Questions

1. In vision, the optical blur that occurs at an image location depends on the accommodation of the lens and on the distance to the point. The blurred image that results is the convolution of the sharp image with the impulse response function that is due to the blur.

What is the analogy in audition, namely what is the relation between the source sound and the sound measured by the ear?

2. In vision, it sometimes happens that you see two images superimposed on each other. This can happen, for example, when you have a dirty glass surface and you see both the dirt on the glass surface as well as the image transmitted through the glass.

The visual system sometimes can perceive the two superimposed images separately, for example, by focusing on the glass (and converging the eyes to bring the glass to zero disparity), or focusing on the background. What is the analogy in audition?

3. Describe the sequence of transformations of a voiced sound that is spoken by one person and heard by another. Start from the glottal pulse train, and go all the way to its coding in the auditory nerve (i.e the code leaving the cochlea).

4. Suppose a ganglion cell in inner ear is tuned to high frequency, say 4 kHz. Suppose the sound source is dominated by a pure tone of frequency 1 kHz. Describe the response (if any) of this ganglion cell. [WARNING: my answer here is somewhat speculative.]

5. Suppose a ganglion cell in the cochlea can produce spikes at a maximum rate of 500 per second. In particular, suppose that the cell requires 2 ms or more between spikes. Describe the autocorrelation function of the cell’s response to a pure 1 kHz tone, assuming that the cell has a non-negligible sensitivity to this this frequency.

FYI: The auto-correlation of a signal is defined is the cross-correlation of the signal with itself. For a signal $s(t)$, the autocorrelation function is:

$$ R(\tau) = \sum_{t=0}^{T-1} s(t) s(t + \tau) $$

where $s(t)$ is discretized time, with values 0 or 1 depending on whether or not there is a spike in bin $t$. 
Solutions

1. The analogy in audition is that the sound pressure signal that arrives at the ear drum when there is a source from a single direction $(\phi, \theta)$ is the temporal convolution of the source sound $I(t)$ with the HRIR function $h(t, \phi, \theta)$ for that direction.

In the case of vision, the eye doesn’t measure the original sharp image. In the case of audition, the ear doesn’t measure the original source signal.

2. In audition, you may have two sounds coming from different directions. These sounds will be superimposed.

To perceptually separate the sounds, it sometimes helps to reorient the head. This can amplify some sounds and attenuate others. (When I am at a very loud event such as a poster session at a conference, I sometimes turn my head slightly away from the poster presenter and plug the ear facing away from him/her).

3. Each glottal pulse is transformed by the articulators (jaw, tongue, oral cavity etc) which is modelled by convolving with some $a(t)$. The sound leaves the mouth and nose and travels through space where it attenuated in strength with distance. The head and ear of the listener then reshape the sound further. This is modelled by convolving with some $h(t)$, the HRIR. The sound wave then contributes to the oscillations of the basilar membrane (gammatone filter). The various mechanical frequency components of the BM vibrations get transduced into chemical signals (neurotransmitters released by hair cells). The times of the peaks of these signals are coded by phase locked volleys of spiking ganglion cells, at least for sounds up to 2 kHz. Beyond 2 kHz, phase locking is not possible.

4. The ganglion cell has relatively weak sensitivity to a 1 kHz sound since it is two octaves away from the peak. But if that sound is loud enough then the 4 kHz sensitive cell will respond to some extent. It just won’t spike with nearly as great a spike frequency as it would if the sound had a 4 kHz component of the same intensity.

The question you might ask yourself is whether the spikes that do occur would be phase locked. I don’t know if they are. I would would speculate that a 4 kHz cell would respond in synch with the peaks of the 1 kHz sound. (The issue here for us isn’t whether this is true or not – since it will have no bearing on the rest of our lives one way or the other. Rather, its an exercise is postulating logically distinct possibilities – which is one of the things we’re learning how to do here.)

5. The autocorrelation function has a sharp peak at $\tau = 0$ since if $s(t)$ has a spike at some $t$ then $s(t + \tau)$ also will have a spike at $t$ when $\tau = 0$ but it will not have a spike at any time $\tau$ near 0 (but different from 0), so the autocorrelation function will be 0 for $\tau$ near 0.

Then, because of phase locking to the 1 kHz sound, one might expect the autocorrelation function to have another another sharp peak at $\tau = i$ ms (i.e. corresponding to peaks in the basilar membrane motion at a period of 1 ms, corresponding to the 1 kHz frequency of the band). However, the question assumes that time between spikes must be at least 2 ms. So the autocorrelation function will have peaks at $\tau = i$ ms where $i \geq 2$. That is, it will be missing the peak at $\tau = 1$ ms.