2

Psychoacoustic Bandwidth Extension for Low Frequencies

2.1 INTRODUCTION

All loudspeakers have a limited frequency range in which they can radiate sound energy at a more or less uniform level. The radiated sound pressure level can be expressed as a function of frequency through the loudspeaker's magnitude response (usually specified for an on-axis measurement). Hi-Fi enthusiasts know that a flat frequency response is very desirable, and it has been shown that the degree of 'flatness' correlates well with perceived quality (Gabrielsson and Lindström [82], Toole [274, 275]). Another desirable feature is that this frequency response is maintained for off-axis radiation, but we will not be concerned with loudspeaker directivity. Whether the loudspeaker's response is flat or not, at low and high frequencies the efficiency always decreases, leading to a *low* (f_l) and *high cut-off frequency* (f_h), usually defined as those frequencies in which the response falls 3 dB below the response at some intermediate reference frequency. Focusing on the low-frequency cut-off point, we can easily derive how the loudspeaker parameters influence its value. By rewriting the expressions derived in Sec. 1.3, we find that the efficiency η in the normal operating range, and f_l, are given by

$$\eta \sim \left(\frac{S}{m}\right)^2, \tag{2.1}$$

$$f_\mathrm{l} = \frac{1}{2\pi}\sqrt{\frac{k_t}{m}}, \tag{2.2}$$

f_l being determined by the resonance frequency (usually denoted as f_0) of the mass-spring system that the loudspeaker is. A high efficiency η necessitates a large cone area S; a low f_l requires a low compliance k_t ('total' compliance: combined suspension and cabinet influence) and/or a large mass m. A low total compliance would necessitate a large cabinet volume; but a large mass greatly decreases the efficiency. For example, to lower the cut-off frequency of an octave by quadrupling the mass, the efficiency would decrease by a factor of 16 (12 dB). Such a measure is not in line with good loudspeaker design (high

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

voltage sensitivity and high power efficiency). For small loudspeakers, in particular, the situation is troublesome: a small cone area, a small mass, and a high compliance lead to high values for the low-frequency cut-off point, and low efficiency. We have already found that lowering f_1 by increasing m is not a viable option; also, lowering k_t is not feasible because it would necessitate a large cabinet volume (which contradicts the loudspeaker being small). The fundamental problem is that good low-frequency sound reproduction requires a large volume velocity, which is very hard to achieve for a small loudspeaker. Beyond this problem of physical origin, the perception of low-frequency sound (lower than, say, 100 Hz) is markedly worse than for intermediate frequencies (see Fig. 1.18). This means that to reproduce a low-frequency tone at equal sensation level relative to a higher frequency tone, the SPL will have to be higher. Some typical loudspeaker responses are shown in Fig. 4.13.

Nevertheless, in many applications, small loudspeakers are unavoidable, because of size and/or cost constraints. In fact, loudspeakers that are large enough to reproduce the lowest audible frequencies (around 20 Hz) at a sufficient level are huge in size and very expensive. Even a more modest goal of good reproduction at 50 Hz is difficult to achieve within the constraints usually encountered in consumer electronics, such as (flat) TV and laptop computers, and also (portable) audio and (in-ear) headphones. Another challenging case is in telephony, in which very small loudspeakers are employed.

The need for higher acoustic output has always existed, especially at low frequencies, ever since the invention of the electrodynamic loudspeaker. Improvements in loudspeaker design have yielded better low-frequency characteristics, the most popular option being vented designs. The vent introduces an additional resonance below the loudspeaker – cabinet resonance, thus extending the low-frequency response. The drawback is that the response falls off twice as fast below the new cut-off frequency, and the temporal behaviour is degraded. Even though this 'bass-reflex' design has a more extended low-frequency response than a conventional loudspeaker and cabinet, Eqns. 2.1–2.2 still hold. As mentioned previously, the fundamental problem is the limited volume velocity that is achievable with a small loudspeaker, and the physical limit cannot be overcome with purely physical modifications of the design. A partial solution to the low-frequency problem has come from BWE and the psychology of hearing. The material presented in this chapter will explain how BWE can be used to improve the bass response of small loudspeakers, and basic algorithms are presented. First, however, we discuss the traditional option of low-frequency emphasis by linear amplification of the bass portion ('bass boosting').

The loudspeaker response can theoretically be inverted using a preceding filter with the inverse of this response, as in Fig. 2.1. In practice, the limiting factors are finite cone excursion and finite power-handling capacity of the loudspeaker. Therefore, this method can only enhance frequencies at or slightly below f_1 (BWE methods can enhance reproduction several octaves below f_1). At high output levels, distortion or even damage and ultimately destruction of the loudspeaker may occur. Also, this solution is very energy inefficient, because of the loudspeaker's intrinsic low efficiency at low frequencies (important for portable devices such as portable audio, cd players, or PDAs). The advantages of this approach are its simplicity and linearity.

Another solution was proposed by Long and Wickersham [164] ('ELF' system). The design specifically drives the loudspeaker below its resonance frequency, by using two

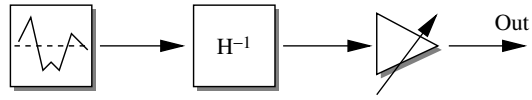

Figure 2.1 A simple circuit for 'inverting' a loudspeaker response. The signal is filtered by H^{-1}, the 'inverse' of the loudspeaker response, scaled and applied to the loudspeaker. Usually the filtering does not use the real inverse of the loudspeaker response, but simply a low-pass filter to boost low-frequency sounds. The subsequent scaling may be manually adjustable or signal dependent

integrators preceding the loudspeaker terminals. The integrators invert the high-pass characteristic of the loudspeaker, and the method is said to work better than 'traditional' amplification. But again, cone excursion and inefficiency are concerns for such an approach.

2.2 PSYCHOACOUSTIC EFFECTS FOR LOW-FREQUENCY ENHANCEMENT OF SMALL LOUDSPEAKER REPRODUCTION

Given the fact that on physical grounds it is impossible to have a good low-frequency loudspeaker response (with small loudspeakers), it is pertinent to ask whether other options are available. One option is to use BWE, with the 'extension' taking part in the auditory system, instead of extending the actual physical bandwidth of the signal. In fact, we sometimes have to reduce the bandwidth of the signal to prevent very low frequencies from entering the loudspeaker, as these cannot be reproduced anyway. If we modify the low-frequency part of the signal in such a way that the auditory system 'fills in' the part that the loudspeaker cannot reproduce, then we have achieved *psychoacoustic* BWE.

2.2.1 PITCH (HARMONIC STRUCTURE)

There are a number of psychoacoustic effects that can possibly be used for this kind of BWE, shown in Fig. 2.2. As before, f_l represents the low-frequency cut-off of the loudspeaker. These options are:

- *Frequency doubling*: By a frequency doubling, we can shift components at frequency f, for which $f_l/2 \leq f \leq f_l$, to frequency f', which is $f_l \leq f' \leq 2f_l$, as in panel (ii) of Fig. 2.2. Because frequency components are now above f_l, they can be reproduced at a higher level, and since the audibility curve slopes downward at low frequencies, the ear is more sensitive in this frequency range as well (Fig. 1.18). It may be expected that the resulting signal will have an increased loudness in the low-frequency range. Frequency doubling is attractive because of its extreme simplicity: a full-wave or half-wave rectification suffices, which is trivial to implement in digital signal processing, or in analog components. The drawback is that the waveform is seriously distorted, and also, the pitch has changed.

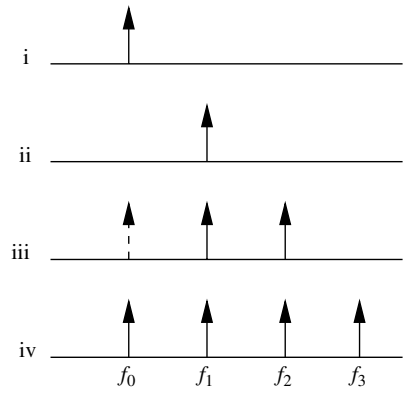

Figure 2.2 Panel (i) shows the frequency representation of a pure-tone signal at f_0 Hz. If $f_0 < f_l$, then the signal can be substituted by the following. Panel (ii): the double frequency at $2f_0$; the pitch has changed, but the loudspeaker will reproduce the signal more efficiently at $2f_0$. Panel (iii): components at $2f_0$ and $3f_0$, which will produce the cubic combination tone, and the difference tone at high levels also , in the cochlea, at f_0. Panel (iv): components at $2f_0$, $3f_0$, $4f_0$ and so on, which will produce a residue pitch at f_0. In panels (iii) and (iv), the dashed arrows are not physically radiated by the loudspeaker, but a pitch corresponding to that frequency is perceived

- *Combination tones*: As was explained in Sec. 1.4.2, non-linearities of the cochlear response generate combination tones (CT) when presented with two-tone stimuli. If pure tones with frequencies f_1 and f_2 enter the cochlea, the generated CT frequencies $f(n)$ correspond to those given in Eqn. 1.87. To elicit a low-frequency pitch at f_0, we need $f_0 = f(1) = 2f_1 - f_2$; this corresponds to the cubic CT, which is highest in level of all CTs. Probably the best option is to have $f_1 = 2f_0$ and $f_2 = 3f_0$. Although the ratio $f_1/f_2 = 1.5$ is unfavourable (the level of the cubic CT at this ratio is very low), at least these two components are harmonically related to f_0. Some advantage might be obtained because the difference tone (DT) (also described in Sec. 1.4.2) $f_2 - f_1$ now coincides with the cubic CT, which might increase the loudness of the f_0 component. The frequencies are shown in panel (iii) of Fig. 2.2. Also, choosing f_1 and f_2 in the manner described above will aid in the perception of virtual pitch, to be described next.
- *Virtual pitch*: Perhaps the most attractive option is to make use of the 'missing fundamental' effect: a special case of residue pitch, also known as virtual pitch. In Sec. 1.4.5, it is shown how the auditory system creates a pitch percept at f_0 if presented with a harmonic series, that is, a tone complex of several frequency components, which have a common fundamental frequency f_0. For this, it is not necessary that the f_0 component is actually present (nor the second, third, etc.).
  For low-frequency psychoacoustic BWE applications, we can substitute an $f < f_l$ by a series kf, $k \geq 2$, to evoke the residue pitch of f, while the loudspeaker does not radiate energy at frequency f. There are many non-linear operations that can be used to

generate a harmonics series that will serve this purpose, as will be presented later in this chapter. Note that at high sound pressure levels, residue pitch and distortion products may occur simultaneously. The spectrum of a harmonic complex with fundamental at f_0 is shown in panel (iv) of Fig. 2.2.

