16

Virtual Hearing

Karl Wiklund and Simon Haykin

McMaster University, Canada

With the introduction of digital hearing aids, designers have been afforded a much greater scope in the nature of the algorithms that can be implemented. Such algorithms include designs meant for speech enhancement, interference cancellation, and so on. However, as the range of these algorithms increases, and as greater attention is given to the problems in real acoustic environments, designers are faced with a mounting problem in terms of how to test the algorithms that they have produced. The existing test procedures such as HINT[1] [20] and SPIN[2] [16] do not adequately reflect the problems that many new algorithms were designed to cope with. However, testing under real conditions must involve the problems of reverberation, different signal types (e.g. speech, music, etc.) and multiple spatially distributed interferers. These interferers may also become active and inactive at random intervals.

The acoustic environment however is not the only obstacle that many researchers face. A further variable in their tests exists in the form of the patients themselves. An algorithm that produces a good result according to a common error metric such as the mean-squared error, may not appreciable improve the patients ability to understand speech. Not only must a proposed algorithm be tested against human patients, but it must be tested against a broad range of patients. Hearing impairment after all comes in many different forms and degrees, so it is essential that the performance of an algorithm be determined for different types of patients. Ideally, one ought to be able to tune the algorithm under test so as to offer the best level of performance for a given patient.

Both of the problems outlined above are not insurmountable. It is possible to conduct tests in real acoustic environments, just as it is possible to arrange for human trials. However, such tests can be expensive, time consuming, and may require specialized equipment or even dedicated laboratory space. In

[1] *HINT* abbreviates *hearing in noise test.*
[2] *SPIN* stands for *speech perception in noise.*

general, the time involved in such testing also limits the turn-around time between designing, testing, and (if necessary) correcting the design.

16.1 Previous Work

Overcoming these difficulties have motivated us to investigate the development of software environments for testing hearing aids. The initial result of such investigation was the R-HINT-E[3] package, which was described in [30] and [34]. This package was developed after careful consideration both of our needs as researchers and of the methods currently used in both architectural acoustics and in the newer field of virtual acoustics. In our case, we were interested in rendering both the acoustic effects of the room and of the human body. In addition, we wished to present the result to a pair of simulated "ears" which would match the physical configuration of the hearing aid under test. The sound impinging on the device was then to be processed according to some user-defined algorithm.

Our particular needs therefore included both the physical realism of the presentation as well as the flexibility of the software simulation. Different source positions and even cocktail-party situations needed to be incorporated into the range of possible scenarios in order for the software to be an effective testing platform. In addition, the effects of different room acoustics needed to be considered, so the inclusion of multiple acoustic environments became an important requirement of the system's design.

The software model we ultimately followed was chosen on the basis of both our requirements and on the practicality of the system. Unlike multimedia based virtual audio systems [27], there was no need for real-time performance, nor was there a strong concern regarding the storage and retrieval of large amounts of data. This did not mean of course, that there were no time-constraints on the simulation's performance. The simulation methods used in architectural acoustics applications (such as ray-tracing or the image-source method) were detailed, but also very slow. Our need for physical accuracy as well as for reasonable performance times led us to use pre-recorded room impulse response measurements.

The recordings themselves were made using custom-built, flat-response loudspeakers. The source signal was an exponentially swept chirp signal [19], which was swept over a range of 0-22050 Hz and a duration of 1486 ms. By recording the system response at the microphones, it was possible to use the known source signal to obtain an estimate of the acoustic impulse response. This was done by straightforward deconvolution in the manner of

$$H\left(e^{j\Omega}\right) = \frac{Y\left(e^{j\Omega}\right)}{S\left(e^{j\Omega}\right)}\,. \tag{16.1}$$

[3] *R-HINT-E* stands for *realistic hearing in noise test environment.*

In the above equation, $Y(e^{j\Omega})$ is the Fourier transform of the measured response, and $S(e^{j\Omega})$ is likewise the Fourier transform of the input source signal. The acoustic impulse response can then be recovered by taking the inverse Fourier transform of $H(e^{j\Omega})$.

For the sake of convenience, we combined the room impulse responses measurements with the measurements of the head-related transfer function by using a KEMAR[4] dummy as our recording platform (see Fig. 16.1). The dummy was supplied by Gennum Corporation, which also supplied the microphone system used in our experiments. This system used three Knowles FG microphones arranged horizontally in each ear of the mannequin. The combined impulse response was then recovered for each microphone using the method of Eq. 16.1.

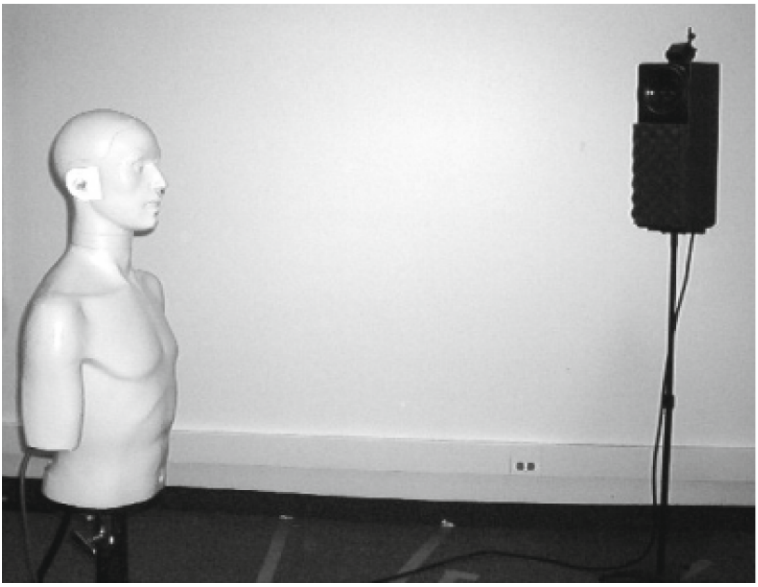

Fig. 16.1. The R-HINT-E recording setup (photo by Ranil Sonnadara).

For the R-HINT-E system, measurements were made from multiple positions in order to approximate realistic acoustic scenarios. In particular, measurements were taken from 12 different angles (0°, 22.5°, 45°, 67.5°, 90°, 135°, 180°, 225°, 270°, 292.5°, 315°, 337.5°), three heights (1'6.5" = 47.0 cm, 4'6" =137.2 cm, 5'5" = 165.1 cm), and from 2 distances (3' = 91.4 cm and 6' = 182.9 cm) for a total of 72 different source-receiver configurations (see Fig. 16.3). These measurements were carried for three different environments representing differing levels of room reverberation. These environments included a small

[4] The term *KEMAR* abbreviates *Knowles electronic manikin for acoustic research.*

room with movable velour drapes, as well as a hard-walled, reverberant lecture room. The first two environments were created in the small room by changing the drape positions from open to closed, where the closed drapes were used to reduce the reverberation time of the room.

Fig. 16.2. A close up of KEMAR's right ear. The three microphones are placed in the ear and arranged in a horizontal fashion (photo by Ranil Sonnadara).

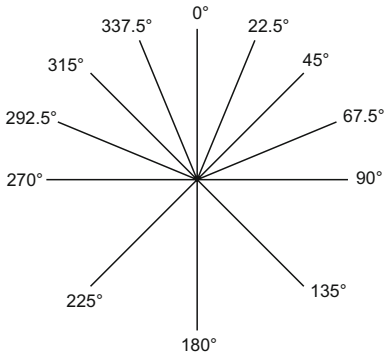

Fig. 16.3. The source positions for the R-HINT-E model.

A GUI[5] front end (see Figs. 16.4 and 16.5) for R-HINT-E was later created using MATLAB, which allowed an experimenter to create custom acoustic scenarios. Using the software we developed, a user could choose the desired room type, source position and source signal in order to create the desired conditions. In addition, multiple signals could be distributed in space to create a 'cocktail party effect' in order to simulate realistic acoustic environments. Additional flexibility was incorporated in the form of allowing the user to test custom hearing aid algorithms within the GUI, as well as allowing the user to provide their own room impulse response (RIR) measurements given a standard format. Using this software also makes it easier to perform large-scale tests involving many different scenarios. This was accomplished through the incorporation of a custom scripting language that could control all of R-HINT-E's major functions.

