented of all four papillae and only located at the oral (anterior) part of the tongue [25].

The taste buds in the papillae can distinguish five tastes of sweet, sour, salt, bitter and umami depending on 3 types of receptor cells [25]. Once a receptor cells in a papillae is activated, the sensory information will be transfered via innervation of three central nerves (CN). The filiform papillae are located at the oral part of the tongue and transfers signals through lingual branch of the trigeminal nerve (CN-V). The fungiform papillae, also at the oral part of the tongue, is innervated by chorda tympani branch of the facial nerve (CN-VII) [24, 25, 31]. The circumvillate and foliate papillae at the pharyngeal part sends information via the glossopharyngeal nerve (CN-IX) [25, 31].

Since the tongue derives during embryonic development from both ectoderm and endoderm linings, similar to the skin and gastrointestinal tract, the tongue contains a stratified squamous epithilium, as well as a moist mucosa like the gastrointestinal tract [30]. The embryonic origin of the oral tongue is from the first pharyngeal arch, ectoderm, while the pharyngeal tongue has an endodermal embryonic development originating from the third and fourth pharyngeal arch. The third pharyngeal arch ends up as the pharyngeal tongue, and the fourth pharyngeal arch provides the vallecula. Because of the different embryonic development, the vallecula is separately innervated by the internal laryngeal nerve [24].

The tongue exists of multiple muscles surrounded by mucous membrane [24]. Several studies state that the tongue is muscular hydrostat, which means that the biomechanics of the tongue are more similar to hydraulic systems relative to mechanical levers known for skeletal muscles [26–29]. Those muscular hydrostat structures change shape and position by deforming local regions [26] and maintain constant volume [28].

The muscles of the tongue can be divided in extrinsic muscles, which are one-sided attached to a bone and insert within the tongue, and intrinsic muscles, which has the origin and insertion in the tongue without any attachment to a bone [24, 28, 29]. Generally, the extrinsic muscles 1) genioglossus, 2) hyoglossus, 3) styloglossus and 4) palatoglossus are responsible for the position and movement of the tongue while the intrinsic muscles 1) superior and 2) inferior longitudinals, 3) vertical and 4) transverse bands alter the shape of the tongue [24, 27–29].

During movement and shaping of the tongue, the detailed contribution of each individual muscles as agonist, antagonist or stabilizer is unknown [29]. All muscles of the tongue are motor innervated by the hypoglossal nerve (CN-XII) [24, 31, 32] , which is subdivided into lateral and medial branches [27]. The lateral-hypoglossal nerve supplies the extrinsic styloglossus and hyoglossus muscles together with the intrinsic superior and inferior longitudinal muscles [27]. The rest of the intrinsic muscles (transverse and vertical) with addition of the genioglossus muscle are innervated by the medial-hyoglossal nerve. The palatoglossus is more essentially an palate muscle innervated by the pharyngeal plexus and forms therefore the only exception of all muscles [24].

Lymph drains from the base of the tongue to bilateral nodes in the neck and from the blade to submental nodes. A third of the drainage of the oral tongue is ipsilateral to submandibular and jugulodigastric nodes while the rest has lymph vessels to bilateral nodes [24].

The tongue is vascularized bilaterally by the lingual arteries originating from the external carotid artery [20]. Additional blood supply is supported by the facial artery and pharyngeal artery [24]. The lingual arteries run symmetrically and no transseptal anastomosis occur between the left and right side. During partial glossectomy, damage to both lingual arteries will result in necrosis of the tongue blade. The anatomical distribution of the lingual arteries should be taken into account when deciding the resection margins and to avoid intraoperative injury of the lingual arteries. Together with other oral structure such as tongue blade, foramen cecum, dorsal (superior) surface, the lingual arteries might be used as anatomical landmarks. From blade to base, the course of the lingual arteries bent into deeper tissue below the dorsal surface as it reaches the pharyngeal part of the tongue [20].

2.2 Technical Background

2.2.1 Two dimensional ultrasound

US imaging is based on the transmission of pressure waves through a medium and receiving the reflected wave with the transducer [33, 34]. US refers to the high frequencies above human hearing (>20000 Hz) [34, 35] caused by oscillations in pressure by piezoelectric crystals in a probe [33]. By applying an electric current with a specific frequency, the piezoelectric crystals change length and create a pressure wave with the corresponding frequency [35]. Following, the electric current stopped and the crystals deform when receiving a pressure wave, which induces electric currents. By altering these phases, the vibrating crystals create longitudinal wave and the reflections contain the information about the medium it transmitted through, which is used to create an image [33, 34].

The velocity or speed of sound (v), frequency (f) and wavelength (λ) of the pressure wave are described by eq. (1). The speed of sound (in m/s) in which a pressure wave transmits through tissue depends on the tissue properties [33]. The frequency (in $\frac{1}{s}$), or oscillations per second [34, 35], depends on the chosen electrical current applied with the transducer [33]. Wavelength (in m) is the distance between two high pressure areas and is depending on both the velocity and frequency, as shown by eq. (1) [33]. This equation shows that the wavelength is inversely proportional to the frequency, so shorter wavelengths result in higher frequencies [34, 35].

$$v/\lambda = f \tag{1}$$

At each boundary, a fraction of the sound will transmit through adjacent tissues and the remaining part will be reflected in infinite directions including back to the probe [33, 35]. This attenuation of the original wave intensity is caused by 1) reflection, 2) absorption, 3) scattering [33] and 4) refraction [35]. 1) Each boundary reflects some of the wave, so boundaries in deeper tissue reached by the transmitted part, receive smaller portions of the original wave and reflect less strong. Time gain compensation increases the intensity of echos further away, to create a more even image [33–35]. 2) Wave energy is partially absorbed by the tissue, since particles start oscillating and produce heat due to friction [35]. The amount of absorption is dependent on the tissue [33]. 3) The wave could scatter in all directions when the pressure wave meets boundaries not perpendicular to the wave's path [33] and at structure much smaller than the wavelength. This refracted part will not reach the transducer probe and is therefore a loss of energy. 4) At the boundary, transmission of the wave energy in a different direction than the original wave, is called refraction. Due to this new direction, the reflection will be in a different direction as well, resulting in no receiving of the reflection [35].

The amount of attenuation depends on the difference in acoustic impedance (Z) of the two tissues [34, 35]. As shown in eq. (2), Z depends on the tissue density ρ and speed of sound v in that tissue [34, 35].

$$Z = \rho \cdot v \tag{2}$$

A large difference in material density is present when comparing soft tissue with air or bone, and so $Z1$ and $Z2$ are different [35]. Difference in Z means attenuation in transmission energy caused by reflection. The sound wave will not reach deeper tissue resulting in dark areas of no information in the ultrasound image (posterior acoustic shadowing) [34, 35]. In clinical practice, collagen and fat are demonstrated as hyperechoic tissues and muscles and fluids such as blood and urine as hypoechoic [33, 34]. In this study, based on the finding van Brouwer de Koning et al. [22], it is assumed that the difference in tongue muscle tissue and TSCC is visible.

To create an 2D image containing many pixels (brightness or B-mode), the location of the reflection and its corresponding intensity is required. The location of the boundary is determined by measuring the elapsed time between a created pulse and the corresponding received sound reflection, assuming that the overall speed of sound in human soft tissue is 1540 m/s [33, 35, 36]. A short travel time corresponds to a location close to the probes surfaces, represented by the upper pixels of the image

and vice versa. The amplitude of the receiving wave determines the intensity the pixel displayed [33, 35]. Mostly, a scale of 256 shades of gray is used to differentiate intensity as a result of different reflections. The image could be optimized by keeping as many shades of gray as possible [33].

The spatial resolution of an ultrasound image is divided into two types: axial and lateral. The axial resolution is the ability to differentiate two points in the direction of the wave [33, 35]. To visualize a structure, it is required that the structure is larger than multiple wavelengths [35]. Therefore, smaller wavelengths come with higher frequencies and so a high frequency probe is required to image smaller structures. The lateral resolution is the ability to distinguish two points at equal distance in two different directions [33, 35].

The best resolution is where the beam converges, also called the focal zone or focal range. This zone is adversely related to the frequency due to attenuation, which in tissue ranges from 0.5 to 1.1 db/cm/MHz [33]. In practice, high frequencies (and short wavelengths) provide more detailed images as lower frequencies reach deeper tissues and expanded the field of imaging [33–35]. Therefore, 5-12 MHz linear probes are used for high resolution assessment of superficial tissues [34], while 2-5 MHz probes are deployed for imaging deeper tissues [35].

2.2.2 Three dimensional ultrasound

With the additional third dimension in US multiple advantages arise. Images to diagnose patients are completely reproducible. 3D US provides a wider ranges of scan planes to analyze because of a reconstructed volume [36]. Even a panoramic view of the region of interest can be made to help surgeons locate their instruments in the target area. 3D US is not dependent on the expertise and knowledge of the operator any more, since it provides full understanding of the distribution of anatomical structures. Obtaining the shape and location of the region of interest (ROI) with 3D US is definitely improved as it enables fast and accurate diagnostics. Finding a precise location during surgery is limited by 2D US, while this extra dimension can visualize a full 3D target area in real-time [37].

In this section, the necessary steps of 3D ultrasound will be described in general. Based on the available techniques at Netherlands Cancer Institute (NCI), those will be explained in further detail.

3D US images can be made in four different ways: 2D array probes, mechanical 3D probes, mechanical localizers and freehand scanners [37].

Volumetric scanning with 2D array probes is possible by steering the sound wave in both azimuth and elevation dimension. The diverging sound wave produced by the 2D array transducer has an pyramidal shape and the reflected waves are processed into integrated 3D images. Adjusting phased array delays serve to steer and focus on the ROI, so the probe could be held at the same location while scanning [37].

A 3D image could also be made with mechanical 3D probes which contain motorized linear transducers acquiring a collection of 2D images. This mechanical 3D probes can rotate, tilt and translate across the target area. Linear scanning acquires parallel images at a consistent slice-distance by adjusting the frame rate. This results in a non-isotropic resolution: in the scanning direction equal to conventional 2D images and in the translation direction dependent on the elevation of the probe [37].

Tilting scanners are fixed on the skin of the patient and obtain a fan of images separated by a specified angle depending on the tilting speed and imaging frame rate. The resolution of this technique is non-isotropic as well, as it degrades when the separation angles increases [37].

To scan the ROI with a rotating scanner, it has to be held statically while it rotates around the central axis of probe. This means that the resolution is depending on the angular distance between images resulting in non-isotropic resolution again. Depending on a convex or linear probe, the resulting 3D images will have a conical or cylindrical shape [37].

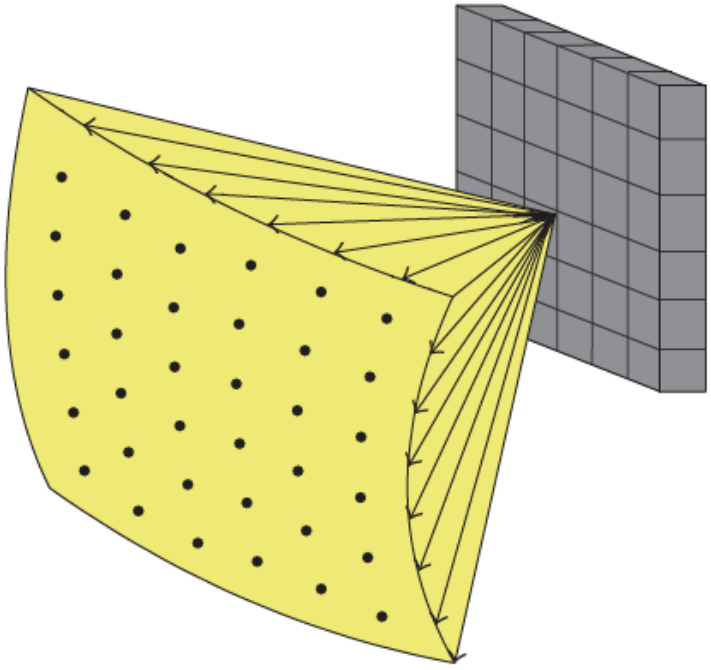


Figure 2. Schematic drawing of imaging a volume by a single crystal in a 2D array probe [37].

Mechanical localizers are motorized and move linearly, tilt and rotate as well as mechanical 3D probes. Instead of a internal motor, mechanical localizers use an external holder to acquire a collection of 2D images using a 1D transducer. To reconstruct the 3D images, the relative position and orientation of the 2D images have to be recorded accurately. However, these localizers are inconvenient since they are heavy and large [37].

A 3D scanning technique more convenient for application is the freehand scanner. This technique provides operators the opportunity to scan the ROI in any direction and position to get optimal anatomical orientation. For reconstruction of 2D images into 3D images, position and orientation parameters are required for each 2D slice [37]. Different device tracking mechanisms are available in clinical setting. However, due to the small operational area during tongue surgery, this research is limited to electromagnetic (EM) tracking.

This tracking mechanism requires a EM transmitter and a receiver located at the probe. Based on EM signal, the required position and orientation of the probe can be calculated which enables 3D reconstruction. The advantage of EM tracking is the relative small sensors and no required line of sight. On the other side, EM interference and metal objects will distort the EM signal and reduce the tracking accuracy [37].

Object reconstruction is an inevitable step and the accuracy and speed are significant for real-time 3D imaging. The main reconstruction algorithms are voxel based methods (VBM) and pixel based methods (PBM) [37]. These reconstruction algorithms will be described in details at chapter 4 within the limits of the available methods of this study.

3
Research Questions

Is 3D ultrasound a feasible technique for intraoperative assessment of surgical resection margins of tongue squamous cell carcinomas?

Sub research questions:

- *What is the impact of the reconstruction algorithm, ultrasound frequency, acquisition method, stabilizer and observer on the accuracy of 3D ultrasound reconstruction?*
- *Is deep learning a feasible technique for fast automatic intra-operative multi-class segmentation of 3D US volumes of resected tongue specimen and tumor?*
- *What is the correlation between the resection margin in tongue tumor specimens assessed by 3D ultrasound and histopathology?*

4

The influence of multiple variables on accurate 3D US reconstructions, a phantom study

4.1 Introduction

The first step to know if 3D US is a feasible and accurate technique for intraoperative assessment of surgical resection margins of tongue squamous cell carcinomas, is to obtain a better understanding about 3D US. This chapter focuses on the acquisition of US images and the reconstruction methods to create 3D US volumes.

US acquisition contains two phases: the acquisition of 2D grey scale US images (B-scans) itself and the reconstruction from 2D B-scans into a 3D US volume [38]. The acquisition of 2D images can be performed by freehand or motorized in a linear, tilt or rotational manner. Freehand scanning results in a non-uniformal distribution of the distance and orientation of the slices relative to each other [39]. The motorized acquisition result in a more regular pattern of the 2D image position. In the final 3D reconstruction, both freehand and motorized method are very similar [40]. The positioning data to reconstruct the 3D volumes could be obtained optically, mechanically, electromagnetically and acoustically [40]. In this study, the position and orientation of the 2D US images relative to the other images is based on the electromagnetic (EM) navigation.

The acquisition and reconstruction phase both include variables which influence the accuracy of the output 3D US volume [40]. The US acquisition phase has variables involving the experimental setup and the settings of the US system. The reconstruction phase has a computational variable, as the reconstruction could be created by several algorithms. It is important to keep in mind that artefacts may be introduced by these algorithms and thereby reduce image quality [39]. The differences between reconstruction algorithms are explained below.

4.1.1 US reconstruction algorithms

Most algorithms are simple and quick so the physician can visualize the 3D US volumes immediately after acquisition [38]. Reconstruction algorithms provided by CustusX, an open-source research platform for image-guided interventions [41], could be divided into two groups, the pixel based algorithms and voxel based algorithms. For both algorithms, a voxel grid has to be filled with values from the acquisition image planes [42]. The algorithms differ in computational time, which, in case of clinical implementation, have to be taken into account. For a wider range of possible algorithms, reference is made to the review of Solberg et al., who describes the benefits and drawbacks of several reconstruction algorithms [40].

Pixel based methods (PBM)

Pixel Nearest Neighbor (PNN) reconstruction iterates over each acquired image plane. Within the image plane, for each pixel the algorithm finds the nearest voxel in the voxel grid and assigns the pixel value to this voxel [37, 38]. Normally, if the voxel already contains a value, multiple contributions are averaged. However, assigning the most recent, the first, or the maximum value is possible as well. After iteration over all image planes, the empty voxels in the voxel grid are filled with a value from neighboring voxels [38, 42, 43]. Several methods are available such as the average of nonzero pixels in 2D planes, average, median, or maximum of nonzero voxels in a 3D local neighborhood or interpolation of the nearest voxels. However, this step may not be necessary when the distance between slices is small enough [40]. Artefacts as a result of this two-step method are visible as a boundary between the voxels with assigned pixel values and voxels filled from the second step [38].

Voxel based methods (VBM)

Many voxel based reconstruction algorithms are available for application. The first is Voxel Nearest Neighbor (VNN) which iterates over the output voxel grid. For each voxel, the nearest image pixel is found and assigned to the voxel [37]. This is a fast method, since the nearest pixel lies on a line normal to the nearest image [38]. If there is no image plane within the maximum radius around the voxel, it is left empty. [40, 42]

A more complex variant of this algorithm is the Voxel Nearest Neighbor2 (VNN2) which does not assign the value of the nearest pixel, but takes all image planes within radius R and assigns a distance-weighted average of the nearest pixels from all images planes to the voxel.

The distance weighted (DW) reconstruction finds the closest 2D pixel on each side of the voxel and applies a bilinear interpolation of the four surrounding pixels before assigning the voxel value [40, 42].

Finally, this distance weighing algorithm is also available with an additional varying Gaussian filter, the anisotropic reconstruction algorithm. This adaptive algorithm keeps details in high-frequency regions and cancels out noise [42].

For all algorithms, artefacts due to reconstruction could occur and can be observed in the voxel array [38].

As described, 3D US acquisition is subject to multiple variables. To know if 3D US is a feasible technique, the impact of these variables on the accuracy of the output volume needs to be investigated. High accuracy with respect to contrast, signal and noise is preferred to differentiate the materials in the US image. Or in clinical setting, differentiate the specimen from tumor tissue. This phantom study investigates the impact of the reconstruction algorithm, acquisition method, US frequency, stabilization rails and observer on the accuracy of US acquisition and reconstruction. The goal of this phantom study is to experimentally substantiate the preferred conditions to acquire accurate 3D US reconstructions of resected tongue tumor specimens.

4.2 Research question

What is the impact of the reconstruction algorithm, ultrasound frequency, acquisition method, stabilizer and observer on the accuracy of 3D ultrasound reconstruction?

Sub research questions:

- *Which US reconstruction algorithm results in the most accurate 3D US reconstruction?*
- *Which sweep-method results in the most accurate 3D US reconstruction?*
- *Which US frequency provides the best resolution for small target volumes?*
- *What acquisition method, freehand or stabilized scanning, provides the most accurate 3D US reconstruction?*
- *What is the inter- and intraobserver variability when performing US acquisition?*

4.3 Method

4.3.1 Materials

A phantom study was designed to evaluate the influence of multiple variables on the 3D reconstruction volumes. To acquire computed tomography (CT) data of the phantom, a Toshiba Acquilion (Canon Medical Acquilion series, Tokyo, Japan) was used. This study included an old prostate phantom (CIRS 053L, Norfolk, USA), previously used to practice prostate biopsy assisted by transrectal ultrasound. The experiments were performed using a BK5000 Ultrasound system (BK Medical, Denmark) in combination with a small intraoperative convex (5-14 MHz) transducer. In addition to the transducer, a 3D printed holder is attached to the transducer. An EM tracking system (Aurora, NDI, Canada) with two associate sensors (six degrees of freedom) were used to provide relative position and orientation coordinates between the two sensors. CustusX enabled combining the 2D US images from the BK US system and the corresponding position and orientation coordinates from the NDI system into 3D US volumes. These reconstructions could be performed separately from the acquisitions. One of the experiments required a stabilization rail, customized to the dimensions of the transducer including a 3D printed holder for one of the EM sensors.

Further image segmentation and quantification of both CT and US data was performed using 3DSlicer, an open source software platform for medical image informatics, image processing, and three-dimensional visualization [44] and customized code written in Python.

4.3.2 Setup of materials

Prior to all the experiments, a specific setup was build which is shown in Fig. 3. First, one EM sensor, the reference sensor, was taped at the outside bottom of specimen scan unit. The NDI EM field generator was placed close to the bucket without any metal objects to minimize any chance of distortion. Then, the bucket was filled with water and the phantom was placed at the bottom of the bucket close to the reference sensor. The second NDI sensor was attached to the US transducer by the 3D printed holder with a clip-on mechanism. To scan the phantom, the transducer was held below the water surface without touching the phantom. By moving the transducer, the whole phantom was observed. During the experiments, variables of the method and US settings were changed, as shown in Table III, to investigate the impact of these variables on the accuracy of reconstruction.

