Journal of Medical Imaging

MedicalImaging.SPIEDigitalLibrary.org

Rapid perceptual processing in two- and three-dimensional prostate images

Melissa Treviño
Baris Turkbey
Bradford J. Wood
Peter A. Pinto
Marcin Czarniecki
Peter L. Choyke
Todd S. Horowitz

SPIE.

Melissa Treviño, Baris Turkbey, Bradford J. Wood, Peter A. Pinto, Marcin Czarniecki, Peter L. Choyke, Todd S. Horowitz, "Rapid perceptual processing in two- and three-dimensional prostate images," *J. Med. Imag.* **7**(2), 022406 (2020), doi: 10.1117/1.JMI.7.2.022406.

Rapid perceptual processing in two- and three-dimensional prostate images

Melissa Treviño,[a,*] Baris Turkbey,[b] Bradford J. Wood,[c] Peter A. Pinto,[d] Marcin Czarniecki,[e] Peter L. Choyke,[b] and Todd S. Horowitz[a]

[a]National Cancer Institute, Basic Biobehavioral and Psychological Sciences Branch, Rockville, Maryland, United States

[b]National Cancer Institute, Molecular Imaging Program, Bethesda, Maryland, United States

[c]National Cancer Institute, Center for Interventional Oncology, Bethesda, Maryland, United States

[d]National Cancer Institute, Urologic Oncology Branch, Bethesda, Maryland, United States

[e]Georgetown University School of Medicine, Washington, DC, United States

Abstract. Radiologists can identify whether a radiograph is abnormal or normal at above chance levels in breast and lung images presented for half a second or less. This early perceptual processing has only been demonstrated in static two-dimensional images (e.g., mammograms). Can radiologists rapidly extract the "gestalt" from more complex imaging modalities? For example, prostate multiparametric magnetic resonance imaging (mpMRI) displays a series of images as a virtual stack and comprises multiple imaging sequences: anatomical information from the T2-weighted (T2W) sequence, functional information from diffusion-weighted imaging, and apparent diffusion coefficient sequences. We first tested rapid perceptual processing in static T2W images then among the two functional sequences. Finally, we examined whether this rapid radiological perception could be observed using T2W multislice imaging. Readers with experience in prostate mpMRI could detect and localize lesions in all sequences after viewing a 500-ms static image. Experienced prostate readers could also detect and localize lesions when viewing multislice image stacks presented as brief movies, with image slices presented at either 48, 96, or 144 ms. The ability to quickly extract the perceptual gestalt may be a general property of expert perception, even in complex imaging modalities. [DOI: 10.1117/1.JMI.7.2.022406]

Keywords: rapid perceptual processing; gestalt; two-dimensional; three-dimensional; prostate; multiparametric MRI.

Paper 19233SSR received Sep. 13, 2019; accepted for publication Dec. 5, 2019; published online Jan. 3, 2020.

1 Introduction

With a single glance, radiologists often have a sense or a hunch that "something" is present in a medical image. This phenomenology corresponds to the notion of "perceptual gist." (The term "gist" is widely used in the visual perception literature to refer to the extraction of global meaning from an image following a brief exposure. However, in radiology, "GIST" refers to "gastrointestinal stromal tumors." In order to avoid confusion, we will use "rapid perceptual processing" or "gestalt.") in cognitive psychology, which refers to the ability to rapidly identify or understand the meaning of a real-world scene given only a very brief exposure.[1,2] In the medical image perception literature, rapid perceptual processing, sometimes termed "gestalt,"[3] has been empirically demonstrated with readers who can distinguish between abnormal and normal images (e.g., mammograms, chest radiographs, and micrographs of cervical cells) better than chance level after presentations of half a second or less.[4–8] To date, this rapid perceptual processing has only been demonstrated in static two-dimensional (2-D) images. Standard practice in radiology is moving to three-dimensional (3-D) sequences and complex modalities that combine

*Address all correspondence to Melissa Treviño, E-mail: melissa.trevino@nih.gov

2329-4302/2020/$28.00

information from multiple imaging techniques obtained in multiple anatomic planes and sections. (We use the term 3-D to refer to multislice images acquired using stacked (or interleaved) 2-D or 3-D acquisition techniques, as information of the organ structure is exhibited in images obtained from either technique. We refrain from using "volumetric" to refer to stacked 2-D or 3-D acquisition images as the term is solely for 3-D acquisition images.) Here, we investigated whether or not radiologists could extract perceptual gestalt from prostate multiparametric magnetic resonance imaging (mpMRI) exams, which exemplify both of these features.

Rapid perceptual processing has been previously demonstrated among different medical professionals using images from chest radiography, mammography, and cytology at image exposure durations ranging from 120 to 2000 ms.[5–8] In Evans et al.'s 2013 investigation,[6] radiologists and cytologists were presented with bilateral mammograms or micrographs of cervical tissue, respectively, for 250 to 2000 ms. Half of the images contained an abnormality. Readers were then asked to localize the abnormality followed by determining whether an abnormity was present on a rating scale. Both radiologists and cytologists were able to detect if the image was abnormal at above chance levels at each presentation duration; for radiologists as well there was no significant detection advantage for longer exposure durations (250 versus 1000 ms). Both groups of readers were also unable to localize the abnormalities. Evans et al. inferred from the above chance performance for detection, but not for localization, that readers were able to rapidly extract the gestalt of an image that was based on a global impression of images but not precise object localization. However, Carrigan et al.[4] argued that this inability to localize may be due to reader inexperience, conservative criteria for scoring localization, and/or the use of abnormalities that did not have a definite visible mass. More convincing evidence that radiologists can use a global signal to make abnormality judgements comes from experiments in which no specific lesion is present in the "abnormal" tissue. For example, Evans et al.[7] showed that radiologists could distinguish between a normal breast image and a mammogram from a breast contralateral to the lesion. In another experiment, they showed radiologists patches of the parenchyma rather than full mammograms. The readers could reliably distinguish between patches from a diseased breast that excluded the lesion (and patches from the breast contralateral to the lesion) and patches from a normal breast. In both cases, the abnormality signal was stronger for images containing a lesion; nevertheless, it is clear that radiologists can use some nonlocalized signal to distinguish between images of normal and abnormal tissue.

These studies validate professional readers' intuitive sense that they can recognize something about an abnormal scan in a single glance. They also allow us to start to infer what kind of visual information these readers use to guide their initial exploration of a case, and what kind of information might be particularly useful for discriminating between diseased and normal tissue. Rapid gestalt perception in real-world scenes has been shown to support identification and localization of objects, deployment of attention to facilitate visual search in a cluttered scene, and guide future eye movements.[9,10] Thus while exposing radiologists to brief glimpses of images is not by any stretch of the imagination a realistic representation of clinical practice, leveraging radiologists' ability to detect an abnormality in a brief glance could lead to improved computer-aided technologies used to assist radiologists as well as develop training protocols that can help promote a strong gestalt signal in order to ultimately improve cancer detection.

The gestalt of 2-D real-world scenes is typically carried by low spatial frequency channels, which convey the structural layout of scenes; subsequently more refined processing then occurs for local or high-frequency information (i.e., boundary edges) used to identify objects.[2] However, this later detailed processing is not required for scene categorization,[11] and this time course of scene processing from coarse to fine is not absolute, can occur in parallel, and is influenced by task demands.[12,13] In contrast, Evans et al.[7] showed that the gestalt signal in mammography is carried by high spatial frequency channels. Evans et al. also tested several potential features, including bilateral breast symmetry and breast density, none of which proved to be important for gestalt processing. Similar studies have yet to be carried out in other medical imaging domains.

Most of what we know about gestalt processing in medical imaging comes from conventional 2-D radiography, including standard mammograms. In contrast, we know nothing about perceptual gestalt in 3-D imaging modalities, such as breast tomosynthesis or mpMRI, used in prostate cancer imaging. In 3-D imaging, a single case consists of a series of image slices through the

body that are assembled into a virtual stack. Radiologists can acquire a 3-D representation of organ structures by scrolling through the stacks. Evaluating prostate mpMRI requires radiologists to scroll through several imaging sequences (or modalities), each highlighting a specific parameter of the prostate. Anatomical or morphological information is derived from (1) T2-weighted (T2W) MRI, while various types of diffusion information come from (2) diffusion-weighted imaging (DWI), displaying movement of water molecules within the tissue and indicative of cancer when diffusion is restricted,[14] (3) apparent diffusion coefficient (ADC) measuring diffusion magnitude with lower values associated with aggressive cancer,[15] and (4) the dynamic contrast-enhanced MRI reflecting blood vessel permeability indicative of cancer.[14] Each sequence contributes to the radiologist's evaluation of whether a clinically significant cancer (Gleason score ≥ 7) is present. Standard workstations present these sequences side-by-side. Radiologists can scroll through each sequence simultaneously or singly.

We hypothesize that rapid perceptual processing is a general property of expert medical image perception, though the relevant channels will probably depend on both imaging sequence and the anatomy and physiology of the medical domain. The aim of this paper is to demonstrate rapid perceptual processing for the first time in a 3-D imaging sequence, specifically prostate mpMRI. However, a negative result with 3-D stimuli would be difficult to interpret, since it might be that prostate mpMRI simply does not support rapid perceptual processing, even in a simple 2-D image. We, therefore, begin with 2-D anatomical (T2W) images, the closest analog to chest radiographs or mammograms, before moving on to functional images, and, finally, 3-D image sequences.

2 Methods

2.1 *Participants*

A power analysis was conducted using R 3.5.3[16] and the "pwr" package.[17] Based on previous work with a reported large effect size ($d = 5.17$),[6] the power analysis indicated a sample size of 2.5 would have 0.80 power to detect whether readers could distinguish between abnormal and normal images higher than chance with an alpha at 0.05. Power was based on number of readers and not images. Table 1 lists demographic and radiologic practice information for each experiment. In experiment 1, we tested 27 readers with prostate mpMRI experience [7 females, average age = 43.44 years, standard deviation (SD) = 9.12] and 5 readers with no experience reading prostate images (3 females, average age = 35.75 years, SD = 12.18 years). Eighteen readers experienced in prostate mpMRI were recruited and assessed at the Radiologic Society of Northern America (RSNA) conference and nine experienced readers completed the task online.

