Stitching of sensor-navigated 3D ultrasound datasets for the determination of large thyroid volumes – a phantom study

Martin Freesmeyer, Leonard Knichel, Christian Kühnel, Thomas Winkens

Clinic of Nuclear Medicine, Jena University Hospital, Jena, Germany

Abstract

Aims: Thyroid volume has to be measured in goiters prior to radioiodine treatment to calculate the needed amount of radioactivity. Modern clinical equipment for ultrasonography shows the trend to smaller probes, so that larger goiters do not fit any longer into the probe’s field of view. This study evaluated the feasibility and accuracy of stitching procedures applied to thyroid volumetric analysis performed using three-dimensional ultrasound (3D-US). Material and methods: Ad hoc thyroid phantoms of different shapes (regular, nodular, thickened isthmus) and volumes (ranging between 50 and 400 mL) were developed. In 15 such phantoms the left and right lobes were separately scanned, and the 3D-US datasets were then assembled (stitched) using predefined landmarks and dedicated software. Volumetric analysis was then assessed via a conventional ellipsoid model (em) and manual tracing (mt). The correlation of measured and reference volumes was determined using Pearson’s correlation coefficients and Bland and Altman limits of agreement. Results: The results showed a high level of agreement (with correlation coefficients ranging from 0.974 to 0.999) for all shapes and volumes tested, including the largest volume of 400 mL. The mt method, although more time consuming, proved superior to the em. Conclusions: Stitching-mediated thyroid volumetric analysis is accurate, and its clinical performance should be investigated in future studies.

Keywords: Stitching; Volumetric analysis; Thyroid phantom; 3D-US; Extended field of view

Introduction

The size and shape of the thyroid gland vary greatly, depending on environmental and individual factors such as iodine availability, age, gender and smoking habits [1,2]. Determination of the thyroid volume is relevant not only for the primary diagnostics of thyroid disorders, but also for the planning of radiotherapy and monitoring of therapeutic effects [3,4]. Ultrasound imaging, mostly two-dimensional (2D-US), is the standard approach for the determination of thyroid volume [5,6]. Important advantages of US include the anatomical resolution, soft-tissue contrast, cost effectiveness, and ubiquitous availability. Other relevant aspects include the absence of contraindications or radiation exposure issues [6,7].

Volume assessment via 2D-US is based on the general assumption that the thyroid lobes have an ideal form and that a rotation ellipsoid model (em) can thus be applied [8]. However, the routine use of em can lead to substantial inter- and intra-observer variability because individual anatomical variants of the thyroid, such as a thickened isthmus or a nodular goiter, can lead to volumetric errors of up to 30% [4,9].

A possible method to overcome these limitations is the use of three-dimensional ultrasound (3D-US), which not only allows for the generation of volume datasets in three dimensions, but also permits data storage for post hoc analyses independent of the actual scanning session [4]. Furthermore, volumetric analysis via 3D-US allows the use of manual tracing (mt), a method that shows very high levels of agreement with the actual thyroid volumes. A drawback of 3D-US, however, is that the size of the thyroid may exceed the field of view (FOV) of the 3D

Received 20.08.2018 Accepted 28.10.2018
Med Ultrason
2018, Vol. 20, No 4, 480-486
Corresponding author: PD Dr. med. Martin Freesmeyer, M.D.
Clinic of Nuclear Medicine
Jena University Hospital
Am Klinikum 1, 07747 Jena, Germany
Phone: +49-3641-9329801
Fax: +49-3641-9329802
E-mail: martin.freesmeyer@med.uni-jena.de

DOI: 10.11152/mu-1687
probe [7], making it impossible to image the whole gland with a single scan [9].

The spatial limitation of the 3D-US can be potentially resolved by *post hoc* assembly of two or more 3D-US datasets using the stitching method, which results in an extended volume or extended FOV (3D-US EFOV). This method, however, has not yet been applied for thyroid imaging or for organ volumetric analysis.

Therefore, the aims of the present study were 1) to investigate whether stitching of two separately scanned 3D-US partial thyroid volumes is feasible and 2) to assess the accuracy of the resulting volumetric 3D-US EFOV. For this purpose, *ad hoc* thyroid phantoms with volume and shape variants were developed based on a previously established phantom [10].

