3

Low-frequency Physical Bandwidth Extension

3.1 INTRODUCTION

Chapter 2 discussed the situation in which the audio signal has a wider low-frequency bandwidth than the loudspeaker. The opposite situation, in which the loudspeaker has a wider bandwidth than the audio signal, can also occur (although this is probably a less common situation). Consider two possibilities:

1. The audio signal was limited in bandwidth owing to the transmission (or storage) channel. In that case, BWE processing should restore, or resynthesize, the missing frequency components as closely as possible. For speech, this occurs during transmission through the telephone network, together with a limitation in the high-frequency bandwidth. Methods to address both the high- and low-frequency limitation for speech applications will be discussed in Chapter 6. For general audio applications other than telephony, this situation is not a common one, and as such it will not further be discussed.
2. The audio signal is full bandwidth (at the low-frequency end) and the additional low frequencies are desired for enhancement only, in which case the loudspeaker must have a suitably low cut-off frequency. Applications of such methods would be, for example, (home) cinema, Hi-Fi, and automotive audio. We will devote the remainder of this chapter to a discussion of this situation.

In most cases, the algorithms for these applications physically extend the low-frequency spectrum of the signal; therefore, we refer to this kind of BWE methods as low-frequency physical BWE methods. Because the loudspeaker is able to reproduce lower frequencies than those present in the audio signal, there is a possibility to lower the perceived pitch.

3.2 PERCEPTUAL CONSIDERATIONS

Assume that a signal $x(t)$ has a low-frequency cut-off of $f_{l,x}$, and the reproducing loudspeaker has a low-frequency cut-off of $f_{l,l} < f_{l,x}$. To increase the apparent bass response

Audio Bandwidth Extension: Application of Psychoacoustics, Signal Processing and Loudspeaker Design. E. Larsen and R. M. Aarts
 ISBN 0-470-85864-8

of the perceived signal, we can utilize the frequency range $f_{1,1}-f_{1,x}$ to add frequency components related to $x(t)$. These lower-frequency components can lower the pitch of the signal, which can be used for bass enhancement.

3.2.1 PITCH (SPECTRAL FINE STRUCTURE)

Say the signal $x(t)$ has a pitch of f_0 Hz, mediated by partials at $n.f_0$ Hz, $n = 1, 2, 3 \ldots$. Reproducing $x(t)$ at a reduced pitch of, say, f_0' is most practical if the harmonics already present in $x(t)$ remain harmonically related to the added low-frequency components. For example, $f_0' = f_0/2$ would be a good choice. Adding the f_0' component to $x(t)$ to create signal $x_2(t)$ would yield a complex tone with a fundamental frequency at f_0' all *even* harmonics. This situation also occurs in low-frequency psychoacoustic BWE using a rectifier as non-linear device (NLD), see Fig. 2.7. In the related discussion, it was remarked that predictions from the auditory image model (AIM [203], see Sec. 1.4.8) indicated two pitch percepts of nearly equal strength. In the notation of this section, there is a pitch at f_0', and slightly weaker, at f_0. This could indicate that either the pitch is ambiguous or that the new fundamental at f_0' is not grouped with the harmonics, leading to two signals being perceived. Note that a common amplitude or frequency modulation of the f_0' partial and the $n.f_0$ partials should facilitate grouping of all partials into a single stream ('common fate' principle, see Sec. 1.4.7).

A less ambiguous situation would occur if instead of only adding the component $f_0' = f_0/2$, components at $(2k+1)f_0' = (k+1/2)f_0$ are also added ($k = 1, 2, 3 \ldots$). In that case, the new signal $x_2(t)$ would contain a fundamental at f_0' and *all* harmonics, akin to the situation in which an integrating NLD is used in the low-frequency psychoacoustic BWE algorithm; see Fig. 2.10 for the AIM pitch prediction of this signal. Such a signal has an unambiguous and strong pitch at f_0' (even without common amplitude and/or frequency modulation). Another possibility is to add a frequency component at $f_0'' = f_0/3$, leading to a pitch that is one-third of the original; this can be extended to a frequency division by 4, 5, and so on. We will not further consider such situations though, and concentrate on the case in which the pitch is lowered by an octave.