In experiment 2, we tested 16 readers with prostate mpMRI experience (5 females, average age = 40.00 years, SD = 10.15). Readers were divided into three groups who viewed either T2W ($n = 6$), DWI ($n = 5$), or ADC images ($n = 5$). In experiment 3, a total of 14 readers with prostate mpMRI experience (4 females, average age = 43.00 years, SD = 10.82) were tested. Readers were divided into three groups who viewed cases presented at either 48 ms/slice (20.8 Hz, $n = 5$), 96 ms/slice (10.4 Hz, $n = 5$), or 144 ms/slice (6.9 Hz, $n = 4$). Readers were recruited from the RSNA conference in Chicago, Illinois, between November 26, 2017 to November 30, 2017 (experiment 1) and November 25, 2018 to November 29, 2018 (experiments 2 and 3). All readers had normal or corrected-to-normal vision, gave informed consent, and were entered in a raffle to win a gift card. All images were anonymized. The study was deemed exempt from institutional review by NIH's Office of Human Subjects Research Protections.

2.2 *Apparatus and Stimuli*

All experiments used deidentified images selected from prostate mpMRI cases. Lesion-present images contained a single cancerous lesion that had been previously diagnosed and biopsied to determine the Gleason score. The Gleason grading system is used to rate prostate tissue from normal to cancerous on a 1 to 10 scale, with 6 being the lowest score for a cancer. The sector maps of the apex, mid, and base regions of the prostate were acquired from the prostate imaging

Table 1 Demographic and radiologic practice characteristics for each experiment.

	Experiment 1		Experiment 2				Experiment 3			
	Experienced	No experience	All	ADC	DWI	T2W	All	48	96	144
Characteristic	($n = 27$)	($n = 5$)	($n = 16$)	($n = 5$)	($n = 5$)	($n = 6$)	($n = 14$)	($n = 5$)	($n = 5$)	($n = 4$)
Gender, *n* (%)										
Female	7 (25.93)	3 (60.00)	5 (31.25)	1 (20.00)	0 (0.00)	4 (66.7)	4 (28.57)	1 (20.00)	3 (60.00)	1 (25.00)
Training level, *n* (%)										
Resident	3 (11.11)	3 (60.00)	4 (25.00)	0 (0.00)	1 (20.00)	3 (50.00)	1 (7.14)	0 (0.00)	0 (0.00)	1 (25.00)
Radiologist	24 (88.89)	2 (40.00)	12 (75.00)	5 (100.00)	4 (80.00)	3 (50.00)	13 (92.86)	5 (100.00)	5 (100.00)	3 (75.00)
Years practicing *M* (SD)	14.42 (8.04)	15.00 (7.07)	13.50 (8.05)	13.40 (10.53)	11.50 (5.74)	16.33 (8.08)	16.35 (11.68)	18.50 (11.86)	16.40 (14.64)	12.67 (8.74)
Age at evaluation										
Years *M* (SD)	43.44 (9.12)	35.75 (12.18)	40.00 (10.15)	43.20 (10.11)	39.40 (8.91)	37.83 (12.16)	43.00 (10.82)	46.40 (12.12)	42.80 (11.39)	39.00 (9.90)
Range	28 to 65	29 to 54	23 to 61	37 to 61	29 to 51	23 to 58	29 to 61	32 to 60	32 to 61	29 to 48
Cases read in the last year (across all subspecialties)										
M	9214	5442	7325	6180	11,880	4483	7435	8900	5260	8325
Range	120 to 50,000	50 to 18,000	300 to 18,200	2000 to 12,000	5200 to 18,200	300 to 7800	1300 to 26,000	2000 to 26,000	1500 to 10,400	1300 to 20,800
Subspecialties, *n* (%)										
Abdominal	27 (100.0)	3 (60.00)	15 (93.75)	5 (100.00)	4 (80.00)	6 (100.00)	11 (78.57)	4 (80.00)	3 (60.00)	4 (100.00)
Breast	2 (7.41)	0 (0.00)	3 (18.75)	1 (20.00)	1 (20.00)	1 (16.67)	2 (14.29)	0 (0.00)	1 (20.00)	1 (25.00)

Table 1 (*Continued*).

	Experiment 1		Experiment 2				Experiment 3			
	Experienced	No experience	All	ADC	DWI	T2W	All	48	96	144
Characteristic	($n = 27$)	($n = 5$)	($n = 16$)	($n = 5$)	($n = 5$)	($n = 6$)	($n = 14$)	($n = 5$)	($n = 5$)	($n = 4$)
Chest	12 (44.44)	1 (20.00)	5 (31.25)	1 (20.00)	2 (40.00)	2 (33.33)	5 (35.71)	3 (60.00)	0 (0.00)	2 (50.00)
General radiology	16 (59.26)	5 (100.00)	7 (43.75)	2 (40.00)	2 (40.00)	3 (50.00)	6 (42.86)	3 (60.00)	1 (20.00)	2 (50.00)
Musculoskeletal	6 (22.22)	2 (40.00)	3 (18.75)	0 (0.00)	1 (20.00)	2 (33.33)	3 (21.43)	1 (20.00)	1 (20.00)	1 (25.00)
Neuro	7 (25.93)	3 (60.00)	3 (18.75)	0 (0.00)	1 (20.00)	2 (33.33)	1 (7.14)	1 (20.00)	0 (0.00)	0 (0.00)
Nuclear	4 (14.81)	1 (20.00)	4 (25.00)	0 (0.00)	2 (40.00)	2 (33.33)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
Pediatric	3 (11.11)	1 (20.00)	3 (18.75)	0 (0.00)	0 (0.00)	3 (50.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
Reading prostate mpMRI										
Years *M* (SD)	8.15 (6.84)	0.00 (0.00)	4.84 (4.63)	4.34 (3.85)	5.50 (4.61)	4.70 (5.89)	9.18 (6.65)	8.30 (7.19)	9.60 (6.62)	9.75 (7.89)
Cases per week										
M (SD)	19.34 (16.99)	0.00 (0.00)	19.27 (20.81)	20.60 (22.50)	10.20 (6.91)	27.00 (27.97)	21.57 (17.45)	27.40 (21.88)	23.40 (17.54)	12.00 (9.27)
Range	0 to 80	—	4 to 75	4 to 60	5 to 20	5 to 75	3 to 60	10 to 60	5 to 50	3 to 25
% of clinical time										
M (SD)	22.33 (23.77)	0.00 (0.00)	22.66 (28.79)	25.60 (39.07)	11.90 (12.52)	29.16 (31.05)	28.14 (28.17)	20.80 (11.65)	42.00 (42.81)	20.00 (17.80)
Range	0 to 95	—	2 to 95	3 to 95	2 to 30	5 to 90	5 to 100	9 to 40	5 to 100	5 to 40

reporting and data system (PI-RADS) by the American College of Radiology under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International license. Sector maps are used in standard practice to localize lesions in the PI-RADS version 2, an assessment and reporting tool for prostate mpMRI. Maps are divided into 12 sectors and sectors are divided into zones: the anterior fibromuscular stroma (AS), the peripheral zone (PZ), the transition zone (TZ), and the central zone (CZ) (see Fig. 1).

PsychoPy software version 3.0[18] was used for in-person (i.e., at RSNA) readers of experiment 1, and jsPsych[19] for online readers of experiment 1 and all for readers of experiments 2 and 3. For in-person readers, stimuli were presented on a 24-in. LG Flatron W2442PA, 1920×1080 pixel, liquid–crystal display operating at a frame rate of 60 Hz and controlled by a Dell Latitude E7240 laptop. Viewing distance was ~57 cm. The average size of lesions was 2.12 deg (SD: 1.09; median: 1.74; range: 0.79 to 6.30 deg). Online readers were asked to perform the task on a desktop or laptop if possible to change screen resolution to 1366×768, to sit a comfortable distance from the monitor, perform the task on either Google Chrome, Safari, or Firefox web browser, maximize their browser window, and close any additional browser windows or tabs. Images in experiment 1 had a native resolution of 512×512 pixels, were resized to 1080×1080, and subtended 29.22 deg $\times$ 29.22 deg on screen. In experiments 2 and 3, images were 600×570 pixels, presented at their native resolution, and subtended 15.89 deg $\times$ 15.07 deg on screen.

In all experiments, lesions were present in half of the cases. Gleason score, zone, and region distributions were all selected to approximate the typical clinical context. Stimuli in experiment 1

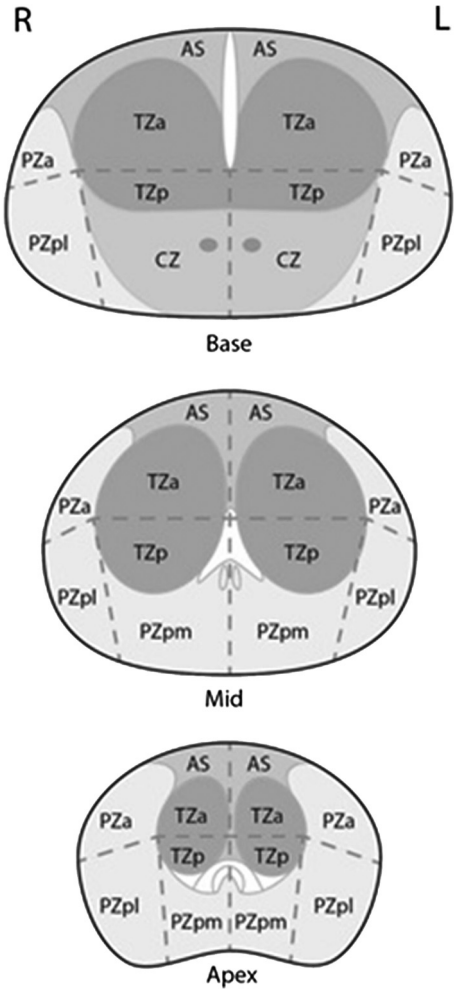
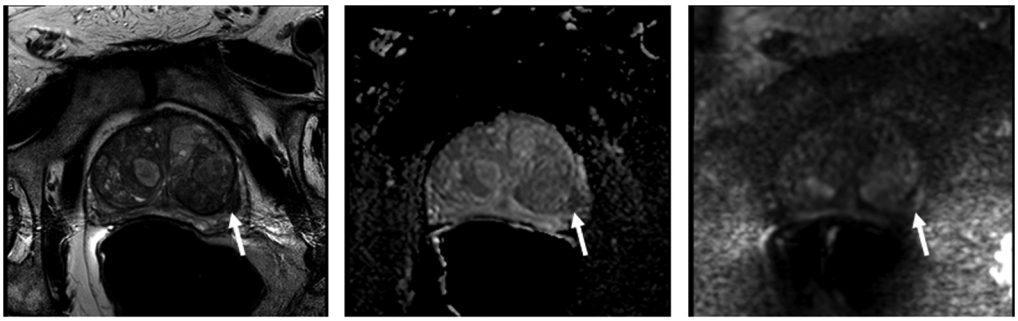


Fig. 1 Prostate sector maps of the apex, mid, and base regions. Maps divide the prostate into 12 sectors and 4 zones: the AS, the PZ, the TZ, and the CZ.



