# 7 Attention Maps—Scientific Tools or Fancy Visualizations?

In the previous two chapters we have explained how to use data samples to estimate oculomotor events (Chapter 5) and how to divide stimulus space into areas of interests (AOIs) to detect further events and define complex representations (Chapter 6). In this chapter, we introduce a representation of eye-tracking data called *attention maps*, which is very often visualized in the form of a *heat map*, but which also has mathematical forms that can be used for quantitative testing.

Although it does not represent attention per se but the spatial distribution of eye-movement data, we consider the term 'attention map' an appropriate homage to the two papers that first introduced them as a general representation form for eye-tracking data: Pomplun *et al.* (1996) ('Attentional landscape') and Wooding, Mugglestone, Purdy, and Gale (2002) ('Fixation map').

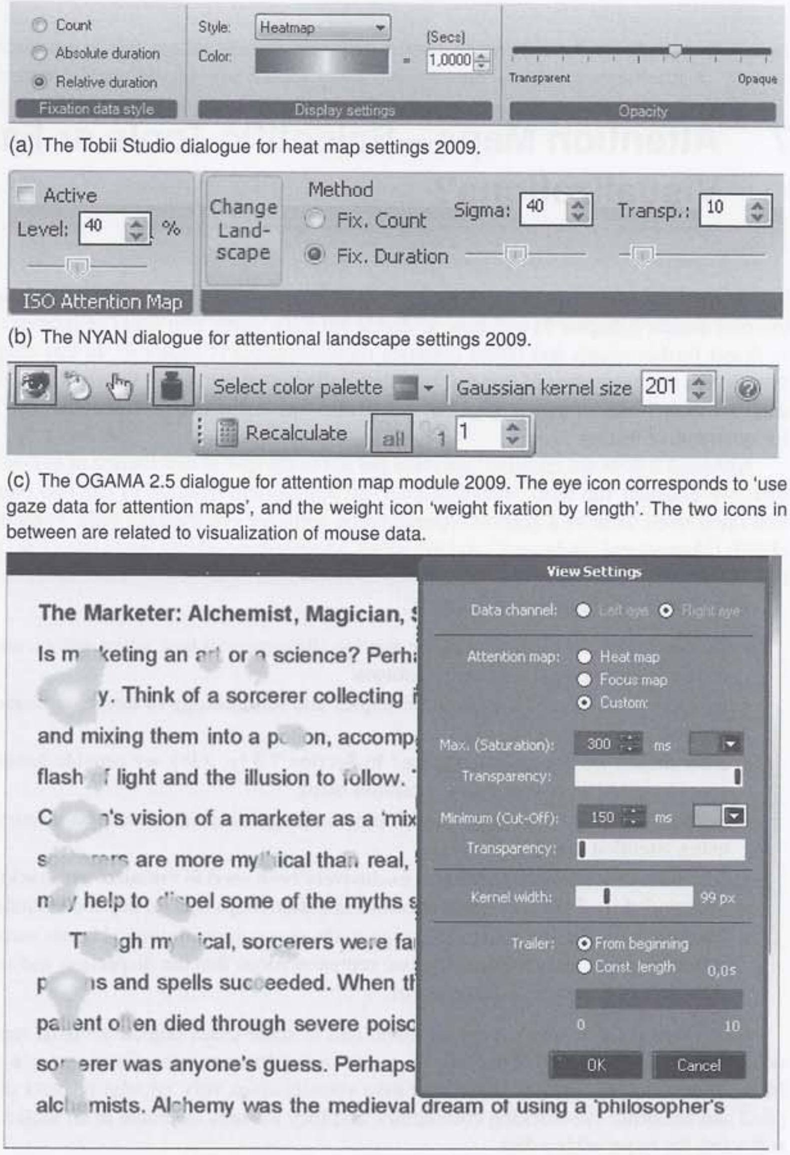
This chapter is organized as follows:

- Section 7.1 (p. 231) introduces the settings dialogue you face when using a number of common commercial software packages.
- In Section 7.2 (p. 233), general principles and terminology of attention maps are described.
- How should we use attention maps? In Section 7.3 (p. 238), we provide hands-on advice to people who want to use attention maps.
- Section 7.4 (p. 239) presents issues and challenges that one should be aware of when using attention maps.
- Attention maps have so far almost exclusively been used to visualize eye-tracking data. Section 7.5 (p. 248) lists applications of attention maps that go beyond visualization.
- The summary Section 7.6 (p. 252) repeats the major methodological issues surrounding attention maps, and reiterates the two representations that the dispersion and similarity measures in Chapter 11 make use of.

Heat maps provide quick, very intuitive, and in some cases objective visual representations of eye-tracking data that naive users and even children can immediately grasp a meaning from. Their intuitiveness has made heat map visualizations very popular in parts of the applied and scientific eye-tracking community, and they are now available in all major analysis softwares for eye-tracking data.

## 7.1 Heat map settings dialogues

When a researcher wants to generate heat maps from recorded data, but not implement the software for doing it, she is faced with software developed by the eye-tracker manufacturer or by independent companies. It is remarkably easy; just click a button and you will immediately see the defaulted heat map visualization. Typically, regions with many fixations or data samples are highlighted with warm colours (red) and regions where few or no people looked



○ Count
○ Absolute duration
● Relative duration
Style: Heatmap
Color: = 1,0000 (Secs)

(a) The Tobii Studio dialogue for heat map settings 2009.

□ Active
Level: 40 %
Method
○ Fix. Count
● Fix. Duration
Sigma: 40
Transp.: 10

(b) The NYAN dialogue for attentional landscape settings 2009.

Gaussian kernel size 201
1

(c) The OGAMA 2.5 dialogue for attention map module 2009. The eye icon corresponds to 'use gaze data for attention maps', and the weight icon 'weight fixation by length'. The two icons in between are related to visualization of mouse data.

The Marketer: Alchemist, Magician,
Is m keting an a or a science? Perh
y. Think of a sorcerer collecting
and mixing them into a p on, accomp
flash f light and the illusion to follow.
C 's vision of a marketer as a 'mix
s rers are more my ical than real,
m y help to spel some of the myths
T gh my ical, sorcerers were fa
p s and spells suc eeded. When th
pa ent o en died through severe poiso
so erer was anyone's guess. Perhaps
alc mists. Alchemy was the medieval dream of using a 'philosopher's

View Settings
Data channel: ● Left eye ● Right eye
Attention map: ● Heat map ● Focus map ● Custom
Max. (Saturation): 300 ms
Transparency:
Minimum (Cut-Off): 150 ms
Transparency:
Kernel width: 99 px
Trailer: ● From beginning ● Const. length 0,0s

(d) The SMI BeGaze dialogue for attention map settings on top of heat map visualization 2009.

**Fig. 7.1** Examples of settings dialogues for attention maps.

are marked with a colder (blue) colour. In the software, there are settings that allow you to vary the appearance of the visualization, and Figures 7.1(a) to 7.1(d) show dialogue boxes from four softwares: Tobii studio, NYAN, OGAMA (Vosskuhler, Nordmeier, Kuchinke, & Jacobs, 2008), and SMI BeGaze.

While the heat maps themselves look very simple and intuitive, it is not always obvious what settings to use in the dialogue boxes. For instance, what is 'Kernel width', what does the size of that kernel have to do with the visualization produced? Is 'Sigma' the same, or something wholly different? What does it mean that a certain colour equals the time of 1 second? What are 'Fixation data styles', and why is it necessary to choose between them? Can the user accept the default settings, whatever they are, or is it a better approach to move the values up and down until the heat map looks good? Is there such a thing as correct settings at all?

## 7.2 Principles and terminology

To understand what the settings in the dialogue boxes are, and why they produce the results they do, we have to understand how attention maps are generated. Two seemingly different but closely related principles have been used to achieve this: gridded AOIs and topological (Gaussian) landscapes.

The gridded AOIs (p. 212) divide space into a set of rectangular AOI cells. When filled with dwell time values, the gridded AOI becomes a dwell map. Although the dwell map as such does not qualify as an attention map visualization, it would if the numbers were converted into colour or intensity.

Gridded AOIs are versatile tools for generating a variety of different attention maps and interesting visualizations to them. To make a *dwell map*, for instance, for each cell (each AOI) in the grid, the total dwell time (p. 389) is calculated, and this value determines the colour of that cell. Figure 7.2 shows such a heat visualization of a dwell map constructed from data recorded from newspaper readers. There are clear hot-spots stretching over several of the cells, indicating lots of gaze at that area. There are also parts that are much cooler (less looked at).

While it is most common to fill the cells (AOIs) with the values for dwell time or number of fixations, gridded AOIs offer several other possibilities to build attention maps. For instance, if we were interested in a map that shows on average *how early* different parts of a scene have been visited, an *entry time map*, we could use the AOI-measure *entry time* (p. 437) as the height of each of the cells in the grid. The map would then show red where participants look early on average, and blue where they arrive late in the trial.