3.2.2 TIMBRE (SPECTRAL ENVELOPE)

Assume a complex tone with harmonic amplitudes a_i at frequencies $i.f_0/2$, with i even. As before, we simplify our modelling of timbre to include only brightness[1], modelled by the spectral centroid C_S (Eqn. 1.95), which for this signal is

$$C_S = f_0 \times \sum_i i a_i^2 / \sum_i a_i^2, \tag{3.1}$$

where the sum runs over all i for which $a_i \neq 0$. After BWE processing, harmonics are added such that the new fundamental is $f_0/2$, and the new harmonic amplitudes are $a_i + gb_i$ ($i = 1, 2, 3, \ldots$); the b_i are the synthetic frequency components and are

[1] In reality, other factors influence timbre, such as the relative phase of partials and temporal envelopes. These factors are neglected to simplify the discussion, and also because they are thought to be less important than the amplitude spectrum of the harmonics, see also Sec. 1.4.6.

determined by the algorithmic details, and g is the gain factor of the b_i. The spectral centroid of the BWE signal C'_S is

$$C'_S = f_0 \times \sum_i i(a_i + gb_i)^2 / \sum_i (a_i + gb_i)^2. \tag{3.2}$$

If the goal is to achieve $C_S = C'_S$, and say the b_i are identically zero for nonzero a_i, and vice versa, then this is achieved if

$$\frac{\sum_i ib_i^2}{\sum_i b_i^2} = \frac{\sum_i ia_i^2}{\sum_i a_i^2}, \tag{3.3}$$

that is, if the individual spectral centroids of the a_i and b_i are identical, independent of the value of g. If the a_i and b_i are nonzero for common i, it is more cumbersome to derive a relationship such that $C'_S = C_S$, in part because the relative phases of the a_i and b_i need to be accounted for as well. It could also be desirable to lower the brightness of the processed signal, such that $C'_S < C_S$. Again, to achieve this it will depend on whether the a_i and b_i are nonzero for different i, or not, and their relative phase in the latter case. Because C_S depends in such a complicated manner on the harmonic structure of the signals, as well as on the processing details, we do not further consider an analysis of these effects. In general, a mildly sloping harmonics spectrum of the b_i should maintain a similar value for C_S and, therefore, a similar timbre. Of course, if only a component at $f_0/2$ is added, without any higher harmonics, C_S will be reduced by an amount that depends on the amplitude of the $f_0/2$ component.

3.2.3 LOUDNESS (AMPLITUDE)

Many of the comments made in Sec. 2.2.3 regarding loudness effects for low-frequency psychoacoustic BWE also apply for low-frequency physical BWE methods, although the effects are in the opposite 'direction'. It will again be useful to refer to the equal-loudness contours of Fig. 1.18. Firstly, we see that adding low-frequency components in the bass frequency range means that the added components will be less audible than the original low partials, due to the upward slope of the equal-loudness contours at low frequencies. A more proper way of analysing this exactly would be to use a loudness model, such as ISO532A or ISO532B, described in Sec. 1.4.4.2. Nonetheless, as acoustic energy is added above threshold, the loudness of the signal will increase. Furthermore, we assume that the loudspeaker has a more or less 'flat' response in the frequency region where the synthetic components are added, so we do need to consider the loudspeaker's response as we did in Sec. 2.2.3 and Eqns. 2.5 and 2.6.

3.3 LOW-FREQUENCY PHYSICAL BANDWIDTH EXTENSION ALGORITHMS

In the previous chapter on low-frequency psychoacoustic BWE, we encountered several different ways to create higher harmonics from a (periodic) signal. We can borrow many of these techniques to also create subharmonics, if we make appropriate changes to the algorithms. Therefore, only a limited number of possible algorithms will be presented, as others will be obvious extensions of what was presented in Chapter 2. First, we will discuss several options on using the synthetic low-frequency components.

3.3.1 SYSTEMS WITH LOW-FREQUENCY EXTENSION

Figure 3.1 presents a low-frequency physical BWE system; comparison with Fig. 2.4 will show that there is a lot of similarity with the low-frequency psychoacoustic BWE system. However, the processing details differ. The input signal $x(t)$ is filtered by FIL1 to obtain the lowest frequency band present in the signal. Although this band will vary over time, for simplicity, a constant band of about an octave can be used. The actual bandpass region will depend on the low-frequency cut-off value f_l of the loudspeaker used, and the frequency division occurring in the non-linear device (NLD). Assuming that this frequency division is a factor of two, the bandpass region of FIL1 can be set to $2f_l$–$4f_l$, for example, for $f_l = 30$ Hz, this would be 60–120 Hz. This bandpass signal is processed by the NLD, to create a fundamental an octave lower than the strongest frequency component at its input, and also the harmonics of this new fundamental. Depending on whether only the new fundamental or also some of its harmonics are desired in the output, filtering by FIL2 will select the desired frequency range. If only the new fundamental is desired, FIL2 will be bandpass between f_l–$2f_l$; if also harmonics are desired, the high-frequency limit of FIL2 should be increased, for example, to $4f_l$. Finally, a gain factor g is applied to create the harmonics signal $x_h(t)$.

