Automated Breast Ultrasound for Ductal Pattern Reconstruction: Ground Truth File Generation and CADe Evaluation

D Manousaki¹, A Panagiotopoulou¹, V Bizimi², M S Haynes³, S Love⁴, and M Kallergi¹
¹Department of Biomedical Engineering, TEIA, GR
²2nd Radiology Department, University Hospital Attikon, Athens, GR
³Jet Propulsion Laboratory (JPL), Pasadena, CA, USA
⁴Dr. Susan Love Research Foundation, Encino, CA, USA

Corresponding Author: Maria Kallergi, kallergi@teiath.gr

Abstract. The purpose of this study was the generation of ground truth files (GTFs) of the breast ducts from 3D images of the Invenia™ Automated Breast Ultrasound System (ABUS) system (GE Healthcare, Little Chalfont, UK) and the application of these GTFs for the optimization of the imaging protocol and the evaluation of a computer aided detection (CADe) algorithm developed for automated duct detection. Six lactating, nursing volunteers were scanned with the ABUS before and right after breastfeeding their infants. An expert in breast ultrasound generated rough outlines of the milk-filled ducts in the transaxial slices of all image volumes and the final GTFs were created by using thresholding and smoothing tools in ImageJ. In addition, a CADe algorithm automatically segmented duct like areas and its results were compared to the expert’s GTFs by estimating true positive fraction (TPF) or % overlap. The CADe output differed significantly from the expert’s but both detected a smaller than expected volume of the ducts due to insufficient contrast (ducts were partially filled with milk), discontinuities, and artifacts. GTFs were used to modify the imaging protocol and improve the CADe method. In conclusion, electronic GTFs provide a valuable tool in the optimization of a tomographic imaging system, the imaging protocol, and the CADe algorithms. Their generation, however, is an extremely time consuming, strenuous process, particularly for multi-slice examinations, and alternatives based on phantoms or simulations are highly desirable.

1. Introduction
The architecture of the female breast is particularly complex making its study difficult. A complete understanding, however, of breast anatomy and the development of study tools are top priorities in the effort to achieve early diagnosis and treatment of breast diseases [1]. About 80% of all breast cancers
begin in the lactiferous ducts (ductal carcinoma) [2]. For that reason, it is necessary to map and understand the behavior of the ducts and how they influence the birth and progression of cancer cells in these areas. In 1840, Astley Cooper created a 3D structure of the ductal system by injecting the breasts of cadavers of women, who had been lactating at the time of death, colored wax, via the nipple, prior to dissection [3]. This allowed the ducts to be clearly distinguished and different lobes to be identified using different colors. After this achievement, researchers tried to find a “bloodless”, in vivo, method to detect the whole network and create a ductal pattern. The most promising approach to date is the use of automated, 3D ultrasound systems. In 2010, Goeding et al reconstructed a small part of the ductal system using a patent 3D ultrasound system [4]. Today, systems like the Invenia™ Automated Breast Ultrasound System (ABUS) system (GE Healthcare, Little Chalfont, UK) offer capabilities that could lead to the desired reconstruction of high-quality 3D image of the breast and its structure. The purpose of this work was to conduct a pilot study with lactating volunteers and determine (a) the capabilities of the new systems for detecting the breast ducts, (b) the potential to reconstruct the entire ductal pattern, and (c) the requirements for the design and development of CADe that could assist in the process. The purpose was achieved by developing GTFs of the ductal pattern from the original images with the aid of an expert in breast ultrasound. GTFs are usually binary versions of the original images where clinical features (e.g., tumors, breast ducts, etc.) are identified and outlined by one or more experts providing thus information on the size and shape of the objects of interest. These files are used for the validation of medical systems and CADe algorithms when a gold standard or absolute ground truth is not available [5].

2. Materials and Methods

Lactating, nursing volunteers were recruited for the study because the milk inside the lactiferous ducts offers a natural contrast medium and was expected to facilitate the process. Six such women were scanned with the Invenia ABUS at Breastlink Orange (Orange, CA, USA) and Dr. Susan Love Research Foundation (Encino, CA, USA). Mean age of the volunteers was 33.5 years and mean lactation time was 10.5 months. Three of the women were primipara and three were multipara. The women, all white, reported no breast disease and no major lactation problems.

Both breasts of five of the six volunteers were imaged before and right after breastfeeding their infants; one volunteer was imaged only before breastfeeding. Three scans were performed for each breast: an anterior-posterior (AP), a lateral (LAT), and a medial (MED) [6]; three volumes of images were generated, each including on average 250 slices. An expert in breast ultrasound generated rough outlines of the milk-filled ducts in the AP slices using ImageJ [https://imagej.nih.gov/ij/]. Final GTFs of the breast ducts were generated by two trained assistants using smoothing and filling tools in ImageJ. Final outlines were again reviewed and approved by the expert. A typical final outline for one slice is shown in Fig. 1.

