

Improving Virtual Sound Source Robustness using Multiresolution Spectral Analysis and Synthesis

John Garas and Piet C.W. Sommen

Eindhoven University of Technology
Ehoog 6.34, P.O.Box 513
5600 MB Eindhoven, The Netherlands.
E-mail: j.garas@ele.tue.nl

A virtual sound source system based on the principle of active noise control is presented. A solution to the robustness problem inherent to this and other virtual source systems is proposed. This solution exploits acoustical and psychoacoustical factors by employing multiresolution spectral analysis. A new multiresolution transformation method, that is suitable for real-time implementation, is developed for this purpose. Performance improvement is shown by simulation.

1 Introduction

Traditional headphone virtual sound source systems are based on filtering a monaural audio signal by a pre-measured pair of Head Related Transfer Functions (HRTFs), to create a virtual source image at a listener's eardrums. In most cases, non-individual HRTFs are used, since it is not practical to measure and store a set of HRTFs for every listener. Moreover, when loudspeakers are used in those systems instead of headphones, the direction information of the loudspeakers have to be cancelled first before creating the new sound image. This requires a cross-talk cancellation filter, also known as a transaural filter. The transaural filter is an inverse of a HRTF matrix, which is a complicated operation to perform in real time.

An alternative approach that is based on the principle of Active Sound Control (ASC) is presented in section 2. A virtual sound image is created using a 2-point active noise

canceller consisting of a primary loudspeaker, two secondary loudspeakers, and two microphones placed in the ear canals of the listener. The secondary loudspeakers are driven by two adaptive filters that are adjusted to minimize the sound pressure at the listener’s eardrums. After convergence, the adaptive filters are frozen and the primary loudspeaker is disconnected. A monaural signal filtered by the frozen filters and played through the two secondary loudspeakers has enough directional information for the listener to perceive the sound to be coming from the primary loudspeaker position. This system not only uses individual’s HRTFs, but also avoids direct matrix inversion by employing adaptive filters.

Virtual sound source systems work well as long as the listener does not move. Small head movements may destroy the sound image completely. This robustness problem is further explained, and a solution to it is proposed in section 3. Our proposed approach is to process signals and transfer functions using decreasing resolution with increasing frequency; which removes spectral fine structures at high frequencies that characterize specific positions in the listening space. A new nonuniform spectral transformation to perform the above mentioned multiresolution analysis and synthesis in real-time is developed in section 4. And finally, performance improvement is shown by simulation in section 5.

In this paper, upper-case characters are used for frequency domain and lower-case characters for time domain signals. Furthermore, boldface characters are used for diagonal matrices and underlined boldface characters for vectors.

2 Virtual sound sources using active sound control

Consider the set-up shown in Fig. 1 where two real loudspeakers LS_L and LS_R are used to create a virtual source located at the position LS_V . The filters $\underline{\mathbf{w}}_L$ and $\underline{\mathbf{w}}_R$ alter the input signals to LS_L and LS_R so that the listener perceives the sound to be coming from LS_V . As discussed in section 1, the filters $\underline{\mathbf{w}}_L$ and $\underline{\mathbf{w}}_R$ are composed of two components: a transaural filter to remove the existence of the real loudspeakers from the listener’s auditory scene and a HRTF pair carrying the directional information of the virtual source at LS_V . The solutions required for the filters $\underline{\mathbf{w}}_L$ and $\underline{\mathbf{w}}_R$ are then given by

$$\begin{bmatrix} \underline{\mathbf{W}}_L \\ \underline{\mathbf{W}}_R \end{bmatrix} = \begin{bmatrix} \mathbf{H}_{ll} & \mathbf{H}_{rl} \\ \mathbf{H}_{lr} & \mathbf{H}_{rr} \end{bmatrix}^{-1} \begin{bmatrix} \underline{\mathbf{H}}_{vl} \\ \underline{\mathbf{H}}_{vr} \end{bmatrix} \quad (1)$$

Direct implementation of (1) requires measuring the six acoustic transfer functions (ATFs), inverse of a $[2 \times 2]$ ATF matrix, and a matrix multiplication. A rather more attractive ap-

proach than direct implementation of (1) is finding the solution to $\underline{\mathbf{w}}_L$ and $\underline{\mathbf{w}}_R$ adaptively. This has the advantage of

- avoiding direct matrix inversion,
- the matrix multiplication operation in (1) is not needed,
- always individual's HRTFs are used,
- most important of all, the set-up becomes similar to that of a 2-point active sound control system. LS_V functions as a primary source while LS_L and LS_R as secondary sources. The required solutions for $\underline{\mathbf{w}}_L$ and $\underline{\mathbf{w}}_R$ are those which minimize the primary source signal at two microphones placed in the listener's ear canals.

The last point means that the problem can be seen as ASC problem instead of the more difficult one of virtual sound source synthesis. This not only allows employing the advanced techniques used in ASC system, but also facilitates describing the system performance in terms of the sound attenuation achieved at the listener's eardrums. The larger this attenuation, the more pronounced the virtual source is perceived. Exploiting this fact, the virtual sound source is generated, in the ASC context, in two stages:

1. **Cancellation stage:** White noise is played through a physical loudspeaker placed at LS_V while $\underline{\mathbf{w}}_L$ and $\underline{\mathbf{w}}_R$ are adjusted to minimize the sound pressure measured at the listener's eardrums. Once this minimum is reached, the coefficients of $\underline{\mathbf{w}}_L$ and $\underline{\mathbf{w}}_R$ are frozen and LS_V is disconnected. At this point, the two filters have the solution given by (1) except for a 180° phase shift.
2. **Emulation stage:** The monaural audio signal is multiplied by -1 (to compensate for the above mentioned 180° phase shift) and filtered through the frozen $\underline{\mathbf{w}}_L$ and $\underline{\mathbf{w}}_R$. The resulting two signals are played through the two real loudspeakers LS_L and LS_R . This has the effect of making the listener perceive the sound to be coming from the position of LS_V .