4.3.3 Data Acquisition

Because the prostate phantom is outdated, the dimensions could be subject to change. Therefore, the actual dimensions of the phantom were scanned by CT (current: 100mA, exposure: 100 mAs, voxel spacing: 0.305 x 0.305 x 0.5mm). These dimensions function as the gold standard dimensions of the phantom.

Table III shows the division of the phantom study into five experiments. The first experiment recorded one acquisition to investigate the reconstruction algorithms.

The second experiment acquired four different recordings to investigate the impact of the amount of sweeps and the acquisition time. Recording one was the already acquired recording from experiment one, scanned in a single sweep of 10 seconds. Recording two was scanned in a single sweep of three seconds, the third recording was scanned in two sweeps (back and forth) of two times three seconds and the last recording was scanned using three sweeps (back-forth-back) all consisting of three seconds with a total recording time of nine seconds.

It is interesting to see how 2D US images recorded by double or triple sweeps were reconstructed in 3D volumes, because the reconstruction algorithms search for close images. Therefore, the analysis

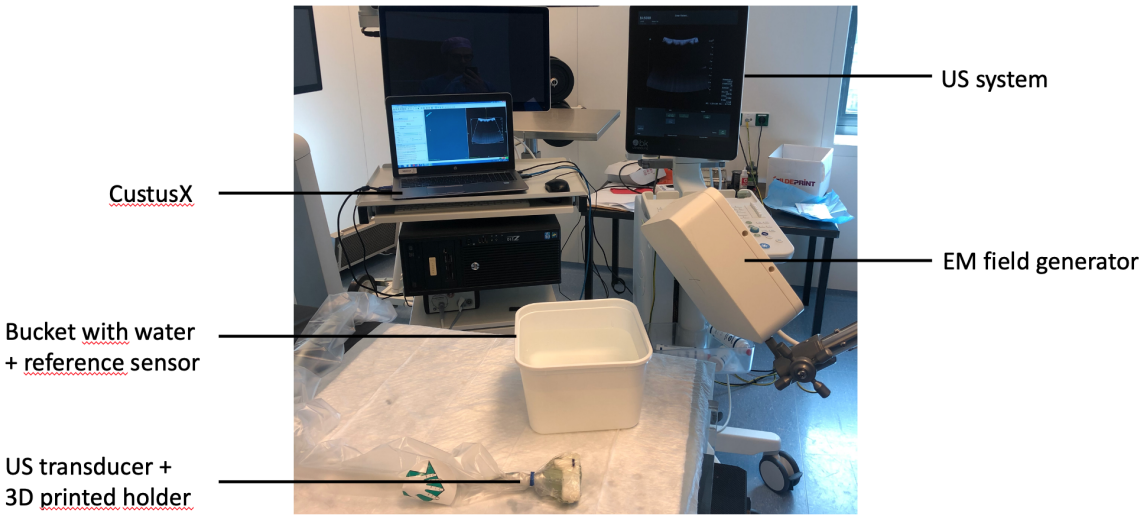


Figure 3. The setup of 3D US acquisition of the phantom.

Table III. The investigated variables changing for each experiment highlighted in red.

Experiment	Reconstr. Algorithm	Sweeping method	Transducer frequency in MHz	Freehand vs Rails	Observer	Width
1	PNN VNN VNN2 DW Anisotropic	Single, 10 s.	10	Freehand	A	100
2	PNN VNN VNN2 DW Anisotropic	Single, 10 s. Single, 3 s. Double, 6 s. Triple, 9 s.	10	Freehand	A	100
3	PNN	Single, 10 s.	5 7.5 10	Freehand	A	100
4	PNN	Single, 10 s.	10	Freehand Rails	A	100
5	PNN	Single, 10 s.	10	Freehand	3x A 3x B 3x C	100

of experiment one was enlarged by investigating the reconstruction algorithm as well for the three recordings acquired in a single, double and triple sweep.

The next experiment was to obtain more insight in the transducer frequency and the resulting image resolution. The transducer frequency for the three different acquisitions was 5, 7.5 and 10 MHz, respectively.

The fourth experiment recorded two acquisitions to investigate the impact of freehand scanning compared to stabilized acquisition. The first recording was scanned by the freehand method followed by a second recording assisted by a stabilization rail.

The final experiment focused on the inter- and intraobserver variability of acquisition. The recordings were repeated three times resulting in three recordings for each observer, so a total of nine recordings.

4.3.4 Data processing and analysis

The recordings of the first and second experiment were reconstructed five times into a 3D volume, respectively by the provided reconstruction algorithms, as described in Section 4.1.1: PNN, VNN, VNN2, DW and anisotropic in CustusX.

The image quality of the 3D US reconstructions was assessed for each experiment evaluating the signal to noise ratio (SNR) and contrast to noise ratio (CNR) , by respectively eq. (3) and (4) [45]. The standard deviation (SD) and mean pixel values (PV) were measured in a region of interest (ROI) and annotated as phantom (P) or background(B) as shown in eq. (4). The phantom ROI was determined in a homogeneous area in the phantom and the background ROI in the surrounding water. Both ROIs share the same spherical segmentation. The PV in US images handle a gray scale from 0 (black) to 255 (white).

$$SNR = \frac{PV}{SD} \qquad (3) \qquad\qquad CNR = \frac{2(PV_P - PV_B)^2}{SD_P^2 + SD_B^2} \qquad (4)$$

Secondly, the dimensions in the mid-slice of the spherical structure were measured along the Anterior-Posterior (AP) , Right-Left (RL) and Inferior-Superior (IS) axis. In case a clear measurement of the dimensions of the structure was not possible due to artefacts, a close estimation was noted.

Finally, the contrast along a scan line was analyzed by determining the height and Full Width at Half Maximum (FWHM) of the peak of the derivative of the pixel intensity at the transition from the phantom to the spherical structure. Herefore, the scan line was drawn in the mid-slice. However, the mid-slice was rotated so the longitudinal axis of the US wave was parallel to the x-axis. Along this scan line, a pixel intensity profile was taken and the derivative computed.

Based on the data analysis presented at Section 4.4, none of the reconstruction algorithms outperformed the others. On the other hand, the acquisition method using a single sweep did result in better numbers. So further recordings of experiments three to five were acquired in a single sweep and were reconstructed by the PNN algorithm, which performed the fasted reconstruction, and analyzed as described above.

4.4 Results

Three different operators performed a total of 19 acquisitions, as shown in Table III. Depending on the experiment, the acquisitions were reconstructed once or five times. A total of 34 3D US volumes were analyzed.

The phantom was used to function as practical simulator of prostate biopsies. Due to the performed biopsies, the material of the phantom is affected leaving hollow spots in the material. These spots - filled with air - result in artefacts in the US images. In addition to the damaged material, several markers were found in the phantom. In the CT volume, these markers are hyperdense structures resulting in artefacts spreading through the imaged volume. Because the spherical structure in the phantom was visibly affected in both the CT and US images, the SNR and CNR were computed in a different ROI that was homogeneous. Secondly, some dimensions were based on estimation instead of accurate manual measurement.

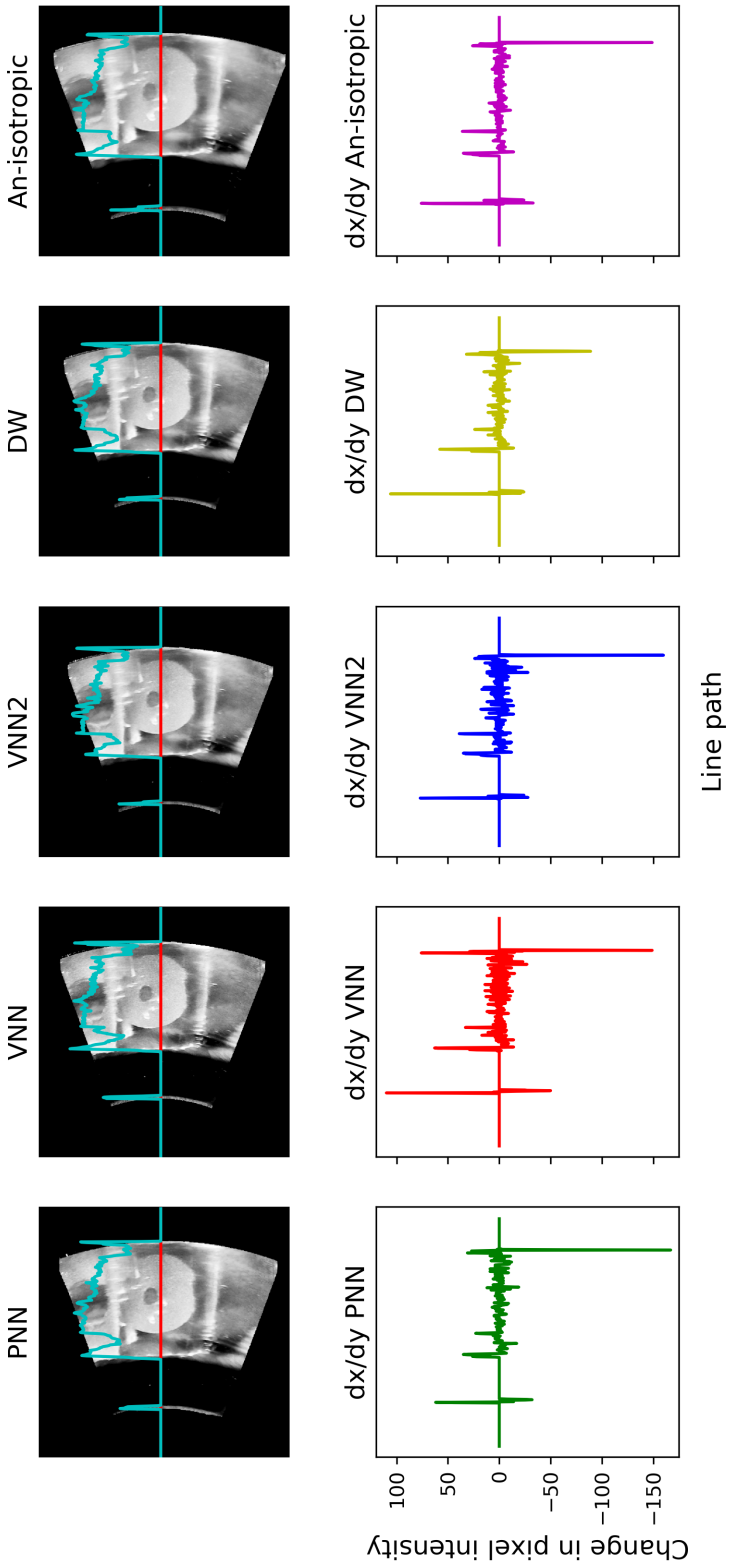


Figure 4. Experiment 1: The top five images show the 3D US mid-slice of the spherical structure in the phantom with the scan line plotted in red and the pixel intensity along this scan line plotted in cyan. The bottom five images show the derivatives of the pixel intensity along the scan line.

4.4.1 Experiment One

Starting with experiment one, the different criteria result in different preferred reconstruction algorithms. The SNR and CNR of all five experiments is shown in Table IV. For both ratios, the volume reconstructed by the anisotropic algorithm results in the highest SNR and CNR.

Table V shows the dimensions of the spherical structure in the phantom in three directions. The computed error (expressed in %) is a deviation between the US and the gold standard CT. Considering the actual measurements only, the anisotropic reconstruction algorithm results in the smallest error of 2.7%.

For the reconstructed volumes, the pixel intensity along a scan line crossing the spherical structure in the phantom of the mid-slice images are shown in Fig. 4. The height and the FWHM of the peaks of the derivatives at the transition from phantom to the spherical structure, are shown in Fig. 5. This graph shows that the smallest FWHM is achieved by the VNN reconstruction algorithm while the highest peak is provided by the VNN2 algorithm.

Table IV. The SNR and CNR for each variable of the five experiments. The numbers in **bold** represent the highest value within each experiment.

Experiment	Variable	SNR	CNR
1	PNN	22.4	979
	VNN	19.2	722
	VNN2	20.1	794
	DW	22.0	952
	Anisotropic	**27,9**	**1415**
2	Single, 10 s.	22.4	979
	Single, 3 s.	**25.8**	**1307**
	Double, 6 s.	14.2	398
	Triple, 9 s.	4.7	43.7
3	5 MHz	12.1	290
	7.5 MHz	**13.5**	**361**
	10 MHz	8.0	124
4	Freehand	**22.4**	**979**
	Rails	6.9	43.9
5: Observer A	Scan 1	20.0	771
	Scan 2	9.4	173
	Scan 3	**20.2**	**801**
5: Observer B	Scan 1	13.8	375
	Scan 2	3.6	25.9
	Scan 3	**20.0**	**795**
5: Observer C	Scan 1	6.6	86.1
	Scan 2	**7.6**	**89.0**
	Scan 3	6.6	53.2

Table V. The measured dimensions of the spherical structure in the phantom for each variable of the five experiments. The distance error compared to the CT is expressed as a percentage. The numbers in **bold** represent the lowest distance error within each experiment. Estimated dimensions are highlighted in red. AP = Anterior-Posterior, RL = Right - Left, IS = Inferior - Superior.

	Direction **CT**	**AP** *mm* (%) 37	**RL** *mm* (%) 43	**IS** *mm* (%) 34
Experiment	**Variable**			
1	PNN	35 (5.4)	38 (11.6)	31 (8.8)
	VNN	35 (5.4)	40 (7.0)	**32** (5.8)
	VNN2	35 (5.4)	39 (9.3)	31 (8.8)
	DW	35 (5.4)	37 (14.0)	**32** (5.8)
	Anisotropic	**36** (2.7)	**44** (2.3)	31 (8.8)
2	10 s. single	35 (5.4)	38 (11.6)	31 (7.8)
	3 s. single	35 (5.4)	40 (7.0)	**32** (5.8)
	6 s. double	**36** (2.7)	36 (16.3)	**32** (5.8)
	9 s. triple	**36** (2.7)	**42** (2.3)	31 (8.8)
3	5 MHz	**36** (2.7)	41 (4.6)	**32** (5.8)
	7.5 MHz	35 (5.4)	41 (4.6)	31 (8.8)
	10 MHz	**36** (2.7)	**43** (0)	**32** (5.8)
4	Freehand	35 (5.4)	38 (11.6)	31 (8.8)
	Rails	**36** (2.7)	**46** (7.0)	**32** (5.8)
5: Observer A	Scan 1	**36** (2.7)	**44** (2.3)	31 (8.8)
	Scan 2	35 (5.4)	46 (7.0)	**33** (2.9)
	Scan 3	**36** (2.7)	46 (7.0)	**33** (2.9)
5: Observer B	Scan 1	**36** (2.7)	**38** (11.6)	31 (8.8)
	Scan 2	33 (10.8)	33 (23.2)	31 (8.8)
	Scan 3	35 (5.4)	33 (23.2)	31 (8.8)
5: Observer C	Scan 1	38 (2.7)	**42** (2.3)	34 (0)
	Scan 2	**37** (0)	39 (9.3)	34 (0)
	Scan 3	38 (2.7)	40 (7.0)	34 (0)

The results of the expansion of experiment one are shown in Appendix 9, Table X. For each of the sweeping methods, the highest SNR and CNR was achieved when the 3D volume was reconstructed by anisotropic algorithm.

In Appendix 9, Table XI shows the dimensions for each reconstruction of the single, double and triple sweep acquisitions. In a single sweep acquisition, the VNN, DW and anisotropic reconstruction algorithm resulted in the lowest error of 2.7%. All reconstruction algorithm provided the lowest error (2.7%) in a double acquisition. In a triple sweep acquisition, the lowest error of 2.7% was the result of the PNN and VNN2 reconstruction algorithm.

For each acquisition method, the highest peak of derivatives was achieved by a different reconstruction algorithm, as shown in Appendix 9, Fig. 22. However, the FWHM is the smallest when reconstructed by the VNN algorithm for all acquisition methods.

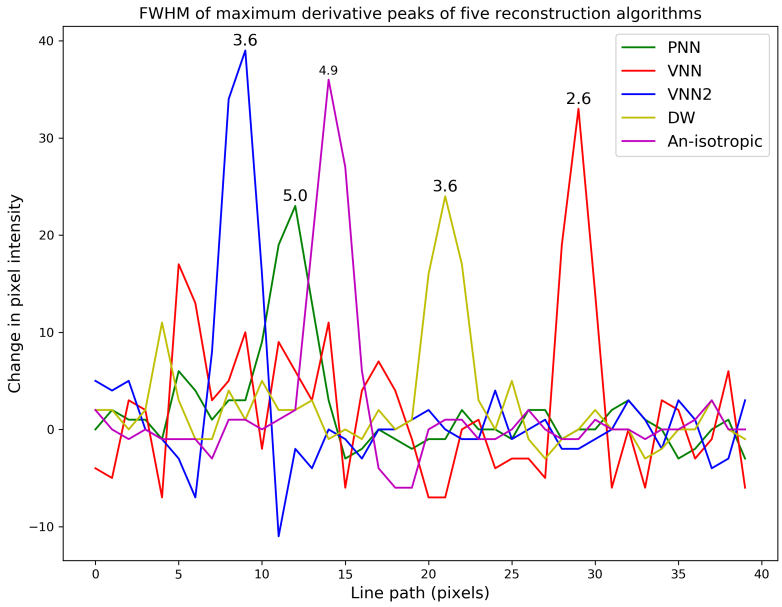


Figure 5. Experiment 1: The derivatives of the pixel intensity along a scan line at the transition from phantom to the spherical structure in five US 3D volumes with different reconstruction algorithms. The FWHM of the maximum peak is plotted at the top of the peak.

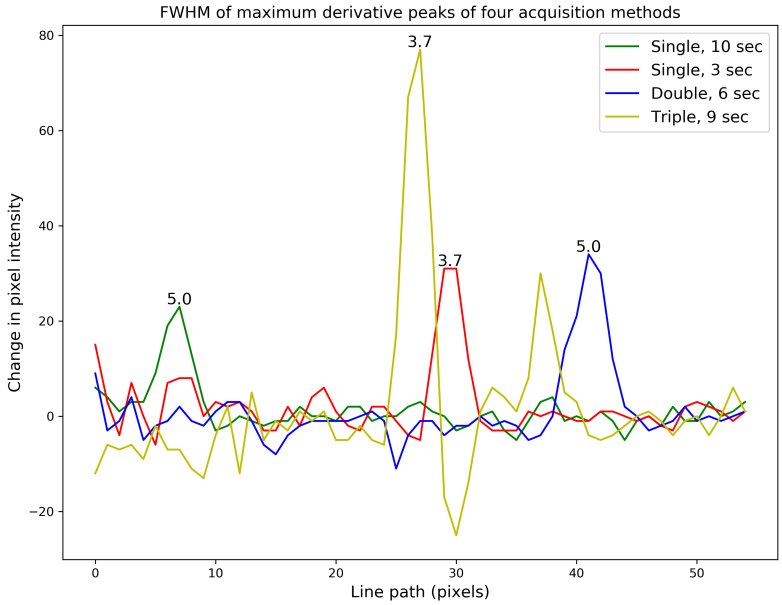


Figure 6. Experiment 2: The derivatives of the pixel intensity along a scan line at the transition from phantom to the spherical structure in four US 3D volumes with different acquisition methods. All 3D US volumes were reconstructed by the PNN algorithm. The FWHM of the maximum peak is plotted at the top of the peak.

4.4.2 Experiment Two

Experiment two was designed to investigate the influence of the sweeping method during US acquisition. Table IV shows the SNR and CNR for all acquisition methods reconstructed by the PNN algorithm. It is noteable that both single (10 and 3 sec.) acquisitions resulted in higher ratios compared to double and triple acquisition methods. In Appendix 9, Table X shows that the single acquisition methods provided higher ratios when reconstructed by the other algorithms as well. Only the anisotropic reconstruction resulted in similar ratios for double sweep acquisition in 6 seconds as the single sweep acquisition in 3 seconds.

The dimensions of the recordings of the four acquisition methods, reconstructed by the PNN algorithm are shown in Table V. Compared to the CT data, the lowest errors were achieved by the double and triple sweep acquisition methods. However, the dimensions of the four recording methods reconstructed by the other algorithms, shown in Table XI, were the lowest error for the single and double acquisition methods. So, the lowest error for each acquisition method is depending on the reconstruction algorithm.