We could also fill the AOI cells with the measure *proportion of participants* (p. 419): for each cell in the grid, the proportion of participants having fixated that cell is calculated. 15% in the cool areas, and perhaps 97.5% in the hottest peak. Heat maps produced by the Eye-tools company, a commercial player using eye tracking to test design, provide this type of heat map (Hernandez, 2007). It has to be properly understood, however: a 'proportion-of-participants map' shows what areas have been looked at by how large a proportion of the participants, but participants who spend much of their time in only one or two parts of the stimuli will make a very small contribution to the map compared to participants who visit many grid cells with only a few fixations in each.

Despite their versatility, gridded AOIs have not become the dominating principle for visualizing heat maps. In fact, none of the settings in Figure 7.1 refer to gridded AOIs. Instead of a grid with sharp borders between cells, smoother, 'landscape-like' representations have become more prominent.

As Figure 7.3(b) illustrates, a heat map looks just as flat as fixations plotted on your stimulus, but the underlying attention map represents a smooth landscape with hills and valleys. An attention map landscape is successively built from the sequence of fixations or raw data, simply by putting a basic form at each point in space where the fixation or raw data sample is

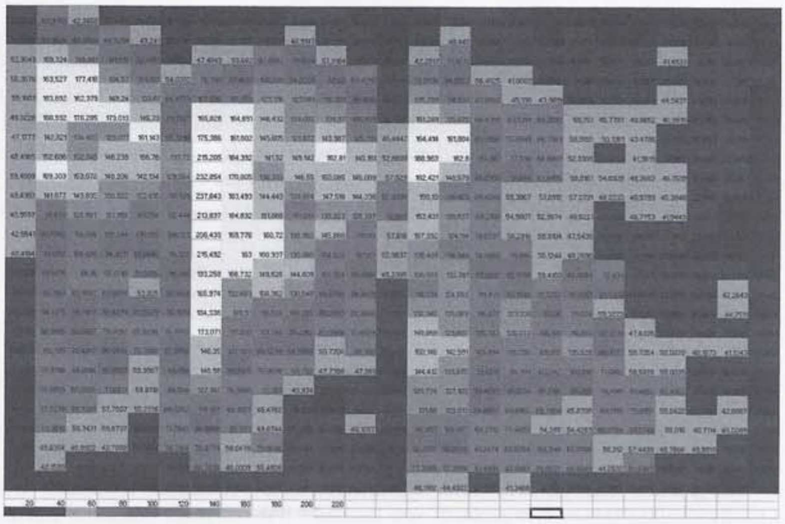
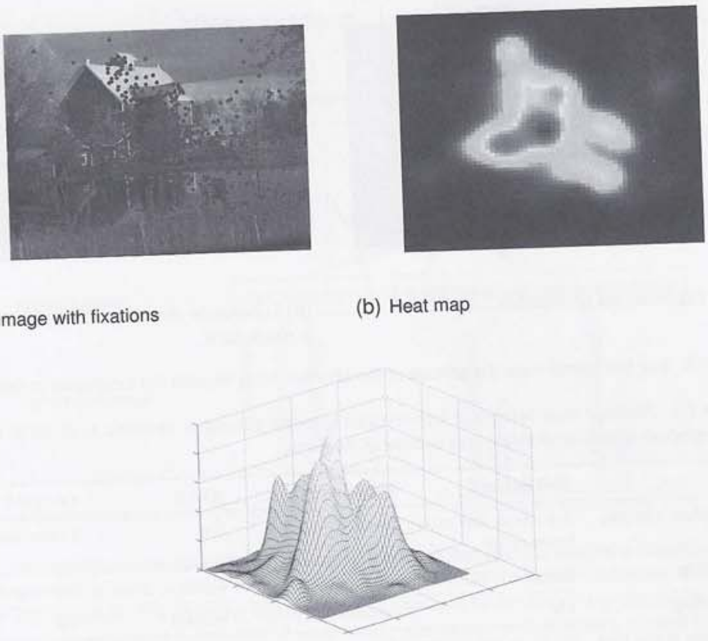


**Fig. 7.2** Gridded AOI heat map. 40 readers of a tabloid newspaper; all participants and all 20 folds. Numbers in grid cells are total dwell time values. Recorded using a head-mounted eye-tracker with Polhemus head tracking at 50 Hz.

located. Various basic forms are being used: Pomplun *et al.* (1996) and Wooding *et al.* (2002) use mathematical constructs called Gaussian functions, while most manufacturer implementations use cone-formed constructs which approximate the top of a Gaussian function. Given that a sufficient number of forms have been dropped on your stimulus, the final maps will look more or less the same, even though the shape of the basic construct varies somewhat (a consequence of the central limit theorem).

For gridded AOI representations, heat map visualizations can easily be made by choosing what colour is attributed to what height values, but borders between cells will be sharp. In the landscape-like representation, smooth colouring is possible. First, we must recognize that in the commercial software, the altitude in an attention map is mostly time, in which case the altitude unit is milliseconds (ms). For example, the location of a fixation can be assigned an altitude value of its duration, say 325 ms, which decides how high the attention map will be in this point. In Figure 7.4, a single fixation hill is shown from the side, with two thresholds on the right side. Suppose we set the lower threshold to 150 ms and blue, and the upper to 300 ms and red. Then, for all altitudes from 150 to 300 ms, the attention map is coloured with hues ranging from blue to red. Using SMI's software package BeGaze 2.2 with similar settings, Figure 7.1(d) illustrates what this can look like. The resulting heat map has hot-spot areas with very distinct borders, because all the areas where the attention map is lower than 150 ms have been made transparent. In the interval between 150 and 300 ms, the colour intensifies, but above 300 ms, the colour remains constant.

Colour (in a heat map), luminance, and transparency (used in Koivunen, Kukkonen, Lahtinen, Rantala, & Sharmin, 2004) can all be mapped to the altitude in the attention map. For instance, with luminance mapping, low altitudes become dark and high become whatever the original picture was at that position. Figure 7.5(b) shows this type of mapping, which we call a *luminance map*. Yet another possibility would be to map the altitude to an image's contrast, to have an image which is sharp and in focus at high altitudes and increasingly more blurred


(a) Image with fixations

(b) Heat map

(c) Attention map

Fig. 7.3 Three different ways to visualize the same eye-tracking data. The fixations are recorded from 12 people viewing the image for five seconds.

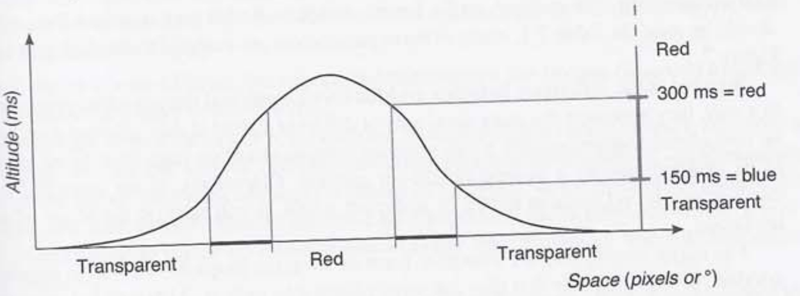
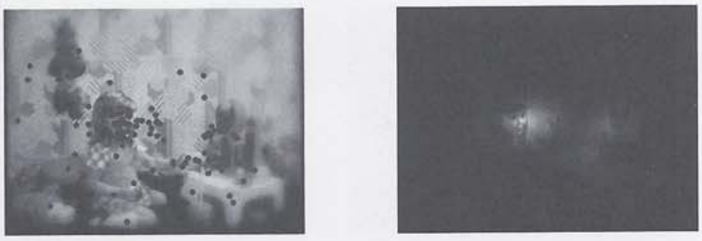


Fig. 7.4 A single Gaussian seen from the side at a point where a 325 ms fixation has landed. The figure shows how a gradual colour dimension is mapped onto altitude so as to generate a 2D heat map visualization.

as we move down the hills into the areas not looked at. Such techniques are explored, e.g. by Geisler and Perry (1998) and Nyström (2008) for the purpose of improved video compression.

There are basically four parameters that decide what your Gaussian attention map will look like. One decides the *mapping from colour to altitude*, another controls the *width of the basic construct*, a third gives you an option to build the attention map with or without taking



(a) Fixations on stimulus

(b) Luminance map; low altitude (few fixations) is made dark.