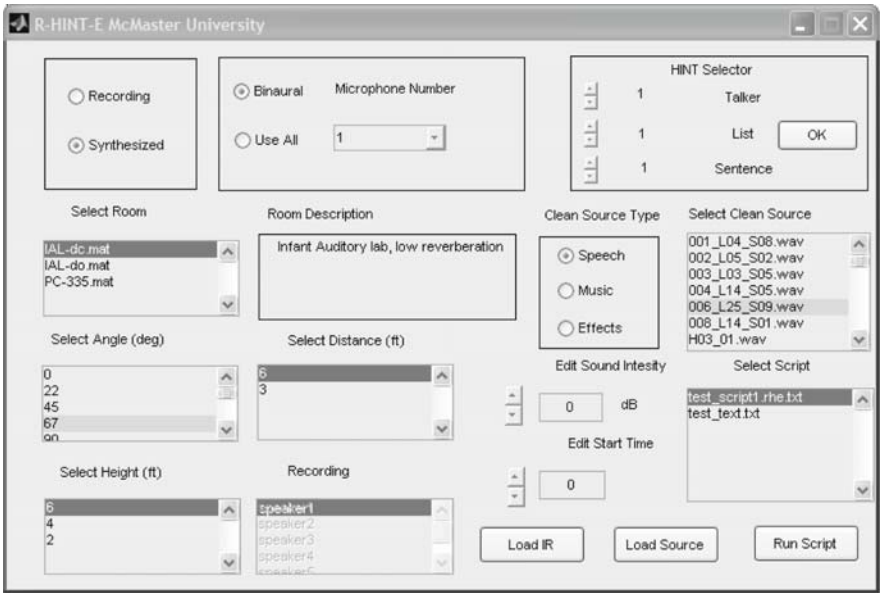

Fig. 16.4. The R-HINT-E main menu.

However, while R-HINT-E is a useful tool with respect to hearing aid research, it is not without its shortcomings. Most important is that by relying on pre-recorded RIRs, we lose a considerable degree of flexibility in the kinds of simulations that can be run. Only a limited number of source-receiver arrangements are possible, and there are even fewer acoustic environments in which one can carry out tests. While it is possible for the user to add new acoustics environments to the R-HINT-E program, this is a time-consuming process,

[5] The term *GUI* abbreviates graphical user interface.

and one that requires the right equipment. Moreover, while R-HINT-E does accurately simulate the room acoustics as well as the processing algorithm of interest, it does not simulate the actual patient beyond the level of the head-related transfer functions or HRTFs. In other words, the simulator cannot provide any information on how useful a processing algorithm might be in terms of patient performance given some level of hearing deficiency.

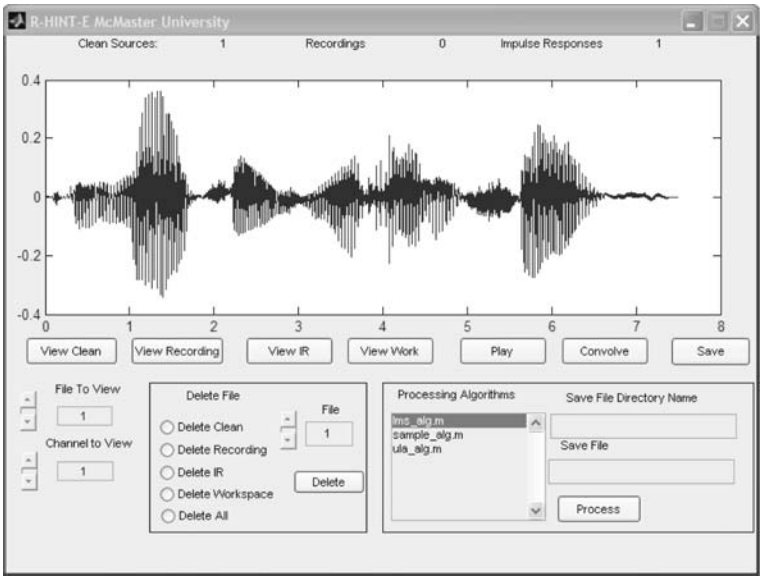

Fig. 16.5. The R-HINT-E processing screen.

Owing to the shortcomings mentioned above, it was decided that the R-HINT-E concept could be extended further to allow not only for greater flexibility in designing acoustic scenarios, but also to encompass the concept of a "virtual patient". Such a virtual patient would involve simulating the pathology of hearing loss in software, which would permit testing and customization of algorithms for specific patterns of hearing deficiency. Ideally, the simulation would also include additional flexibility by incorporating a range of HRTFs, which are known to vary widely from patient to patient.

16.2 VirtualHearing

The VirtualHearing software that we have developed in response to these needs attempts to incorporate all the aforementioned simulation requirements into a single package. That is, it contains modules that allow for the simulation of room acoustics, patient HRTFs, and finally the patient itself. In addition,

a further module also makes it possible for the end user to test their own hearing aid algorithms by loading them as pre-compiled binary files. The basic software model thus follows Fig. 16.6.

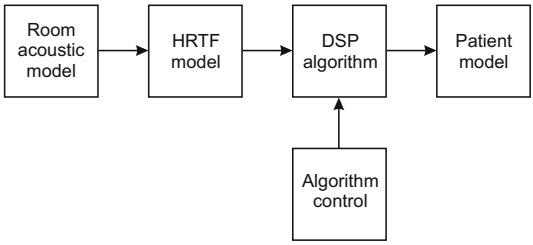

Fig. 16.6. The VirtualHearing software model. If no DSP algorithm is specified, then the output of the HRTF model is sent directly to the patient model.

As Fig. 16.6 implies, this model is a fairly simple one, but allows for individual modules to be specified independently of each other, and to be replaced if necessary in the course of software maintenence. The nature of the individual modules will be discussed in the following sections.

16.3 Room Acoustic Model

The simulation of virtual environments is a fairly well-developed field, and plays an important role in both entertainment and architectural applications. The different requirements of these applications mean that specific implementations will be forced to trade off accuracy vs. speed. Many entertainment applications for example need real-time or close to real-time rendering of acoustic scenes, while on the other hand accuracy is of paramount importance for architects considering the acoustic properties of their designs. The needs of our own software however fall in between these two extremes. In order for the room impulse responses to have a realistic effect on the designers signal processing algorithms, greater accuracy is needed than in the case of multimedia applications where the standard of accuracy is simply perceptual realism. On the other hand, while we do not have the same real-time demands that some multimedia algorithms have, a practical desktop simulator cannot afford to take the time needed to run the most accurate simulations.

In order to meet these constraints, it was decided that a hybrid simulator would be used to implement the room acoustics module. Such an arrangement was chosen because it allowed us to divide the room impulse response into two portions: the early reflections and the late reverberation, which in turn allowed for the use of algorithms best suited for each component. The use of

this method was suggested in [27] by Savioja for use in the DIVA[6] multimedia system [15,26]. Unlike the DIVA system however, which has a greater concern for the architectural acoustics of concert halls, we have chosen to use a much simpler modelling geometry than was implemented in that system. For our purposes, it was deemed sufficient to model rooms as simple rectangular prisms (see Fig. 16.7). The reason for this is that we are solely interested in the

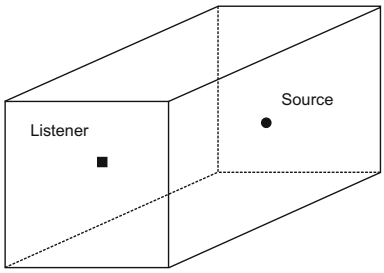

Fig. 16.7. All simulated rooms are modelled as rectangular prisms.

effects of reverberation on hearing, which can be captured sufficiently in simple rooms. Including more complicated geometries would have added nothing to this part of the simulation while greatly increasing the difficulties both in programming and in handling the front-end user interface.

The so-called early reflections are modelled essentially as direct reflections from the source to the receiver. The well known image-source [3] method was used for this effect owing to the simplicity of its implementation and its accuracy. In this method, the reflected sounds are modelled as separate, virtual sources located somewhere outside the bounds of the room (see Fig. 16.8). By adding up the calculated reflection strengths as a series of weighted, time-delayed delta functions, the rooms impulse response can be estimated for any given source-receiver pair.

Working from this model, it is a fairly simple matter to include the attenuation effects resulting from both the distance and the reflection. Additionally, the use of this method also allows us to include the directional effects of the calculated reflections. That is, given the direction from which the virtual source appears to be coming from, it can be convolved with the HRTF for that head direction. Doing this provides a better model for the confusion of binaural cues by room reverberation.

Unfortunately, while the image-source method does provide a good approximation of real room impulse responses, it suffers from a significant problem in that the computational complexity of the algorithm increases exponentially with the order of the reflections to be included in the model. As a result, the use of this algorithm is generally cut off after some more or less arbitrary

[6] *DIVA* abbreviates *digital interactive virtual acoustics*.

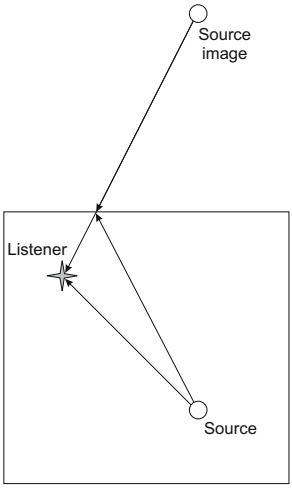

Fig. 16.8. The reflected sound is modelled as a separate source.

reflection order has been passed. Needless to say, such a cut off does not provide a good representation of the real room impulse response, since it ignores the late reverberation.