Figure 6 shows that the triple sweep acquisition method provided the highest peak of the derivatives and the smallest FWHM of 3.7 pixels. Analyzing the acquisition methods reconstructed by the other algorithms, as shown in Appendix 9 Fig. 22, the highest peak for each acquisition method was again depending on the reconstruction algorithm. The smallest FWHM was provided by the single sweep acquisition method, except for the reconstruction by the anisotropic algorithm which showed the smallest FWHM for the triple sweep acquisition method.

Also, a second peak of the derivative occurred when the triple sweep acquisition method was used. In the mid-slice image of the triple sweep acquisition method, as shown in Fig. 7, some sort of checkerboard effect could be noticed which was not visible in reconstruction acquired by the single and double sweep acquisition methods. This checkerboard effect with multiple intensity transitions resulted in an additional peak of the derivative.

Finally, the single sweep for 3 seconds acquisition method resulted in the highest SNR and CNR. The dimensions did not show an out-performing acquisition method. The FWHM was the lowest when the volume was acquired by the single sweep acquisition method. Also, no checkerboard effect arose.

4.4.3 Experiment Three

Experiment three focused on the influence of transducer frequency on accurate 3D US acquisition. Table IV shows that a US frequency of 7.5 MHz provided the highest SNR and CNR. Looking at the dimensions in Table V, the lowest error of 2.7% (36 mm) in AP direction was achieved by 5 and 10 MHz. Figure 8 displays that the smallest FWHM (2.9 pixels) and highest peak of the derivative of pixel intensity at the transition from phantom to spherical structure was the result of acquisition with 10 MHz.

Summarized, each of the transducer frequencies provided the lowest or highest numbers at one of the analysis criteria.

4.4.4 Experiment Four

To understand the influence of freehand compared to stabilized US scanning experiment four was set up. Table IV shows that the freehand acquisition results in higher SNR and CNR compared to the acquisition assisted by a stabilization rails. However, the dimensions of the spherical structure show lower errors in all directions with a minimum 2.7% when the volume was acquired by stabilization rails, as shown in Table V. Focusing on the peak of the derivatives of the pixel intensity along a scan line, plotted in Fig. 9, acquisition assisted by stabilization rails resulted in a higher peak and a smaller FWHM compared to the freehand acquisition method.

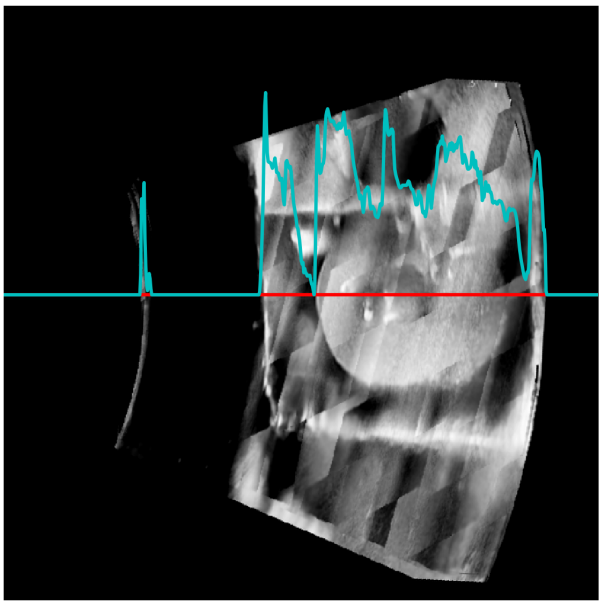


Figure 7. The 3D US mid-slice of the spherical structure in the phantom with the scan line plotted in red and the pixel intensity along this scan line plotted in cyan. The 3D US volume is acquired using three sweeps and reconstructed by the PNN algorithm. A checkerboard artefact is visible throughout the image.

4.4.5 Experiment Five

The final experiment investigated the observer variability of 3D US acquisition. As Table IV presents, large deviations between the three scans occur for both observer A and B. Observer C showed deviations between the scans, but substantially less than the other observers. Between the observers, the SNR and CNR showed major differences as the maximum SNR and CNR deviated from 7.6 to 20.2 and 89 to 801, respectively.

Table V shows that differences in errors of the AP dimension within the observer's acquisitions ranged from 2.7% (observer A and C) to 8.1% (observer B). Between the observers the maximum difference in error of the dimensions is 13.5% (4 mm).

The FWHM and height of the peaks of the derivatives, shown in Fig. 10, showed a large variety for observer B only. The smallest FWHM ranged from 2.7 to 4.3 pixels and the highest peak ranged from 60 to 90 between the observers.

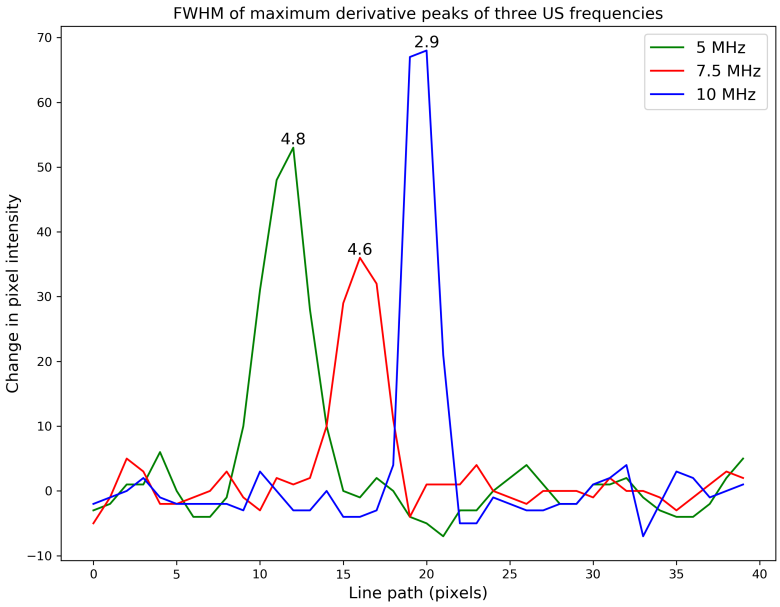


Figure 8. Experiment 3: The derivatives of the pixel intensity along a scan line at the transition from phantom to the spherical structure in three 3D US volumes with different US frequencies. The FWHM of the maximum peak is plotted at the top of the peak.

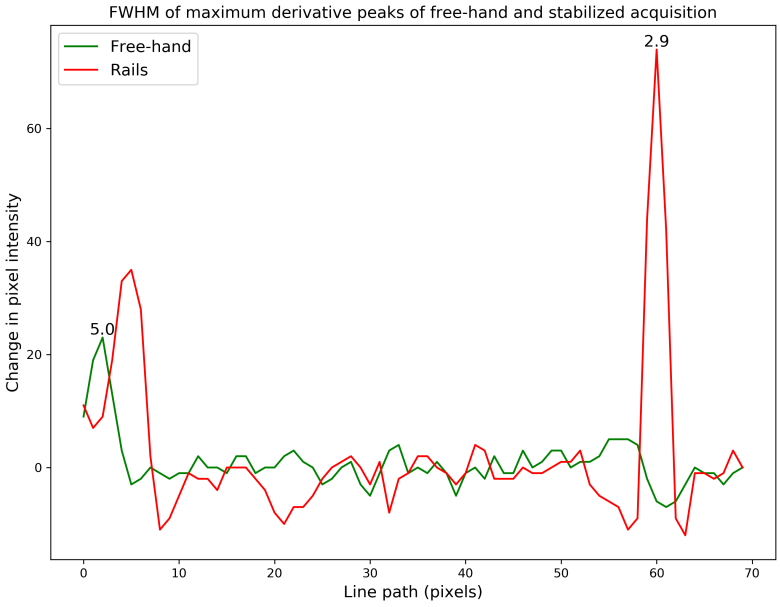


Figure 9. Experiment 4: The derivatives of the pixel intensity along a scan line at the transition from phantom to the spherical structure in two 3D US volumes acquired with and without stabilization. The FWHM of the maximum peak is plotted at the top of the peak.

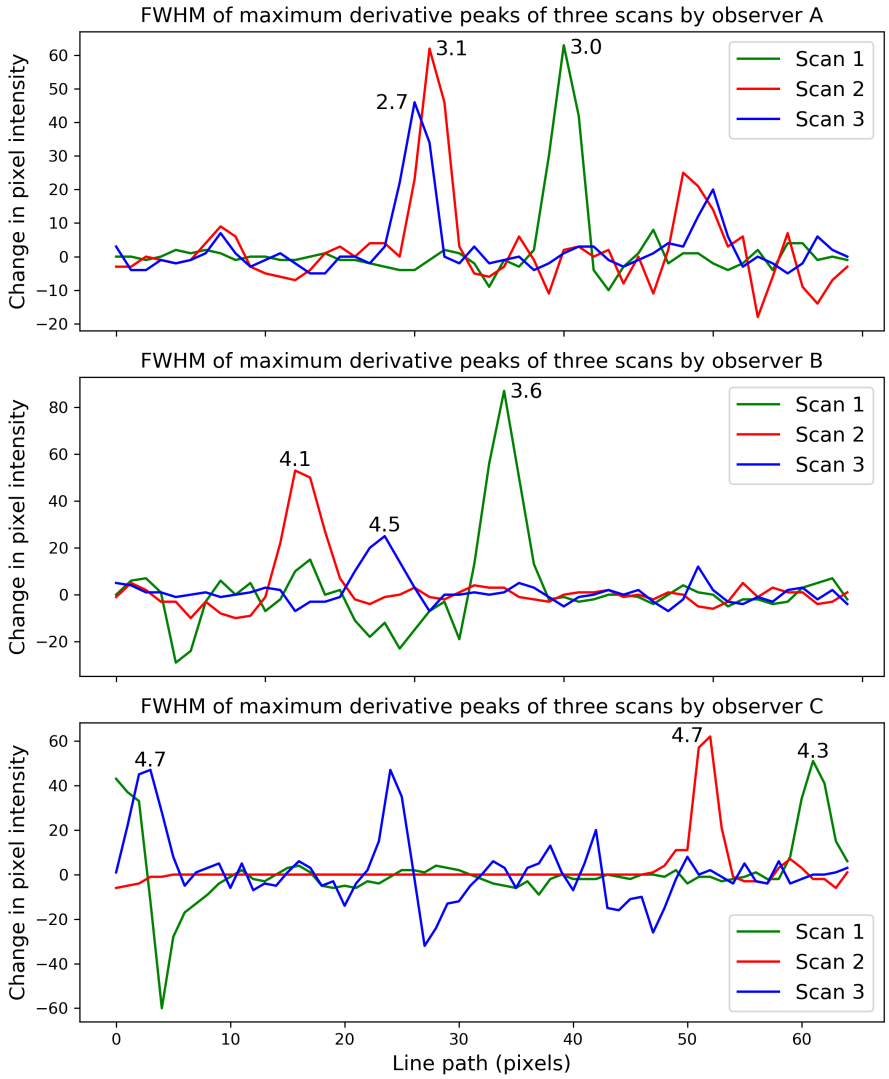


Figure 10. Experiment 5: The derivatives of the pixel intensity along a scan line at the transition from phantom to the spherical structure in 3D US volumes. The top, middle and bottom figures represents the acquired 3D US volumes by observer A, B and C, respectively. The FWHM of the maximum peak is plotted at the top of the peak.

4.5 Discussion

The present study was designed to determine the impact of the reconstruction algorithm, US frequency, acquisition method, stabilizer and observer on the accuracy of 3D US reconstruction. The main findings of the experiments were that 1) none of the reconstruction algorithms was out-performing the others, 2) the single sweep acquisition method provides better SNR, CNR and contrast along a scan line and reconstruction artefacts did not occur, 3) there was no US frequency resulting in better 3D US reconstruction, 4) acquisition assisted by a stabilization rails provides more accurate reconstructions and 5) there was an inter- and intraobserver variability during acquisition and reconstruction of 3D US volumes.

Inferring from these findings, the best acquisition method would be scanning using a single sweep method assisted by rails and performed by a single operator. Based on existing theory, the preferred US frequency is 10 MHz or higher, resulting in a higher spatial resolution. Contrary to the expectations, this study could not identify a reconstruction algorithm performing better than others. This suggests that acquisition variables do have a clear impact, while the reconstruction variable does not provide a preference with regard to the clinical problem of accurate 3D US reconstruction.

Compared to previous studies in the literature, 3D reconstructed volumes were created using similar tracking, scanning and reconstruction systems [38, 39, 46]. However, the impact of the variables of US acquisition and reconstruction was unknown. This study showed the impact of several variables on the acquisition and reconstruction of 2D US images into a 3D US volume.

The results from this study confirm the presence of checkerboard artefacts after reconstruction, which were not present during acquisition. As indicated by several previous studies [38, 39], these artefacts created due to reconstruction were visible in the voxel array. On top of that, this study showed the differences in image quality and contrast of 3D US volumes in various conditions of frequency, acquisition methods and reconstruction algorithms.

This study provided more details about the influence of US acquisition and reconstruction methods on the accuracy of 3D US volumes, where others focused on the clinical implementation of 3D US.

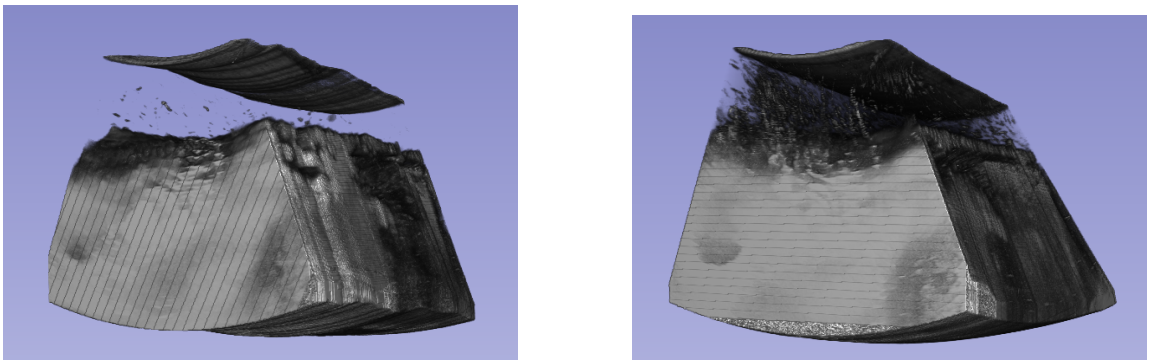
The first research question was set up with the aim of assessing the impact of the reconstruction algorithm on the accuracy of 3D US reconstruction. Interpreting the results from experiment one, the highest SNR and CNR at the anisotropic reconstruction could be explained by the description of algorithm in Section 4.1. The anisotropic algorithm applies a Gaussian filter on top of the distance weighted sum of the surrounding pixels to assign the voxel value [42]. This Gaussian filter reduces the amount of noise, creating a lower SD in the measured ROI resulting in the highest SNR and CNR.

Including all acquisition methods, the anisotropic reconstruction was not the best-performing algorithm for the criteria dimensions and FWHM.

The 3D volumes have a voxel size of 0.21 x 0.21 x 0.21 mm, so manually selecting a neighboring pixel twice due to intraobserver variability causes a deviation of 0.42 mm already. Therefore, it can be concluded that the presented outcomes from the dimension criteria are not reliable to substantiate statements of a best performing algorithm.

The FWHM was the smallest when reconstructed by the VNN algorithm for all acquisition methods. Compared to the other VBM, VNN is the only algorithm not averaging values but directly assigning the nearest image pixel. The averaging algorithms eventually create a smooth transition while VNN relies on the actual pixel value resulting in a steep transition, in other words: a small FWHM.

Eventually, the description of algorithms explains the differences in best performing algorithm for each criteria but, unexpectedly, a preferred algorithm for clinical implementation cannot be inferred from this study.



(a) Freehand **(b)** Assisted by rails

Figure 11. Rendered volumes of the 3D US reconstruction acquired by (a) Freehand and (b) assisted by rails. The 2D US images show less deviation relative to each other when acquisition is assisted by rails.

The second question in this study sought to determine which sweep method results in the most accurate 3D US reconstruction. Taking a closer look at the results of experiment two, the SNR and CNR were the highest in single sweep acquisition. The presence of checkerboard artefacts after reconstruction by multi-sweep acquisition results in areas of high and low pixel intensity within a select ROI. This large range of pixel intensity decreases the mean value and increases the SD in the ROI resulting in lower SNR and CNR.

The checkerboard artefacts affect the derivative of the pixel intensity as well. A second peak represents the transition to the area of high intensity within the spherical structure. Therefore, this study proved that acquiring US data using a multi-sweep method introduces artefacts and that a single sweep method is preferred in 3D US acquisition in a clinical setting.

The third question in this research was to know which US frequency provides the best resolution for small target volumes. The theory about ultrasound waves is that a higher US frequency provides better spatial resolution but reaches superficial tissue, while lower frequency reaches deeper tissue but has less spatial resolution. This phenomenon is confirmed by the results shown in Fig. 8, where both the highest peak and smallest FWHM are achieved with the highest US frequency, 10 MHz. The SNR and CNR increase by changing US frequency from 5 to 7.5 MHz. However, increasing to 10 MHz does not extend this relationship. It is hard to interpret these results based on an underlying mechanism. Based on the theory, a US frequency of 10 MHz would provide the best resolution for small superficial volumes.

The fourth objective of this study was to identify whether freehand or stabilized acquisition provides the most accurate 3D US reconstruction. It is expected that the stabilization rails prevent irregularities, such as shaking, during the US acquisition. If the rails succeeded cannot be derived from the SNR and CNR, dimensions or the contrast along a scan line. Comparing the rendered volumes in Fig. 11, freehand acquisition showed deviation in the alignment of 2D images due to shaking, which is not visible in case of acquisition assisted by rails.

However, a contradiction can be noticed between freehand acquisition, achieving higher SNR and CNR, and acquisition assisted by rails resulting in smaller FWHM and higher peaks of the derivatives at the transition from phantom to the spherical structure. So, the impact of stabilization rails while acquiring the US data is visible in the 3D rendered volume, but the analysis could not substantiate this finding.

With respect to the final research question, a variability was found in all criteria between the observers as well as within. During the acquisition all circumstances remained equal except the

trajectory, including shaking, and duration of acquisition. This is visible in length of the trajectory and the amount of slices, resulting in different voxel sizes (all around 0.2 x 0.2 x 0.2 mm) after reconstruction. A transition of a certain length could be visualized in a different amount of voxels. So, when less voxels represent change in pixel intensity, the FWHM becomes smaller. However, this mechanism could not explain the large difference in SNR and CNR. Since the inter- and intraobserver variability is certainly present, US acquisition in clinical setting should be performed by a single operator and always take into account the intraobserver variability.

The findings of this study are clinically relevant because intra-operative assessment of resection margins by 3D US could be contributing to improving image guided therapy. Then, surgeons could perform re-excisions within the current operation, resulting in less requirements of adjunct therapy. This study provided a better understanding of 3D US which is necessary for further research about the correlation between the resection margins assessed by 3D US and histopathology.

This study had some strengths and limitations. The first strength is that the setup of the experiments was created in a clinical setting. To implement this setup for clinical usage to investigate tumor margins, no changes have to be made. So any conclusion from this phantom study could be applied clinically and is not subject to environmental changes when scanning fresh specimens. Another strength of the method was to select a preferred acquisition method and reconstruction algorithm based on the first two experiments prior to acquisition of experiments three to five.

However, the old prostate phantom did introduce some limitations. The artefacts due to the damage of the phantom ensured that the SNR and CNR were measured at a different location and the dimensions were estimated instead of measured. Because of these estimations by manual selection, which are subject to observer variability, the dimensions became a weak validating criteria. Only the AP direction was measured, leaving 2/3 of the morphological dimensions useless. Also, the phantom did not provide small structures to analyze the spatial resolution for various reconstruction algorithms. Secondly, the contrast along a scan line is an analysis performed along a single line. The analysis shows a local contrast within a single image at one transition and is therefore not representative for the entire volume. However, the phantom contains a spherical structure meaning that the contrast should be equal along the scan line crossing the structure in all directions. Third, the setup was built to prevent occurrence of random errors. However, the bed at the operating theater and surrounding equipment are made of materials which could possibly distort the electromagnetic field. Incorrect tracking of the location and orientation of 2D images might be the result from these distortions.

In the end, some limitations were solved by changing the method and others were taken into account when interpreting the results. Eventually, we do not expect those limitations to have major impact on the conclusions as stated above.

4.6 Conclusion

3D US acquisition is subject to several variables during both US acquisition and reconstruction. The goal of this study was to experimentally substantiate the preferred settings to acquire an accurate 3D reconstruction of resected tongue tumor specimens. The experiments in this study have shown that this could be accomplished by an acquisition using a single sweep method assisted by rails, applying the highest US frequency possible and performed by one operator. Unexpectedly, a preferred reconstruction algorithm could not be found based on these experiments.