**Fig. 7.5** In a luminance map, the altitude of the attention map decides the *luminance* of the visualization.

**Table 7.1** Attention map settings in four analysis software packages; versions as of 2009. Fillings are as yet restricted to fixation duration and number of fixations.

| | SMI BeGaze | Tobii Studio | NYAN | OGAMA |
|---|---|---|---|---|
| Colour mapping | Saturation, and minimum cutoff | colour = time | n/a | Colour palette |
| Width | Kernel width | n/a | Sigma | Gaussian kernel size |
| Filling | n/a | Fixation data style | Method | Yes |
| Does it use both fixations and raw samples? | Fixations are always used for images, raw samples always for video. | Yes | Fixations are always used. | Fixations are always used. |

fixation *duration* into account, and a fourth decides whether raw data samples or fixations should be used. In Table 7.1, some of these parameters are mapped to the dialogue settings in Figure 7.1.

So what is the difference between gridded dwell maps and the smoother representations? In a way, they represent the same data, only at different spatial scales; gridded dwell maps can be considered mathematically a sub-sampled smooth attention map after being interpolated to its original size by a nearest neighbour method. Conversely, if we convolve a gridded dwell map with a Gaussian kernel of sufficient width, it will take on the shape of a smooth landscape.

The major reason that the smoother form of attention maps have come to dominate over gridded AOIs is probably that they are more pleasing to look at. Although being aesthetically pleasing is a seemingly superficial reason, the advent of smooth-looking heat maps helped to radically expand the eye-tracker market in favour of the manufacturer who could first deliver them as part of their software.[24] Another reason could be that the gridded dwell maps provide only a coarse representation of the original data, since fixations or raw data samples are quantized to the grid elements, and thus contain less information about the original data than do Gaussian based attention maps. A less explored reason is that attention maps built from a

[24] Qualitative interviews that the authors have made with salespeople and applied practioners clearly show the craze for heat maps in the mid 2000s. Several salespeople report on customers asking only for the ability to produce heat maps before ordering. Applied practioners have told us how they have tried to sell quantitive AOI analysis to customers in the advertisment and web business, but that their customers returned their analysis and asked for heat maps.

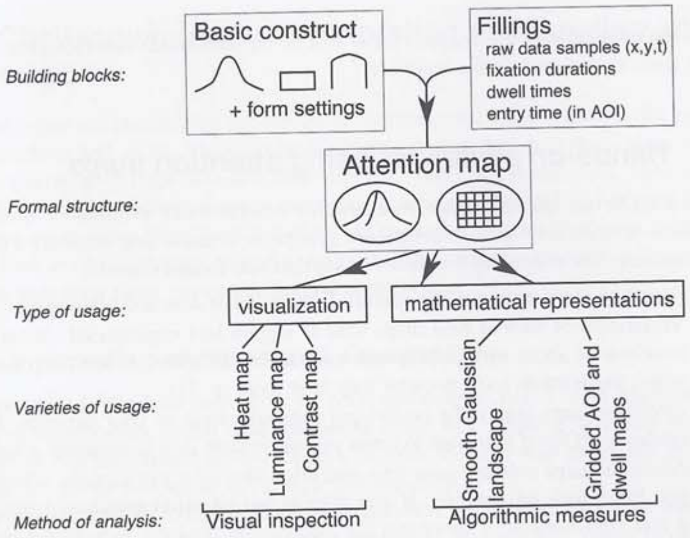


**Fig. 7.6** Attention maps are created using a *basic construct*, often (but not always) a Gaussian function, and a *filling* into the construct. The filling can be number of fixations, fixation durations, dwell time, or any other AOI measure. The resulting *attention map* is a smooth landscape or a gridded AOI partition of space, that can be *visualized* in a number of ways (heat maps being the best known), or used in *measures* of e.g. position dispersion and similarity.

Gaussian function can be analysed with a suite of powerful mathematical tools; however, in principle this is also true of gridded AOIs.

Figure 7.6 summarizes this introduction to the representations we call attention maps. It shows that attention maps:

- Consist of two different, but related, representations: the smooth *Gaussian landscapes*, and the hard-edge *gridded AOIs*.
- Are built from a variety of *basic constructs*, including the basic Gaussian function, the grid cell and the various cylinder-like constructs that manufacturer implementations make use of. One such construct is placed or 'dropped' at the position of each fixation or raw data sample. This is not the case for gridded AOIs however, where a grid is simply used to divide up stimulus-space, rather than 'dropping' constructs at specific positions.
- Have *settings* that pertain to the basic constructs rather than to the full and finished map itself. When changing construct settings such as number of cells (in a gridded AOI), 'Kernel width' (of a cylinder or cone) or 'sigma' (of a Gaussian function), the full map only changes because the single basic constructs change.
- Have *fillings*, which are other eye-tracking measures, such as fixation durations or dwell time, that fill the basic constructs and decide their height. More of the measure means a higher basic construct. The full attention map gets its shape from the additive effect when many basic constructs of varying heights pile up.
- Appear in *visualizations* as heat maps, luminance maps or contrast maps, and in *measures* as their basic mathematical forms, and in normalized versions called probability density functions (pdf:s).

In the remainder of the chapter, focus will be on the smooth representations of attention maps, since they dominate the literature.

## 7.3 Hands-on advice for using attention maps

If you plan to use attention maps to exemplify results, make exploratory investigations of your data, perform statistical tests between groups, or because heat maps are a deliverable to your customer, the following list shows issues that you should consider.

- Attention maps represent the *spatial distribution of data* and nothing else.
- Visualizations such as heat maps tend to inspire less experienced viewers to jump to conclusions about *why* participants look at the hot spots. A heat map can only show where participants look, not why they look there (p. 71).
- Attention maps ignore the underlying semantic areas of your stimulus. Use areas of interests (AOIs) if you want to relate eye-movement data to semantic areas.
- Attention maps collapse over time and often also over participants, which means that you lose much information. If you plan to use attention map-based statistics, see to it that your hypothesis compares the information retained, namely the overall spatial distribution.
- Visualizations such as heat maps can exemplify, support, and even nuance quantitative results, but should be published on their own only after careful consideration.
- There are virtually no guidelines or systematic investigations of the effects of settings for the colour mapping. A $\sigma$ value of around 2° visual angle (diameter) will give an indication of what is looked at with the point of highest visual acuity—the fovea, but note that this can be misleading because items can still be attended in peripheral vision outside of this area. Blignaut (2010) has recently addressed this issue by incorporating perceptual span into attention map settings.
- Do not make a habit of changing values up and down between heat map visualizations of your different groups, participants, or conditions. If you do, your visualizations will not be comparable. In fact, you will be producing art rather than data visualizations. Select one setting and stick to it throughout all the heat maps you make, so you know that any difference that you see in the heat map is actually an effect in your data and not in your settings.
- When you publish your attention map visualization, always report type of eye-tracker and its sampling frequency, analysis software version, settings for the basic constructs, settings for the mapping of colour, luminance or contrast to the height of the map, the time segment that the visualization was created from, whether you use fixations rather than raw data samples, and the criteria for fixation detection.
- Make sure to have sufficient amounts of data for your visualizations.
- If you have data from a low-speed eye-tracker with not so good precision, use fixations for building your attention maps. If you have a high-end eye-tracker, you can also use raw data.
- If you are having participants look at a central fixation cross before stimuli onset, make sure to remove the first fixation in your recordings. Otherwise your attention map may have an artificially large hot-spot in the centre.
- Attention maps are much more versatile than generally believed, and can also be used scientifically. Figure 7.6 illustrates the range of possibilities.

Addressing applied users of eye tracking, Bojko (2009) lists similar but not identical advice.