As a result, a second algorithm was used in conjunction with the image-source method in order to simulate the late reverberation component of the room impulse response. This second algorithm has as a starting point the fact that the late reverberation can essentially be modelled as a diffuse sound field, instead of being made up of a series of discrete reflections. To that end, we investigated the use of various recursive digital filter algorithms [24, 31], and ultimately decided on the one developed by Vaananen et al. [32].

All such algorithms have a roughly similar structure and are required to meet specific criteria relating to the behavior of reverberant sound fields. These criteria are outlined in [27]. In particular, a good digital reverberator algorithm must meet the following criteria:

1. A spectrally dense pattern of reflection with an exponential decay of energy in the time domain.
2. Higher frequencies must have shorter reverberation times than lower frequencies.
3. The late reverberation as perceived by a binaural listener should be partially incoherent.

Of the algorithms we studied, it was found that the Vaananen reverberator not only met these criteria, but it also used less memory than other methods while allowing a much faster growth in reflection density.

The Vaananen algorithm consists of a series of parallel filter blocks and delay lines as shown in Fig. 16.9. The individual filter blocks $H_k(z)$ and $A_k(z)$ are a low-pass filter and an all-pass comb filter respectively:

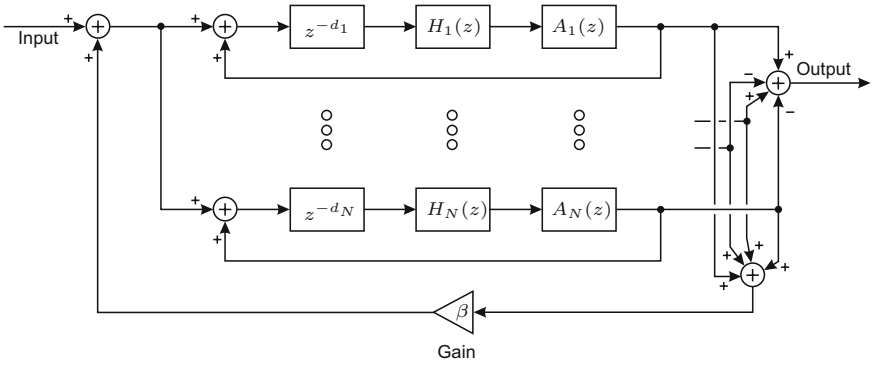

Fig. 16.9. The structure of the Vaananen reverberator consists of a series of parallel delay loops. Each loop contains a lowpass filter and an all-pass comb filter.

$$H_k(z) = \frac{g_k\,(1-b_k)\,z^{-1}}{1-b_k\,z^{-1}} \tag{16.2}$$

$$A_k(z) = \frac{-\frac{1}{2}+z^{-M_k}}{1-\frac{1}{2}\,z^{-M_k}}\,. \tag{16.3}$$

The purpose of the low-pass filter $H_k(z)$ is to mirror the low-pass characteristics of air, while the all-pass comb filter $A_k(z)$ helps to diffuse the reflections, making them more irregular. These filters are of very simple construction and are shown in Fig. 16.10, for a given kth filter in the loop. These filters are governed by three different parameters, b_k and g_k, and the lengths of the individual delay lines.

These parameters are dependent on the nature of the room; in particular they depend on both the room size and the average reflectivity of the rooms walls. The filter delay lengths for example, are based strictly on the room dimensions, with the smallest value being equal to the largest dimension of the room. The lengths of subsequent delay lines are increased, but in an irregular fashion in order to prevent colouration of the reverberant signal.

The remaining parameters are calculated using the reverberation time T_{60}, which can be estimated for the room using Sabines formula [22]

$$T_{60} = K_{T_{60}}\,\frac{V}{\sum\limits_i \alpha_i\,A_i}\,, \tag{16.4}$$

where V is the volume of the room, while a_i and A_i are respectively, the reflection coefficient and area of the ith wall. The coefficient $K_{T_{60}}$ is

$$K_{T_{60}} = 0.161\frac{\mathrm{s}}{\mathrm{m}}\,. \tag{16.5}$$

From the reverberation time, the delay line gains g_i can be calculated [31] using Eq. 16.6, while the low-pass coefficients b_i are calculated using Eq. 16.7:

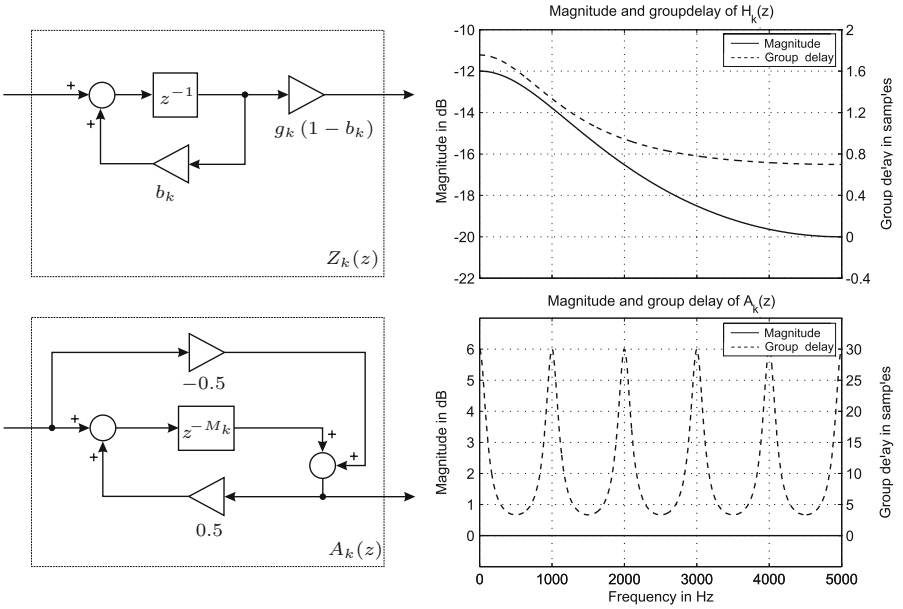

Fig. 16.10. Block diagrams for the low-pass $H_k(z)$ and all-pass diffusive filter $A_k(z)$. $d_k = 1000$, $f_\mathrm{s} = 10000$ Hz, $T_{60} = 0.5$ seconds, $\epsilon = 0.6$

$$g_i = 10^{-\frac{3\,d_i}{f_\mathrm{s}\,T_{60}}}, \tag{16.6}$$

$$b_i = 1 - \frac{2}{1 + g_i^{\left(1-\frac{1}{\epsilon}\right)}}. \tag{16.7}$$

The quantity d_i in Eq. 16.6 is equal to the delay length of the ith delay line in Fig. 16.9, while f_s is the sampling frequency. In Eq. 16.7, the symbol ϵ represents the ratio of the reverberation time for frequencies $f = f_\mathrm{s}/2$ to $f = 0$ Hz. This value cannot be computed directly since it is dependant on the filter coefficients described above; it must instead be specified by the designer of the simulation. For our own purposes, we found that a value of $\epsilon \approx 0.6$ gave results that provided a reasonable approximation the results to be had from several pre-recorded rooms (see Fig. 16.11 and Fig. 16.12).

16.4 HRTF Simulation

The combination of the image-source method with the Vaananen reverberator described above provides an adequate simulation of the acoustics of simple rooms. In addition to this though, the acoustic signal perceived by a human

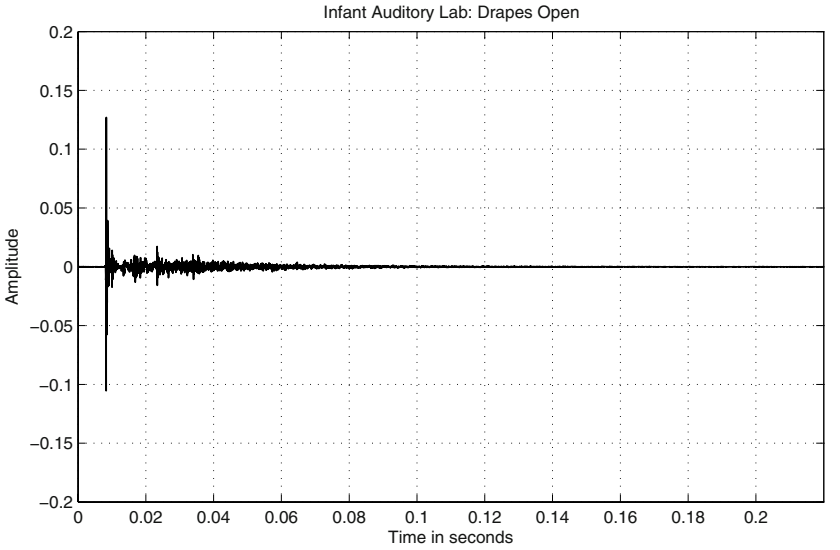

Fig. 16.11. The measured impulse response and HRTF for a real room.