At this point, there are several options. One option is indicated by the dash-dotted line in Fig. 3.1, where $x_h(t)$ is fed to a 'low-frequency effects' (LFE) loudspeaker (analogous to the sixth channel in a 5.1 surround sound system)[2]. The other option is to add $x_h(t)$ back to a delayed version of $x(t)$, where the delay of $x(t)$ should match the filtering delay of $x_h(t)$[3], yielding output signal $y(t)$. A standard crossover network can then be used to

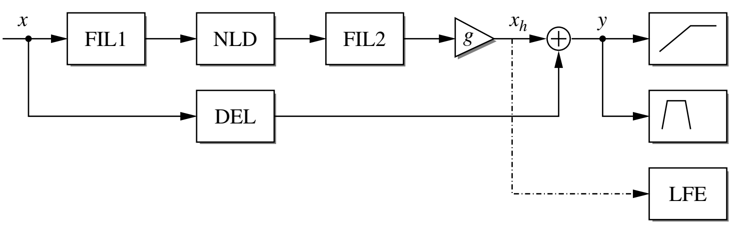

Figure 3.1 Low-frequency physical BWE system. FIL1 and FIL2 are bandpass filters, NLD is a non-linear device, g is a scaling factor. The input signal $x(t)$ is processed to yield a harmonics signal $x_h(t)$, which can be fed directly to a low-frequency effects (LFE) channel. Alternatively, it can be added back to $x(t)$ and applied to a set of loudspeakers, including a subwoofer

[2] By obvious modification to the filter bandpass regions and the NLD, extremely low frequencies that are below, say, 20 Hz can be generated. These frequencies can be used to drive *shakers* that transmit tactile vibrations.

[3] As for the low-frequency psychoacoustic BWE system, filtering for low-frequency physical BWE should preferentially be done with linear-phase filters (see Sec. 2.3.3), such that all frequency components of $x_h(t)$ will be equally delayed.

drive the available loudspeaker system, for example, a subwoofer for the frequency range 30–100 Hz, and an accompanying full-range system.

Note that we have described the low-frequency physical BWE system here as adding very low frequency components, down to 20 or 30 Hz. Proper reproduction of such frequencies requires very large and expensive loudspeaker systems. However, the frequency regions can be scaled to higher values (say to 100 Hz), such that implementation on smaller loudspeaker systems is also possible. Of course, for bass enhancement purposes, the final effects are usually more dramatic at very low frequencies.

3.3.2 NON-LINEAR DEVICE

Non-linear processing for low-frequency physical BWE applications aims to lower the frequency content of the available signal. Many of the comments made in Sec. 2.3.2 with regard to non-linear processing for low-frequency psychoacoustic BWE algorithms can be applied here as well. In particular, it is preferable to have NLDs that are level independent (homogeneous systems, see Sec. 1.1). With regard to the actual implementation of the NLDs, the various possibilities discussed in Sec. 2.3.2 can often be used here, with slight modifications such that not only higher harmonics but also the subharmonic is generated. To avoid duplicating a lot of material, we present only a brief discussion of several NLD options here.

An important difference with low-frequency psychoacoustic BWE is that for low-frequency physical BWE, we can choose to add only the halved fundamental, thereby creating a harmonic series with $f_0/2$, f_0, $2f_0$, and so on. As discussed in Sec. 3.2.1, the $f_0/2$ component may not group too well with the other partials, although common amplitude or frequency modulation of all partials will facilitate grouping (Sec. 1.4.7). On the other hand, the harmonics generated by the NLD (as discussed below) can all be added back to the main signal. In such a case, it is possible that a harmonics series of for example, $f_0/2$, f_0, $3f_0/2$, $2f_0$, and so on, is generated, in which case a strong pitch at $f_0/2$ is always perceived. The perceptual effect of low-frequency physical BWE therefore depends on how many synthetic harmonics are added back to the main signal.