**Materials and methods**

**Phantoms**

Fifteen water-filled, heart-shaped balloons were used to produce phantoms with a shape resembling the thyroid gland in 15 predefined volumes (50, 75, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, and 400 mL). For the measurements, the phantoms were placed into a double-walled plastic container filled with water (fig 1). The phantoms were submersed 2 cm below the water surface to enable US scanning. To maintain the phantoms in this container, four small pieces of plastic were inserted into the balloons and then ligated with two threads per pole to the wall of the inner plastic container (fig 2). To reproduce the shapes of the lobes and the typically narrow isthmus of a healthy thyroid, a clamp was applied to the phantom as shown in fig 1 and 2.

The unladen weight (tare) of each empty phantom was measured using a precision scale. Subsequently, the balloons were filled with pre-boiled water to minimize
the formation of air bubbles. The net volumes derived from the weight measurements represented the reference standard for the assessment of the volumetric accuracy of the different US methods. The phantoms were weighed again at the end of the experiments to exclude leakage during the experiments.

After completion of the first measurement cycle with normally shaped thyroid phantoms (fig 2d), a second cycle was performed with deformed lobes as a model of nodular goiter (fig 2e). In the third investigation cycle, the clamp was removed to simulate a model of a thickened isthmus (fig 2f).

**Ultrasonography scans**

All scans were performed with the same US equipment (Logiq E9, GE Medical Systems, Milwaukee, WI, USA) with the application of two different sensor-navigated 3D-US probes: the convex C2-9 probe (C-probe) and the linear ML6-15 matrix probe (L-probe) (fig 1c and 1d). The phantom was placed approximately 2 cm from a magnetic transmitter. Two positional sensors were fixed to the probe to allow recording of the position and movement of the US probe in the magnetic field during scanning. The equipment reconstructed a 3D volume dataset based on the coordinates of the probes and the acquired 2D images. Scans of the right and left lobes of the phantoms were performed separately with inclusion of the isthmus, ensuring that each scan included the three ipsilateral and three contralateral landmarks (fig 3). During the use of the L-probe, a virtual convex mode was used to enlarge the FOV of the primary scan. The depth, focus number, and focus position were adapted to optimize the quality of the image of the corresponding phantom volume, whereas the remaining parameters of the probe setting were kept constant.

**Digital Imaging and Communications in Medicine (DICOM) export**

Storage of the 3D-US datasets in the US equipment was performed with a dedicated DICOM standard for 3D-US (Enhanced Ultrasound Volume Storage; EUVS) as a stack of transverse slices. The stored datasets were finally transferred to an external storage medium.

**Stitching and post-stitching data modification**

The 3D-US datasets of the left and right phantom lobes were subsequently loaded to the stitching module of PMOD software (PMOD, Version 3.409, PMOD Technologies Ltd., Zurich, Switzerland), displayed in a half-transparent manner and oriented to each other so that all six landmarks became superimposed in the axial and sagittal planes (fig 3c and 3d). A combined extended 3D-US dataset, now including the whole phantom with both lobes, was then generated and stored (fig 3e). To reduce the time effort of the subsequent volumetric analyses using mt, the extended 3D-US datasets were converted from a transverse slice thickness of 0.21-0.4 mm to a uniform thickness of 3.0 mm using PMOD software and stored again [10].

**Volumetric analysis**

Volumetric analysis for each phantom was performed by an experienced investigator. All acquired datasets were assessed by both the ellipsoid formula (length$_{max}$ x width$_{max}$ x depth$_{max}$ x 0.5) and the mt procedure. The PMOD software tool “image viewing and volume of interest analysis” was used to evaluate the volumes.

In the case of the em, volumetric analysis was performed separately for each lobe starting from the sagittal standard planes to identify the slices with the maximum length. The transverse plane was then oriented perpendicularly to the sagittal plane to measure the maximum width and depth. For the models with a thickened isthmus, the isthmus volume was divided and assigned to the volume of the right and left lobe so that no separate measurement was performed.

