Supplementary Materials:

We have conducted a reliability study with MRIs of the left and right parotid glands with 15 participants. To evaluate the inter-observer reliability, two observers (one radiologist and one radiation oncologist) were asked to independently contour the parotid glands on the 15 pre-RT and 15 3-month post-RT MRIs. Both observers/physicians had over 10 years of experiences and one observer was blinded to the contours of the other observer. There are usually 15 to 20 transverse MR slices (images) of the parotid gland. No additional information such as patient or treatment characteristics was provided to the observers. The observers contoured the parotid glands slice by slice following standard clinical protocols.

The variations of the volume difference were calculated for assessment of consistency among measurements by the two observers. The small volume difference in left and right parotid volume contoured by the physicians is demonstrated in Fig. 1. Between the two observers, the left and right mean absolute volume percentage difference of the parotid glands was 6.86%±1.82% and 7.01% ±2.08% for the pre-RT MRI and 7.15%±2.65% and 8.20% ±1.91% for the 3-month post-RT MRI.

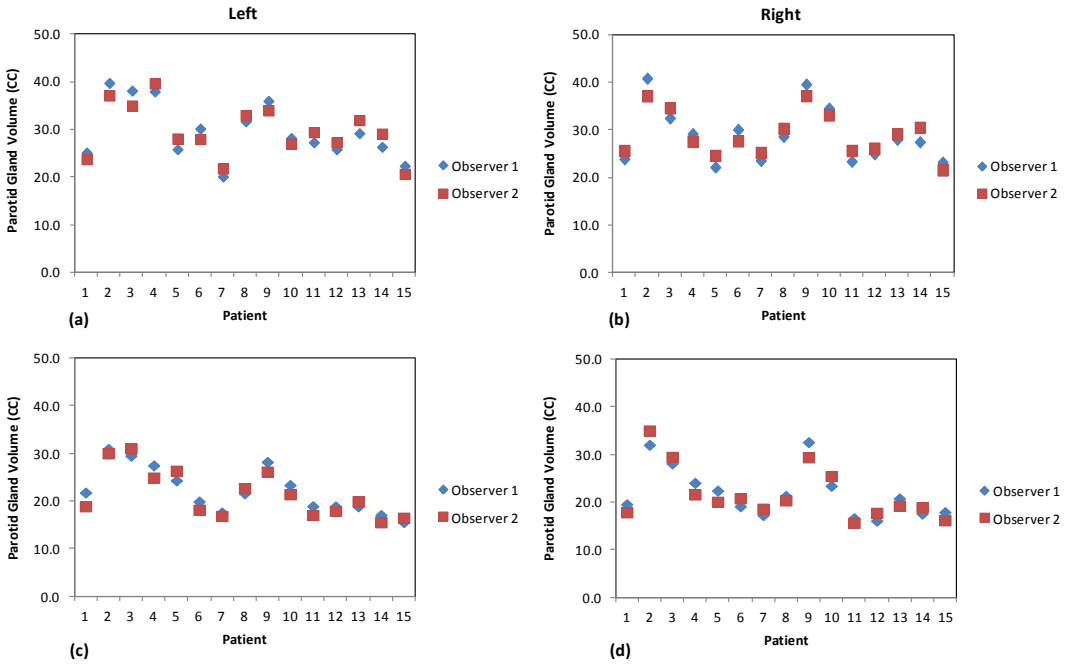

Figure 1. Inter-observer reliability is demonstrated by the agreement between two observers' parotid contours on the pre-RT (the 1st row) and the 3-month post-RT (the 2nd row) MRIs: (a) and (c) left parotid, and (b) and (d) right parotid gland.

Even though the physicians' manual contours of the parotid glands are consistent, we further evaluate the effects of manual contours on the resulted automatic segmentations. Figure 2 demonstrates the automatic-segmented parotid-gland volumes at 3-month post-RT follow-up based on the two observers' manual contours. The left and right mean absolute volume percentage difference of the parotid glands using two different sets of manual contours for the 3-month post-RT MRI was 8.01%±1.67% and 8.58%±1.87%.

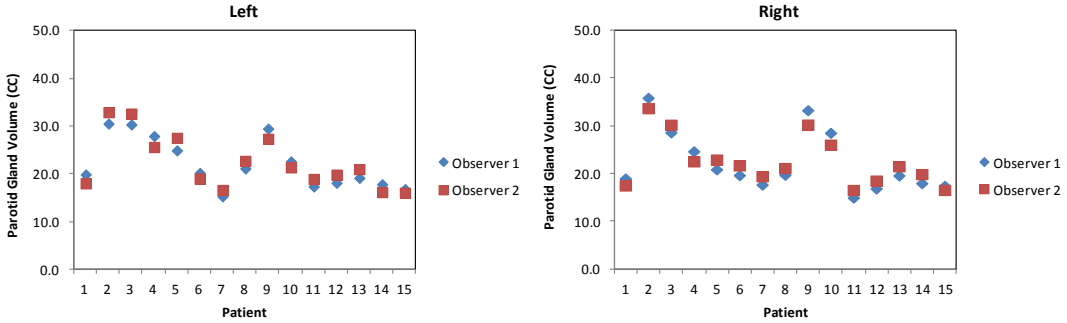

Figure 2. Automated parotid segmentation results based on observers 1 and 2: left parotid (a) and right parotid (b).

In summary, the inter-observer reliability study shows consistency in physicians' manual baseline contours, which are used as the ground truth, as well as in automatic parotid segmentations using different set of baseline contours.