When and Why Might a Computer-aided Detection (CAD) System Interfere with Visual Search? An Eye-tracking Study

Trafton Drew, MS, PhD, Corbin Cunningham, BS, Jeremy M. Wolfe, MS, PhD

Rational and Objectives: Computer-aided detection (CAD) systems are intended to improve performance. This study investigates how CAD might actually interfere with a visual search task. This is a laboratory study with implications for clinical use of CAD.

Methods: Forty-seven naive observers in two studies were asked to search for a target, embedded in $1/f^{2.4}$ noise while we monitored their eye movements. For some observers, a CAD system marked 75% of targets and 10% of distractors, whereas other observers completed the study without CAD. In experiment 1, the CAD system's primary function was to tell observers where the target might be. In experiment 2, CAD provided information about target identity.

Results: In experiment 1, there was a significant enhancement of observer sensitivity in the presence of CAD ($t(22) = 4.74$, $P < .001$), but there was also a substantial cost. Targets that were not marked by the CAD system were missed more frequently than equivalent targets in no-CAD blocks of the experiment ($t(22) = 7.02$, $P < .001$). Experiment 2 showed no behavioral benefit from CAD, but also no significant cost on sensitivity to unmarked targets ($t(22) = 0.6$, P = NS). Finally, in both experiments, CAD produced reliable changes in eye movements: CAD observers examined a lower total percentage of the search area than the no-CAD observers (experiment 1: $t(48) = 3.05$, $P < .005$; experiment 2: $t(50) = 7.31$, $P < .001$).

Conclusions: CAD signals do not combine with observers' unaided performance in a straightforward manner. CAD can engender a sense of certainty that can lead to incomplete search and elevated chances of missing unmarked stimuli.

Key Words: Observer performance; computer-aided detection; image perception.

Computer-aided detection (CAD) algorithms are designed to assist radiologists during medical image interpretation. For instance, in mammography, a typical CAD system marks potential abnormalities on the image to encourage additional evaluation by the radiologist before the radiologist makes a final recommendation. In the United States, CAD is currently used on nearly 75% of all mammograms (1). Several large studies have assessed the efficacy of CAD (2,3). Although most studies show that hit rate increases when CAD is introduced to a practice, false alarm rate also tends to increase, making it unclear whether the benefits of CAD outweigh the costs (4,5). From a signal detection perspective, the relatively small benefit of CAD is surprising because the CAD system should be increasing the total amount of information available to the radiologists, yielding increased performance. The size of the hypothetical benefit would be larger if CAD and radiologists were making use of independent signals and smaller if they are using the same noisy signals. Even if CAD and radiologists are not independent, the hypothetical benefit seems to be larger than what is observed (3). That the use of CAD produces only modest improvement in signal detection measures such as area under the receiver operating characteristic (ROC) curve suggests that radiologists are unable to optimally combine the information conveyed by the CAD system and information they gather from the image itself.

In the current study, we use eye-tracking to study the costs and benefits of the presence of a simultaneous CAD system. The laboratory task we created was designed to emulate critical aspects of a typical radiologic search for a difficult to find target. In both experiments, half of the observers completed the experiment without a CAD system, whereas the other half searched the same trials with the help of our artificial CAD system that marked 75% of all targets

Acad Radiol 2012; 19:1260–1267

From the Visual Attention Lab, Department of Surgery, Brigham & Women's Hospital, 64 Sidney St. Suite. 170, Cambridge, MA 0213-4170 (T.D., J.W.); Harvard Medical School, Boston, MA (T.D., C.C., J.M.W.). Received April 9, 2012; accepted May 8, 2012. **Address correspondence to:** T.D. e-mail: tdrew1@rics.bwh.harvard.edu

http://dx.doi.org/10.1016/j.acra.2012.05.013

and 10% of nontargets. In experiment 1, targets were difficult to find because they were embedded in a field of noise. Here, the CAD system primarily aided target detection (CADe). In experiment 2, we manipulated the appearance of our target "Ts" and distractor "Ls," making the Ts and Ls more similar to each other. At the same time, we decreased the opacity of the background noise the items so that the items were easier to find. Our intent was to keep the overall difficulty roughly the same across the two experiments. In this case, the CAD system primarily aided target diagnosis (CADx).