The mt method was applied via the mt of the phantom contours in parallel slices of transverse image slices (fig 4a). In accordance with the em, the left and right
lobes were assessed separately. Only the model with a thickened isthmus was evaluated with a whole contour covering the right lobe, left lobe, and isthmus. Volume calculation was automatically performed by the software.

**Statistical analysis**

The relationship between the calculated and reference volumes of the thyroid phantoms was quantified by Pearson’s correlation coefficient. The correspondence of the calculated and reference volumes was then quantified using a slightly modified version of the Bland and Altman method [11]. The difference between measured and reference volumes was calculated as the relative difference, i.e., as a percentage of the reference volume. These relative differences are described by the mean, standard deviation (SD), median, and range (minimum, maximum). To assess the magnitude of systematic differences, the 95% confidence intervals of the mean were also calculated. When the confidence interval did not include a zero value, the corresponding scanning method was considered to have a significant systematic error (significance level 0.05, overestimation or underestimation of the reference volume). In addition, calculated and reference volumes were compared using the limits of agreement (mean + 1.96*SD) recommended by Bland and Altman [11]. This interval includes 95% of all differences, under the assumption that the differences are normally distributed. All calculations were performed with Stata/IC 13.1 for Windows (StataCorp LP, College Station, TX, USA).

**Results**

Phantoms of different sizes and shapes proved practical and suitable for the planned investigations and all phantoms could be investigated according to the intent of the study. Stitching of the obtained datasets via PMOD also proved feasible. The phantoms, fully captured using 3D-US EFOV, could be successfully subjected to volumetric analysis with both the em and mt methods. In two cases, a leak in the phantoms was detected because of a difference in weight measured before and after the final scan. In both cases, the experiments were repeated.

Acquisition of a sensor-navigated 3D-US dataset lasted 3 to 5 min, irrespective of the probe used. The mt measurements required approximately 15-20 min per phantom, and the em measurements required 1-2 min.

The correlation coefficients of the calculated and reference volumes as well as the mean values, SDs, 95% confidence intervals, and limits of agreement of relative differences for all three phantom shapes (regular, nodular, thickened isthmus) are presented in table I. The high correlation coefficients (range: 0.974-0.999) for all methods investigated – irrespective of the phantom shape – indicated a very strong positive relationship between the calculated and reference volumes.

Although the general level of correspondence was very high, some discrepancies were found between the calculated and reference volumes (fig 5). In general, volumetric analysis via mt showed markedly smaller deviations from the reference volumes than the em method. The em was associated with clearly higher SDs and wider 95% confidence intervals. The volumetric analy-
ses with the C-probe were associated with a significant underestimation of values obtained with the em for the nodular shape and thickened isthmus. The mt method also showed some deviations, i.e., a mild but significant underestimation of phantoms with a regular or nodular shape. Volumetric analysis using the em was clearly impacted by the anatomical variants of the phantom, whereas the mt method was largely unaffected by the shape of the phantoms (fig. 5).

**Discussions**

Volumetric analysis of the thyroid by 2D-US has been largely established in clinical routine practice [9], but 3D-US approaches are also conceivable. In 3D-US, the resulting 3D datasets can be reconstructed in the three standard dimensions and evaluated post hoc, in accordance with computed tomography (CT) and magnetic resonance imaging (MRI) [12]. Indeed, the interpretation of 3D-US images may be less prone to inter-observer variability, given the clearer and more objective images [13]. However, 3D-US presents some limitations [12]. The limited FOV in particular allows only imaging of relatively small volumes compared to CT and MRI [14]. This not only reduces the interpretation of the anatomical context, but also limits the measurement of large volumes, such as in cases of severely enlarged thyroids.

