Expertise in Interactions of Perceptual and Conceptual Processing

Junya Morita[1] (morita@cog.human.nagoya-u.ac.jp), Kazuhisa Miwa[2] (miwa@cog.human.nagoya-u.ac.jp)
Takayuki Kitasaka[2] (kitasaka@is.nagoya-u.ac.jp), Kensaku Mori[2] (kensaku@is.nagoya-u.ac.jp)
Yasuhito Suenaga[2] (suenaga@is.nagoya-u.ac.jp), Shingo Iwano[3] (iwano45@med.nagoya-u.ac.jp)
Mitsuru Ikeda[4] (a40495a@nucc.cc.nagoya-u.ac.jp), Takeo Ishigaki[3] (i45265a@nucc.cc.nagoya-u.ac.jp)

[1]Graduate School of Human Informatics. Nagoya University. Furo-cho. Chigusa-ku. Nagoya, Japan.
[2]Graduate School of Information Science. Nagoya University. Furo-cho. Chigusa-ku. Nagoya. Japan.
[3]Graduate School of Medicine. Nagoya University. 65, Tsurumai-cho, Showa-ku, Nagoya. Japan.
[4]School of Health Science. Nagoya University. 1-20, Daikou-Minami 1-chome. Higashi-ku. Nagoya. Japan.

Abstract

This paper is an investigation on the development of expertise in 3D-CT diagnosis. The protocols data obtained in the experiment were analyzed by using a Japanese morphological analysis system called ChaSen. The results of the protocols analysis revealed differences between the expert radiologists and the novice residents in (1) the number of verbalized words, (2) the variations in perceptual and conceptual vocabularies, and (3) the frequencies of cycles between perceptual and conceptual processing. The results suggest that expertise in medical image diagnosis involves not only the development of both perceptual and conceptual processing, but also the development of an ability to coordinate between perceptual and conceptual processing.

Keywords: Protocol Analysis; Image Diagnosis; Morphological Analysis.

Introduction

Medical image diagnosis is a task in which a physician makes a diagnosis while viewing medical images such as radiography or 3D-CT (Computed Tomography) images. Many studies have so far been conducted to explore cognitive factors underlying the task, employing methods of eye tracking and think-aloud protocols analysis. We considered that it is generally important in the theoretical and empirical development of cognitive science to understand the cognitive process behind the task. Prior to presenting our study, we briefly review previous findings of medical image diagnosis.

The first finding of medical image diagnosis relates to cognitive components of medical image diagnosis. In image diagnosis, a physician perceives abnormal findings on a medical image, and makes a diagnosis about what disease affects the patient. The former activity mainly concerns "perceptual processing," including a visual search and object-recognition process. The latter activity relates to "conceptual processing," including a decision-making and a hypothesis-testing process.

Traditionally, the process of medical image diagnosis has been regarded as a one-way, bottom-up process, where perceptual processing follows conceptual processing. In fact, conventional medical reports follow a format in which "findings" and "impressions" are separately described: physicians usually write down their findings followed by impressions.

However, contrary to the above traditional view, many researchers have agreed on the interactive aspects of diagnostic process (e.g., Krupinski, 2003). Empirically, Norman, Brooks, Coblentz, and Babcook (1992) demonstrated that perceived features are dramatically influenced by clinical charts in which any previous disease of the patients is indicated. Lesgold, Rubinson, Feltovitch, Glaser, Klopfer, and Wang (1988) also conducted protocols analysis studies, which confirmed the interactive cycles between bottom-up and top-down processes, where an initial hypothesis is immediately triggered after a first glance at medical images, followed by searching for abnormalities in the medical images.

The second previous finding of medical image diagnosis is on the development of expertise, which significantly improves the two components of medical image diagnosis (Woods, 1999). For example, Myles-Worsley, Johnston, and Simons (1988) demonstrated that expert radiologists could discriminate abnormal X-ray films from normal ones within 500 msec. Additionally, Lesgold et al confirmed that expert radiologists reported more findings, verbalized more hypotheses, and showed more and longer reasoning chains than novice residents did.