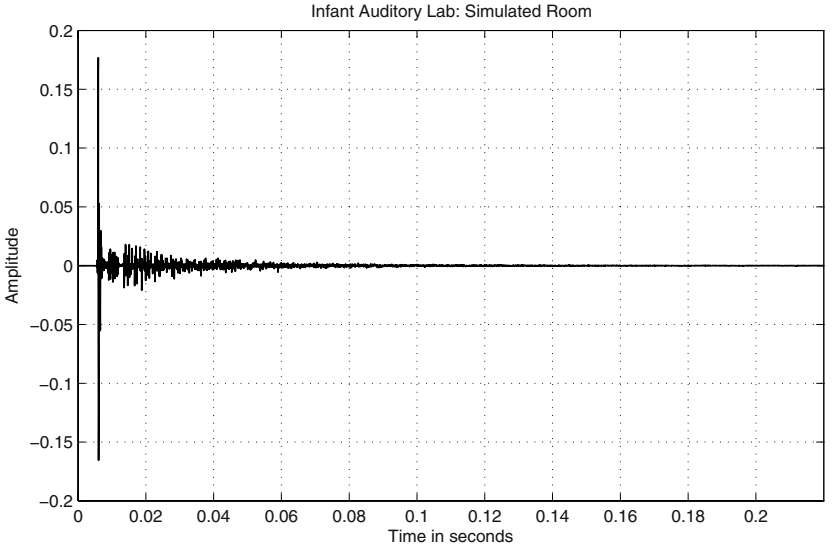

Fig. 16.12. The simulated room impulse response and HRTF using "best guess" parameters to approximate the room used in Fig. 16.11.

listener is also strongly affected by the listeners own HRTFs. These binaural transfer functions arise from the scattering of acoustic energy off of the listeners head, torso and outer ears (pinnae), and are vital in forming the auditory cues needed for the acoustic source localization [1,28]. The inclusion of HRTF effects is therefore vital for any hearing aid simulation, especially for the development of binaural algorithms.

As it happens though, there is considerable variation among the HRTFs for individual listeners, a fact that arises from the diversity of ear and body shapes [28]. In addition, the HRTF also depends on the direction and range of the sound source impinging on the listener. It was important therefore, during the design of this component of the simulator to ensure that these facts were adequately represented in the software. In practical terms this meant first making sure that it was possible to represent the direction dependent HRTFs in sufficient density to approximate real-life scenarios. In addition, it was desirable to include the possibility of simulating different body and pinna shapes.

In the VirtualHearing software package, this problem was solved by making use of a database of pre-recorded HRTFs, albeit one that was more extensive than that used in our previous R-HINT-E project [30,34]. This database was compiled by V.R. Algazi et al [2] at the U.C. Davis CIPIC laboratory, and is publicly available online at http://interface.cipic.ucdavis.edu. The database is comprised of HRTF measurements taken for 43 different individuals, plus two KEMAR sets using both large and small pinnae. For each individual, the measurements include 25 different azimuths and 20 different elevations, with each HRTF being 200 samples in length and sampled at rate of $f_s = 44.1$ kHz.

To make use of this database in our software package, a menu can be accessed that allows the user to specify the subject to use as the simulated listener, while at the same time viewing his/her anthropometric data (see Fig. 16.13). Additionally, the KEMAR models may also be selected instead of using the human subjects provided in the database. A separate options screen allows the user to choose whether to use the binaural HRTFs, or the HRTF corresponding to just one ear. This allows for the development of binaural processing algorithms which use the inputs from both ears in order to enhance the input signal, as opposed to most conventional devices today where the hearing aids in either ear operate independently.

Given the wide range of possible positions of course, one cannot expect even a large HRTF database to include all conceivable directions. In order for there to be adequate coverage of the three-dimensional space for the HRTFs, it is necessary to use interpolation in order to approximate the required HRTF from the ones existing in the database. For this application, we have followed the lead of Huopaniemi [14] in implementing Begault's [4] bilinear interpolation scheme. This method uses the four nearest neighbors of the desired point in order to form the approximation, and is accomplished via

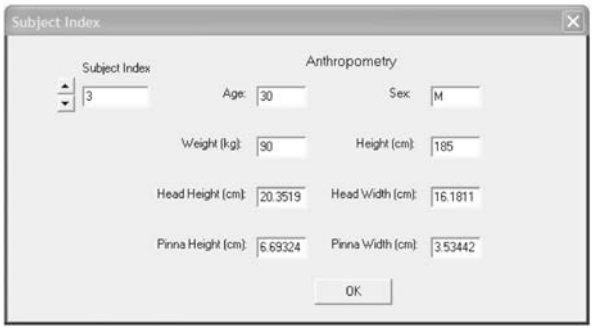

Fig. 16.13. The subject selection window from VirtualHearing allows the user to browse the subjects from whom the HRTFs were selected. Anthropemetric data is also displayed.

$$H_{\mathrm{d}}\left(e^{j\Omega}\right) = \left(1-c_\theta\right)\left(1-c_\phi\right)H_1\left(e^{j\Omega}\right) + c_\theta\left(1-c_\phi\right)H_2\left(e^{j\Omega}\right) + c_\theta\, c_\phi\, H_3\left(e^{j\Omega}\right) + \left(1-c_\theta\right)c_\phi\, H_4\left(e^{j\Omega}\right)\,. \tag{16.8}$$

In the above formula, the known HRTFs ($H_1(e^{j\Omega})$ through $H_4(e^{j\Omega})$) surround the desired HRTF $H_{\mathrm{d}}(e^{j\Omega})$, which is situated at the point $(\theta,\, \phi)$ on the azimuth/elevation grid shown in Fig. 16.14. The interpolation coefficients are computed using

$$c_\theta = \frac{\theta_{\mathrm{d}} \bmod \theta_{\mathrm{grid}}}{\theta_{\mathrm{grid}}}\,, \tag{16.9}$$

$$c_\phi = \frac{\phi_{\mathrm{d}} \bmod \theta_{\mathrm{grid}}}{\theta_{\mathrm{grid}}}\,. \tag{16.10}$$

16.5 Neural Model

In addition to the acoustic modelling discussed in the previous section, the other major task of the software is to model the expected cochlear response of a listener immersed in the acoustic scenario specified by the user. In order for such a simulation to be useful in the evaluation of hearing aids, it ought to match the responses of a real cochlea subjected to the same stimulus. This is not a trivial task however, given the complexity of the human auditory system, which is both non-linear and time-varying in its responses to acoustic stimuli.

In mammals, the detection of acoustic stimuli and their transduction into neural signals occurs in several stages. Vibrations in the air impinge on the eardrum, inducing movement that is transmitted to the cochlea via the action of several small bones connecting the two structures. Within the cochlea,

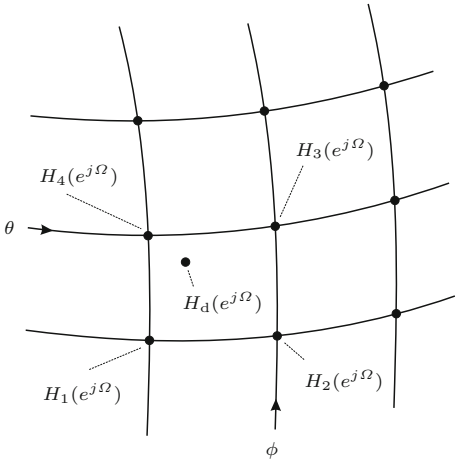

Fig. 16.14. The HRTF for point $H_{\mathrm{d}}(e^{j\Omega})$ must be interpolated from its four nearest neighbours.

the connecting bones induce pressure waves that create small displacements of the basilar membrane [10]. Owing to the mechanical properties of this membrane, local resonances exist that follow a tonotopic pattern [9]. That is, different parts of the basilar membrane respond to different frequencies, with low frequencies causing vibrations at the base of the membrane, and high frequencies at the apex. Detection of these displacements and their conversion into the electrical signals used by the nervous system is carried out by the hair cells that lie along the length of the basilar membrane. As a result of the localized nature of the basilar membrane response, each group of hair cells responds only to a select band of frequencies.