## 7.4 Challenging issues: interpreting and building attention maps

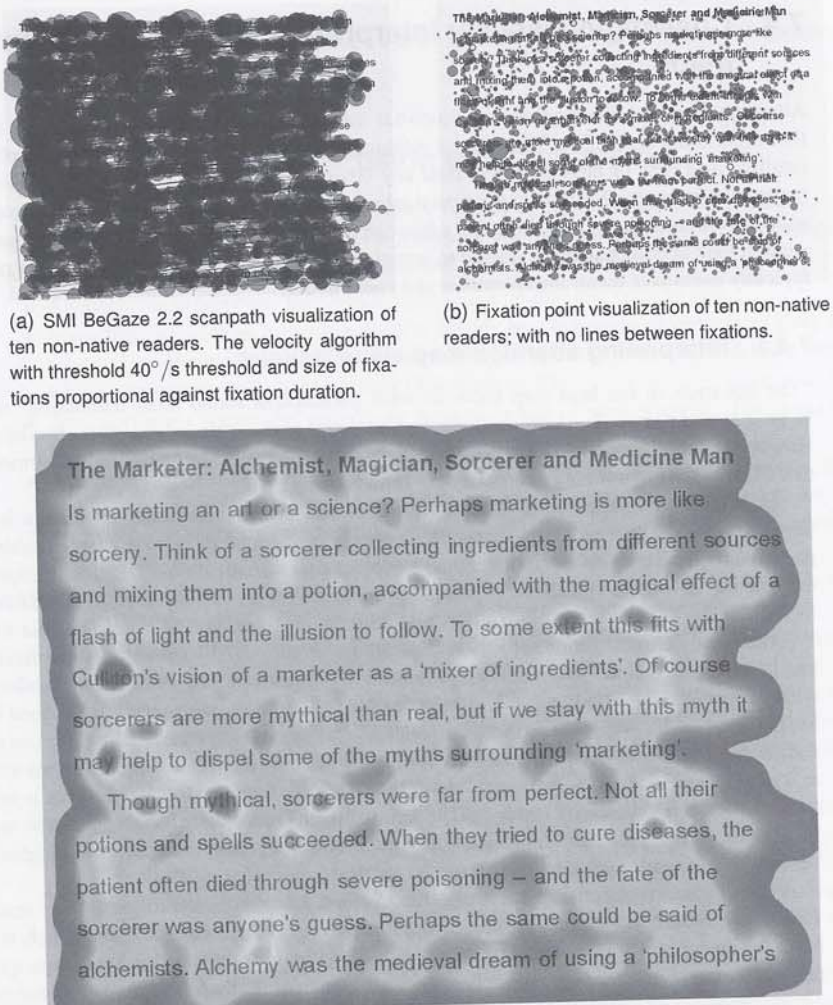
Attention maps are challenging for somewhat different reasons compared with event detection algorithms and AOIs. Their major advantage is that they provide a more simple and intuitive overview of large data sets than any other form of visualization. This is also their major disadvantage: the heat maps in particular look so simple that it is tempting to immediately draw conclusions from them that often can not and should not be drawn. In this section, we will look at two challenges related to attention maps, and present them in the order people typically encounter them: interpretation and construction.

### 7.4.1 Interpreting attention map visualizations

"The hot-spots in the heat map show us what participants found most interesting" seems to be a straightforward, sound observation one could make from a heat map visualization. However, is this assumption always correct, or does the heat map really tell us something else? Let us start by looking at an example from reading.

Notwithstanding the complications with attention maps, their major advantage is that they very quickly give an easily digestible overview of the total data from a large number of participants. This they do much better than any other data visualizations. Take for example the ten non-native readers visualized as scanpaths in Figure 7.7(a). Such a scanpath visualization looks like a complete mess of fixations and connection lines, and it is not possible to see if some words attracted more attention than others. Figure 7.7(b) shows only the fixations and no connection lines. The circles representing fixations have also been made smaller and with thinner lines. It is difficult but not impossible to see that some words have indeed been looked at more than others. The words 'magician' and 'sorcerer' in the heading, for instance, and several of the first words in each line. We cannot see how these individual fixations add up to a total group gaze on the words, however, nor compare between the words. This is where the attention map shows its strength. The heat map in Figure 7.7(c) shows that the words 'Culliton's', 'sorcerers', and 'may' have attracted more total gaze than the words higher up, but also that 'extent' and 'this fits' appear under the highest peaks.

Now, reading researchers have altogether avoided using attention maps in their studies. This is not only because attention maps ignore the important temporal order which is the basis for detecting for instance regressions. Nor is the major problem that heat maps ignore the small effects that reading researchers typically look for. Rather, heat maps and the other visualizations of attention maps do not provide any method for systematic and statistical comparison between conditions. How should we reason over a heat-map visualization? Suppose that we are linguists pondering over the heat map in Figure 7.7(c). We can see that certain words are hot, and, since we are curious about the result, immediately try to look for explanations: 'Culliton's' could be an unknown proper name, and 'mythical', 'sorcerers', and 'diseases' are infrequent words that the readers had to take some time to look at before they could understand them. And the hot-spot on 'this fits' must have something to do with the reference being difficult to understand, must it not? 'Though mythical' is perhaps also a difficult sentence start for these non-native readers? Throughout the data collection phase, we have looked forward to finding effects in the data of difficulties in understanding the text, and now we see them in the red-hot areas of the heat map. The hot-spots become confirmatory examples for our first explanation of whatever word or construction happens to be underneath it, making this study at best *exploratory* or at worst a confusing *fishing trip* (p. 66). We fail to look at the cool blue areas where less reader attention has been spent, and where similar references exist and where infrequent words are hidden counterexamples. The heat map sim-



(a) SMI BeGaze 2.2 scanpath visualization of ten non-native readers. The velocity algorithm with threshold 40°/s threshold and size of fixations proportional against fixation duration.

(b) Fixation point visualization of ten non-native readers; with no lines between fixations.

The Marketer: Alchemist, Magician, Sorcerer and Medicine Man

Is marketing an art or a science? Perhaps marketing is more like sorcery. Think of a sorcerer collecting ingredients from different sources and mixing them into a potion, accompanied with the magical effect of a flash of light and the illusion to follow. To some extent this fits with Culliton's vision of a marketer as a 'mixer of ingredients'. Of course sorcerers are more mythical than real, but if we stay with this myth it may help to dispel some of the myths surrounding 'marketing'.

Though mythical, sorcerers were far from perfect. Not all their potions and spells succeeded. When they tried to cure diseases, the patient often died through severe poisoning – and the fate of the sorcerer was anyone's guess. Perhaps the same could be said of alchemists. Alchemy was the medieval dream of using a 'philosopher's

(c) SMI BeGaze 2.2 heat map visualization of ten non-native readers. Kernel width 99 pixels, maximal saturation 300 ms, and minimum saturation 0 ms.

**Fig. 7.7** Heat maps give a good overview over data. Data from 10 non-native students of economy reading a management text, while being recorded with a towermounted high-end system.

ply does not invite systematic comparisons between all confirmatory and counter-examples, and so does not fit into the scientific method. Instead the heat map tends to invite to post-hoc interpretations that favour the researchers own hopes or favourite theory.

Are heat maps more helpful if we use them only to characterize the general viewing behaviour, rather than to explain it? In the eye tracking for web usability field, heat map visualizations are very commonly used to describe the general outcome of an eye tracking study. For example, Nielsen refers to the well-quoted so-called F-pattern (Nielsen, 2006): "Eyetracking visualizations show that users often read web pages in an F-shaped pattern: two

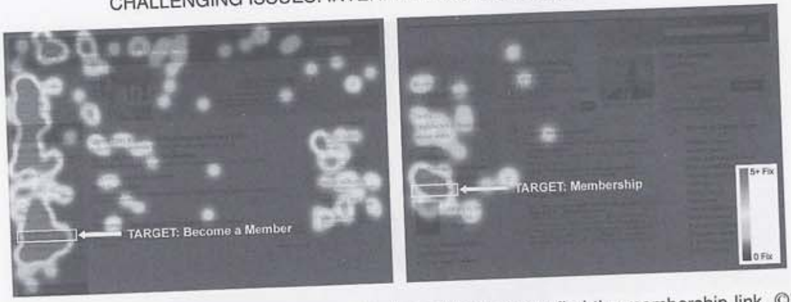


**Fig. 7.8** Two designs of the same web page. Participant task was to find the membership link. ©UPA, *Journal of Usability Studies,* Volume 1, Issue 3, p. 117; Bojko, A.A. (2006), Using eye tracking to compare web page designs: A case study.

horizontal stripes followed by a vertical stripe". Wulff (2007) uses heat maps to investigate how viewers explore a web page and concludes:

> Only the very first links are looked at and our test confirms—as do some Google tests—that the user does not read much on the page. The first 1-2 hits are read to the end. However, lower down in the list, it is only the first part of the line that is looked at. At the bottom we find only scattered glances. It is important that the most relevant hits are placed at the top of the page as is also the case in Google.

Similarly, Bojko (2006) compares two web page designs (Figure 7.8), using the heat map only to show that in the one case the gazes of participants were focused only on the task-relevant target, while in the other web page design, gazes were spread out. As long as the conclusion refers to *the viewing behaviour itself*, rather than *explanations to it*, the heat map is not tricking us. Compare this with the use of a heat map over a web page to conclude that: "Bottom half of page—ineffective line-height spacing and lack of white-space reduce reading. Most of the content is being missed".[25] Yes, the content is being missed by many; this a heat map can show. But to prove that it is missed *because of* ineffective line-height spacing rather then due to position, reader interest in the content, or long-term evolved reader expectations of where to look at web pages; that requires a much larger undertaking in terms of experimental design (p. 71).

In summary, the hot-spots in a heat map can point out the regions that attracted people's gazes, but it is highly speculative to draw any conclusions of what made people look there solely based on the heat map; it could be because these regions are interesting, but also because they are confusing, and people need to look twice to get the message.

### Attention maps and effect sizes

Figure 7.9 shows a much quoted teaser from a commercial company promoting their eye-tracking services. The heat maps appear to show that when the woman looks at the product, the participants—real customers according to the company web page—look towards the product too. We can easily compare the two conditions, and immediately see the difference in heat maps, can we not?