3.3.2.1 Rectifier

In Sec. 2.3.2.2, a rectifier as an NLD was introduced for low-frequency psychoacoustic BWE applications. It was seen that the rectifier predominantly generates the double-frequency component of the strongest input frequency component, that its temporal response is good, and that the amount of intermodulation distortion can be quite large if more than one frequency component of comparable amplitude are contained in the input. To modify this algorithm such that it generates half the input frequency, we can simply set the output to zero for alternating periods of the input, as in Fig. 3.2, which can be done effectively by detecting zero crossings of the input signal.

3.3.2.2 Integrator

In Sec. 2.3.2.3, an integrator as an NLD was introduced. In contrast to the rectifier, all (odd and even) harmonics of the input frequency are generated. The amount of intermodulation distortion was generally low, although the temporal response was slightly worse (in terms

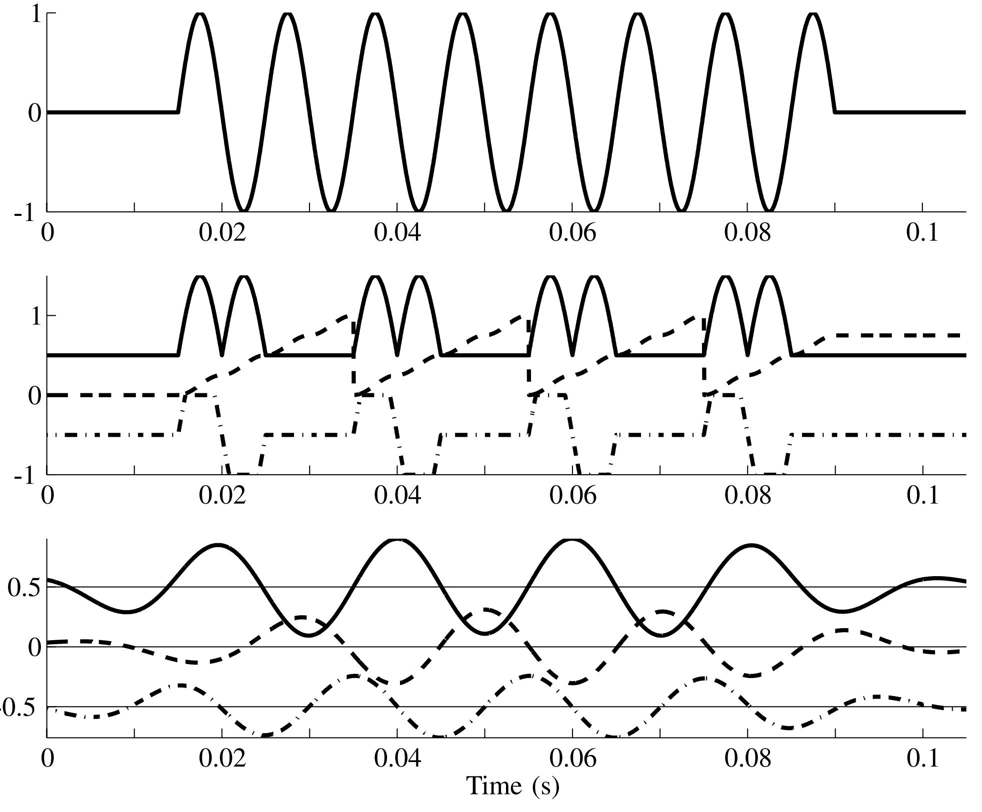

Figure 3.2 Low-frequency physical BWE processing with various NLDs. The upper panel shows a tone burst at 100 Hz. The middle panel shows the effect of rectification (solid line, offset by +0.5), integration (dashed line, no offset), and clipping (dash-dotted line, offset by −0.5). Note that period doubling occurs, because the signal is zeroed on every other period for the rectifier and clipper, while for the integrator the reset occurs after two input periods. The lower panel shows the effect of linear-phase filtering around 50 Hz, such that only the fundamental frequency is retained

of distortion of temporal envelope). To use an integrator for halving the frequency of a signal, the only necessary modification is to reset the output of the integrator to zero after two input periods have occurred. The input signal periods can again be detected by observing zero crossings. The integrated output for a pure-tone input is plotted as a dashed line in the middle panel of Fig. 3.2. Linear-phase filtering this signal around 50 Hz leads to a signal as shown by a dashed line in the lower panel of the same figure.

