5

High-frequency Bandwidth Extension for Audio

5.1 INTRODUCTION

The previous chapters focused on BWE methods to extend the (perceived) low-frequency content of reproduced audio signals. The bandwidth limitation in those cases was primarily due to the transducer. Bandwidth limitation can also occur in the transmission channel (in which, for the moment, we also include signal storage). A familiar example is the telephone channel, which has a bandwidth of about 3 kHz. Speech signals transmitted through this channel are audibly bandlimited, because the bandwidth of natural speech is about 8 kHz. Methods to extend the bandwidth of speech, primarily intended for telephony applications, will be discussed in Chapter 6.

Bandwidth limitation in the transmission channel can also occur if perceptual audio coders are used at high compression ratios; for example, in recent years MPEG1 layer-3 (a.k.a. MP3) has become a tremendously popular format for audio storage and transmission. Perceptual audio coders achieve high coding efficiencies because they attempt to store signal information with a low resolution, just 'sufficiently high' for the human auditory system. For the MP3 scheme, this means, in practice, that a significant amount of distortion is introduced in the signal, but the distortion spectrum is designed such that it remains inaudible, for example, is *masked* (see Sec. 1.4.4.5). This in achieved by analysing the short-term power spectrum of the audio signal and using a masking model to compute the masked threshold, that is, a frequency-dependent curve below which distortion components will be masked by the audio signal. This is turn determines, per frequency band, how many bits are needed to code the audio signal. Now, for very high compression ratios, or equivalently, very low bit rates, the coding algorithm is not able to keep all of the distortion below the masked threshold for the full-bandwidth signal. Typically, the bandwidth is then reduced at the high-frequency end, such that the specified bit rate can be achieved for the bandwidth-limited signal, while at the same time keeping the distortion below the masked threshold. The drawback is that high frequencies are lost, resulting in a 'muffled' sound percept. Reviews of audio coding and audio signal processing can be found in, for example, Bosi and Goldberg [37] and Kahrs and Brandenburg [139]. Low-bit-rate perceptual audio coders are now being extensively used

Audio Bandwidth Extension: Application of Psychoacoustics, Signal Processing and Loudspeaker Design. E. Larsen and R. M. Aarts
© 2004 John Wiley & Sons, Ltd ISBN 0-470-85864-8

for audio storage on the Internet, distribution through Internet radio, satellite radio, and private use such as with personal computers, MP3-players, and the like.

To enhance the reproduction of high-frequency bandwidth-limited audio, BWE processing can be applied. Given that the reproduction system can reproduce high frequencies, which is usually not a problem, synthetic frequency components can be added to the signal, improving the quality thereof. Of course, the added synthetic frequency components should be derived from the available bandwidth-limited signal, and for this two approaches have been developed:

- In the first approach, the BWE algorithm is *blind*, that is, it has no information regarding the missing high-frequency components. Thus, only assumptions on the statistics of audio signals (Sec. 1.2) can be used to design such systems. The main advantages of using this approach are that such a BWE system can be applied to a wide class of signals (specifically, both to music and speech) and that there are no requirements on the signal format, because the only required information is the actual signal waveform (or spectrum). This means that such methods could also be used to enhance the quality of old (analog) recordings. It also appears that computationally efficient algorithms can be realized. The drawback is that the quality of the bandwidth-extended output signal is significantly lower than that of the original full-bandwidth signal, even though it is higher than the bandwidth-limited signal. This is due to the lack of information about the missing high frequencies. This approach is the topic of Sec. 5.4.
- In the second approach, the BWE algorithm does have a priori knowledge regarding the missing high-frequency components. This allows for a much more exact reconstruction of the original full-bandwidth signal than is possible with the blind approach, and therefore the quality of the bandwidth-extended signal can be (near) transparent, that is, indistinguishable, from the original full-bandwidth signal. The high quality of the output signal is obviously the main advantage of this approach. The drawback is that some provisions need to be taken to provide the BWE algorithm with the requisite a priori information. A successful approach is that of ‘spectral band replication’ (SBR), the topic of Sec. 5.5. This is a method that works in combination with an ‘ordinary’ audio codec[1], and stores (at a very low bit rate) some specific information about high frequencies in the coded audio stream; in this way the overall required bit rate can be significantly reduced. If the appropriate decoder is used, this information is utilized to reconstruct the missing high frequencies. A decoder that cannot use the additional information only decodes the low-frequency band, thereby insuring forward and backward compatibility. SBR has been used to enhance the coding efficiency of MP3, leading to MP3Pro, and also AAC (Advanced Audio Coding), leading to aacPlus.

A more conceptual difference between the two approaches mentioned here is that the first attempts to extend the bandwidth of a signal that is, for whatever reason, bandlimited, while the second attempt purposely limits the bandwidth of the signal, but does it in such a way that at the output a high-quality full-bandwidth signal can be recreated.

Besides creating a ‘brighter’, more natural sound percept, high-frequency BWE processing can potentially enhance localization of sound sources as well. Bronkhorst [40]

[1] ‘Codec’ is the concatenation of the words coder/decoder, referring to both the coding and decoding algorithms used for a particular coding scheme.

found that spectral cues above 7 kHz have a significant effect on localization performance – by reducing the number of front/back confusions, and enabling the listener to localize a sound source not only more accurately but also more quickly and with less head movements. See also (Chapter 8) the patent by Dempsey on p. 257.

In Sec. 5.6, a method (BWE instantaneous compression) is discussed that would properly be categorized as a blind high-frequency BWE algorithm, but it is discussed separately as it was not originally designed for BWE purposes. The original purpose of this algorithm was to enhance reproduction of centre and surround channel signals in multi-channel sound systems. Signals in these channels are often at a somewhat low level, and the algorithm was designed as a simple means to boost their level, while preventing distortion at high signal levels. The particular nature of the processing also extends the high-frequency content of the signal spectrum, and as such it is also a BWE algorithm that works well in this particular application; it is not generally applicable as are the methods of Secs. 5.4 and 5.5. Section 5.3 discusses the perceptual aspects of high-frequency BWE methods.

First, in Sec. 5.2, we briefly show that traditional methods (in particular, deconvolution) to overcome bandwidth limitations in transmission channels are not suitable for the applications as discussed in this introduction.

5.2 THE LIMITS OF DECONVOLUTION

If a wideband signal $x(t)$ is passed through a linear system $h(t)$ having, for example, a low-pass characteristic, the filtered signal $x_l(t)$ has a reduced bandwidth. We can write

$$x_l(t) = x(t) * h(t)\,, \tag{5.1}$$

where $*$ denotes convolution.

If the received signal $x_l(t)$ is used to reconstruct an estimate $\hat{x}(t)$ of the original $x(t)$, we need to find a filter $g(t)$ such that $x(t) * g(t) = \delta(t - \tau)$, with $\tau > 0$. In that case, we would have

$$\hat{x}(t) = x_l(t) * g(t) = x(t) * f(t) * g(t) = x(t) * \delta(t - \tau) = x(t - \tau)\,, \tag{5.2}$$

a perfect reconstruction, up to a finite time delay. The only condition on $g(t)$ is that it must be stable, which means that all its poles must lie within the unit circle (Sec. 1.1). Now $g(t)$ is simply the inverse of $f(t)$ (up to the time delay τ), which means that all of $f(t)$'s zeros must lie within the unit circle, implying that $f(t)$ be minimum phase. Because $f(t)$ must also be stable, its poles will also lie inside the unit circle, and by the same argument as before, $g(t)$ must then also be minimum phase. So we find that the inverse of $f(t)$ can only be obtained if it is minimum phase; the inverse filter $g(t)$ will then also be minimum phase. If $f(t)$ is not minimum phase, a stable inverse filter does not exist. See, for example, Neely and Allen [184] for a more elaborate discussion, in the context of inverting room impulse responses. The process of obtaining $\hat{x}(t)$ from $x_l(t)$ is called deconvolution, or inversion. In practice, there is a complicating factor in that the received signal $x_l(t)$ will be corrupted by additive noise, wherefore Eqn. 5.1 becomes

$$x_l(t) = x(t) * f(t) + \epsilon(t), \tag{5.3}$$

where $\epsilon(t)$ is the noise. The optimal filter, in the sense that $\hat{x}(t) - x(t)$ is minimized in least-squares sense, is then given in the frequency domain as (e.g. Berkhout *et al.* [29], [30])

$$G(f) = \frac{F^*(f)}{|F(f)|^2 + \sigma_\epsilon^2(f)}, \tag{5.4}$$

with $\sigma_\epsilon^2(f)$ the frequency-dependent variance (power) of $\epsilon(t)$, and $F^*(f)$ the complex conjugate of the Fourier transform of $f(t)$. This optimal filter is also called a Wiener filter, although the solution can be improved if the short-term spectrum $\Xi(f, k)$ of $x(t)$ is known (signal frame k), in which case the time-varying Wiener filter becomes

$$G(f,\ k) = \frac{\Xi(f, k)F^*(f)}{\Xi(f, k)|F(f)|^2 + \sigma_\epsilon^2(f)}. \tag{5.5}$$

Of course, $\Xi(f, k)$ is not known, but a long-term average spectrum $\overline{X}(f)$ might be known or estimable, and could be used as well. Then the estimated spectrum $\hat{X}(f)$ becomes

$$\hat{X}(f) = \frac{\{\overline{X}(f)\}^2|F(f)|^2}{\overline{X}(f)|F(f)|^2 + \sigma_\epsilon^2(f)}. \tag{5.6}$$