(a) (b) (c)

Fig. 2 An example of a Gleason 9 lesion (arrow) is demonstrated in the left peripheral zone posterolateral in (a) a T2W image, (b) an ADC image, and (c) a DWI image.

were 100 T2W images. In experiment 2, stimuli were 100 T2W, DWI, and ADC images (see Fig. 2). The same cases were used across all three image sequences and the T2W cases were the same as those used in experiment 1. In experiments 1 and 2, lesion-present images had Gleason scores of 6 ($n = 10$), 7 ($n = 15$), 8 ($n = 15$), and 9 ($n = 10$). 70% of lesions were located in PZ and 30% in TZ, reflecting the prevalence in the clinic.[20] Lesions were distributed across the apex ($n = 20$), mid ($n = 15$), and base ($n = 15$) regions of the prostate. In experiment 3, stimuli were 56 cases each comprising a stack of 26 T2W images. Lesions had Gleason scores of 6 ($n = 8$), 7 ($n = 15$), 8 ($n = 3$), 9 ($n = 1$), and 10 ($n = 1$). 71% of lesions were located in PZ and 29% in TZ. Lesions were located in the apex ($n = 9$), mid ($n = 10$), and base ($n = 9$) regions of the prostate.

2.3 *Experimental Design and Procedure*

Experiment 1 was based on the paradigm employed by Evans et al.[6] There were two groups of readers (experience with prostate mpMRI, $n = 27$; no experience, $n = 5$). All readers completed 100 trials. The trial sequence is depicted in Fig. 3(a). Each trial began with a fixation cross presented for 500 ms, followed by a randomly selected static T2W image for 500 ms. The prostate image was then replaced with a sector map, which covered the same area as the prostate image. Readers then were asked to localize the cancerous lesion on the prostate sector map, whether or not they thought there was a lesion, by selecting one or more of the 12 possible sectors. Finally, readers rated how confident they were that a lesion was present on a scale from 0 to 100, with 0 denoting there was no lesion and 100 denoting there was certainly a lesion. An additional set of experienced readers ($n = 9$) participated online and completed a slightly modified paradigm: after localizing, readers were given a two-alternative forced choice question, "Was there a cancerous lesion?" and answered by mouse clicking on either the "yes" or "no" response buttons [see Fig. 3(b)]. [We modified the procedure in response to evidence that confidence ratings reflect perceptual variability rather than signal strength.[21] This work suggests that signal detection analyses (e.g., d') should be computed using the two-alternative forced choice values rather than confidence intervals.]. Readers then rated their confidence in their answer to the previous question: "Was there a cancerous lesion?" on a scale from 0 to 100 with 0 indicating "not confident" and 100 "confident." All readers completed 100 trials.

Experiment 2 used the modified paradigm from experiment 1 [Fig. 3(b)]. There were three groups of readers, each of whom was presented with stimuli from a different mpMRI sequence: T2W, $n = 6$; DWI, $n = 5$; or ADC, $n = 5$. All readers completed 100 trials. The independent group design was chosen partly to reduce reader burden, but also because all three sequences came from the same 100 cases. Therefore, if we had used multiple images from the same case, there might have been contamination between trials. Experiment 3 was identical except that, instead of a single static image, the stimulus was a stack of 26 T2W slices presented in movie form. Presentation duration was varied between three groups. Readers viewed cases presented at either 48 ms/slice (20.8 Hz, $n = 5$), 96 ms/slice (10.4 Hz, $n = 5$), or 144 ms/slice (6.9 Hz, $n = 4$). All readers completed 56 trials. We reduced the number of cases again to minimize

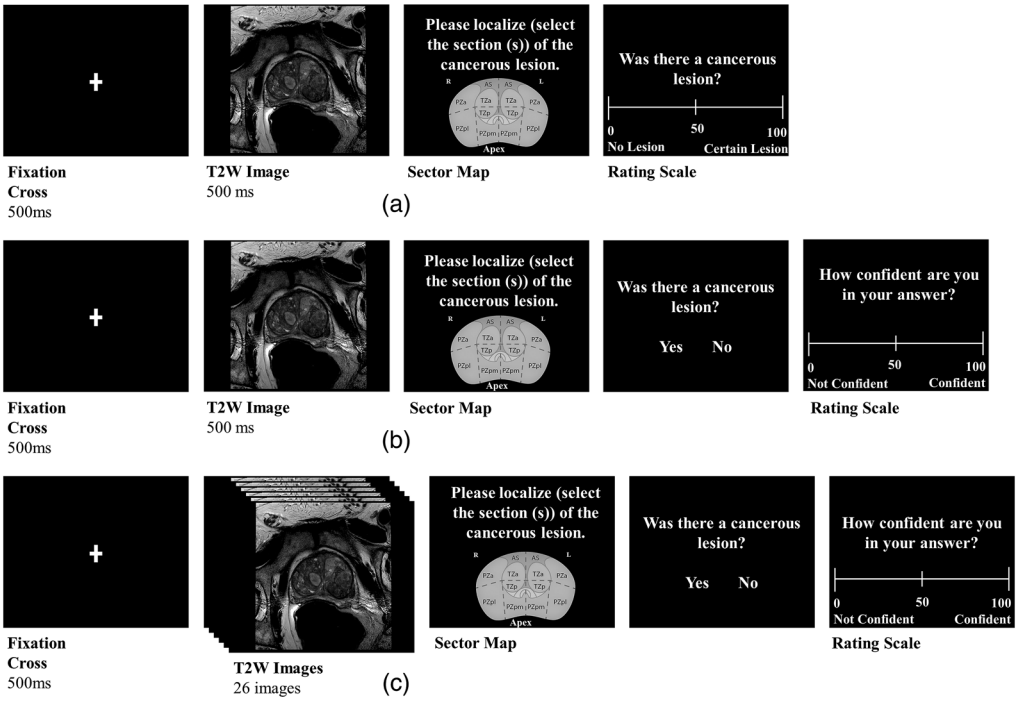


Fig. 3 (a) The temporal sequence of events defining a trial for RSNA readers in experiment 1. At the beginning of a trial, a fixation cross appeared for 500 ms, followed for 500 ms by a T2W image of the prostate. A prostate sector map was then presented, where readers were asked to localize the cancerous lesion. Readers then rated whether a cancerous lesion was present on a rating scale. (b) The trial sequence for online readers in experiment 1 and readers for experiment 2. After a 500-ms fixation cross, online readers in experiment 1 viewed a 500-ms T2W image of the prostate and readers in experiment 2 were presented with either a T2W, DWI, or ADC image of the prostate for 500 ms. Readers then localized the cancerous lesion on a prostate sector map. After the sector map, readers determined whether a lesion was presented with either a "yes" or "no" button response. A rating scale then appeared in which readers rated their confidence in their previous "yes" or "no" answer on a rating scale. (c) The temporal sequence of a trial in experiment 3. Following the 500-ms fixation cross, a case comprising a stack of T2W image slices was presented; image slices were presented at either 48, 96, or 144 ms. After each case, identical to experiment 2, a sector map appeared where readers localized the lesion then indicated whether a cancerous lesion was presented and then gave a confidence rating.

reader burden, as the movies in experiment 3 took longer than the single images in experiments 1 and 2.

2.4 *Statistical Analyses*

Analysis of the z-ROC curves for all experiments revealed that the z-slope was not equal to 1.0 (M [95% confidence interval (CI)]: = 0.78 [0.72, 1.50]), so we report d_a as our measure of discriminability instead of d'.[22] We also computed the area under the receiver operating characteristic (ROC) curve (AUC) for each reader using the IMRMC software version 4.0.3.[23] Brandon D. Gallas (BDG) analysis was used to calculate AUC for small sample sizes ($n \leq 6$).[24] BDG is a multireader ROC analysis that is a variant of the three-sample U-statistic test and assumes the test statistic follows a t-distribution for smaller sample sizes.[24] Localization performance was the reader's probability of correctly selecting the primary sector(s) in which the cancerous lesion was located on trials where readers correctly indicated that a cancerous lesion was present (for the in-person readers in experiment 1, we used those trials where confidence was ≥50). Chance localization performance was determined by averaging across trials the probability of each sector being a primary sector out of the total number of primary sectors. Chance was 17.22% for experiments 1 and 2 and 17.77% for experiment 3.

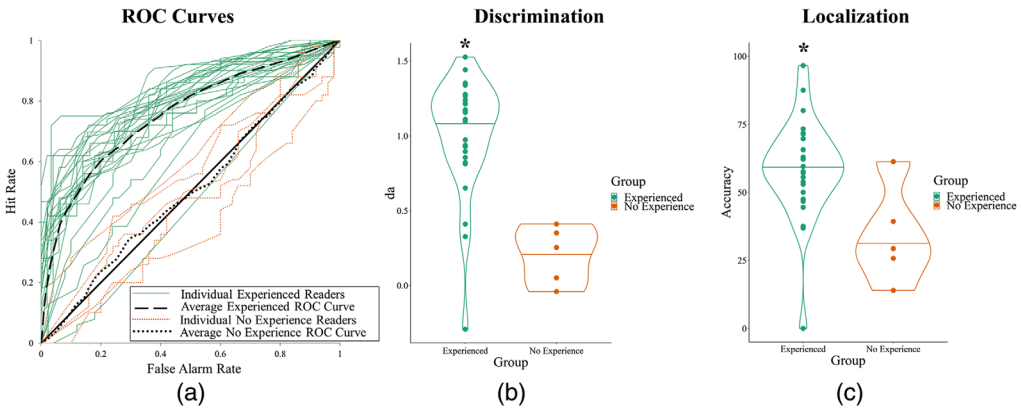


Fig. 4 Results for experiment 1. (a) ROC curves of readers' performances for each group: readers with experience or no experience with prostate mpMRI. (b) Discriminability performance for both groups measured in d_a. (c) Localization performance for both groups.

3 Results

3.1 *Experiment 1*

Results for experiment 1 are shown in Fig. 4 and broken down by various categories in Table 2.