Unlike traditional ASC systems, the above system has to control a signal covering the whole audio frequency range. Therefore, the system has to operate on much higher sampling frequency than traditional ASC systems. And since the adaptive solution is, implicitly, the convolution of an ATF and the inverse of another ATF, it has a long and complex impulse response. Therefore, real-time implementation of the system imposes high requirements of hardware resources. To meet the real-time requirements, these adaptive filters are implemented using block processing techniques. Efficient implementation of the system using block frequency domain adaptive filters (BFDAF) is described in [1].

In such implementation, the monaural input signal is divided in overlapping blocks of samples. These blocks are transformed to frequency domain using the fast Fourier transform (FFT). The convolution, required for calculating the filter output, and correlation, required for the coefficients update, are performed in frequency domain where these operations map to element-wise multiplications. In this paper, we are considering replacing the FFT with a similar transformation that helps solving the robustness problem described in the next section.

3 The Robustness Problem

The adaptive virtual source generator described above works just fine as long as the listener does not move his/her head through the whole process of cancellation and emulation. Small head movements may destroy the sound image completely. This is expected, since the solutions for \underline{w}_L and \underline{w}_R are dependent on the six involved ATFs as given by (1). Those ATFs are, in turn, very position dependent. This position dependency is further frequency dependent. A movement of 2 cm corresponds to a complete wavelength of a 17 kHz signal while it is just 0.01 wavelength of a 170 Hz signal. Therefore, larger amplitude errors are encountered at high frequencies than low frequencies.

It is the goal of this paper to improve system robustness to *small* listener movements. Large movements have to be solved by updating the adaptive filters on-line, which will be the topic of future research. But to prevent the system from continuously updating these filters, and never reaching a steady state, the virtual source image has to be kept intact within small movements.

Our approach in achieving this goal is replacing the constant spectral resolution filters by multiresolution ones. This is done by processing signals and transfer functions using decreasing resolution with increasing frequency; which removes spectral fine structures at high frequencies that characterize specific positions. This has two effects:

- The first is that the transaural filter component of the adaptive filter solution becomes a spatial average solution. This is because the impulse responses of a room at adjacent positions in space are similar at low frequencies but may vary significantly at high frequencies. Furthermore, when the listener moves a little, his HRTFs differ only at high frequencies.
- The second effect however, is reducing the spectral localization cues embedded in the acoustic path from LS_V to the ears. This has rather a negligible effect on the perceived source location which is justified by two facts:

- the human auditory system performs similar nonuniform spectral analysis, therefore similar inaccuracies,
- human localization resolution that depend on high frequency cues (elevation resolution for instance) are rather poor compared to binaural cues.

It is shown in section 5 that the proposed technique improves system robustness significantly. But first, a multiresolution transformation that is suitable for real time applications, such as that considered here, is developed.

4 Nonuniform spectral analysis and synthesis

Recognising the importance of using nonuniform spectral analysis in audio applications, researchers began in the early seventies to look for a reliable method to perform such analysis. A brief summary of the published methods is given in subsection 4.2. Unfortunately, none of these methods meets the real-time requirements. Therefore, a new multiresolution transform is developed in subsection 4.3. To be able to judge those algorithms, we first define the requirements for a multiresolution transform to be suitable for real-time applications.

4.1 Real-time requirements

For a multiresolution transformation to be of any use in real-time applications, it has to satisfy the following requirements:

- The transform should be linear, unitary and invertible (nonsingular).
- It must support the convolution property. This means that convolution in time domain can still be carried out as element-wise multiplication in the transform domain.
- The discovery of the fast Fourier algorithm (FFT) made it possible to perform the convolution operation faster in frequency domain. The multiresolution transform has to maintain this speed advantage.

Although the fast wavelet transform meets the above requirements, it is not suitable for the application in hand, since it gives too coarse resolution at higher frequencies due to its binary grid. Alternative methods to perform multiresolution spectral analysis are discussed in the next subsection.

4.2 Existing Techniques

In 1971 Oppenheim et al. [2] introduced the digital frequency warping method. They showed that calculating the DFT of a predistorted time sequence gives samples of the spectrum on a nonlinear frequency scale. Although this method uses the speed of the FFT to transform the distorted sequence to the frequency domain, it requires $N(3M - 1)$ real operations to calculate M distorted samples from N input time samples. This is proportional to N^2 for $M = N$ which makes it less attractive for real-time applications.

Several authors have published methods to perform nonuniform spectral analysis using block processing algorithms. Harris [3] processed the FFT output with spectral windows of constant time duration but adjustable bandwidths centered at the nearest FFT bin to the required analysis frequency. This method does not alter the time resolution of the signal. Therefore, it is information destructive and has no inverse transformation.

Another group of authors concentrated on calculating a constant-Q (proportional bandwidth) spectral analysis and synthesis. Youngberg and Boll [5] gave a constant-Q integral transform, which is a generalization of the Short Time Fourier Transform (STFT), by letting the analysis window be a variable in the product of time and frequency. Gambardella [6] showed that if a signal undergoes a short time spectral analysis via a continuous set of constant-Q bandpass filters, this process can be mathematically represented through an integral transform that can be inverted by means of the Mellin transform. Kates [7] used an exponentially decaying window whose argument is a constant times the product of time and frequency to calculate the constant-Q integral. For this specific window, the integral can be evaluated using the chirp z-transform. And finally, Brown [10] has developed a discrete version of the constant-Q integral transform but, unfortunately, this method has no inverse. All these methods can be viewed as a constant-Q filter bank and involve a high degree of complexity to be used in real-time filtering applications.