For high signal-to-noise ratio (SNR), the Wiener filter can be approximated as

$$G(f) \approx F^{-1}(f), \quad \sigma_\epsilon^2(f) \ll |F(f)|^2, \tag{5.7}$$

such that $g(t)$ is simply the inverse of $f(t)$ as we found previously, and does not depend on the signal spectrum. For low SNR this becomes, for the cases of Eqns. 5.4 and 5.5, respectively

$$\left.\begin{array}{rcl} G(f) & \approx & F^*(f)/\sigma_\epsilon^2(f) \\ G(f,\ k) & \approx & \Xi(f, k)F^*(f)/\sigma_\epsilon^2(f) \end{array}\right\} \quad \sigma_\epsilon^2(f) \gg |F(f)|^2, \tag{5.8}$$

which is the matched filter, as also known from signal detection theory, and does depend on the signal spectrum. For a given situation, the approximations in Eqns. 5.7–5.8 can be valid in different frequency bands. So those frequency bands having a low SNR will be strongly attenuated, and the effect of $f(t)$ can only be inverted if the SNR is high or intermediate. The conclusion is that for bandlimiting operations, deconvolution is not very effective, because the bandlimited frequency regions will usually have a poor SNR, and the signal in those bands can therefore not be retrieved. This is true for a telephone network, but, in particular, also for perceptually coded audio in which high frequencies have been eliminated to reduce the required bit rate. In those cases, the high-frequency band does not contain any useful signal any more, and other (non-linear) methods must be used to restore (some of) the original signal parts. Because in audio applications there is in general very little a priori information about the nature of the signals, more specialized deconvolution methods are not practical.

As an example, we use the situation of speech transmission through the telephone network. Although a full treatment of speech enhancement processing is deferred until Chapter 6, we use this example because (1) the bandlimitation (being the telephone network) is very well defined, and (2) the long-term average spectrum of speech is known. For general audio applications, both the bandlimitation as well as the signal spectrum are not well defined (e.g. have a large variability). It should however be understood that the forthcoming arguments regarding the limitation of deconvolution apply equally well to speech transmission through the telephone network as to more general situations. Thus, consider Fig. 5.1, which shows the approximate filtering characteristic of the telephone channel $F(f)$ (see also Chapter 6 and Fig. 6.1) in (a) (solid line), together with an assumed white noise spectrum $\sigma_\epsilon(f)$ at -30 dB (dashed line). Application of Eqn. 5.4 leads to a deconvolution filter $G_1(f)$, the magnitude of which is shown in (b) (solid line). Taking into account the long-term average spectrum of speech $S(f)$, shown in (c) (solid line, calculated using Eqn. 1.24 and assuming an average male pitch of 125 Hz), we find using Eqn. 5.5 a different deconvolution filter $G_2(f)$, also shown in Fig. 5.1 (b) (dashed line). We see that $G_2(f)$ has a broad peak around 150 Hz and a narrower peak at about 4 kHz. The final received speech signal $S_\mathrm{r}(f) = S(f)F(f)G_2(f)$ has a long-term average spectrum as shown in (c) (dashed line). In contrast to the speech signal, $S_0(f)$ would be received without any deconvolution filtering, shown as the dash-dotted line in (c), the effect of $G_2(f)$ is to accurately invert the telephone channel filtering down to about 200 Hz and up to about 4 kHz. Although this is an improvement, a lot of energy in the speech signal is still not recovered, as is clear from Fig. 5.1. In particular high frequencies, above 4 kHz (as contained mostly in fricatives such as /s/ and /f/), are not well reproduced. Note that the Wiener filter is optimal from a signal's point of view, and does not necessarily yield that linear filter that gives the best perceptual result.

We do not give examples for the case of the music signals, as music spectra are highly variable, and the bandlimiting is highly dependent on the coding algorithm and bit rate used (assuming the bandlimitation occurs through perceptual coding). The reader will understand that similar arguments as those made above for the example of speech through the telephone channel lead to similar conclusions in that deconvolution is not a practical method to restore highly bandlimited music signals.

5.3 PERCEPTUAL CONSIDERATIONS

In this section, the characteristics the synthesized high-frequency components should have to properly enhance audio reproduction, by considering pitch, timbre, and loudness are discussed. These considerations are useful to know what aspects of the available narrowband signal should be reflected in the high band; for the high-frequency BWE codec of Sec. 5.5 it also useful as it helps to realize what information the encoder should store (and what is irrelevant), to be used as a priori information for the decoder.

5.3.1 PITCH (HARMONIC STRUCTURE)

The pitch of a complex tone is determined by the frequencies of the constituent partials (harmonics), see Sec. 1.4.5. The strongest pitch percepts are obtained when low-order (resolved) harmonics are present, but complex tones with only high-order (unresolved)

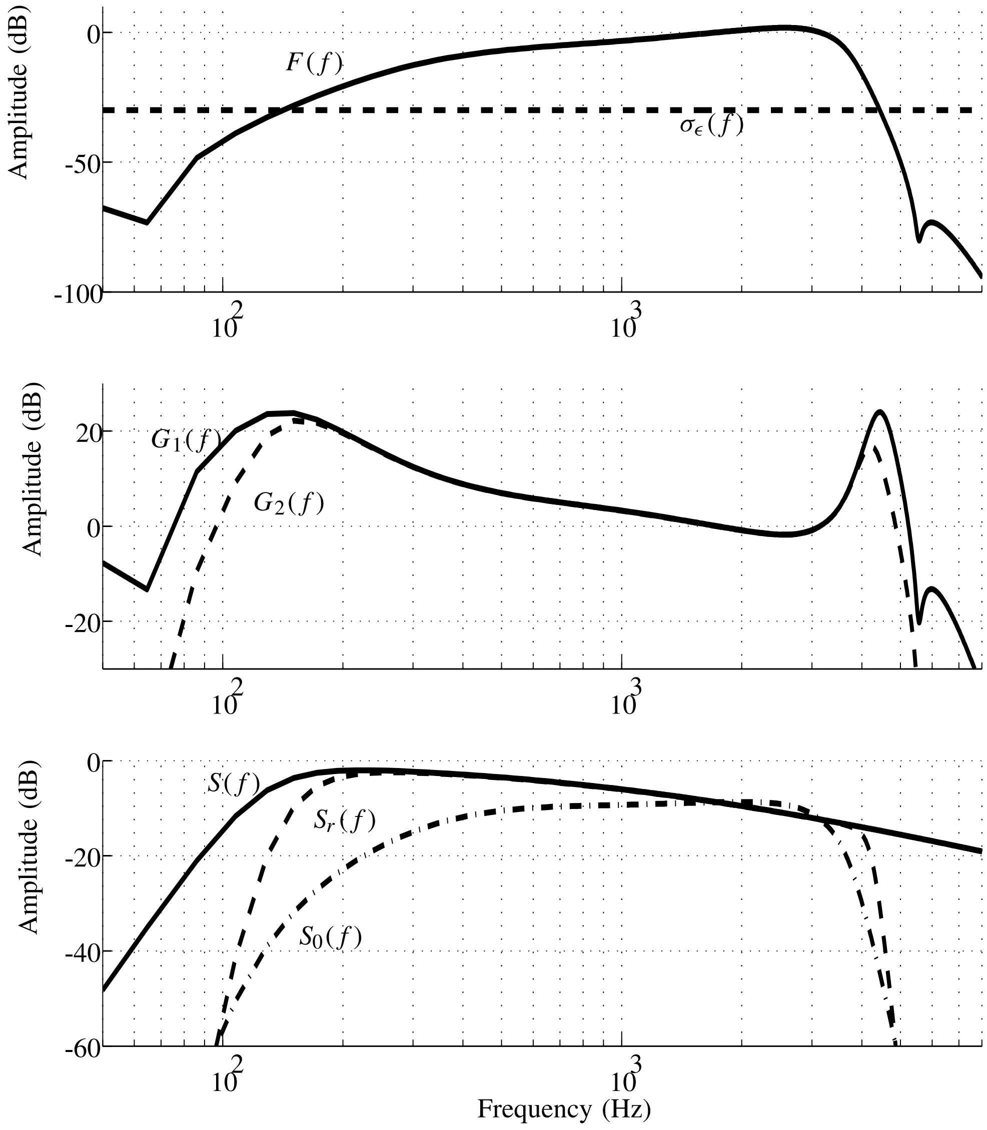

Figure 5.1 (a) Shows the approximate filter characteristic $F(f)$ of the telephone network (solid line), and an assumed white noise spectrum $\sigma_\epsilon(f)$ at -30 dB. (b) Shows the optimal filter $G_1(f)$ to invert this channel effect according to Eqn. 5.4 (solid line); taking into account the long-term average spectrum of speech $S(f)$ (shown in (c), solid line), the optimal filter $G_2(f)$ becomes as is shown by the dashed line (b), using Eqn. 5.5. Finally, (c) shows the long-term average spectrum of natural speech $S(f)$ and the received speech $S_0(f)$ without deconvolution filtering (dash-dotted line), and also for the received speech $S_\mathrm{r}(f)$ signal using the Wiener filter of Eqn. 5.5. It can be seen that the Wiener filter improves reproduction down to 200 Hz and up to about 4 kHz, but that a significant portion of speech energy (mainly that above 4 kHz) is not recovered

harmonics also yield pitch percepts, although weaker. Ritsma [227] determined that harmonics 3–5 are dominant in the perception of pitch. Such low-order harmonics would in most cases fall outside the frequency range in which synthetic frequency components are generated by high-frequency BWE algorithms, as the lower limit of this range is typically 4 kHz at least, but may be up to over 10 kHz. Therefore, frequency components added by high-frequency BWE would typically be unresolved harmonics.