3.1.1 *Detection*

In experiment 1, readers with mpMRI experience were able to discriminate between normal and abnormal images with better than chance ($d_a = 1.01$, $p < 0.0001$; AUC $= 0.76$, $p < 0.0001$), but readers without previous mpMRI experience were at chance ($d_a = 0.02$, $p > 0.884$; AUC $= 0.50$, $p > 0.939$). This finding generalizes the phenomenon of rapid perceptual processing to the mpMRI domain.

3.1.2 *Localization*

Experienced readers were able to localize lesions better than chance (58.72%, $p < 0.0001$), whereas readers without mpMRI experience were not (33.89%, $p < 0.103$).

3.1.3 *PZ versus TZ*

In radiologic practice, lesions are more prevalent and easier to detect in PZ compared to TZ. Our results with experienced readers conformed to this pattern, both for detection [d_a: PZ $= 1.08$, $p < 0.0001$, TZ $= 0.85$, $p < 0.0001$; AUC: PZ $= 0.77$, $p < 0.0001$, TZ $= 0.72$, $p < 0.0001$] and localization; in fact, localization in TZ did not differ from chance (PZ $= 69.79\%$, $p < 0.0001$; TZ $= 29.29\%$, $p > 0.160$). Readers without mpMRI experience were unable to detect or localize in either zone [d_a: PZ $= 0.02$, $p > 0.874$, TZ $= 0.02$, $p > 0.892$; AUC: PZ $= 0.51$, $p > 0.918$, TZ $= 0.50$, $p > 0.993$; localization accuracy: PZ $= 29.99\%$, $p > 0.137$, TZ $= 43.19\%$, $p > 0.062$].

3.1.4 *Effect of experience*

Experience had a moderate effect on rapid perceptual processing. Among readers with mpMRI experience, those with more than 10 years of experience reading prostate images detected best [d_a: $t(25) = -1.85$, $p < 0.039$; AUC: $t(25) = -1.88$, $p < 0.036$] and were the only subgroup to localize lesions in the TZ. There were no appreciable differences between those with 1 to 10 years' experience and those with <1 year of experience [d_a: $t(16) = 0.30$, $p > 0.766$; AUC: $t(16) = 0.26$, $p > 0.800$]. Readers who read one case or less per week performed comparable to

Table 2 Results for experiment 1.

	d_a			AUC			Localization %		
	M (95% CI)	*T*	*p*	*M* (95% CI)	*t*	*p*	*M* (95% CI)	*t*	*p*
Experienced ($n = 27$)	1.01 (0.86, 1.16)	13.62	<0.0001	0.76 (0.70, 0.81)	8.56	<0.0001	58.72 (51.89, 65.55)	11.91	<0.0001
RSNA ($n = 18$)	0.95 (0.77, 1.13)	10.30	<0.0001	0.74 (0.68, 0.81)	7.59	<0.0001	59.42 (50.43, 68.41)	9.20	<0.0001
Online ($n = 9$)	1.13 (0.89, 1.36)	9.27	<0.0001	0.78 (0.69, 0.88)	5.80	<0.0001	57.31 (46.82, 67.80)	7.50	<0.0001
Radiologists ($n = 24$)	1.01 (0.84, 1.17)	12.06	<0.0001	0.75 (0.69, 0.81)	8.15	<0.0001	58.00 (50.39, 65.61)	10.50	<0.0001
Residents ($n = 3$)	1.05 (0.93, 1.17)	17.43	0.003	0.77 (0.71, 0.82)	25.83	<0.0001	64.48 (57.46, 71.50)	13.21	<0.0001
≤1 year reading prostate ($n = 5$)	0.97 (0.73, 1.20)	8.17	0.001	0.74 (0.67, 0.81)	6.51	<0.0001	59.54 (51.64, 67.44)	10.51	<0.0001
>1 and <10 years reading prostate ($n = 13$)	0.90 (0.64, 1.15)	6.82	<0.0001	0.73 (0.65, 0.82)	5.47	<0.0001	54.19 (42.49, 65.89)	6.19	<0.0001
≥10 years reading prostate ($n = 9$)	1.20 (1.05, 1.34)	16.28	<0.0001	0.80 (0.75, 0.84)	13.25	<0.0001	64.81 (54.40, 75.22)	8.96	<0.0001
≤1 prostate case per week ($n = 4$)	0.40 (−0.06, 0.87)	1.70	0.187	0.61 (0.43, 0.79)	1.90	0.153	44.31 (15.23, 73.39)	1.83	0.165
No experience ($n = 5$)	0.02 (−0.24, 0.28)	0.15	0.885	0.50 (0.40, 0.61)	7.04	0.940	33.89 (18.28, 49.50)	2.09	0.104
Radiologists ($n = 2$)	−0.03 (−0.78, 0.72)	−0.08	0.951	0.48 (0.03, 0.94)	2.00	0.892	43.50 (8.63, 78.37)	1.48	0.379
Residents ($n = 3$)	0.05 (−0.14, 0.25)	0.54	0.642	0.52 (0.40, 0.63)	3.57	0.684	27.48 (13.01, 41.95)	1.39	0.299

Table 2 (*Continued*).

	PZ d_a			PZ AUC			PZ localization %		
	M (95% CI)	*t*	*p*	*M* (95% CI)	*t*	*p*	*M* (95% CI)	*t*	*p*
Experienced ($n = 27$)	1.08 (0.93, 1.24)	13.74	<0.0001	0.77 (0.70, 0.84)	7.52	<0.0001	69.79 (62.59, 76.99)	15.03	<0.0001
RSNA ($n = 18$)	1.07 (0.87, 1.27)	10.46	<0.0001	0.77 (0.69, 0.84)	7.21	<0.0001	70.11 (60.55, 79.67)	11.40	<0.0001
Online ($n = 9$)	1.11 (0.86, 1.36)	8.79	<0.0001	0.78 (0.68, 0.88)	5.30	<0.0001	69.14 (58.31, 79.97)	9.89	<0.0001
Radiologists ($n = 24$)	1.07 (0.89, 1.24)	12.17	<0.0001	0.77 (0.70, 0.84)	7.22	<0.0001	68.78 (60.85, 76.71)	13.41	<0.0001
Residents ($n = 3$)	1.22 (1.02, 1.41)	12.12	0.007	0.79 (0.72, 0.87)	18.59	<0.0001	77.82 (65.55, 90.09)	10.11	0.010
≤1 year reading prostate ($n = 5$)	1.13 (0.86, 1.40)	8.18	0.001	0.77 (0.71, 0.83)	17.50	<0.0001	66.69 (58.66, 74.72)	12.73	<0.0001
>1 and <10 years reading prostate ($n = 13$)	0.93 (0.66, 1.21)	6.63	<0.0001	0.74 (0.65, 0.83)	5.09	<0.0001	69.04 (55.83, 82.25)	8.09	<0.0001
≥10 years reading prostate ($n = 9$)	1.28 (1.16, 1.39)	21.81	<0.0001	0.82 (0.75, 0.88)	9.47	<0.0001	72.58 (62.43, 82.73)	11.20	<0.0001
≤1 prostate case per week ($n = 4$)	0.41 (−0.10, 0.92)	1.59	0.211	0.62 (0.40, 0.84)	3.77	0.194	52.46 (17.46, 87.46)	2.12	0.124
No experience ($n = 5$)	0.02 (−0.25, 0.30)	0.17	0.875	0.51 (0.39, 0.62)	6.94	0.919	29.99 (13.61, 46.37)	1.85	0.138
Radiologists ($n = 2$)	−0.02 (−0.74, 0.69)	−0.07	0.958	0.48 (0.03, 0.94)	2.00	0.897	40.26 (8.38, 72.14)	1.59	0.359
Residents ($n = 3$)	0.06 (−0.23, 0.34)	0.38	0.739	0.52 (0.32, 0.72)	2.64	0.720	23.14 (−13.07, 59.35)	0.93	0.452

Table 2 (*Continued*).

	TZ d_a			TZ AUC			TZ localization %		
	M (95% CI)	*t*	*p*	*M* (95% CI)	*t*	*p*	*M* (95% CI)	*t*	*p*
Experienced ($n = 27$)	0.85 (0.67, 1.02)	9.63	<0.0001	0.72 (0.65, 0.79)	6.09	<0.0001	29.29 (18.36, 40.22)	1.44	0.161
RSNA ($n = 18$)	0.71 (0.52, 0.90)	7.29	<0.0001	0.69 (0.61, 0.77)	4.73	<0.0001	30.09 (15.88, 44.30)	1.22	0.239
Online ($n = 9$)	1.13 (0.84, 1.14)	7.82	<0.0001	0.79 (0.68, 0.91)	5.10	<0.0001	27.69 (10.27, 45.11)	0.72	0.489
Radiologists ($n = 24$)	0.86 (0.67, 1.04)	9.13	<0.0001	0.72 (0.65, 0.80)	5.79	<0.0001	29.31 (17.14, 41.48)	1.3	0.207
Residents ($n = 3$)	0.75 (0.18, 1.32)	2.57	0.124	0.70 (0.43, 0.96)	3.02	0.099	29.17 (10.33, 48.01)	0.82	0.497
≤1 year reading prostate ($n = 5$)	0.64 (0.34, 0.95)	4.11	0.015	0.67 (0.57, 0.76)	19.11	0.002	36.57 (21.10, 52.04)	1.94	0.124
>1 and <10 years reading prostate ($n = 13$)	0.78 (0.54, 1.01)	6.41	<0.0001	0.72 (0.63, 0.81)	4.93	<0.0001	17.97 (0.63, 35.31)	−0.37	0.717
≥10 years reading prostate ($n = 9$)	1.06 (0.74, 1.38)	6.42	<0.0001	0.75 (0.67, 0.83)	6.18	<0.0001	41.60 (25.14, 58.06)	2.42	0.042
≤1 prostate case per week ($n = 4$)	0.36 (0.00, 0.71)	1.94	0.147	0.60 (0.56, 0.63)	3.00	0.012	22.77 (−6.36, 51.90)	0.10	0.925
No experience ($n = 5$)	0.02 (−0.26, 0.30)	0.14	0.893	0.50 (0.38, 0.62)	8.71	0.994	43.19 (26.33, 60.05)	2.55	0.063
Radiologists ($n = 2$)	−0.04 (−0.83, 0.75)	−0.10	0.940	0.48 (0.04, 0.93)	2.00	0.880	36.98 (−9.79, 83.75)	1.39	0.397
Residents ($n = 3$)	0.06 (−0.16, 0.28)	0.54	0.645	0.51 (0.39, 0.63)	5.50	0.823	52.50 (16.49, 88.51)	2.88	0.102

readers without any mpMRI experience in both detection [d_a: $t(7) = -1.43$, $p > 0.194$; AUC: $t(7) = 0.05$, $p > 0.959$] and localization [$t(7) = -0.66$, $p > 0.531$].