5

3D US volume segmentation by deep learning network UNet

5.1 Introduction

Surgeons aim to remove a tongue tumor with a minimal resection margin of 5 mm. However, Smits et al. show that 85% of the resection margins is <5 mm [47]. Another study shows a local recurrence rate of 32,7% in patients with oral squamous cell carcinoma (OSCC) [18]. Currently, histopathological assessment of the resection margins provides the only feedback about the accurateness of resection [13]. Therefore, intra-operative assessment of resection margins of tongue tumors is recommended to obtain a minimal resection margin of 5 mm [12, 15].

US proves to be a feasible technique to assess resection margins intra-operatively. Using 2D US, Brouwer de Koning et al. confirm a correlation of the margins assessed by US and histopathology. The whole specimen was examined. Based on the operator's decision a 2D intersection with the closest margin was chosen [22]. However, measuring the margins or segmenting the tumor for all 2D US images is time-consuming and not feasible during surgery. Therefore, 3D volume segmentation of specimen and tumor is required to improve accurate intraoperative assessment of resected tumor margins.

Segmenting of the tongue tumor is challenging given the different geometry and US pixel intensity in each specimen. Due to varying size, shape and anatomical location in oral cavity, the orientation of the specimen during US acquisition differs. Also, noise and the appearance of US artefacts could lead to difficult differentiation between specimen and tumor. Therefore, a radiologist is required during surgery.

Manual segmentation would be an obvious technique. However, it is often accompanied by observer variability. Furthermore, because it is time-consuming, this technique is not feasible during surgery. Semi-automatic algorithms are faster but require manual initialization. Stevenson et al. show that the random-walker algorithm provides equal results compared to manual segmentation [48]. Recently, studies show promising result by applying deep learning using convolutional neural networks (CNN) to segment the first trimester placenta in 3D US automatically [49].

Zhu et al. show that UNet [50], a fully CNN and widely used in medical image segmentation, can automatically segment tongue contour in US images for speech research. With the mean sum distance (MSD) as evaluation metric, which is the comparison of two curves without point-wise alignment, an accuracy of 3.5 pixels MSD was achieved [51]. Even when the training data size was small (1% of the original dataset), a reasonable accuracy of 5-6 pixels in MSD can be achieved by implementing data augmentation [51]. A CNN, trained on segmented data of tongue tumors volumes, could provide fully automated resected tongue specimen and tongue tumor segmentation in 3D US for the first time.

In this study, an open-source CNN, UNet [50], was used to segment resected tongue specimens

and tongue tumors in 3D US volumes. The labels for supervised learning were manually provided by a radiologist. Evaluation of the results was based on the Dice similarity coefficient (DSC). The goal of this study is to show as a proof of concept that deep learning is a feasible technique for fast automatic multi-class segmentation of tongue specimen and tumor in 3D freehand US volumes.

5.1.1 Semi-automatic algorithm

Prior to the implementation of fully automatic segmentation algorithms, some techniques require the manual input from an operator such as region growing and K-means clustering. The region growing algorithm is basically a region starting at an initial point and expanding towards neighboring pixel or voxels that have similar gray values until all surrounding pixels do not meet the properties of the region [52]. The initial point and growth settings have a large influence at the performance of this algorithm. These settings should be changed over and over when performing this algorithm repetitively with data being inconsistent. Also, segmentation by this algorithm is difficult and often inaccurate in case the target area consists of a wide distribution of gray values. Then, the algorithm stops without reaching the edges of the target area or over-expands due to the absence of clear edges as shown in Fig. 12.

K-means clustering is an unsupervised machine learning algorithm which separates an image or volume into K clusters. This iterative algorithm minimizes the distance between the image pixels [53]. Figure 13 shows an original 2D US image and the partitioning of pixels into 3, 5 and 8 clusters. Again, a large distribution of gray values results in inaccurate segmentation and requires several post-processing steps, for example selecting the cluster or the combination of clusters representing the target area.

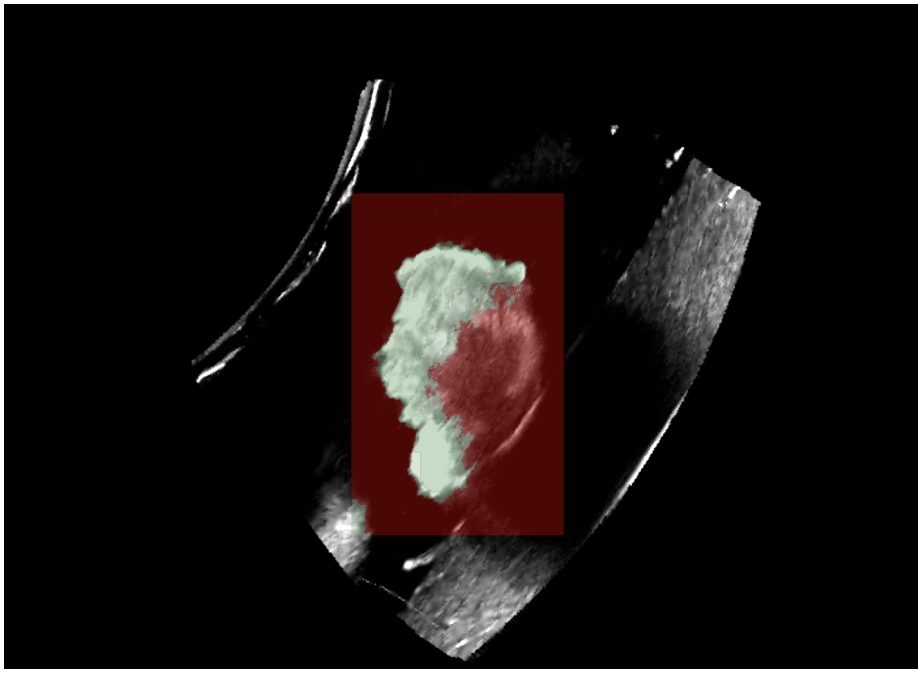


Figure 12. A 2D US image of a tongue specimen including a squamous cell carcinoma incorrectly segmented by the region grow algorithm. Due to the absence of clear edges of the tumor, the region grow algorithm expanded in the background until a preset maximum was reached, visualized by the red area. The specimen annotated in green shows that the algorithm actually could be limited at the edges of the specimen, where large deviations in pixel intensities are present.

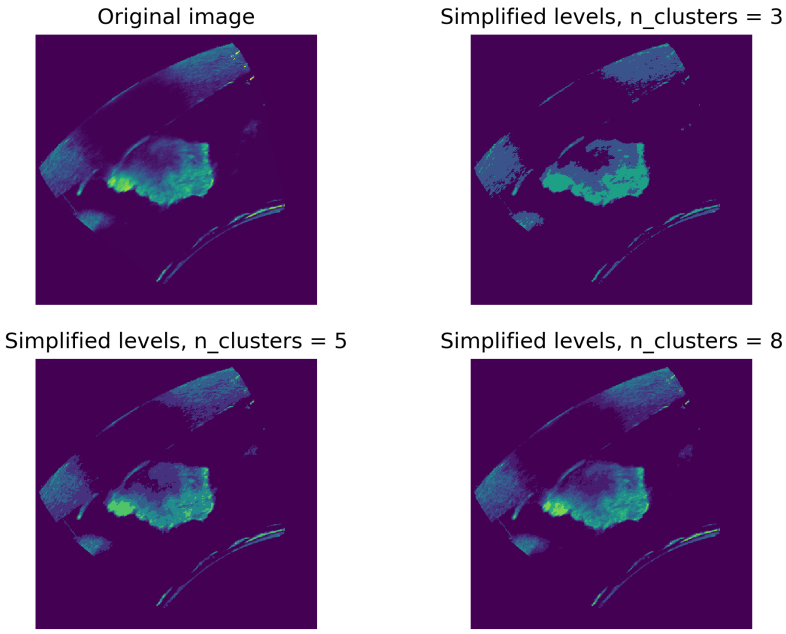


Figure 13. K-Means Clustering for 3, 5 and 8 clusters.

5.1.2 Deep learning

A well-known technique within artificial intelligence is machine learning (ML) , which constructs analytic algorithms to learn predictions from examples in data [54][55]. In case outputs of interest are known, the predictive models learn associations between inputs and outputs, which is called supervised learning. Modern ML, using a predictive model containing multiple hidden layers, is called deep learning (DL) [56]. DL can explore more complex (non)linear patterns in the data and enables to deal with increased volume and complexity of data such as medical images [55]. A popular DL algorithm for image segmentation and image classification is the CNN, which handles large numbers of inputs compared to traditional ML algorithms. A CNN is widely used for imaging analysis, since image data contains large numbers of pixels as inputs [55][56].

5.2 Research question

Is deep learning a feasible technique for fast automatic intra-operative multi-class segmentation of 3D US volumes of resected tongue specimen and tumor?

Sub research questions:

- *What is the variability between ground truth annotations by a radiologist within the same specimen?*
- *How accurate is deep learning in segmentation of tongue specimen and tumor in intra-operatively acquired 3D US volumes?*
- *What is the feasibility of intra-operative multi-class segmentation predictions by deep learning?*

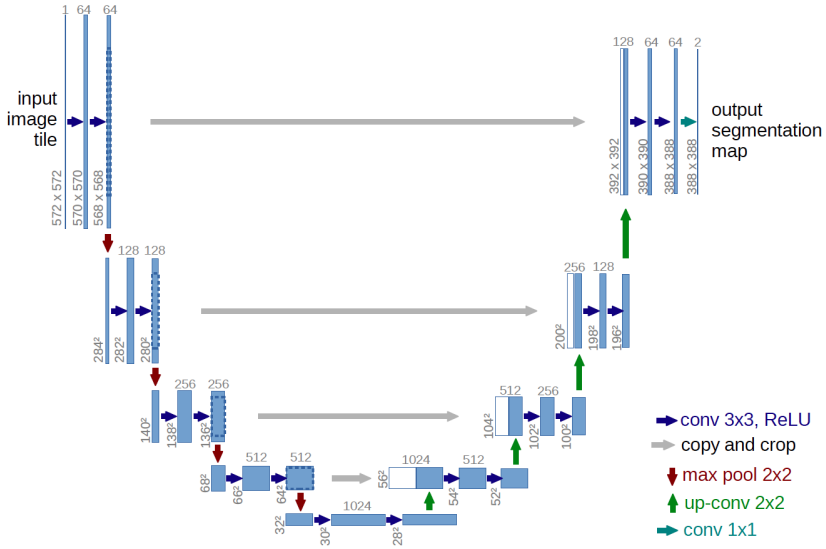


Figure 14. Visualization of the UNet architecture [50]. The numbers in gray represent example sizes of the input and output images, and the amount of features maps throughout the model.

5.3 Method

This section describes data acquisition, the adopted UNet architecture, the data pre- and post-processing steps, the training strategy and performed experiments.

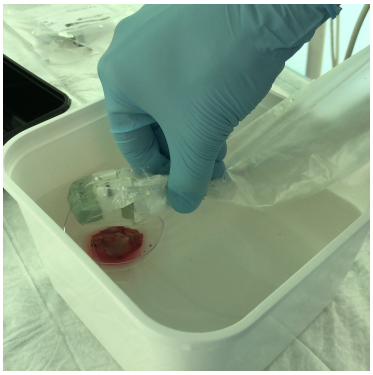
In this prospective study, nine patients were included based on the diagnosis of a TSCC and surgery as therapy. The clinical tumor stages were: T1 (one cases), T2 (five cases), T3 (two cases) and one case was a residue after radiotherapy for an initial T2. The patients were treated according to standard protocol and if required, they received adjuvant therapy as well. The study had full local ethical approval from the institutional research board of the Netherlands Cancer Institute.

5.3.1 Materials and data acquisition

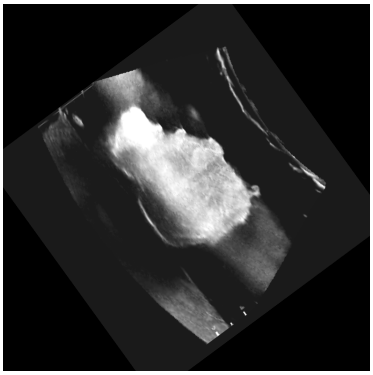
For this study, 2D US data were acquired intra-operatively on nine subjects using a BK5000 ultrasound system (BK Medical, Denmark) and a small intraoperative convex (5-14 MHz) transducer, as shown in Fig. 15c. The operator used a freehand sweep with a transducer frequency of 10 MHz. The position and orientation of the transducer were measured and recorded by an EM tracking system (Aurora, NDI, Canada) with two sensors (six degrees of freedom), one attached to the transducer and the other located at the bottom of the setup, as reference. The 2D US images were reconstructed into 3D US volumes, as shown in Fig 15d-e, for fast and easy annotation using a Pixel Nearest Neighbor algorithm in CustusX, an open-source navigation platform for image guided therapy [42]. From each subject, several acquisitions from different directions were performed which resulted in a total of 69 3D US volumes. Some acquisitions were excluded because the field of view did not capture the complete specimen or extreme movement of the transducer resulted in useless reconstructed volumes, ending up with 44 3D US volumes included for this study. The pre- and post-processing of the US data was performed with Python. The ground truth labels of the specimen and tumor were annotated, as shown in Fig. 15g, by a radiologist in 3DSlicer, an open-source software platform for medical image informatics, image processing, and three-dimensional visualization [44]. The UNet model was trained using a Tesla K80 with 17 GB of video memory.

(a) In-vivo tongue and tumor prior to surgical excision.

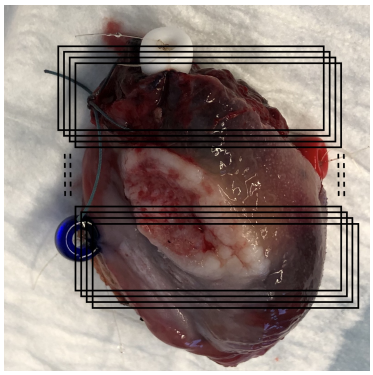
(b) Ex-vivo tongue specimen including tumor. The orientation was indicated with red, white and blue markers.



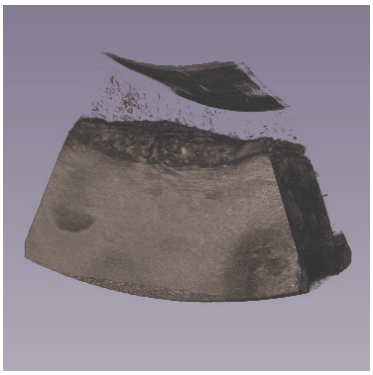
(c) Acquisition of the specimen with a navigated US transducer.



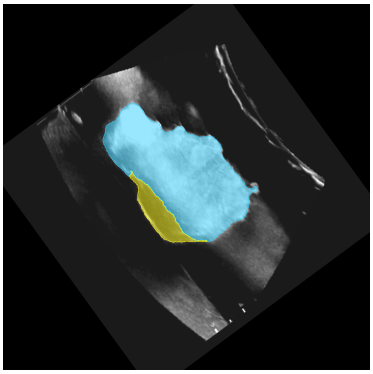
(d) A 2D US image in gray scale of a cross section of the specimen.



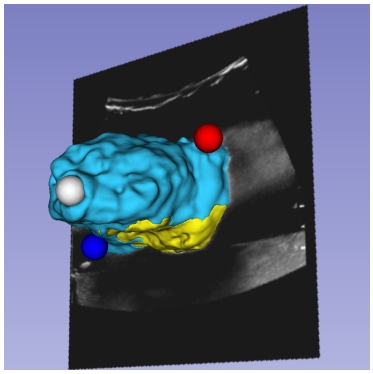
(e) The black boxes represent the multiple 2D US images which were acquired in sub-figure (c).



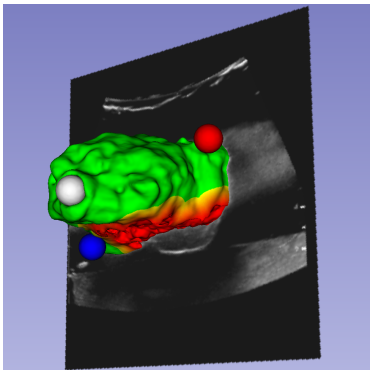
(f) The multiple 2D images were reconstructed into a 3D US volume visible in this rendered volume.



(g) A US image in gray scale of a cross section of the 3D volume. The specimen is segmented in light blue and the tumor in yellow.



(h) A 3D view of the segmented specimen (light blue) and tumor (yellow). A US image in gray scale at the relative location within the 3D US volume.



(i) A 3D view of the segmented specimen and tumor. The color of the specimen's surface represents the resection margin in 3D from 0 to 5 mm, in respectively red to green.

Figure 15. All steps from surgical resection till 3D visualization of 5 mm margins around the tumor. If good performance in clinical setting is proved, the trained UNet models could replace the manual segmentation in sub-figure (g).

5.3.2 UNet

For this study, the UNet architecture, as shown in Fig. 14, was adopted. Since this study used the cascade training strategy, explained in section 5.3.3, this UNet architecture was used to create two identical UNet models, the model specimen and the model tumor. It consists of a down-sampling pathway, up-sampling pathway and skip-connections to reuse low-level features in higher levels. The down-sampling pathway contains convolutional blocks and max-pooling layers repetitively. The up-sampling pathway was built from de-convolutional layer and convolutional blocks. The following settings were used for the UNet model. Each convolutional block was made up of 3x3 conv + rectified linear unit (ReLU) + 3x3 conv + ReLU + 2x2 max pool. The components in the de-convolutional block are: 2x2 up-conv + 3x3 conv + ReLU + 3x3 conv + ReLU. The movement of the convolutional filter had a stride of one in both dimensions and to ensure that the output image has the same size as the input image, padding was activated for all convolution operations. After each convolutional block, the number of feature maps was doubled ranging from 8 to 128, and halved during the up-sampling pathway due to each de-convolutional blocks. The final layer consists of a 1x1 convolutional layer with a sigmoid activation. The weights of the network were updated using an Adam optimizer with an initial learning rate of 0.001 and decays with 10% in case the validation loss did not decrease for 5 epochs. Depending on the experiment, see section 5.3.8, the models had a binary cross-entropy or dice coefficient loss function.

5.3.3 Cascade Strategy

This study proposes a strategy adopted from the cascade strategy [57–59], where a complex multi-class segmentation problem is split into multiple simple binary segmentation problems. This is also known as coarse-to-fine medical image segmentation [57, 58]. This strategy is popular because of the class imbalance problem, which is known in medical images [57]. As mentioned above, two models (specimen and tumor) were built and each trained upon their own dataset. The cascade strategy decreases the class imbalance problem for tongue tumor segmentation by the following steps, listed referring to Fig. 16: c) predict the pixels containing specimen including tumor, d) compute a ROI around the prediction of model specimen, e) crop the input images by the computed ROI, g) predict the pixels containing tumor only in cropped input images, i) combine the predictions, where the tumor is correctly re-located in the predicted specimen.

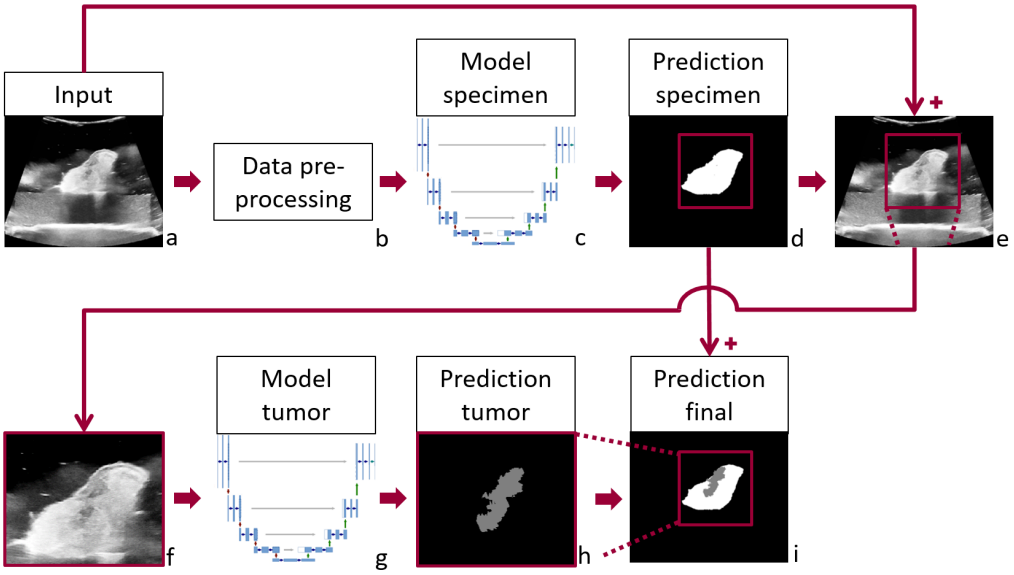


Figure 16. Schematic overview visualizing all steps of predicting specimen and tumor in a cascade strategic fashion. a) Original images as input for model specimen, b) Pre-process the input images by resizing to slices x 256 x 256 x channel, normalizing and binarizing, c) Predict the pixels containing specimen, d) Compute a ROI (256x256) around the predicted specimen, e) Crop the original image in the size of the ROI, f) Cropped images as input for model tumor, g) Predict the pixels containing tumor, h) Re-locate the ROI back into the original position in the prediction specimen and i) A final multi-class prediction.