Following the above studies, we conducted a protocols analysis study, which was designed to achieve two goals. The first goal was to investigate the development of expertise in 3D-CT diagnosis. There are many protocols analysis studies on medical image diagnosis (Lesgold, et al., 1988; Raufaste, Eyrolle, and Mariné, 1998; Rogers, 1996; Azevedo, and Lajoie, 1988). However the authors found no prior protocols analysis studies on 3D-CT diagnosis. Recently, medical images mainly used for diagnosis have been changing to 3D-CT images, which have distinct features, compared with radiography. Basically, CT scanners generate cross-sectional images of a human body, and 3D human anatomical structures are reconstructed by piling them up. Although 3D information provided by CT images enables physicians to observe image features precisely, such large amounts of information considerably increase the complexity of the task.

The second aim of our study was to reveal the nature of the interactions between perceptual and conceptual processing. Although many researchers agree on the interactive aspects of medical image diagnosis, there have been only a few quantitative investigations into the interactive processes of perceptual and conceptual processing. More importantly, it is still unclear how the development of expertise influences the interactions of the two components (e.g., Norman, et.al, 1992). Therefore, we quantify the interactions between perceptual and conceptual processing by developing our own protocols analysis method, and attempt to demonstrate the effects of expertise on the interactions of the two components.

Method

In order to investigate medical image diagnosis in a realistic context, the experiment was performed in a room located in the radiology department at Nagoya University, where participants in our experiment usually work.

In the experiment, participants were required to make "differential diagnoses of lung nodules (malignant/benign)." We considered that this task has an advantage for investigating the interactions between perceptual and conceptual processing. In order to make a differential diagnosis of lung nodules, physicians are required to consider several diseases entities, such as *pneumonia*, *tuberculosis*, *benign tumors*, *malignant tumors*, etc. Additionally, since these diseases show common features on medical images, physicians need to consider complex conjunctions of several physical features [e.g., *"spiculation," "over 1 cm," "round shape," and "converging blood-vessels."*] Therefore, we thought that not only perceptual processing but also conceptual processing is inevitably involved in solving the task.

Materials

Cases We chose case materials from a research database, which consisted of cases whose diagnoses had already been determined by operations, biopsies, or follow-up examinations. The cases were randomly chosen, all of which contain at least one nodular lesion. In a later section, we refer the most significant nodular lesion in each case as a *target lesion*. The selected cases consisted of eight benign and six malignant cases. The benign cases had been diagnosed as *tuberculosis*, *organizing pneumonia*, *amyloidosis*, or *benign tumor*, while the malignant cases included a variety of lung cancers, such as *well differentiated carcinoma* and *squamous cell carcinoma*.

CT data sets Each of the cases consisted of three types of CT data sets; we refer hereafter to these three types of CT data as *lung window CT*, *mediastinal window CT*, and *high-resolution CT*.

The lung window CT data set means the CT data with a window level of -600 H.U., a window width of 1800 H.U, and a slice thickness of 5-10 mm, and shows the overall lung area composed of 30 to 50 slices. By using this type of data set, a physician judges the location of a target lesion and observes a base diseases of the lung area such as *emphysema* or *interstitial pneumonia*.

The mediastinal window CT data set is the same as the lung window CT data set, except that display conditions are adjusted to show the mediastinal area clearly (window level, 50 H.U; window width, 300 H.U). By using this type of image, physicians investigate abnormalities in *mediastinal regions* and *axillary regions*. This data set is used to confirm the presence of calcifications inside the target lesion.

The high-resolution CT means the data set focusing on the target lesion (resolution, about 300 μm; slice thickness, 0.5 to 2 mm). Usually, physicians use this type of image to investigate important features of the target lesion (*density, shape*) and relations to lung tissues (*blood vessels, bronchi, and pleural membranes*).

Clinical histories In this experiment, no clinical information other that CT images was presented because the previous studies indicated significant influence of clinical histories on the accuracy of diagnosis.

Devices for viewing CT images The CT images were presented by using the tools usually employed by the participants (a workstation, two medical monochrome LCDs, and a DICOM viewer). Each of the two LCDs is able to display several slices at the same time, and the operations can use the workstation' mouse to transform those slices displayed. Furthermore, the DICOM viewer provides several supporting tools, such as a tuning function of window level, a measurement tool for sizes and densities of shadows, and a synchronization function for different types of CT data sets. In this experiment, participants were allowed to use theses tools without any constraints.