3.1.5 *Effect of testing modality*

We compared the performance of the readers tested in person at RSNA and those tested online with the modified protocol. No differences were found in d_a [$t(25) = 1.11$, $p > 0.276$], AUC [$t(25) = 0.86$, $p > 0.391$] or localization performance [$t(25) = 0.28$, $p > 0.780$].

3.2 *Experiment 2*

Results for experiment 2 are shown in Fig. 5 and broken down by various categories in Table 4 in Sec. 5.

3.2.1 *Detection*

Experiment 2 used only readers with mpMRI experience. Collapsing across sequences, readers were able to discriminate abnormal images from normal images with better than chance ($d_a = 0.94$, $p < 0.0001$; AUC $= 0.74$, $p < 0.0001$). This replicates experiment 1. Breaking detection performance down by sequence, we found that detection was better than chance for all sequences. The ADC sequence yielded a better performance than the other two sequences, though this was only statistically significant compared to DWI [ADC versus DWI: $t(8) = 2.68$, $p < 0.029$; ADC versus T2W: $t(9) = 1.28$, $p > 0.233$; T2W versus DWI: $t(9) = 0.08$, $p > 0.934$].

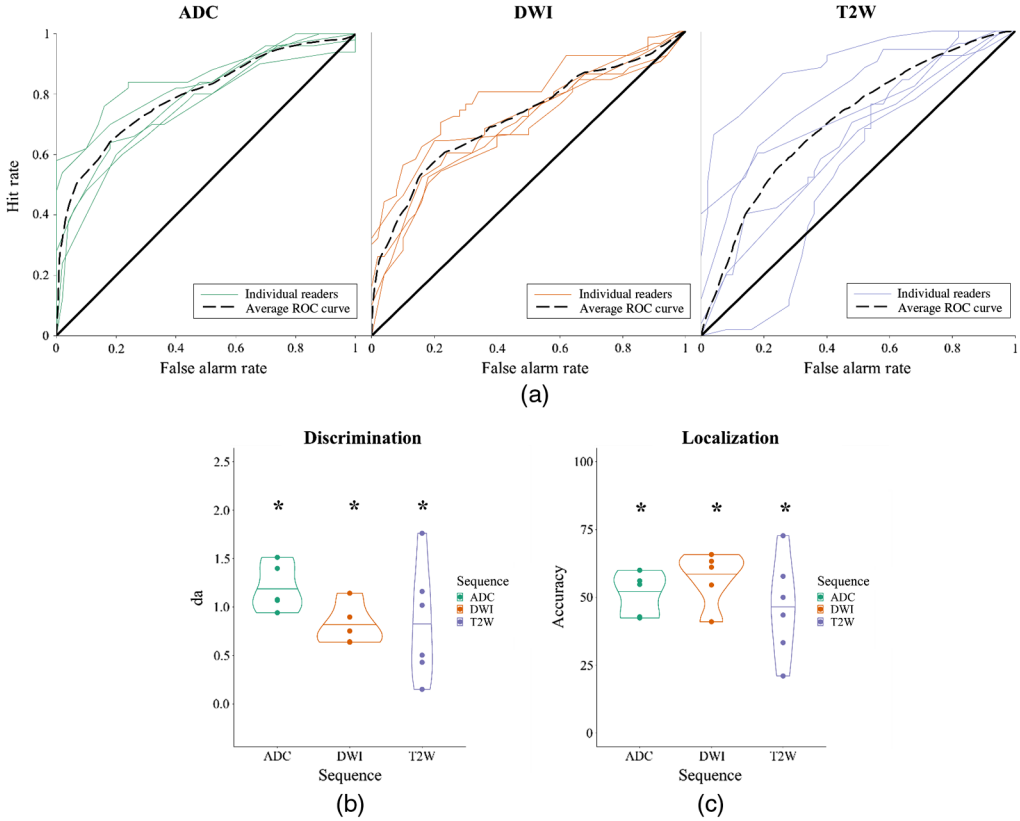


Fig. 5 Results for experiment 2. (a) ROC curves of individual readers' performances for each sequence condition: ADC, DWI, and T2W. (b) Discriminability performance for each condition measured in d_a. (c) Localization performance for all three conditions. Asterisks denote $p < 0.05$.

3.2.2 *Localization*

As in experiment 1, readers were able to localize lesions better than chance (51.27%, $p < 0.0001$) across all sequences. Localization was better than chance in all sequences. Unlike detection performance, localization was best for the DWI sequence, though this advantage was not statistically significant [ADC versus DWI: $t(8) = 1.04$, $p > 0.330$; ADC versus T2W: $t(9) = 1.18$, $p > 0.268$; T2W versus DWI: $t(9) = 1.18$, $p > 0.268$].

3.2.3 *Correlation*

Because the images used for the three sequences in experiment 2 came from the same set of cases, we can ask whether the same cases are easy in every sequence. It is possible that the three sequences carry completely independent information, so that, for example, the gestalt "pops out" in the ADC image for case 1, but not the T2W and DWI images, whereas for case 2 it is the T2W image that carries the most gestalt information, and so forth. Alternatively, the three sequences might be completely redundant from a gestalt processing point of view. To answer this question, we looked at the correlation among the three sequences. Overall, cases with more advanced cancer are more likely to have detectable lesions. Cancer severity is measured with Gleason scores, which we had for each of our cases (see below for analysis of performance by Gleason score). Therefore, we report the partial correlations (ρ) while controlling for Gleason score and performance on the third sequence (Table 3). The ρ scores were weak to moderate, between 0.12 and 0.31, indicating that the three sequences are neither completely independent nor completely redundant. Each contributes some additional information. The two functional sequences (ADC and DWI) were correlated with each other, as well as ADC and the anatomical (T2W) sequence, but T2W and DWI were not significantly correlated.

3.3 *Experiment 3*

Results for experiment 3 are shown in Fig. 6 and broken down by various categories in Table 5 in Sec. 6.

3.3.1 *Detection*

Readers were able to discriminate between normal and abnormal stacks at above chance levels ($d_a = 0.55$, $p < 0.0001$; AUC $= 0.65$, $p < 0.0002$), though performance was lower than what we had observed in previous experiments with static stimuli. We expected that detection would improve for slower presentation rates, which correspond to longer exposure times. Surprisingly, this was not the case; presentation rates did not seem to affect detection performance [$F(2, 11) = 0.22$, $p > 0.807$]. However, lesion visibility is not completely correlated with presentation rate, since some lesions are only visible in a single slice, some in two, and so forth. Multiplying the number of slices by the frame duration yields lesion duration, which is plotted in Fig. 6(c). (Note: Discriminability was measured in d' due to lack of slopes to compute d_a.) For example, a lesion that was visible in three slices and was presented for 48 ms per slice was visible for a total of 144 ms and produces a d' of 0.61. Lesion duration did not significantly affect performance [$F(2,18) = 0.97$, $p > 0.397$].

Table 3 Partial correlations, controlling for Gleason score and third sequence performance.

	T2W	DWI	ADC
T2W	—	0	0.29*
DWI		—	0.31*
ADC			—

*$p < 0.01$.

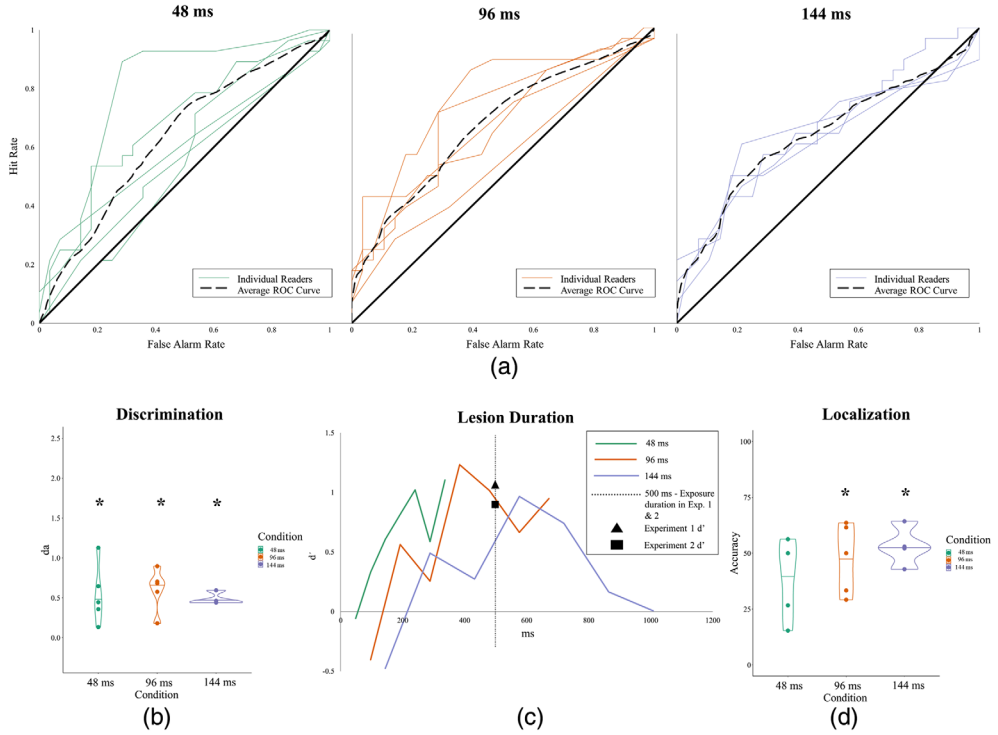


Fig. 6 Results for experiment 3. (a) ROC curves of individual readers' performances for each condition: 48, 96, and 144 ms. (b) Discriminability performance for each condition measured in d_a. (c) Discriminability performance of lesion duration for each condition measured in d'. (d) Localization performance for all conditions. Asterisks denote $p < 0.05$.

3.3.2 *Localization*

Localization overall was better than chance (46.30%, $p < 0.0001$). However, in the 48 ms condition, localization performance failed to significantly exceed chance (39.66%, $p = 0.050$). Similar to detection, localization performance did not differ as presentation rates slowed [$F(2,11) = 0.91$, $p > 0.429$].

3.4 *Effect of Gleason Score (Experiments 1 and 2)*

Figure 7 shows detection and localization as a function of Gleason score for experienced readers in experiment 1 and readers in the T2W condition of experiment 2 (see also Table 6 in Sec. 7). Detection is, in general, better for higher Gleason scores (more severe lesions), but the trend is weak and nonmonotonic. However, localization shows a much stronger trend.