3.3.2.3 Clipper

In Sec. 2.3.2.4, a clipper as an NLD was introduced. This device generates odd harmonics of the input frequency. The clipper was shown to be very robust against intermodulation distortion, and to have a good temporal response. Again, a slight modification will allow the clipper to be used for low-frequency physical BWE purposes. As with the rectifier discussed previously, alternating periods of the input signal are clipped (at a clipping level of half-maximum amplitude, in this case ± 0.5), the other periods being zeroed. The resulting signal is shown as the dash-dotted line in the middle panel of Fig. 3.2. Linear-phase filtering around 50 Hz leads to the signal shown by the dash-dotted line in the lower panel of the same figure.

For low-frequency psychoacoustic BWE algorithms, it was discussed at length how to enhance performance by making the clipping level adaptive with respect to the level of the input signal. For low-frequency physical BWE, this might also have some advantage, and a similar strategy could be employed, although this has not been validated by actual listening tests.

3.3.2.4 Low-frequency Physical Bandwidth Extension with Frequency Tracking

An alternative method of generating subharmonics is by using a frequency tracker, as discussed in Sec. 2.4, refer to Fig. 2.25. This algorithm generates any desired harmonics spectrum, with a fundamental that is based on the strongest frequency component contained in the input spectrum (the tracked frequency). Advantages are that there is no intermodulation distortion and that harmonics are only generated if the input is periodic. Also, the frequency tracker is implemented in a computationally very efficient manner. The disadvantage is that the frequency tracker needs finite time to converge to the actual signal frequency, and that errors may occur if multiple input frequencies or additive noise is present. However, the particular method of frequency tracking was shown to adapt quickly and that some corrections for the tracked frequency are possible if a few statistics of any additive noise are known or can be estimated.

Assuming that the input contains frequency ω_0, and is correctly estimated by the frequency tracker, the harmonics generator (HG in Fig. 2.25) will generate a signal $x_h(t)$ as

$$x_h(t) = \sum_{k=1}^{N} A_k \sin(k\frac{\omega_0}{2}t), \tag{3.4}$$

If one desired to merely add the halved fundamental, then $N = 1$; otherwise, $N > 1$ and a desired harmonics spectrum can be generated. Because the synthetic frequency components are explicitly created, there is no intermodulation distortion.

3.3.2.5 Inclusion of Higher Harmonics

From the lower panel of Fig. 3.2, it is quite obvious that the resulting signal (representing the signal $x_h(t)$ in Fig. 3.1) does not depend a great deal on the particular choice of NLD. This is because in all cases only the halved fundamental (in this case 50 Hz) was extracted by FIL2 of Fig. 3.2. The various NLDs (rectifier, integrator, clipper) differ most in the amplitude and phase spectrum of the generated harmonics, thus if some of the harmonics are retained by FIL2, then the output signals would differ more. To illustrate this, Fig. 3.3 shows signals $x_h(t)$ as they would be obtained if FIL2 were a linear-phase filter with a bandpass of 40–160 Hz (all the other processing steps are identical to those in the first