Thus, there are three options to increase (apparent) bass reproduction below a loudspeaker's cut-off frequency. Note that the original low-frequency components below f_l can be either removed by appropriate filtering or retained.

2.2.2 TIMBRE (SPECTRAL ENVELOPE)

In addition to a correct pitch, the extended frequency components (harmonics of f_0) should have a timbre that is close to what would be perceived over an ideal loudspeaker. In Sec. 1.4.6, it was discussed that timbre depends on magnitude spectrum more than on phase spectrum, and the spectral centroid C_S was introduced as an objective metric to represent the subjective quality of 'brightness' of a sound.

For an analysis of low-frequency psychoacoustic BWE, we apply the simple wideband formulation as in Eqn. 1.95 to compute the spectral centroid $C_{S,0}$ for an input pure tone of frequency f_0 and amplitude a_0 and the spectral centroid $C_{S,1}$ for a synthetic signal generated by BWE, that consists, say, of fundamental and harmonics 1–5 (as discussed in Sec. 2.2.1), with amplitudes $a_0 \ldots a_5$:

$$C_{S,0} = (f_0 a_0^2)/a_0^2 = f_0, \tag{2.3}$$

$$C_{S,1} = \left(\sum_{i=0}^{5} f_i a_i^2\right) / \sum_{i=0}^{5} a_i^2 = f_0 \times \left(\sum_{i=0}^{5} i a_i^2\right) / \sum_{i=0} a_i^2 = f_0 \times \alpha, \quad 1 \leq \alpha \leq 6. \tag{2.4}$$

The conclusion is that $C_{S,0} \neq C_{S,1} \forall \alpha > 1$, that is, the brightness of the input signal will never be equal to the brightness of the output signal, unless all harmonic amplitudes are zero. One can easily verify this by listening to the two above signals and concluding that the pitch of both will always be equal, but the timbre will never be. We can deduce from Eqn. 2.4 that the timbre of the harmonics signal will be closest to that of the pure-tone input if the low-order harmonics are relatively larger in amplitude. In the limit that $a_0 \ll a_i$ ($i = 1 \ldots 5$), the two timbres will be indistinguishable (neglecting other factors, which may influence timbre, such as phase spectrum), but in that case, there is no bandwidth extension taking place. In practice, this means that there has to be a compromise between a large BWE effect (weakly decaying harmonics spectrum) and a good timbre match (strongly decaying harmonics spectrum). How this compromise is achieved is discussed later.

2.2.3 LOUDNESS (AMPLITUDE) AND TONE DURATION

After pitch and timbre, the last perceptual variable to control is loudness, measured in phones (Sec. 1.4.4). To keep matters simple, we will only consider the influence of

intensity and frequency on loudness here. For this, we refer to the equal-loudness contours by Fletcher and Munson [73] (Fig. 1.18). For signals below about 500 Hz, we can state that equal-intensity signals will sound louder if the frequency content is higher. For low-frequency psychoacoustic BWE applications, this is a favourable circumstance, as the algorithms typically replace low-frequency signals with higher harmonics. The equal-loudness contours were measured for steady-state pure tones, so to assess the loudness of time-varying complex signals (such as music or speech), it would be better to use more sophisticated models, such as those by Stevens [257], Paulus and Zwicker [205], Zwicker [310], or Glasberg and Moore [90] (discussed in Sec. 1.4.4.2). Using any of these methods, we could compute the loudness of a pure-tone signal s_0 of frequency f_0 as perceived through a 'perfect' loudspeaker, as $L(s_0)$. We must also take into account the other frequencies' components present in the signal, which we represent by signal s_m. The loudness of the entire signal, perceived through a perfect loudspeaker, would then be $l_0 = L(s_0 + s_m)$. BWE processing is applied to s_0, which creates a signal s_0', and the loudness of the total signal would be $l_0' = L(gs_0' + s_m)$, where s_0 is scaled by factor g. However, this signal is reproduced over a non-ideal loudspeaker that has an average response of h_0' in the frequency region of s_0' and a response of 1 in the frequency region of s_m. Then the perceived loudness will be $l_\mathrm{h}' = L(gh_0's_0' + s_m)$. So, we should have

$$L(s_0 + s_m) = L(gh_0's_0' + s_m). \tag{2.5}$$

The increase in loudness ΔL that the processed signal would have relative to the unprocessed signal, played over the same loudspeaker, would be

$$\Delta L = L(gh_0's_0' + s_m) - L(h_0s_0 + h_ms_m) \approx L(gh_0's_0' + s_m) - L(h_ms_m), \quad \text{if } h_0s_0 \ll h_ms_m; \tag{2.6}$$

h_0 is the loudspeaker response in the frequency region of s_0, which, being below the loudspeaker's resonance frequency, is assumed to be negligible. Because L is a complicated non-linear function, we cannot give a closed-form expression for g (Eqn. 2.5) in terms of the other variables, nor for ΔL (Eqn. 2.6). This analysis does show us that beside a model for L (loudness perception of complex tones), we also need to know the characteristics of the loudspeaker, at least in the bass frequency range. A few attempts have been made to create appropriate loudness models and algorithms for low-frequency psychoacoustic BWE applications, which will be discussed in Sec. 2.3.4.

The frequency dependence of both the equal-loudness contours and the loudspeaker response can also influence the perceived duration of tones. This will be illustrated with respect to Fig. 2.3. A 50-Hz decaying tone is shown in part (a), as produced by, for example, a bass drum. The amplitude of the signal is determined by the loudspeaker's response at 50 Hz. The dashed line indicates the minimum audible field at 50 Hz. The perceived duration of the tone is indicated by the horizontal line, and equals T_1 s. Say 50 Hz lies below the cut-off frequency of the loudspeaker and a low-frequency psychoacoustic BWE algorithm is applied to enhance the bass response, resulting in the signal shown in the Fig. 2.3 (b). The BWE algorithm has created harmonics (say, 100 and 150 Hz) at which frequencies the loudspeaker will be more efficient. This is shown by the increased

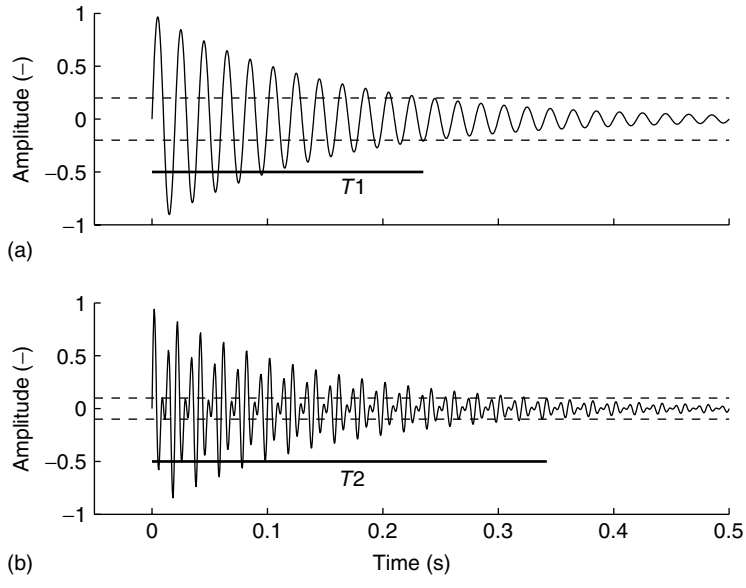

Figure 2.3 (a) Shows a decaying signal at 50 Hz. The loudspeaker characteristics and the minimum audible field at this frequency cause signals below the dashed line to be inaudible; thus, the tone has a duration T_1 of about 0.23 s. (b) Shows the signal with the 50-Hz component replaced by components at 100 and 150 Hz. The higher efficiency of the loudspeaker and the lower value of the minimum audible field at these frequencies cause the signal level where the tone becomes inaudible to be lower; thus, the tone has a duration of T_2 of about 0.34 s now

amplitude of the signal. As the equal-loudness contours slope downward at low frequencies, the minimum audible field has decreased[1]. Together, these two effects increase the perceived duration of the tone from T_1 s to T_2 s, in the given example from 0.23 to 0.34 s (of course, the loudness of the tone also increases greatly). This increase in tone duration can, for some repertoire, sound artificial. In fact, it merely shows that the BWE extension is doing its job well, as the same repertoire reproduced on a high-quality subwoofer has the same long duration bass notes. The artificial aspect of the perception on a small loudspeaker system with BWE is probably due to the fact that one does not expect good bass reproduction for such systems.

Careful inspection of the equal-loudness contours will reveal that the spacing of the contours is not constant. Rather, the contours are more 'compressed' for very low frequencies than for higher frequencies, if we restrict our attention to the bass frequency range. The consequence is that if we vary the level of two pure tones of unequal frequency by the same amount, then the loudness variation of the two will be unequal. The lower

[1] In most cases, the lowest audible intensity is not determined by the minimum audible field, but rather by masking effects of ambient noise. Therefore, the increased sensitivity of the ear at higher frequencies would not usually influence tone duration.

frequency tone will appear to have the greater loudness variation. For low-frequency psychoacoustic BWE, this could imply that if very low-frequency components are replaced by higher frequency components, loudness variations decrease. Gan *et al.* [83] have taken this into account in their BWE algorithm, as will be discussed in Sec. 2.3.4.

2.3 LOW-FREQUENCY PSYCHOACOUSTIC BANDWIDTH EXTENSION ALGORITHMS

2.3.1 OVERVIEW

We will discuss low-frequency psychoacoustic BWE algorithms using the general structure shown in Fig. 2.4; compare this structure to the general BWE framework introduced in Chapter I.3 as Fig. I.2. The implementation is in the time domain, which has the benefit of computational efficiency. Frequency-domain algorithms would be possible, but suffer the drawback that it would be difficult to achieve the desired frequency resolution while at the same time keeping the analysis window sufficiently short to satisfy stationarity of the input signal.