4 Discussion

Professional readers' skilled ability to rapidly extract the gestalt of a medical image (i.e., discerning between normal and abnormal) has been demonstrated over the years with readers viewing static mammograms, chest radiographs, and cytology slides.[4–8] Experiments 1 and 2 provide further evidence that this early processing is a general property of expert medical image perception and not a specific feature of conventional imaging techniques; additionally, we demonstrate that a variety of factors (e.g., radiologic experience, sequence modality, and cancer characteristics) can affect readers' gestalt perception. Our main objective was to assess if rapid perceptual processing could be empirically observed in radiologists using 3-D imaging modalities. 3-D imaging is increasingly relied upon to improve detection and diagnostic accuracy in many different organs.[25–27] Experiment 3 demonstrates that experts viewing 3-D imaging can rapidly extract the gestalt from this new modality.

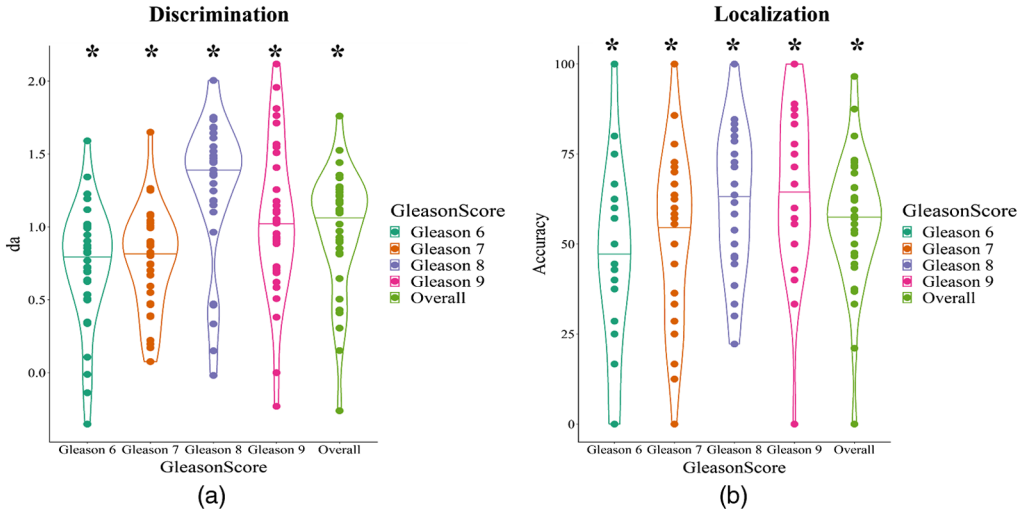


Fig. 7 (a) Discriminability performance as a function of Gleason score for experienced readers in experiment 1 and readers in the T2W condition of experiment 2. (b) Localization performance as a function of Gleason score. Asterisks denote $p < 0.05$.

The aim of experiment 1 was to test the hypothesis that rapid perceptual processing, defined as above-chance detection performance following a brief exposure, could be demonstrated using 2-D T2W mpMRI images of the prostate. Experienced readers could detect and localize a cancerous lesion after viewing an image for only half a second. Lesion detection ($d_a = 1.01$) (Experienced readers for experiment 1 had a d' of 1.06.) was comparable to gestalt performance in 2-D mammography ($d' = \sim 1$).[6] Lesion detection in the PZ was easier than in TZ, and localization was only possible for the PZ lesions and not for the TZ. The superior performance for PZ lesions may be due to experienced readers' expectations of prevalence, as the majority of prostate cancers occur in the PZ; moreover, TZ lesions have a low signal intensity on T2W images, making them difficult to detect.[28,29] Only experienced readers with 10 or more years reading prostate images were able to localize TZ lesions, while residents were unable to detect or localize these lesions. Readers who had no prior experience reading prostate cancer and experienced readers that read one or less prostate case a week were unable to detect or localize lesions over all zones.

Prostate lesions vary in clinical significance, as indexed by the Gleason score. We might expect more severe lesions to be more visually salient, and indeed Gleason 6 [A Gleason score 6 is well-differentiated (resembles normal cells), has a small volume of <0.5 cm^3, and is contained to the prostate.[30] Clinically, it is low-risk, rarely develops into an aggressive cancer, and is typically monitored through active surveillance.[31]], whose cells resemble more normal cells, are less likely to be detected in MRI than lesions with higher Gleason scores as in Refs. 32 and 33. In our experiments, experienced readers were able to detect and localize lesions in briefly presented images across all Gleason categories, including Gleason 6. Though readers were not asked to rate the clinical significance of the lesion (done in the clinical setting using PI-RADS version 2), they demonstrated detection and localization capabilities for even the more difficult to detect lesions.

MpMRI combines both anatomical information from the T2W sequence and functional information from the DWI and the ADC sequences. In experiment 2, we investigated rapid perceptual processing for these sequences. For each of the three sequences (i.e., T2W, DWI, and ADC), readers were able to detect and localize lesions after viewing an image for half a second. ADC had the highest detection performance ($d_a = 1.20$) followed by T2W ($d_a = 0.84$) and DWI ($d_a = 0.81$), though only ADC was found to be significantly higher than DWI. These results indicate that the ADC sequence may generate the strongest gestalt signal, but both anatomical and functional sequences can contribute to mpMRI gestalt. The inferior performance in the T2W group, relative to experiment 1, may be attributable to the subset of residents ($n = 3$, $d_a = 0.36$) who were not proficient in reading prostate mpMRI based on the number of years reading and viewing prostate cases each week (years reading prostate: $M = 0.40$, SD $= 0.53$; prostate cases each week: $M = 12.5$, SD $= 10.61$) compared to the practicing radiologists ($n = 3$, $d_a = 1.31$) (years reading prostate: $M = 9.00$, SD $= 5.57$; prostate cases each week: $M = 36.66$, SD $= 34.03$).

Moderate correlations between sequences suggest that information is partly overlapping, but each sequence provides some unique information to aid in rapid perceptual processing. Under normal viewing conditions, detection sensitivity has been shown to be higher for viewing DWI (59%) than T2W (48%),[34] but increased detection performance is achieved when viewing sequences jointly rather than in isolation,[35–38] suggesting gestalt perception could be stronger when all sequences are viewed together. However, this would depend on the relative contributions of foveal vision to rapid perceptual processing. In the experiments we report here, the prostate would have been centered in the field of vision, and in some cases, the lesion may have been foveated. In the standard mpMRI configuration, this could only be true for one of the sequences (at most) in a brief presentation.

In experiment 3, we measured rapid perceptual processing for 3-D images, presented as brief movies. We predicted that lesion detection would be above chance, and that performance would improve with slice duration (i.e., performance would be inversely related to presentation rate). The first prediction was confirmed, as detection and localization were overall well above chance. Readers viewing images for all viewing conditions performed well above chance at detection and were above chance level for localization for the 96-and 128-ms conditions. We did not observe the predicted performance advantage for slower presentation rates. Although we lack the power to compare presentation rates, the pattern of the preliminary results suggests that when lesion exposure was held constant, performance was better for faster presentation rates. Conversely, when holding the frame rate constant, exposure duration made little difference, though detection was poor for very brief or very long exposures, which may suggest that visual information about the lesion itself is less important than understanding the anatomical context around the lesion. This would be consistent with work on the importance of background statistics in camouflage breaking[39] and mammography.[40,41] Slow presentation rates may inhibit comprehension of the 3-D structure of the gland, perhaps due to the lack of motion cues. The optimal frame rate for processing 3-D anatomical information may be 20 Hz or better; however, future studies are needed.

A major limitation of the present study was power. The experiments were powered to detect differences from chance rather than differences between conditions. As a result, we cannot conclusively determine which mpMRI sequence in experiment 2 (i.e., ADC, DWI, and T2W) or which frame rate (i.e., 48, 96, and 144 ms) in experiment 3 is most advantageous for producing a gestalt signal. Displays for online and in-person readers were not calibrated to digital imaging and communications in medicine grayscale standard display function; therefore, image quality may have varied across displays. Benign lesions were not included in lesion-present trials, as our focus was to investigate gestalt processing for a cancerous lesion as those are the most detrimental for the patient and to utilize a metric to systemically vary lesions (i.e., Gleason score). As a result, we do not know if rapid perpetual processing exists when readers view benign lesions or if it is a characteristic of only cancerous lesions.

Understanding the perceptual and cognitive mechanisms underlying radiologists' performances is central to efforts to improve detection and diagnostic accuracy. Here, we focus on the phenomenon of rapid perceptual processing of medical images. Determining the factors that promote a strong gestalt signal has the potential to inform in the development of diagnostic tools and training protocols to channel the benefits of rapid visual processing. This study generalizes the phenomenon of rapid perceptual processing to mpMRI of the prostate and to 3-D imaging modalities. Our data indicate that radiologic experience is critical to developing gestalt processing ability. Both anatomical and functional sequences can contribute to gestalt perception. The effects of cancer severity (i.e., Gleason score) and lesion exposure duration were surprisingly weak. Our results also suggest an unexpected superiority for faster presentation rates, holding exposure duration constant.

These findings open up a number of potential lines of future investigation. For example, is prostate gestalt carried by the same high spatial frequency channels as with static mammograms? [7] Are the critical spatial frequency channels the same across the different mpMRI sequences or do the functional sequences convey information at lower spatial frequencies? How is information from different imaging sequences combined? Do radiologists intuitively scroll through 3-D image stacks at optimal frame rates? If not, can we improve detection by suggesting faster (or slower) scrolling? Answers to these questions will inform our efforts to help radiologists improve their performance.

5 Appendix A

Table 4 shows the discriminability and localization performance for experiment 2. In experiment 2, three groups of readers, all with experience in reading prostate mpMRI, viewed either static images from the T2W, DWI or ADC sequence.