Devices for writing medical reports To enhance the reality of the experimental situation, we asked participants to write formal medical reports. The medical reports were written using a reporting system that provides two text forms: "findings" and "impressions".

Participants

Ten participants were recruited from the radiology department at Nagoya University, including five experts and five novices. All of the experts were radiologists, who were on the academic staff of the radiology department. They had five to twenty years experience in image diagnosis. On the other hand, the novices were residents and graduate students of the radiology department. They were physicians who had completed the degree of undergraduate medicine, and had less than two years experience in image diagnosis.

Procedure

The experiment required a total of two to four hours, which was divided into the following four stages.

1. Instructions Each participant was given the following instructions: "imagine the situations where abnormal findings were detected as a result of screening tests. Your task is to make differential diagnoses of the detected abnormal findings." Following this, the participant was also instructed to verbalize all of their thoughts without filtering them.

2. A Practice Task Each participant diagnosed one of the benign cases while being prompted to talk aloud. If the participant did not talk aloud for more than about ten seconds, the experimenter prompted the participants by an encouragement such as "please continue to talk aloud." The data obtained in the practice task were excluded from the later analysis.

3. Main tasks In the main tasks, each participant made diagnoses on thirteen cases that included seven benign and six malignant cases. The presentation order was randomized between the participants. For each case, the participants investigated the CT images and wrote a medical report about abnormal findings and suspected diseases. Since writing a medical report proceeded almost simultaneously with investigation of the CT images, we did not discriminate the two activities in the later analysis.

4. Interviews Following the diagnosis of each case, an experimenter interviewed the participants. Preventing the participants from a specific bias in answering, the experimenter used only the following three questions.

1. Explain their own diagnostic processes. The participant was asked to report how he/she discovered the findings written down in the medical reports.

2. Report findings that were not written in the reports. The participant was prompted to report findings that he/she did not write down in the medical reports.

3. Rate the probability of malignancy. The participant was asked to rate how strong he/she felt the target lesion was malignant (0: absolutely benign to 10: absolutely malignant).

Table 1: Definitions of tags

Category	Subcategory	Examples
Percept	Size	small, large, cm
	Shape	round, polygon
	Outline	spiculation, borderline
	Number	singly, multiple, several
	Inside	dark, grand-glass-opacity
	Relation	continue, attract
Concept	Disease	tuberculosis, Lung cancer
	Surgery	post surgery, cut off
	Artifact	gravity,
	Treatment	follow-up, biospy
	Judgment	benign-malignancy
	Others	meta, infarction, emphysema
Region	Lung	right/left-lung, S1, S2...
	Mediastinum	heart, main-artery
	Abdomen	liver, pancreas
	Bronchus	B1, B2...
	Blood vein	blood-vein
	Pleura	major-fissure, chest-wall
Goal	Image	window-level, HRCT
	Action	diagnosis, differentiate

Table 2: Outputs of ChaSen with a new dictionary

Example 1			Example 2		
Words	Sub	Cat	Words	Sub	Cat
S8	lung	reg	pneumonia	disease	con
S10	lung	reg	trace	other	con
right-lung	lung	reg	pneumonia	disease	con
S10	lung	reg	benign-tumor	disease	con
light	dense	per	trace	others	con
nodule	shape	per	malignant	disease	con
spicula	line	per	deny	judge	con

Data analysis

Recorded data We recorded all verbalizations and videotaped the images and texts on the LCDs. The verbalizations and the texts in the medical reports were transcribed. Therefore, the protocols data included: (1) the verbalizations during the main tasks, (2) the texts written in the medical reports, and (3) the verbalizations during the interviews. Since the first and second types of data are supposed to represent the participants' thinking process in the main tasks, we analyzed two types of data all together to explore their diagnostic processes. On the other hand, the third type of data were retrospective reports. Therefore, we used the data in analysis only when we investigated overall trends of their diagnoses.

