Svenja Ipsen*, Ralf Bruder, Verónica García-Vázquez, Achim Schweikard and Floris Ernst

Assessment of 4D Ultrasound Systems for Image-guided Radiation Therapy – Image Quality, Framerates and CT Artifacts

Abstract: 4D ultrasound (4D US) is gaining relevance as a tracking method in radiation therapy (RT) with modern matrix array probes offering new possibilities for real-time target detection. However, for clinical implementation of US-guided RT, image quality, volumetric framerate and artifacts caused by the probe's presence during planning and / or setup computed tomography (CT) must be quantified. We compared three diagnostic 4D US systems with matrix array probes using a commercial wire phantom to measure spatial resolution as well as a calibration and a torso phantom to assess different image quality metrics. CT artifacts were quantified in the torso phantom by calculating the total variation and percentage of affected voxels between a reference CT scan and CT scans with probes in place. We found that state-of-the-art 4D US systems with small probes can fit inside the CT bore and cause fewer metal artifacts than larger probes. US image quality varies between systems and is task-dependent. Volume sizes and framerates are much higher than the commercial guidance solution for US-guided RT, warranting further investigation regarding clinical performance for image guidance.

Keywords: 4D US, image guidance, motion compensation, radiation therapy, system comparison.

<https://doi.org/10.1515/cdbme-2019-0062>

1 Introduction

In recent years, real-time volumetric ultrasound (4D US) imaging has seen major improvements in volume rates and image quality. It is being investigated not only for diagnostic but also interventional tasks as an image guidance method, for example for target localization in radiation therapy (RT).

The only commercial 4D US system currently available for motion compensation in RT is the *Clarity Autoscan*

system (Elekta AB, Stockholm, Sweden) using a mechanical 3D probe with volumetric framerates of 1-2 Hz for prostate localization [1,2]. Respiratory motion tracking is not feasible with this system due to the low temporal resolution [3].

Novel 4D US systems with 2D matrix array probes can achieve much higher framerates at wider fields of view (FOVs) than mechanical probes. However, online access to the image data is mandatory for real-time guidance. The commercial diagnostic 4D US systems *iU22*, *Epiq 7* (Philips Healthcare, Best, Netherlands), *Vivid 7* and *E9* (GE Healthcare, Chicago, USA) can be accessed with either custom or manufacturer-provided real-time streaming interfaces and are currently under investigation for motion compensation in RT [4–6].

Additionally, integration into the RT workflow requires registration between the current target position measured with US during treatment and the CT scan used for treatment planning (CT_{plan}). To avoid an error-prone inter-modality registration, a pre-treatment planning US (US_{plan}) is recommended [7]. *Clarity* acquires US_{plan} and CT_{plan} in direct succession, but for abdominal targets such as liver or pancreas, organ motion and deformation caused by probe pressure can lead to substantial mismatches and thus higher errors [7]. For optimal anatomical correspondence, US_{plan} and CT_{plan} should be acquired simultaneously, but large US probe dimensions, short probe cables and metal artifacts pose challenges for simultaneous acquisition inside the CT bore.

The aim of this study is the assessment of current diagnostic 4D US systems with online data access regarding their suitability for motion compensation in RT, including image quality, volumetric framerates and CT artifacts.

2 Methods

2.1 US systems and phantoms

Three 4D US systems were compared: (1) *Vivid 7 Dimension* with 3V-D probe (GE Healthcare, Chicago, USA), (2) *Vivid E95* with 4Vc-D / 4V-D (GE) and (3) *Epiq 7* with X6-1 (Philips Healthcare, Best, Netherlands). A wire phantom *P_{Wire}*

*Corresponding author: **Svenja Ipsen:** Institute for Robotics and Cognitive Systems, Universität zu Lübeck, Ratzeburger Allee 160, Lübeck, Germany, e-mail: ipsen@rob.uni-luebeck.de

Ralf Bruder, Verónica García-Vázquez, Achim Schweikard, Floris Ernst: Institute for Robotics and Cognitive Systems, Universität zu Lübeck, Lübeck, Germany

Figure 1: (a) Phantoms and CT setup utilized in this study (top – wire and calibration phantoms, bottom – torso phantom in experimental setup with probe fixation for CT artifact assessment). (b) US scans and properties measured in the phantoms.