Table 4 Results for experiment 2.

	d_a			AUC			Localization %		
	M (95% CI)	*t*	*p*	*M* (95% CI)	*t*	*p*	*M* (95% CI)	*t*	*P*
All ($n = 16$)	0.94 (0.74, 1.15)	9.01	<0.0001	0.74 (0.67, 0.80)	45.55	<0.0001	51.27 (44.78, 57.76)	10.29	<0.0001
T2W ($n = 6$)	0.84 (0.37, 1.31)	3.48	0.018	0.70 (0.56, 0.84)	6.04	0.012	46.39 (31.82, 60.96)	3.92	0.011
Radiologists ($n = 3$)	1.31 (0.87, 1.76)	5.78	0.029	0.81 (0.58, 1.03)	2.83	0.028	57.99 (41.43, 74.55)	4.83	0.040
Residents ($n = 3$)	0.36 (0.15, 0.57)	3.38	0.078	0.60 (0.50, 0.69)	6.51	0.045	34.80 (18.36, 51.24)	2.09	0.171
≤1 year reading prostate ($n = 3$)	0.36 (0.15, 0.57)	3.38	0.078	0.60 (0.50, 0.69)	6.51	0.045	34.80 (18.36, 51.24)	2.09	0.171
>1 and <10 years reading prostate ($n = 2$)	1.09 (0.95, 1.23)	15.27	0.042	0.77 (0.58, 0.95)	2.45	0.027	58.10 (29.44, 86.76)	1.91	0.197
≥10 years reading prostate ($n = 1$)	1.76	—	—	0.89	—	—	57.78	—	—
DWI ($n = 5$)	0.81 (0.63, 1.00)	8.61	0.001	0.71 (0.63, 0.80)	39.04	<0.0001	57.16 (48.45, 65.87)	8.98	0.001
Radiologists ($n = 4$)	0.86 (0.65, 1.07)	7.96	0.004	0.73 (0.64, 0.81)	22.04	<0.0001	57.81 (46.69, 68.93)	7.15	0.006
Residents ($n = 1$)	0.64	—	—	0.67	—	—	54.55	—	—
≤1 year reading prostate ($n = 1$)	0.64	—	—	0.67	—	—	54.55	—	—
>1 and <10 years reading prostate ($n = 3$)	0.76 (0.62, 0.91)	10.35	0.009	0.70 (0.61, 0.79)	49.47	<0.0001	55.16 (41.25, 69.07)	7.53	0.005
≥10 years reading prostate ($n = 1$)	1.14	—	—	0.86	—	—	65.79	—	—
ADC ($n = 5$)	1.20 (0.99, 1.41)	11.04	<0.0001	0.80 (0.73, 0.87)	19.13	<0.0001	51.23 (44.16, 58.30)	9.42	0.001
≤1 year reading prostate ($n = 2$)	1.01 (0.88, 1.14)	14.78	0.043	0.76 (0.70, 0.82)	20.12	<0.0001	51.21 (33.98, 68.44)	3.87	0.161
>1 and <10 years reading prostate ($n = 2$)	1.23 (0.91, 1.56)	7.48	0.085	0.80 (0.62, 0.98)	2.00	0.018	55.43 (54.13, 56.73)	7.78	0.016
≥10 years reading prostate ($n = 1$)	1.51	—	—	—	—	—	42.86	—	—

Table 4 (*Continued*).

	PZ d_a			PZ AUC			PZ Localization %		
	M (95% CI)	*t*	*p*	*M* (95% CI)	*t*	*p*	*M* (95% CI)	*t*	*P*
All ($n = 16$)	0.92 (0.72, 1.13)	8.86	<0.0001	0.73 (0.66, 0.81)	45.77	<0.0001	53.49 (45.37, 61.61)	9.40	<0.0001
T2W ($n = 6$)	0.86 (0.42, 1.29)	3.87	0.012	0.71 (0.58, 0.85)	6.52	0.008	49.87 (28.99, 70.75)	3.32	0.021
Radiologists ($n = 3$)	1.29 (0.90, 1.68)	6.46	0.023	0.81 (0.66, 0.96)	3.64	0.007	69.10 (50.25, 87.95)	5.67	0.030
Residents ($n = 3$)	0.42 (0.17, 0.68)	3.25	0.083	0.61 (0.50, 0.72)	6.34	0.045	30.64 (10.52, 50.76)	1.57	0.257
≤1 year reading prostate ($n = 3$)	0.42 (0.17, 0.68)	3.25	0.083	0.61 (0.50, 0.72)	6.34	0.045	30.64 (10.52, 50.76)	1.57	0.257
>1 and <10 years reading prostate ($n = 2$)	1.09 (1.04, 1.14)	41.72	0.015	0.77 (0.68, 0.86)	7.16	<0.0001	63.33 (37.19, 89.47)	3.66	0.170
≥10 years reading prostate ($n = 1$)	1.69	—	—	0.89	—	—	80.65	—	—
DWI ($n = 5$)	0.79 (0.52, 1.06)	5.75	0.005	0.71 (0.60, 0.82)	23.49	0.001	57.54 (48.24, 66.84)	9.07	0.001
Radiologists ($n = 4$)	0.86 (0.55, 1.17)	5.46	0.012	0.72 (0.61, 0.84)	12.69	0.001	58.29 (46.44, 70.14)	7.24	0.005
Residents ($n = 1$)	0.53	—	—	0.65	—	—	54.55	—	—
≤1 year reading prostate ($n = 1$)	0.53	—	—	0.65	—	—	54.55	—	—
>1 and <10 years reading prostate ($n = 3$)	0.70 (0.63, 0.77)	20.28	0.002	0.69 (0.59, 0.79)	40.96	<0.0001	53.91 (42.36, 65.46)	6.68	0.022
≥10 years reading prostate ($n = 1$)	1.32	—	—	0.83	—	—	71.43	—	—
ADC ($n = 5$)	1.13 (0.85, 1.41)	7.79	0.001	0.78 (0.69, 0.88)	14.36	<0.0001	53.78 (49.27, 58.29)	17.02	<0.0001
≤1 year reading prostate ($n = 2$)	0.91 (0.71, 1.12)	8.77	0.072	0.74 (0.66, 0.82)	10.19	<0.0001	52.66 (42.79, 62.53)	7.57	0.084
>1 and <10 years reading prostate ($n = 2$)	1.24 (0.57, 1.90)	3.63	0.171	0.80 (0.47, 1.13)	2.00	0.060	56.79 (50.48, 63.10)	13.14	0.048
≥10 years reading prostate ($n = 1$)	1.35	—	—	0.83	—	—	50	—	—

Table 4 (*Continued*).

	TZ d_a			TZ AUC			TZ Localization %		
	M (95% CI)	*t*	*p*	*M* (95% CI)	*t*	*p*	*M* (95% CI)	*t*	*P*
All ($n = 16$)	0.97 (0.71, 1.22)	7.53	<0.0001	0.74 (0.64, 0.85)	23.29	<0.0001	46.27 (36.98, 55.56)	5.28	<0.0001
T2W ($n = 6$)	0.74 (0.20, 1.29)	2.66	0.045	0.68 (0.52, 0.85)	6.86	0.036	39.31 (19.25, 59.37)	1.76	0.138
Radiologists ($n = 3$)	1.24 (0.49, 1.99)	3.23	0.084	0.80 (0.56, 1.04)	3.15	0.030	33.33 (0.66, 66.00)	0.72	0.544
Residents ($n = 3$)	0.25 (0.19, 0.31)	7.76	0.016	0.56 (0.46, 0.67)	6.73	0.179	45.29 (16.88, 73.70)	1.66	0.239
≤1 year reading prostate ($n = 3$)	0.25 (0.19, 0.31)	7.76	0.016	0.56 (0.46, 0.67)	6.73	0.179	45.29 (16.88, 73.70)	1.66	0.239
>1 and <10 years reading prostate ($n = 2$)	0.92 (0.20, 1.63)	2.52	0.240	0.74 (0.42, 1.07)	2.00	0.084	46.43 (11.44, 81.42)	1.41	0.393
≥10 years reading prostate ($n = 1$)	1.88	—	—	0.91	—	—	7.14	—	—
DWI ($n = 5$)	0.90 (0.68, 1.11)	8.05	0.001	0.73 (0.57, 0.89)	15.21	0.008	55.15 (44.08, 66.22)	6.00	0.004
Radiologists ($n = 4$)	0.90 (0.61, 1.18)	6.24	0.008	0.73 (0.57, 0.89)	14.18	0.009	55.30 (41.01, 69.59)	4.67	0.019
Residents ($n = 1$)	0.89	—	—	0.72	—	—	54.55	—	—
≤1 year reading prostate ($n = 1$)	0.89	—	—	0.72	—	—	54.55	—	—
>1 and <10 years reading prostate ($n = 3$)	0.94 (0.56, 1.32)	4.86	0.040	0.74 (0.55, 0.93)	14.41	0.018	57.06 (37.46, 76.66)	3.58	0.070
≥10 years reading prostate ($n = 1$)	0.76	—	—	0.71	—	—	50	—	—
ADC ($n = 5$)	1.30 (1.00, 1.60)	8.56	0.001	0.83 (0.73, 0.94)	16.23	<0.0001	45.76 (32.45, 59.07)	3.61	0.023
≤1 year reading prostate ($n = 2$)	1.24 (1.08, 1.39)	15.56	0.041	0.81 (0.70, 0.93)	11.62	<0.0001	48.81 (18.47, 79.15)	1.78	0.326
>1 and <10 years reading prostate ($n = 2$)	1.09 (0.81, 1.38)	7.53	0.084	0.81 (0.55, 1.06)	2.61	0.035	51.30 (39.85, 62.75)	5.14	0.122
≥10 years reading prostate ($n = 1$)	1.86	—	—	0.91	—	—	28.57	—	—

6 Appendix B

Table 5 shows the discriminability and localization performance for experiment 3. In experiment 3, three groups of readers, with experience in prostate mpMRI, viewed cases that comprised a stack of T2W slices. Image slices were presented at either 48 ms, 96, or 144 ms.