Protocols analysis The protocols data obtained in our experiment were so large that the traditional hand-coding protocols analysis was practically difficult. In order to ensure reliability of coding, and to conduct detailed quantitative analysis, we developed a semi-automatic protocols analysis method, in which the Japanese morphological analysis system ChaSen was used. ChaSen is a standard tool for text analysis and text mining in Japan. The system automatically converts plain texts to word sequences, using dictionaries of words and grammar. In this analysis, we directly described semantic tags in ChaSen's word dictionary, ChaSen outputted tagged words sequentially. The coding procedure comprised the following five stages.

1. Morphological analysis with the default dictionary. First, all of the protocol data were input into ChaSen. ChaSen then analyzed the data with the default word dictionary, and output 104,473 words.

2. Selecting the words. Most of the words output by the above procedure were syncategorematic terms (e.g., prepositions), or words that did not directly relate to the diagnostic activities (e.g., conjunctions, fillers). Therefore, these kinds of word were eliminated in the later analysis.

3. Creating a new dictionary. We created a new word dictionary comprising the words selected by the above procedure. Additionally, technical terms that were not appropriately discriminated by the default dictionary were registered into the dictionary.

4. Marking semantic tags. A semantic tag was labeled in each of the words. Examples of the tags are shown in Table 1. The tags were divided into four main categories: *percept* (338 words), *concept* (143 words), *region* (166 words), and *goal* (15 words). The *percept* concerns a vocabulary of perceptual features, which can be directly observed from the CT images. The *concept* indicates a word concerning physiological or pathological features on the CT images, such as a disease name and a surgery method. The *region* indicates lung area or an organization of the lung, which is a technical term of anatomy, and the *goal* indicates a word concerning a type of CT images or a word relating the task that the physicians performed.

5. Morphological analysis with the new dictionary. After deletion of the default dictionary, morphological analysis was again conducted with the new dictionary. ChaSen with the

Table 3: Performance [Mean (SD)]

	Accuracy score	Time (seconds)
Novices	1.05 (2.32)	539 (87)
Experts	1.55 (2.46)	491 (121)

new dictionary output a total of 19,678 words (13,984 in the main tasks; 5,694 in the interviews). Table 2 shows examples of the output words.

Statistical tests In this paper, a unit of statistical tests was the average value for each case (n = 13).

Results and Discussion

1. Performance

Prior to presenting the results of protocols analysis, we indicate overall performances of the experts and novices. Table 3 shows (a) accuracy score, and (b) time (seconds) to complete medical reports.

(a) Accuracy score The accuracy score was calculated by subtracting five points from a participant's rating score of malignancy. In the cases where the target lesion was benign, the scores were reversed. Thus, the scores ranged from −5 to 5, with higher scores indicate more confident correct diagnoses, and lower scores more confident incorrect diagnoses. A dependent groups t test revealed that the experts made more accurate diagnoses than the novices did [$t(12) = 1.80, p < .05$].

(b) Required time Past research indicated that the development of expertise reduces the time to make a diagnosis (e.g., Azevedo and Lajoie, 1998). From this, we tried to confirm the difference between the experts and the novices in the time taken to complete the medical reports. A dependent groups t test revealed that the experts made diagnoses faster than the novices [$t(12) = 1.81, p < .05$].

The above two results confirmed that the experts were superior to the novices in 3D-CT diagnosis.

2. Results of protocols analysis

In this section, we show three results of protocols analysis: (a) overall patterns of tagged words, (b) patterns of tagged words in the earlier and later phases of the diagnostic process, and (3) frequencies of the interactive cycles of perceptual and conceptual processing. The first analysis aimed to reconfirm the previous findings of X-ray film diagnosis, while the second and third analyses concerned the interactive process of the perceptual and conceptual processing[1].

(a) Overall patterns of tagged words Table 4 shows two scores of four tag types: (1) mean numbers of words per case, and (2) mean numbers of different types of word per case. The difference of the first and second scores is in whether duplicated identical words were counted or not. The first score indicates the total amount of protocols data, whereas the second score indicates the variety of protocols data. We compared the experts with the novices in each of the scores for each of the tag types.

