Chapter 3

Sound Perception and Room Effects

In an information-rich virtual reality environment, the user is immersed in a world containing many objects providing that information. Given the finite computational resources of any computer system, optimization is required to ensure that the most important information is presented to the user as clearly as possible and in a timely fashion [Herder and Cohen, 1997a].

3.1 Conceptual model

A conceptual model, diagrammed in Figure 3.1, has three layers. The top layer spans the known human-machine interface. A middle layer distinguishes between perceivable space/sensor system and motor system/feedback [Sanders and McCormick, 1987, p. 46–47] [MacKenzie, 1995, p. 438]. The models layer completes the classification. Typical user interfaces do not consider in a dynamic way the perceivable space of the user; only at the design stage is this estimated, and it is statically frozen by programming. More advanced user interfaces measure user performance, or can switch between different modes to support the users. Examples of such adaptation include classification of user expertise and then offering more functions in menus and less or different help texts [Sukaviriya and Foley, 1993]. Research in adaptive user interfaces has been done with focus on various aspects including adaptive automatic display layout [Stille *et al.*, 1996] and information filtering for pilots based on workload [Mulgund and Zacharias, 1996].

A virtual reality user interface can provide more information to a user than WIMP-based interfaces by using broader sensory channels. Also, more information about an immersed user can be obtained by the system and used

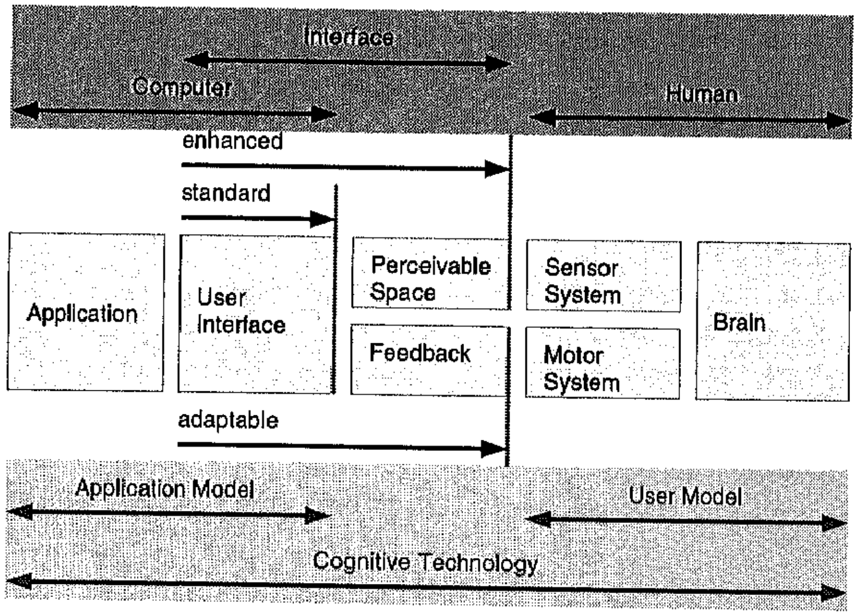

Figure 3.1: Concept — Interface — Models

to enhance the user interface in a dynamic fashion. Examples are head- or eye-tracking for finding out users' focus area.

3.2 Resource management

Human interface resources can be classified using different taxonomies. One such organization classifies according to resources provided by the computer. Another is to classify by the ability of the user to perceive information [Ovan and Havens, 1993].

It is not necessary to compute (i.e., use system resources) display data which the user cannot perceive because of occlusion, masking, or low level. Rendering devices have limited capabilities. A sound spatialization backend (e.g., Acoustetron) can render (i.e., put monaural sound sources into three-dimensional space) only a limited number of mixels (e.g., eight channels) simultaneously. The responsibility of the resource manager is to determine the set of computable renderable data $R_{\text{computation}}$, which is a subset of both $R_{\text{displayable}}$ and $R_{\text{perceivable}}$.

$$R_{\text{computation}} \subseteq R_{\text{displayable}} \cap R_{\text{perceivable}}$$

$R_{\text{displayable}}$ is the set of resources which are available for the display.

$R_{\text{perceivable}}$ is the set of resources which the user can perceive. At a given point in time, $R_{\text{computation}}$ is optimal if there is no larger set which fulfills the requirements. This is not necessarily the best solution over a long period in time because allocated resources cannot be freed immediately, so that in a dynamically changing scene, a non-optimal solution for a short time period could give, on average, a better impression.

3.3 Audio resources

The responsibility of the audio module in a VR environment is to present sound sources as well as possible, but any practical system has only limited resources, including spatialization channels (mixels), MIDI/audio channels, and processing power. A sound spatialization resource manager [Herder and Cohen, 1997b] controls sound resources and optimizes fidelity (presence) under given conditions. For that, a priority scheme based on psychoacoustics is needed. Parameters for spatialization priorities include:

- intensity, calculated from volume and distance,
- orientation, for non-uniform radiation patterns,
- occluding objects,
- frequency spectra (low frequencies are harder to localize), and
- expected activity.

Objects which are spatially and perceptually close together (depending on distance and direction) can be clustered. Sources that cannot be spatialized separately may be mixed as ambient sources.

3.4 Room simulation

Localization means not only directionalization, whence a source emits, but also range estimation and a "feeling" of the space. The auditory environment gives a context to the situation, the space, and helps to orient the user in it.[1]

Head-related transfer functions are usually measured only at a fixed distance. The convolution of a low reverberant signal with HRTFs gives good

[1]The reverberation missing in an anechoic chamber does not allow usual localization just by using the auditory senses. People in a dark anechoic room experience a kind of "lost in space" sensation which can even approach fear.

	spatialization	localization
media	hardware/software	auditory senses
function	directionalization (i.e., θ, ϕ) distance	orientation
space	room effects	presence

Table 3.1: Spatialization/localization taxonomy

directional cues but not distance cues. It is necessary to add room reverberation to the processed signal [Anderson and Casey, 1997]. Such a processing scheme is shown in Figure 3.2.

Role of reverberation in distance perception In visual perception, if one sees a larger object far and a smaller object near, their relation can be recognized from context [Ishikawa *et al.*, 1998]. Aural perception includes analogous effects. The magnitude of a sound is source intensity, having a physical, objective value. But perceived source level is loudness, a mental percept. If a source with small intensity is close to a listener, it is recognized as such through reverberation, which is analogous to the context (objects of known size) in visual perception. Since reverberation is the environmental context, it must be constant across distance, depending only on source intensity.