A particular uncertainty lingers on: how big is this difference, or *effect size*, really? The heat (i.e. the altitude in the attention maps) on the two product bottles may differ less than the heat map, with all its variable settings, suggests. The effect size in Figure 7.9 is much

[25] Citation from `http://blog.eyetools.com/2005/02/the-new-washington-post-homepage-design-an-eyetools-eyetracking-analysis.html`, retrieved Aug 31, 2010. The company owning the webpage is called eye-tools, and claim to be "The inventors of eyetracking heatmapping".

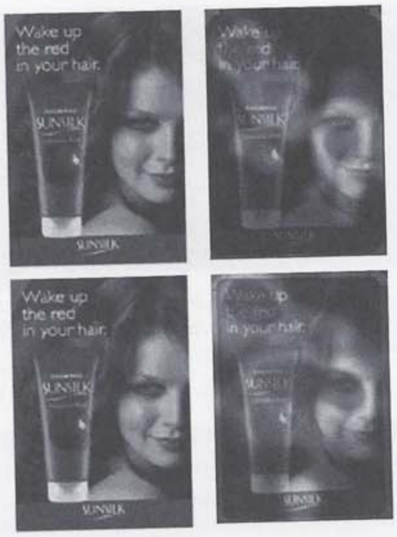


**Fig. 7.9** Is this heat map convincing evidence that gaze direction in a face of the picture alters viewer gaze towards the product? From `http://www.bunnyfoot.com/articles/not_focus_groups.htm`, with kind permission from Bunnyfoot and Think Eyetracking.

easier to calculate using AOIs and the total dwell time measure (p. 389). The product bottle in Figure 7.9 would make an excellent AOI, and the statistical test should be easy. If this marketing analysis company had wanted to impress us, why not present dwell time data in AOIs instead, or at least as a complement to the heat map? The reason could be in their customers; applied eye-tracking practitioners that we have interviewed claim that most of their customers strongly prefer heat map visualizations to conclusions based on statistics.

In other cases, positioning AOIs onto the stimulus image is not always possible. In an elegant example, Pomplun *et al.* (1996) used attention maps to alter the luminance of an image to promote a certain perceptual interpretation (we call this luminance-based attention map a *luminance map*). Using pictures with two ambiguous interpretations, eye movements were measured and participants were asked to press one of two different buttons to indicate which interpretation they currently perceived. Attention maps built on the collected eye movements belonging to each interpretation were used to modify the luminance such that frequently viewed regions were retained in high luminance, whereas other regions were reduced in luminance. Hence, two new versions of each stimulus image were generated, one for each subjective interpretation. Figure 7.10 exemplifies data for one stimulus image. The difference between the two luminance maps can quickly be seen in the figures, and we tend to easily accept that these images show that the subjective experience of an ambiguous picture is closely related to gaze position. However, on closer inspection, it is also obvious that there is a considerable overlap between the two luminance maps. In fact, the only part of the king in 7.10(b) that is not visible in the animal version 7.10(c), is the small area around the mouth. If we want to calculate effect sizes (how big the difference really is) using a dwell time measure, how do we position the AOIs? On the mouth, or on the king as a whole, or somewhere in between? Every position is a poor compromise. What we would want is a comparison that takes the attention map as a whole and compares it to another attention map as a whole.

Lacking a statistical method for comparing attentions maps, Pomplun *et al.* (1996) had

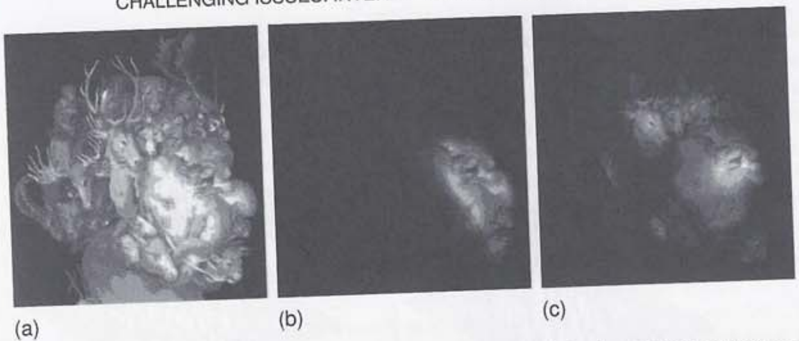


**Fig. 7.10** The painting 'Earth' by Giuseppe Arcimboldo. a: Original: b: luminance maps of data samples from participants who (clicked to) report that they see the king. c: luminance map of data from participants who report that they see the animals. Reprinted from *Perception, 25*(8), Pomplun M, Ritter H, Velichkovsky B, Disambiguating complex visual information: Towards communication of personal views of a scene, ©1996, pp. 931–948 with permission from Pion Limited, London.

other participants look at the luminance maps in Figures 7.10(b) and 7.10(c), and judge which interpretation they experienced in each. This difference proved to be strong. Such indirect tests are rarely applicable, however.

Bojko (2006) compares two web designs using heat maps, as a complement to the other two measures: entry time in number of fixations (defined on page 437), and total duration on the page. Participants had various tasks, for instance to find the 'How to become a member' link on the page. The purpose of using heat maps was to visualize the spread of search across the page up until the correct link was found. Figure 7.8 on page 241 shows heat maps that indicate a fairly clear difference between the two designs, but how do we quantify the difference? As the heat maps are used to illustrate the spread of search, AOI analysis could not easily do the job, but the relatively new dispersion and similarity measures in Chapter 11 would be able to quantify the deviation we see in the heat maps of Figure 7.8, adding an effect size to what is otherwise only a difference in visualizations.

There is one virtually never used benefit of attention maps, in relation to effect size measures with AOIs and other quantitative statistical tools: suppose that we use AOI measures, and get a significant effect that participants indeed look more at one part of the image than another, or more at an object in one image than in the other (as in Figure 7.9). Normally, a significant effect makes us trust that the effect is robust and worthy of publication. But a significant effect only shows that the difference is systematic and stable, not its magnitude. An attention map visualization such as a heat map allows us to estimate the magnitude of the effect in the context of the entire data: are the peaks in the selected areas really that big in comparison to the rest of the data?

### 7.4.2 How many fixations/participants?

As mentioned earlier in the chapter, attention maps are helpful to visualize large amounts of eye-tracking data. But when is it useful to switch from e.g. a scanpath view to a heat map? How many fixations/participants do we need in order to produce a robust visualization, i.e. one that does not change significantly if more data is added to it, or if the same amount of data from another group of viewers is used to produce it? Unfortunately, there is no single answer to these questions. Pernice and Nielsen (2009, p. 19) propose that usability studies should contain at least 30 participants if heat maps are "the main deliverables", and/or used in drawing conclusions. To motivate this heuristic, Pernice and Nielsen (2009) correlate a heat

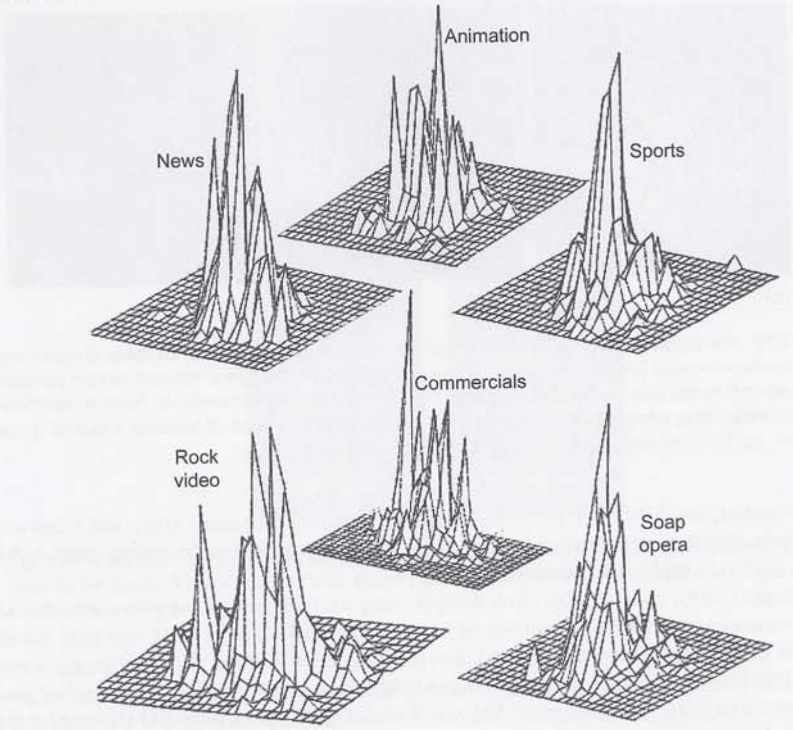


**Fig. 7.11** Attention map visualization showing the cumulative gaze distribution from three participants watching five videos. Reprinted with permission from Elias, Sherwin, and Wise (1984), ©1984, SMPTE.

map built from 60 viewers with another heat map built from a subset of these 60 viewers, and conclude that a subset of 30 people suffices to obtain a robust correlation between the maps. However, in the general case, the number of participants required closely depends on:

- your experimental hypothesis and the exact implementation of the heat map
- whether you have recorded a heterogeneous participant group
- the width of your basic construct (e.g. $\sigma$ in a Gaussian function)
- whether the task at hand leads to increased inter-participant gaze dispersion

## 7.4.3 How attention maps are built

Yet another challenge is to decide what raw material to build the attention map from, i.e. the 'basic construct'. As we saw earlier in the chapter, different constructs are used across analysis software. The good news is that if you have a large enough set of data, most attention map implementations will produce heat maps that look more or less the same.