(Model 83, Dansk Fantom Service, Frederikssund, Denmark) with a wire diameter of 0.2 mm was used to quantify spatial resolution. Image quality was assessed in a commercial calibration phantom P_{Calib} with a central spherical target (CIRS Inc., Norfolk, USA) and an anthropomorphic torso phantom P_{Torso} (FAST Trauma Training Model, CAE Inc. Blue Phantom, Montreal, Canada) customized with tumor masses (kidney, liver). The phantoms are shown in Figure 1.

2.2 US properties

Maximum volume sizes, volumetric acquisition framerates (not including streaming) and image quality metrics were assessed at similar acquisition settings for two different FOV sizes at 200 mm depth: $FOV_1 = 50^\circ$ (azimuth) $\times$ 30° (elevation) and $FOV_2 = 90^\circ \times 90^\circ$. At the time of this study, the 4Vc-D probe was not yet available for real-time volume acquisition and is therefore not included in this section.

As an indicator for spatial resolution, the mean full width at half maximum ($FWHM$) was measured in P_{Wire} acquired from two orientations (Figure 1b) over the 1D intensity profiles along the three main directions (axial, azimuth, elevation) [8]. Image quality was assessed by calculating the speckle signal-to-noise ratio ($SSNR$)

$$SSNR = \frac{\mu_{ROI}}{\sigma_{ROI}} \quad (1)$$

for homogeneous regions of interest (ROIs) in P_{Calib} and in P_{Torso} (kidney, liver), with μ_{ROI} and σ_{ROI} referring to the mean and standard deviation of the intensities inside the ROI. The contrast-to-noise ratio (CNR)

$$CNR = \frac{|\mu_{TG} - \mu_{BG}|}{\sqrt{\sigma_{TG}^2 + \sigma_{BG}^2}} \quad (2)$$

was calculated between a target (TG) and homogeneous background (BG) signal in the spherical target (P_{Calib}) and in an artificial kidney tumor (P_{Torso}) [8].

2.3 CT artifacts

Probe-induced CT artifacts were quantified in P_{Torso} with the probe fixed in a medial position perpendicular to the phantom surface (Figure 1a) using a *Somatom Definition AS* CT scanner (Siemens Healthcare GmbH, Erlangen, Germany) and a standard RT protocol with a voxel size of $0.9 \times 0.9 \times 1 \text{ mm}^3$ and no artifact reduction.

Two reference scans were obtained to determine the baseline for intensity variations σ_{ref} between static scans acquired of the same object in an identical pose. The artifact magnitude was calculated as the mean absolute difference (MAD) between the mean reference and each probe scan. Artifacts were assessed over the entire volume and in different ROIs (Figure 1b) by calculating the percentage of affected voxels with $MAD > 2 * \sigma_{ref}$ and the total variation (TV). TV is a measure to quantify local gradient variations, with high values indicating large variations as often seen for metal streaking artifacts [9].

3 Results and Discussion

The largest of the investigated probes was 3V-D with a length of 13.4 cm (tip to cable connection), weighing 275 g without cable. 4Vc-D was the lightest probe (90 g) with the smallest cable diameter (6 mm). X6-1 was the shortest probe (10.0 cm) but also had the largest footprint ($6.4 \times 4.0 \text{ cm}^2$). 4V-D provided the longest cable (245 cm), with other cables between 197 cm (X6-1) and 212 cm (4Vc-D).

3.1 US properties

The volume sizes, voxel spacings and volumetric framerates are summarized in Table 1. The Vivid 7 (3V-D probe), being the oldest of the three systems, had the smallest maximum achievable FOV size (FOV_1) with a volumetric framerate of 13.8 Hz. For a similarly sized volume, the E95 (4V-D) was more than twice as fast as the Epiq 7 (X6-1, 20.9 vs. 8.0 Hz). It should be noted that the reported framerates only cover image acquisition and reconstruction on the US systems, thus not including latencies from streaming or further processing.