We assume that the input signal $x(t)$ is a complex tone with fundamental f_0. If $x(t)$ is bandlimited, only a finite number N_b of harmonics will be present at integer multiples of f_0. The high-frequency BWE algorithm should add additional frequency components at kf_0, with $k = N_b + 1,\ N_b + 2,\ N_b + 3,\ \ldots$. Of course, without a priori information, the correct amplitudes of these harmonics are unknown, but in most cases a gradually decaying amplitude spectrum is suitable. By using a proper harmonics generator (non-linear device, see Sec. 5.4.1), it can be ensured that the harmonic amplitudes do indeed gradually decay.

If the generated frequency components do not fall onto the regular $k = N_b + 1,\ N_b + 2,\ N_b + 3,\ \ldots$ pattern, a variety of effects could occur. If the spacing between the partials is incorrect, or if the partials are shifted by an amount not equal to a multiple of f_0, the added harmonics will elicit a pitch at a different frequency. The signal comprising these high harmonics would then be heard separately from the original complex tone, that is, the signal bands segregate (see Sec. 1.4.7). Such a translation of the harmonic 'grid' could occur, for example, if the higher harmonics are generated through spectral folding (Sec. 6.3.3.1).

Not all musical signals contain harmonically related frequency components however, and particularly at higher frequencies noise-like signals can occur, for example, percussion. In such cases, the extended frequency spectrum should also be noise-like.

5.3.2 TIMBRE (SPECTRAL ENVELOPE)

The explicit goal of high-frequency BWE is to extend the spectral envelope to high frequencies, thereby modifying the sound's timbre; the loss of high frequencies is the reason that music and speech sounds muffled. As discussed in Sec. 1.4.6, timbre depends on a number of variables, including amplitude spectrum, phase spectrum, and temporal envelope (in particular, attack and decay times).

- The temporal envelope in a high-frequency band should be broadly similar to that in a low-frequency band for most typical audio signals, so it would suffice if the BWE algorithm ensures a more or less linear relationship between the two. This will be the case if the non-linear device (NLD) is a homogenous system (Sec. 1.1.1).
- The phase spectrum is considered to be the most unimportant aspect in high-frequency BWE applications. The lower limit of generated frequency components is typically 4 kHz (but possibly much higher), and at these frequencies the auditory system is fairly insensitive to phase. It is possible that phase changes of adjacent frequency components lead to modifications of the interference pattern produced at particular locations on the basilar membrane (BM). Because auditory filters broaden with increasing frequency (e.g. being 672 Hz wide at 4 kHz and 888 Hz wide at 8 kHz, according to Eqn. 1.88), the likelihood of such interactions increases at high frequencies. It is conceivable that certain phase changes could significantly change the overall amplitude of vibration

at particular locations of the BM, and this would cause a change in neural response (because neural response is directly linked to the amplitude of the BM vibration). In this way, the amplitude spectrum as sensed by the auditory system could be modified by phase changes of the physical signal spectrum. These effects are difficult to predict, in particular because the signals arriving at the ears have a more or less random phase (in the sense of a deterministically chaotic system) owing to the variations of the impulse responses between loudspeaker and ears. Also, a study by Plomp and Steeneken [210] found the effect of phase spectrum on timbre to be small compared to the effect of amplitude spectrum, although in that study complex tones containing low-order (<10) harmonics were used. For all these reasons, it is considered impractical and not necessary to control for the phase of the synthetic high-frequency components.

- The amplitude spectrum, directly determined by the amount of high frequencies added through the high-frequency BWE algorithm, is known to be important in determining timbre. As before in Secs. 2.2.2 and 3.2.2, we model only the brightness aspect of timbre that is closely linked to amplitude spectrum. Brightness is modelled by the spectral centroid C_S (Eqn. 1.95 in Sec. 1.4.6), with higher values for C_S implying a brighter sound percept.

  If the original full-bandwidth signal is known, the BWE algorithm should obviously be designed such that the reconstructed high frequencies closely match the original high frequencies in amplitude spectrum. For blind BWE algorithms, the original high-frequency spectrum is unknown, and the best approach is to smoothly ‘extrapolate’ the signal spectrum to high frequencies. Typically, audio signals have gradually decaying spectra (although resonances do occur). Such a ‘smooth’ extrapolation can be ensured by a proper choice of the NLD, as is discussed in Sec. 5.4.1

5.3.3 LOUDNESS (AMPLITUDE)

The loudness of the harmonics signal is directly related to its amplitude. However, if properly generated, the added harmonics will not be perceived separately, but as integral part of the original narrowband signal (e.g. grouping will occur). This would also be the case if the extended signal does not consist of regularly spaced harmonics, but is noise-like. Therefore, we should consider the effect on the loudness of the, originally narrowband, tone when adding higher harmonics.

Standardized loudness models such as ISO532A and ISO532B compute loudness on the basis of the long-term amplitude spectrum of the signal. The amplitude spectrum is specified in narrowbands (e.g. one-third octave bands for ISO532A and 0.1 Bark bands for ISO532B) followed by an integration over frequency, also allowing for masking effects. The details of each procedure differ and are explained in Sec. 1.4.4.2. The main conclusion is that accepted models of loudness perception only take into account the amplitude spectrum to compute loudness. Even Glasberg and Moore’s [90] more recent loudness model that can be used for time-varying signals only takes the short-term amplitude spectrum into account. So if a high-frequency BWE algorithm exactly reconstructs the amplitude spectrum of the high-frequency band, the reconstructed signal should have the same loudness as the original full-bandwidth signal. This is of course only possible for BWE algorithms that employ a priori information. For blind algorithms, the reconstructed high-frequency band will deviate from the original high-frequency band. Depending on the pattern and magnitude of these deviations, loudness of the reconstructed signal will not be identical to the original signal's loudness. However, some of these deviations might not be perceptible, as masking effects can reduce or eliminate the contributions of some frequency bands to the total loudness. Also, the largest contributions to loudness will derive from intermediate frequency bands, around 1–4 kHz, where absolute thresholds are lowest (ear is most sensitive). For typical high-frequency BWE applications, the synthesized high frequencies will have a lower limit of at least 4 kHz, and possibly much higher, so the entire contribution of the synthesized high frequencies is probably fairly small anyway.

5.3.4 EFFECTS OF HEARING LOSS

Figure 1.19 shows hearing loss for a group of otologically normal males of various ages (20–70 years) in terms of the 50th percentile points, as a function of frequency. For frequencies below about 1 kHz, the loss remains below 12 dB, but above 1 kHz the amount of loss increases rapidly. At 4 kHz, the amount of hearing loss for a 70-year-old male is, on average, about 42 dB, and at 8 kHz (the highest frequency that was included) this is as much as 60 dB. The same trend is observed for females, although somewhat smaller values for hearing loss are typical. The implication is that older persons, on average, will not perceive high frequencies contained in speech and music signals. The situation is probably more aggravated for the latter category, as music signals contain more energy at higher frequencies than does speech. For these persons, high-frequency bandwidth limitation might not be perceivable at all, and conversely, they might not detect any enhancement of the high-frequency spectrum obtained through high-frequency BWE processing. This is confirmed by the fact that for A/B tests of high-frequency BWE systems younger listeners often perceive clear enhancement, but no or little difference is detectable for older listeners (who did not use hearing aids). From the limited experience gained through informal tests of typical implementations of high-frequency BWE systems, this seems to be the case for persons in the age group of approximately 40 to 60 years old (and presumably older persons as well, although no listeners in that age group had been tested).

For individual listeners, this problem could be (at least partly) overcome by a linear filter that emphasizes signal energy in those regions where hearing loss is severe. But in almost all practical applications such flexibility is not implementable, and probably not even desirable, as reproduced signals can be intended for a group of listeners. In the latter case, a design must be sought that is the best 'on average', and definitely not annoying for any single listener. In practice, this probably means that a high-frequency BWE system would be designed to sound as good as possible for persons with no, or little, hearing loss (typically younger listeners); persons with high-frequency hearing loss (typically older listeners) will therefore, on average, benefit less or not at all, from high-frequency BWE processing (unless the hearing loss is properly compensated for by a hearing aid).

5.3.5 CONCLUSIONS

For a high-frequency BWE algorithm to resynthesize a signal with correct timbre and loudness, it suffices to match the spectral envelope of the original full-bandwidth signal. Correct reproduction of the spectral fine structure is essential for a proper grouping of the synthesized high frequencies with the low frequencies. Persons with significant high-frequency hearing loss will not benefit from high-frequency BWE methods.

5.4 HIGH-FREQUENCY BANDWIDTH EXTENSION FOR AUDIO

Although this whole chapter is devoted to high-frequency BWE for audio applications, the title of this section reflects that here we will discuss implementations analogous to the general structures discussed in Chapters 2 to 3. Other sections in this chapter present alternate structures for high-frequency BWE algorithms. The general structure presented here is shown in Fig. 5.2. Note the correspondence with BWE structures for low-frequency psychoacoustic BWE (Fig. 2.4) and low-frequency physical BWE (Fig. 3.1). Again, there are two branches, the lower of which simply delays the input signal $x(t)$ such that it is later added exactly in phase with the processed signal from the upper branch. The processing consists of two filters and a non-linear device (NLD). The first filter, FIL1, extracts the highest octave present in $x(t)$, which is then the input for the NLD. The non-linear processing generates a harmonics signal, which is filtered by FIL2 to obtain a suitable spectrum. After scaling, the resulting signal is added back to $x(t)$ to yield the bandwidth-extended output $y(t)$. In the remainder of this section, we will explore the various processing steps in more detail.