Figure 7.11 shows what are perhaps the first visualizations of attention maps. The maps summarise the gaze distribution from three viewers watching video clips with different contents, and the authors conclude that the viewers seem to have a strong bias towards the centre of the display on which the videos were shown. However, the implementation is not described in the paper.

Another early usage of attention maps for scientific purposes, Pomplun *et al*. (1996), used Gaussian functions as basic constructs. Metaphorically, a Gaussian function is equivalent to a little hill of sand. Adding a new fixation to an attention map more or less equals pouring another bucket of sand over the stimulus picture. Gaussians are not said to be poured, however, but are dropped onto the stimulus image. As an alternative metaphor, the Gaussian can be seen as a rubber sheet, which with a given elasticity is being dropped on a pole located at the position of a fixation or raw sample point. The height of the pole would then correspond to the fixation duration.

The SMI BeGaze 2.1 implementation of attention maps first scales each image pixel in proportion to the durations of all fixations landing on it. Typically, this results in a very sparse 'fixation hit map', since only a small proportion of the pixels have been 'hit'. Then this 'hit map' is convolved with a Gaussian kernel with a certain width; a wider kernel gives a smoother, less pointy appearance to the attention map. The reason for dividing the attention map creation into two steps—one to produce the 'fixation hit map' and one to perform the convolution—is simply because the latter can be done directly on the computers graphics card, and therefore speeds up the software significantly and shows attention map visualization in real time, e.g. as video replay. The end result, however, is largely the same as using the Gaussian functions directly. The Tobii Studio software version of 2009 does not appear to have this setting, nor is it clear from the Tobii documentation of that time whether they use Gaussians or some other mathematical form. The heat map visualizations suggest that the Tobii Studio of 2009 uses cones with soft tips and constant widths.

In fact, using Gaussians is not the only way to mathematically represent the buckets of sand poured for each fixation. Wooding *et al*. (2002) proposes using *cylinders* to approximate the Gaussian. For small numbers of participants, cylinders give attention maps with very sharp edges, but when the number of participants is large, the many cylinders soften each other and the difference from a real Gaussian landscape will diminish. The Tobii cones of 2009 have softer edges, but do not gradually continue out into the periphery as does the Gaussian. Also, neither cylinders nor cones share the mathematical elegance of the Gaussian.

Formally, the Gaussian is defined as

$$G(x,y) = \exp\left(-\frac{(x-x_i)^2+(y-y_i)^2}{2\sigma^2}\right) \tag{7.1}$$

where $(x_i, y_i)$ is the centre point of the fixation or data sample at which we drop it. $(x, y)$ typically span the dimensions of the stimulus image. The overall attention map is generated by dropping one Gaussian function at each such centre point, and then summing all such dropped Gaussian functions into a single function.

There is only one important choice to make in the definition of the Gaussian, namely its *variance*, $\sigma^2$. In our sand metaphor, $\sigma^2$ is a measure of the sand's resistance to sliding. A small value on the variance $\sigma^2$ gives thin pointy Gaussian peaks, as though the sand resisted sliding to the sides and only formed a higher and higher peak at the very position where it was poured (Figure 7.12(a)). A high value on $\sigma^2$ makes the sand slide very much to the side, so that the Gaussians of different fixations completely merge and a map with a much smoother surface will be generated (Figure 7.12(b)). If the $\sigma^2$ is high enough, there is no resistance to sliding, and the Gaussians will be so wide and low that the attention map will look as though we were pouring water rather than sand: completely flat. If we look at the heat map visualization, a very low $\sigma^2$ setting will give many but narrow intense hot-spots with large cool areas between them. Very high values will yield a uniformly coloured heat map with one or two vague and very distributed hot-spots. Sigma corresponds to kernel width in manufacturer software.

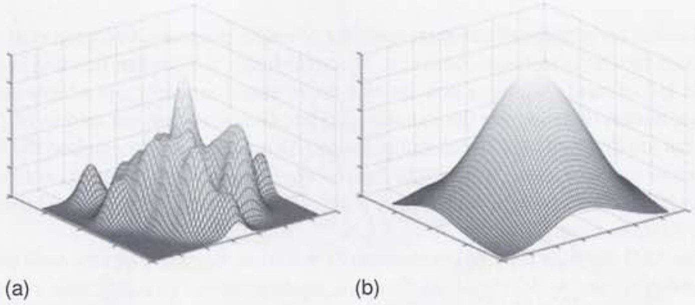


**Fig. 7.12** Pointy and smooth Gaussian-based attention maps generated from the same eye-tracking data, but with different $\sigma$.

So what is so great about the Gaussian function? There are at least three reasons why we should use them:

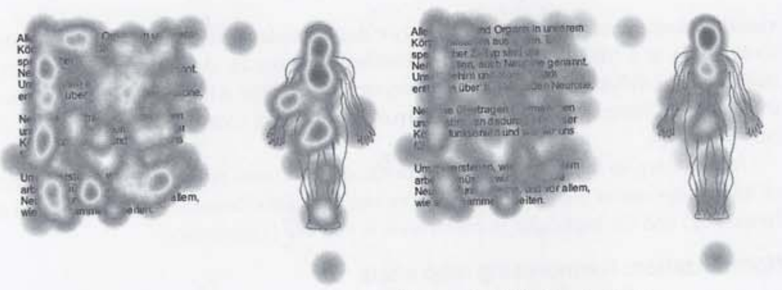
- The Gaussian can be used to *approximate the acuity fall-off* away from the point of fixation. It can then be argued that the value of the attention map corresponds to what a viewer can visually resolve at that position. Taken one step further, the attention map can be thought to reflect the amount of attention a viewer spends at each image location. As we already know, fixation is not identical to attention, nor does the Gaussian perfectly match the fall-off in acuity,[26] but seen from a mathematical perspective it is one of the simplest, most well-known and well-behaved functions there is that approximates this fall-off.
- A smooth, Gaussian-based attention map can be used as a probability density function that gives a good estimate of where a new viewer, having the same image and task, is likely to look.
- Because of its desirable properties, a powerful suite of mathematical tools can be used to derive measures from the Gaussian-based attention map (p. 359–376).

The few researchers who have used Gaussian-based attention maps have suggested that the variance $\sigma^2$ should be set so that at the half-height of the Gaussian peak, the width of the hill at that altitude corresponds to the size of the foveal projection on the stimulus image (Rajashekar, Cormack, & Bovik, 2004; Nyström, Novak, & Holmqvist, 2004). The foveal size is approximately two degrees of visual angle (p. 21). If the geometrical set-up of the experiment is known (or can be estimated), the 2° variance setting can easily be calculated. As mentioned above, this setting of $\sigma^2$ does not make the attention map mimic the fall-off of acuity on the retina precisely, but it does approximate both the size of the foveal projection and the further decrease in acuity outside. There is in fact not much point in precise modelling of retinal acuity, since attention maps almost always incorporate several participants whose data are not perfectly aligned. The variance has also been adjusted to compensate for the lack in precision of the eye-tracker; the higher the precision the smaller the variance and vice versa. Following this rule, data from a remote eye-tracker would produce a smoother attention map than data collected from a tower-mounted high-end system.

## Raw data samples or fixations?

Either raw data samples or fixations can be used to generate the attention map. The attention map, and the visualizations made from it, will differ depending on your choice. This is

[26] The true fall-off appears to follow an exponentially decreasing function (Geisler & Perry, 1998).



(a) Filled with number of fixations

(b) Filled with fixation duration

**Fig. 7.13** The role of filling with fixation duration becomes clearer when fixation durations differ in different parts of the scene. Here fixations on the text are short, but longer on the image, in particular the face. Data recorded using a very common but slightly outdated 50 Hz remote and visualized using Tobii Studio.

particularly true for the Gaussian attention map, so we will focus on it here.