In all of the tag types except *region*, the number of words in the experts was higher than that of the novices [*percept*, $t(12) = 1.96, p < .05$; *concept*, $t(12) = 2.61, p < .01$; *region*, $t(12) = 1.55, ns.$,; *goal*, $t(12) = 3.51, p < .01$]. There are also significant differences among the types of word in all of the tag types [*percept*, $t(12) = 7.52, p < .01$; *concept*, $t(12) = 8.94, p < .01$; *region*, $t(12) = 2.96, p < .05$; *goal*, $t(12) = 5.11, p < .01$].

Roughly speaking, the above results are consistent with the previous findings of X-ray film diagnosis. As noted earlier, Lesgold et al (1988) confirmed that experts verbalized more types of finding and more types of hypothesis than novices. Also, Rufaste, Eyrolle, & Marine (1996) revealed that semantic networks constructed from experts' verbalizations were richer than those from novices' verbalizations.

(b) Tagged words in the earlier and later diagnostic processes Our main area of interest was to understand the nature of the interactions between perceptual and conceptual processing. Therefore, we investigated the transition of words from the earlier to later phases of diagnostic process. That is, each word sequence was divided at the center of the process, and we counted words appearing in the earlier and later phases. The result is shown in Figure 1. In this analysis, we counted only words tagged as *percept* and *concept*, disregarding words tagged as *region* and *goal*.

We conducted two 2 × 2 expertise (within) × phases (within) one-way analyses of variance with each of *percept* and *concept* as dependent measures. The ANOVA with *percept* revealed that words tagged as *percept* significantly decreased from the earlier to later phases regardless of expertise [a main effect of the phases, $F(1, 12) = 9.43, p < .01$; a main effect of the expertise, $F(1, 12) = 1.74, ns.$; an interaction between the expertise and the phases, $F(1, 12) = 0.09, ns.$,]. On the other hand, the ANOVA with *concept* revealed not only a significant main effect of the phases [$F(1, 12) = 14.42, p < .01$], but also a significant main effect of the expertise [$F(1, 12) = 7.37, p < .05$], and a significant interaction between the expertise and the phases [$F(1, 12) = 6.80, p < .05$]. As results of analyses of simple main effects, it was confirmed that the number of words tagged as *concept* increased from the earlier to later phases, both among the experts [$F(1, 12) = 8.25, p < .05$] and for the novices [$F(1, 12) = 18.79$]. Furthermore, it was also confirmed that the experts more frequently reported words tagged as *concept* than the novices in the earlier phase [$F(1, 12) = 20.31, p < .01$.], but not in the later phase [$F(1, 12) = 1.09, ns.$].

The above results suggest that the process of medical image diagnosis is basically a serial bottom-up process, which is consistent with the *traditional view of medical image diag-*

[1] The first analysis was based on all of the protocols data including the verbalizations in the interviews, whereas the second and third analyses were based on the protocols data in the main task excluding the verbalizations in the interviews.

Table 4: Overall patterns of tagged words [Mean (SD)]

		Percepts	Concepts	Regions	Goals
Novices	Words	61.89 (15.75)	24.21 (6.93)	41.98 (8.29)	11.32 (2.07)
	Types	26.78 (5.19)	9.80 (2.00)	21.06 (3.50)	4.29 (0.62)
Experts	Words	71.69 (16.12)	31.29 (9.29)	45.69 (8.55)	14.38 (3.06)
	Types	34.16 (6.44)	12.41 (2.54)	23.56 (5.02)	5.04 (1.01)

nosis noted earlier (as a decrease of *percept* and an increase of *concept* from the earlier to later phases). However, the significant interaction in *concept*, which indicates that the experts frequently reported words tagged as *concept* in the earlier phase, suggests that the development of expertise changes the methods of interactions of the two components.

(c) Cycles of perceptual and conceptual processing To investigate the interactive process of medical image diagnosis more directly, we analyzed local transition patterns of word sequences. That is, we distinguished segments concerning perceptual processing from segments concerning conceptual processing. We assumed the earlier segments as continuous appearances of *percept* (e.g., Example 1 of Table 2), and the latter as continuous appearances of *concept* (Example 2 of Table 2). *Region* and *goal* were disregarded in this analysis because these two tag types seemed to relate to both perceptual and conceptual processing. Following this, we counted the number of transitions from the percept segments to the concept segments, which roughly indicates the frequency of cycles between perceptual and conceptual processing. The number of cycles, C, was counted using the following procedure.