These hair cells themselves may be divided into two types of cell, inner and outer, each of which plays a distinct role in the process of transducing mechanical motions into electrical impulses. The actual transduction in fact, is only carried out by the inner hair cells, while the purpose of the outer hair cells is to act as a feedback mechanism in order to boost the detectability of the incoming acoustic signal, and to sharpen the level of tuning [8]. It is important to note that neither the tonotopic map of the basilar membrane, nor the role of the outer hair cells, are produced by entirely passive mechanisms. The outer hair cells themselves in fact are thought to produce forces that lead to amplification of the stimuli [8, 23]. Also noteworthy is the fact that the ability of the outer hair cells to provide sharp tuning is dependent on the intensity of the stimuli. The tuning is sharpest near the threshold of reception, and broadens as the sound level increases.

The role of the hair cells thus has important implications for modelling hearing loss. Damage to both types of hair cell can occur, but the effects on human auditory perception depend on what amount of damage has been done

to each group, and where on the basilar membrane the losses have occurred. Damage to the outer hair cells for example, will result in both broadened tuning curves and elevated reception thresholds. On the other hand, damage to the inner cells will weaken the ability of the cells to transduce the mechanical stimuli, and will therefore raise the reception thresholds.

In order to produce a computational model of the human auditory system and the effects of hair cell loss, it is possible to think in terms of varying levels of simulation detail. At its most basic level, for example the basilar membrane can be thought of as a filterbank, where the individual filters are described by the frequency-domain gamma-tone function

$$G(\omega) \propto \left[\frac{1}{1 + j\tau(\omega - \omega_{\mathrm{CF}})}\right]^{\gamma} e^{-j\omega\alpha} \tag{16.11}$$

These auditory filters are model-based approximations to the linear revcor functions which were derived empirically [21] and are parametrized by three main quantities. The first of these is ω_{CF}, which simply represents the centre frequency of the filter. The parameter τ controls the filter's time-decay and bandwidth, while α controls the time delay introduced by the filter, which can also be co-opted to model additional system delays if needed. The remaining parameter, γ is simply an empirically-derived value, and does not directly relate to the more physically meaningful parameters.

A simple model of neural transduction can be also be included by passing the bandpass filtered signals through a half-wave rectifier, which reflects the fact that the hair cells only transduce vibrations in one direction. Such a model is sufficient when only general details about the auditory peripheral system are needed, such as in [7] and [13]. For the sort of detailed modeling of the human auditory system that our "virtual patient" system requires however, these kind of approximations are inadequate. The model described above for example, does not include the non-linear time-varying effects that are known to be present in the human auditory system. These effects include for example, the changes in auditory filter bandwidth with input sound pressure level, or the changes in auditory behavior brought about by hearing impairment. For this reason, we have chosen to employ a more detailed model of the human auditory system, which addresses these needs.

In particular, we have incorporated the Bruce-Zilaney [5, 36] model of the cochlea, which is itself a modification of the model developed first by Carney [6] and later modified by Zhang [35]. This model can be broken down in to five main processing blocks: a middle ear filter, a time-varying gamma-tone filter, an OHC[7] control path, an IHC[8] transduction model, and synaptic model. These blocks are organized in the manner shown in Fig. 16.15, where the input is an acoustic stimulus measured in Pascals. The output of the model

[7] *OHC* abbreviates *outer hair cell*.
[8] *IHC* stands for *inner hair cell*.

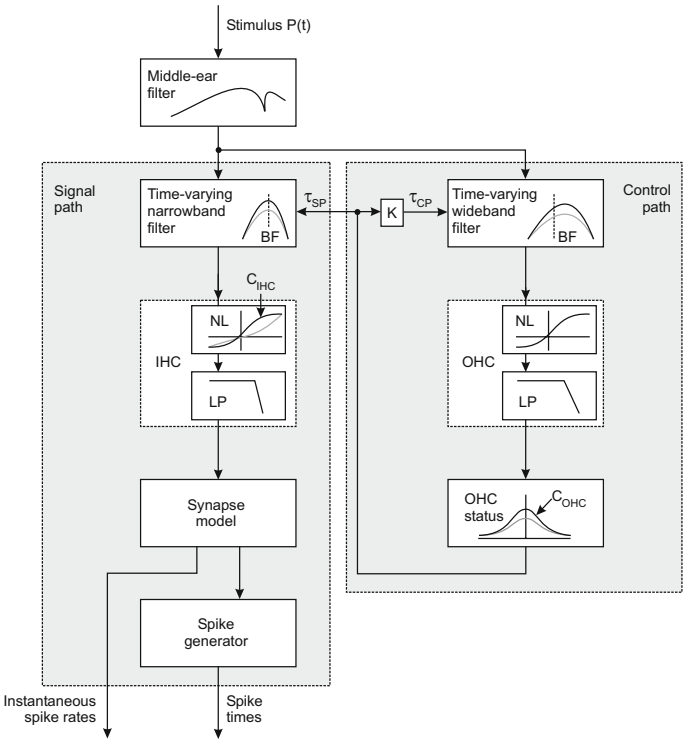

Fig. 16.15. Block diagram of the neural model.

consists of two signals, which include a binary spike train as well as a signal of instantaneous spike rates.

The first actual component of the block, the middle ear filter, is meant to model the filtering aspects of the human ear canal. This is most important for describing wide-band signals, where the relative amplitudes of different frequency components may be changed. The middle ear block may be implemented using a digital filter representation, which is fully described in Appendix A of [5].

Like the simple models discussed earlier in this section, the cochlear model that we make use of also incorporates a bank of gamma-tone filters in order to model the frequency selectivity of the human ear. However, unlike those models, the model of Zilaney and Bruce [36] generates a data-driven control signal that affects the tuning of the gamma-tone filters. This section makes up the control path shown in Fig. 16.15, and models the affects of the OHCs on filter tuning. Specifically, the acoustic input signal is asymmetrically band-pass filtered about the gammatone filter's centre frequency. Afterwards, the resultant signal is rectified and lowpass filtered in a model approximation of the OHC's input/output behavior. An additional nonlinear function maps the

OHC outputs to the gammatone filter's new time constant, which results in a subsequent change in filter bandwidth.

In addition to modeling the time-varying properties of the cochlear filter, this model of the OHC control path is useful in two other significant ways. Firstly, the asymmetric bandpass filtering of the acoustic input allows the incorporation of two-tone suppression effects without needing to directly implement the interactions between separate cochlear filters [35]. In addition, the non-linear function that relates the OHC output to the filter time constants may be easily parametrized in order to describe the effects of OHC impairment. Specifically, all that needs to be done is to scale the function output by some constant C_{OHC} with

$$0 \leq C_{\mathrm{OHC}} \leq 1\,. \tag{16.12}$$

This parameter represents the degree of OHC impairment. Thus a C_{OHC} value of 1 indicates healthy functionality, while a value of 0 indicates total impairment. This results in the modified time constant output [5]

$$\tau_{\mathrm{sp,impaired}}(n) = C_{\mathrm{OHC}}\left(\tau_{\mathrm{sp}}(n) - \tau_{\mathrm{wide}}\right) + \tau_{\mathrm{wide}}\,, \tag{16.13}$$

where $\tau_{\mathrm{sp,impaired}}(n)$ is the new time constant, $\tau_{\mathrm{sp}}(n)$ is the non-impaired time constant, and τ_{wide} is a constant value that reflects the time constant of the cochlear filter in the absence of any OHC tuning.

In contrast to the role of the OHCs as a control mechanism, the purpose of the IHCs is to transduce the mechanical stimulus from the basilar membrane into an electrical potential. As a result, the modelling of this portion of the cochlea is rather simpler than the previous section. To simulate the transduction process, a logarithmic function is used as a half-wave rectifier, the output of which is low pass filtered. The rectifier is modelled on a similar rectification process known to exist in the cochlea, while the lowpass filter, which has a cutoff frequency of 3800 Hz, simulates the loss of synchrony capture that occurs in the auditory nerve as the stimulus frequency increases. The modelling of the inner hair cell impairment is also quite simple. The input to the IHC block can simply be scaled by the parameter C_{IHC} [5] with

$$0 \leq C_{\mathrm{IHC}} \leq 1\,. \tag{16.14}$$

As in the case of OHC impairment, a value of one indicates healthy functioning, while a value of zero indicates a total loss of function.

The remaining portion of the cochlear simulation, the synapse model is based on a discrete-time adaptation [6] of Westerman and Smith's three store diffusion model of the cochlear synapse [33]. This block is followed by a time-varying Poisson discharge generator, which takes as its input the instantaneous rate values provided by the discrete-time synapse model. The spike generator also incorporates the effect of refractoriness by keeping track of the time since the last discharge, and modifying the discharge probabilities accordingly [6].

While this cochlear model is fairly complex, it does capture much of the phenomena related to hearing. In particular, as been shown above, it is capable of modelling the effects of hearing loss beyond simply elevating the reception thresholds. Unfortunately, the parameters associated with the modelling of hearing loss are not directly related to measurable quantities, and new procedures need to be developed before they can be estimated from patient testing. It is likely however, that a combination of auditory threshold measurements combined with loudness growth profiles should be able to characterize the inner and outer hair cell parameters [18]. Further work on this subject needs to be carried out in order for this idea to be realizable.