Note that the signal $x(t)$ must have enough 'empty' bandwidth at the high-frequency end to synthesize the higher harmonics. At a sample rate f_s, the highest frequency present in the signal is maximally equal to the Nyquist frequency, $f_N = f_s/2$. If $x(t)$ contains energy at frequencies higher than $f_N/2 = f_s/4$, then $x(t)$ first needs to be upsampled. In all cases, there must be at least one additional octave above the highest frequency of $x(t)$.

5.4.1 NON-LINEAR DEVICE

To ensure that the synthetic high-frequency band covaries in amplitude with the bandlimited input signal, it is necessary to use a harmonics generator that is homogenous, i.e. scales the output proportional to the input (Sec. 1.1.1). This also has the beneficial property that the relative amount of harmonics generated is independent of the input level. We intend to extend the bandwidth of $x(t)$ by one octave; this yields a significantly

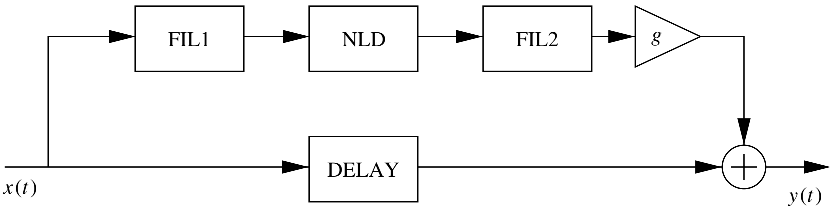

Figure 5.2 High-frequency BWE system. FIL1 extracts the highest octave present in $x(t)$, and harmonics of this signal are generated by NLD. The harmonics spectrum is shaped by FIL2. The harmonics signal is then scaled and added to the delayed input signal to form the bandwidth-extended signal $y(t)$

brighter percept, yet does not suffer from artefacts that have sometimes been observed when extending the bandwidth even further.

As FIL1 extracts the highest octave of $x_(t)$, which we denote as $x_0(t)$, the NLD must therefore double the frequencies present in $x_0(t)$, leading to the harmonics signal $x_\mathrm{h}(t)$. Frequency doubling can be very efficiently done through rectification, as the spectrum of a rectified pure tone consists mainly of its double frequency; the other components are higher even harmonics, but these decay by 12 dB per octave (Eqn. 2.19 in Sec. 2.3.2.2) and are thus quite weak. A rectifier is also a homogenous system. The intermodulation distortion of a rectifier, given a two-tone input signal, was analysed in Sec. 2.3.2.2, and displayed in Fig. 2.9. It was shown that the relative amount of intermodulation distortion could be fairly high if multiple components of comparable amplitude are present in the input signal. The results of this analysis cannot be directly used for high-frequency BWE though, as in all cases the used expression was for continuous-time implementations of the rectifier. As those sections dealt with low-frequency psychoacoustic BWE applications, that approach was valid, because the frequencies of interest were at least two orders of magnitude smaller than the sample rate. It can be shown that in the limit of very high sample rates, expressions for NLD output spectra in continuous and discrete time are equal. In the present case however, the frequencies of interest are in the same order of magnitude as the sample rate. Specifically, if the spectrum of $x_0(t)$ lies in the range $[f_\mathrm{s}/8,\ f_\mathrm{s}/4]$, which would be fairly typical, high harmonics could only be added up to $f_\mathrm{s}/2$ maximally. Thus, harmonics higher than the second harmonic cannot even exist in such cases. This obviously alters the expressions for the output spectrum of a rectifier, and also necessitates a re-evaluation of the relative amount of intermodulation distortion, given a multiple-component input. The latter also differs from low-frequency psychoacoustic BWE applications in another aspect, namely, for low-frequency psychoacoustic BWE it is probably reasonable to assume that no more than two frequency components will be present at the input to the NLD, as FIL1 in that case is typically a band-pass filter with a bandwidth of about 50 or 100 Hz. For high-frequency BWE applications, FIL1 is a band-pass filter with a bandwidth of several thousand hertz, and will in most cases contain many frequency components. The correct expression for computing the output spectrum of a rectifier in discrete time, given an arbitrary periodic input signal, is Eqn. 2.116. The expression is fairly complex, and with the added variability of input signals (in terms of number of components, and their frequencies), we have not derived expressions to evaluate the amount of intermodulation distortion energy relative to the amount of harmonic energy, as was done for the simpler case in Sec. 2.3.2.2. Rather, the quality of the bandwidth-extended signal has been judged perceptually, using a variety of repertoire, and the performance is generally considered to be good.

5.4.2 FILTERING

As with low-frequency psychoacoustic BWE and low-frequency physical BWE systems, for high-frequency BWE the signal applied to the NLD needs to be a specific frequency band, and the output of the NLD has to be shaped properly to yield a proper timbre. Thus, it is necessary to use filters before and after the NLD, as in Fig. 5.2.

For both reasons mentioned in Sec. 2.3.3.3, it is beneficial to use linear-phase filters. The first reason was that non-linear-phase filters can lead to interference between the processed (harmonics) signal and the original bandlimited signal, in the limited frequency band where FIL1 and FIL2 overlap. As this is only a small frequency region, this might not be as important as with low-frequency psychoacoustic BWE systems. The other reason is that non-linear-phase filters can give rise to large variations in group delay, which might lead the synthetic high-frequency signal to group poorly with the lower-frequency bandlimited input signal. If FIL1 and FIL2 are both linear phase, their processing delay can be exactly compensated for by a delay of the input signal, such that both harmonics signal and input signal can be added in phase to form the bandwidth-extended output signal. Because the bandwidths of FIL1 and FIL2 are fairly large compared to the sample rate, it is feasible to implement these using FIR filters. Alternatively, IIR filters can be used in the method as described in Sec. 2.3.3.3.

Because a typical application for a blind high-frequency BWE system would be to enhance bandlimited signals as received from, for example, Internet radio, the bandwidth of the incoming signal is not known a priori. Therefore, the passbands of FIL1 and FIL2 need to be adjustable to be able to adapt to whatever the momentary signal bandwidth is. Two methods could be used to implement high-frequency BWE, depending on the bandwidth of the input signal. The first method simply assumes that the signal bandwidth is equal to the Nyquist frequency, that is, half the sample rate. Therefore, the input signal is first up-sampled by a factor of 2, after which the additional octave is ‘filled’ with the synthesized higher frequencies. Although this method is not guaranteed to work because of the simple assumptions, in practice, for Internet Radio applications it has demonstrated to work quite well. A second, in principle more reliable, method is to analyse the energy content of the signal in various frequency bands, for example, through a number of broad band-pass filters. In most cases, high-frequency BWE will be applied to perceptually coded audio, and in those cases the bandwidth of the signal can be detected by analysing the coefficients of the encoded audio stream directly.

5.4.2.1 Filter 1

Assume that the bandwidth of the input signal is known (or estimated), and the highest frequency component present is f_{h}. Further, assume that the sample rate $f_{\mathrm{s}} \geq 4f_{\mathrm{h}}$, possibly through upsampling prior to BWE processing. As the NLD, being a rectifier, generates second harmonics of the input signal, FIL1 should be band pass between $f_{\mathrm{h}}/2 - f_{\mathrm{h}}$. The high-pass flank of FIL1 can be designed as a second-order filter, while for the low-pass flank a somewhat higher order, say fourth order, is better. This prevents frequencies $f' > f_{\mathrm{s}}/4$ from entering the NLD (only if $f_{\mathrm{h}} \approx f_{\mathrm{s}}/4$); if such frequencies did enter the NLD, they would end up as aliased components at low frequencies, because $2f' > f_{\mathrm{s}}/2$. Although any low-frequency component generated by the NLD would be filtered out by FIL2, it is generally beneficial to keep the number of frequency components entering the NLD as small as possible, to minimize intermodulation distortion.

FIL1 could, in principle, be implemented as a filterbank, with each output driving a separate NLD, the aim of which would be to minimize intermodulation distortion. Some informal testing revealed that this strategy does not seem to lead to a significantly better-quality signal, however.

5.4.2.2 Filter 2

The input signal for FIL2 is the harmonics signal as processed by the NLD. Because the input of the NLD is a frequency band $f_h/2-f_h$ (the highest octave present in the bandlimited input signal), the NLD output consists primarily of the second harmonics of these components, that is, the frequency band f_h-2f_h. However, there will be intermodulation distortion components at frequencies below f_h, which have to be eliminated. Therefore, FIL2 has a high-pass flank of at least fourth order. This ensures that the synthesized frequency components are only added at frequencies higher than those contained in the input signal. Depending on the sample rate f_s, a low-pass flank may or may not be required. If $f_s = 4f_h$, then the harmonics signal extends maximally up to $2f_h = f_s/2$, and a low-pass flank is not required. If the sample rate is higher, a low-pass flank can be implemented at a cut-off frequency of $2f_h$. The order of this low-pass flank can be quite low, as the harmonics signal generated by the NLD (rectifier) decays rapidly.

Figure 5.3 shows an example implementation of both FIL1 and FIL2. The input signal has a bandwidth $f_h = 4$ kHz, and the sample rate has been converted to 16 kHz. FIL1 is a Butterworth band-pass filter from 2 to 4 kHz, with a second-order high-pass flank and an eighth-order low-pass flank. FIL2 is a high-pass filter at 4 kHz, with an eighth-order flank. It is not necessary to implement a low-pass flank for FIL2, as frequencies higher than 8 kHz do not exist (as the sample rate is 16 kHz).