Table 5 Results for experiment 3.

	d_a			AUC			Localization %		
	M (95% CI)	*t*	*p*	*M* (95% CI)	*t*	*P*	*M* (95% CI)	*t*	*P*
All ($n = 14$)	0.55 (0.41, 0.69)	7.83	<0.0001	0.65 (0.56, 0.74)	43.05	0.001	46.30 (38.49, 54.11)	7.16	<0.0001
48 ($n = 5$)	0.54 (0.21, 0.87)	3.23	0.032	0.63 (0.50, 0.77)	11.31	0.049	39.66 (24.18, 55.14)	2.77	0.050
>1 and <10 years reading prostate ($n = 3$)	0.64 (0.07, 1.20)	2.21	0.157	0.67 (0.41, 0.93)	3.45	0.127	40.54 (15.63, 65.45)	1.79	0.215
≥10 years reading prostate ($n = 2$)	0.40 (0.32, 0.49)	9.24	0.069	0.58 (0.47, 0.68)	6.82	0.118	38.33 (15.46, 61.20)	1.76	0.329
96 ($n = 5$)	0.61 (0.38, 0.84)	5.12	0.007	0.67 (0.56, 0.78)	16.73	0.004	47.53 (33.66, 61.40)	4.21	0.014
>1 and <10 years reading prostate ($n = 3$)	0.59 (0.17, 1.01)	2.78	0.109	0.66 (0.47, 0.86)	4.63	0.084	47.60 (27.96, 67.24)	2.98	0.097
≥10 years reading prostate ($n = 2$)	0.63 (0.52, 0.73)	11.78	0.054	0.69 (0.56, 0.82)	2.00	0.026	47.44 (19.81, 75.07)	2.10	0.283
144 ($n = 4$)	0.48 (0.41, 0.56)	13.08	0.001	0.64 (0.53, 0.75)	31.22	0.013	53.06 (44.47, 61.65)	8.05	0.004
Radiologists ($n = 3$)	0.45 (0.43, 0.46)	58.93	<0.0001	0.63 (0.52, 0.75)	27.68	0.022	53.36 (41.23, 65.49)	5.75	0.029
Residents ($n = 1$)	0.59	—	—	0.66	—	—	52.17	—	—
>1 and <10 years reading prostate ($n = 2$)	0.52 (0.37, 0.67)	6.71	0.094	0.64 (0.53, 0.76)	20.92	0.018	58.23 (46.37, 70.09)	6.68	0.095
≥10 years reading prostate ($n = 2$)	0.45 (0.43, 0.47)	39.62	0.016	0.64 (0.53, 0.74)	13.72	0.013	47.90 (38.02, 57.78)	5.97	0.106

7 Appendix C

Table 6 shows the results of discriminability and localization performance as a function of Gleason score for experienced readers in experiment 1 and readers in the T2W condition of experiment 2.

Table 6 Results of discriminability and localization performance as a function of Gleason score for experienced readers in experiment 1 and readers in the T2W condition of experiment 2.

	d_a			Localization %		
	M (95% CI)	*t*	*p*	*M* (95% CI)	*T*	*p*
Gleason 6	0.74 (0.60, 0.88)	10.12	<0.0001	45.19 (36.31, 54.07)	5.71	<0.0001
Gleason 7	0.77 (0.66, 0.89)	12.78	<0.0001	52.13 (44.00, 60.10)	8.47	<0.0001
Gleason 8	1.19 (0.95, 1.43)	9.61	<0.0001	60.01 (52.18, 67.84)	10.85	<0.0001
Gleason 9	1.05 (0.87, 1.23)	11.45	<0.0001	66.38 (58.65, 74.11)	12.68	<0.0001
Combined	0.98 (0.83, 1.12)	13.27	<0.0001	56.48 (50.17, 62.78)	12.2	<0.0001

Disclosures

The authors have no competing interests to declare and do not have any relevant disclosures.

References

1. M. C. Potter and B. A. Faulconer, "Time to understand pictures and words," *Nature* **253**, 437–438 (1975).
2. P. G. Schyns and A. Oliva, "From blobs to boundary edges: evidence for time- and spatial-scale-dependent scene recognition," *Psychol. Sci.* **5**, 195–200 (1994).
3. E. A. Krupinski and R. S. Weinstein, "Changes in visual search patterns of pathology residents as they gain experience," *Proc. SPIE* **7966**, 79660P (2011).
4. A. J. Carrigan, S. G. Wardle, and A. N. Rich, "Finding cancer in mammograms: if you know it's there, do you know where?" *Cognit. Res. Princ. Implic.* **3**, 10 (2018).
5. D. P. Carmody, C. F. Nodine, and H. L. Kundel, "An analysis of perceptual and cognitive factors in radiographic interpretation," *Perception* **9**, 339–344 (1980).
6. K. K. Evans et al., "The gist of the abnormal: above-chance medical decision making in the blink of an eye," *Psychon. Bull. Rev.* **20**, 1170–1175 (2013).
7. K. K. Evans et al., "A half-second glimpse often lets radiologists identify breast cancer cases even when viewing the mammogram of the opposite breast," *Proc. Natl. Acad. Sci. U. S. A.* **113**, 10292–10297 (2016).
8. H. L. Kundel and C. F. Nodine, "Interpreting chest radiographs without visual search," *Radiology* **116**, 527–532 (1975).
9. A. Oliva et al., "Top-down control of visual attention in object detection," in *Proc. Int. Conf. Image Process.* (2003).
10. M. S. Castelhano and J. M. Henderson, "Initial scene representations facilitate eye movement guidance in visual search," *J. Exp. Psychol. Hum. Percept. Perform.* **33**, 753–763 (2007).
11. A. Oliva and A. Torralba, "Modeling the shape of the scene: a holistic representation of the spatial envelope," *Int. J. Comput. Vision* **42**, 145–175 (2001).
12. A. Oliva and P. G. Schyns, "Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli," *Cognit. Psychol.* **34**, 72–107 (1997).
13. A. Oliva and A. Torralba, "Building the gist of a scene: the role of global image features in recognition," *Prog. Brain Res.* **155**, 23–36 (2006).

14. M. D. Greer, P. L. Choyke, and B. Turkbey, "PI-RADSv2: how we do it," *J. Magn. Reson. Imaging* **46**, 11–23 (2017).
15. T. Hambrock et al., "Relationship between apparent diffusion coefficients at 3.0-T MR imaging and Gleason grade in peripheral zone prostate cancer," *Radiology* **259**, 453–461 (2011).
16. R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria (2019).
17. S. Champely et al., "pwr: basic functions for power analysis," R Foundation for Statistical Computing, Vienna, Austria (2018).
18. J. Peirce et al., "PsychoPy2: experiments in behavior made easy," *Behav. Res. Methods* **51**, 195–203 (2019).
19. J. R. de Leeuw, "jsPsych: a JavaScript library for creating behavioral experiments in a web browser," *Behav. Res. Methods* **47**, 1–12 (2015).
20. J. C. Weinreb et al., "PI-RADS prostate imaging—reporting and data system: 2015, version 2," *Eur. Urol.* **69**, 16–40 (2016).
21. A. Bertana et al., "Confidence predicts variability but not biases in perceptual decisions," *J. Vision* **18**, 1047–1047 (2018).
22. N. A. Macmillan and C. D. Creelman, *Detection Theory: A User's Guide*, Lawrence Erlbaum Associates, Mahwah, New Jersey (2005).
23. B. D. Gallas et al., "A framework for random-effects ROC analysis: biases with the bootstrap and other variance estimators," *Commun. Stat. Theory Methods* **38**, 2586–2603 (2009).
24. N. A. Obuchowski, B. D. Gallas, and S. L. Hillis, "Multi-reader ROC studies with split-plot designs: a comparison of statistical methods," *Acad. Radiol.* **19**, 1508–1517 (2012).
25. H. Abdi et al., "Multiparametric magnetic resonance imaging enhances detection of significant tumor in patients on active surveillance for prostate cancer," *Urology* **85**, 423–429 (2015).
26. N. Houssami and P. Skaane, "Overview of the evidence on digital breast tomosynthesis in breast cancer detection," *The Breast* **22**, 101–108 (2013).
27. D. T. Oberlin et al., "Dramatic increase in the utilization of multiparametric magnetic resonance imaging for detection and management of prostate cancer," *Abdom. Radiol.* **42**, 1255–1258 (2017).
28. A. Kayhan et al., "Multi-parametric MR imaging of transition zone prostate cancer: imaging features, detection and staging," *World J. Radiol.* **2**, 180–187 (2010).
29. J. McNeal, "Normal histology of the prostate," *Am. J. Surg. Pathol.* **12**, 619–633 (1988).
30. G. Ploussard et al., "The contemporary concept of significant versus insignificant prostate cancer," *Eur. Urol.* **60**, 291–303 (2011).
31. M. Komisarenko, L. J. Martin, and A. Finelli, "Active surveillance review: contemporary selection criteria, follow-up, compliance and outcomes," *Transl. Androl. Urol.* **7**, 243–255 (2018).
32. J. H. Ellis et al., "MR imaging and sonography of early prostatic cancer: pathologic and imaging features that influence identification and diagnosis," *Am. J. Roentgenol.* **162**, 865–872 (1994).
33. S. Ikonen et al., "Magnetic resonance imaging of prostatic cancer: does detection vary between high and low Gleason score tumors?" *The Prostate* **43**, 43–48 (2000).
34. S. Gaur et al., "A multireader exploratory evaluation of individual pulse sequence cancer detection on prostate multiparametric magnetic resonance imaging (MRI)," *Acad. Radiol.* **26**, 5–14 (2019).
35. S. Gaur et al., "Can computer-aided diagnosis assist in the identification of prostate cancer on prostate MRI? A multi-center, multi-reader investigation," *Oncotarget* **9**, 33804–33817 (2018).
36. H. Miao, H. Fukatsu, and T. Ishigaki, "Prostate cancer detection with 3-T MRI: comparison of diffusion-weighted and T2-weighted imaging," *Eur. J. Radiol.* **61**, 297–302 (2007).
37. P. Preziosi et al., "Enhancement patterns of prostate cancer in dynamic MRI," *Eur. Radiol.* **13**, 925–930 (2003).
38. B. Zelhof et al., "Correlation of diffusion-weighted magnetic resonance data with cellularity in prostate cancer," *BJU Int.* **103**, 883–888 (2009).

39. X. Chen and J. Hegdé, "Learning to break camouflage by learning the background," *Psychol. Sci.* **23**, 1395–1403 (2012).
40. J. Hegde, "Role of statistical learning in radiological diagnosis of cancer," in *Med. Image Perception Soc. Conf. XVI*, Vol. 16, p. 17 (2015).
41. E. Kompaniez et al., "Adaptation aftereffects in the perception of radiological images," *PLoS One* **8**, e76175 (2013).

Melissa Treviño is a postdoctoral fellow in the Basic Biobehavioral and Psychological Sciences Branch at the U.S. National Cancer Institute. She work focuses on applying basic visual perception and cognition principles to investigate and identify properties that foster early perceptual processing of the radiologic search process. Her work also focuses on bridging the gap between cognitive science and clinical neuropsychology to advance cognitive measures used to assess cancer patients and survivors.

Todd S. Horowitz is a cognitive psychologist and program director in the Cancer Control and Population Sciences Division at the National Cancer Institute. He has published over 70 peer-reviewed research papers. Currently, he is working to engage cognitive psychologists and vision scientists with problems in cancer control, such as improving medical image interpretation, studying the cognitive effects of cancer and cancer treatments, and improving the effectiveness of visual health communications.

Biographies of the other authors are not available.