The essential element of the structure in Fig. 2.4 is the *non-linear device* (NLD), which converts frequencies below the loudspeaker cut-off frequency f_l to frequencies above f_l. The NLD will be chosen such that the pitch of the input signal is preserved, which will be the case if the output frequencies are harmonically related to the input frequencies (Sec. 2.2.1). In Sec. 2.3.2, we will present several options for the NLD. The filters FIL1 and FIL2, placed respectively before and after the NLD, serve two functions. FIL1 ensures that only frequencies below f_l enter the NLD; it is assumed that higher frequencies are reproduced properly by the loudspeaker, and therefore should not be modified. The filter characteristics therefore depend mainly on f_l. FIL2 does the spectral envelope shaping of the complex signal produced by the NLD. Its characteristics do not depend heavily on the loudspeaker, but on the implementation of the NLD, in particular, the relative

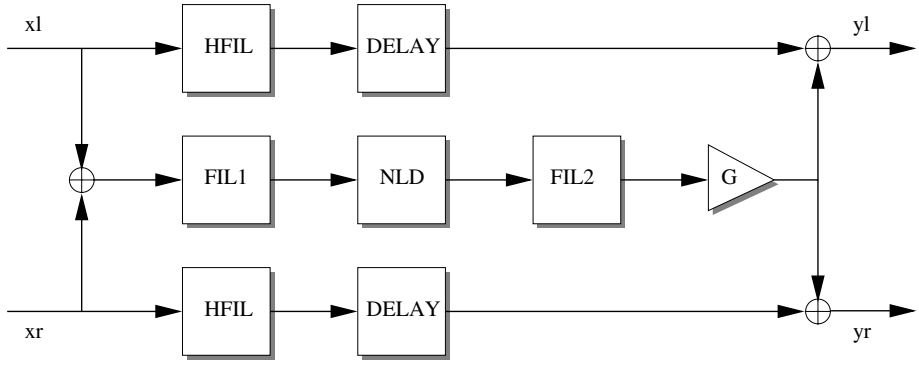

Figure 2.4 Overview of low-frequency psychoacoustic BWE in a time-domain algorithm. The harmonics are generated by the non-linear device (NLD), with appropriate filtering by FIL1 and FIL2. After scaling, the extended signal is added back to the input signals, which is delayed and possibly high-pass filtered (HFIL)

amplitudes of generated harmonics. FIL2 attempts to control the timbre of the synthetic bass signal. In Sec. 2.3.3, we will go into more detail regarding the characteristics of FIL1 and FIL2. Finally, the harmonics signal must be scaled such that an appropriate loudness is achieved, after which it is added back to the input signal. The gain may be fixed or, in more complex algorithms, adaptively determined by characteristics of the input and output signals, as will be discussed in Sec. 2.3.4. The higher frequencies of the input signals are usually passed straight through to the output, although a high-pass filter (with cut-off frequency of approximately f_1) may be applied to eliminate very low-frequency components. The rationale for this is that these very low components are not audible anyway (due to the poor loudspeaker response at those frequencies and the high audibility threshold), but do contribute to cone excursion of the loudspeaker. By removing these components, the cone excursion is decreased, which can be beneficial for the quality of the reproduced signal.

The structure of Fig. 2.4 shows that for a stereo input signal, processing is done on the summed input. This is because low-frequency content is usually identical in both channels. Also, localization is quite poor at very low frequencies, wherefore the actual distribution in left and right output channels is irrelevant.

2.3.2 NON-LINEAR DEVICE

The essential element of low-frequency psychoacoustic BWE algorithms is the non-linear device (NLD), but non-linearity is a very general property, and there are many kinds of non-linear functions. For any type of BWE, we usually require amplitude linearity, such that the relation between input and output signals is independent of level. Thus, in Vaidyanathan's [278] terminology (see Sec. 1.1), we prefer NLDs to be homogeneous systems. In this section, we review several NLD implementations that can be useful for low-frequency psychoacoustic BWE.

In discussing the various NLDs and their characteristics, we shall make reference to the auditory image model (Patterson *et al.* [203]) discussed in Sec. 1.4.8, which we use to assess the pitch of complex tones.

2.3.2.1 Multiplier

Spectral characteristics Figure 2.5 shows an NLD whereby the input signal is repeatedly multiplied with itself, producing a harmonic series. Although this system is not homogeneous, it has the advantage that one can control at the outset the number of harmonics created, and their relative amplitudes. Because the output spectrum is under direct control, a shaping filter (FIL2 in Fig. 2.4) is not necessary. The problem that this multiplying NLD is non-linear in amplitude can be solved by using an automatic gain control (AGC) before the NLD, which will scale the signal level to a reference value. After the NLD, the inverse gain will be applied to restore the signal to its original level. In this way, the whole system has effectively become homogeneous.

We begin the analysis by assuming a pure-tone input signal x of frequency f_0 at the reference level (defined to be 1). Thus,

$$x(t) = \sin(2\pi f_0 t). \tag{2.7}$$

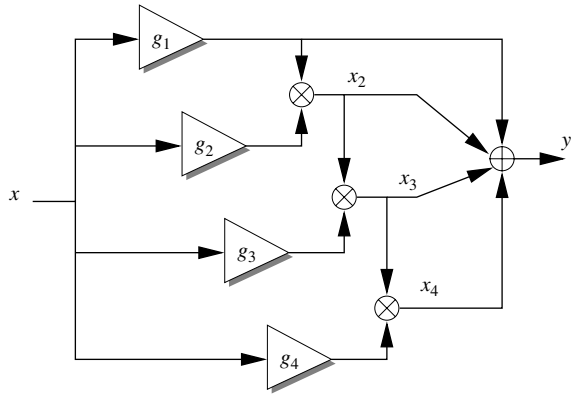

Figure 2.5 Harmonics generation by multiplication. The input signal x is repeatedly multiplied by itself to generate the harmonics. For a pure-tone input, the weights g_i can be chosen such that the output signal has a prespecified spectrum, according to Eqn. 2.11. The example shown here generates four harmonics, but more or less numbers is also possible

After multiplying a scaled x with a scaled replica, we get x_2, as

$$x_2(t) = \frac{g_1 g_2}{2}[1 - \cos(2 \times 2\pi f_0 t)]. \tag{2.8}$$

The frequency doubling is apparent; g_1, g_2 are the scaling factors. If the heterodyning option of Sec. 2.2.1 is chosen, the second harmonic is all we need and this concludes the processing of the NLD (but in this case a more effective NLD is a rectifier, which will be discussed later in this section). If, however, we propose to use the virtual pitch option of Sec. 2.2.1, we will need to generate at least an additional two to three harmonics. By multiplying x_2 with another scaled replica of x, the third harmonic x_3 can be created, and so on. The output signal y, assuming that three harmonics above f_0 are generated, will become

$$y(t) = h_0 + \sum_{i=1}^{2} \left[h_{2i-1} \sin((2i-1) \times 2\pi f_0 t) + h_{2i} \cos(2i \times 2\pi f_0 t) \right] \tag{2.9}$$

the h_i being the scale factors given as

$$\begin{aligned} h_0 &= \frac{g_1 g_2}{2}\left[1 + \frac{1}{4} g_3 g_4\right], \\ h_1 &= g_1\left[1 + \frac{3}{4} g_2 g_3\right], \\ h_2 &= -\frac{g_1 g_2}{2}\left[1 + g_3 g_4\right], \end{aligned} \tag{2.10}$$

$$h_3 = -\frac{g_1 g_2 g_3}{4},$$
$$h_4 = \frac{g_1 g_2 g_3 g_4}{8},$$

If we prespecify what the amplitudes h_i should be (the amplitude of h_0 is the uninteresting DC term, which will be filtered out later, thus we do not care about its value), then we must choose the scaling factors g_i such that

$$\begin{aligned} g_1 &= h_1 + 3\,h_3, \\ g_2 &= -2\frac{h_2 + 4h_4}{h_1 + 3h_3}, \quad h_1 \neq -3h_3, \\ g_3 &= \frac{2h_3}{h_2 + 4h_4}, \quad h_2 \neq -4h_4 \\ g_4 &= -2\frac{h_4}{h_3}, \quad h_3 \neq 0 \end{aligned} \tag{2.11}$$

As long as none of the numerators of Eqn. 2.11 are zero, the scale factors are well behaved and can be chosen to yield any desired harmonics spectrum. If, for example, all harmonics amplitudes are to be +1, then we have $g_1 = 4$, $g_2 = -10/4$, $g_3 = 2/5$, and $g_4 = -2$. Note that the even harmonics will be $\pi/2$ rad out of phase with the odd harmonics (Eqn. 2.9). For the multiplying NLD, we do not give an example of an AIM simulation (Sec. 1.4.8), as the number and amplitudes of harmonics is not fixed in this case: they both depend on the g_i and the level of multiplication used.

Intermodulation distortion Non-linear devices exhibit the so-called intermodulation distortion: the presence of frequency components in the output that are not harmonically related to frequency components in the input. It is the interaction of the input frequency components that produces these intermodulation distortion products. The frequencies at which they occur are the sum and difference frequencies of the input components, but the amplitudes depend on the kind of non-linearity and also on the amplitudes of the input frequency components.