16.6 The Software and Interface

The software to implement the above modules was written in C++, while a graphical user interface was developed using Borland C++ Builder 5. The interface so created allows the end-user a considerable degree of control in deciding the parameters of the simulation, and the nature of the acoustic scenario to test under. This ensures a high degree of repeatability between experiments, and also allows the user to save and load test scenarios that are commonly used.

The interface screen, which is shown in Fig. 16.16 demonstrates some of this functionality. In particular, this screen deals with the basic elements of the acoustic scenario. From here the user can specify the size of the room, the acoustic reflectivity of the walls, as well as the number and placement of the sources. The placement of the sensor (listener) is also handled here, as well the sensors orientation. For the purposes of orientation, it is always assumed that 0° of azimuth corresponds to the direction of the X-axis, while 0° of elevation corresponds to the subject looking straight ahead in an ordinary upright position.

As Fig. 16.16 shows, the user can load any source file he or she wishes provided it is in the standard PCM[9] .wav format, and sampled at 44.1 kHz. This option allows the user to preview the sound before adding it to the simulation. Additional sources can be added simply by clicking the Add Source button after specifying their filenames and positions. The loudness in dB SPL[10] as well as the time-delay of each source can also be chosen, reflecting the fact that sources may start and stop at different times, and be active with different intensities. Sources that have been previously added to the simulation may be viewed and managed by using the Delete Source button, which calls up the menu in shown in Fig. 16.17.

The user can also specify whether the source is distributed in space or not (see Fig. 16.18). This may be desirable since most real sound sources are not

[9] *PCM* stands for *pulse code modulation.*
[10] *SPL* abbreviates *sound pressure level.*

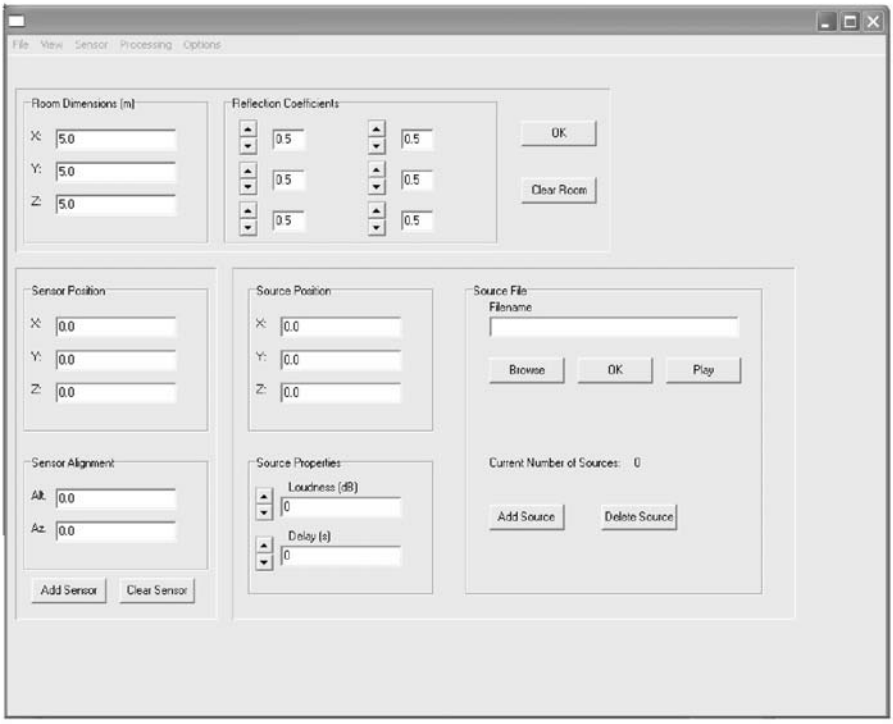

Fig. 16.16. The VirtualHearing main screen. This screen allows the user to specify most of the details regarding the acoustic scenario.

point sources, but are generated over some area or volume. A stereo speaker for example radiates acoustic energy over its whole surface rather than just from a point in its centre. While such a distinction may not matter for many applications, it has been noted that modelling distributed sources as point sources may not allow for proper testing of some spatial processing algorithms. As a result, we have chosen to include the possibility of distributed sources, and to model them as an array of point sources [17]. Such an approach retains both the necessary realism as well as limiting computational complexity.

A source that has been designated as being spatially distributed is centred at the source position previously specified. The user can designate the number of subsources, as well the over all size of the distribution. In addition, the user may also choose from one of several distribution types. At the moment, these include flat surfaces in the XY, XZ, or YZ planes, a cube, or a random cloud, although more types could be added in the future.

With the acoustic environment specified, the user can also choose other simulation parameters to control. In particular, the nature of the hearing loss suffered by the hypothetical patient can be controlled by setting the inner

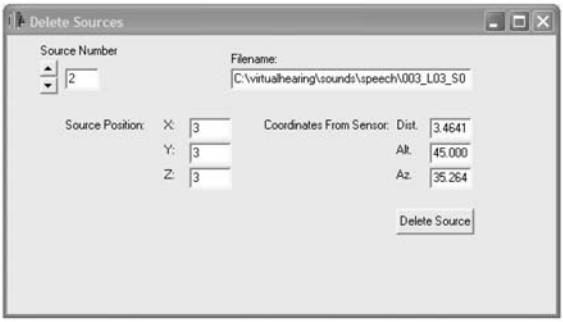

Fig. 16.17. This screen allows the user to browse the selected sources, and to delete some if so desired.

and outer hair cell parameters as shown in Fig. 16.19. For each of the patient centre frequencies, the user can control the hearing impairment levels using the sliding buttons shown above. It is also possible to move through the centre frequencies, and to specify different levels of impairment for each frequency.

The user also has several other simulation options that may be specified (see Fig. 16.20). In particular, it needs to be decided whether or not the simulation should be binaural or monaural. Since most hearing aid algorithms in use today are strictly monaural, this is the default option. In addition, the user is also allowed to specify how many of the early reflections should be considered as being directional. That is, given the placement of the virtual sources, one can decide how many of the impinging virtual sounds need to be convolved with the HRTF for the corresponding angle.

The final simulation option allows the end user to receive the simulation output in terms of a neural spike train in addition to the usual time series of instantaneous spike rates. While this option is only of very limited interest to those developing hearing aids, it is of use to those interested in the

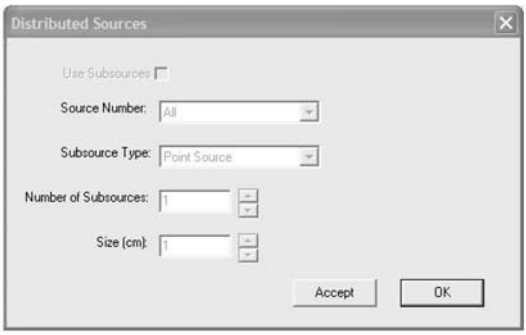

Fig. 16.18. If the user wishes, any individual source may be distributed in space instead of being modelled as a point source.

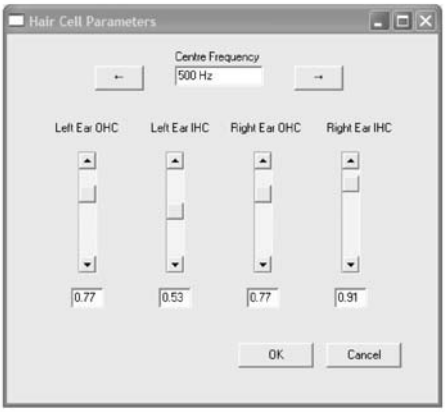

Fig. 16.19. The simulation currently considers seven centre frequencies, which reflect positions on the basilar membrane. The OHC and IHC parameters which specify the level of hearing impairment can be controlled from the menu shown above.

neurobiological processing of sound. Since the real auditory system operates on the basis of spike trains rather than rate signals, study of this system and how it processes sound must be based on realistic inputs. The VirtualHearing simulator is a useful platform in this regard given that it encompasses not only a realistic auditory model, but also offers an easy to use interface for managing signals and environmental effects.