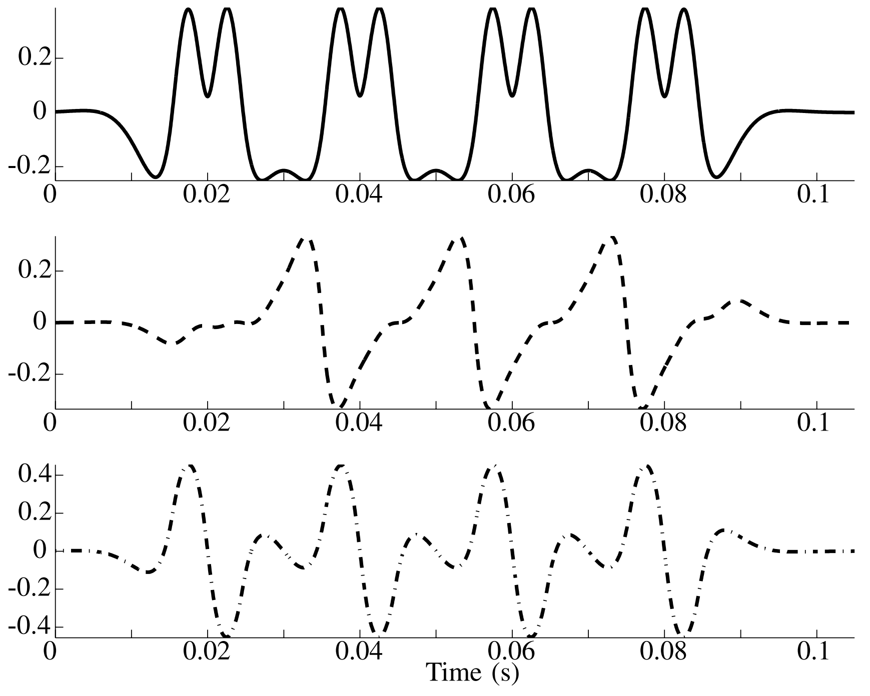

Figure 3.3 Low-frequency physical BWE processing with various NLDs; input signal is assumed to be a gated 100-Hz tone as in the upper panel of Fig. 3.2. Here, the output signal $x_h(t)$ (refer to Fig. 3.1) is shown for three different NLDs: rectifier (upper panel), integrator (middle panel), and clipper (lower panel). Owing to filtering of the spectrum generated by the NLD, only harmonics at 50, 100, and 150 Hz are appreciably present in the output signals. Note that the different spectra generated by the NLDs causes the observed differences in waveforms

example). Thus, the partials at 50, 100, and 150 Hz would be passed completely, while higher partials would be strongly attenuated by FIL2. The interference of these three partials creates the differences visible in Fig. 3.3, where signals by rectifier, integrator, and clipper are displayed as solid, dashed, and dash-dotted lines respectively.

3.3.3 FILTERING

Using the same kind of analysis as was done in Sec. 2.3.3.3 for low-frequency psychoacoustic BWE algorithms, it can be shown that it is preferable to use a linear-phase implementation for filters FIL1 and FIL2, for two reasons:

- The final output signal of the low-frequency physical BWE algorithm will be obtained by scaling $x_h(t)$ and adding it back to $x(t)$ (either electrically or acoustically, see Fig. 3.1), so common frequency components will interfere either constructively or destructively. In the example of Figs. 3.2 and 3.3, there is a synthetic 100-Hz component in $x_h(t)$ as generated by the rectifier and the integrator (but not the clipper, which generates only odd harmonics of 50 Hz), which adds to the original 100-Hz fundamental present in $x(t)$. Figure 2.21 shows examples in which synthetic and original frequency bands interfere to create an irregular amplitude spectrum in case of non-linear-phase filters. For low-frequency physical BWE algorithms, the situation would be similar, if beside the halved fundamental, higher harmonics are also added, as these would overlap in frequency range with the original signal. If only the halved fundamental is added, there is no frequency overlap between $x(t)$ and $x_h(t)$, and there should be no objection to use a non-linear-phase filter from this particular point of view.

- Because of the filtering by FIL1 and FIL2, signal $x_h(t)$ will be delayed with respect to $x(t)$, by an amount that depends on the filter orders (higher filter order giving larger delays). In Sec. 1.4.7 on auditory scene analysis, it was explained that an onset delay between frequency components can lead to segregation, that is, $x(t)$ and $x_h(t)$ being separately perceived as two different streams. This was analysed for low-frequency psychoacoustic BWE in detail in Sec. 2.3.3.3, and typical values for total group delay *variation over frequency* of components of $x_h(t)$ was shown to be on the order of 10 ms (Fig. 2.22), which combines the effect of both filters. Such delays can be assumed to be detectable in principle (Zera and Green [304]), and there is some circumstantial evidence from informal listening tests that such can indeed lead to segregation. To avoid such problems, linear-phase filters will not lead to group delay variations, and if $x(t)$ is delayed (using a delay line) by the same amount as $x_h(t)$ (the required amount of delay can be easily computed if the filters FIL1 and FIL2 are known), both signals can be added exactly in phase. Note that in this case it does not matter whether $x(t)$ and $x_h(t)$ have overlapping frequency components or not.

The standard way to implement linear-phase filters with FIR structures presents a problem for low-frequency physical BWE in that the bandpass region is usually a very small fraction of the sample rate. Also, the filter's low cut-off frequency can be very small, even smaller than for low-frequency psychoacoustic BWE algorithms. An FIR filter would have to be of very high order (many taps) to implement such a specification, leading to a high computational burden. A more efficient way is to use an IIR filter that is designed to have an amplitude spectrum that is the square root of the desired specification. The filter can be applied once in forward time, after which the filtered signal is time reversed and filtered using the same filter. This output is then time reversed again. The result is zero phase shift (as phase changes of the first filter are canceled by the second time-reversed filter) and an attenuation that is the square of the IIR filter's amplitude spectrum. This was also discussed in Sec. 2.3.3.3 and is discussed elaborately in, for example, Powell and Chau [213].