Assume a signal s with two frequency components, at f_1 and f_2 (which are not themselves harmonically related), with amplitudes 1 and $0 \leq a \leq 1$. The magnitude of the Fourier representation is given by delta functions, which are

$$s \overset{|\mathcal{F}|}{\rightarrow} [f_1] + a[f_2]. \tag{2.12}$$

Here, we have used a shorthand notation on the right-hand side of the arrow, where $a[f_2]$ indicates a frequency component at frequency f_2 with amplitude a, and so on. Multiplying s with itself yields

$$\begin{aligned} s^2 \overset{|\mathcal{F}|}{\rightarrow} (*) \quad & [2f_1] + a^2[2f_2] \\ (\dagger) \quad & 2a\left([f_1 - f_2] + [f_1 + f_2]\right). \end{aligned} \tag{2.13}$$

The harmonic components are indicated by $(*)$ and the intermodulation distortion components by $(\dagger)$. We can define a harmonic-to-distortion energy ratio ς (akin to signal-to-noise ratio)

$$\varsigma = 10\log\frac{\sum f_{(*)}^2}{\sum f_{(\dagger)}^2} \tag{2.14}$$

which, for s^2 would be

$$\begin{aligned}\varsigma_2 &= 10\log\left[\frac{1^2+(a^2)^2}{2\times(2a)^2}\right] = 10\log\left[\frac{1+a^4}{8a^2}\right] \\ &= -10\log 8 + 20\log\frac{1}{a} \quad \text{for } a \ll 1.\end{aligned} \tag{2.15}$$

High ς_2 can be achieved for $a \ll 1$; the minimum value (worst-case distortion) of $\varsigma_2 = -6.0\,\text{dB}$ occurs for $a = 1$. Continuing along the same lines, we now multiply s with itself twice, and obtain

$$\begin{aligned}s^3 \overset{|\mathcal{F}|}{\rightarrow} (*) \quad & \frac{1}{4}((3+6a^2)[f_1] + [3f_1] + 3a(2+a^2)[f_2] + a^3[3f_3]) \\ (\dagger) \quad & \frac{3a}{4}(a([f_1-2f_2]+[f1+2f_2]) + [2f_1-f_2] + [2f_1+f_2]).\end{aligned} \tag{2.16}$$

For s^3, the harmonic-to-distortion energy ratio ς_3 becomes

$$\begin{aligned}\varsigma_3 &= 10\log\left[\frac{23a^6+18a^4+36a^2+5}{9a^2(a^2+1)}\right] \\ &= -10\log\frac{9}{5} + 20\log\frac{1}{a} \quad \text{for } a \ll 1.\end{aligned} \tag{2.17}$$

The lowest value $\varsigma_3 = 6.4\,\text{dB}$ occurs for $a = 0.832$; compared to the minimum value of ς_2, this is a much better situation. For s^4, we compute ς_4 directly as

$$\begin{aligned}\varsigma_4 &= 10\log\left[\frac{17a^8+288a^4+17}{a^2(176a^4+36a^2+176)}\right] \\ &= -10\log\frac{176}{17} + 20\log\frac{1}{a} \quad \text{for } a \ll 1.\end{aligned} \tag{2.18}$$

The worst-case distortion value is now $\varsigma_4 = -1.5\,\text{dB}$, at $a = 0.545$. This analysis can be continued for all necessary levels of multiplication (depending on how many harmonics are desired), but the point should be adequately illustrated. The 'grand total' ς_t, that is, that of the weighted sum of the s^i, with weighting factors g_i, is not a simple combination of the ς_i, but has to be recomputed by adding all harmonic and distortion energies and taking the ratio. The final result will be a function of a and the g_i, and may serve to choose a particular combination of g_i that will maximize ς_t, subject to some constraints on the

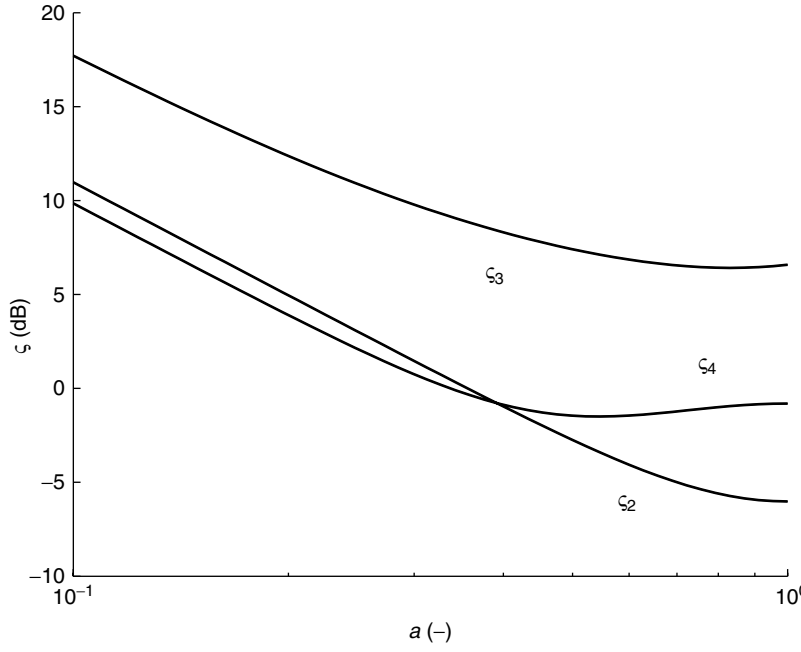

Figure 2.6 The harmonic-to-distortion energy ratio ς for the output signal of a multiplying non-linearity, given a two-tone input with amplitudes 1 and a, as a function of a. The subscripts indicate the highest harmonic number at the output, for example one ς_2 applied to one level of multiplication, where the double frequency is generated

desired harmonic amplitudes h_i (Eqns. 2.10 and 2.11). In Fig. 2.6, $\varsigma_2 - \varsigma_4$ are plotted as a function of a.

The above analysis is of course incomplete in the sense that situations with more than two input frequency components can occur in practice. The analysis becomes very tedious for such involved cases, though. Also, the metric ς can only be regarded as a very crude approximation to subjective quality, for it ignores that some of the weaker components may be masked by the stronger components. As the distortion components are generally smaller than the harmonic components, one could argue that ς overestimates the effects of distortion: in practice, some of the distortion components will be masked. In other words, a high value for ς is always good, but a low value is not necessarily very bad. It would be interesting to study this with a psychoacoustic model, using for example the two-component signal illustrated in the above analysis.

2.3.2.2 Rectifier

Spectral characteristics A very efficient method of harmonics generation is by rectification; either half-wave or full-wave. Both analog and digital implementations are trivial, and another favourable aspect is that rectification is a homogeneous operation. Of course, as a whole the system is non-linear, and the output frequency components

are mostly double those at the input. To compute the harmonics spectrum exactly, we apply a pure-tone input signal of frequency f_0, and compute the Fourier series b_k of the full-wave-rectified signal

$$b_k = f_0 \int_0^{1/f_0} |\sin(2\pi f_0 t)|\, \mathrm{e}^{-2\pi i k f_0 t}\, \mathrm{dt} = \begin{cases} \frac{2}{\pi(1-k^2)} & \text{for even } k, \\ 0 & \text{for odd } k. \end{cases} \tag{2.19}$$

The resulting spectrum consists of only the even harmonics of f_0, which implies that the fundamental frequency of the output signal is now $2f_0$. Perceptually, this means that the synthetic bass sounds an octave too high, compared to the input signal. However, the increase in bass perception using this kind of low-frequency psychoacoustic BWE can still be attractive (mainly because of the efficient implementation). The harmonics spectrum decays quite rapidly, at -12 dB per octave.

Figure 2.7 shows the AIM calculations for a 200-Hz rectified signal, with the input component added back. Thus, the signal contains 200, 400, 800, 1200 Hz, and so on.

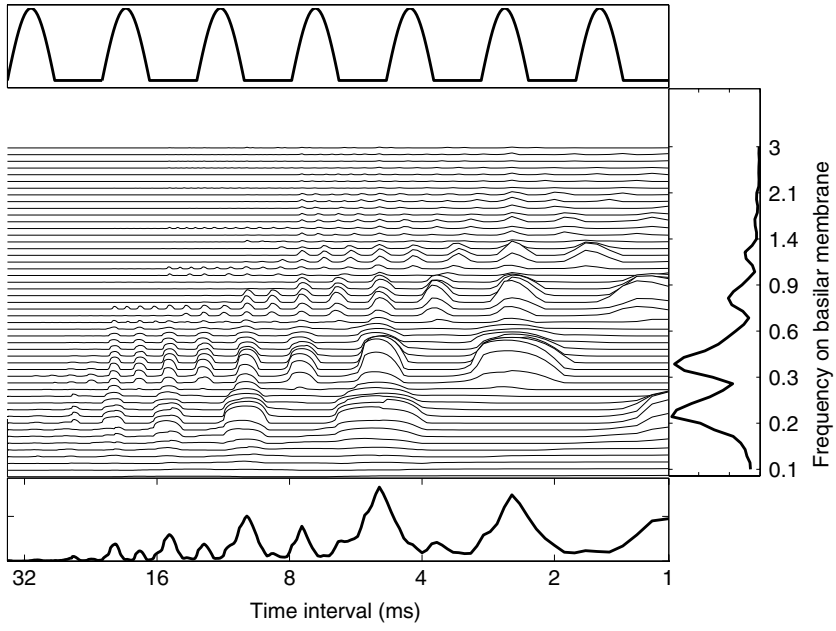

Figure 2.7 AIM calculations for a rectified pure tone with frequency $f_0 = 200$ Hz; the pure tone is added back to the rectified signal, giving a complex tone that includes the fundamental and all even harmonics. The strongest peak in the frequency-collapsed plot (lower panel) occurs at a lag of 5 ms (200 Hz). However, there is also a strong peak at a lag of 2.5 ms (400 Hz). This might indicate an ambiguous pitch or a signal that is segregated into two percepts

Although the 200-Hz component would not normally be present in the output of the rectifying NLD, it is added in this analysis, as without it the pitch percept would be unambiguously at 400 Hz. This situation might occur when using a small loudspeaker that cannot reproduce 200 Hz at a sufficient level. It is of interest to see what the effect of adding back the original 200-Hz component might be on the perceived pitch. The lower panel shows a dominant peak at a lag of 5 ms, corresponding to 200 Hz. There is another peak of almost equal amplitude, however, at a lag of 2.5 ms, corresponding to 400 Hz. This may imply an ambiguous pitch percept, or a failure of the 200-Hz component to group with the harmonic series, leading to two auditory objects with different pitches. If it is indeed a grouping problem, then common amplitude modulation of all components (as would occur in practice) could increase the likelihood that one auditory object is perceived, instead of two.