3.5 Audio rendering based on an image model

A common technique for including early reflections in audio rendering is to use an image model [Begault, 1994, p. 184] [Kendall and Martens, 1984]. An extension of this is the cellular approach for modeling room acoustics [Dhillon, 1994]. For the rendering two passes are necessary. In a first pass, rays are sent out from sound sources to the sound sinks. This process is analogous to ray-tracing in graphics. From those rays, representatives are chosen using a selection process. Along those significant rays, filter functions are calculated for radiation, reflection, and occlusion. In a second pass, the filter functions are applied to the audio signals. As optimization, the filter function can be compiled to one filter function. Figure 3.3 shows the image-based rendering process. The filter functions require local information. The radiation function is based on radiation direction and radiation pattern. A reflection filter function uses material data, surface normal and incoming direction. The occlusion function [Tsingos and Gascuel, 1997] uses the ratio

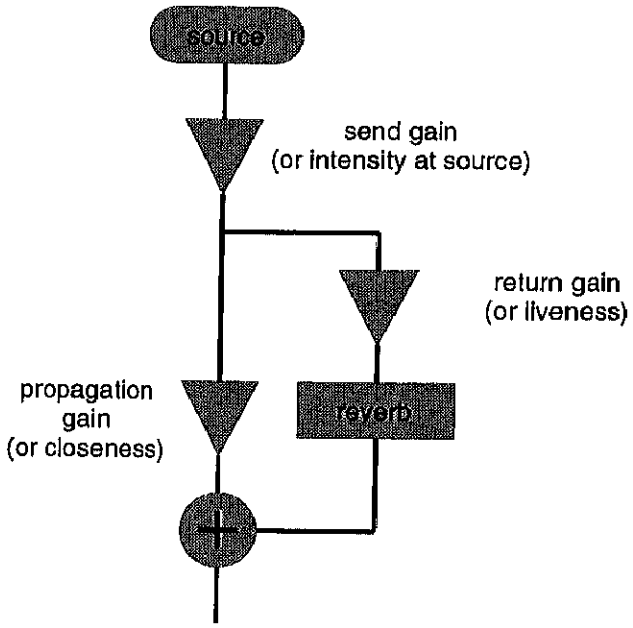

Figure 3.2: Source loudness amplification

between room volume and occluder volume. Part of the necessary data can be provided in a system with integrated graphics and audio rendering by the graphics process.

The complexity of sound rendering based on ray-tracing is illustrated in the following paragraph. Equation 3.1 shows the calculation of a signal along one ray s_{r_i} by applying the sink filter function L_k, reflection and occlusion filter functions F_{h_i}, and source radiation filter function R_i to a source signal s_i. The index h_i runs from 1_i to n_i. The cardinal n_i depends on the number of occlusions and reflections as well on the maximum numbers of filter functions, determined by system capabilities or user tuning, for one ray (usually limited). A signal of a source reaches the sink along different paths (rays). Therefore the signal contributed by a source s_{o_j} is given in Equation 3.2, which is the summation over all rays. The calculation of a signal for an output channel s_k is given in Equation 3.3 by adding the signal for room reverberation and contributions of all sound sources. For a headphone based system the range for k is two.

$$s_{r_i} = L_k * F_{1_i} * \ldots * F_{n_i} * R_i * s_i \tag{3.1}$$
$$s_{o_j} = \sum_{i \in rays} s_{r_i} \tag{3.2}$$

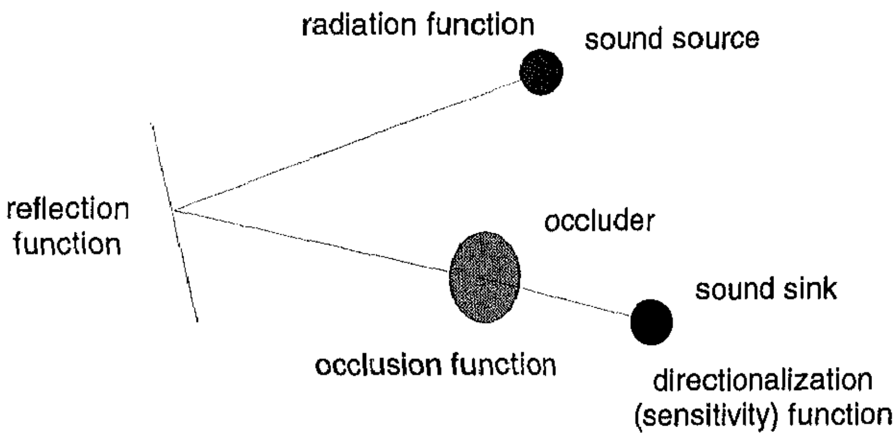

Figure 3.3: Image-based rendering

$$s_k = s_r + \sum_{j \in sources} s_{o_j} \tag{3.3}$$

The following table lists the used symbols for calculating the output signal:

R	source signal radiation function
F	reflection or occlusion filter function
L	sink filter function (for directionalization; in case of headphones: HRTF)
k	output channel
s_i	raw source input signal
s_{r_i}	signal along one ray
s_{o_j}	signal contribution for a source
s_k	signal for an output channel
s_r	signal for room reverberation

3.6 Early reflections

Early reflections, especially first-order reflections, contribute to localizability and space awareness. First- and higher-order reflections can be simulated using source an image model [Wenzel, 1994] based on mirror-like symmetries. First- and second-order reflections based on the source image model have been implemented in a spatial reverberator [Kendall and Martens, 1984, p. 118–120]. The image model assumes specular reflections and practical realtime systems limit the order to two or three [Begault, 1994, p. 184–186].

Implementing reflections using the image source model like that mentioned in the previous section does not scale well for complex environments.

A very promising approach [Funkhouser *et al.*, 1998] based on beam-tracing[2] depends mostly on the complexity of the local environment. Currently, this approach works only for fixed source locations, but extensions are expected.

3.7 Perceptual space

Section 3.1 introduced the notion of a sensory space of the user giving a global context, but did not explore it further. For spatialization resource management (Chapter 2), criteria are necessary to make the allocation and clustering (Chapter 4) decisions. In [Herder and Cohen, 1997b], application programmers' priority was used as criteria, coupled with a simple physical model based on distance and intensity. This does not adequately capture the capabilities of the human auditory system, which has different abilities to perceive sound depending on environment and direction. This richness is well-known, often measured using audible limens, also known as "just noticeable differences" (JND) [Williams, 1994, p. 100] [Blauert, 1996, p. 16]. A similar term is "minimum audible angle" (MAA) [Begault, 1994, p. 40]. This space is inhomogeneous. If distances based on Euclidean space are used as criteria for sound spatialization management, then errors are introduced as the perceptual capabilities of a human listener are not considered. A mapping from the Euclidean space of the virtual reality environment (object space) to perceptual space, which equalizes the audible limen, can solve this problem. In this space all decisions concerning spatialization resource management, like relevance calculation and clustering, are done. From this space we can get back into the object space by keeping references to the objects. This is more efficient and accurate than trying to define a function which would do the back-mapping. This perceptual space is context- and application-dependent. For example, head-tracking changes auditory resolution.

The perceptual space is used for answering following important questions of the resource management:

- Is a given audio event/signal audible[3]?
- Is a given audio event/signal masked by another signal?
- Is a given audio event/signal localizable separately from another audio event/signal?

[2] a beam covers a collection of rays

[3] "Audible" here means that the signal is louder then the noise/reverberation level of the system.

- Does a first- (or higher-) order reflection contribute to the perception of a scene?
- Does an occluder contribute to the perception of a scene?

3.7.1 Construction of the perceptual space

Head-related transfer functions (HRTFs) are measured and are in common use for sound spatialization. HRTFs can also be used to calculate the intensity at the ear for a given signal with a certain frequency range at a certain source location relative to the listener position.[4] Other studies suggest loudness and audibility functions dependent on frequency [Schiffman, 1995, p. 351–353] [Blauert, 1996, p. 120]. Thresholds for audibility are insufficient for determining actual audibility because the signal might have a level lower than the room noise.