MATERIALS AND METHODS

Observers were instructed to search for a target letter, T, among distractor, Ls. All of the stimuli were embedded in a 16.5° square texture of cloudlike $1/f^{2.4}$ noise (Fig 1). This noise roughly simulates the spatial frequency of radiologic images. Mammograms, for example, can be roughly characterized as $1/f^3$ stimuli (6). The similarity to real medical images is not critical in this case. The noise was merely designed to make the search task more demanding. The stimuli consisted of Ts and Ls of a random orientation that were made up of two perpendicular lines slightly offset from each other. These stimuli allowed us to manipulate the difficulty of differentiating targets and distractors by changing the offset of bars comprising these items. Ts and Ls subtended 1.35° visual angle. CAD marks were pink circles with a diameter of 1.5°. Target and distractor locations were chosen at random from a 4 × 4 grid of possible locations. Position within this grid was randomly jittered (up to 0.25°) to avoid predictable locations (Fig 1).

Observers were instructed to click on the T when detected and to click on an "absent" button if no target was found. Half of the trials contained a single target. A confidence rating was collected at the conclusion of each trial using a 6-point scale, with 6 denoting highest confidence in target presence and 1, lowest. On CAD blocks, observers were instructed to use the CAD to help them find the target; however, they were told that CAD would sometimes miss the target or mark a distractor. In this artificial situation, we could set the performance of our simulated CAD to any level. In this case, our CAD marked the target 75% of the time and marked 10% of the distractor Ls; equivalent to a d-prime value of 1.95. Each trial contained an average of 5 Ls (range 0–15), meaning that the CAD made an average of 0.5 false-positive marks per image. CAD marks appeared simultaneously with stimulus onset. This differs from the Food and Drug Administration–approved protocol of showing CAD marking after an initial CAD-free reading.

Both experiments employed a between-subjects design in which half of the observers were assigned to a CAD condition and the other half to a no-CAD condition. Observers in both conditions began with a 50 trial practice block that did not contain CAD markings. This was followed by a block of 100 experimental trials. All observers saw the same 150

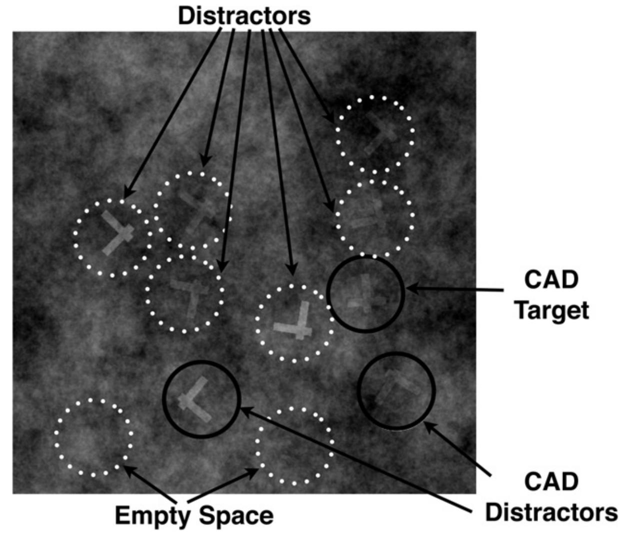

Figure 1. Representative example of the search stimulus. *Dotted circles* represent predefined interest areas that were not visible during the experiment.

"cases" though the order of cases was different for each observer. In the CAD condition, the 100 experimental trials had CAD marks added. We then compared performance across observers in the CAD/no-CAD block. This design allowed us to equate the amount of experience the observers had with our task when they undertook the critical CAD/no-CAD block of trials.

Differences between Experiments 1 and 2

Experiments 1 and 2 differed in the opacity of the $1/f^{2.4}$ noise and the similarity between targets and distractors. Higher noise opacity makes the items harder to detect. Increased similarity makes targets harder to discriminate from distractors. The effects of these manipulations are not independent because noise also makes the items harder to discriminate. However, separately manipulating these two factors allows us to produce two tasks with similar performance for different reasons. Experiment 1 had high noise and low similarity between targets and distractors, whereas experiment 2 had lower noise and higher similarity between targets and distractors. Thus, the targets in experiment 1 were difficult to detect but easy to "diagnose." Here CAD would aid detection (CADe). The targets in experiment 2 were easy to detect and hard to identify. In this case, CAD would aid diagnosis (CADx).

Observers

Twenty-three observers were tested in experiment 1 and 24 in experiment 2. Observers ranged in age from 18 to 54 (average = 24.3, standard deviation = 5.7, 11 male). All had at least 20/25 acuity (with correction as needed) and could pass the Ishihara Color-Blindness test. All gave informed consent and were paid $10/hour for their time.