If you have reason not to trust the fixation algorithm, either because it is simply not a good algorithm, or because you are uncertain of its settings, or because you use animated stimuli that elicit smooth pursuit (Chapter 5), then using raw data to generate the attention map is the only proper alternative. If the attention map is built from raw data samples, then each sample contributes with the same small height, simply because a data sample is always of the same duration.

Commercial implementations of attention maps typically use fixations as input when building attention maps. This way, the input is preprocessed, avoiding adding saccades and various noise to the visualization. Using fixations as input is beneficial also in terms of computational processing, since fixations are substantially less voluminous than raw data; about 250 times smaller when recorded at 1000 Hz. Consequently, it becomes quicker to generate and update an attention map that is built with fixations instead of raw data.

With fixations, there is a choice: either each fixation yields the same height in the map, or we scale the fixation with its duration. Using the manufacturers' terminology, this is often referred to as 'unscaled' or 'scaled' heat maps. In an attention map filled by fixation duration, each peak will represent *fixation time* on an area and therefore look more like an attention map generated from data samples than will an attention map filled by numbers of fixations. The peaks in the attention map then rather represent the *number of fixations* in an area. As dwell time is more related to the level of cognitive processing and number of fixations rather than to repeated interest, the two attention maps also represent different types of analysis. For instance, Figure 7.10 from Pomplun *et al.* (1996) uses attention maps filled with fixation duration to visualize differences in processing.  In practice, filling attention maps with fixation duration does not generally differ much from filling with number of fixations. Only if fixation durations differ, as in Figure 7.13 where fixations are short on the text and long on the image, does this choice makes a difference.

There are situations when the choice between using raw data and fixations becomes particularly relevant. If you are using an eye-tracker with low sampling frequency and poor precision, individual data samples can significantly impact the appearance of the heat map, making it look very different from one generated from fixations. This case is nicely illustrated by Bojko (2009), who recommends users to always generate heat maps from fixations. If you on the other hand use data recorded at higher speed and good precision, heat maps look more or less the same regardless of how they were generated.

One exception is if the width of the basic construct is small; Figure 7.14 shows heat map visualizations generated from the same raw data samples using the unprocessed data directly (column 1) and filled with fixation durations (column 2). As illustrated in the figure, the magnitude of difference between the rows depends on the width of the basic construct ($\sigma$ in the Gaussian function), with a very narrow width more clearly emphasizing the difference between maps.

In summary, to fully understand an attention map visualization, a certain minimum amount of knowledge about the nature of the eye-tracking data (accuracy, precision, and sampling frequency) and the particular implementation you use is necessary.

### Normalization: harmonizing map scale

Depending on the later usage of an attention map *normalization* is sometimes necessary. Normalization changes the scale of the attention map, but retains its form. Two kinds of normalization have been used by researchers, with slightly different properties.

1. The simpler type of normalization is done by scaling the altitude of the attention map, so that the maximum peak has the altitude value 1. This allows two attention maps to be compared even if one has a higher maximum peak than the other, which is the case if it is built on a much larger participant population or on longer recordings with more fixations than the other (Wooding, 2002a).
2. The other type of normalization scales the altitude so that the volume underneath the attention map (the total volume of sand poured) is 1 unit. This allow for the same kinds of comparisons across maps as the simpler normalization, but additionally opens up for use of the normalized attention map as a *probability density function* (pdf).

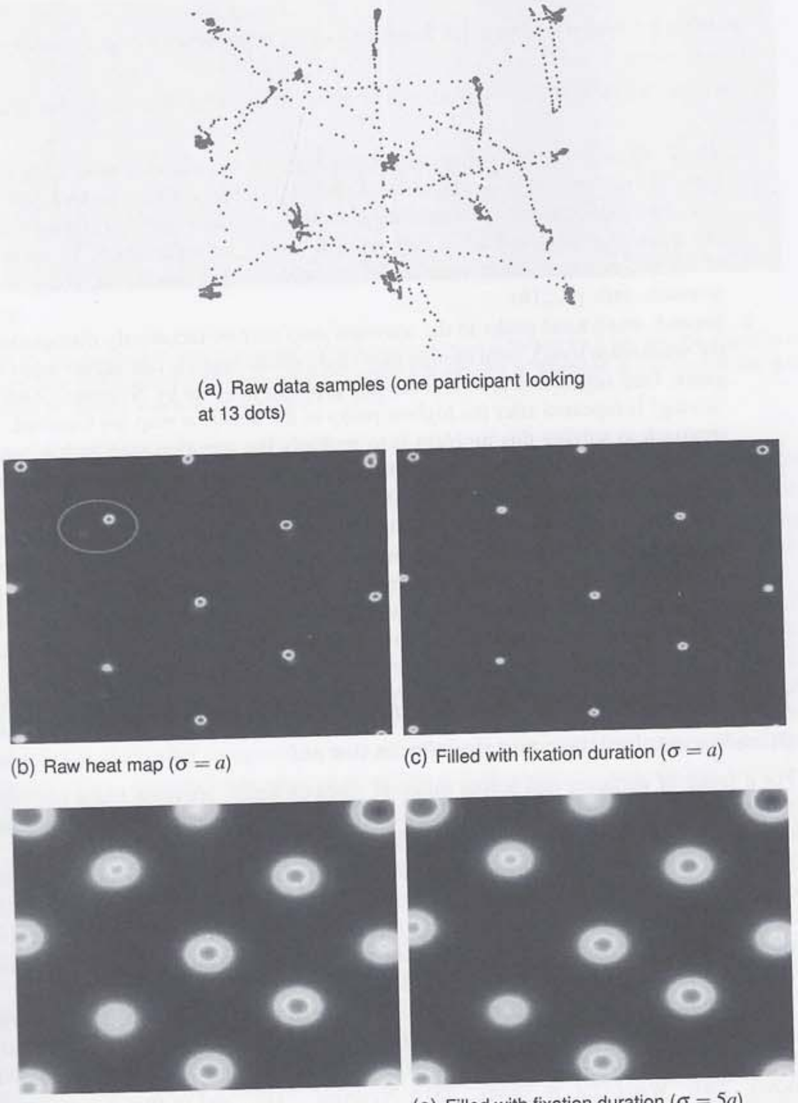
The pdf representation of attention maps has a range of properties that allow us to use well-established mathematical tools from information theory, such as the Kullback-Leibler distance (KLD) on page 376. Normalization is particularly important when two attention maps are to be compared with each other, removing differences due to unequal viewing times.

## 7.5 Usage of attention maps other than for visualization

More than for any other purpose, attention maps are used to visualize data. Visualization is what commercial eye-tracking software offers for us to do with attention maps. But there are other possibilities: you can use the attention map to manipulate the stimulus in a variety of ways, and look at effects of manipulation. Under certain conditions, attention maps may also be used to define AOIs. Moreover, viewing behaviour can be *quantified* using attention maps; new measures based on attention maps are continuously finding their way into the analysis of eye-tracking data.

### 7.5.1 Using attention maps to define AOIs

Attention maps give "an opportunity for defining objectively the principal regions of interest of observers when they view an image", according to Wooding (2002a). There are two ways to describe the operation that does this. Imagine that the sand metaphor of the attention map that we formed above has solidified, and that it is flooded to a level so that only some of the peaks are above the waterline, as mountainous islands in an archipelago. In this metaphor, the coastline of each island defines the form of an area of interest. An equivalent metaphor with no water is to slice the attention map at the same altitude as reached by the flood, and let the slice areas comprise the AOIs. We then project the slice areas (or the equivalent coastlines) down to the stimulus image at the bottom of the attention map. Given an attention map with eye-tracking data, two things decide the size and form of the generated AOIs:



(a) Raw data samples (one participant looking at 13 dots)

(b) Raw heat map ($\sigma = a$)

(c) Filled with fixation duration ($\sigma = a$)

(d) Raw heat map ($\sigma = 5a$)

(e) Filled with fixation duration ($\sigma = 5a$)

**Fig. 7.14** Heat map visualization from raw data samples (column 1) and filled with fixation duration (column 2). Notice how raw data samples from very short periods of stillness outside of the central fixation point give rise to bleak spots (for instance inside the ring) in the raw heat map at setting $\sigma = a$, while at $\sigma = 5a$, these raw outlying data samples are hidden under the wide Gaussian distribution. Eye movements were tracked at 500 Hz with a tower-mounted eye-tracker.

- The level of the slice/flood. Wooding (2002a) does not suggest how this value should be selected, nor does anyone else. The AOIs become smaller (and typically fewer) as the level of water increases, i.e. as the height of the slice increases.

- What value of $\sigma$ is selected in Equation (7.1); a large value of $\sigma$ gives wider AOIs, and vice versa.

Notice also, that the attention map method to define AOIs suffers from a number of disadvantages.