The volumes of the breasts and the volumes of the detected ducts were estimated from the images and the GTFs and were compared to the literature. In addition, statistical differences were determined between various pairs of measurements using the student t-test and level of significance at 0.5. In parallel, a basic CADe algorithm was developed by scientists at the Jet Propulsion Laboratory (JPL) (Pasadena, CA, USA) that generated outlines of the duct areas based on 3D Non Local-Means filtering and adaptive thresholding criteria [7,8]. The GTFs were used to evaluate the algorithm’s performance and guide future design and development. Evaluation was done by estimating the true positive fraction (TPF) or % overlap between GTFs and CADe outlines in MATLAB. A representative overlay of the two results for the same slice is shown in Fig. 2.
3. Results

Eleven sets of GTFs were generated; six before and five after lactation. The process was completed in a 12-month period and required 143 person-hours to review 5,851 images and generate outlines. The whole process was time-consuming and strenuous to the expert due to several factors beyond the work load. Difficulties in the interpretation of the images by the expert and the detection of the ducts included: (a) artifacts due to the compression applied to the breast by the transducer (too little compression generated fuzzy areas, too much compression caused the ducts to collapse), (b) the ducts were partially filled with milk and, hence, there was not sufficient contrast through the entire duct system and (c) the limited training of the expert on the specific ABUS. In addition, the expert reviewed only the AP scans on a research workstation with research software (ImageJ) and not on the clinical workstation of ABUS. It is possible that the addition of the other two image volumes, i.e., the LAT and MED scans, would have provided supplementary information particularly in the areas toward the edges of the breast that presented significant shadowing.

The analysis of the measurements showed that the detected duct volumes were small compared to the entire breast area (≤1%) confirming that only a small part of the ductal system was detected with this protocol. The average right breast volume was 573.7 ml (SD=217.1 ml) and average left volume was 605.3 ml (SD=241.7 ml). Based on the pre-lactation GTFs, the left breast showed significantly more milk-filled ducts than the right breast in 5 of the 6 volunteers (p=0.046). Such differences were not

Figure 1. A typical output of expert’s outline of the ducts on the AP slices of the ABUS.

Figure 2. MATLAB algorithm output comparing expert’s GTFs and CADe detections. Dark: True Positive (overlap) Grey: False Positive (CADe) Light Grey: False Negative (expert)
observed post-lactation. An analysis of the upper and lower halves of each breast showed that more milk-filled ducts were detected in the lower half than the upper half in 5 out of 6 volunteers. The observed asymmetries in the ductal distribution are in agreement with prior reports [4] but the causes remain unknown. Table 1 lists the p-values estimated by testing the significance of the differences between left and right breast, pre- and post-lactation imaging, and between upper and lower halves of the breast. Statistical significance was set at the 0.05 level.

Table 1. Results of t-test’s p-value for various pairs of volume measurements.

| Pair                                           | p-value |
|------------------------------------------------|---------|
| Left breast volume > Right breast volume       | 0.080   |
| Expert’s left ductal volume > Expert’s right ductal volume (pre) | 0.046   |
| Expert’s left ductal volume > Expert’s right ductal volume (post) | 0.450   |
| Expert’s pre ductal volume > Expert’s post ductal volume (right breast) | 0.030   |
| Expert’s pre ductal volume > Expert’s post ductal volume (left breast) | 0.020   |
| Expert’s lower half ductal volume > Expert’s upper half ductal volume (pre, right) | 0.100   |
| Expert’s lower half ductal volume > Expert’s upper half ductal volume (pre, left) | 0.220   |
| Expert’s lower half ductal volume > Expert’s upper half ductal volume (post, right) | 0.310   |
| Expert’s lower half ductal volume > Expert’s upper half ductal volume (post, left) | 0.410   |

Finally, CADe results were significantly different from the GTFs. Average TPF value was 18% and ranged from 0% to 39% for either the left or right breast.

4. Conclusions
The new ABUS systems offer significant advantages over the standard ultrasound systems in breast imaging. In this work, the systems were tested for a purpose other than the detection of abnormalities, that of detecting the ducts of the breast. The results showed that it may be possible to generate the breast ductal pattern via in-vivo imaging with an ABUS but attention needs to be given to (a) the imaging protocol, (b) the compression level, (c) the positioning of the transducer, (d) the training of the experts, (e) the development of CADe, and (f) the evaluation process. The first three factors have a major impact on image quality, the presence of artifacts, and the preservation of duct contrast. The training of the experts impacts the false positive and false negative rate. Based on this experience, a simultaneous training on an ABUS and a handheld ultrasound system is proposed. The role of CADe methods for this application is critical. Their development and evaluation, however, cannot be based solely on the generation of experts’ GTFs because this is a time-consuming, laborious, and subjective process. New breast phantoms with duct like structures are necessary, which, along with simulation techniques, could reduce the burden on the experts and provide more objective metrics of assessment.

5. References
[1] Mills D, Gordon E J, Casano A, et al. 2011 J. Physiol. Biochem. 67 621
[2] Ferris-James D M, Iuanow E, Mehta T S, et al. 2012 Radiographics 32 1009
[3] Cooper, Sir Astley Paston 1840 Plate VI: Ducts and glandules” (2005). On the anatomy of the breast, by Sir Astley Paston Cooper, Paper 16. http://jdc.jefferson.edu/cooper/16
[4] Gooding M J, Finlay J, Shipley J A, et al 2010 J. Ultrasound Med. 29 95
[5] Kallergi M, Hersh M, and Manohar A 2005 Handbook of Biomedical Image Analysis: Volume 2: Segmentation Models Part B, ed J S Suri, D L Wilson, and S Laxminarayan (Kluwer Academic, Plenum Publishers, New York) p. 183
[6] Brem R F, Tabár L, Duffy S W, et al. 2015 Radiology 274 663
[7] Coupé P, Hellier P, Kervrann Ch, and Barillot Ch 2009 IEEE Trans. Image Process. 18 2221
[8] Dougherty G 2010 Biomed. Imaging Interv. J. 6 e32