Mitra et al. [13] have recently introduced the Nonuniform Discrete Fourier Transform (NDFT). The NDFT evaluates the z-transform at arbitrary located distinct points in the complex z-plane. They showed that the NDFT is invertible and can be used in filtering operations. But since the NDFT lacks the symmetry of the Fourier transform, it can not be calculated by a fast method like the FFT and therefore requires an order of N^2 complex operations.

Unfortunately, none of the above mentioned methods (and others not mentioned here) satisfy all the real-time requirements of subsection 4.1. In the next subsection, a multiresolution transformation that is suitable for real-time applications is developed.

4.3 Time-Frequency Scaling (TFS)

In this subsection we introduce a new nonuniform spectral analysis and synthesis method that requires only an FFT and possibly a resampler. In block processing system, the resampler is not needed and the complexity is that of an FFT.

4.3.1 TFS Formulation

The new method, which we refer to as Time-Frequency Scaling (TFS), is based on Clark's one-dimensional nonuniform sampling theorem [12], and the scaling property of the Fourier transform. The nonuniform sampling theorem states that a band limited function of one variable $f(t)$ sampled at sampling moments t_n that are not necessarily equally spaced (Fig. 2 (a)) can be completely reconstructed from its samples; provided that there exists a one-to-one continuous stretching or compressing transformation $\tau = \gamma(t)$ that maps $f(t)$ to $g(\tau)$ (Fig. 2 (b)) with sampling moments $nT = \gamma(t_n)$. The signal $g(\tau)$ is uniformly sampled on the time axis τ and the classical sampling theorem holds. Therefore $f(t)$ can be reconstructed from its nonuniform samples $f(t_n)$ using the following formula

$$f(t) = \sum_{n=-\infty}^{\infty} f(t_n) \frac{\sin \left[\frac{\pi}{T} (\gamma(t) - nT) \right]}{\left[\frac{\pi}{T} (\gamma(t) - nT) \right]} \quad (2)$$

The relation between the spectra of $f(t)$ and $g(\tau)$ is given by the scaling property of the Fourier transform which, for a constant scaler α , is given by

$$f(\alpha t) \Leftrightarrow \frac{1}{\alpha} F\left(\frac{\omega}{\alpha}\right) \quad (3)$$

It states that if the time axis is stretched (compressed) by a factor α , the frequency axis is compressed (stretched) by the same factor so that the product of time and frequency is always constant. Applying the scaling property to the nonuniform sampling case shows that the scaling factor γ is a function of time and thus, the frequency axis will be stretched or compressed nonlinearly. To see that this holds, let the frequency variables corresponding to t and τ be ω and Ω respectively. Then $g(\tau)$ is given by

$$g(\tau) = \frac{1}{2\pi} \int_{-\frac{\pi}{T}}^{+\frac{\pi}{T}} G(\Omega) e^{j\Omega\tau} d\Omega \quad (4)$$

where $G(\Omega)$ is the Fourier transformation of $g(\tau)$. Noting that $g(\tau) = g(\gamma(t)) = f(t)$, (4) can be written as

$$f(t) = \frac{1}{2\pi} \int_{-\frac{\pi}{T}}^{+\frac{\pi}{T}} G(\Omega) e^{\frac{j\Omega\tau t}{t}} d\Omega \quad (5)$$

substituting $\omega = \frac{\Omega\tau}{t}$, (5) can be written in the form

$$f(t) = \frac{1}{2\pi} \int_{(-\frac{\pi}{T})(\frac{\tau}{t})}^{(+\frac{\pi}{T})(\frac{\tau}{t})} \frac{1}{\tau/t} G\left(\frac{\omega}{\tau/t}\right) e^{j\omega t} d\omega \quad (6)$$

which shows that the analysis frequencies, as well as the bandwidth, of the nonuniformly sampled signal is scaled by the ratio $\tau/t = \gamma(t)/t$. This ratio controls the mapping (or warping) between Ω and ω .

The above theory can be readily applied in real-time to calculate the multiresolution spectra of time signals. For this, we consider two cases: The first when the signals to be processed are already uniformly sampled digital signals. The second when block processing is used on continuous time signals.

Nonuniform spectral analysis of a band limited signal sampled uniformly can be obtained using the following three steps:

1. Define a one-to-one continuous mapping function that satisfies the nonuniform sampling theorem.
2. Resample the signal using the previously selected mapping function. This may be done using an interpolation algorithm or much faster using a hardware resampler as described in subsection 4.3.3.
3. Perform an FFT transformation on the nonuniformly sampled function.

Note that the nonuniform sampling theorem ensures that no aliasing occurs, and the original signal can be recovered by first taking the inverse FFT, and then resampling from the nonuniform to the uniform domain.

In cases where all processing is performed on blocks of data, as in the case of implementing the virtual source generator using block frequency domain adaptive filters [1], and it is required to perform nonuniform spectral analysis in the whole system, it is wiser to sample the signals from the continuous time domain directly to the nonuniform domain. This has to be performed such that the nonuniform sampling pattern is repeated every block. In this case, the resampling operation is not required and the multiresolution spectra are obtained by performing an FFT transformation directly on the sampled sequence.