5.3.4 Loss

The outputs of the network are a representation of the probability of each pixel to belong to a certain label. In the US images, the tumor and specimen pixels occupy a small region causing a class imbalance between the target areas and the background. This imbalance results in predictions strongly biased towards the background and target areas are missed or partially detected [51, 60]. The accuracy has not been a correct evaluation metric in such case, since it would remain high because of the background largely present in labels and correctly predicted. Previous approaches restored this imbalance by re-weighting the pixels belonging to the foreground [60]. Another option is the Dice Similarity Coefficient (DSC), which only takes the predicted foreground and ground truth foreground into account. The DSC is calculated by eq. 5 [57], where TP is the number of true positive predicted pixels, FP is the number of false positive predicted pixels and FN indicates the number of false negative predicted pixels. This coefficient is ranging from 0 to 1 and should be maximized.

$$DSC = \frac{2TP}{FP + 2TP + FN} \tag{5}$$

During the training process, the soft Dice loss function, eq. 6, should be minimized [51, 58]:

$$\mathbf{L}_{dice} = \frac{2\sum_{i=1}^{N} o_i l_i + \epsilon}{\sum_{i=1}^{N} o_i + \sum_{i=1}^{N} l_i + \epsilon} \tag{6}$$

where the sums run over the N voxels, of the predicted output volume $o_i \in O$ and the ground truth volume $l_i \in L$. o_i is the probability between 0 and 1, and $l_i = 0$ when i is not in the ground truth and 1 if i is. A smoothing factor ϵ is set to 1, to smooth the loss function and avoid zero division.

5.3.5 Dataset splitting

The ground truth labels were created by manual segmentation of the tongue specimen and tumor by a radiologist. These annotations were exported as labelmap, containing the labels 0, 1 and 2 for background, tumor and specimen, respectively. From the 44 included 3D US volumes the patients were divided over datasets based on the amount of included 3D US volumes from the same patient, so the training dataset contained patient 2, 5, 8 and 9, the validation dataset contained patients 4 and 6 and patients 3 and 7 functioned as the test dataset. The 3D volume structure of both the inputs and labels were split into 2D images and tagged with an ID, corresponding to the inputs and labels. Only 2D images presenting >300 pixels of ground truth specimen have been used, since images with less pixels contribute less to the training of the model. These images with less pixels were present at the head and tail of the 3D US volume. The included 2D images were captured from the center of the 3D US volumes and accounted for 65-70% of the total number of slices within the datasets.

The cascade strategy required two datasets for two UNet models. From the total dataset, dataset specimen and dataset tumor were made, while maintaining the distribution of the patients into training, validation and test datasets. Figure 17 visualizes the process of creating the datasets specimen and tumor.

In the first dataset, containing all the 2D images and labels presenting >300 pixels of ground truth, the annotations were binarized, so 0 representing background and 1 representing specimen including tumor. In total, dataset specimen contained 12648 2D images divided into training (8233 images, 65%), validation (2658 images, 21%) and test sets (1757 images, 14%).

To create the dataset tumor, several pre-processing steps had to be performed. In case the original 2D image and label in dataset specimen presented the annotation tumor, the minimum and maximum x- and y-value of the specimen including tumor were found. The center of these values

was computed and a fixed region of interest (ROI), corresponding to the input size of model tumor, was cropped from both the original input images and labels. Now, the annotations were binarized again, with 0 representing background including specimen and 1 representing tumor only. Finally, dataset tumor only contained cropped images and labels which present tumor annotation. In total, dataset tumor contained 5718 2D images divided into training (4284, 75%), validation (886, 15%) and test sets (548, 10%).

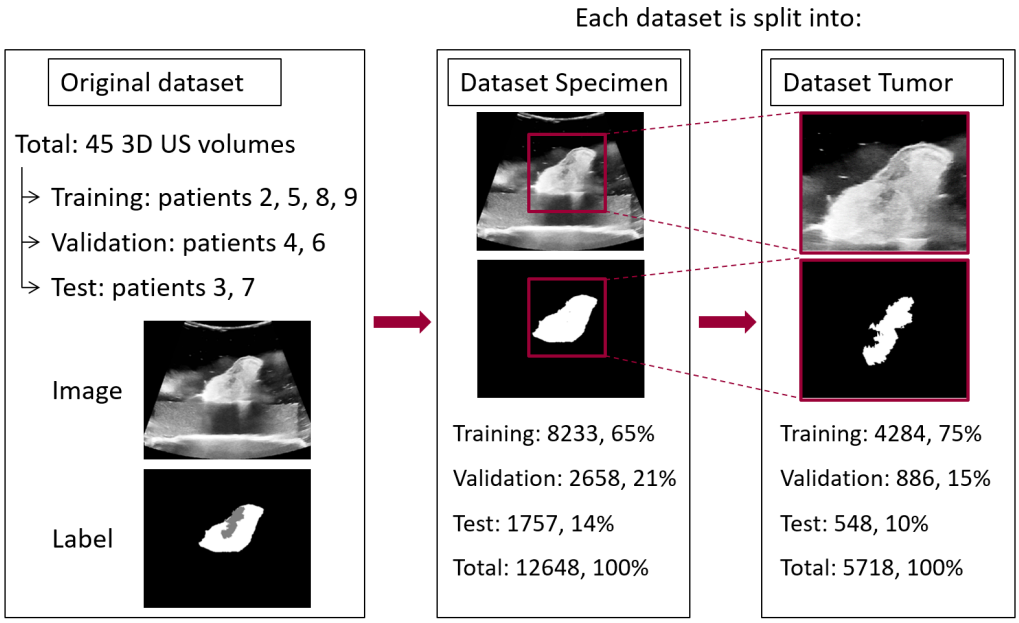


Figure 17. Schematic overview of creating datasets specimen and tumor.

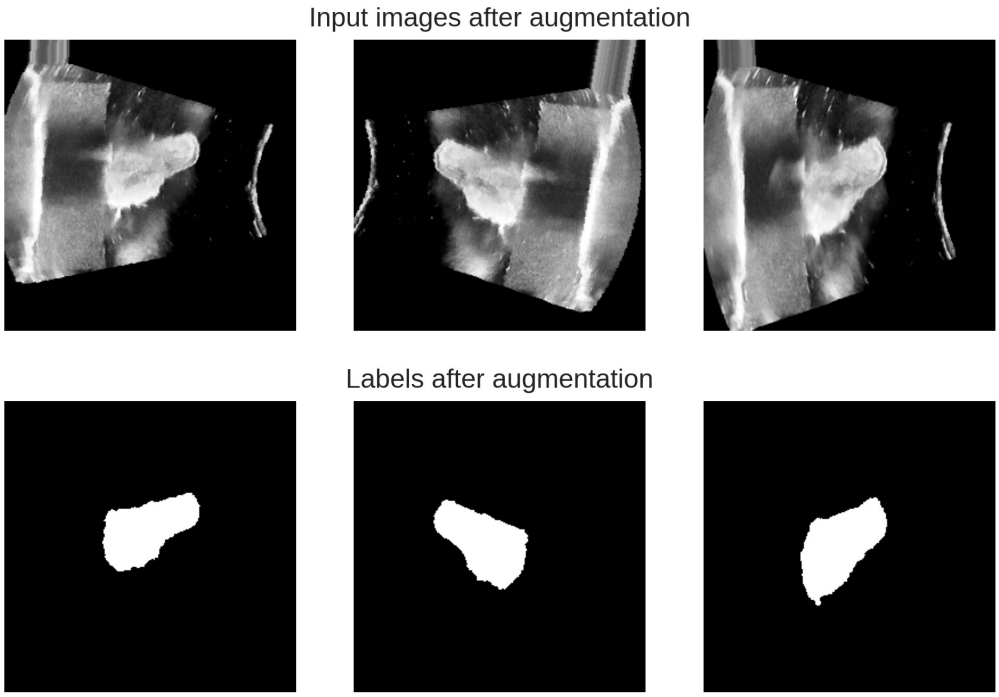


Figure 18. Examples of inputs and labels after augmentation from dataset specimen.

5.3.6 Data pre-processing and augmentation

Pre-processing of the data in both datasets, equal for input images and labels, was performed per batch and consists of normalization and resizing into 1 slice x 256 pixels x 256 pixels x 1 channel. Data augmentation was randomly applied to generate additional diverse data and avoid overfitting by the following transformations: rotation (range: -30°, +30°), horizontal and vertical shift (range: -10, +10%), zooming (range: 70-130%), brightness shift (range: 50-100%) and horizontal and vertical flip. Figure 18 shows some examples of augmented data.

5.3.7 Evaluation

The quality of manual segmentation of the specimen and tumor as annotations by a radiologist were evaluated by comparing the 3D volumes of the ground truth annotations for each patient.

Since DSC is used as key evaluation metric within segmentation challenges, such as the BraTS (Brain Tumor Segmentation) challenge [59], the DSC was computed for each region to evaluate the performance of both models.

5.3.8 Experiments

To prove the effect of data augmentation on overfitting during the training of the models, in the first experiment both models specimen and tumor were trained with and without application of the described data augmentation. The second experiment investigated the influence of the loss function in both models specimen and tumor. The dice loss function is effective in alleviating class imbalance. The results of this loss function were compared with the results of the binary cross-entropy (BCE) loss function, common in binary classification problems, calculated as followed:

$$BCE = -\frac{1}{N}\sum_{i=1}^{N} g_i \cdot \log(\hat{g}_i) + (1 - g_i) \cdot \log(1 - \hat{g}_i) \tag{7}$$

with N the number of pixels, g_i the ground truth label and $\hat{g}_i$ the predicted output.

For all experiments, the data were divided randomly into batches of 32 images. The models were trained end-to-end for 50 epochs.

In the end, the final predictions of all slices were combined into a 3D volume. These volumes were exported to 3DSlicer. Here, the distance between the tumor and specimen was computed in 3D and visualized by a colormap ranging from 0 (red) to 5 mm (green), as shown in Fig. 15i.

5.4 Results

In this section, the intraobserver variability in ground truth annotations and the achievements of the models is presented.

From the nine patients who underwent oral surgery of the tongue, a total of 69 3D US volumes were acquired from the specimens of which 44 volumes were included in this study. All 3D US volumes from patient 1 have been excluded, since the specimen was not entirely observed in the field of view of the 3D US.

Table VI shows the variability of 3D volumes of the ground truth annotations within each patient. What stands out in the table is the large standard deviation (SD) for specimen and tumor. For example, patient 3 has a SD of 39% and 72% of the average volume for specimen and tumor, respectively.

Table VI. The average volumes of the annotation within each patient.

		# included 3D volumes	**Average volume specimen in mm3 (SD)**	**Average volume tumor in mm3 (SD)**
Patient	**1**	excluded	excluded	excluded
	2	5	18222 (±328)	766 (±316)
	3	4	21513 (±8482)	121 (±87)
	4	4	47104 (±902)	708 (±268)
	5	7	30066 (±2506)	9075 (±2520)
	6	6	19020 (±1495)	2017(±1084)
	7	3	13729 (±469)	1290 (±163)
	8	7	36485 (±1407)	2276 (±792)
	9	8	12782 (±435)	1059 (±260)

The performance of both models specimen and tumor is shown in Table VII. Figure 19 shows a prediction compared to the corresponding annotated ground truth and the original US image from the best performing network combination. Predicting a test volume took approximately 12 seconds.

Table VII. The DSC of Model Specimen and Model Tumor. The bold values represent the highest DSC.

Model Specimen		
	Data Augmentation	
Loss	True	False
Binary Cross-entropy	**86%**	68%
Dice coefficient	76%	77%

Model Tumor		
	Data Augmentation	
Loss	True	False
Binary Cross-entropy	9%	0%
Dice coefficient	**18%**	0%

(a) Input image (b) Ground truth (c) Prediction

Figure 19. Comparison of the input image, ground truth and final prediction.

5.5 Discussion

The goal of this study was to show as a proof of concept that deep learning is a feasible technique for fast automatic multi-class segmentation of tongue specimen and tumor in 3D freehand US volumes. The main findings of this study were: 1) the radiologist had difficulties with annotating the ground truth in the 3D US volumes, 2) the BCE out-performed the Dice loss function in segmentation of the specimen but not the tumor in 3D US data, 3) data augmentation provided more accurate results in segmenting the tumor. From these findings, one can infer that the current models' accuracies were not sufficient for implementation in clinical setting. Clinical implementation might be able once the models' predictions significantly correlate to the histopathological findings, the gold standard, and the assessment by 3D US fits into the intra-operative workflow. Higher accuracy could be obtained by increasing the dataset, improving data acquisition, assisting the radiologist with accurate annotating based on histopathology slices, combining the models into one so it could be trained as one, although the individual impact is unknown. It is speculated that deep learning could perform automated 3D segmentation of tongue specimen and tumor in 3D US volumes if improvements were carried out.

Comparing to previous studies, the findings of this study showed that the BCE loss achieved a higher DSC than the Dice loss for larger areas, and vice versa for smaller regions, which confirms the results of Wang et al. [57]. Applying data augmentation showed minimal effect on model specimen while the DSC largely improved on model tumor because of data augmentation. Zhu et al. had similar findings of models with data augmentation which outperformed models without data augmentation, only in case of a minimal training data [51]. Besides confirmation of the impact of loss function and data augmentation, this study was the first segmenting 3D US volumes of tongue resected specimens for resection margin assessment by deep learning. To accomplish this, a complete intra-operative workflow from data acquisition to intuitive colormap visualization was set up. The colormap provided the surgeon feedback about the obtained resection margins.

The first objective of this study was to gain insights in the variability in annotating ground truth in 3D US volumes by a radiologist. A small deviation between the annotated volumes of one patient has been expected, since manual segmentation is hard to reproduce accurately. Surprisingly, the SD of volumes within one patient, shown in Table VI, ranges 1-39 %, 12-72 % of the average volume

for specimen and tumor annotations, respectively. Low reproducibility of manual segmentation could not cause such largely diverging annotated volumes. A possible explanation might be the intra-observer variability during US acquisition and reconstruction, resulting in large differences between the acquired volumes of one patient. Another reason could be the low US frequency of 10 MHz used during data acquisition. With a 10 MHz US transducer, the borders of small tumors, such as 2 mm tumor thickness following the histopathology reports, were hard to detect by the radiologist. High frequency US transducers should provide more details when acquiring small parts such as the resected specimens from this study. Summarizing, the intra-observer variability during US acquisition and additionally a low US transducer of 10 MHz were likely the reason of the high variability in the annotated volumes by a radiologist.

The second question in this research was to investigate the accuracy of deep learning in segmentation of tongue specimen and tumor in intra-operatively acquired 3D US volumes. The two investigated variables which influence this accuracy were the loss functions and the application of data augmentation. The results from Table VII:Model specimen showed that the BCE loss function provided a better DSC over the Dice loss function when data augmentation was applied. It was expected that the data contained large class-imbalance between background and specimen including tumor. Previous studies stated that a Dice loss function is able to deal with class-imbalance [57], as this function does not take true negative predictions into account, whereas the BCE loss function does. However, BCE loss function achieved higher DSC so it can be concluded that the class-imbalance between background and specimen including tumor was overestimated.

Contrary, the BCE loss functions in model specimen resulted in lower DSC compared to the Dice loss function without applying data augmentation. During predictions of the seven test volumes from patient 3 and 7, large differences in the predictions between the patients occurred. The deviations in annotated volumes within patient 3 were very hard to predict, which decrease the average prediction DSC of all 7 test volumes. When only the test volumes of patient 7 were predicted, the BCE loss function without application of data augmentation would have resulted in a average DSC of 79%.

Focusing on model tumor, the results were as expected. The tumor was a small region compared to the background including specimen, resulting in a class-imbalance by which the Dice loss should provided the best DSC. And to take into account the large SD of the average volume of the tumor within patient 3, the average DSC would have been 43% when only the test volumes of patient 7 were predicted.

Applying data augmentation resulted in a minimal impact at the test DSC for both loss functions when predicting the specimen, as shown in Table VII. Data augmentation creates more diverse data so training the model while applying this technique should result in a more generalized model. An explanation could be that dataset specimen, containing 12648 images, was sufficient to train a generalized model. The impact of data augmentation on model tumor was more evident. The only difference between the two datasets was the total amount of images, 12648 relative to 5718 images in dataset tumor. Concluding, application of data augmentation was effective in case the dataset did not contain a sufficient amount of images.

Another aspect to discuss is the 0% accuracy of model tumor for both loss functions when data augmentation was not applied. This suggests that the 5718 2D US images in dataset tumor contain deficient information to train the model in segmenting the tumor. However, applying data augmentation provided 9% and 18% accuracy suggesting that the effects of rotating, shifting, flipping, zooming and brightness adjustments makes the information in the 2D US images sufficient to predict tumor pixels correctly.

Overall answering the second question, this study gave more insight in the impact of the loss function and data augmentation on the final model's accuracy. However, due to a small group of included patients, cross-validation has not been applied, since the datasets were too small to be generalized. By applying cross-validation, totally different datasets would be created, resulting in different outcomes which could not be compared. Therefore, the chance remains that some results were exceptions.

The final objective was to understand the feasibility of intra-operative multi-class segmentation predictions by deep learning. In other words, what are the requirements to perform this technique intra-operatively and how long requires deep learning for a prediction? Once the models were trained, they could be transferred to any computer at the operating room with the required open-source supporting software to run a prediction. Since all 3D US acquisitions had already been performed at the operating theater, no changes in materials were required to perform a prediction. The time required for predicting one test volume was approximately 12 seconds. With the additional steps of computing the distances between specimen and tumor in order to create a clinical intuitive colormap to support the surgeon, a total of five minutes could be assumed. Since surgeons requested to provide the feedback within 30 minutes, deep learning could be a feasible technique for intra-operative multi-class segmentation.

The findings contribute to the improvement of surgical resection of tongue squamous cell carcinoma (TSCC). Intra-operative assessment by 2D US was already possible, since Brouwer et al. proved the significant relationship between the resection margins measured on 2D US and histopathological slice [22]. In this study, the assessment of resection margins was extended from 2D US images to 3D US volumes, to ensure that the measured margin on US is the actual minimal margin. The correlation between the measured margins on 3D US and histopathological slices should be investigated again. If a significant relationship between the two modalities could be proved, intra-operative 3D US assessment based on fully automatic segmentation of the tumor should be implemented to reduce the recurrence rates and excision with resection margins <5mm. Therefore, further research focusing on the correlation between the measured resection margins on 3D US and histopathological slices is recommended.

Pointing out several strengths of this study, the first one was that data acquisition was performed in an intra-operative clinical setting. Conclusions from this study were representative for findings after clinical implementation. Secondly, annotating the ground truth in 3D was helpful for the radiologist. The information of subsequent slices provided additional insight in the orientation of the specimen and the location of the tumor. Also, investigating the influence of two loss functions on the final prediction was a strong point in the method of this study. In advance, binary class-imbalance was taken into account, so the Dice loss functions, effective in dealing with class-imbalance, could be investigated. Last but not least, adopting the state-of-the-art cascade strategy could be seen as a strong point to deal with the complex multi-class segmentation problem in this study. This strategy easily deals with the class-imbalance between tumor and background including specimen by cropping the input images. Also, more insights were provided about class specific segmentation difficulties. Now it is clear which segmentation problem requires more improvements.

On the other hand, this study was subject to some limitations. The major limitation of this study was the small group of patients. The final results from this study were based on only two patients. Therefore, the findings could not be generalized. However, as a proof of concept, this research provided indications about the possibility that deep learning might be a feasible technique for fast automatic multi-class segmentation of tongue specimen and tumor in 3D freehand US volumes.

Additionally, the models would have been trained better when the datasets were extended. Also, more patients would facilitate the possibility to perform cross-validation. This will reduce the chance of findings by coincidence.

Besides the impact on training, the adopted UNet model could be converted from 2D to 3D. Then, inputs will be entire 3D US volumes instead of 2D images and therefore segmentation of the specimen and tumor could be more accurate when information of subsequent slices will be included.