Besides the question of audibility, an important factor is localization blur, which depends on frequency and orientation of the listener [Blauert, 1996, p. 40–50]. Multiplying the values of both functions gives loudness dependent of distance, orientation (i.e., location), and frequency. The resulting function can be used for guessing the audibility and maskability of audio events. The localizability has be modeled separately.

3.8 Role of the frequency band in directionalization

The frequency band of a sound source plays an important role in the ability of an user to perceive a sound source from a certain direction. A listener is able to differentiate sound sources in the horizontal plane using interaural time difference (ITD) and interaural level difference (ILD). The cone of confusion [Handel, 1989, pp. 110–111] [Blauert, 1996, p. 179], shown in Figure 3.4, describes the locus of points with the same ITD and ILD. Spectral differences allow listeners to discriminate sound sources on the cone of confusion. The spectral content of a sound is correlated how well a sound source can be spatialized. Even for narrow-band signals, predictions of ability to locate a sound source can be made [Blauert, 1968].

Table 3.2 shows qualitative relations between frequency band and ability to directionalize sound sources. Telephone equipment used for voice communication have cutoff frequency at around 5 kHz which reduces the required bandwidth.

[4] This is a convolution of the signal with the HRTF for the specified location.

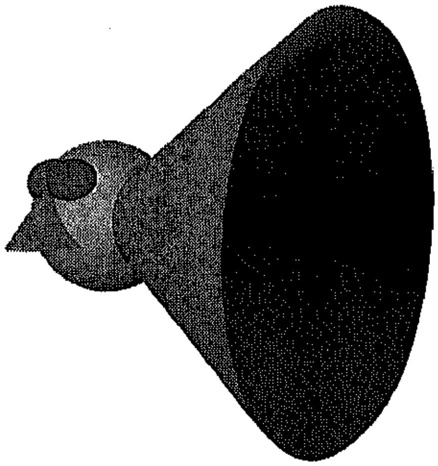

Figure 3.4: Cone of confusion; locus of points with the same ITD and ILD

frequency range	application	front $\leftrightarrow$ back confusion	elevation discrimination
low	voice, conferencing	high	low
high	music, environmental sounds	low	high

Table 3.2: Frequency band determines limits of spatial resolution

Harmonic distortion and interaction with other objects can generate higher frequencies.

3.8.1 Localization error depending on target direction

Figure 3.5 shows horizontal and vertical unsigned localization errors for a broadband signal (measured values are taken from [Makous and Middlebrooks, 1990]). The ellipses axes of the cones denote error in azimuth and elevation. A generalizing of the data suggests that the error in azimuth to the front is small, growing larger to the sides, while error in elevation decreases at the sides. The back has much higher error than the front.

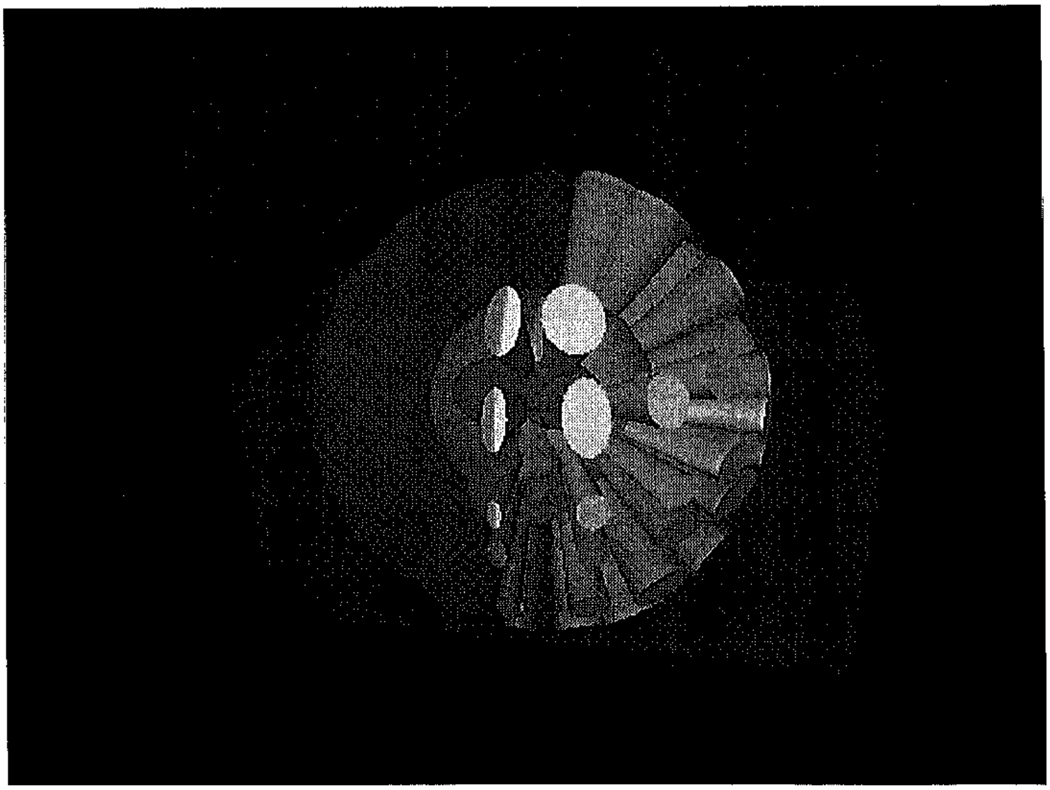

Figure 3.5: Horizontal and vertical unsigned localization errors for a broadband signal; ellipses axes denote error in azimuth and elevation

3.8.2 Elevation discrimination

In a listening experiment at the University of Aizu [Suzuki, 1999], elevation discrimination was investigated for two types of stimuli. Voice samples had a bandwidth of 5 kHz and an explosion sample had a bandwidth of 10 kHz. The experiment was done with 5 subjects in an anechoic chamber for five elevation angles and four major azimuth angles. The influence of reverberation on an identification task was studied by presenting reverberation through two separate loudspeakers.

The results are summarized in Figure 3.6. Sensitivity is shown using the d' metric [Green and Swets, 1966, p. 405], the difference of the Z distribution of the hit rate and false alarm rate, which excludes biases. Elevation angles were better estimated without reverberation. At lateral angles the identification performance was better than in the front or back. High frequency stimuli could be better identified.

3.9 Sound occluder

In previous research [Tsingos and Gascuel, 1997], Fresnel zones[5] had been used to calculate a filter function of a given occluder along a sound path. Visibility methods like those used in computer graphics allow efficient calculation [Tsingos and Gascuel, 1997]. In other work, an occlusion volume was suggested [Ellis, 1998] as an extension for the VRML97 specification [Bell *et al.*, 1997]. This newly proposed `SoundOcclusion` node represents the volume as a sphere that defines an infinite impulse response filter (IIR).

In the interest of improving audio rendering efficiency, a simplified filtering model was developed [Martens *et al.*, 1999]. Two perceptually salient components of occluder acoustics were identified that could be directly related to the geometry and orientation of a simple occluder. Actual occluder impulse responses measured in an anechoic chamber resembled the responses of a model incorporating only a variable delay line and a first-order (one-pole, one-zero) filter. The filtering model has the attractive feature that either the reflecting or the occluding effects of an obstruction can be simulated through changes in continuously variable parameters, requiring no change in the filter structure.