As shown in Figure 2, $FWHM$ in azimuth and elevation directions was lowest for the E95 (3.6 / 3.8 mm), while the Epiq 7 had the lowest $FWHM$ in axial direction (1.9 mm). Overall, the $FWHM$ in FOV_1 was 3.1 mm, 2.8 mm and 3.0 mm for Vivid 7, E95 and Epiq 7, respectively. Furthermore, the newer systems allowed acquisition of the large maximum FOV_2 at 3 Hz (E95) and 2 Hz (Epiq 7). In this case, the mean $FWHM$ increased to 3.1 mm (E95) and 3.4 mm (Epiq 7).

In all scenarios, the Epiq 7 showed superior $SSNR$, most pronounced in the liver with values more than three times higher than Vivid 7 and E95. Upon visual inspection, the Epiq 7 volumes appeared smoother in homogeneous regions while the GE systems produced grainier volumes (Table 1). The Vivid 7 had the lowest $SSNR$ in all investigated structures as can be seen in Figure 2.

The CNR was found to be target-dependent with the highest value for the E95 in the kidney tumor (1.2 vs. 0.5) in P_{Torso} but a lower value in the spherical target (P_{Calib}) compared to the Epiq 7 (6.1 vs. 8.5). This suggests that tumor visibility might differ between different US systems and tumor types and there might not be one ideal system for all target localization tasks. While Vivid 7 and E95 are cardiac systems focusing on high framerates, the Epiq 7 (X6-1 probe) is aimed at abdominal organs with lower requirements on

temporal resolution and more focus on visualization quality.

3.2 CT artifacts

Axial slices intersecting the probes showed the strongest artifacts. Global TV increase was between 1.9% (4Vc-D) and 4.6% (X6-1), most pronounced in the ROI directly below the probe (Figure 1b) ranging from 166% (4Vc-D) to 390% (X6-1). At 15 cm depth, this was reduced to between 10% (4Vc-D) and 41% (3V-D), down to 0.1% (4Vc-D) and 3.6% (3V-D) at a slice position 6.5 cm away from the probe.

As shown in Figure 2, the percentage of affected voxels was highest for the largest, heaviest probe (19.5%, 3V-D) and lowest for the lightest 4Vc-D (10.9%). The magnitude of the deviations from the reference scan is up to 86% less for the 4Vc-D relative to the 3V-D in the region directly below the probe affected by the strongest artifacts.

CT reconstruction with metal artefact reduction will be tested in the future to further enhance image quality. The results also show that metal artifacts are mainly affecting those CT slices containing the probe. By placing the probe away from the target region or tilted at an angle relative to the surface, contouring of anatomical structures might be unaffected and treatment planning could be possible despite the presence of the probe during CT acquisition.

Furthermore, there is an ongoing trend toward probe miniaturization, with the newest probe 4Vc-D being much smaller and weighing about one third of the older 4V-D model. Probe dimensions, composition and cable lengths will play an important role for positioning flexibility in the planning and treatment phases since smaller, with lighter probes containing less metal causing fewer artifacts. These properties are also relevant for robotic probe manipulation where probe weight and cable length could interfere and even limit the automatic positioning process.

Table 1: Voxel / volume sizes and volumetric framerates of three 4D US systems with matrix array probes.