If n's tag value $= n - 1$*'s tag value*
  Then $n = n + 1$
  Else
    If n's tag value = "percept" or "concept"
      Then{
        back to previous "percept" or "concept"
        If n' tag value = previous tag value
          Then $n = n + 1$
          Else $\{C = C + 1, n = n + 1\}\}$
      Else $n = n + 1$

Table 5 shows the result of this analysis. *Number* indicates a mean number of cycles per case, while *proportion* indicates a mean proportion of the number of cycles to the total amount of words per case. Therefore, *proportion* indicates how long a particular segment type continued until the other type of segment appeared. Two dependent group t tests revealed that the cycles of the expert were more frequent and more rapid than those of the novices [number of cycles, $t(12) = 3.39, p < .01$; proportion of cycles, $t(12) = 2.83, p < .05$]. The results suggest that the development of expertise changes how two components interact with each other. That is, the experts switched the two components more frequently, whereas the novices continued the sequence longer in each component.

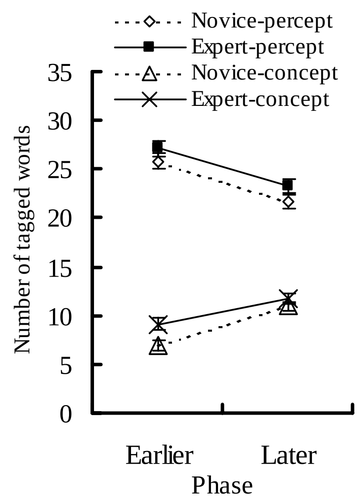

Figure 1: Tagged words in the earlier and later diagnostic processes. *Note.* Error bars represent one standard error of mean.

Table 5: Cycles [Mean (SD)]

	Number	Proportion
Novices	15.73 (6.03)	0.150 (0.04)
Experts	19.33 (7.09)	0.177 (0.05)

General Discussion

In this study, we investigated the development of expertise in 3D-CT image diagnosis. Roughly speaking, our results are consistent with the previous findings of X-ray film diagnosis (Lesgold et. al., 1988; Raufaste, Eyrolle, & Mariné, 1998) in that the experts exhibited superior diagnostic performance. The results of protocols analysis also demonstrated that the development of expertise changes both perceptual and conceptual processing.

Furthermore, our study revealed effects of expertise on the interactions between the two components, which had been unclear in the previous studies. Overall, both the experts and the novices made their entire diagnoses in the bottom-up manner; however, the experts engaged in conceptual processing in the earlier stages. They produced more frequent cycles of the two components. Although further investigation is required, the clear results presented in this paper have a strong implication for understanding the nature of expertise in medical image diagnosis.

We believe that our results have implications for the general goal of cognitive science studies, such as *deep understanding of human cognition*. Through evolutions of cognitive science, the interactions between perception (external world, or data) and concept (schema, or hypothesis) have been frequently discussed (e.g., Simon, and Lea, 1974; Neisser, 1978; Goldstone, Steyvers, Spencer-Smith, and Kersten, 2000). In particular, we consider that the process presented in this paper is similar to those of design and scientific discovery. For example, the design studies revealed that "novel perception and novel concept are generated through the interactive process of perceptual and conceptual activities" (Suwa, Gero, and Purcell, 2000). Additionally, beginning with the dual-space search model, proposed by Simon, and Lea (1974), the studies of scientific discovery intensively explored the interactions of data search and hypothesis generation, which provide opportunities of historical discoveries. Furthermore, we speculate that hypothesis testing strategies such as the negative test strategy are important skills for medical image diagnosis. In fact, one of the experts gave a verbalization that suggests the use of negative-tests such as "because the shadow apparently looked malignant, I searched features that belonged to benign cases" (similar findings in Pani, Chariker, and Fell, 2004).

Additionally, we consider that our study has a methodological implication in protocols analysis studies. In this study, the morphological analysis tool made it possible to analyze large amounts of protocols data, and successfully quantified the interactions between perceptual and conceptual processing. Although there are some limitations in our methods, we believe that further elaborations of the methods will make it possible to understand the details of the cognitive process.