What does acoustical occlusion sound like? When an object is interposed between a source and a sink, and that object is not acoustically transparent, it typically reduces the sound pressure level (SPL) that can be measured at the sink position. It also changes the spectral energy distribution, and

[5] Fresnel zones are volumes enclosed between ellipsoids.

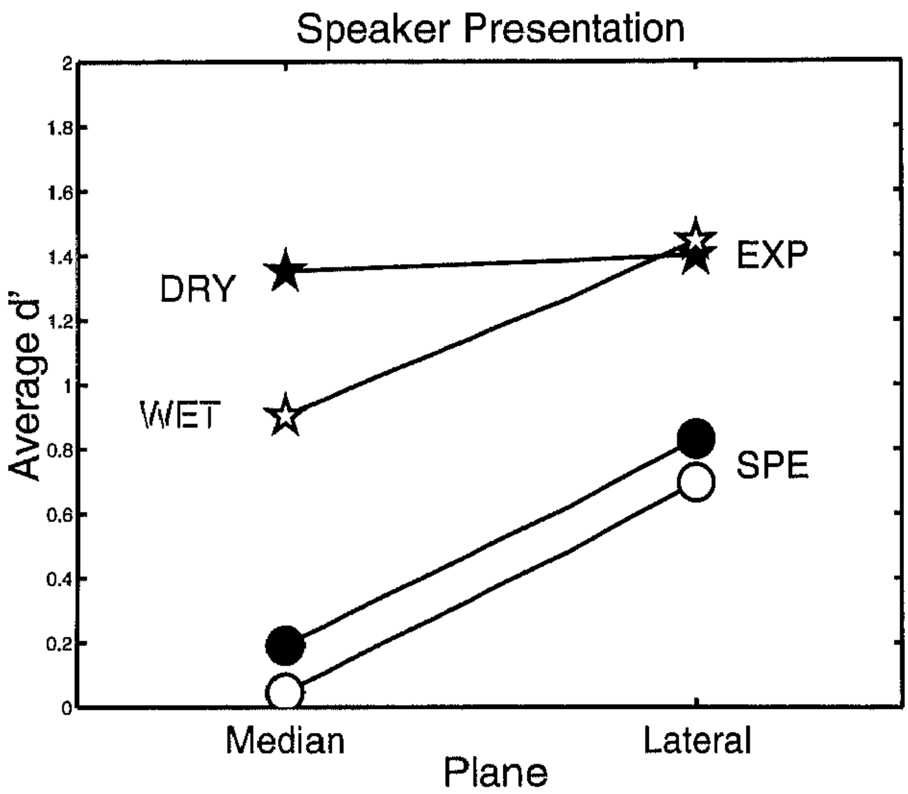

Figure 3.6: Sensitivity for elevation discrimination depends on content and direction: d' measures the sensitivity; stars (EXP) are the results for the explosion samples; circles (SPE) are the results for speech samples; full symbols (DRY) are results without reverberation

usually smears the energy over time as well. But when are the changes to a sound source audibly characteristic of occlusion? In Section 3.10 this general question was reduced to the specific experimental question, "When can a human listener confidently detect the presence of an occluder in a blind listening test?" Because the SPL at the source was varied from trial to trial, the SPL at the receiver position was an unreliable cue for the detection of occlusion in that test. Under these circumstances, occlusion effects were somewhat difficult to identify. One goal was to develop a filtering model that could effectively synthesize signals that would sound more clearly as if they resulted from occlusion. Another goal was to develop efficient means for rendering the spatial image of an occluded virtual sound source in order to minimize computational cost while maximizing effect audibility.

In contrast to the goal of reducing rendering cost by selectively eliminating reflectors and occluders from the rendering equation [Martens and Herder, 1999], the motivation was to simplify the rendering algorithm, and thereby allow more sonic effects to be rendered.

An alternative approach to rendering the complex spectro-temporal phenomena associated with occlusion is to reduce the effects to simple changes in sound source gain. This approach was taken by [Takala and Hahn, 1992], who set gain via a scaling factor proportional to the amount of occlusion. What is the rationale for such gross simplification? In many multimedia applications, as well as in conventional audio production for sound effects, the purpose of audio rendering is to illustrate an event rather than to attempt an acoustically accurate simulation [Takala and Hahn, 1992]. Given that occlusion effects can be difficult to identify in blind listening tests [Martens and Herder, 1999], there seems to be some justification for such extreme rendering simplifications. On the other hand, there are changes in tone color of the direct sound that are characteristic of some occlusion effects. In contrast to highly detailed physical solutions, however, such as those based upon Fresnel zones [Tsingos and Gascuel, 1998], a solution was desired that was driven rather by perceptually-defined specifications. As a more detailed alternative to the simplification that reduces occlusion effects to changes in gain, this section presents a filtering model of low computational cost that can produce distinctive auditory spatial images associated with identifiable occlusion effects.

First, many acoustical measurements of an actual occluder were made in an anechoic chamber in order to characterize such stimuli. Second, a digital filtering model was designed to capture the variation in occluder impulse responses that was observed as the occluder size, position, and orientation were varied.

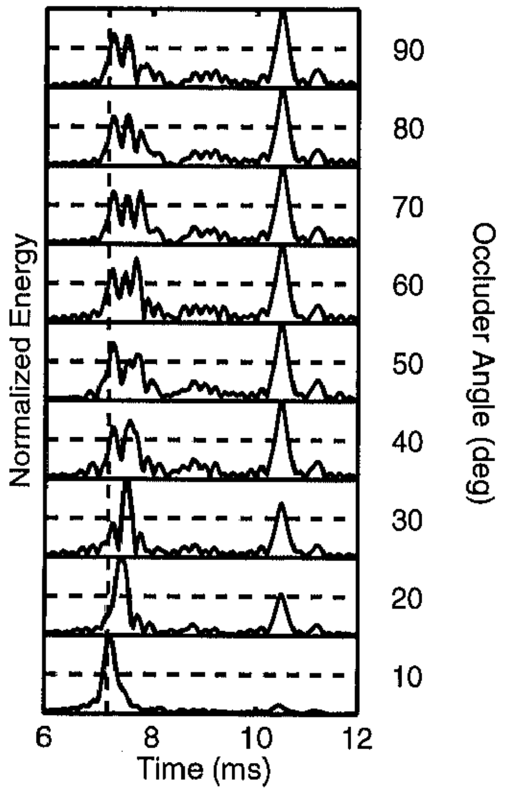

Figure 3.7: Time-domain responses at nine occluder angles

3.9.1 Acoustical measurement of occlusion

The acoustical responses of obstructions of various sizes were measured at various positions and orientations relative to the location of a loudspeaker and a microphone. Figure 3.7 shows the changes in the response of a small, rectangular board of 40 by 90 cm as its orientation varied in relation to a sound source and sink that were each located 100 cm from the center of the board. The curves plot normalized energy over time for the response at nine occluder angles. The 90° angle provided the maximum obstruction to sound in this set of measurements, while the 10° angle presented the smallest silhouette along the line of sight between source and sink. The vertical dashed line marks the time at which the normalized energy peaks for an unobstructed sound arriving at the microphone position via direct path from the loudspeaker. Note that the amplitude of the initial peak observed at around 7 ms decreases with occluder angle while that of the secondary peak observed at around 10.5 ms increases with occluder angle (with increasing obstruction). The absolute peak energy is also declining with elevation, though that detail is obscured in the figure by normalizing energy for each measurement according to the maximum of each curve.