Temporal characteristics Beside frequency characteristics, temporal behaviour is important as well. In particular, it is desirable that the temporal envelope of the signal remains as close to the original as possible. If, for example, the attack time of an impulsive sound is increased, this can be very noticeable. The temporal behaviour of the rectifying NLD is satisfactory, refer to Fig. 2.8. A 5-cycle 50-Hz tone burst is shown in (a). The rectifying

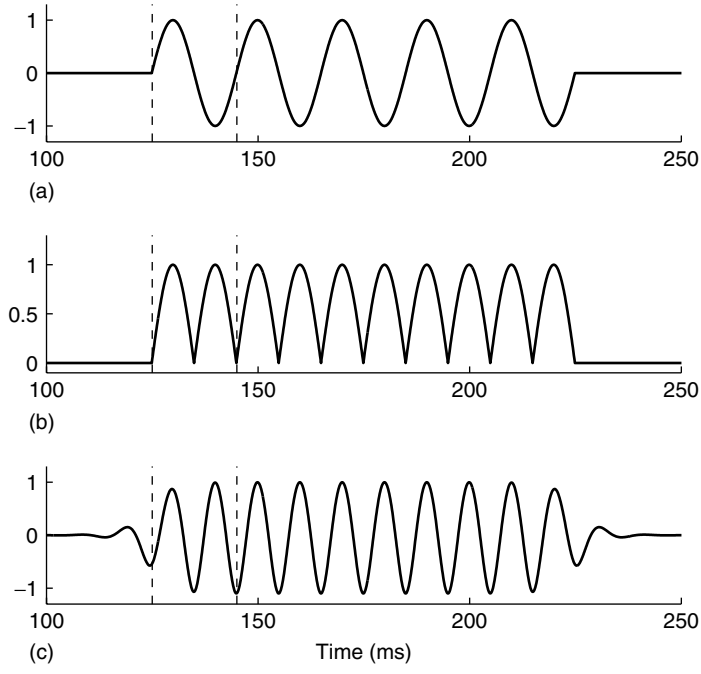

Figure 2.8 (a) Shows 5 cycles of a 50-Hz tone as input signal to a rectifying NLD, the output of which is shown in (b). (c) Shows the result after bandpass filtering between 70 and 150 Hz. The filtered output signal reaches full amplitude in the first cycle, which is beneficial for perceptual quality

NLD produces the signal shown in (b), after which bandpass filtering between 70 and 150 Hz is done (linear phase, with delay compensation), shown in (c). The filtered output reaches maximum amplitude within the first cycle.

Intermodulation distortion The robustness of this NLD to intermodulation distortion when presented with two-tone stimuli, or more complex input spectra, can be assessed using expressions derived in the appendix of this chapter, which are taken from Larsen and Aarts [156]. Although the rectifier is a non-linear system, the full output spectrum can be computed conveniently for arbitrary periodic input signals. Consider a signal $f(t)$, with Fourier series coefficients a_k. The rectified output signal has Fourier series coefficients b_k, which are given by

$$b_k = (2t_0 - 1)a_k - \sum_{n \neq k} \frac{n a_n}{i\pi k(k-n)}[1 - \mathrm{e}^{i2\pi(n-k)t_0}] \quad (n \in \mathbb{Z},\ k \neq 0), \tag{2.20}$$

where it is assumed that the input signal has period 1, with a single zero crossing at t_0. Note that for a pure tone ($a_1 = 1/2i$, $a_{-1} = -1/2i$, $t_0 = 1/2$) the result of Eqn. 2.19 is obtained. The more general case, in which the period of the signal is arbitrary and with an arbitrary number of zero crossings per period, is presented in the appendix at the end of this chapter. There, it is also shown that for large k the b_k are mainly determined by the slope of f at its zero crossings.

To quantify the amount of harmonic energy versus intermodulation distortion energy, we can again use the metric ς (Eqn. 2.14). A complicating factor in evaluating Eqn. 2.20, or its more general form as given in the appendix (Eqn. 2.93), is that the period, or rather the fundamental frequency f_0, of the input signal is a function of the input frequency components. Assume an input signal with frequencies f_1 and f_2 and amplitudes 1 and $0 \leq a \leq 1$; f_0 is the greatest common divisor of f_1 and f_2, for example, 10 Hz for input frequencies of 50 and 70 Hz, which are then the fifth and seventh harmonics: $a_{\pm 5} = \pm 1/(2i)$ and $a_{\pm 7} = \pm a/(2i)$ in Eqn. 2.93. In accounting the harmonic energy of the output signal, we sum all $b^2_{\pm 5n}$ and $b^2_{\pm 7n}$, $n \in \mathbb{Z}$. We numerically compute ς for a number of representative cases. Hence, given f_1 and f_2 (not themselves harmonically related) with amplitudes 1 and a, we first compute all zero crossings in (0, 1] and then apply Eqn. 2.93. The harmonic energy is computed as stated above, and all excess energy is considered as originating from intermodulation distortion. Figure 2.9 shows ς for various $f1/f2$ ratios; $\varsigma(x, y)$ indicates the harmonic-to-distortion energy ratio for frequencies x and y (and multiples thereof, e.g. $\varsigma(2, 3)$ would be valid for frequencies 40 and 60 Hz, and 50 and 75 Hz, etc.). It appears that ς does not depend heavily on the ratio f_1/f_2. As ς is quite large for small a, the rectifier will perform well if there is one frequency component that dominates all others; however, if there are two (or more) components of comparable amplitude, then distortion will be severe.

2.3.2.3 Integrator

Spectral characteristics Another efficient method of generating harmonics is by integrating the rectified input signal, and resetting the output to zero after each second zero crossing. A discrete-time algorithm would thus process an input signal $x(n)$ into output

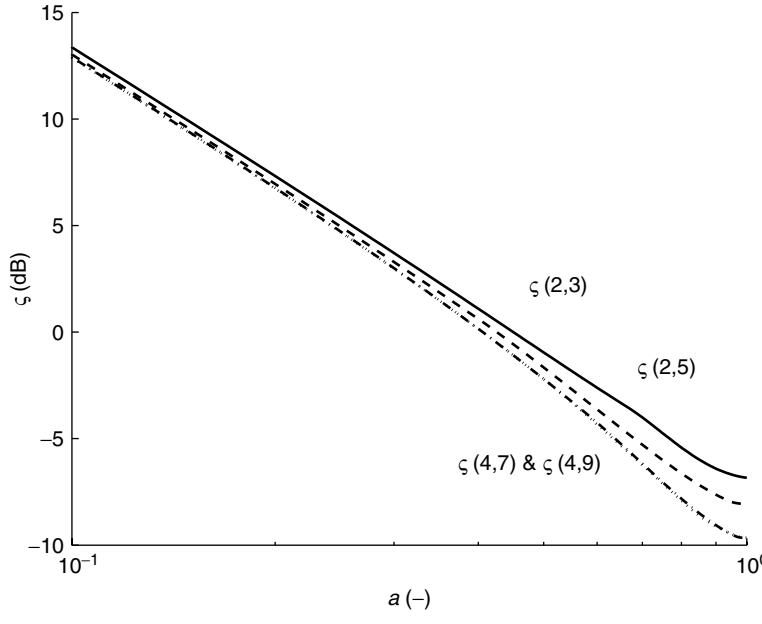

Figure 2.9 For a two-tone input with frequency ratios (f_1, f_2), and amplitudes 1 and a, Eqn. 2.93 can be used to compute the harmonic-to-intermodulation distortion energy ratio ς, for the rectifying NLD

$y(n)$ according to

$$y(n) = \begin{cases} 0 & \text{if} \quad z(x(n)) = 1 \text{ and } x(n) - x(n-1) > 0, \\ y(n-1) + c|x(n)| & \text{otherwise.} \end{cases} \tag{2.21}$$

c is a constant of integration and z is a function that detects zero crossings. In Eqn. 2.21, the output is reset to zero at positive-going (positive derivative of $x(n)$) zero crossings, but negative-going is also possible. The output signal will have the same fundamental frequency as the input signal, and for a pure-tone input will resemble a saw-tooth waveform. The integration has a low-pass filtering effect, but the discontinuities in the output due to the resetting create a strong harmonics spectrum. As is true for the rectifier, the integrator is a homogenous system, that is, input and output amplitudes are linearly related. Assuming a pure tone of frequency 1, the output signal will be (continuous-time)

$$y(t) = \begin{cases} \frac{2}{\pi}(1 - \cos(2\pi t)) & \text{for} \quad t \leq \frac{1}{2}, \\ \frac{2}{\pi}(3 + \cos(2\pi t)) & \text{for} \quad t > \frac{1}{2}. \end{cases} \tag{2.22}$$

Here we define $t \in [0, 1)$, which is the periodicity interval of $y(t)$. The Fourier series coefficients c_k then follow as

$$c_k = \frac{2k^2 + (-1)^k - 1}{i \times 2k\pi(k^2 - 1)}. \tag{2.23}$$

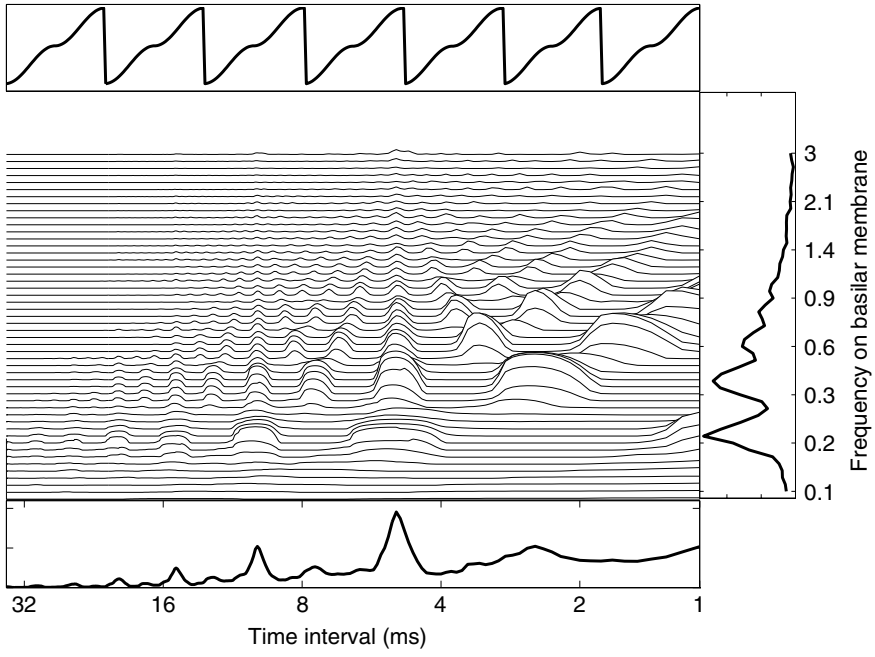

Figure 2.10 AIM calculation for the output of the integrating NLD for a 200-Hz pure-tone input. There is a single large peak at a lag of 5 ms, indicating a strong pitch percept at 200 Hz