4.3.2 Example

To demonstrate the multiresolution properties of the TFS method, the mapping function $\gamma(\cdot) = \log_b(\cdot)$, where b is any suitable base, is used. The nonuniform sampling moments are $t_n = \gamma^{-1}(nT) = b^{nT}$ and the corresponding uniform moments $\tau_n = \log_b(t_n) = nT$. This mapping function, with $b = e$ and $T = 1/44100$, is applied to a time signal composed of four sine waves as shown in Fig. 3 (a) with frequency components 400, 600, 18000 and 18200 Hz. The FFT of 1024 exponential samples is shown in Fig. 3 (b) while the FFT of the uniform samples, taken in the same time period as in the exponential case, is shown in Fig. 3 (c). Comparing the spectra in Fig. 3 (b) and (c) shows that the resolution in case (b) decreases as the frequency increases while the resolution is constant in case (c).

4.3.3 Hardware Resampler

In most practical cases, the audio signals to be analysed or processed are available in continuous time form, and a resampler is not needed. In cases where signals are available only in digital form, the resampling operation can be readily implemented in hardware. This can be done using an A/D converter connected back to back with an D/A converter. To transform a signal from a uniform sampling domain to a nonuniform one an D/A converter clocked at the uniform sampling moments is first used to transform the signal to the continuous domain which is then resampled by an A/D converter clocked at the nonuniform sampling moments. Similarly, to transform a signal from the nonuniform sampling domain to the uniform domain, the D/A is clocked at the nonuniform sampling moments while the A/D is clocked at the uniform sampling moments as given by (2).

Yet, in cases where a hardware sampler can not be used, any of the well known interpolation algorithms may be employed. In these cases, the transform complexity depends on the complexity of the interpolation algorithm in use; which ranges from $O(N)$ for cubic interpolation to $O(N^2)$ for spline algorithms. Simulations have shown that good results can be obtained using cubic interpolation when the signal is first oversampled by a factor four.

5 Simulations

In this section, it is shown that the robustness of the virtual sound source generator, presented in section 2, can be improved by using multiresolution spectral analysis. This is done using the one-point noise cancellation experiment set-up shown in Fig.4. In this set-up, a simplified version of the virtual source generator at one ear is simulated. A white

noise signal x is filtered through an acoustic impulse response \underline{h}_p simulating the sound pressure due to a physical loudspeaker LS_V at one eardrum of the listener. The filter \underline{w} is calculated to minimize the sound pressure at MIC through the secondary source and \underline{h}_s . The MIC output is then fed to a cochlea model (gammatone filter bank [14]) consisting of 64 channels and covering the frequency range from 20 kHz to 20 Hz. The output of each fourth channel of these 64 channels are half-wave rectified, low-pass filtered and plotted against time for inspection. These plots represent the probability of firing along the auditory nerve due to the sound signal received at the listener's eardrum represented here by the microphone MIC.

The optimum solution for the filter \underline{w} in this case is a simple version of (1) and is given (in frequency domain) by

$$\underline{W} = \mathbf{H}_s^{-1} \cdot \underline{H}_p \quad (7)$$

where the diagonal matrix \mathbf{H}_s and the vector \underline{H}_p represent the spectrum of the corresponding acoustic impulse responses. From this equation, it is clear that the filter length must be large enough to accommodate this solution. The degree of attenuation in sound pressure at MIC depends on the length of the filter \underline{w} . A longer filter achieves higher attenuation, and therefore better emulation of the virtual source.

Using the FFT algorithm to calculate \mathbf{H}_s and \underline{H}_p from \underline{h}_s and \underline{h}_p produces constant resolution filter \underline{W} . On the other hand, using multiresolution transform to obtain \mathbf{H}_s and \underline{H}_p will produce a multiresolution solution for \underline{W} . To show the performance improvement obtained using multiresolution technique, the following experiment is performed:

Four acoustic impulse responses, each of 512 samples, were calculated using the source image program *Room* [16]. A room of dimensions [$L_x = 5, L_y = 4, L_z = 3$] meters and reverberation time of 0.3238 seconds has been used. The four transfer functions are as follows (see Fig. 5 - 8):

- \underline{h}_{p0} : from primary source at [2, 3, 1.5] to MIC position at [4, 3, 1.5]
- \underline{h}_{s0} : from secondary source at [4, 2, 1.5] to MIC position at [4, 3, 1.5]
- \underline{h}_{p1} : from primary source at [2, 3, 1.5] to MIC position at [4, 3, 1.52]
- \underline{h}_{s1} : from secondary source at [4, 2, 1.5] to MIC position at [4, 3, 1.52]

\underline{h}_{p1} and \underline{h}_{s1} represent moving MIC 2 cm vertically in space from its initial position. The filter \underline{W} of 1024 coefficients was calculated in two different ways: using constant and multiresolution spectral analysis. Only the initial MIC position has been used in calculating \underline{W} such that $\underline{W} = \mathbf{H}_{s0}^{-1} \cdot \underline{H}_{p0}$. The simulation shown in Fig. 4 was ran twice; once using the constant resolution \underline{W} and the second time using the multiresolution

W. At the middle of the simulation, MIC is moved 2 *cm* from its initial position (the simulation switched to $\underline{\mathbf{h}}_{p1}$ and $\underline{\mathbf{h}}_{s1}$ in place of $\underline{\mathbf{h}}_{p0}$ and $\underline{\mathbf{h}}_{s0}$). The output of these two simulation runs is shown in Fig. 9. From this figure, we notice the following:

- The constant resolution solution suffers from MIC movements only at high frequencies, where the movement is comparable to the wave length. This is consistent with the motivation of using multiresolution approach as discussed in section 3.
- The multiresolution solution produces much better attenuation at almost all frequencies. At low frequencies, higher resolution is used and therefore more accurate solution is obtained. At high frequencies, coarser resolution is used which matches that of the cochlea model and prevents the effects of the spectral fine structure at those frequencies.
- Spatial averaging at high frequencies is clear when using coarse resolution as discussed in section 3. This averaging effect makes the solution robuster to MIC movements, and therefore makes the virtual source generator robuster to small listener movements.
- Constant resolution could not achieve any attenuation at some frequencies (such as channels 40 and 44) even before moving MIC, while multiresolution could. This may occur if the spectrum of $\underline{\mathbf{h}}_s$ contains a deep notch at these frequencies, causing inversion problems. Since coarser resolution is equivalent to averaging the spectra, this produces less notches and a better solution for W can be obtained.