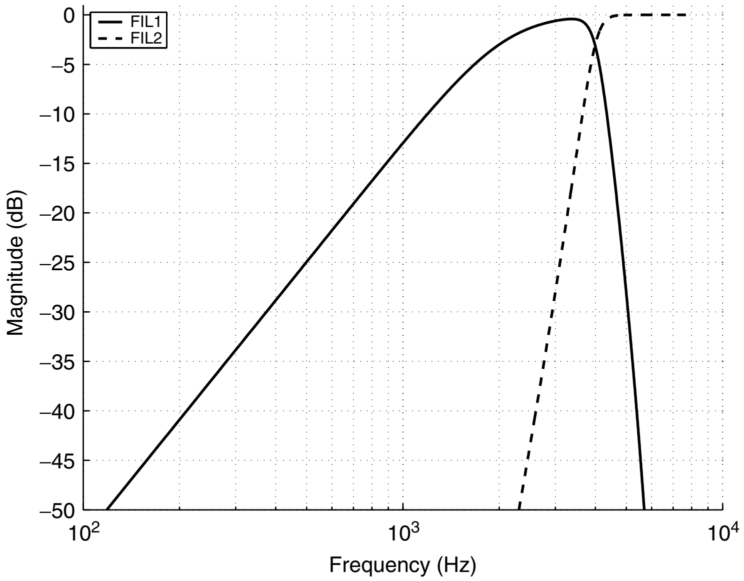

Figure 5.3 Example implementations for FIL1 and FIL2 of a high-frequency BWE system. It is assumed that the highest frequency present in the input signal is 4 kHz, such that FIL1 extracts the highest octave therein. The NLD predominantly generates second harmonics of these components, after which FIL2 ensures that any intermodulation distortion component below 4 kHz is removed

5.4.3 GAIN OF HARMONICS SIGNAL

As the high-frequency spectrum is not known a priori in a blind high-frequency BWE system, the gain of the high-frequency spectrum relative to the low-frequency spectrum is unknown. In practice, a gain value must be chosen that sounds well 'on average'. It is thus inevitable that on many occasions the high-frequency spectrum is either too strong or too weak, compared to the actual high-frequency spectrum. However, these deviations are not excessive, and in nearly all cases the bandwidth-extended signal is judged as more natural compared to the bandlimited signal.

A conceptually simple improvement would be to use an adaptive gain. As there is no a priori information, the control signal for the gain variations would have to be derived from the bandlimited input signal. Some preliminary experiments indicated that this could lead to a more accurate high-frequency percept. Specifically, the gain control signal was derived by matching the energy of the artificially generated high-frequency band to the energy of the actual high-frequency band, short (~20 ms) time frames. This gain control signal was then used to scale the synthesized frequency components. Obviously, this is not possible in an actual application, but it demonstrated that it is possible to improve the quality of the described blind high-frequency BWE algorithm by relatively simple means. Figure 5.4 shows an example, in which a 10-s fragment of pop music, bandlimited to 11 kHz, was processed by the described high-frequency BWE algorithm (FIL1: 5.5–11 kHz, FIL2: high pass at 11 kHz, NLD: rectifier); signal energy in 20-ms frames was computed for the synthetic high frequencies. This was compared to the actual signal energy above 11 kHz (the full-bandwidth signal was also available), the result of which is shown in the figure.

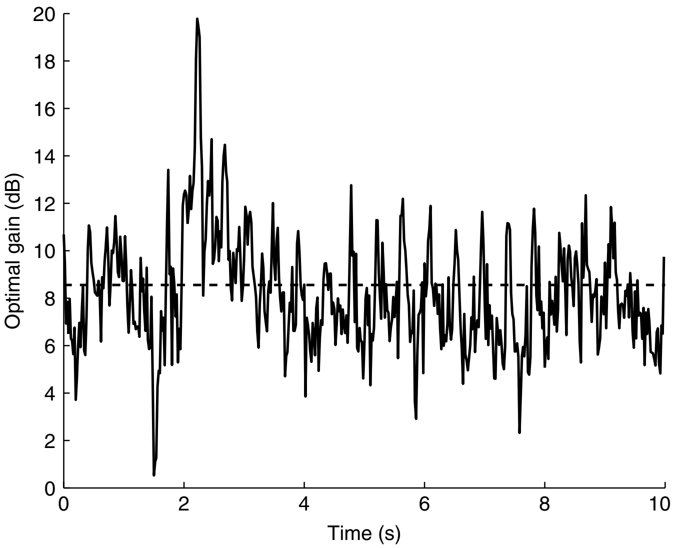

Figure 5.4 'Optimal gain' calculated for a 10-s pop music signal. The gain was derived by comparing the actual signal energy above 11 kHz with the energy of the synthetic high-frequency signal as generated by the high-frequency BWE algorithm. The dashed line indicates the mean (~8.2 dB)

This 'optimal gain' (in the sense that it matches the high-frequency energy) varies around a fairly stable mean value of ~8.2 dB, although occasional large deviations occur (e.g. around 2.5 s). Although the 'mean optimal gain value' varies per repertoire, a 'grand mean value' can be chosen such that fairly good results are obtained for most repertoire. Note that appreciation of individual listeners will vary, which may in part be due to differences in hearing loss at high frequencies (Sec. 5.3.4).

The problem in deriving a practical gain control signal using only information from the bandlimited input signal might conceptually be solved in similar fashion as is done for speech BWE algorithms. Section 6.5 describes what features of the narrowband speech signal are thought to carry some information regarding the high-frequency spectral envelope. For speech, this approach works reasonably well, but it might be much more difficult for music, as music signals have a much larger range of variability than speech signals. Also, the amount of speech bandwidth limitation is well defined through the telephone channel (being about 300–3400 Hz), but this is not the case for the more general situation where bandwidth limitation occurs through perceptual coding. Depending on the bit rate and the coder implementation, the bandwidth can vary from less than 4 kHz to full bandwidth (~22 kHz). For each degree of bandlimitation, another set of parameters would have to be defined to translate narrowband signal features to high-band spectral envelope. Therefore, it remains to be seen if such an approach could work, while remaining practically feasible, for general audio applications.

5.5 SPECTRAL BAND REPLICATION (SBR)

Spectral Band Replication (SBR) is a technique to enhance the efficiency of perceptual audio codecs (Ekstrand [111], Kunz [153], Schug *et al.* [241]). High-frequency components of an audio signal are reconstructed from low-frequency components by the decoder, such that the encoder need only encode the low-frequency part. In this fashion, a bit-rate reduction can be achieved while maintaining subjective audio quality. The basic idea of SBR is based on the observation that characteristics of high-band signals typically exhibit quite a high correlation with those of the lowband signals. Therefore, it is often possible to replace the high band with a transposed version of the lowband, avoiding the need to transmit the high-band signal at all. This can obviously reduce the required bit rate. SBR encodes a bandlimited version of the audio signal using conventional means, and then recreates the high band in the decoder. The difference with blind methods, such as those discussed in Sec. 5.4, is that the encoder provides a very small amount of additional control information (5–10% of the total), which the decoder uses to shape the high-band spectrum. This process is illustrated in Fig. 5.5. The control information is multiplexed with the encoded data into a single bitstream; the decoder first de-multiplexes the bitstream, decodes the lowband signal, and uses a high-frequency BWE algorithm to recreate the high-band signal, thereby using the control data to optimize the BWE processing.

The most important part of the SBR data is the information describing the spectral envelope of the original high-band signal (Dietz *et al.*[61]). Its main design goal is to use it as an equalizer without introducing annoying aliasing artefacts, and to provide good spectral and time resolution. The core algorithm of SBR consists of a 64-band, complex-valued polyphase filterbank (QMF). At the encoder side, an analysis QMF is

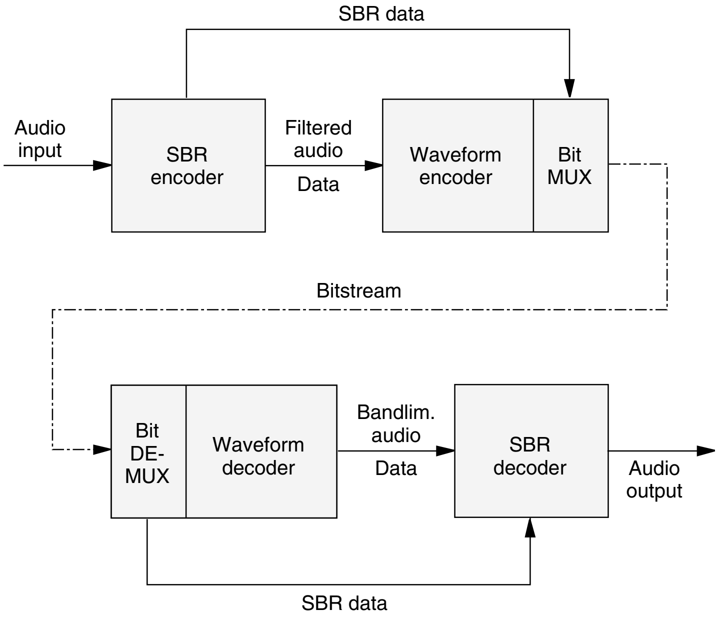

Figure 5.5 Spectral band replication (SBR) acts as pre-processor at the encoder and as post-processor at the decoder. A small amount of control data is provided along with the encoded lowband signal, which the decoder uses to optimize the high-frequency BWE algorithm

Table 5.1 SBR data rates for a number of example configurations. The total bit rate (audio coding and SBR data) is shown in the left-hand column, the number of used frequency bands in the middle column, and the SBR data rate in the right-hand column

Bit rate mono [kb/s]	SBR freq. range # QMF bands	SBR data rate [kb/s]
16	21	1.2
24	24	2.0
32	29	2.5
48	32	3.5

used to obtain energy samples of the original input signal's high band, which are used as reference values for the envelope adjustment at the decoder side. In order to keep the overhead low, the bitstream format of aacPlus[2] allows to group the QMF bands into scalefactor bands. By using a Bark-scale-oriented approach, grouping frequency bands may result in wider scalefactor bands the higher the frequency gets. Table 5.1, from

[2] The combination of AAC with SBR is named aacPlus, which is a registered trademark of Coding Technologies.