Thus, the harmonics spectrum comprises all (even and odd) harmonics, decaying at a rate of -6 dB per octave for large k, and should give rise to a strong pitch at the fundamental. Therefore, we might expect the integrator to be a good algorithm to serve as NLD for low-frequency psychoacoustic BWE. This is confirmed by Fig. 2.10, which shows an AIM calculation for the integrating NLD output of a 200-Hz pure-tone input. The presence of all harmonics in the output signal yields one dominant peak at a lag of $\tau = 5$ ms (200-Hz pitch). Even if the fundamental is removed by filtering, or by a poor loudspeaker response at low frequencies, the 200-Hz pitch remains strong and unambiguous, as is shown in Fig. 2.11. The AIM calculation was done on a high-pass filtered (300 Hz cut-off) version of the integrated 200-Hz pure tone.

Temporal characteristics An analysis of the temporal characteristics shows that the output lags the input, refer to Fig. 2.12. A 5-cycle 50-Hz tone burst is shown in (a). The integrating NLD produces the signal shown in (b), after which bandpass filtering between 70 and 150 Hz is done, shown in (c). This filtered output signal remains small during the first cycle, and reaches a significant amplitude only at the end of the first cycle (where the integrator resets the output to zero). The combined effect is that the attack time of the input signal is increased, and also that the onset is delayed. This could be particularly detrimental if higher harmonics of the 50-Hz fundamental are present. More will be said

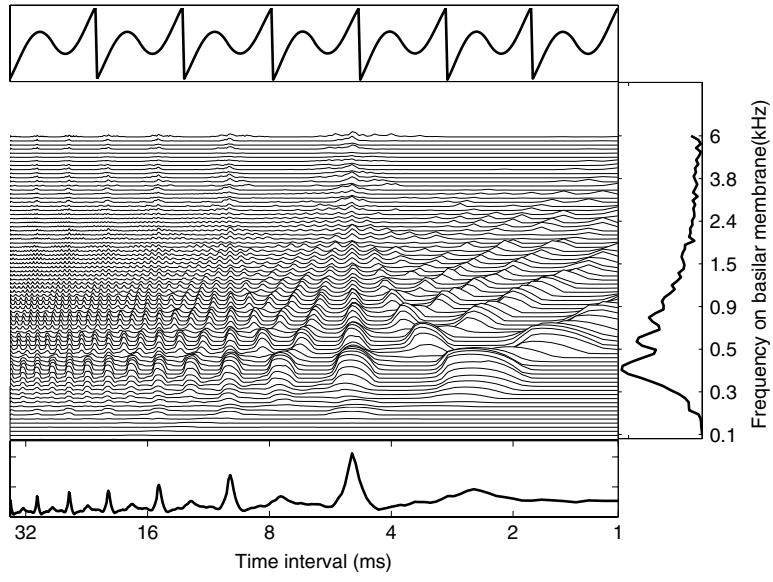

Figure 2.11 AIM calculation for the output of the integrating NLD for a 200-Hz pure-tone input, in which the 200-Hz component is filtered out (simulating the transfer function of a small loudspeaker). Even though the fundamental is not present in the spectrum, the remaining harmonics yield a single large peak at a lag of 5 ms, indicating a strong pitch percept at 200 Hz

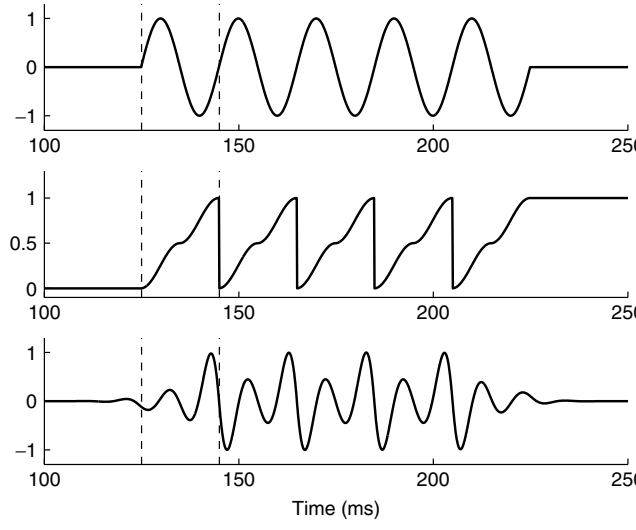

Figure 2.12 (a) Shows 5 cycles of a 50-Hz tone as input signal to an integrating NLD, the output of which is shown in (b). (c) Shows the result after bandpass filtering between 70 and 150 Hz. The filtered output signal rises more slowly than the input, and only reaches a significant amplitude at the end of the first cycle. This can lead to a degraded percept of the signal

about temporal characteristics when discussing the phase characteristics of the filters of low-frequency psychoacoustic BWE, in Sec. 2.3.3.3.

The percept is that signals with fast attacks and/or decays sound less 'tight'.

Intermodulation distortion As for the rectifier, the output signal spectrum for arbitrary periodic input signals can be computed conveniently; the derivation is presented in the appendix. The resulting Fourier series c_k for a given input Fourier series a_k is

$$c_0 = \int_0^1 (1-t)|f(t)|\,\mathrm{dt}, \tag{2.24}$$

$$c_k = \frac{b_k - \alpha_0}{i2\pi k}, \quad k \neq 0, \tag{2.25}$$

for the special case that we assume $f_0 = 1$ and there is one zero crossing in the interval $[0, 1]$. The magnitude of the integrator output just before resetting is α_0; the b_k are the Fourier series coefficients of the rectified output signal, and are given in Sec. 2.3.2.2. The b_k decay as $1/k$, so for large k, c_k will be proportional to α_0/k. Equation 2.108 (appendix) can be used to assess the relative amount of intermodulation distortion energy given a two-tone input signal (with possibly multiple zero crossings in the periodicity interval). An analytic solution is not available as the parameter α_0 depends on the frequencies and amplitudes of the input frequency components in a complicated non-linear way. Thus, we resort to numerical methods to compute ς, the harmonic-to-intermodulation distortion energy ratio, as in Sec. 2.3.2.2. Results are shown in Fig. 2.13 for a few f_1, f_2 combinations, where the amplitude of f_1 is always 1, and the amplitude of f_2 is $0 \leq a \leq 1$. In comparison to Fig. 2.9, which plots ς for various f_1/f_2 ratios using the rectifying NLD, the integrating NLD is seen to be significantly more robust against intermodulation distortion, as the ς's are considerably larger. In fact, $\varsigma > 0$ for almost all a. The graphs in Fig. 2.13 display a number of 'knee points', where ς suddenly decays more rapidly with increasing amplitude a of the f_2 component. This occurs because at particular values of a, additional zero crossings are created in the periodicity interval, which cause large changes in the output spectrum (see also Eqn. 2.108 in the appendix).

2.3.2.4 Clipper

Spectral characteristics A convenient way to generate a harmonics signal with only odd harmonics is by means of a limiter or a clipper. The limiter output signal g_l in response to an input f is

$$g_l(t) = \begin{cases} 1 & \text{if} \quad f(t) \geq 0, \\ -1 & \text{if} \quad f(t) < 0. \end{cases} \tag{2.26}$$

For the clipper, the output signal g_c is

$$g_c(t) = \begin{cases} f(t) & \text{if} \quad |f(t)| \leq l_c \\ l_c & \text{if} \quad f(t) > l_c, \\ -l_c & \text{if} \quad f(t) < -l_c, \end{cases} \tag{2.27}$$

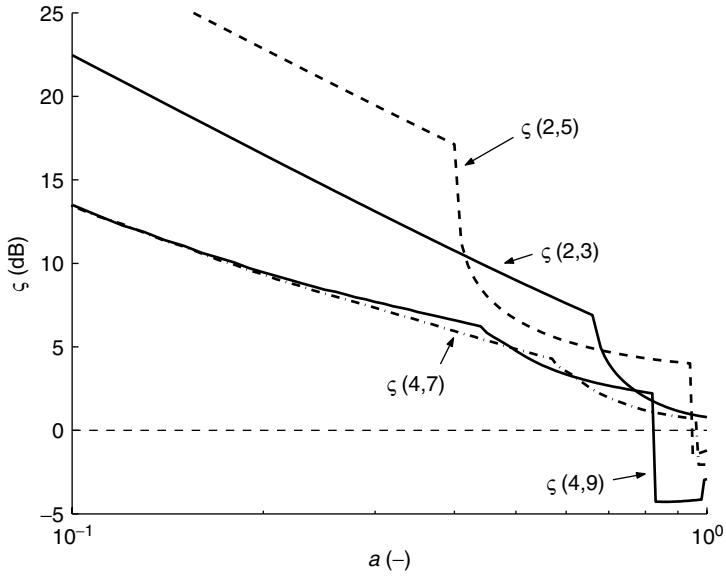

Figure 2.13 For a two-tone input with frequency ratios (f_1, f_2), and amplitudes 1 and a, Eqn. 2.108 can be used to compute the harmonic-to-intermodulation distortion energy ratio ς, for the integrating NLD

where l_c is the *clipping level*, here taken to be symmetrical around zero. One could also define different clipping levels for positive and negative signal values. Both the limiter and the clipper will generate odd harmonics of a pure-tone input signal, but are not directly suitable for BWE applications, as they are not homogeneous systems. For the limiter, this is because the output level is always ± 1, a highly non-linear characteristic. This can be overcome by detecting the envelope of the input signal and scaling the limited signal appropriately. For the clipper, the situation is a little bit more complicated. For low input levels, there may be no clipping at all, if $|f|$ does not exceed l_c, thus $g_c = f$. At intermediate input levels, moderate clipping will occur, with the desired harmonics generation. At very high input levels, such that mostly $|f| \gg l_c$, the clipper becomes a limiter (with output $\pm l_c$). The characteristics of a clipper vary significantly as the input level varies. Again, this can be overcome, or at least reduced, by scaling l_c in response to the level of f. In fact, by doing this in a special way, the clipper has demonstrated very good subjective results in the low-frequency psychoacoustic BWE application – we will elaborate on this later. As the subjective performance of the clipper is generally superior to that of the limiter, we will focus on the clipper in the remainder of this section. In Chapter 5, on high-frequency BWE of audio, we will introduce the 'soft' clipper, an operation that does not have a 'hard' threshold above which the output signal is not allowed to rise, but rather a mild compression of the input as the input level increases. The clipping as discussed in this section is hard clipping.