6 Conclusions

The active sound control principle (ASC) is used to create a virtual sound source image in an arbitrary position in the listening space. This approach not only exploits the advanced techniques employed in ASC systems, but also facilitates using the attenuation achieved at the listener's ears as a physical measure of virtual source perception.

The robustness to listener movements is a known problem associated with virtual source systems. Multiresolution spectral analysis is shown to be an effective solution to small listener movements. For this purpose, a multiresolution transform that can be implemented in real time is needed. A suitable transform that employs only an FFT, and possibly a resampler, is developed in this paper.

Simulation shows that a multiresolution transformation with coarse resolution at high frequencies and high resolution at low frequencies is superior to constant resolution methods. It not only improves system robustness against listener movements, but also produces

better attenuation at almost all frequencies and therefore assists in better perception of a virtual source.

References

- [1] Piet C.W. Sommen, Ronald M. Aarts Alexander W.M Mathijssen, John Garas and Haiyan He, *Efficient Frequency Domain Filtered-x Realization of Phantom Sources*, Proceedings of the 8th annual workshop on circuits, systems and signal processing, ProRISC 1997 pp. 589-595. Also available from www.ses.ele.tue.nl/signalen/personen/garas/abstracts.html
- [2] Alan V. Oppenheim, Don Johnson and Kenneth Steiglitz, *Computation of Spectra with Unequal Resolution Using the Fast Fourier Transform*, Proc. IEEE, Vol. 59, pp. 299-301, FEBRUARY. 1971.
- [3] Fredric J. Harris, *High-resolution Spectral Analysis with Arbitrary Spectral Centers and Arbitrary Spectral Resolutions*, J. Comput. & Elect. Eng. Vol. 3, pp. 171-191, 1976.
- [4] Howard D. Helms, *Power Spectra Obtained From Exponentially Increasing Spacings of Sampling Positions and Frequencies*, IEEE trans. on Acoust., Speech and Signal Processing, Vol. ASSP-24, No. 1, pp. 63-71, FEBRUARY 1976.
- [5] James E. Youngberg and Steven F. Boll, *Constant-Q Signal Analysis and Synthesis*, Proceedings of the IEEE Int. Conf. on Acoust., Speech and Sig. Proc., pp. 375-378 APRIL 1978.
- [6] G. Gambardella, *The Mellin Transforms and Constant-Q Spectral Analysis*, J. Acoustical Society of America, Vol. 66, No. 3, pp. 913-915, SEPTEMBER 1979.
- [7] James M. Kates, *Constant-Q Analysis Using the Chirp Z-Transform*, Proceedings of the IEEE Int. Conf. on Acoust., Speech and Sig. Proc., CH-1379, pp. 314-317, 1978.
- [8] Dale T. Teaney, Victor L. Moruzzi and Frederick C. Mintzer, *The Tempered Fourier transform*, J. Acoustical Society of America, Vol. 67, No. 6, pp. 2063-2067, JUNE 1980.
- [9] James M. Kates, *An Auditory Spectral Analysis Model Using the Chirp Z-Transform*, IEEE trans. on Acoust., Speech and Signal Processing, Vol. ASSP-31, No. 1, pp. 148-156, FEBRUARY 1983.

- [10] Judith C. Brown, *Calculation of a Constant Q spectral Transform*, J. Acoustical Society of America, Vol. 89, No. 1, pp. 425-434, JANUARY 1991.
- [11] Judith C. Brown, *An efficient Algorithm for the Calculation of a Constant Q Transform*, J. Acoustical Society of America, Vol. 92, No. 5, pp. 2698-2701, NOVEMBER 1992.
- [12] James J. Clark, Matthew R. Palmer, and Peter D. Lawrence, *A Transformation Method for the Reconstruction of Functions from Nonuniformly Spaced Samples*, IEEE trans. on Acoust., Speech and Signal Processing, Vol. ASSP-33, No. 4, pp. 1151-1165, OCTOBER 1985.
- [13] S. K. Mitra, S. Chakrabarti and E. Abreu, *Nonuniform Discrete Fourier Transform and Its Application in Signal Processing*, Proc. EUSIPCO '92, Sixth Euro. Signal Processing Conf., Brussels, Belgium, Vol.2, pp. 909-912, AUGUST 1992.
- [14] R. D. Patterson, K. Robinson, J. Holdsworth, D. McKeown, C.Zhang, and M. H. Allerhand, *Complex sounds and auditory images*, Auditory Physiology and Perception, (Eds.) Y Cazals, L. Demany, K.Horner, Pergamon, Oxford, 1992, pp. 429-446.
- [15] Alan V. Oppenheim and Ronald W. Schaffer, *Discrete-time Signal Processing*, Prentice-Hall, Englewood Cliffs, 1989, ISBN 0-13-216-771-9.
- [16] John Garas, *Room Impulse Response 2.1*,
www.ses.ele.tue.nl/signalen/personen/garas/room/room.html