Dietz *et al.* [61], shows typical SBR data rates for a number of example configurations. The SBR method is obviously non-blind, as control parameters are used to create the high-frequency signal; however, a blind mode is possible as well, as explained in the patent discussed in Chapter 8.

In some cases, subjectively unsatisfactory results are produced when the low- and high-frequency bands are weakly correlated. This can occur with signals that are predominantly harmonic in the low-frequency range, but more noise-like in the high-frequency range (or vice versa), for example, having tonal instruments at low frequencies together with a hi-hat or cymbals at high frequencies. In such cases, additional information is encoded to indicate the need for synthesizing additional noise or additional tonal components at the decoder, such that the reconstructed high band will be similar to the original.

The combination of SBR technology with the conventional waveform audio coder standardized in MPEG, Advanced Audio Coding (AAC), is discussed in Ehret *et al.* [62]. With this enhanced audio coding scheme, called aacPlus, it is possible to achieve high-quality stereo audio at bit rates as low as 40 kb/s. The structure of the aacPlus decoder is shown in Fig. 5.6. After demultiplexing the aacPlus bitstream, the standard AAC bitstream is converted into a bandlimited audio signal. Then the SBR decoder generates high frequencies from the QMF-filtered bandlimited audio, ensuring a proper spectral envelope by using the SBR data. The high- and low-band QMF signals are then synthesized into a full-bandwidth output signal.

SBR technology is especially interesting in applications in which very high compression efficiency is desired, usually motivated by cost or physical limitations. Examples of such application areas are digital broadcasting and mobile applications. An overview of the latest developments with respect to the standardization process of aacPlus within MPEG-4 and subjective verification results are given in Ehret *et al.* [62], while implementations are described in Homm *et al.* [112].

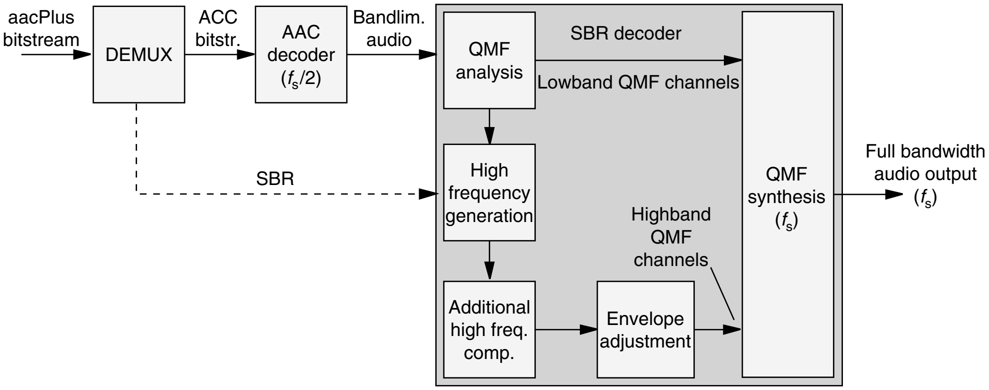

Figure 5.6 Block diagram of the aacPlus decoder. After demultiplexing the aacPlus bitstream, the standard AAC bitstream is converted into a bandlimited audio signal. Then the SBR decoder generates high frequencies from the QMF-filtered bandlimited audio, ensuring a proper spectral envelope by using the SBR data. The high- and low-band QMF signals are then synthesized into a full-bandwidth output signal

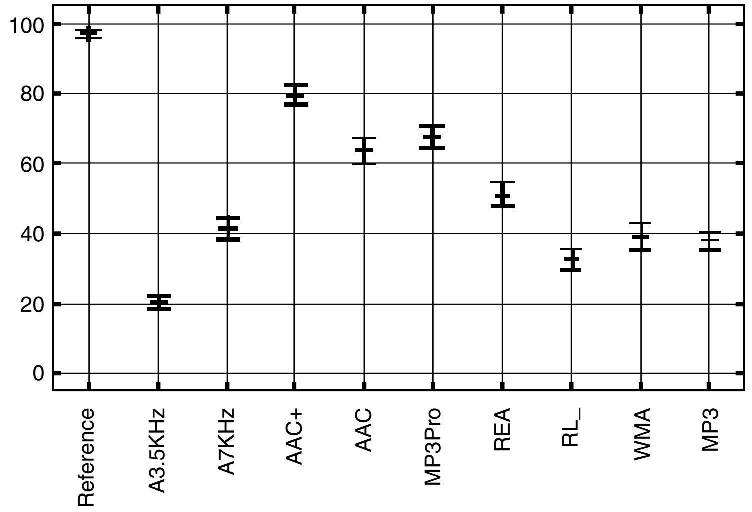

Figure 5.7 MUSHRA test results for various codecs at a stereo bit rate of 48 kb/s. The best result was obtained by aacPlus, followed by MP3Pro, which both use SBR technology

aaCPlus has been subjectively evaluated by several independent listening test sites. The results of all these tests have shown that aacPlus is a very good codec. For example, Fig. 5.7 shows the results of a MUSHRA test[3], carried out in the course of the EBU Internet audio evaluation, in which several audio-coding schemes were compared (cited by Dietz *et al.* [61]). Eight codecs were tested, and their results can be compared to two standards, namely the original, or reference signal (score nearly 100), and a 3.5-kHz low-pass filtered signal (score 20). The figure shows the result of the test for a stereo bit rate of 48 kb/s. The aacPlus decoder was judged as yielding the highest quality signal, with a score of 80 (on the border between 'good' and 'excellent'). The average score of the other codecs was about 50 ('fair'). The score of the core AAC codec was about 65. Another codec that has been integrated with SBR is MP3, called MP3Pro (Gröschel *et al.* [101]). Figure 5.7 shows that MP3Pro has the second-highest test score, nearly 70. The core MP3 coder received a score of just below 40.

5.6 HIGH-FREQUENCY BANDWIDTH EXTENSION BY INSTANTANEOUS COMPRESSION

5.6.1 INTRODUCTION AND ALGORITHM

A special form of BWE can be achieved by audio compression. This approach is especially suitable for multi-channel sound reproduction, in which the processed signals are predominantly speech or special effects. One of the disadvantages of multi-channel material is

[3] The MUSHRA scale range is 0–100, where 0–20 means 'bad', 20–40, 'poor', 40–60, 'fair', 60–80, 'good' and 80–100, 'excellent audio quality'.

that the surround-sound signal is often at a very low level. If the surround signal is simply linearly increased, it can become too dominant, or even lead to audible distortion in either the amplifier or loudspeaker. The same is true for the centre signal, which is often used for dialogues. Here we develop and analyse in instantaneous compression, algorithm that can enhance signals for centre and surround channels. This method is not generally applicable, because for music it does not yield good results; therefore the algorithm is not applied to the left/right loudspeakers of the multi-channel system.

Whereas the initial goal of the described compression algorithm was to overcome the problems of low signal level in centre and surround channels, it was also realized that it is a special kind of a BWE system. At high signal levels, where compression is most active, harmonic frequencies are generated and add some ‘brilliance’ to the sound. In contrast to what is normally desired of a BWE system, this ‘BWE compressor’ is not a homogeneous system (i.e. it does not scale its output proportionally to its input, see Sec. 1.1.1), and it is most effective at high signal levels (while e.g. low-frequency psychoacoustic BWE should be more effective at low signal levels, see Sec. 2.3.4). It is also different from other BWE algorithms as it uses the entire bandwidth of the input signal to generate harmonics, that is, the BWE compressor consists only of a non-linear device (NLD), without pre- or post-processing.

The BWE compressor uses a function that has a gain at low and moderate signal levels, but an attenuation at high signal levels. It is different from more usual compressors in that it is memoryless, that is, it is an instantaneous compressor. Any anti-symmetric monotonous function with a positive but decreasing derivative can be used in principle. During experiments, it appeared that the function

$$y(x) = c_1 \tanh(c_2 x), \tag{5.9}$$

plotted in Fig. 5.8 (for $c_1 = c_2 = 1$) is a suitable choice. The constant c_1 determines the maximum output level and c_2 determines the gain at low signal levels. During experiments, it appeared that for $|x| \leq 1$ suitable values for these constants are $c_1 = 0.763$, and $c_2 = 4.19$. For these values, the instantaneous input–output function is shown in Fig. 5.9. Using the Taylor series expansion,

$$\tanh(x) = x - \frac{x^3}{3} + \frac{2x^5}{15} - \frac{17x^7}{315} + \cdots \quad \text{for } |x| < \pi/2, \tag{5.10}$$

we get for small x that $y/x = c_1 c_2 \approx 10\,\text{dB}$ for the given values of c_1, c_2.