The spectral characteristics of a clipped sine depend greatly on the clipping level l_c. As a special case of the more general situation described in the appendix (Sec. 2.6.4), the

Fourier series coefficients of the clipped sine are, for the fundamental

$$a_1 = \frac{(2t_1 + \sin 2t_1)}{\pi}, \tag{2.28}$$

(using $t_1 = \sin^{-1} l_c$), and for the odd harmonics (even harmonics are zero because of the symmetry)

$$a_{2n+1} = \frac{\sin 2(n+1)t_1}{\pi(n+1)(2n+1)} + \frac{\sin 2nt_1}{\pi n(2n+1)} \tag{2.29}$$

As l_c approaches 0, we find as limiting case (for all values of n)

$$a_n \approx \frac{4l_c}{\pi n}, \tag{2.30}$$

which is also the result for the limiter (at level l_c). Figure 2.14 shows the Fourier series coefficients according to Eqns. 2.28 and 2.29 for $0 <= l_c <= 1$.

Figure 2.15 shows an AIM calculation for a clipped 200-Hz pure tone. There is a clear peak at $\tau = 5$ ms, corresponding to a pitch percept of 200 Hz. There are two rather large and broad peaks at smaller lags, and these become very prominent if the fundamental frequency is removed, for example, by high-pass filtering at 300 Hz, as shown in Fig. 2.16. The sharpest peak still occurs at $\tau = 5$ ms (200-Hz pitch), but the other peaks (around 3 and 2 ms) are very large as well. The reason these peaks occur is that whereas only a 200-Hz fundamental 'fits' the given harmonic series perfectly, there are other possible

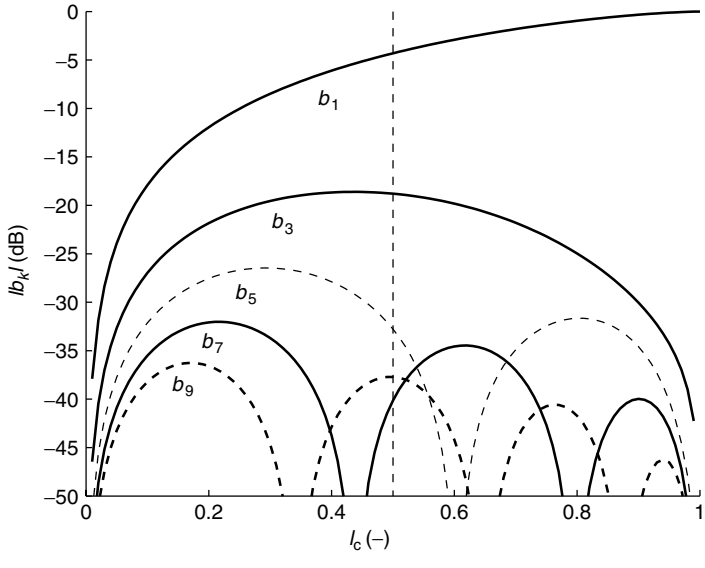

Figure 2.14 Magnitudes of harmonics of a clipped sine; clipping level l_c. Harmonic k is indicated as b_k; note that for $k \geq 5$, magnitudes may be zero for some l_c. The dashed vertical line indicates the commonly used value of 0.5 for the clipping level

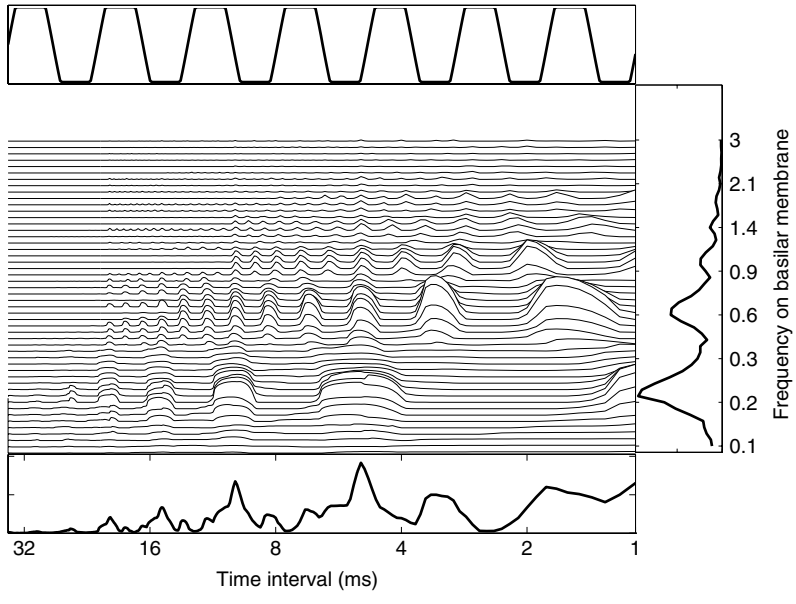

Figure 2.15 AIM calculations for a complex tone consisting of f_0 with odd harmonics (3, 5, and 7)

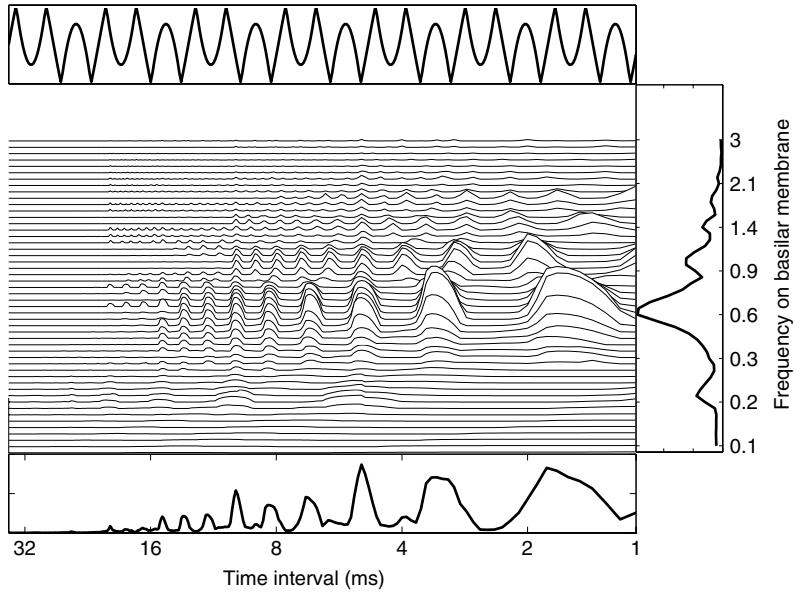

Figure 2.16 AIM calculations for a complex tone consisting of odd harmonics (3, 5, and 7), without fundamental f_0

fundamentals that fit 'reasonably' well. Especially, if we only consider the odd harmonics 3, 5, and 7 (which are in the dominance region for pitch perception (Ritsma [227]), then a 333.3- or a 500-Hz fundamental fits the harmonic at 1000 Hz exactly and roughly matches the 600 and 1400 Hz harmonics. Because the fit is not exact, pitches corresponding to either 333.3 or 500 Hz would be quite vague. But it would seem that the signal as a whole would not have a well-defined pitch, as would for example, the output signal of an integrator (Sec. 2.3.2.3 and Figs. 2.10–2.11). Thus, when a clipping NLD is used in low-frequency psychoacoustic BWE, and the fundamental frequency is not reproduced, then the resulting pitch may not be very strong at the original fundamental.

Temporal characteristics Figure 2.17 shows a 50-Hz signal (a), which clipped at $l_c = 0.5$; the resulting signal is shown in (b) (amplitude normalized). The clipped signal is filtered between 70 and 150 Hz, as shown in (c). The filtered output reaches maximum amplitude within the first cycle, thus the temporal characteristics are good.