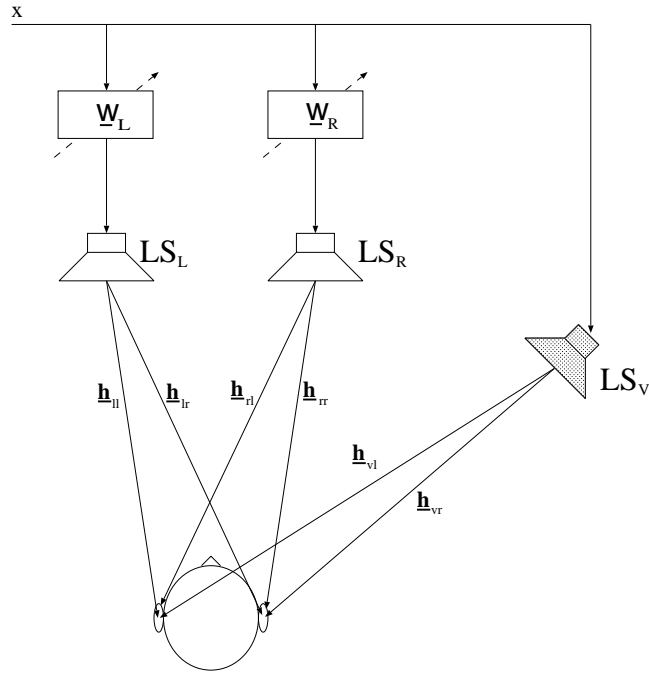


Figure 1: *Set-up for virtual sound source generation.*

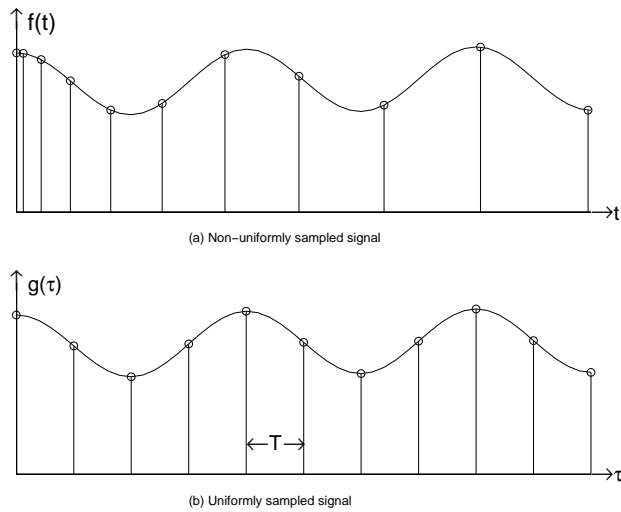


Figure 2: *A function $f(t)$ sampled at sampling moments t_n not equally spaced can be transformed to a uniformly sampled function $g(\tau)$ by applying a compression or expansion mapping.*

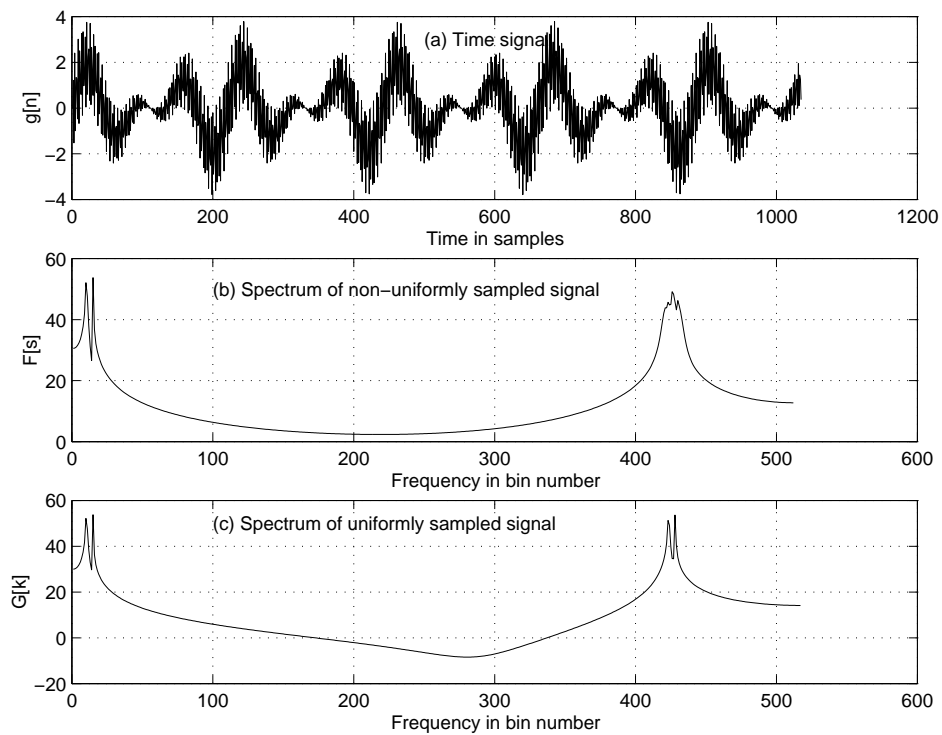


Figure 3: A time signal (a), its nonuniform spectra (b) and its uniform spectra (c).

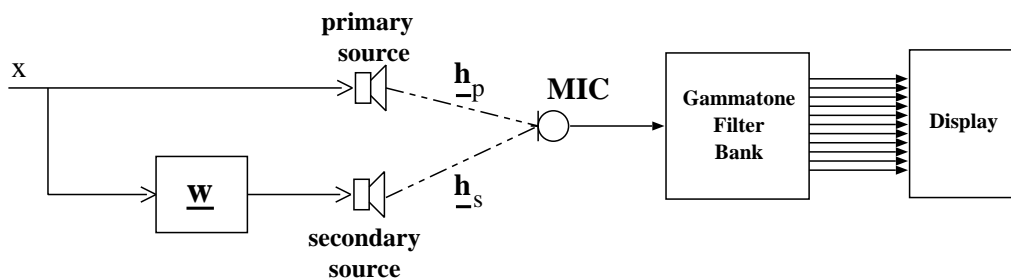


Figure 4: Experiment set-up.