5.6.2 *ANALYSIS OF HARMONICS GENERATION*

In order to study the bandwidth extension of the NLD given by Eqn. 5.9, we assume an input signal $x(t) = A\sin(2\pi t)$, and calculate the coefficients b_n of the Fourier series of $y(x)$

$$\tanh(A \sin 2\pi t) = \sum_{n=0}^{\infty} b_n \sin 2\pi(2n+1)t, \tag{5.11}$$

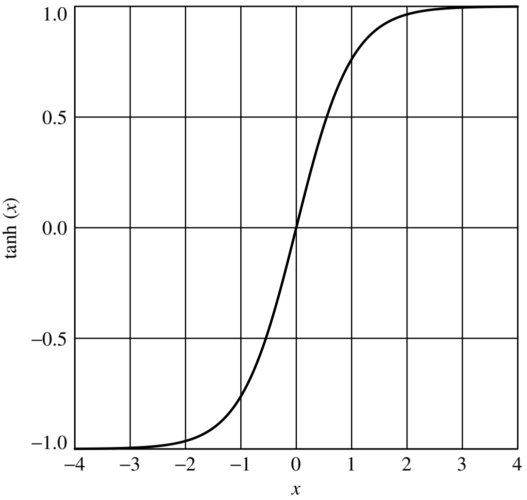

Figure 5.8 The function tanh(x), used as BWE compressor (Eqn. 5.9)

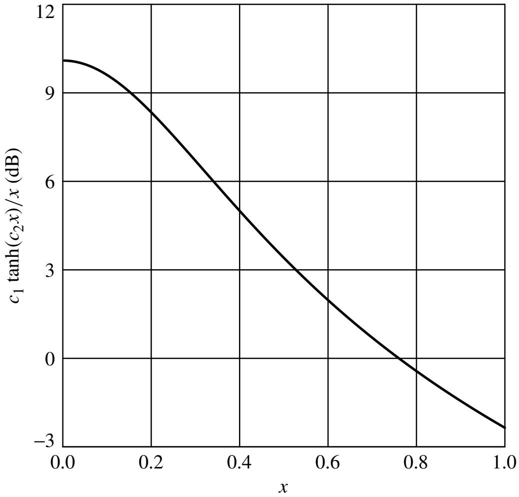

Figure 5.9 The instantaneous input–output ‘transfer function’ of the BWE compressor: $c_1 \tanh(c_2x)/x$, for $c_1 = 0.763$ and $c_2 = 4.19$. These values are suitable for signals that are ±1 at full scale

After some calculations[4] involving the calculus of residues, we find

$$b_n = \frac{8}{A}\sum_{k=0}^{\infty}\frac{1}{u_k^{2n}(1+u_k^2)}, \tag{5.12}$$

[4] Private communication with A.J.E.M. Janssen, Dec. 2002.

and

$$u_k = \frac{\pi(k+1/2)}{A} + \left(1 + \left(\frac{\pi(k+1/2)}{A}\right)^2\right)^{\frac{1}{2}}, \quad k \in \mathbb{N}_0. \tag{5.13}$$

For large A, we can approximate the b_n by ignoring terms containing powers of $1/A$ higher than five, which yields

$$b_n = \frac{4}{\pi}\frac{1}{2n+1} - \frac{(2n+1)\pi}{6A^2} - \frac{7(2n+1)\pi^3}{60A^4}(2 - \frac{1}{6}n(n+1)), \quad n \in \mathbb{N}_0. \tag{5.14}$$

For very large A, the output signal will tend to a square wave, which can be made explicit by taking $\lim_{A\to\infty}$ and showing

$$\lim_{A\to\infty} b_n = \frac{8}{A}\sum_{k=0}^{\infty}\frac{1}{u_k^{2n}}\frac{1}{(1+u_k^2)} = \frac{4}{\pi}\frac{1}{2n+1}, \tag{5.15}$$

so that we get, as we should, the familiar Fourier series coefficients of a square wave. On the other hand, for very small A, the output and input signals are proportional, because we get $b_0 = A$ and $b_k = 0$ for $k \in \mathbb{N}\backslash\{0\}$. This was also directly obvious from Eqn. 5.10.

5.6.3 IMPLEMENTATION

With analog components, Eqn. 5.9 can be easily implemented, using a long-tail pair with two transistors. On a digital platform, there are several possibilities. If the platform used is capable of directly implementing Eqn. 5.9, this would be the easiest way. If this is not the case, Eqn. 5.9 can be approximated by a power series. The Taylor series expansion of Eqn. 5.10 is not suitable, since this is only accurate for small $|x|$, while we are interested in the range $|c_2x| \leq 1$ (where $|x| \leq 1$). Therefore, we use a power series with an ℓ_∞−norm using a NAG [182] routine, based on a Chebyshev approximation (Barrodale and Phillips [25]). It appears impractical to use only one polynomial for the whole range, and therefore we use (for the case that $c_2 = 4.2$) two ranges, namely $|c_2x| \leq 1$ and $|c_2x| > 1$. This yields the following result

$$\tanh(x) \approx \hat{y}(x) = x\sum_{k=0}^{3} a_k x^{2k} \qquad \text{for} \quad |x| \leq 1 \tag{5.16}$$

and

$$\tanh(x) \approx \hat{y}(x) = \text{sign}(x)\sum_{k=0}^{7} b_k |x|^k \qquad \text{for} \quad 1 < |x| \leq 4.2 \tag{5.17}$$

where $z = c_2x$.

The order of the approximation is chosen such that the maximum error is equal to about 2^{-15}, which is suitable for 16-bit systems. If a lower or higher degree of approximation

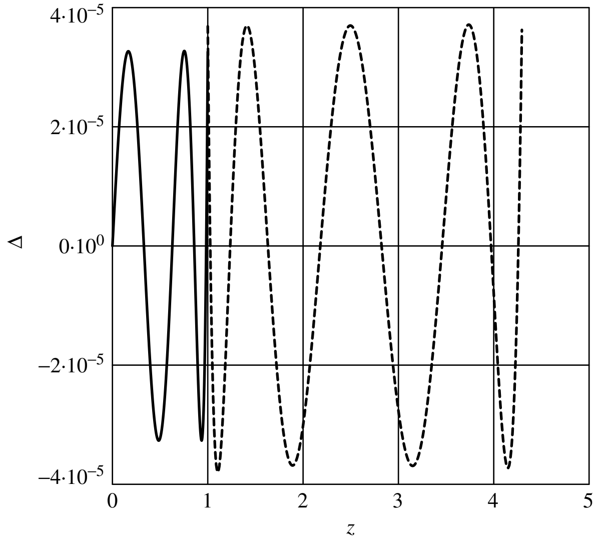

Figure 5.10 The error $\Delta = \tanh(z) - \hat{y}(z)$. The solid line shows the approximation error of Eqn. 5.16; the dotted line shows the approximation error of Eqn. 5.17. The coefficients a_k and b_k are as given in Sec. 5.6.5

is required, all coefficients a_k and b_k have to be recomputed again, since, as opposed to a Taylor series, the coefficients of a Chebyshev approximation depend on the order of the approximation. The approximation error, using Eqns. 5.16 and 5.17, and the coefficients a_k and b_k as given in Sec. 5.6.5, is plotted in Fig. 5.10.

5.6.4 EXAMPLES

Here we present some example signals and their processed versions, to illustrate the effects of BWE compression processing. Figure 5.11 shows four histograms displaying the amplitude distribution of two different input and output signals. All signal values were contained in $[-1,\ 1]$, and the processing used coefficients $c_1 = 0.763$ and $c_2 = 4.19$. The histograms have 50 equally spaced bins, and the value displayed for each bin is the log (base 10) of the number of occurrences that the signal value was in the bin range. Figure 5.11 (a) shows the amplitude distribution for a 60-s fragment of a pop music signal, which is nearly full scale, assuming that the transducer limits are ±1. A linear amplification of this signal would lead to clipping distortion. Part (c) shows the signal distribution after BWE compression. It is obvious that the maximum values of the signal have been reduced to $c_1 = 0.763$, and that the overall distribution has become flatter (as low-valued samples of the signal have been amplified). Part (b) displays the amplitude distribution for the same signal as in part (a), but scaled down by a factor of 10. BWE compression leads to an amplitude distribution as shown in part (d). Because of the small signal values, the compressor operates in its linear region, and as

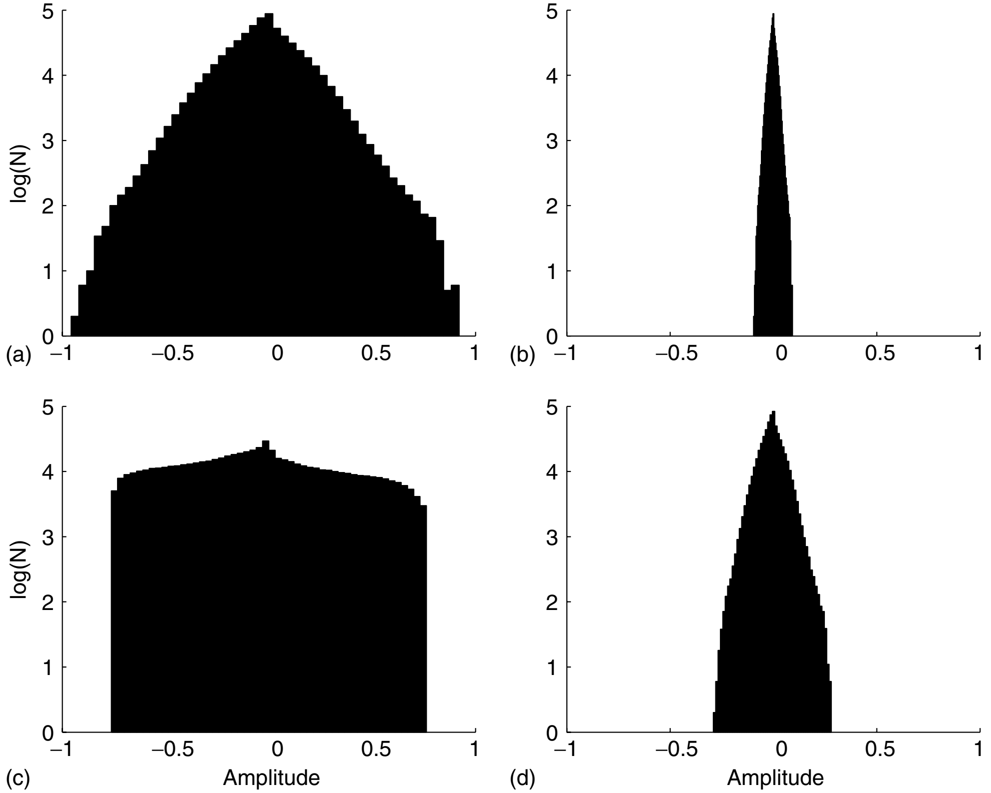

Figure 5.11 Histograms of a 60-s excerpt of music. Histogram values indicate number of occurrences per bin (log–base 10) of the signal value. Part (a) shows a signal that is close to full scale, and part (c) shows the BWE-compressed version. Note that the maximum signal value has decreased and the distribution has become flatter. Part (b) shows the signal of part (a), but scaled by a factor of 0.1. Part (d) is its BWE-compressed version, which shows only a linear scaling and no change in shape of the distribution (no flattening)

a result the signal distribution does not change shape (it has not become flatter as in part (b)).