Finally we would like to assert contributions of our study from the viewpoint of cognitive engineering. This study is part of a larger project that is being conducted in collaboration with researchers of image-processing engineering and radiologists, with the aim of developing intelligent systems that support the diagnostic process (details in Morita, Miwa, Kitasaka, Mori, Suenaga, Iwano, Ikeda, and Ishigaki, 2004). So far, image-processing engineering has developed elaborated tools that mainly support physicians' perceptual processing (e.g., Mori, Hasegawa, Suenaga, & Toriwaki, 2000). We believe that the combination of image-processing engineering and cognitive scientific analysis will make possible to create innovative tools for supporting the interactive process in medical image diagnosis.

Acknowledgments

This work was supported in part by the Grant-In-Aid for 21st COE program provided by the Ministry of Education, Culture, Sports, Science and Technology of the Japanese government. The authors thank the physicians of Nagoya University Hospital for participating in the experiment.

References

Azevedo, R, & Lajoie, S. P., (1998). The cognitive basis for the design of a mammography interpretation tutor. *International journal of artificial intelligence in education*, *9*, 32–44.

Goldstone, R.L., Steyvers, M., Spencer-Smith, J., & Kersten, A., (2000). Interactions between Perceptual and Conceptual Learning, In E. Diettrich & A.B.Markman(Eds.) Cognitive Dynamics: Conceptual Change in Humans and Machines., 191–228.

Lesgold, A., Rubinson, H., Feltovitch, P., Glaser, R., Klopfer, D., & Wang, Y., (1988). Expertise in a complex skill: Diagnosing X-ray pictures. In M. Chi, R. Glaser, & M. Farr, M. (Eds.), *The nature of expertise*, Erlbaum, Hillsdale, NJ. 311–342.

Krupinski, E., (2003). The future of image perception in radiology: Synergy between humans and computers. *Academic radiology*, *10*, 1–3.

Mori, K., Hasegawa, J., Suenaga, Y., & Toriwaki, J., (2000). Automated anatomical labeling of the bronchial branch and its application to the virtual bronchoscopy system. *IEEE Transactions on Medical Imaging*. *19*, 103–114.

Morita, J., Miwa, K. Kitasaka, T., Mori, K., Suenaga, Y., Iwano, S., Ikeda, M., & Ishigaki, T., (2004). Chance Discovery in Image Diagnosis: Analysis of Perceptual Cycles. *In proceedings of the first European Workshop on Chance Discovery, 16th European Conference on Artificial Intelligence*. 162–171.

Myles-Worsley, M., Johnston, W. A., & Simons, M. A., (1988). The influence of expertise on X-ray image processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 553–557.

Neisser, U., (1978). Cognition and reality. W.H. Freeman and Company.

Norman, G. R., Brooks, L. R., Coblentz, C. L., & Babcook, C. J., (1992). The correlation of feature identification and category judgments in diagnostic radiology. *Memory & Cognition*, *20*, 344–355.

Pani, J. R., Chariker, J. H., & Fell, R. D., (2004). Toward a theory of qualitative vidual reasoning in microanatomy, *In proceedings of the eighteenth International workshop on qualitative reasoning*.

Raufaste, E., Eyrolle, H., & Mariné, C., (1998). Pertinence generation in radiological diagnosis: Spreading activation and the nature of expertise. *Cognitive Science*, *22*, 517–546.

Rogers, E., (1996). A Study of Visual Reasoning in Medical Diagnosis. *In proceedings of the eiteenth annual conference of the cognitive science society*, Lawrence Erlbaum: Mahwah, NJ, 213–218.

Suwa, M., Gero, J. & Purcell, T. (2000) Unexpected discoveries and S-invention of design requirements: Important vehicles for a design process. *Design Studies*, *21*, 539–567.

Simon, H. A., & Lea, G., (1974). Problem solving and rule induction: unified view. In L. Gregg, (eds.), *Knowledge and Cognition*, Erlbaum, Hillsdale, NJ. 105–127.

Woods, B. P., (1999) Visual Expertise. *Radiology*, *211*, 1–3.