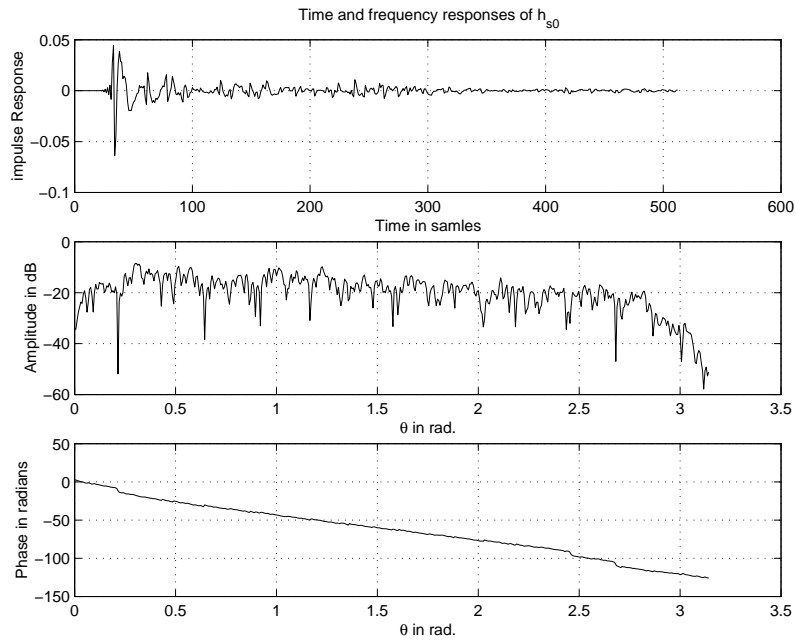


Figure 5: *Initial secondary acoustic path.*

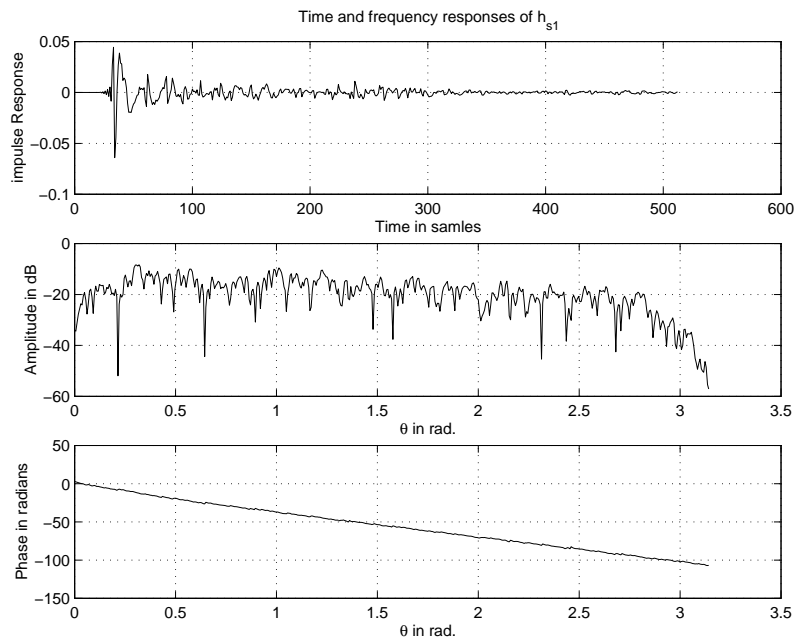


Figure 6: *Secondary acoustic path after moving MIC 2 cm vertically.*

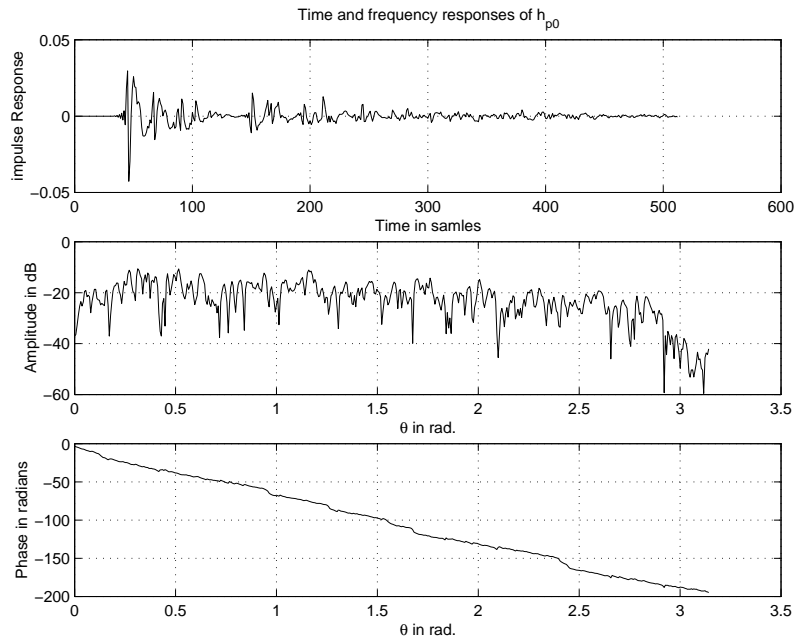


Figure 7: *Initial primary acoustic path.*

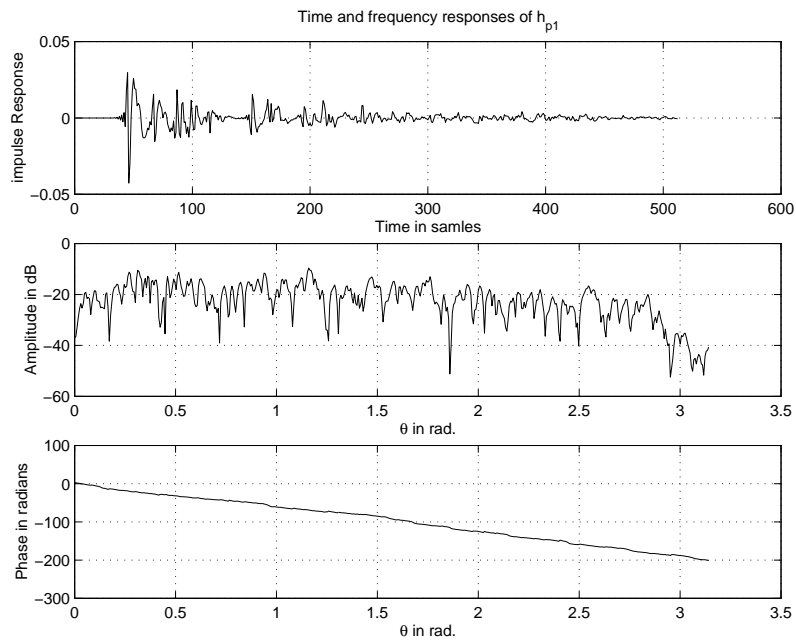