The problem of augmenting the anatomical view is not limited to the 3D-US but also concerns 2D-US. To overcome this limitation, different techniques have been developed, one of which is the EFOV. Using 2D-US, for example, panoramic views can be created along the vertical direction of the US probe, and the image information based on given characteristic structures can be assembled into a unified image [7]. The images obtained can be immediately seen on the US monitor, and reconstruction is not necessary [7,15]. Another approach used in 3D-US is the enlargement of the detectable anatomical context through the assembly (stitching) of two or more 3D datasets. This strategy has been investigated in different fields (gynecology, urology, gastrointestinal, cardiology, and orthopedics) [7,12,13,15-18], but not yet in thyroid studies. Additionally, the above studies did not focus on volumetric analysis but rather on distance measurements or comparisons between 3D-US EFOV and CT data. In contrast, the present study focused on the feasibility of stitching 3D-US datasets specifically for thyroid volumetric analysis using ad hoc-developed phantoms. Different phantoms are available for US investigations [12,14,16,17,19] and another group also recently developed a flexible balloon phantom of the thyroid that allows for the simulation of nodular shape alterations and isthmus thickening [10]. In the present study, this original model was further adapted to increase the volume range from 200 mL to 400 mL; additionally, ventral echogenic landmarks were placed to facilitate the subsequent stitching procedure, and a support rod was added with the intent of maintaining the anthropomorphic shape of the thyroid phantom and retaining a trachea dummy in place.

Orientation of 3D-US datasets before stitching can be performed automatically or semi-automatically [12,13,16,17,19-21]; however, in this study, we opted for a manual approach based on the phantom landmarks, with a time span of not more than 2 min. The results show

<table>
<thead>
<tr>
<th>Probe type</th>
<th>Method</th>
<th>Phantom shape</th>
<th>Rho*</th>
<th>Mean ± SD</th>
<th>Median</th>
<th>Range</th>
<th>95% CI</th>
<th>Mean ± 1.96*SD (LOA)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Convex</strong></td>
<td>em</td>
<td>regular</td>
<td>0.9939</td>
<td>-1.1 ± 5.2</td>
<td>-4.0</td>
<td>-8.2 - 5.7</td>
<td>-4.0, 1.8</td>
<td>-11.4, 9.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>nodular</td>
<td>0.9934</td>
<td>-5.9 ± 6.9</td>
<td>-8.1</td>
<td>-12.5 - 8.4</td>
<td>-9.8, -2.1</td>
<td>-19.4, 7.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>thick isthmus</td>
<td>0.9944</td>
<td>-9.6 ± 5.6</td>
<td>-10.2</td>
<td>-16.5 - 7.5</td>
<td>-12.6, -6.5</td>
<td>-20.5, 1.3</td>
</tr>
<tr>
<td>mt</td>
<td>regular</td>
<td></td>
<td>0.9993</td>
<td>-2.8 ± 2.4</td>
<td>-2.4</td>
<td>-7.7 - 1.3</td>
<td>-4.1, -1.4</td>
<td>-7.5, 1.9</td>
</tr>
<tr>
<td></td>
<td>nodular</td>
<td></td>
<td>0.9987</td>
<td>-1.9 ± 3.0</td>
<td>-1.1</td>
<td>-7.0 - 4.3</td>
<td>-3.6, 0.2</td>
<td>-7.8, 4.0</td>
</tr>
<tr>
<td></td>
<td>thick isthmus</td>
<td></td>
<td>0.9984</td>
<td>-1.3 ± 3.0</td>
<td>-1.3</td>
<td>-6.8 - 5.0</td>
<td>-2.9, 0.4</td>
<td>-7.1, 4.6</td>
</tr>
<tr>
<td><strong>Linear</strong></td>
<td>em</td>
<td>regular</td>
<td>0.9933</td>
<td>2.8 ± 5.5</td>
<td>4.9</td>
<td>-6.9 - 7.8</td>
<td>-0.3, 5.8</td>
<td>-8.0, 13.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>nodular</td>
<td>0.9811</td>
<td>1.3 ± 8.8</td>
<td>7.5</td>
<td>-10.0 - 12.3</td>
<td>-3.6, 6.2</td>
<td>-16.0, 18.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>thick isthmus</td>
<td>0.9774</td>
<td>-2.6 ± 9.7</td>
<td>-7.7</td>
<td>-11.8 - 13.3</td>
<td>-7.9, 2.8</td>
<td>-21.5, 16.4</td>
</tr>
<tr>
<td>mt</td>
<td>regular</td>
<td></td>
<td>0.9982</td>
<td>0.5 ± 3.1</td>
<td>0.6</td>
<td>-4.7 - 5.0</td>
<td>-1.2, 2.2</td>
<td>-5.5, 6.5</td>
</tr>
<tr>
<td></td>
<td>nodular</td>
<td></td>
<td>0.9991</td>
<td>0.3 ± 2.8</td>
<td>0.5</td>
<td>-7.0 - 3.8</td>
<td>-1.2, 1.8</td>
<td>-5.1, 5.7</td>
</tr>
<tr>
<td></td>
<td>thick isthmus</td>
<td></td>
<td>0.9987</td>
<td>1.4 ± 3.1</td>
<td>2.3</td>
<td>-6.6 - 4.8</td>
<td>-0.3, 3.1</td>
<td>-4.7, 7.4</td>
</tr>
</tbody>
</table>