The magnitude response over frequency for the nine occluder angles (not shown here) showed substantial attenuation in a broad band centered at 5 kHz, but showed virtually no attenuation around 8 kHz. At 10°, very little attenuation was observed at frequencies below 8 kHz. The small, rectangular occluder almost always shows a boost in gain at frequencies ranging from .1 to 1 kHz, even at a 20° angle that presents a very small silhouette along the line of sight between microphone and speaker.

3.9.2 Filter design for occlusion and reflection simulation

The filtering model was designed to capture the sonic effects of a modeled obstruction, whether the sound source is reflected or occluded by that obstruction. Because obstructions located near the direct-sound path are always switching between these two states, a comprehensive solution was required. The filter operates by tapping the audio input buffer containing the sound-source audio samples either earlier or later than the tap for the direct-sound signal. If the obstruction is oriented as a reflector, then the obstruction signal is delayed relative to the direct sound, and receives a high-pass emphasis. If, however, the obstruction is oriented as an occluder, then the obstruction signal will lead the direct sound, and is given a low-frequency emphasis.

Characteristic magnitude responses resulting when obstruction signals are

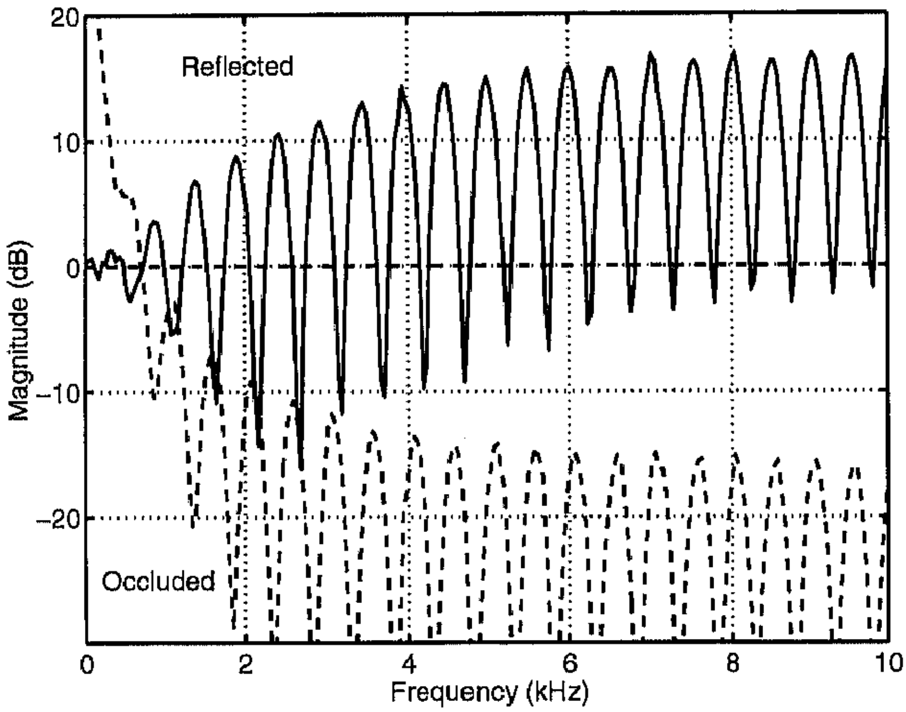

Figure 3.8: Filter model magnitude response for reflected and occluded sound

combined with direct-sound signals are shown in Figure 3.8. For the relatively small obstruction simulated here, the upper, solid curve shows that the high-frequency portion of the direct sound is reflected at higher gain. In contrast, the lower, dashed curve shows that the low-frequency portion of the occluded direct sound is amplified in the model, somewhat exaggerating the increase in gain observed in measured responses. In both cases, the depth of the comb pattern in the magnitude response increases with increasing frequency.

Attenuation of DC energy for early reflection simulation is common [Kendall *et al.*, 1986], but a single filter structure that can smoothly transform reflection effects into occlusion effects is novel and advantageous.

The filtering model described here allows for smooth transition from occluder to reflector state, easing processing control by handling the two states between which obstructions are always switching. Perceptual evaluation confirmed that the model enables the creation of a continuous range of obstruction effects that also produces satisfying auditory spatial imagery.

A criteria for removing an occluder from a rendering process is suggested in the next section.

3.10 Perceptual criteria for eliminating reflectors and occluders from the rendering of environmental sound

This section discusses determination of effective means for choosing which components of the rendering of a reflector or a occluder would provide the most audible differences for spatial sound imagery. Rather than begin with an analytic approach that attempts to predict audible differences on the basis of objective parameters, subjective tests of how audibly different the rendering result may be heard to be when that result includes two types of sound obstruction (reflectors and occluders) give an approximate answer. Single-channel recordings of 90 short speech sounds were made in an anechoic chamber in the presence and absence of these two types of obstructions, and as the angle of those obstructions varied over a 90° range (see Figure 3.9).

In two listening experiments, these recordings were reproduced over a single loudspeaker in the anechoic chamber, and listeners were asked to rate how confident they were that the recording of each of these 90 stimuli included an obstruction. In a parallel experiment, listeners recorded confidence ratings for the presence of an occluder when the stimulus presentation included simulated reverberation. Because the sound source varied from trial to trial in these experiments, confidence ratings were scattered, but clearly modulated

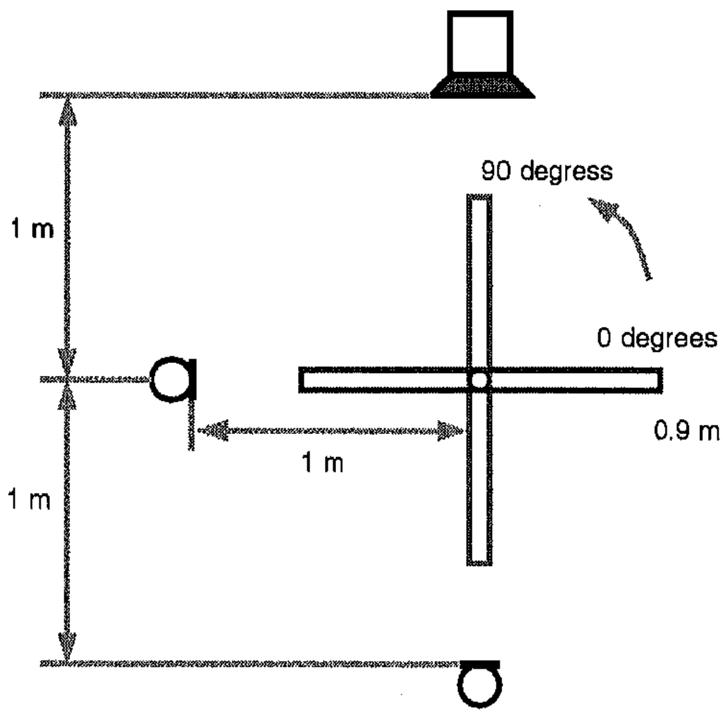

Figure 3.9: Reflector and occluder recording setup

by obstruction parameters. Although obstruction identification was already difficult under anechoic listening conditions, embedding the test signal in simulated reverberation did not reduce identification performance. Thus, at least for obstruction effects that operate within the first few milliseconds of direct sound arrival time, the obtained confidence ratings provide a basis for an evaluation function that predicts which reflectors and occluders might be most important for rendering.