Once the acoustic environment and simulation options are chosen, the simulation itself can be run by clicking on the Processing menu and running the appropriate simulation: environmental, DSP, or neural. Naturally, neither of the last two options can be run before the environmental simulation has been

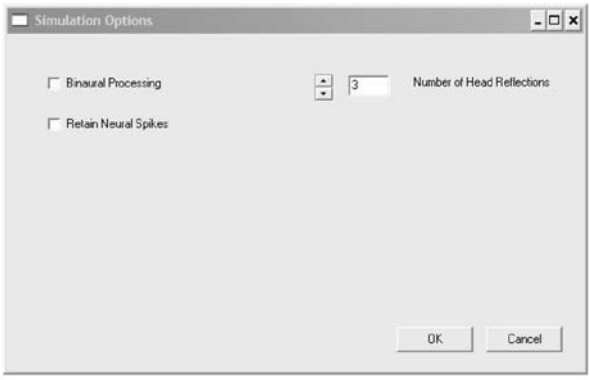

Fig. 16.20. The Simulation menu allows the user to choose several options regarding the simulation output.

completed. In addition, while the neural simulator can be run immediately if the user wishes to do so, the user also has the option of running a custom DSP algorithm that fills the role of the hearing aid processor. User defined algorithms must be in the form of Borland-compatible .dll files. Instructions for creating such files can be found in [12] as well as in the final documentation of our software. A template file will also be provided to assist developers in creating files that will be compatible with the VirtualHearing software. Once created, a custom DSP algorithm can be accessed by clicking File|Load Processor, which calls up the relevant file menu. Similarly, a processor that has already been loaded can be removed by clicking File|Clear Processor, or by clearing the simulation altogether (by clicking File|New).

The results of the simulations can be saved for further analysis depending on the application that the user is interested in. Sound files are saved in PCM .wav format at the standard sampling rate of 44.1 kHz. The other outputs such as the instantaneous spike rates and the spike trains themselves are saved as text files. In this case, there are multiple output files corresponding to the centre frequency as well as the binaural option. For these files, the user is simply asked for a basic file name, to which the particulars of the data are appended. For example, given the base file name out.dat, the program will create several files: out_R0.dat, out_L0.dat, out_R1.dat, and so on. The "R" and "L" suffixes designate the right and left responses, while the numerical entries designate the index (and ultimately the centre frequency) of the particular nerve fibre.

16.7 Software Testing

In order for the software to be useful, it was necessary for us to have confidence that the outputs successfully matched data gathered in real world environments. As a result, it was necessary for us to validate the results. Fortunately, this was unnecessary in the case of the neural model, as extensive comparisons had already been made by the original authors, which indicated its soundness as a model. This meant that only the acoustic modules needed to be examined for accuracy.

The testing procedure we chose to carry out consisted of a listening test given to four volunteers (one male and three females) with unimpaired hearing. For this test, sixty different HINT speech sentences were produced, each filtered by the impulse response of a reverberant room. Thirty of these impulse responses were measured in real rooms, while the remaining thirty were simulated in software using estimates of the room parameters. The sixty speech sentences were presented in random order through a set of headphones to each subject, who asked to rank the "naturalness" of the result on a scale of one to seven. In other words, the subjects were asked to rank their confidence in whether they felt that the reverberation in the signal had been introduced through natural or artificial means. A rank of 1 on this scale meant that

the presented sentence sounded wholly artificial, while a rank of 7 indicated that the subjects felt that the reverberation present in the sentence sounded completely natural.

Summed over all participants the results of the listening trial are shown in Tab. 16.1. In this table, we have taken the rankings of all participants and broken them down according to personal preferences with respect to the realism of the presented sound. Thus, the number of times the participants rated either a simulated or a real sound a given "naturalness" score is recorded below.

Table 16.1. These are the rankings summed over all the test subjects.

	Rank						
Type	1	2	3	4	5	6	7
Simulated	1	3	7	20	22	34	33
Real	0	3	33	14	18	28	24

These results indicate a slight preference in belief that the simulated sounds possessed a greater degree of "naturalness" than did the sounds created using the measured impulse responses. This means that there was little perceptual difference between the real and simulated sounds, and thus the simulation is capable of creating perceptually realistic stimuli. It should be noted that the measured room impulse responses used in this trial were also tested against actual room recordings in [30], and no statistically significant preference was found on the part of the subjects. In other words, we could be confident that the measured room impulse responses provided an accurate model of the room acoustics and did not introduce perceptual artifacts.

A further point of comparison between the natural and artificial room responses was the spectrum of the impulse responses. In keeping with the required properties of artificial reverberation discussed in Sec. 16.3, the spectrum should be broad band, although the high frequency parts of the signal should decrease over time. By comparing the windowed FFTs of the room impulse responses, we can see that this property does indeed hold (see Figs. 16.21 -16.24).

16.8 Future Work and Conclusions

While considerable effort has been put into the creation of this simulator as a tool for researchers, future developments could enhance this software's value. In particular, it may be desirable to include objective speech intelligibility

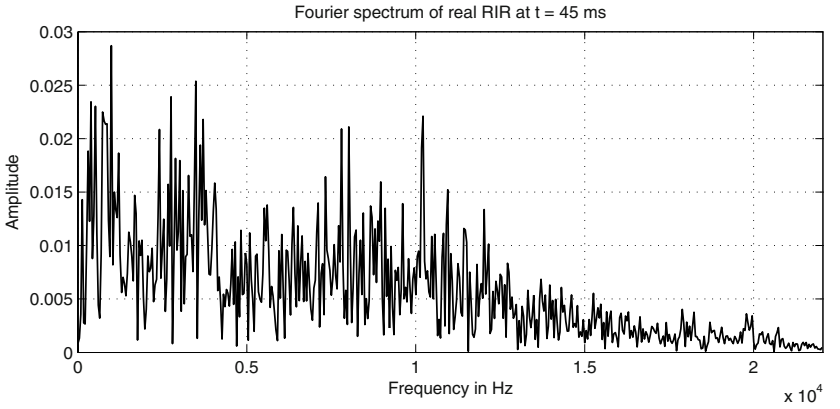

Fig. 16.21. Fourier spectrum of the real room impulse starting at t=45 ms.

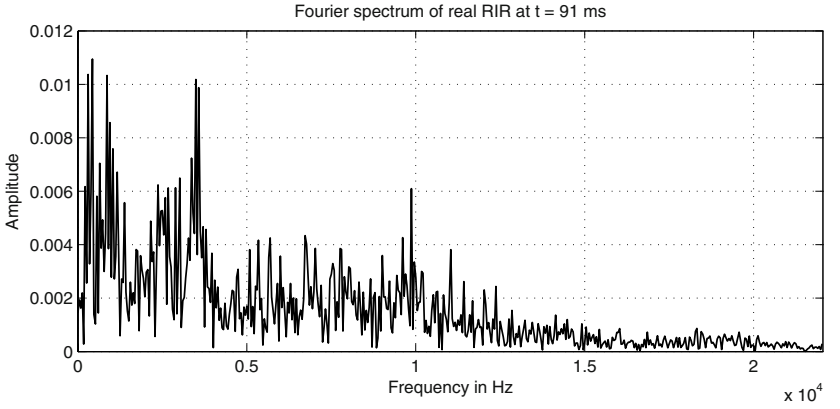

Fig. 16.22. Fourier spectrum of the real room impulse response starting at t=91 ms.

metrics along with automatic comparisons using these metrics. In addition, some thought may also be given to expanding the range of geometries offered in the simulator. Currently, only a binaural simulation is possible. Additional time and effort however, could allow our HRTF models to include multiple microphones, or different receiver geometries. Suggestions from users and potential users may also help shape the development of this software.

Of particular interest to the authors is the development of a hearing test that will allow us to use an audiogram or modified audiogram to estimate the inner and outer hair cell losses. Currently, these parameters are somewhat divorced from the quantities that are measureable on an audiogram. More accurate estimates though could improve algorithm design and fitting by more closely representing the effects of the cochlear damage on the incoming signal.

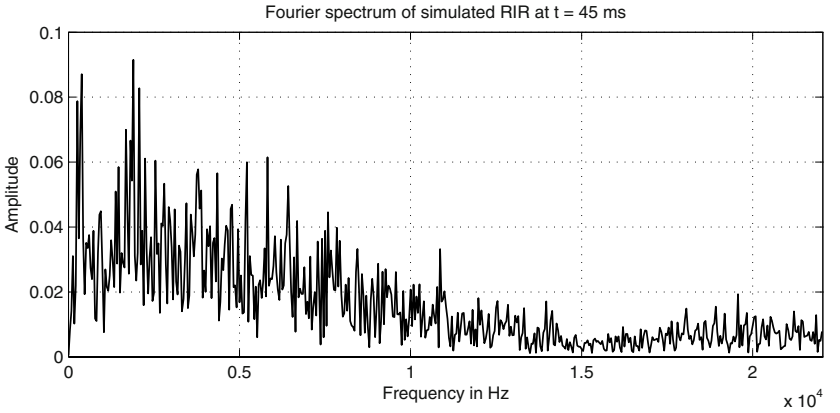

Fig. 16.23. Fourier spectrum of the simulated room impulse response starting at t=45 ms.