Figure 8: *Primary acoustic path after moving MIC 2 cm vertically.*

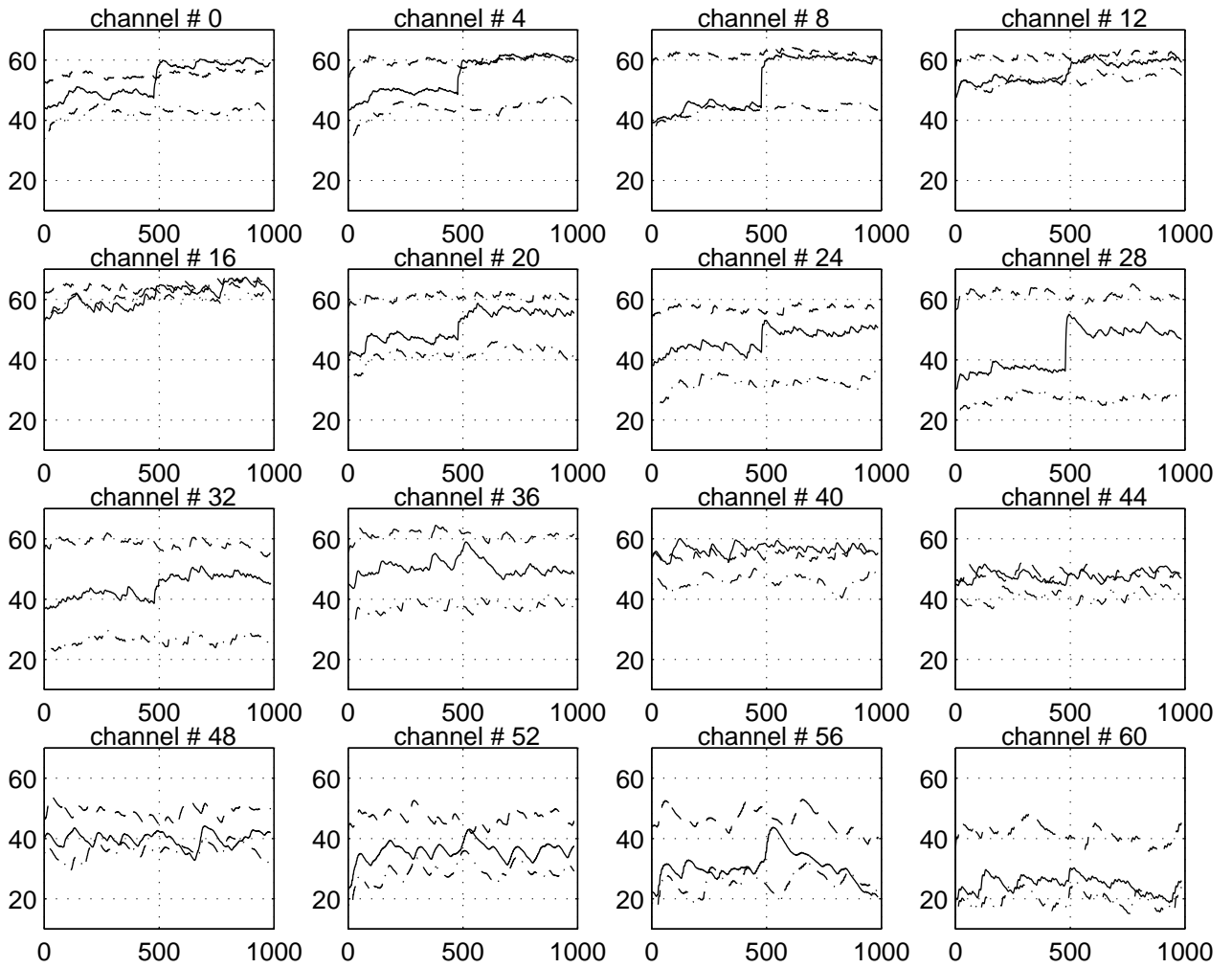


Figure 9: *Output of the cochlea model: Lower channel number corresponds to higher frequencies. (— —) sound pressure level (SPL) before cancellation, (—) SPL after cancellation using uniform spectral analysis and (— · —) SPL after cancellation using nonuniform spectral analysis.*