It is also instructive to visualize the modifications generated by BWE compression in the time–frequency domain. Figure 5.12 (a) shows the spectrogram of the first 10 s of the input signal (the amplitude distribution of which is shown in Fig. 5.11) (a); in all spectrograms, black indicates high energy, and white indicates low energy (dB scale). Note that there is a gradual roll-off above 3 kHz and an abrupt high-frequency limit at about 5 kHz. The spectrogram of the BWE-compressed signal, shown in part (b), displays much more energy in the high-frequency region, up to the Nyquist frequency. The transients (recognizable as dark vertical lines) are clearly enhanced. Also, some of the complex tones have enhanced harmonics, for example, the harmonics below 1000 Hz, just before

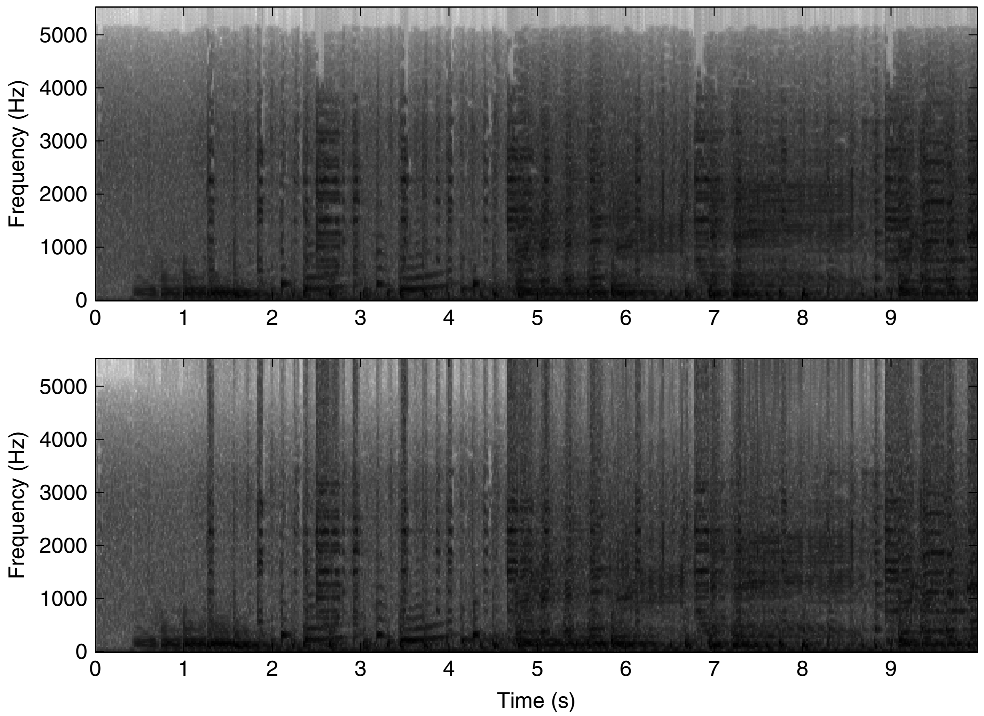

Figure 5.12 Spectrograms of a 10-s excerpt of music (the amplitude distribution of which is shown in Fig. 5.11) (a), and its BWE-compressed output in part (b). The input signal is somewhat bandlimited, whereas the output shows enhanced transients and more high-frequency content

4 s; although difficult to see in this fashion, it can be clearly observed when switching the spectrograms on a computer screen.

Figure 5.13 shows two spectrograms, the upper one displaying the time–frequency energy distribution of the same input signal as previously, but scaled down by a factor of 10 (as in Fig. 5.11 (b)). Note that all signals have been normalized before time–frequency analysis, such that any changes in the spectrograms are not due to overall level effects, but indicate relative changes in energy distribution in the time–frequency domain. Figure 5.13 (b) shows the BWE-compressed output, and exhibits only a modest enhancement of high frequencies, as we would expect, given that for low levels the compressor operates in its (near) linear regime. Also, the enhancement of low harmonics is not as pronounced as for the higher-level signal of Fig. 5.12.

5.6.5 APPROXIMATION OF THE FUNCTION tanh(**Z**)

In order to derive an approximation of the function tanh(x), a power series with an ℓ_∞–norm and a NAG [182] routine, based on a Chebyshev approximation (Barrodale

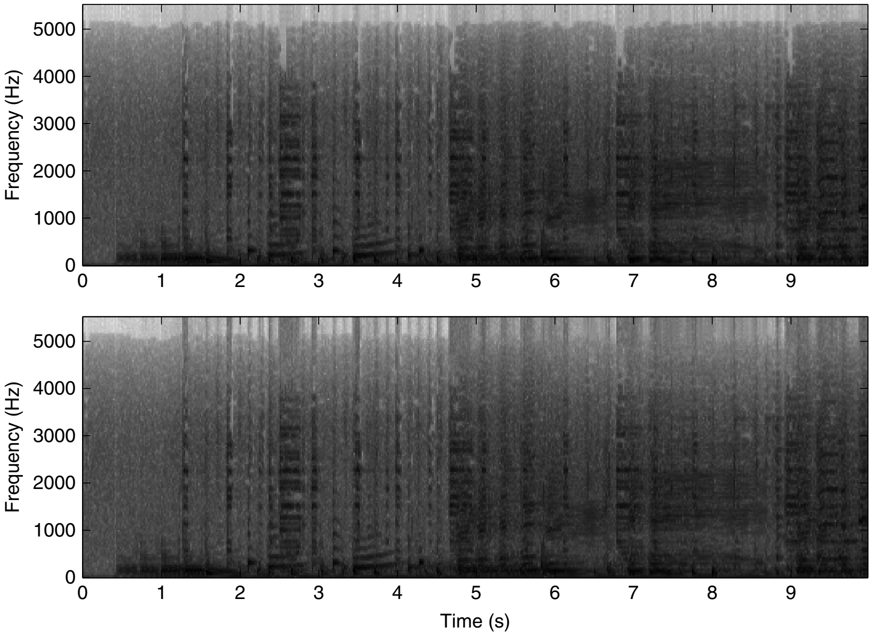

Figure 5.13 Spectrograms of a 10-s excerpt of music, the same as in Fig. 5.12 (a), but scaled down in amplitude by a factor of 10. Its BWE-compressed output is shown in part (b). Because of the low signal level of the input, the compressor operates in its (near) linear regime, and there is only a modest enhancement of high frequencies; overall the two spectrograms are very similar (much more so than the spectrograms of Fig. 5.12)

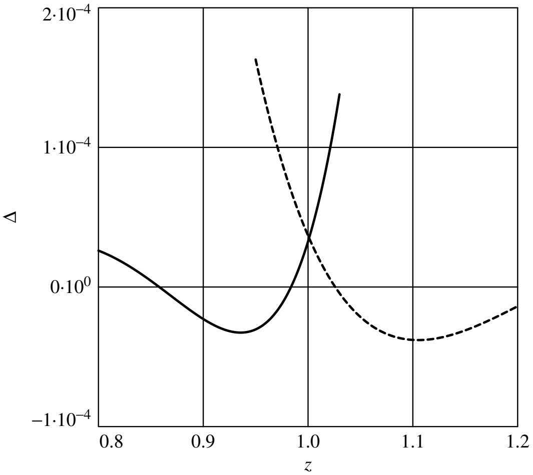

Figure 5.14 The error $\Delta = \tanh(z) - \hat{y}$; same as Fig. 5.10, but zoomed in around $z \approx 1$

and Phillips [25]), was used. Using Eqn. 5.16 and 5.17, we get the following algorithm:

```
x =~c_2*x

IF |x| <= 1 THEN

x2 = z^2
y=c_1*x*(0.9997 + x2*(-0.3289 + x2*(0.1154 --~x2*0.02465)))

ELSE

xa = |x|
y = c_1*sign(x)*(-0.1694 + xa*(1.6489 + xa*(-0.9587 + xa*(0.2713
    + xa*(-0.02786 + xa*(-0.003742 + ...
xa*(0.001199 + xa*(-0.00008518)))))))

END
```

This yields an approximation error as plotted in Fig. 5.14. To avoid a discontinuity in the transition area between both approximations ($x \approx 1$), the coefficients are chosen such that the sign of the errors for both approximations are the same, and the magnitudes are about equal.