Secondly, it was unfortunate that the annotations of the ground truth by the radiologist showed a large SD of the average volumes within a patient. This inconsistency did not contribute in training an accurate model. Because of the low frequency US transducer (10 MHz) some 3D US volumes lack details in case the tumor was very small. Simply introducing a high frequency US transducer

would have provided these details and could help the radiologist annotating more correctly. Another reason for the large SD was the lack of feedback towards the radiologist about the accuracy of annotations. The 3D US volumes of unknown tissue without specific anatomical landmarks impeded the radiologist to annotate with certainty. Taking a closer look at Fig. 19 a, it could be speculated that the model might be more accurate in segmenting the tumor than the actual ground truth annotation. Annotations based on the gold standard of histopathology slices could ensure more accuracy. This means that the entire specimen needs to be cut into slices with a known inter-slice distance. Annotating the specimen and tumor in the histopathology slices and remodel these slices into a 3D histopathological model would provide gold standard information. Registration of the 3D histopathological model towards the 3D US model including the gold standard annotations could support the radiologist in annotating with certainty.

Finally, future research should focus on: 1) implementing a 3D UNet when more data is acquired, 2) remodeling the histopathological slices into a 3D model and registering towards the 3D US models, 3) improving the implementation of the cascade strategy. With a 3D UNet, the information of subsequent slices will be included which could help segmenting the specimen and tumor in the 3D US volumes. Currently, the models were trained independently while the actual cascade strategy as published in previous literature connects the models so the output of model 1 will be propagated forward into model 2 and both models will be optimized at once during training [57–59].

5.6 Conclusion

This study was set out as a proof of concept to show that deep learning is a feasible technique for fast automatic multi-class segmentation of tongue specimen and tumor in 3D freehand US volumes. The results of this investigation showed that two UNet models trained on 12648 and 5418 slices, respectively, from 44 3D US volumes could reach a DSC of 86% segmentation of the specimen and 18% DSC in segmenting the tumor. This implicates that fast automatic segmentation of tongue specimen and tumor in 3D US volume by deep learning is a feasible technique for intra-operative assessment of resection margins. This is the first study adopting state-of-the-art methodology, to assess resected tumors with 3D freehand US volumes and fast automatic segmentation of multiple regions from these 3D US volumes. In addition, a full intra-operative workflow was created from data acquisition to intuitive visualization of close resection margins using a colormap.

6

The correlation between the resection margin assessed by 3D US and histopathology

6.1 Introduction

Complete surgical excision of TSCC has been very important, since close resection margins are related to poor prognosis in terms of local recurrence and 5 years of survival [5, 11, 19]. Opinions about the distance representing clear or close margins are divided [13, 14]. However, in most literature <1 mm, 1-5 mm and >5 mm measured in histopathological slices are considered as positive, close and clear resection margins, respectively [2, 9, 13, 15–17]. Unfortunately, histological assessment is performed post-operatively [13], while intra-operative assessment of the resection margins in tongue tumors is highly preferred [12, 15]. Several studies investigate techniques such as frozen section and intra-operative imaging modalities to meet this preference.

Brouwer de Koning et al. prove that 2D US is a feasible intra-operative assessment technique and significantly correlates to histopathology [22]. Nonetheless, the chance remains that the actual closest resection margin is missed as the 2D US is limited to an intersection view only. The previous section 5 tried to prove the potential of intra-operative 3D US to assess resections margins, and so providing complete feedback about the entire specimen to the surgeon. However, very little is currently known about the correlation between the resection margins found by intra-operative 3D US and histopathology. When correlation is proven, intra-operative 3D US could be implemented clinically and prevent potential close margins.

Applying 3D US in other clinical cases, such as assessing the response of tumor angiogenesis in breast cancer patients undergoing neoadjuvant chemotherapy, 3D US proved to be effective when contrast was enhanced [61]. Also, Hashad et al. show that 3D US was able to accurately diagnose adenomyosis in 59% of the patients [62]. To the best of our knowledge, the correlation between resection margins assessed by 3D US and histopathology has not been reported before.

For this research, in nine patients the resection margins of TSCC were assessed intra-operatively by 3D US and post-operatively by histopathology. The correlation between the measurements by 3D US and histopathology was computed by the Pearson correlation coefficient. The aim of this study is to explore the correlation between resection margins assessed by 3D US and histopathology.

6.2 Research question

What is the correlation between the resection margin in tongue tumor specimens assessed by 3D ultrasound and histopathology?

6.3 Method

In this section, the subjects, study design, performed measurements and analysis are described.

6.3.1 Subjects

Nine patients who underwent surgical treatment at the Netherlands Cancer Institute were included. The group consisted of five men and four women. The average age was 72.5 years, ranging between 49 and 86 years. The tumor was located either at the tongue blade (six times right, 2 times left) or tongue base (left). Clinically the tumor stages, following the 8th edition of the AJCC Cancer Staging Manual [10], were: T1 (one cases), T2 (five cases), T3 (two cases) and one case was a residue after radiotherapy. The institutional research board of the Netherlands Cancer Institute provided full local ethical approval to this study.

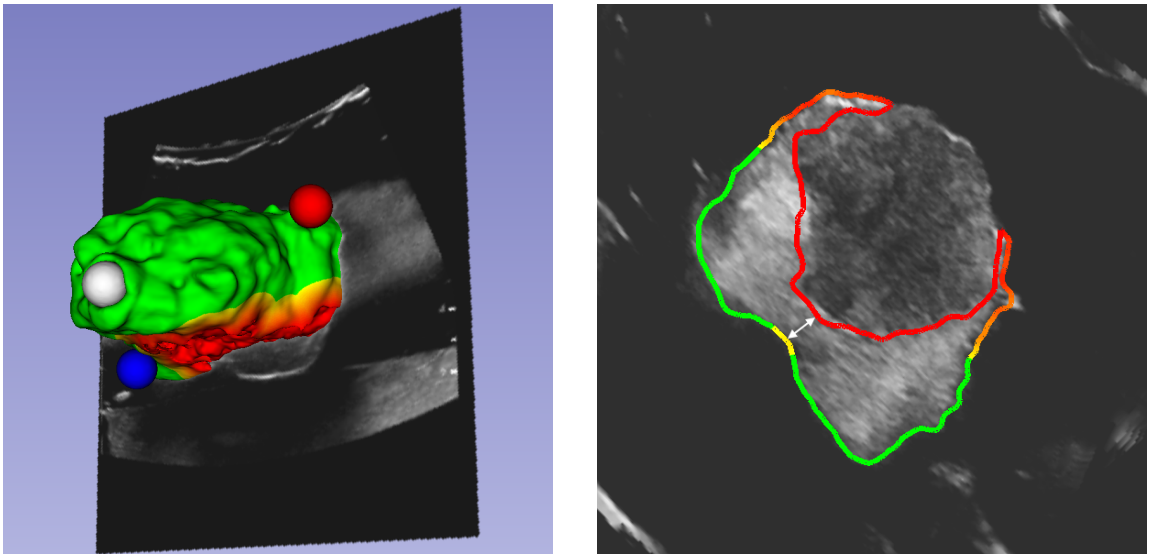
6.3.2 Methods of measurement

From each patient, the resected tongue specimen was acquired as 3D US volume between five and 12 times per patient. 3D US volumes were excluded in case the specimen exceeded the field of view or severe shaking of the US transducer occurred during acquisition. Data acquisition was performed as described in section 5.3.1.

The specimen and tumor within the 3D US volumes were annotated by a radiologist in 3DSlicer. A 3D distance map between the specimen and tumor was computed, which visualizes the resection margins surrounding the tumor, as shown in Fig. 20a. A customized colormap from red to yellow was created to represent the resection margins between 0 and a user defined distance, respectively. The areas at a distance equal to or larger than the user defined distance (e.g. 5 mm) were represented in green. For each 3D US volume, the user defined distance was adjusted until the specimen on the resection plane turned yellow. Than, the corresponding distance was noted as closest resection margin, as shown in Fig 20b. Finally, the average resection margin within each patient was computed.

The resected tissues were treated by a pathologist following standard protocol. First, the specimen was inked followed by fixation in formalin for 24 hours. Slides of 4 mm were cut from the specimen before embedding those slides in paraffin. Finally, 4 um sections were cut from the 4 mm slides, stained with haemotoxylin-eosin dye (HE) and mounted on histopathological glass. The closest resection margin measured by the pathologist was extracted from the histopathological report.

The correlation between the resection margins measured at the annotated ground truth in 3D US and the histopathological slices was determined by performing the Pearson correlation coefficient. Since only two patients were distributed as test group in chapter 5, a correlation between the resection margins measured at the deep learning predicted segments and histopathological slices was not calculated.



(a) 3D colormap **(b)** Measuring a resection margin in an intersecting slide

Figure 20. a) 3D colormap representing the resection margins on the specimens surface. The 2D US image is shown intersecting the specimen and tumor. b) A 2D US image with the distance colormap and the measured distance in mm.

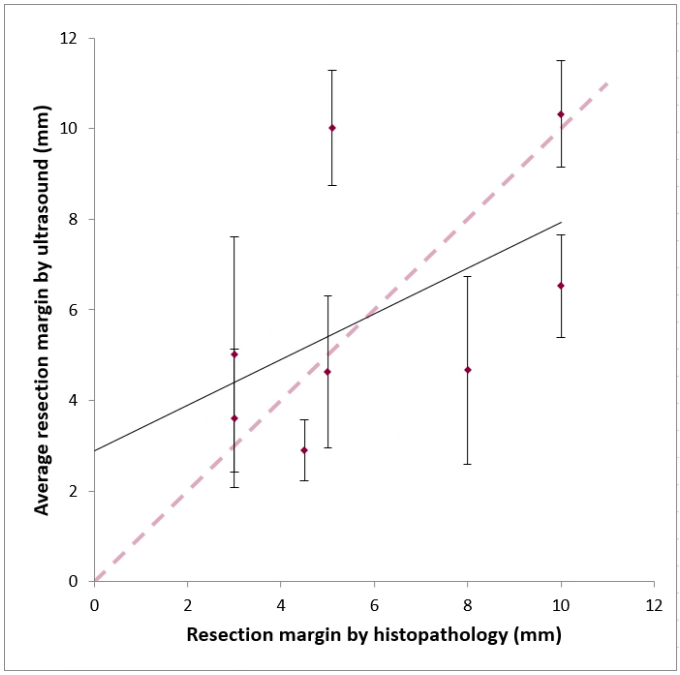


Figure 21. The relationship between resection margin by histopathological slice and 3D ultrasound. The error bars represent the SD of the average resection margins by ultrasound of each patient. The red dotted line represents a correlation of $Y = x$. There was no statistically significant correlation between the measurements (n=8). $R = 0.518$, $Y = 0.5054x + 2.8889$

Table VIII. The minimal resection margins measured by different methods and the absolute differences between those methods.

Case	TNM stage	Tumor location	Average resection margin by 3D US in mm (SD)	Resection margin by histopathology in mm	Absolute difference between 3D US and histopathology in mm
1	pT3pN3b	Tongue blade right	excluded	excluded	excluded
2	pT2N1	Tongue blade right	4.7 (±2.1)	8.0	3.3
3	pT1N0M0	Tongue blade right	10 (±1.2)	10	0.3
4	Residue	Tongue base left	4.6 (±1.7)	5.0	0.4
5	pT4aN2b	Tongue blade right	3.6 (±1.5)	3.0	0.6
6	pT2N0	Tongue blade left	5.0 (±2.6)	3.0	2.0
7	pT2N2c	Tongue blade right	2.9 (±0.7)	4.5	1.6
8	pT2N2b	Tongue blade left	10 (±1.3)	5.1	4.9
9	pT1N0	Tongue blade right	6.5 (±1.1)	10	3.5
		Average	6.0 (±2.6)	6.1 (±2.7)	2.1 (±1.6)

6.4 Results

Table VIII shows the pathological TNM stages and tumor location of the included patients. Patient 1 (man, clinical stage: T3) has been excluded since the specimen was not entirely observed in the field of view of the 3D US. The data analysis of all patients is shown in Table VIII. The mean difference between 3D US and histopathology was -0.1 mm (SD: 2.6 mm). Figure 21 shows the correlation between the measurements by 3D US and histopathology. The Pearson correlation coefficient showed no statistically significant correlation between these measurements ($R = 0.518, p = 0.187$).

6.5 Discussion

This study set out with the aim of exploring the correlation between resection margins assessed by 3D US and histopathology. The main finding of this study was that the measurements of resection margins by 3D US and histopathology do not correlate statistically significant. Inferring from this finding, 3D US could not provide correct intra-operative feedback to the surgeon. However, with a small sample size (n=8), caution must be applied when interpreting this result. Also, certainty about measuring the resection margin by 3D US and histopathology at the same location could not be guaranteed. More insights about the measuring location could suggest that assessment of resection margins by 3D US and histopathology correlates.

Novel in the present study was the use of 3D US to assess the resection margin and calculating its correlation with histopathology. Comparing the correlation between measurements by 2D US and histopathology with other studies, Shintani et al. evaluated the lesion pre-operatively by measuring tumor thickness and found significant correlations between 2D US and histological sections [63]. Others found a Pearson correlation between assessing tumor thickness in TSCC by 2D US and histopathology of $R = 0.80$ [64]. The utility of 3D US was proved by Lunardelli da Silva et al., as the correlation between 3D US and histopathology in diagnosing endometriosis was 72.5% (n=40) [65].

The initial objective of this study was to identify the correlation between the resection margins assessed by 3D US and histopathology in tongue tumor specimens. There were three likely causes for the statistically not significant correlation. First, the result has been expected since it is probably related to the small sample size (n=8).

Secondly, this finding may be explained by the large SD of the average resection margin by 3D US. These large SD were probably the outcome of the manual annotation by the radiologist. Section 5 shows large SD in the average volume of the annotations by the radiologist. Since the resection margins by 3D US were assessed based on the same annotations, a large SD of the average resection margin by 3D US has been expected.

Finally, it is uncertain that the resection margins assessed by 3D US and histopathology were measured at the same location. The current data did not provide the location of measuring the resection margin by histopathology. Therefore, the absolute differences between 3D US and histopathology could be correct and accurate. By providing the location of the resection margin by histopathology additionally, recalculating the correlation could result in statistically significant. Reconstruction of the histopathological sections into a 3D model could provided this information.

Additionally, the inter-observer variability among histopathologists has to be taken into account [66, 67]. In the current method, a single histopathologist assessed the resection margins. Because of this degree of variability, a certain difference between measured resection margins by 3D US and histopathology was expected. Correctly assessing the resection margin by histopathology should therefore by performed repetitively by a single or multiple operator(s) followed by computation of the degree of agreement among assessments.

Answering the research question, the correlation between assessing the resection margin in tongue tumor specimens by 3D US and histopathology was not statistically significant.

The findings of this study contribute to the improvement of surgical excision of TSCC by providing feedback to the surgeon about the obtained resection margins. Unfortunately, there was no statistically significant correlation between measurement of the resection margin by 3D US and histopathology. However, the correlation between measurement of the resection margin by 3D US and histopathology should be calculated again, after improving the 3D US data acquisition, reconstructing histopathological section into 3D models and obtaining a large patient group. It is speculated that the correlation between the resection margins by 3D US and histopathology could be

statistically significant as 2D US is [22]. For now, physicians can already be provided of intra-operative feedback by assessing the resection margin in 2D US.

Addressing strong points and limitations, the strength of this study was the set up which provides simple implementation of new data. Recalculating the correlation would be easy when sufficient data is available.

This research was limited by the small sample size, the large SD of the average resection margin by 3D US which was already mentioned in section 5.5 and the absence of the location measuring the resection margin by histopathology. Possible reduction of these limitations is already mentioned above or in section 5.5.

6.6 Conclusion

The purpose of the current study was to determine the correlation between resection margins measured by 3D US and histopathology in tongue tumor specimens. The result of this research show that assessment of the resection margin in tongue tumor specimens by 3D US and histopathology do not correlate statistically significant. This study should be repeated using more included patients, accurate and consistent annotations in 3D US and the location of the measured resection margin of both 3D US and histopathology. Then, in case correlation is statistically significant, this technique could be implemented to assess the resection margin intra-operatively which helps us reaching the goal of minimizing the resection margins <5 mm.

7
General Conclusion

In 2018, the prevalence of tongue cancer was 2069 in the Netherlands [1]. Surgical treatment is mostly common, however in 85% of the cases the obtained resection margin is below the minimum of 5 mm [17, 47]. Currently, post-operative histopathological assessment is the only way to confirm the resection margin. A tool to assess the resection margin intra-operatively and provide feedback to the surgeon is highly preferred. Previously, investigation showed that 2D US was successful and recommended extension to 3D US since only a single plane can be observed [22]. An invasive growth pattern of the tumor could result in involved margins elsewhere other than the observed plane. This thesis describes three consecutive studies which attempted to solve the clinical problem and create an intra-operative tool to assess resection margins of tongue squamous cell carcinomas based on 3D US.

To reach this goal, the first section aimed to experimentally substantiate the preferred conditions to acquire accurate 3D US reconstructions of resected tongue tumor specimens. The results showed that accurate 3D reconstruction of resected tongue tumor specimens could be accomplished by an acquisition using a single sweep method assisted by rails, applying the highest US frequency possible and performed by one operator. Unexpectedly, a preferred reconstruction algorithm could not be found.

Secondly, the goal was to show as a proof of concept that deep learning is a feasible technique for fast automatic multi-class segmentation of tongue specimen and tumor in 3D freehand US volumes. This study implicates that fast automatic segmentation of tongue specimen and tumor in 3D US volume by deep learning is a feasible technique for intra-operative assessment of resection margins, based on a DSC of 86% and 43% DSC in segmenting the specimen and tumor, respectively. Additionally, a full intra-operative workflow was created from data acquisition to intuitive visualization of close resection margins using a colormap.

The goal of the final study was to determine the correlation between resection margins in tongue tumor specimens assessed by 3D US and histopathology. Unfortunately, this study found that assessment of resection margins in tongue tumor specimens by 3D US and histopathology do not correlate statistically significant.

The overall project resulted in a clinical relevancy that for now, physicians can already be provided of intra-operative feedback by assessing the resection margin in 2D US besides the physical examination. The resection margins of the entire specimen could be determined by manual segmentation if the additional time and possible variabilities are taken into account.

Future research should focus on the improvement of data acquisition, by utilizing a high frequency US transducer and stabilization rails, and remodeling the histopathological slices into a 3D model and registering towards the 3D US models. This would probably result in more accurate annotations by the radiologist. It is expected, as a consequence, that the deep learning models will become more

accurate in predicting specimen and tumor in the 3D US volumes. Eventually, it is speculated that recalculating the correlation between the resection margin in tongue tumor specimens by 3D US and histopathology could be statistically significant. Additionally, research should focus on adequate orientation of involved margins to the in-situ resection field. This would be as response to act upon the found involved margins and perform a secondary resection to rectify and prevent local recurrence.

References

[1] Integraal kankercentrum Nederland. *Cijfers over kanker*. 2019. url: https://www.cijfersoverkanker.nl/nkr/index.

[2] Klaus-Dietrich Wolff, Markus Follmann, and Alexander Nast. "The Diagnosis and Treatment of Oral Cavity Cancer". In: *Deutsches Aerzteblatt Online* 109.48 (2012). doi: 10.3238/arztebl.2012.0829.

[3] Mohamed A.F. Mourad and Mahmoud M. Higazi. "MRI prognostic factors of tongue cancer: Potential predictors of cervical lymph nodes metastases". In: *Radiology and Oncology* 53.1 (2019), pp. 49–56. issn: 15813207. doi: 10.2478/raon-2019-0012.

[4] Thamirys Dantas Nóbrega et al. "Clinicopathological evaluation and survival of patients with squamous cell carcinoma of the tongue". In: *Medicina Oral Patologia Oral y Cirugia Bucal* 23.5 (2018), e579–e587. issn: 16986946. doi: 10.4317/medoral.22421.

[5] Kiran B. Jadhav and Nidhi Gupta. "Clinicopathological prognostic implicators of oral squamous cell carcinoma: Need to understand and revise". In: *North American Journal of Medical Sciences* 5.12 (2013), pp. 671–679. issn: 22501541. doi: 10.4103/1947-2714.123239.

[6] J. Sciubba. "The importance of early diagnosis and treatment". In: *American Journal of Clinical Dermatology* (2001), p. 13. issn: 00034819. doi: 10.7326/0003-4819-51-6-1427.

[7] Kittipong Dhanuthai et al. "Oral cancer: A multicenter study". In: *Medicina Oral Patologia Oral y Cirugia Bucal* 23.1 (2018), e23–e29. issn: 16986946. doi: 10.4317/medoral.21999.