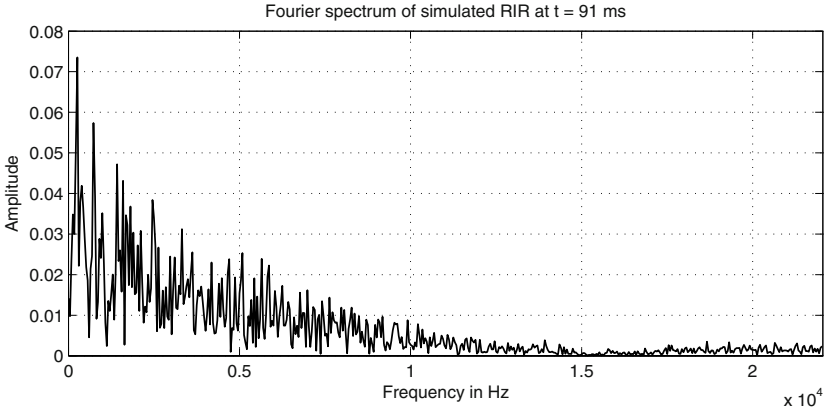

Fig. 16.24. Fourier Spectrum of the simulated room impulse response starting at t=91 ms.

We feel that such an estimation procedure is not out of reach, and that it would be a valuable addition to the tools available to the audiologist.

However, in spite of the possible improvements outline above, we are confident that the VirtualHearing software will prove to be a useful tool for the designers of hearing aid algorithms, as well as those interested in the neurobiological aspects of hearing. Previous experience with the R-HINT-E platform found that it was a useful tool for generating testing data, even if it was somewhat inflexible. The new VirtualHearing software offers a much more flexible simulation environment that encompasses many important problems facing designers, and it does so while ensuring that the simulation is all-inclusive. This encompassing of acoustics, neurobiology and user-defined DSP algorithms will make this software a valuable tool for many researchers.

References

1. V. R. Algazi, R. O. Duda, R. P. Morrison, D. M. Thompson: Structural composition and decomposition of HRTFs, *Proc. WASPAA '01*, 103–106, New Paltz, NY, USA, Oct. 2001.
2. V. R. Algazi, R. O. Duda, D. M. Thompson, C. Avendano: The CIPIC HRTF database, *Proc. WASPAA '01*, 99–102, New Paltz, NY, USA, Oct. 2001.
3. J. B. Allen, D. A. Berkley: Image method for efficiently simulating small-room acoustics, *J. Acoust. Soc. Am.*, **65**(4), 943–950, 1979.
4. D. R. Begault: *3-D Sound for Virtual Reality and Multimedia*, Cambridge, MA, USA: Morgan Kaufmann, 1994.
5. I. C. Bruce, M. B. Sachs, E. D. Young: An auditory periphery model of the effects of acoustic trauma on auditory nerve responses, *J. Acoust. Soc. Am.*, **113**(1), 369–388, 2003.
6. L. H. Carney: A model for the response of low-frequency auditory nerve fibers in cat, *J. Acoust. Soc. Am.*, **93**(1), 401–417, 1993.
7. T. Chi, P. Ru, S. A. Shamma: Multiresolution spectrotemporal analysis of complex sounds, *J. Acoust. Soc. Am.*, **118**(21), 887–906, 2005.
8. P. Dallos: The active cochlea, *Journal of Neuroscience*, **12**, 4575–4585, 1992.
9. R. Fettiplace, C. M. Hackney: The sensory and motor roles of auditory hair cells, *Nature Reviews: Neuroscience*, **7**, 19–29, 2006.
10. S. A. Gelfand: Hearing: *An Introduction to Psychological and Physiological Acoustics*, London, UK: Informa Healthcare, 2004.
11. M. Hauenstein: Application of Meddis inner hair-cell model to the prediction of subjective speech quality, *Proc. ICASSP '98*, **1**, 545–548, Seattle, WA, USA, 1998.
12. J. Hollingworth, et al.: *C++ Builder 5 Developers's Guide,* Indianapolis, IN, USA, Sams Publishing, 2001.
13. G. Hu and D. Wang: Monaural speech segregation based on pitch tracking and amplitude modulation, *IEEE Trans. on Neural Networks*, **15**(5), 1135–1150, September 2004.
14. J. Huopaniemi: *Virtual Acoustics and 3-D Sound in Multimedia Signal Processing*, PhD thesis, Helsinki University of Technology, Helsinky, Finland, 1999.
15. J. Huopaniemi, L. Saioja, T. Tkala: DIVA virtual audio reality system, *Proc. ICAD '96*, 111–116, Palo Alto, CA, USA, 1996.
16. D. N. Kalikow, K. N. Stevens, L. L. Eliot: Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability, *J. Acoust. Soc. Am.*, **61**(5), 1337–1351, 1977.
17. A. L. Lalime, M. E. Johnson: Development of an efficient binaural simulation for the analysis of structural acoustic data, *NASA Technical Report*, CR-2002-211753, 2003.
18. B. C. J. Moore, B. R. Glasberg: A revised model of loudness perception applied to cochlear hearing loss, *Hearing Research*, **188**(1), 70–88, 2004.
19. S. Müller, P. Massarani: Transfer function measurement with sweeps, *Journal of the Audio Engineering Society,* **49**(6), 443–471, June 2001.
20. M. J. Nilsson, S. D. Soli, J. A. Sullivan: Development of a hearing in noise test for the measurement of speech reception thresholds in quiet and in noise, *J. Acoust. Soc. Am.*, **95**(2), 1085–1099, 1994.
21. R. Patterson, et al.: SVOS Final Report: The Auditory Filter Bank, *Technical Report 2341,* Cambridge, UK: MRC Applied Psychology Unit, 1988.

22. A. D. Pierce: Acoustics: *An Introduction to its Physical Principles and Applications*, Woodbury, NY, USA: The Acoustical Society of America, 1991.
23. L. Robles, M. A. Ruggerio: Mechanics of the mammalian cochlea, *Physiological Reviews*, **81**, 1305–1352, 2001.
24. D. Rocchesso, J. O. Smith: Circulant and elliptic feedback delay networks for artificial reverberation, *IEEE Trans. Speech and Audio Process.*, **5**(1), 51–63, 1997.
25. M. B. Sachs, I. C. Bruce, R. L. Miller, E. D. Young: Biological basis of hearing aid design, *Annals of Biomedical Engineering*, **30**(2), 157–168, 2002.
26. L. Savioja, J. Huopaniemi, T. Lokki, R. Vaananen: Virtual environment simulation – Advances in the DIVA project, *Proc. ICAD '97*, 43–46, Palo Alto, CA, USA, 1997.
27. L. Savioja: *Modelling Techniques for Virtual Acoustics*, PhD thesis, Helsinki University of Technology, 1999.
28. R. Shilling, B. Shinn-Cunningham: Virtual auditory displays, in K. M. Stanney (ed.), *Handbook of Virtual Environments: Design, Implementation, and Applications*, 65–92, Mahwah, NJ, USA: Lawrence Erlbaum, 2001.
29. C. Taishih, et al.: Multiresolution spectrotemporal analysis of complex sounds, *J. Acoust. Soc. Am.*, **118**(2), 887–906, 2005.
30. L. J. Trainor, K. Winklung, J. Bondy, S. Gupta, S. Becker, I. C. Bruce, S. Haykin: Development of a flexible, realistic hearing in noise test environment (R-HINT-E), *Signal Processing*, **84**(2), 299–309, 2004.
31. R. Vaananen: *Efficient Modeling and Simulation of Room Reverberation*, Masters Thesis, Helsinki University of Technology, 1997.
32. R. Vaananen, V. Valimaki, J. Huopaniemi, M. Karjalainen: Efficient and parametric reverberator for room acoustics modeling, *Proc. ICMC '97*, 200–203, Thessaloniki, Greece, 1997.
33. L. A. Westerman, R. L. Smith: A diffusion model of the transient response of the cochlear inner hair cell synapse, *J. Acoust. Soc. Am.*, **83**(6), 2266–2276, 1988.
34. K. Wiklund, R. Sonnadara, L. Trainor, S. Haykin: R-HINT-E: a realistic hearing in noise test environment, *Proc. ICASSP '04*, **4**, 5–8, Montreal, Canada, 2004.
35. X. Zhang, M. G. Heinz, I. C. Bruce, L. H. Carney: A phenomenological model for the responses of auditory-nerve fibers: I. Nonlinear tuning with compression and suppression, *J. Acoust. Soc. Am.*, **109**(2), 648–670, 2001.
36. M. S. Zilany, I. C. Bruce: Modeling auditory-nerve responses for high sound pressure levels in the normal and impaired auditory periphery, *J. Acoust. Soc. Am.*, **120**(3), 1446–1466, 2006.