Resource management in the rendering of spatial sound [Herder and Cohen, 1997b] is important for efficient synthesis of realistic sound imagery. Complex environments can contain many sonic obstructions, including objects that occlude direct sound (interposed between sound source and sink), and objects that reflect sound (producing an additional sound signal that delays, attenuates, and combines with the direct sound traveling from source to sink). Of course, some of these sonic effects are not significant, but predicting which will be perceptible is a difficult problem [Begault, 1996]. Besides the indirect sound associated with the walls, ceiling, and floor of a modeled space, spatial audio rendering for complex environments must include considerations of components for objects located within the enclosure, such as reflectors and occluders. In contrast to enclosure-related reflections that arrive at the listening position with longer delays, the contribution of such sonic obstructions to auditory spatial image formation is predominantly focussed on characteristics of the source, such as tone coloration [Barron, 1971].

Two questions were addressed in the experiments reported here. The first question concerns the audibility of obstruction effects when correlated parameters such as loudness are unreliable cues to the presence or absence of an obstruction: Are obstruction effects *per se* identifiable in blind listening tests? The second question concerns the potential masking effects that subsequently arriving indirect sound might have on the audibility of obstruction effects: Are obstruction effects less identifiable when the presentation includes (simulated) reverberation?

3.10.1 Method

For the first two experiments, a constructed set of 360 stimuli for four separate listening sessions was used for testing, in which the angle of an obstruction (i.e., reflector or occluder) was varied over a 90° range. The presence or absence of these obstructions was randomly varied from trial to trial, there being a 50% probability that a given trial would contain only the unobstructed direct sound. The stimuli were presented in an anechoic chamber over a small loudspeaker located directly ahead of the listener in a distance of 230 cm. The sound source used was also different for every trial, taken from a set of anechoic speech samples (Japanese numbers ranging from 1 to 90) that had been recorded by one of five human speakers. The listening sessions were conducted separately. In each listening session of 90 stimuli, the initial angle of the obstruction was zero degrees (which means perpendicular incidence of the direct sound on the obstruction surface). In order to minimize uncertainty, and thereby maximize the likelihood of correct identification of the obstruction, the obstruction angle was changed gradually from trial to trial. So in each successive trial, the obstruction was rotated by one degree relative to the previous trial's obstruction angle. Thus the 45th trial contained a reflector angle that most closely provided a specular reflection for the sound source (i.e., the angle formed by incident and reflected sound rays was split by the surface normal of the reflector). In the second and final listening session under each condition, the complementary set of trials was presented. For example, if in the first session a reflection was presented in the 10th trial at a reflection angle of 10° , then in the second session the dry source would be presented in the 10th trial. Listeners were asked to estimate the presence or absence of an obstruction using a scale from 1 to 5. If they were confident that the obstruction was present, then the response category reported was "5." If they were confident that the obstruction was absent, then the response category reported was "1."

For the third experiment, a unique sound source was again employed on each trial; however, only four obstruction configurations were included in the stimulus set. Here again, listeners rated their confidence that an occluder had been present or absent, but they heard the sound stimulus in three reverberant contexts: under anechoic conditions, with the addition of reverberation simulating a large room, and with reverberation simulating a small room.

3.10.2 Results

Figure 3.10 shows the average confidence ratings made by four listeners in the presence (circular symbols) and absence (diamond symbols) of a reflector, plotted as a function of reflector angle. A third-order polynomial was fit to the average ratings for each of these conditions, and the distance between these two curves gives an indication of how easily the sound of the reflector could be identified. Note that the highest ratings were given for an incidence angle of 45° (close to a specular reflection).

Figure 3.11 shows the average confidence ratings made in the presence and absence of an occluder, again plotted as a function of the angle of the obstruction, and using (circular symbols for occluder presence and diamond symbols for absence). However, in this case, the zero degree angle corresponds to the

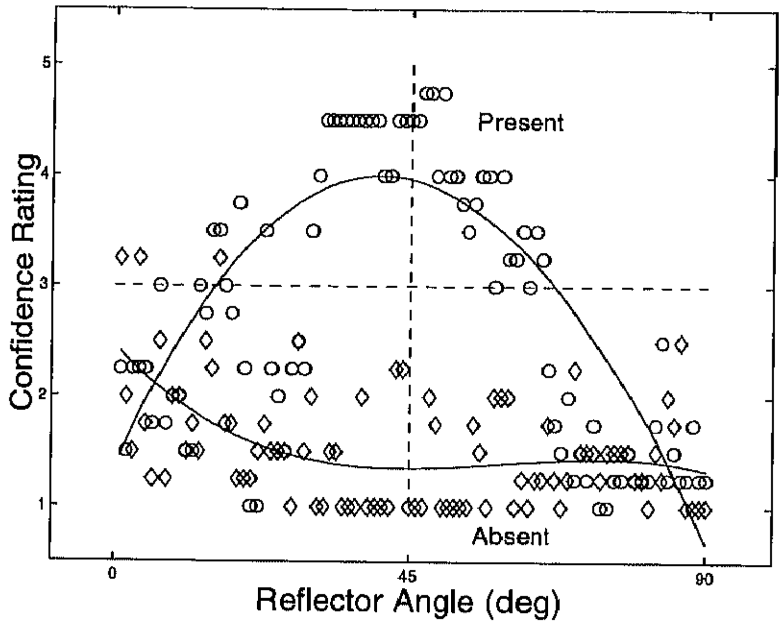

Figure 3.10: Identification of reflector presence

spatial configuration in which the maximum width of the occluder is interposed between source and sink. Thus, ratings were highest for low angles (i.e., greatest blocking). The smooth curves fit to these data converge only at the highest angles (near 90° incidence).

Figure 3.12 shows the relative frequency of each response for four listeners of the third experiment. With no evidence of a ceiling or floor effect, it appears that identification performance is not degraded by the presence of simulated reverberation.