3.3.4 GAIN OF HARMONICS SIGNAL

Again, the situation is very analogous to the case of low-frequency psychoacoustic BWE applications. Gain (or scaling) $g(t)$ of $x_h(t)$ prior to addition to $x(t)$ to form the output $y(t)$ (Fig. 3.1) could be fixed, frequency adaptive, and/or output-level adaptive. For a simple implementation, $g(t)$ could simply be a constant value, such that the loudness balance of the synthetic harmonics signal and the original signal is subjectively appropriate, and does not lead to distortion at high signal levels.

A slightly more subtle scheme could be adopted analogous to Gan *et al.* [83], which takes into account the equal-loudness level contours (Sec. 2.3.4.2 and Fig. 1.18) to expand the envelope of the generated harmonics signal. While Gan *et al.* designed this scheme for low-frequency psychoacoustic BWE applications, the same procedure could be used for a low-frequency physical BWE algorithm. In a simple implementation of their scheme, the envelope of the harmonics signal $x_h(t)$ would be expanded by a factor of about 1.10 (intended for fundamental frequencies in the range 40–100 Hz; the expansion ratio varies with fundamental frequency, the quoted value is an average). A more sophisticated approach implements a different expansion ratio for each harmonic, reflecting the change in loudness growth over frequency. A scheme like this, in which each harmonic is scaled separately, is attractive if harmonics are generated individually, such as in Gan *et al.*ś modulation technique, or the method described in Sec. 2.4 and 3.3.2.4 (using a frequency tracker). In the latter case, expansion ratios can be determined even more accurately, as the fundamental frequency is known (estimated), which is not implemented in Gan *et al.*ś original algorithm.

An output-level-adaptive gain could be implemented entirely analogous to that described in Sec. 2.3.4.3, and illustrated in Fig. 2.24. This scheme employs a feedback loop from the output $y(t)$ to control $g(t)$. In normal circumstances, $g(t)$ has a fixed value, but if the level of $y(t)$ exceeds a predetermined value, $g(t)$ should be decreased very quickly to prevent distortion occurring in the loudspeaker. While $y(t)$ is below the threshold level, $g(t)$ can be slowly increased back to its nominal value. This increase should occur slowly to prevent envelope distortion of $y(t)$ (i.e. on a time scale of a few seconds). An additional advantage of such an automatic gain control (AGC) scheme is that low-level bass sounds are maximally amplified, while high-level bass sounds are attenuated. This is a good match to the audibility characteristics of low-frequency sounds, which have very low loudness at low-to-intermediate sound pressure level, but which increase in loudness very rapidly (more so than intermediate frequency sounds) with increasing sound pressure level.

3.4 LOW-FREQUENCY PHYSICAL BANDWIDTH EXTENSION COMBINED WITH LOW-FREQUENCY PSYCHOACOUSTIC BANDWIDTH EXTENSION

We have so far presented several options for physical extension of the low-frequency spectrum of an audio signal. In Chapter 2, we have presented the same for a psychoacoustic bandwidth extension. These two concepts can be combined such that the bass pitch in audio signal is lowered (usually by an octave), but in such a way that very low frequency components are not radiated nor ever physically present. This would permit application to smaller loudspeaker systems.

Beside the obvious cascade of a low-frequency physical BWE system followed by a low-frequency psychoacoustic BWE system, this concept can be more efficiently implemented by the ‘standard’ low-frequency physical BWE approach of Fig. 3.1. The only modification is that FIL2 should have a higher low cut-off frequency, such that the halved fundamental is not actually present in $x_h(t)$. For example, a complex tone with a 70-Hz fundamental would yield a synthetic signal after non-linear processing that has a fundamental at 35 Hz, including harmonics (the spectrum of which depends on the particular choice of the NLD). When using a small loudspeaker, with a low-frequency resonance at 100 Hz, FIL2 could be implemented as a bandpass filter between 100–200 Hz. The result is that the pitch of the bass tone has been lowered to 35 Hz, which is effected by frequency components >100 Hz. In other words, a signal without very low pitched tones has been modified such that it is perceived as having very low pitched tones, using a loudspeaker that cannot reproduce very low frequencies.