Rho*: correlation coefficient according to Bravais-Pearson analyses; 95% CI: 95% confidence interval of the mean; LOA: Limit of agreement; em: ellipsoid mode; mt: manual tracing
that the manual adjustment and the stitching of separately obtained 3D-US datasets acquired via sensor-navigated probe was possible and in fact accurate.

Compared to a previous phantom study performed with volumes up to 200 mL and without the use of stitching [10], the present data showed even higher levels of agreement with the reference volumes – irrespective of whether the em or the mt method was used – indicating that 3D-US EFOV may be a promising option for thyroid volumetric analysis.

The present data are consistent with previous studies showing that mt volumetric analysis is associated with far smaller deviations from the expected values than the em method [4,9,10,22,23]. Systematic underestimations and overestimations are also smaller with mt. Larger deviations of the ellipsoid-based volumes occur particularly in the models of nodular shape alterations and isthmus thickening. These imprecisions can be prevented by the use of mt in 3D-US EFOV [4], given that mt allows for a much better shape-adapted contouring of the target. Moreover, the present study also confirms that these advantages are achieved at the cost of substantially greater time and effort [9,24,25].

Compared to the L-probe, in the 3D-US images, the C-probe showed contours with more artifacts, leading to broader images of the phantom walls (results not shown) due to the lower frequency and geometry of the transducer. In this case, the exact placement of the landmarks for the ellipsoid method and the delineation of the contours are less clear. Consistent with a previous study [10], the ellipsoid landmarks and the contours for mt were placed in the middle of the sonographically depicted wall structure, rather than in the inner or outer border. Consequently, all plots obtained using the C-probe showed slightly lower values than the data obtained with the L-probe. In addition, the use of the C-probe was characterized by significant underestimations of the reference volumes in three groups, while this did not occur when the L-probe was used. Although it may be expected that the artifact-prone wall representation leads to larger limits of agreement, no such differences were found.

Additional analyses of intra- and inter-observer variability were not attempted in this study, in accordance with similar studies by other groups [9,26] and because the results obtained with the mt method were unambiguous.

The present study has the following limitations: 1) the time effort of mt was substantially greater than that of em volumetric analysis, therefore, its use in the clinical setting is not routinely practicable and technical developments are needed to optimize this aspect; 2) consistent with previous studies, the 3D-US datasets were interpolated to 3-mm slice thickness in the transverse dimension to limit the time effort of the mt method [10]. While the image quality was slightly worsened due to the partial volume effect, a significant negative impact on the results was not expected because the wall structures were easily identifiable; 3) the clinical relevance of phantom-generated results is yet to be determined, and clinical studies are warranted for this purpose. A potential problem in human studies is the definition of appropriate anatomical landmarks that are easily identifiable in separate scans.

Conclusions

The generation of a 3D-US EFOV via the stitching of two separate 3D-US datasets obtained via sensor-navigated probes proved feasible and accurate, leading to a correct volumetric analysis of even large thyroid phantoms (up to 400 mL). The quality of the data obtained via mt was superior (but the method was more time consuming) compared to the em approach. The present thyroid phantom study provides useful information for further investigations on the potential use of 3D-US in the clinical setting of thyroid disease.

Acknowledgements: The authors gratefully acknowledge Hiltrud Niggemann for statistical analyses and Dr. Ernesta Palombo-Kinne for revision of the manuscript. GE Healthcare, Milwaukee, WI, USA is gratefully acknowledged for providing the hardware (LOGIQ E9) for this study.

Conflict of interest: none

References