3.10.3 Discussion

How can these data be used to determine which reflectors and occluders are most important to render? As stated before, the intention is to render only those that combine with the direct sound to provide an identifiably different spatial impression in comparison to the image of the direct sound alone. An arbitrary threshold confidence rating of "3" (marked by the horizontal dotted line shown in Figures 3.10 and 3.11) to choose which obstructions should be rendered might be used, but this approach does not take into account the number of rendering elements for which resources are available. An alternative is to choose the obstructions for rendering from a set containing both reflectors and occluders, using the standardized confidence rating scale as a

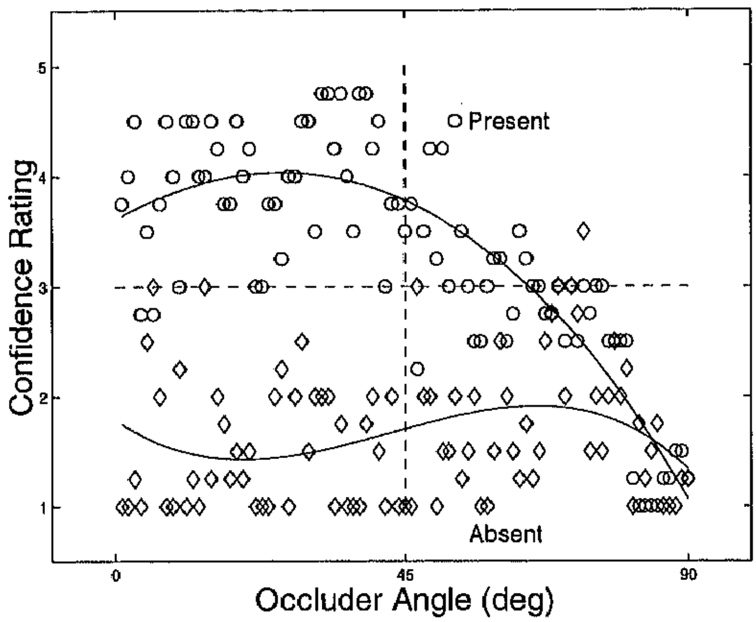

Figure 3.11: Identification of occluder presence

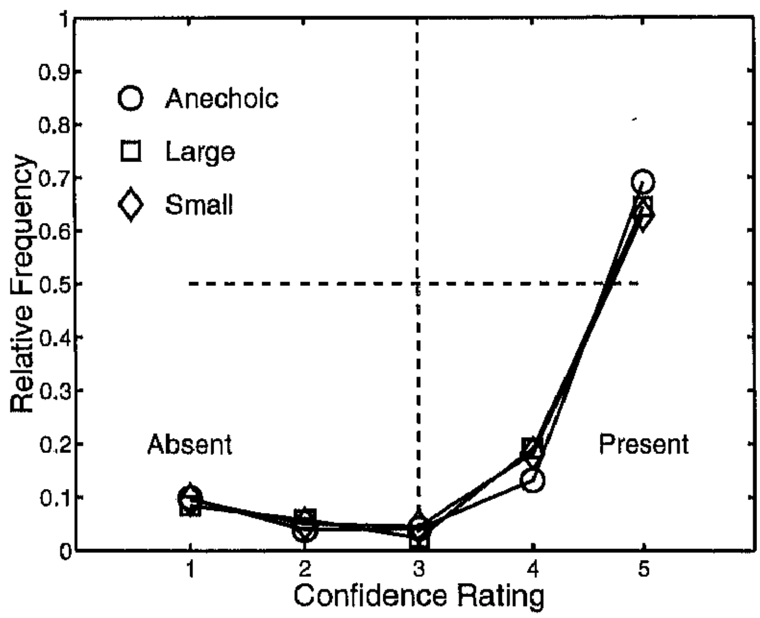

Figure 3.12: Identification of occluder presence in reverberation

common means for setting the criterion.

3.11 Sound spatialization with processing elements

A hardware architecture for sound spatialization proposed by [Ikedo and Martens, 1999] has processing elements for each sound path. The design includes obstruction transfer function. Such a system has processing elements as shown in Figure 3.13.

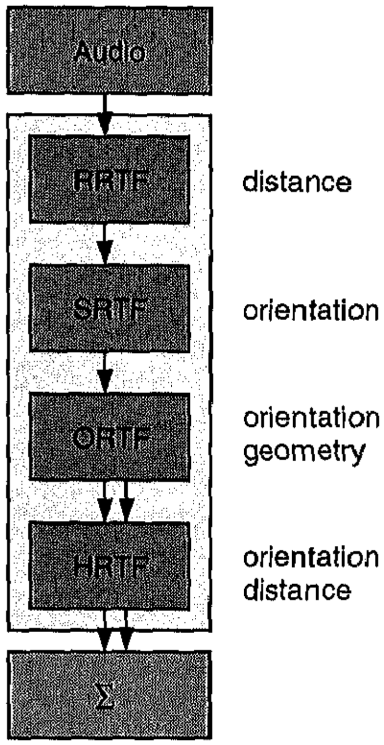

Figure 3.13: Processing element for one sound pass including control parameters

Range radiation transfer functions (RRTFs) model distance effects, parameterized by the distance between source and sink. Source radiation transfer functions model the radiation pattern of the sound source with orientation of the source relative to the sink as parameter. The hardware architecture can handle only a single obstruction. In case the obstruction is a reflection, then for each reflection, one processing element is used. The obstruction (occluder or reflector) transfer function (ORTF) has as parameters the geometry and orientation. The head-related transfer function (HRTF) uses orientation of the sink and distance as parameter. The responsibility of the resource manager is to allocate those processing elements and distribute control parameters.

3.12 Resource management for occluding objects

A processing element like that suggested in the previous section can handle only a limited number of occluders (i.e., one). A resource management system can select relevant occluders using Algorithm 3 and perceptual criteria investigated in Section 3.10.

Algorithm 3 Selecting occluder for sound sink path

```
for each sink in active sinks do
  for each source in active sources of sink do
    find all occluders along sink sound path
    calculate relevance based on perceptual criteria and geometry
    sort by relevance
    select most relevant occluders as resources are available
  end for
end for
```

Bibliography

[Anderson and Casey, 1997] David B. Anderson and Michael A. Casey. The sound dimension. *IEEE Spectrum*, 34(3):46–50, March 1997.

[Barron, 1971] M. F. E. Barron. The subjective effects of first reflections in concert halls — the need for lateral reflections. *Journal of Sound and Vibration*, 15:475–94, 1971.

[Begault, 1994] Durand R. Begault. *3-D Sound for Virtual Reality and Multimedia.* Academic Press, 1994. ISBN 0-12-084735-3.

[Begault, 1996] Durand R. Begault. Audible and inaudible early reflections: thresholds for auralization system design. In *Audio Engineering Society 100th Convention*, Copenhagen, 1996. Preprint 4244.

[Bell *et al.*, 1997] Gavin Bell, Rikk Carey, and Chris Marrin. ISO/IEC 14772-1:1997: The Virtual Reality Modeling Language (VRML97), 1997. `http://www.vrml.org/Specifications/VRML97/`.

[Blauert, 1968] Jens Blauert. Ein Beitrag zur Theorie des Vorwaerts - Rueckwaerts - Eindruckes beim Hoeren. In *The 6th International Congress on Acoustics*, pages A – 45–48, Tokyo, Japan, August 1968. (In German).

[Blauert, 1996] Jens Blauert. *Spatial Hearing: The Psychophysics of Human Sound Localization.* MIT Press, revised edition, 1996. ISBN 0-262-02413-6.

[Dhillon, 1994] Navdeep S. Dhillon. Cellular approach for modeling room acoustics: A framework for implementations based on the ray tracing algorithm. Master's thesis, University of Wisconsin - Madison, August 1994. `http://home1.gte.net/dhillos/camra/index.html`.

[Ellis, 1998] Sean Ellis. Towards More Realistic Sound in VRML. In *VRML 98*, pages 95–100, Monterey CA, USA, 1998.