	FOV_1 (50° x 30°, 200 mm depth)			FOV_2 (90° x 90°, 200 mm depth)	
	Vivid 7 (3V-D)	E95 (4V-D)	Epiq 7 (X6-1)	E95 (4V-D)	Epiq 7 (X6-1)
Example slice
Voxel size (mm)	1.4 x 1.0 x 1.4	1.4 x 1.0 x 1.4	0.9 x 0.8 x 1.3	1.5 x 1.0 x 1.5	0.8 x 0.8 x 1.3
Volume size (mm)	154 x 200 x 111	154 x 201 x 104	164 x 203 x 107	283 x 200 x 283	295 x 207 x 297
Framerate (Hz)	13.8	20.9	8.0	3.0	2.0

Figure 2: Quantification of US image properties and CT artifacts. *FWHM* was measured in a wire phantom (left), *SSNR* in homogeneous phantom regions (center) and the percentage of affected voxels and mean absolute differences in CT scans (right).

4 Conclusion

All investigated 4D US systems achieved higher volume framerates for comparable volume sizes than the currently available commercial system. At the same time, newer systems (E95, Epiq 7) can cover larger FOVs which could facilitate the generation of deformation fields over entire organs. While the E95 showed the highest spatiotemporal resolution in this study, the Epiq 7 provided a higher *SSNR*. Considering RT workflow integration, smaller probes (e.g. 4Vc-D) can fit inside the CT bore and cause up to 86% less artifacts than older, bigger probes.

In the future, the systems will be compared regarding their real-time tracking capabilities and *in-vivo* performance. Combined with improved metal artefact reduction during reconstruction, simultaneous acquisition of US and CT could potentially increase registration accuracy in US-guided RT.

Author Statement

Research funding: Parts of this work were supported by the German Federal Ministry of Education and Research (grant no. 13GW0228B), the German Research Foundation (grant no. ER 817/1-1), the Ministry of Economic Affairs, Employment, Transport and Technology of Schleswig-Holstein. Conflict of interest: The 4Vc-D probe was provided by GE healthcare. Ethical approval: The conducted research is not related to either human or animals use.

References

[1] Lachaine M, Falco T. Intrafractional prostate motion

management with the Clarity Autoscan system. *Med Phys Int.* 2013;1(1):72–80.

- [2] Grimwood A, McNair HA, O'Shea TP, Gilroy S, Thomas K, Bamber JC, Tree AC, Harris EJ. In vivo validation of Elekta's Clarity Autoscan for ultrasound-based intrafraction motion estimation of the prostate during radiation therapy. *Int J Radiat Oncol.* 2018;102(4):912–21.
- [3] Harris EJ, Miller NR, Bamber JC, Symonds-Taylor JRN, Evans PM. The effect of object speed and direction on the performance of 3D speckle tracking using a 3D swept-volume ultrasound probe. *Phys Med Biol.* 2011;56(22):7127–43.
- [4] Schlosser J, Gong RH, Bruder R, Schweikard A, Jang S, Henrie J, Kamaya A, Koong A, Chang DT, Hristov D. Robotic intrafractional US guidance for liver SABR: System design, beam avoidance, and clinical imaging. *Med Phys.* 2016;43(11):5951–63.
- [5] Øye OK, Wein W, Ulvang DM, Matre K, Viola I. Real time image-based tracking of 4D ultrasound data. *Med Image Comput Comput Assist Interv.* 2012;15(Pt 1):447–54.
- [6] Bruder R, Ernst F, Schläfer A, Schweikard A. A framework for real-time target tracking in IGRT using three-dimensional ultrasound. In: *Proceedings of the 25th International Congress and Exhibition on Computer Assisted Radiology and Surgery (CARS).* Berlin, Germany: IJCARS; 2011. p. 306–7.
- [7] van der Meer S, Seravalli E, Fontanarosa D, Bloemen-van Gorp EJ., Verhaegen F. Consequences of intermodality registration errors for intramodality 3D ultrasound IGRT. *Technol Cancer Res Treat.* 2015;15(4):632–8.
- [8] Thijssen JM, Van Wijk MC, Cuypers MHM. Performance testing of medical echo/Doppler equipment. *Eur J Ultrasound.* 2002;15(3):151–64.
- [9] Wu M, Keil A, Constantin D, Star-Lack J, Zhu L, Fahrig R. Metal artifact correction for x-ray computed tomography using kV and selective MV imaging. *Med Phys.* 2014;41(12):121910.