[8] César Rivera. "Essentials of oral cancer". In: *International Journal of Clinical and Experimental Pathology* 8.9 (2015), pp. 11884–11894. issn: 19362625. doi: 10.5281/zenodo.192487.

[9] R. Sankaranarayanan. "Cancer". In: *Disease control priorities* 3.9 (2013), pp. 85–100. issn: 1098-6596. doi: 10.1017/CBO9781107415324.004.

[10] William M. Lydiatt et al. "Head and neck cancers-major changes in the American Joint Committee on cancer eighth edition cancer staging manual". In: *CA: A Cancer Journal for Clinicians* 67.2 (2017), pp. 122–137. issn: 1542-4863. doi: 10.3322/caac.21389.

[11] AKIHIKO MIYAWAKI et al. "Intraoperative frozen section histological analysis of resection samples is useful for the control of primary lesions in patients with oral squamous cell carcinoma". In: *Molecular and Clinical Oncology* 3.1 (2015), pp. 55–62. issn: 2049-9450. doi: 10.3892/mco.2014.409.

[12] Stefan C.A. Steens et al. "Evaluation of tongue squamous cell carcinoma resection margins using ex-vivo MR". In: *International Journal of Computer Assisted Radiology and Surgery* 12.5 (2017), pp. 821–828. issn: 18616429. doi: 10.1007/s11548-017-1524-6.

[13] David N. Sutton et al. "The prognostic implications of the surgical margin in oral squamous cell carcinoma". In: *International Journal of Oral and Maxillofacial Surgery* 32.1 (2003), pp. 30–34. issn: 09015027. doi: 10.1054/ijom.2002.0313.

[14] Matteo Alicandri-Ciufelli et al. "Surgical margins in head and neck squamous cell carcinoma: What is 'close'?" In: *European Archives of Oto-Rhino-Laryngology* 270.10 (2013), pp. 2603–2609. issn: 09374477. doi: 10.1007/s00405-012-2317-8.

[15] Susan G. Brouwer de Koning et al. "Toward assessment of resection margins using hyperspectral diffuse reflection imaging (400–1,700 nm) during tongue cancer surgery". In: *Lasers in Surgery and Medicine* 2 (2019), pp. 1–7. issn: 0196-8092. doi: 10.1002/lsm.23161.

[16] Julia Anne Woolgar and Asterios Triantafyllou. “A histopathological appraisal of surgical margins in oral and oropharyngeal cancer resection specimens”. In: *Oral Oncology* 41.10 (2005), pp. 1034–1043. issn: 13688375. doi: 10.1016/j.oraloncology.2005.06.008.

[17] Tim. Helliwell and Julia. Woolgar. “Standards and datasets for reporting cancers. Dataset for histopathology reporting of mucosal malignancies of the oral cavity”. In: *The Royal College of Pathologists* November 2013 (2013), pp. 1–32.

[18] Bo Wang et al. “The recurrence and survival of oral squamous cell carcinoma: A report of 275 cases”. In: *Chinese Journal of Cancer* 32.11 (2013), pp. 614–618. issn: 1000467X. doi: 10.5732/cjc.012.10219.

[19] M. Weijers et al. “The status of the deep surgical margins in tongue and floor of mouth squamous cell carcinoma and risk of local recurrence; an analysis of 68 patients”. In: *International Journal of Oral and Maxillofacial Surgery* 33.2 (2004), pp. 146–149. issn: 09015027. doi: 10.1054/ijom.2002.0469.

[20] Mi Jin Mun et al. “Histopathologic evaluations of the lingual artery in healthy tongue of adult cadaver”. In: *Clinical and Experimental Otorhinolaryngology* 9.3 (2016), pp. 257–262. issn: 20050720. doi: 10.21053/ceo.2015.01137.

[21] S. G. Brouwer de Koning et al. “The oral cavity tumor thickness: Measurement accuracy and consequences for tumor staging”. In: *European Journal of Surgical Oncology* 45.11 (2019), pp. 2131–2136. issn: 15322157. doi: 10.1016/j.ejso.2019.06.005. url: https://doi.org/10.1016/j.ejso.2019.06.005.

[22] S G Brouwer de Koning and M B Karakullukcu. “Ultrasound aids in intra-operative deep resection margin assessment of squamous cell carcinoma of the tongue”. In: *British Journal of Oral & Maxillofacial Surgery* ().

[23] Michael L. Hinni, Matthew A. Zarka, and Joseph M. Hoxworth. “Margin mapping in transoral surgery for head and neck cancer”. In: *Laryngoscope* 123.5 (2013), pp. 1190–1198. issn: 0023852X. doi: 10.1002/lary.23900.

[24] Cheng K. Ong and Vincent F.H. Chong. “Imaging of tongue carcinoma”. In: *Cancer Imaging* 6.1 (2006), pp. 186–193. issn: 14707330. doi: 10.1102/1470-7330.2006.0029.

[25] Stephen A. Gravina, Gregory L. Yep, and Mehmood Khan. “Human biology of taste”. In: *Annals of Saudi Medicine* 33.3 (2013), pp. 217–222. issn: 02564947. doi: 10.5144/0256-4947.2013.217.

[26] Maureen Stone et al. “Structure and variability in human tongue muscle anatomy”. In: *Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization* 6.5 (2018), pp. 499–507. issn: 21681171. doi: 10.1080/21681163.2016.1162752.

[27] Liancai Mu. “Human tongue neuroanatomy: Nerve supply and Motor endplates”. In: *Clinical anatomy* 23.1 (2006), p. 27. issn: 15378276. doi: 10.1002/ca.21011..

[28] M. Kayalioglu. “Roles of intrinsic and extrinsic tongue muscles in feeding: electromyographic study in pigs”. In: *Oral Biology* 23.1 (2008), pp. 1–7. issn: 15378276. doi: 10.1038/jid.2014.371.

[29] I. Sanders. “A 3-Dimensional atlas of human tongue muscles”. In: *Anatomical reconstrcution* (2014), p. 14. issn: 15378276. doi: 10.1002/ar.22711..

[30] Charlotte M. Mistretta and Archana Kumari. “Tongue and Taste Organ Biology and Function: Homeostasis Maintained by Hedgehog Signaling”. In: *Annual Review of Physiology* 79.1 (2017), pp. 335–356. issn: 0066-4278. doi: 10.1146/annurev-physiol-022516-034202.

[31] Soo Jeong Hong et al. “Tongue growth during prenatal development in Korean fetuses and embryos”. In: *Journal of Pathology and Translational Medicine* 49.6 (2015), pp. 497–510. issn: 23837845. doi: 10.4132/jptm.2015.09.17.

[32] Bruno Bordoni et al. “The Anatomical Relationships of the Tongue with the Body System”. In: *Cureus* 10.12 (2018), pp. 1–7. issn: 2168-8184. doi: 10.7759/cureus.3695.

[33] Amy J. Thorsen and Gaio E. Lakin. “Basic physics of ultrasonography”. In: *Seminars in Colon and Rectal Surgery* 21.4 (2010), pp. 186–190. issn: 10431489. doi: 10.1053/j.scrs.2010.09.001.

[34] Fikri M. Abu-Zidan, Ashraf F. Hefny, and Peter Corr. “Clinical ultrasound physics”. In: *Journal of Emergencies, Trauma and Shock* 4.4 (2011), pp. 501–503. issn: 09742700. doi: 10.4103/0974-2700.86646.

[35] Alexander EJ Powles et al. “Physics of ultrasound”. In: *Anaesthesia and Intensive Care Medicine* 19.4 (2018), pp. 202–205. issn: 18787584. doi: 10.1016/j.mpaic.2018.01.005. url: https://doi.org/10.1016/j.mpaic.2018.01.005.

[36] G. Karlsson. “The Physics of Ultrasound and Some Recent Techniques Used”. In: *Contemporary Interventional Ultrasonography in Urology* (2009), pp. 103–112. doi: 10.1007/978-1-84800-217-3.

[37] Qinghua Huang and Zhaozheng Zeng. “A Review on Real-Time 3D Ultrasound Imaging Technology”. In: *BioMed Research International* 2017 (2017). issn: 23146141. doi: 10.1155/2017/6027029.

[38] Robert Rohling, Andrew Gee, and Laurence Berman. “A comparison of freehand three-dimensional ultrasound reconstruction techniques”. In: *Medical Image Analysis* 3.4 (1999), pp. 339–359. issn: 13618415. doi: 10.1016/S1361-8415(99)80028-0.

[39] D. Miller et al. “Comparison of different reconstruction algorithms for three-dimensional ultrasound imaging in a neurosurgical setting”. In: *International Journal of medical robotics and computer assisted surgery* (2012), p. 12. doi: 10.1002/rcs.

[40] Ole Vegard Solberg et al. “Freehand 3D Ultrasound Reconstruction Algorithms-A Review”. In: *Ultrasound in Medicine and Biology* 33.7 (2007), pp. 991–1009. issn: 03015629. doi: 10.1016/j.ultrasmedbio.2007.02.015.

[41] Christian Askeland et al. “CustusX: an open-source research platform for image-guided therapy”. In: *International Journal of Computer Assisted Radiology and Surgery* 11.4 (2016), pp. 505–519. issn: 18616429. doi: 10.1007/s11548-015-1292-0.

[42] CustusX. *CustusX Developer Documentation*. url: https://www.custusx.org/uploads/developer_doc/nightly/index.html.

[43] Ole Vegard Solberg et al. “3D ultrasound reconstruction algorithms from analog and digital data”. In: *Ultrasonics* 51.4 (2011), pp. 405–419. issn: 0041624X. doi: 10.1016/j.ultras.2010.11.007.

[44] A Fedorov. “3D Slicer as an Image Computing Platform for the Quantative Imaging Network”. In: *Magn. Reson. Imaging* (2012), p. 28. doi: 10.1016/j.mri.2012.05.001.3D. url: http://www.cabdirect.org/abstracts/19941607462.html;jsessionid=3B9A6CF71444FF0AA753CB8CBBCA3B59.

[45] Mercy Afadzi et al. “Image Quality Measured From Ultra-Low Dose Chest Computed Tomography Examination Protocols Using 6 Different Iterative Reconstructions From 4 Vendors, a Phantom Study”. In: *Journal of computer assisted tomography* 44.1 (2020), pp. 95–101. issn: 15323145. doi: 10.1097/RCT.0000000000000947.

[46] D. Liao et al. “Analysis of surface geometry of the human stomach using real-time 3-D ultrasonography in vivo”. In: *Neurogastroenterology and Motility* 16.3 (2004), pp. 315–324. issn: 13501925. doi: 10.1111/j.1365-2982.2004.00522.x.

[47] Roeland W.H. Smits et al. “Resection margins in oral cancer surgery: Room for improvement”. In: *Head & Neck* 38.S1 (Apr. 2016). Ed. by David W. Eisele, E2197–E2203. issn: 10433074. doi: 10.1002/hed.24075. url: http://doi.wiley.com/10.1002/hed.24075.

[48] Gordon N. Stevenson et al. “3-D Ultrasound Segmentation of the Placenta Using the Random Walker Algorithm: Reliability and Agreement”. In: *Ultrasound in Medicine and Biology* 41.12 (2015), pp. 3182–3193. issn: 1879291X. doi: 10.1016/j.ultrasmedbio.2015.07.021.

[49] Padraig Looney et al. “Automatic 3D ultrasound segmentation of the first trimester placenta using deep learning”. In: *Proceedings - International Symposium on Biomedical Imaging* (2017), pp. 279–282. issn: 19458452. doi: 10.1109/ISBI.2017.7950519.

[50] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-net: Convolutional networks for biomedical image segmentation”. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 9351 (2015), pp. 234–241. issn: 16113349. doi: 10.1007/978-3-319-24574-4{_}28.

[51] Jian Zhu, Will Styler, and Ian Calloway. “A CNN-based tool for automatic tongue contour tracking in ultrasound images”. In: (2019), pp. 1–6. url: http://arxiv.org/abs/1907.10210.

[52] Huiyan Jiang et al. “A region growing vessel segmentation algorithm based on spectrum information”. In: *Computational and Mathematical Methods in Medicine* 2013 (2013), pp. 1–9. issn: 17486718. doi: 10.1155/2013/743870.

[53] Kh Rezaee and J Haddadnia. “Designing an Algorithm for Cancerous Tissue Segmentation Using Adaptive K-means Cluttering and Discrete Wavelet Transform”. In: *J Biomed Phys Eng.* (2013), pp. 93–104.

[54] T. Panch. “Artificial intelligence, machine learning and health systems”. In: *Journal of Global Health* 8.2 (2018), pp. 1–8. issn: 2047-2978. doi: 10.7189/jogh.08.020303.

[55] F. Jiang et al. “Artificial intelligence in healthcare: Past, present and future”. In: *Stroke and Vascular Neurology* 2.4 (2017), pp. 230–243. issn: 20598696. doi: 10.1136/svn-2017-000101.

[56] O. Gupta. “Distributed learning of deep neural network over multiple agents”. In: *Journal of Network and Computer Applications* 116.May (2018), pp. 1–8. issn: 10958592. doi: 10.1016/j.jnca.2018.05.003.

[57] Liansheng Wang et al. “Nested dilation networks for brain tumor segmentation based on magnetic resonance imaging”. In: *Frontiers in Neuroscience* 13.APR (2019), pp. 1–14. issn: 1662453X. doi: 10.3389/fnins.2019.00285.

[58] Zeyu Jiang et al. “Two-Stage Cascaded U-Net: 1st Place Solution to BraTS Challenge 2019 Segmentation Task”. In: *Crimi a., Bakas S. BrainLesion: Glioma, Mulitple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2019* 11992 (2019), pp. 231–241. doi: 10.1007/978-3-030-46640-4. url: https://doi.org/10.1007/978-3-030-46640-4_22.

[59] Xinchao Cheng et al. “Memory-efficient cascade 3d u-net for brain tumor segmentation”. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 11992 LNCS.December 2019 (2020), pp. 242–253. issn: 16113349. doi: 10.1007/978-3-030-46640-4{_}23.

[60] Fausto Milletari, Nassir Navab, and Seyed Ahmad Ahmadi. “V-Net: Fully convolutional neural networks for volumetric medical image segmentation”. In: *Proceedings - 2016 4th International Conference on 3D Vision, 3DV 2016* (2016), pp. 565–571. doi: 10.1109/3DV.2016.79.

[61] Wan Ru Jia et al. “Three-dimensional Contrast-enhanced Ultrasound in Response Assessment for Breast Cancer: A Comparison with Dynamic Contrast-enhanced Magnetic Resonance Imaging and Pathology”. In: *Scientific Reports* 6.April (2016), pp. 1–10. issn: 20452322. doi: 10.1038/srep33832. url: http://dx.doi.org/10.1038/srep33832.

[62] Ahmed M. Hashad, Nashwa E. Hassan, and Ahmed E. Elbohoty. “3D Ultrasonography Compared with Magnetic Resonance Imaging for the Diagnosis of Adenomyosis”. In: *The Egyptian Journal of Hospital Medicine* 69.8 (2017), pp. 3123–3133. issn: 16872002. doi: 10.12816/0042864.

[63] Satoru Shintani et al. “The usefulness of intraoral ultrasonography in the evaluation of oral cancer”. In: *International Journal of Oral and Maxillofacial Surgery* 30.2 (2001), pp. 139–143. issn: 09015027. doi: 10.1054/ijom.2000.0035.

[64] A. Yesuratnam et al. “Preoperative evaluation of oral tongue squamous cell carcinoma with intraoral ultrasound and magnetic resonance imaging - Comparison with histopathological tumour thickness and accuracy in guiding patient management”. In: *International Journal of Oral and Maxillofacial Surgery* 43.7 (2014), pp. 787–794. issn: 13990020. doi: 10.1016/j.ijom.2013.12.009. url: http://dx.doi.org/10.1016/j.ijom.2013.12.009.

[65] Maria Cecilia Lunardelli da Silva and Doryane Maria dos Reis Lima. “Correlation of the three-dimensional ultrasound findings with pathology in patients with deep pelvic infiltrating endometriosis submitted to surgery”. In: *Journal of Coloproctology* 36.2 (2016), pp. 69–74.

[66] Adam K. Glaser et al. “Light-sheet microscopy for slide-free non-destructive pathology of large clinical specimens”. In: *Nature Biomedical Engineering* 1.7 (2017). issn: 2157846X. doi: 10.1038/s41551-017-0084.

[67] Gary Tozbikian, Edi Brogi, and Christina E. Vallejo. “Atypical Ductal Hyperplasia Bordering on Ductal Carcinoma In Situ: Interobserver Variability and Outcomes in 105 Cases”. In: *Physiology & behavior* 176.12 (2017), pp. 139–148. doi: 10.1016/j.physbeh.2017.03.040.

[68] Mahul B. Amin et al., eds. *AJCC Cancer Staging Manual*. 8th. Springer International Publishing, 2017, pp. XVII, 1032. isbn: 978-3-319-40617-6.

8
Appendix A

Table IX. Lip, oral cavity, and non-HPV oropharynx stages following the 8th edition of the AJCC Cancer Staging Manual [68]

AJCC stage	Stage grouping	Lip, oral cavity and non-HPV oropharynx stage description*
0	Tis N0 M0	The cancer is still within the epithelium (the top layer of cells lining the oral cavity and oropharynx) and has not yet grown into deeper layers. It has not spread to nearby lymph nodes (N0) or distant sites (M0). This stage is also known as carcinoma in situ (Tis).
I	T1 N0 M0	The cancer is 2 cm (about ¾ inch) or smaller. It's not growing into nearby tissues (T1). It has not spread to nearby lymph nodes (N0) or to distant sites (M0).
II	T2 N0 M0	The cancer is larger than 2 cm but no larger than 4 cm (about 1½ inch). It's not growing into nearby tissues (T2). It has not spread to nearby lymph nodes (N0) or to distant sites (M0).
III	T3 N0 M0	The cancer is larger than 4 cm (T3). For cancers of the oropharynx, T3 also includes tumors that are growing into the epiglottis (the base of the tongue). It has not spread to nearby lymph nodes (N0) or to distant sites (M0).
	OR	
	T1, T2, T3 N1 M0	The cancer is any size and may have grown into nearby structures if oropharynx cancer(T1-T3) AND has spread to 1 lymph node on the same side as the primary tumor. The cancer has not grown outside of the lymph node and the lymph node is no larger than 3 cm (about 1¼ inch) (N1). It has not spread to distant sites (M0).

To be continued on the next page...

AJCC stage	Stage grouping	Lip, oral cavity and non-HPV oropharynx stage description*
IVa	T4a N0 or N1 M0	The cancer is any size and is growing into nearby structures such as:For lip cancers: nearby bone, the inferior alveolar nerve (the nerve to the jawbone), the floor of the mouth, or the skin of the chin or nose (T4a)For oral cavity cancers: the bones of the jaw or face, deep muscle of the tongue, skin of the face, or the maxillary sinus (T4a)For oropharyngeal cancers: the larynx (voice box), the tongue muscle, or bones such as the medial pterygoid, the hard palate, or the jaw (T4a).This is known as moderately advanced local disease (T4a).,AND either of the following:It has not spread to nearby lymph nodes (N0)It has spread to 1 lymph node on the same side as the primary tumor, but has not grown outside of the lymph node and the lymph node is no larger than 3 cm (about 1¼ inch) (N1).It has not spread to distant sites (M0).
	OR	
	T1, T2, T3 or T4a N2 M0	The cancer is any size and may have grown into nearby structures (T0-T4a). It has not spread to distant organs (M0). It has spread to one of the following:1 lymph node on the same side as the primary tumor, but it has not grown outside of the lymph node and the lymph node is larger than 3 cm but not larger than 6 cm (about 2½ inches) (N2a) OR It has spread to more than 1 lymph node on the same side as the primary tumor, but it has not grown outside of any of the lymph nodes and none are larger than 6 cm (N2b) OR It has spread to 1 or more lymph nodes either on the opposite side of the primary tumor or on both sides of the neck, but has not grown outside any of the lymph nodes and none are larger than 6 cm (N2c).

To be continued on the next page...

AJCC stage	Stage grouping	Lip, oral cavity and non-HPV oropharynx stage description*
IVb	Any T N3 M0	The cancer is any size and may have grown into nearby soft tissues or structures (Any T) AND any of the following:It has spread to 1 lymph node that's larger than 6 cm but has not grown outside of the lymph node (N3a) OR It has spread to 1 lymph node that's larger than 3 cm and has clearly grown outside the lymph node (N3b) OR It has spread to more than 1 lymph node on the same side, the opposite side, or both sides of the primary cancer with growth outside of the lymph node(s) (N3b) OR It has spread to 1 lymph node on the opposite side of the primary cancer that's 3 cm or smaller and has grown outside of the lymph node (N3b). It has not spread to distant organs (M0).
	OR	
	T4b Any N M0	The cancer is any size and is growing into nearby structures such as the base of the skull or other bones nearby, or it surrounds the carotid artery. This is known as very advanced local disease (T4b). It might or might not have spread to nearby lymph nodes (Any N). It has not spread to distant organs (M0).
IVc	Any T Any N M1	The cancer is any size and may have grown into nearby soft tissues or structures (Any T) AND it might or might not have spread to nearby lymph nodes (Any N). It has spread to distant sites such as the lungs (M1).