1. The AOIs do not necessarily encapsulate objects in the stimulus image, and if they do, parts of the objects may be outside of the AOIs. Just like with the gridded AOIs on page 212, the AOIs generated from attention maps do not correspond to semantic entities in the stimulus image, and when they do, it is largely by coincidence. In contrast, most of the analyses that would make use of the AOIs assume that the AOIs are meaningful semantic units (p. 216).
2. Second, small local peaks in the attention map may be incorrectly disregarded (under the water/slice level), even though they could define regions that attract many viewers' gazes. One suggestion to overcome this problem is given by Nyström (2008), where 'slicing' is repeated after the highest peaks in the attention map are removed. Another approach to solving this problem is to multiply the attention map with a 'squashing' function which remaps the attention map, such that values above average grow, whereas other values are suppressed (Itti, 2004).
3. Third, using an attention map to define AOIs often means defining the AOIs in a post-hoc manner. Whether this is a problem or not depends on the precise research question, and what AOI measures are then employed using the generated AOIs (p. 216). Probably, measures utilizing transition and string representations of data make more sense to use with these AOIs than dwell time and the various fixation number measures.

## 7.5.2 Attention maps as image and data processing tools

### Stimulus manipulation and the effects thereof

For a range of purposes and across different research fields, attention maps has been used to manipulate an image such that luminance, resolution, frequency, or contrast are modified according to the height dimension of the map (see e.g. Pomplun *et al.*, 1996; Wooding, 2002b; Nyström, 2008). We have already seen some examples of this earlier in the chapter. For example, Figure 7.10 is one of the earliest uses of attention maps to manipulate images. After having produced the luminance maps in this figure, Pomplun *et al.* had another group of viewers looking at them, and found that the perceptual interpretation participants saw was very much influenced by image manipulation.

Attention map-based image manipulation has been widely used in both engineering and psychology; in engineering to speed up graphics rendering and improve image and video communications (see e.g. Parkhurst & Niebur, 2002; Kortum & Geisler, 1996; Wang & Bovik, 2001; Wang, Lu, & Bovik, 2003; Nyström, 2008), and in psychology to measure the perceptual span in scene perception (Loschky, McConkie, Yang, & Miller, 2001) and to investigate the possibility of manipulating viewers' gaze behaviour and perception (Dorr, Jarodzka, & Barth, 2010; Dorr, Vig, Gegenfurtner, Martinetz, & Barth, 2008), to name just a few examples.

Girod (1988) argued that if you know where a person looks, this information can be used to make image and video communications more efficient. Since people cannot see with high detail in their peripheral visual field, it would be enough to have high quality only where the person looks and reduce the quality elsewhere. Such quality reduction typically makes a video smaller in size after compression. More recently, this way of manipulating video quality has been dubbed 'foveation', and has been adapted by a number of people working with image and video compression (Juday & Fisher, 1989; Stelmach & Tam, 1994; Kortum & Geisler,

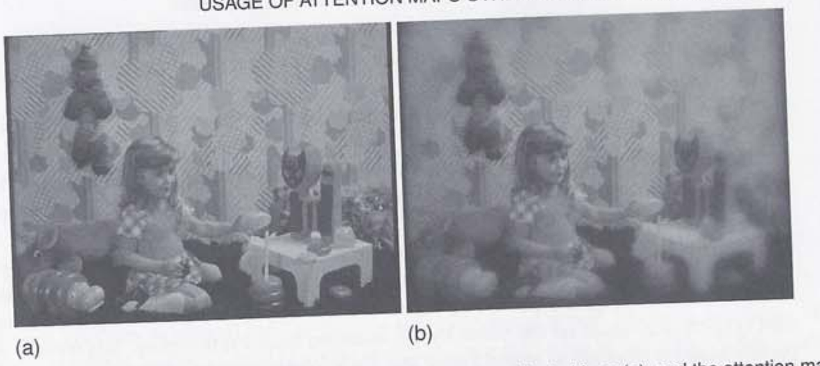


**Fig. 7.15** An example of image foveation. Fixations have been collected from (a), and the attention map created from the fixations has been used to control the amount of foveation in (b). Notice the quality reduction in regions where no or few people fixated.

1996; Geisler & Perry, 1998; Duchowski, 2000; Wang & Bovik, 2001; Wang *et al.*, 2003; Bergström, 2003; Nyström, 2008). Foveation-based techniques have been used both when a person's gaze position is measured online by an eye-tracker and offline when eye movements from one of several people are recorded beforehand. An example of the latter is given by Nyström (2008) where video regions are foveated in accordance with the attention map. Regions with small heights in the map were lowpass filtered more heavily than other regions, resulting in blurred quality with little contrast while regions corresponding to the peaks in the map were left in full quality. An example of such offline filtering is shown in Figure 7.15, where the left figure shows the original image and the right figure illustrates the image after foveation. In Nyström (2008), foveation showed a reduction in the number of bits needed to represent the video after compression, without a reduction in subjective quality. Watching the foveated videos, viewers tended to look at regions in high quality (where original viewers had looked) and avoided gazing at regions that has been heavily lowpass-filtered.

With the resolution (which typically reflects the quality) of an image being reduced in regions where people do not look, there is an obvious computational gain when generating foveated images instead of images with a high, uniform resolution (which is the typical case). Parkhurst and Niebur (2002) found this gain to reach 8.7 for a typical viewing set-up on a computer screen. This means that it requires 8.7 times more computational resources (or 8.7 times more time with the same resources at hand) to render an image with uniform resolution compared to a foveated image.

Besides the potential benefit of using attention maps to improve the efficiency in technical systems, there are several areas in psychology where such image manipulations have been shown to be useful. Loschky *et al.* (2001), for example, used gaze-contingent foveation to investigate how much you can reduce the peripheral quality of an image before it becomes visible to the observer, and also how such manipulations affect eye-movement parameters such as fixation duration. Moderate quality reductions did not affect either the impression of image quality or eye-movement behaviour. However, with an increasing degree of peripheral blur, quality judgements became lower, fixation duration increased, while saccade length decreased. For this set-up to work in real time, the image quality must be updated very quickly every time a viewer changes his position of gaze, otherwise quality impairments will be visible after a saccade. Loschky and Wolverton (2007) found that if the update is complete within 60 ms after the saccade has landed, viewers will not notice that change. These results tell us that caution should be taken if you plan to use foveated images in your experiment; depending on how the quality reductions are implemented and how quick your system is to update and display the foveated images, perception and eye movements can be affected in various ways.

Image manipulation based on attention maps has opened up some interesting possibilities to study the effects of gaze guidance, which refers to the hypothesis that manipulations can be used to control or guide a viewer's gaze towards certain parts of a display. There are several situations in which gaze guidance would be interesting to study. For example, eye movements from experts can be recorded on instruction videos in medicine, human factors, biological classification, and other fields where it is important to look at the correct position to make a correct diagnosis or classification based on visual observations. Novices studying foveated video examples based on an expert model's eye movements, are guided to the relevant (expert-like) areas on the video by the attention map. Furthermore, this trained viewing pattern enables students to attend faster and for a longer time to relevant areas in new—not foveated—videos and interpert these more correctly, compared to students learning from non-foveated videos (Dorr *et al.*, 2010; Jarodzka, Balslev, *et al.*, 2010; Jarodzka, Van Gog, Dorr, Scheiter, & Gerjets, forthcoming).

### 7.5.3 Using attention maps in measures

So far, attention maps predominantly become a tool for *visualizing* eye-tracking data. But attention maps are by no means limited to visualization, but can be used to *quantify* and *statistically test* the distribution of a group of fixations or data samples, or the position similarity between two different groups of fixations, even over time. A number of measures now exist that do this, either using just the set of points, Gaussians, or gridded AOIs, see Chapter 11.

## 7.6 Summary: attention map representations

Attention maps are used to describe the overall spatial distribution of eye-tracking data. They typically come in one of three representations:

- The **gridded AOI** with its grid size settings and its various fillings.
- The **Gaussian landscape** with its $\sigma$-setting.
- The normalized version of either of these, known as a **probability density function** or pdf.

Attention map visualizations, in particular heat maps, are today widely used to illustrate large amounts of eye-tracking data. With the current range of software tools, attention maps can be built in a variety of ways, each having its own concept of what an attention map is. They are so visually appealing that less careful viewers may easily misinterpret them, but these visualizations are nevertheless important complements to quantitative data. Settings for attention map visualizations always contain a width or $\sigma$ but are otherwise diverse: before you make any decisions on the basis of a heat map, check carefully how it has been calculated and what settings were used. If you publish a heat map, always report the form and size of the basic contructs, whether height is scaled by fixation durations or any other measure, whether raw data or fixation position were used, and the precise mapping of colour to altitude. Otherwise your heat map may be a piece of art suited for gallery exhibitions but hardly a tool for the scientific study of human behaviour.