[Funkhouser *et al.*, 1998] Thomas Funkhouser, Ingrid Carlbom, Gary Elko, Gopal Pingali, Mohan Sondhi, and Jim West. A beam tracing approach to acoustic modeling for interactive virtual environments. In *SIGGRAPH 98 conference*, held in Orlando, Florida, July 1998.

[Green and Swets, 1966] David M. Green and John A. Swets. *Signal Detection Theory and Psychophysics.* John Wiley & Sons, 1988 reprint edition, 1966. ISBN 0-932146-23-6.

[Handel, 1989] Stephen Handel. *Listening: An Introduction to the Perception of Auditory Events.* MIT Press, 1989. ISBN 0-262-08179-2.

[Herder and Cohen, 1997a] Jens Herder and Michael Cohen. Enhancing perspicuity of objects in virtual reality environments. In *CT'97 — Second Int. Cognitive Technology Conf.*, pages 228–237. IEEE, IEEE Press, August 1997. ISBN 0-8186-8084-9.

[Herder and Cohen, 1997b] Jens Herder and Michael Cohen. Sound Spatialization Resource Management in Virtual Reality Environments. In *ASVA'97 — Int. Symp. on Simulation, Visualization and Auralization for Acoustic Research and Education*, pages 407–414, Tokyo, Japan, April 1997. The Acoustical Society of Japan (ASJ).

[Ikedo and Martens, 1999] Tsuneo Ikedo and William L. Martens. Technologies for creating realistic sights and sounds in future multimedia systems. 1999. in preparation.

[Ishikawa *et al.*, 1998] Kimitaka Ishikawa, Minefumi Hirose, and Jens Herder. A sound spatialization server for a speaker array as an integrated part of a virtual environment. In IEEE *YUFORIC Germany 1998*, Stuttgart, June 1998. `http://www-ci.u-aizu.ac.jp/~herder/publications/ve98-spatial-server/`.

[Kendall and Martens, 1984] Gary S. Kendall and William L. Martens. Simulating the cues of spatial hearing in natural environments. In *ICMC: Proc. Intnl. Comp. Music Conf.*, pages 111–126, Paris, 1984. Computer Music Association.

[Kendall *et al.*, 1986] Gary S. Kendall, William L. Martens, Daniel J. Freed, M. Derek Ludwig, and Richard W. Karstens. Image model reverberation from recirculating delays. *Proceedings of the Audio Engineering Society 81st Convention*, 1986. Preprint No. 2408 (C-17).

[MacKenzie, 1995] I. Scott MacKenzie. Input devices and interaction techniques for advanced computing. In Woodrow Barfield and Thomas A. Furness III, editors, *Virtual Environments and Advanced Interface Design*, pages 436–470. Oxford University Press, 1995. ISBN 0-19-507555-2.

[Makous and Middlebrooks, 1990] James C. Makous and John C. Middlebrooks. Two-dimensional sound localization by human listeners. *JASA*, 87(5):2188–2200, May 1990.

[Martens and Herder, 1999] William L. Martens and Jens Herder. Perceptual criteria for eliminating reflectors and occluders from the rendering of environmental sound. In *Proc. Joint Meeting of the* 137^{th} *Regular Meeting of the Acoustical Society of America and the* 2^{nd} *Convention of the European Acoustics Association: Forum Acusticum*, page CDROM, Berlin, March 1999. Acoustical Society of America (ASA), and European Acoustics Association (EAA). Signal Processing in Acoustics and Psychological and Pysiological Acoustics: Auditory Displays, 1pSP2.

[Martens *et al.*, 1999] William L. Martens, Jens Herder, and Yoshiki Shiba. A filtering model for efficient rendering of the spatial image of an occluded virtual sound source. In *Proc. Joint Meeting of the* 137^{th} *Regular Meeting of the Acoustical Society of America and the* 2^{nd} *Convention of the European Acoustics Association: Forum Acusticum*, page CDROM, Berlin, March 1999. Acoustical Society of America (ASA), and European Acoustics Association (EAA). Signal Processing in Acoustics and Psychological and Pysiological Acoustics: Auditory Displays, 1pSP7.

[Mulgund and Zacharias, 1996] Sandeep S. Mulgund and Greg L. Zacharias. A situation-driven adaptive pilot/vehicle interface. In *Proceedings: 3rd Annual Symposium on Human Interaction with Complex Systems — HICS'96*, pages 193–198. IEEE, IEEE Computer Society Press, 1996. Dayton, August 25-28.

[Ovan and Havens, 1993] Russell Ovan and William S. Havens. Intelligent mediation: An architecture for the real-time allocation of interface resources. In *Proceedings of the 1993 International Workshop on Intelligent User Interfaces*, pages 55–61. ACM SIGCHI, ACM Press, 1993. Orlando, Florida, January 4-7, 1993.

[Sanders and McCormick, 1987] Mark S. Sanders and Ernest J. McCormick. *Human Factors in Engineering and Design.* McGraw-Hill, New York, sixth edition, 1987. ISBN 0-07-044903-1.

[Schiffman, 1995] Harvey Richard Schiffman. *Sensation and Perception: An Integrated Approach*, chapter 13, pages 350–363. John Wiley & Sons, Inc., 4th edition, 1995. ISBN 0-471-58620-X.

[Stille *et al.*, 1996] Stefan Stille, Shailey Minocha, and Rolf Ernst. A^2DL — An Adaptive Automatic Display Layout System. In *Proc. 3rd Annual Symposium on Human Interaction with Complex Systems — HICS'96*, pages 243–250. IEEE, IEEE Computer Society Press, 1996. Dayton, August 25-28.

[Sukaviriya and Foley, 1993] Piyawadee "Noi" Sukaviriya and James D. Foley. Supporting adaptive interfaces in a knowledge-based user interface invironment. In *Proceedings of the 1993 International Workshop on Intelligent User Interfaces*, pages 107–113. ACM SIGCHI, ACM Press, 1993. Orlando, Florida, January 4-7, 1993.

[Suzuki, 1999] Ikumi Suzuki. Elevation identification performance for real sound sources vs. HRTF-processed virtual sources. Bachelor thesis, University of Aizu, 1999.

[Takala and Hahn, 1992] Tapio Takala and James Hahn. Sound rendering. *Computer Graphics*, 26(2):211–220, July 1992.

[Tsingos and Gascuel, 1997] Nicolas Tsingos and Jean-Dominique Gascuel. Soundtracks for computer animation: Sound rendering in dynamic environments with occlusion. In *Graphics Interface '97*, pages 9–16, 1997.

[Tsingos and Gascuel, 1998] Nicolas Tsingos and Jean-Dominique Gascuel. Fast rendering of sound occlusion and diffraction effects for virtual acoustic environments. In *the 104th Convention of the AES*, Copenhagen, May 1998. Preprint 4699 (P4-7).

[Wenzel, 1994] Elizabeth M. Wenzel. Spatial sound and sonification. In Gregory Kramer, editor, *Auditory Display*, volume XVIII, pages 127–150. Addison-Wesley, April 1994. ISBN 0-0201-62603-9.

[Williams, 1994] Sheila M. Williams. Perceptual principles in sound grouping. In Gregory Kramer, editor, *Auditory Display*, volume XVIII, pages 95–125. Addison-Wesley, April 1994. ISBN 0-0201-62603-9.