* The following additional categories are not described in the table above:
TX: Main tumor cannot be assessed due to lack of information.
T0: No evidence of a primary tumor.
NX: Regional lymph nodes cannot be assessed due to lack of information.

9 Appendix B

Table X. The SNR and CNR for each acquisition method of the five reconstruction algorithms. The numbers in **bold** represent the highest value within each reconstruction algorithm. Results of experiment two.

Reconstruction algorithm	Acquisition method	SNR	CNR
PNN	Single, 10 sec.	22.4	979
	Single, 3 sec.	**25.8**	**1307**
	Double, 6 sec.	14.2	398
	Triple, 9 sec.	4.7	43.7
VNN	Single, 10 sec.	19.2	722
	Single, 3 sec.	**22.8**	**1008**
	Double, 6 sec.	13.3	345
	Triple, 9 sec.	4.9	46.8
VNN2	Single, 10 sec.	20.1	794
	Single, 3 sec.	**28.4**	**1540**
	Double, 6 sec.	20.9	865
	Triple, 9 sec.	9.2	166
DW	Single, 10 sec.	22.0	952
	Single, 3 sec.	**28.0**	**1502**
	Double, 6 sec.	22.3	986
	Triple, 9 sec.	8.8	152
Anisotropic	Single, 10 sec.	27.9	1416
	Single, 3 sec.	**34.5**	2167
	Double, 6 sec.	**34.5**	**2351**
	Triple, 9 sec.	19.3	737

Table XI. The measured dimensions of the spherical structure in the phantom for each reconstruction algorithm of three acquisition methods. The distance error compared to the CT is expressed as percentage. Estimated dimensions are highlighted in red. AP = Anterior-Posterior, RL = Right - Left, IS = Inferior - Superior.

	Direction **CT**	**AP** *mm* (%) 37	**RL** *mm* (%) 43	**IS** *mm* (%) 34
Acquisition method	**Reconstruction method**			
Single, 10 sec.	PNN	35 (5.4)	38 (11.6)	40 (17.6)
	VNN	35 (5.4)	40 (7.0)	**32** (5.8)
	VNN2	35 (5.4)	39 (9.3)	31 (8.8)
	DW	35 (5.4)	37 (14.0)	**32** (5.8)
	Anisotropic	**36 (2.7)**	**44** (2.3)	31 (8.8)
Single, 3 sec.	PNN	35 (5.4)	40 (7.0)	**32** (5.8)
	VNN	**36** (2.7)	41 (4.6)	**32** (5.8)
	VNN2	35 (5.4)	**43** (0)	**32** (5.8)
	DW	**36** (2.7)	42 (2.3)	31 (8.8)
	Anisotropic	**36** (2.7)	44 (2.3)	31 (8.8)
Double, 6 sec.	PNN	**36** (2.7)	36 (16.3)	**32** (5.8)
	VNN	**36** (2.7)	36 (16.3)	31 (8.8)
	VNN2	**36** (2.7)	40 (7.0)	31 (8.8)
	DW	**36** (2.7)	36 (16.3)	31 (8.8)
	Anisotropic	**36** (2.7)	**42** (2.3)	31 (8.8)
Triple, 9 sec.	PNN	**36** (2.7)	42 (2.3)	31 (8.8)
	VNN	35 (5.4)	**43** (0)	**32** (5.8)
	VNN2	**36** (2.7)	45 (4.6)	**32** (5.8)
	DW	35 (5.4)	46 (7.0)	**32** (5.8)
	Anisotropic	35 (5.4)	44 (2.3)	**32** (5.8)

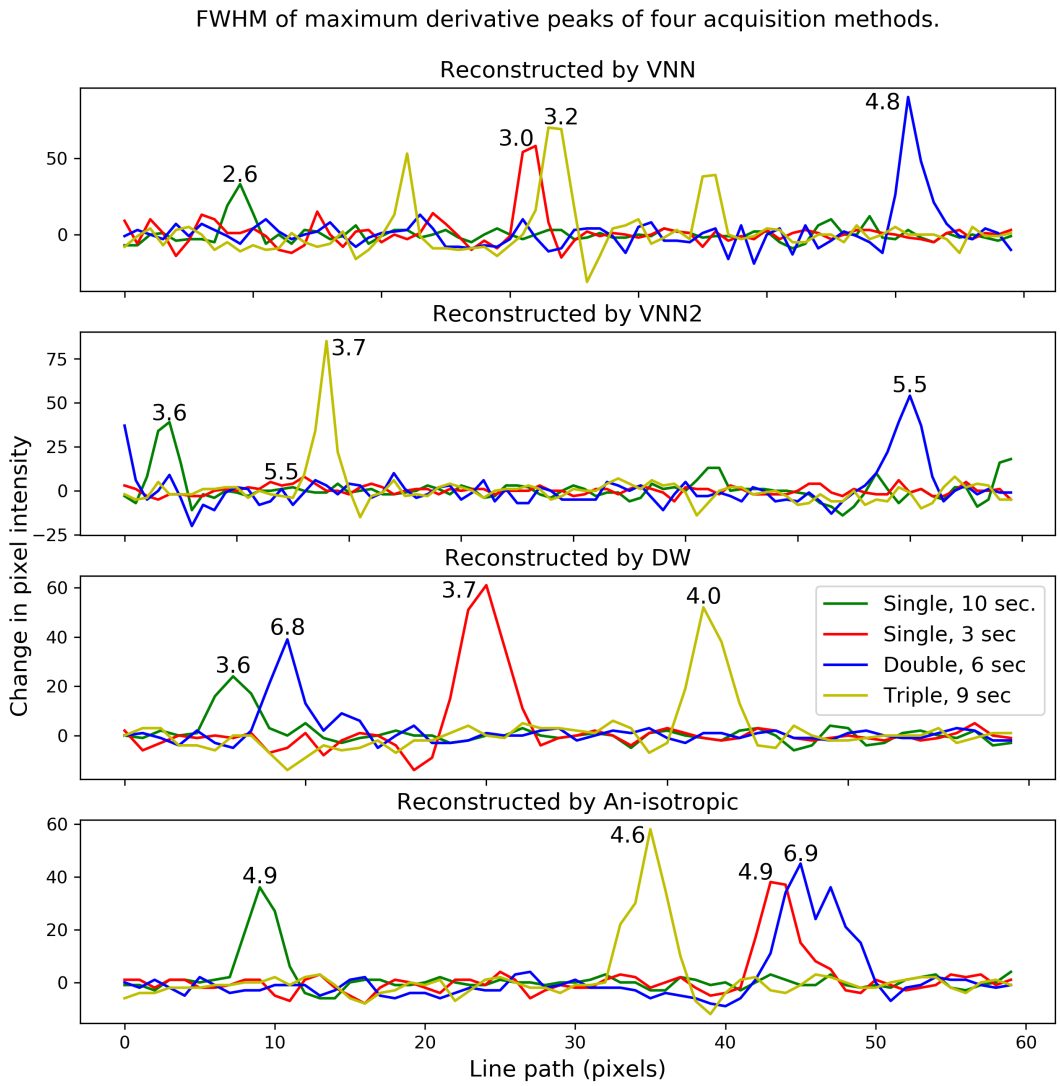


Figure 22. The derivatives of the pixel intensity along a scan line at the transition from phantom to the spherical structure in four 3D US volumes with different acquisition methods. From top to bottom the 3D US volumes were reconstructed, respectively by the VNN, VNN2, DW and anisotropic algorithm.

10
Appendix C

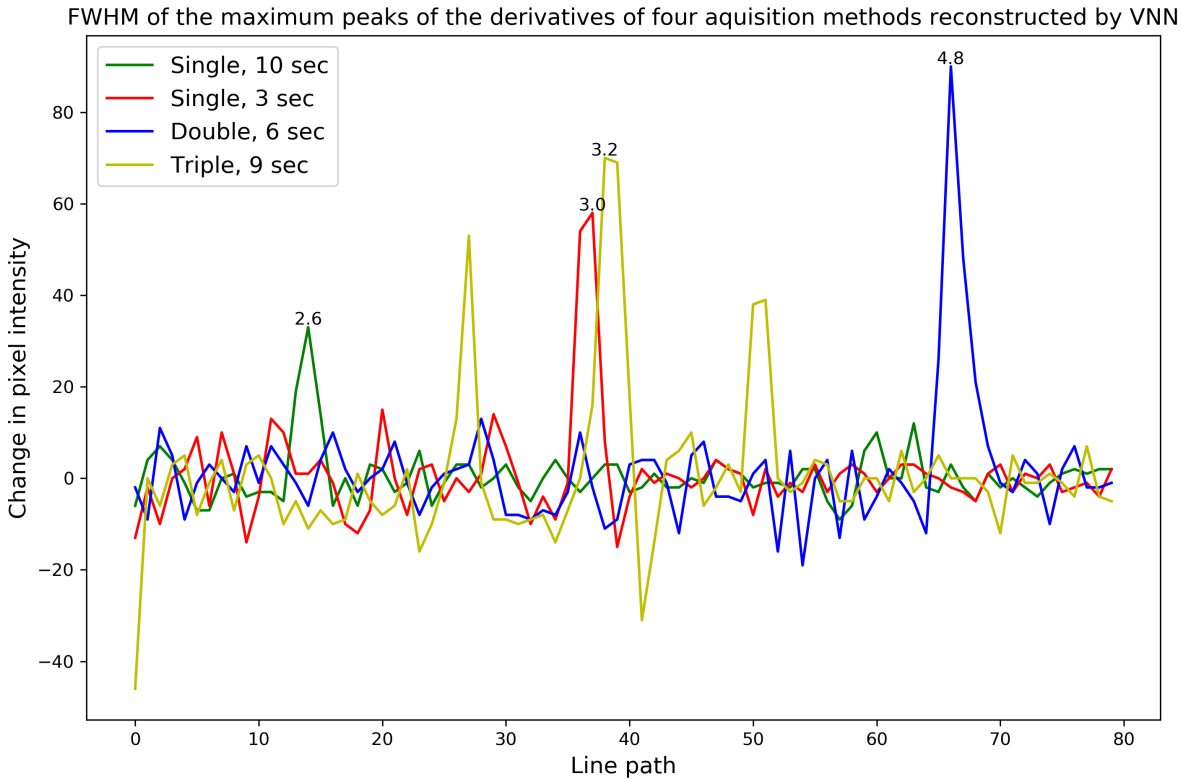


Figure 23. Experiment 2: The derivatives of the pixel intensity along a scan line at the transition from phantom to the spherical structure in four US 3D volumes with different acquisition methods. All 3D US volumes were reconstructed by the VNN algorithm. The FWHM of the maximum peak is plotted at the top of the peak.

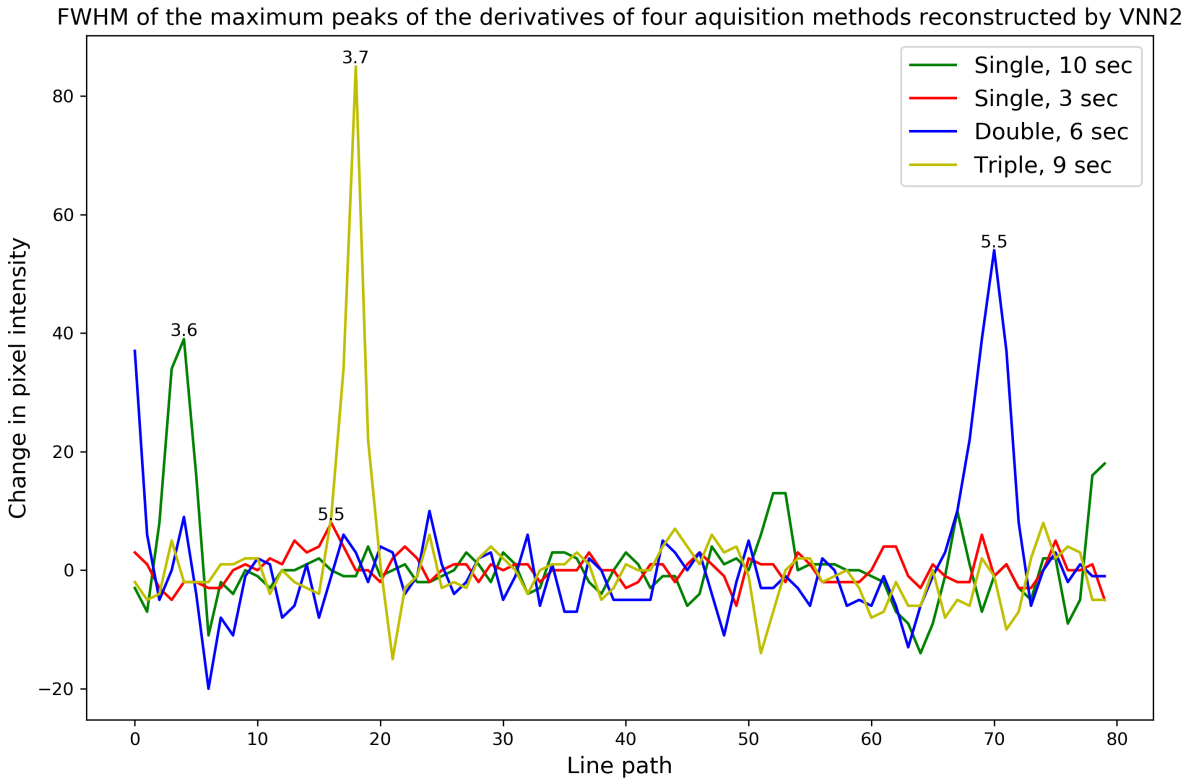


Figure 24. Experiment 2: The derivatives of the pixel intensity along a scan line at the transition from phantom to the spherical structure in four US 3D volumes with different acquisition methods. All 3D US volumes were reconstructed by the VNN2 algorithm. The FWHM of the maximum peak is plotted at the top of the peak.

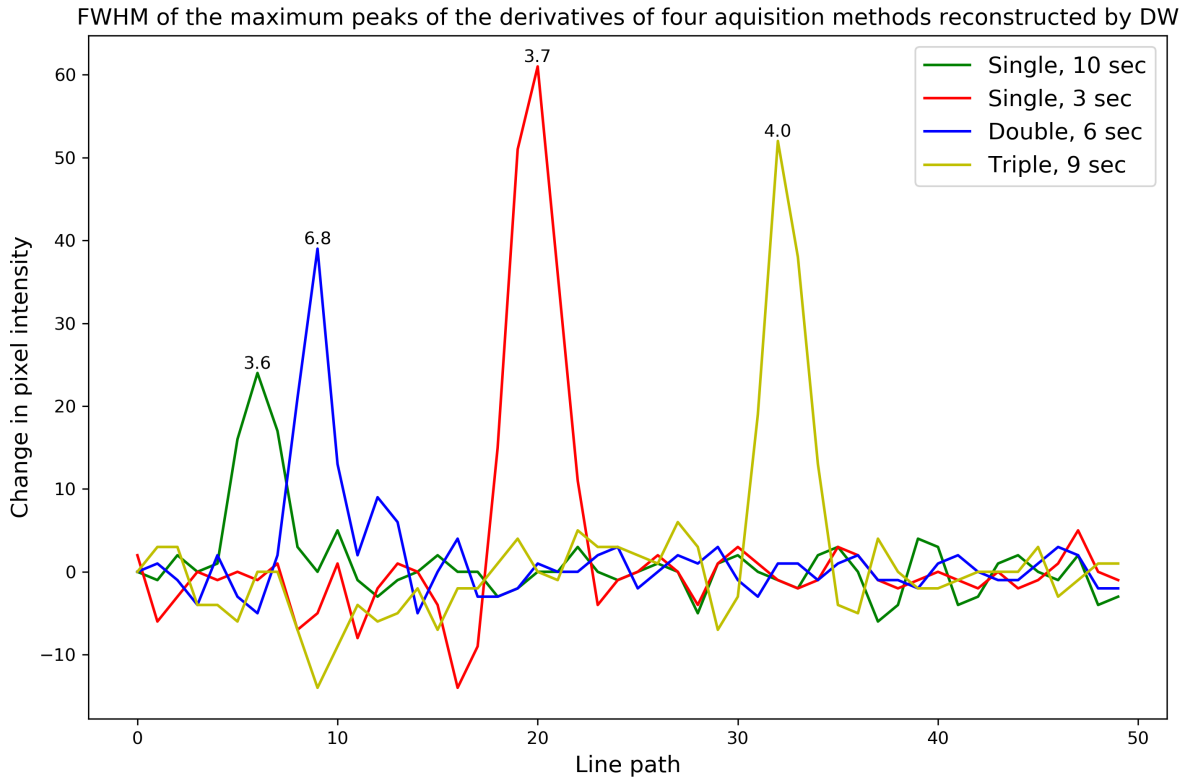


Figure 25. Experiment 2: The derivatives of the pixel intensity along a scan line at the transition from phantom to the spherical structure in four US 3D volumes with different acquisition methods. All 3D US volumes were reconstructed by the DW algorithm. The FWHM of the maximum peak is plotted at the top of the peak.

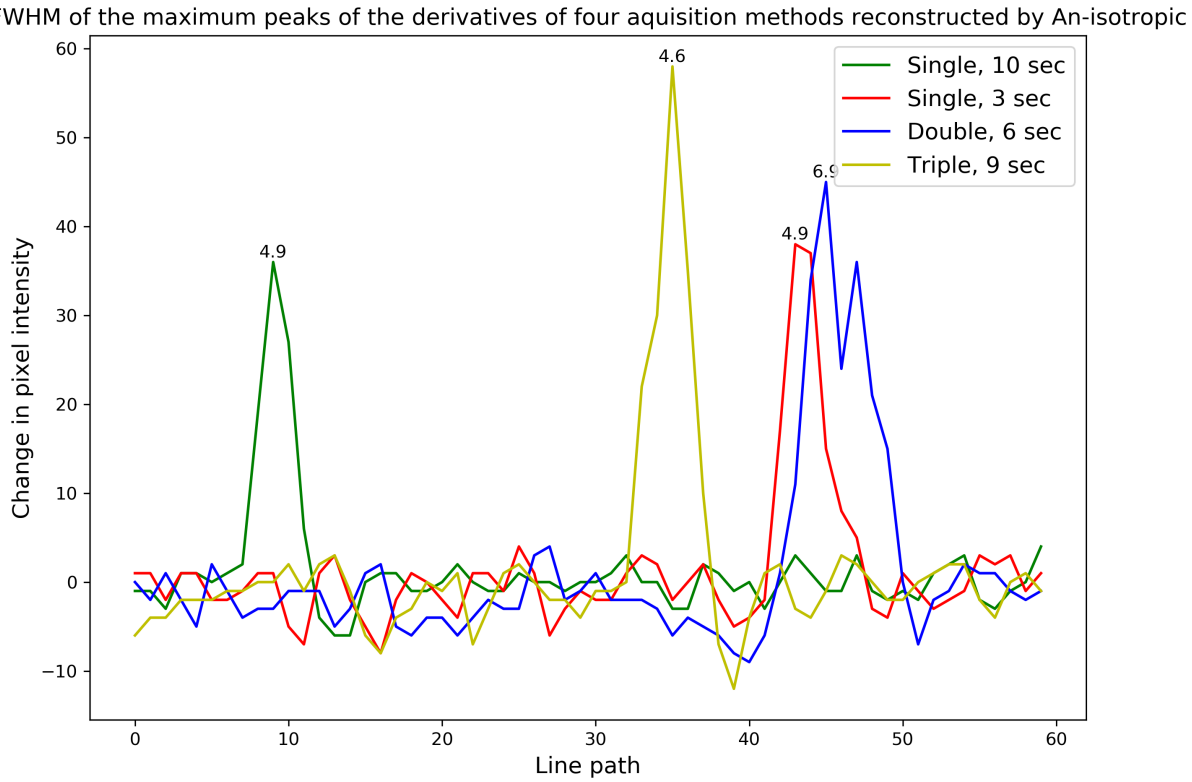


Figure 26. Experiment 2: The derivatives of the pixel intensity along a scan line at the transition from phantom to the spherical structure in four US 3D volumes with different acquisition methods. All 3D US volumes were reconstructed by the anisotropic algorithm. The FWHM of the maximum peak is plotted at the top of the peak.