Intermodulation distortion The general expression for the output spectrum Fourier coefficients b_n of a clipped periodic input signal with Fourier coefficients a_n is given in the appendix, Sec. 2.6.4. The result for a signal with period 2π is

$$b_n = \frac{1}{2\pi} \sum_{k=1}^{K} \sum_{m=-\infty}^{\infty} m a_m \frac{e^{i(m-n)\beta_k} - e^{i(m-n)\alpha_k}}{i(m-n)}, \tag{2.31}$$

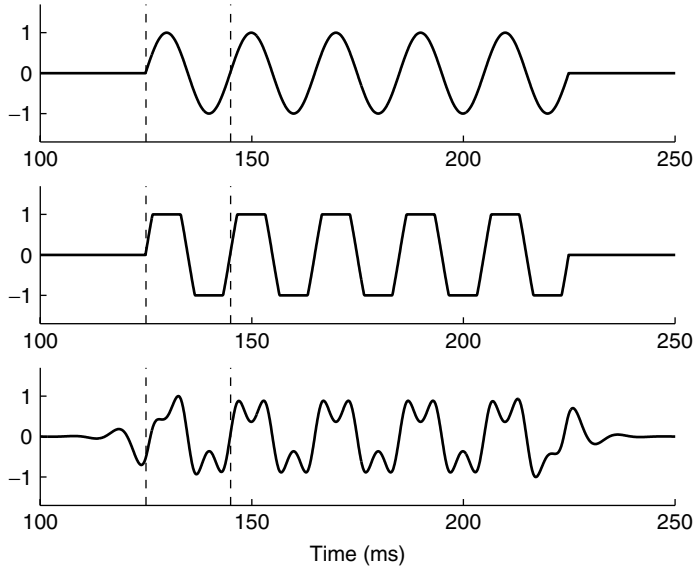

Figure 2.17 (a) This shows 5 cycles of a 50 Hz tone as input signal to a clipping NLD, the output of which is shown in (b). (c) This shows the result after bandpass filtering between 70 and 150 Hz. The filtered output reaches maximum amplitude in the first cycle, which is beneficial to perceived quality

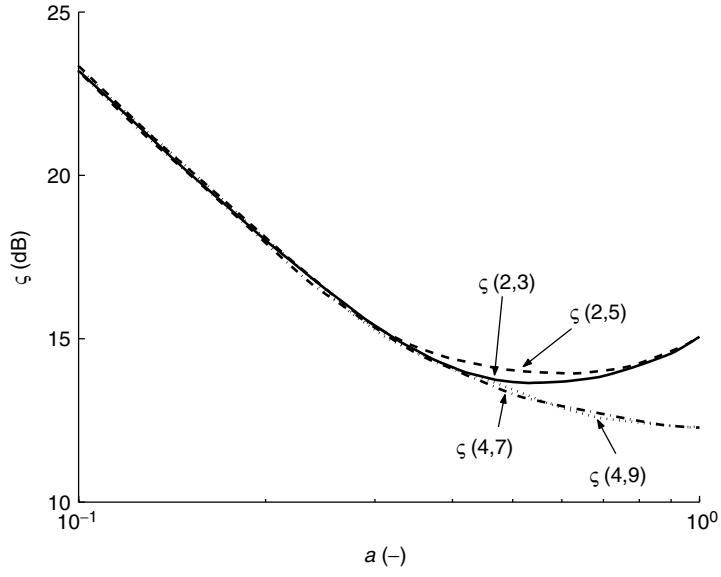

Figure 2.18 For a two-tone input with frequency ratios (f_1, f_2), and amplitudes 1 and a, the harmonic-to-intermodulation distortion energy ratio ς for the clipping non-linearity, at a clipping level of 0.5 (signal is normalized to the range [-1, +1])

with the α_k, β_k $(k = 1 \ldots K)$ determining the intervals $[\alpha_k, \beta_k]$ where the signal is not clipped; these will depend on the clipping level l_c and the amplitude of the signal.

This analysis was performed for a two-tone input signal, with frequencies f_1 and f_2, amplitudes 1 and $a = [0, 1]$. Because a clipping non-linearity is not a homogeneous system, all signals were normalized to the range $[-1, +1]$, and then clipped at a level of 0.5. Without this normalization ς, values are about 2 dB lower for $a \approx 1$. For the normalized clipped signals, ς is shown in Fig. 2.18. It is obvious that the values are significantly higher than for any of the preceding non-linearities discussed (multiplier, rectifier, integrator). Presumably, this is due to the fact that during portions in which the signals are clipped, the output remains fixed at the clipping level, and the influence of the interfering frequency components is thus minimized. The effect of clipping level l_c (for a fixed two-tone input signal) is not large for $l_c < 0.5$: ς drops by only a few dB as $l_c \downarrow 0$ (indicating that a limiting non-linearity performs slightly worse than a clipping non-linearity with respect to intermodulation distortion). As $\lim_{l_c \uparrow 1}$, $\lim_{\varsigma \to \infty}$.

Input-level-dependent clipping level It was already mentioned that the clipping level l_c should be scaled according to the envelope of its input signal f. Here, we shall discuss how this scaling can be implemented, following a method proposed by C. Polisset. The basic idea is to follow the envelope of f with different time constants during the attack and decay of the waveform. We define the nominal clipping level $l_{\mathrm{N},c}$ such that for stationary input signals

$$l_c = l_{\mathrm{N},c} \max |f(t)|, \tag{2.32}$$

for example, $l_{\mathrm{N},c} = 1/2$ would be a typical choice, such that the clipping level is half of the maximum absolute value of f. The time dependence of l_c can then be defined as (assuming a sample rate of $1/T_\mathrm{s}$)

$$l_c(t) = \begin{cases} a l_c(t - T_\mathrm{s}) & \text{if} \quad |f(t)| \leq l_c(t)/l_{\mathrm{N},c}, \\ l_{\mathrm{N},c}|f(t)| & \text{if} \quad |f(t)| > l_c(t)/l_{\mathrm{N},c}. \end{cases} \tag{2.33}$$

Such a dependence will cause l_c to follow without delay, any increase in amplitude of f, but to decrease at a maximum rate given by the parameter a of Eqn. 2.33. For stability reasons, $0 < a < 1$. To achieve a specified decay time $\tau_{1/2}$ (duration in which l_c will halve in value), a is given by

$$a = \mathrm{e}^{-\ln 2 \frac{\tau_{1/2}}{T_\mathrm{s}}}. \tag{2.34}$$

The limited rate of decrease of l_c is purposefully chosen such that typical musical signals decay much faster. An illustration hereof is given in the upper panel of Fig. 2.19. The solid line shows a decaying musical signal, and the dashed line is the associated clipping level. During the initial phase of the signal, where the envelope rises fast, the clipping level is increased to half of the envelope. During the next phase, the level of the signal is sustained, causing a stationary envelope. The clipping level is also stationary. The harmonics spectra during the various phases are shown schematically in the accompanying insets of Fig. 2.19. The last part of the signal is the decay, where the clipping level decays at a much slower rate than the signal envelope. Therefore, before the signal reaches zero amplitude, the clipping level has exceeded the signal level. From that point on, no harmonics will be generated. Effectively, after the sustained period of the signal, the harmonics spectrum slowly changes from its maximum strength to complete absence (no harmonics).

The time constant for the decay of l_c is usually taken in the order of a few seconds. This will present a problem if the audio signal decreases its overall level rapidly, because no harmonics generation will occur until l_c has decayed sufficiently. Such cases do not often occur though, in particular not in modern music, which typically has a very low dynamic range. The clipping level usually does not vary over a great range.

The perceptual effect of the varying harmonics strength seems to be generally beneficial to the low-frequency psychoacoustic BWE application. Some reasons for this observed benefit might be:

- Most instruments have time-varying spectral characteristics, and the level-dependent clipper might emulate these characteristics better than other NLDs.
- During the final part of the signal, there is no harmonics generation at all, and thus the output of the level-dependent clipper is equal to its input. This means that the latter part of the output signal will have a much lower audibility than if a full harmonics spectrum were generated. The 'tone-lengthening' effect mentioned in Sec. 2.2.3 might thus be decreased, or prevented altogether. This means the audibility of the output signal has been greatly increased relative to the input, but tone duration has not markedly changed. One could assume that this is preferred by listeners.

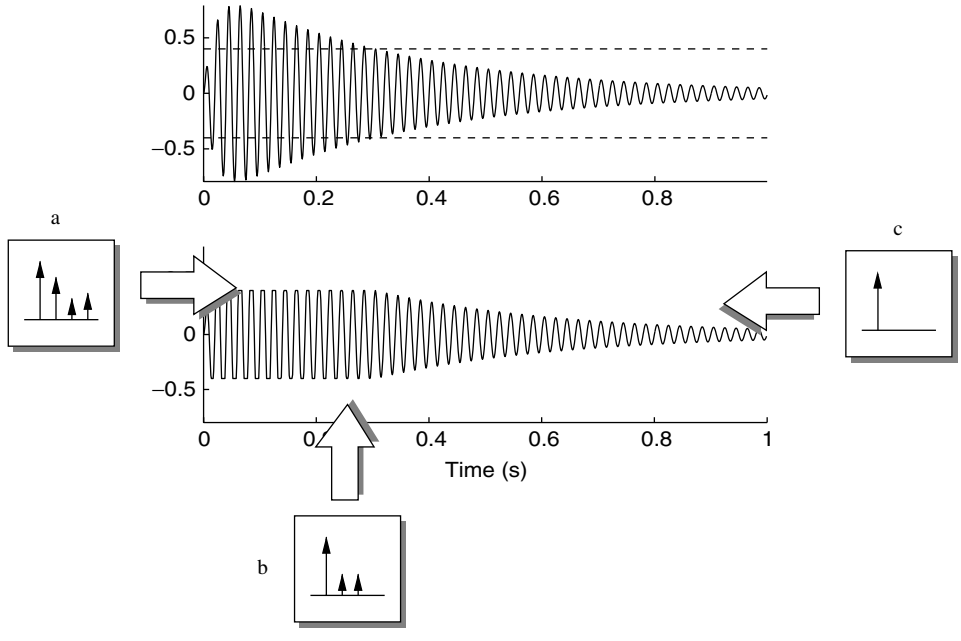

Figure 2.19 A transient signal and associated clipping level (dashed line). As the signal level decreases, the clipping level decreases at a much slower rate (it is shown constant in this figure). This causes the harmonics spectrum to change during the lifetime of the signal, as is shown in the accompanying insets. Inset a shows the harmonics at maximum strength at the beginning of the signal; inset b shows a moderate strength signal when the clipping level becomes relatively high; inset c shows the harmonics spectrum during the decay, where the signal level is below the clipping level. The only component left is the fundamental frequency

- Because the latter parts of the output signal have a weak or non-existent harmonics spectrum, the overall timbre remains closer to that of the input. Even though this comes at a cost of reduced loudness, the closer-matching timbre might be preferred by listeners.

The time-varying harmonics spectrum is a natural effect of a clipping non-linearity, and the method of varying the clipping level according to Eqn. 2.33 gives an appropriate spectro-temporal characteristic to the clipper. It would be much harder to implement something similar for a rectifying (Sec. 2.3.2.2) or integrating (Sec. 2.3.2.3) non-linearity. These NLDs produce a harmonics spectrum that is independent of level, and the only way to create a time-varying harmonics spectrum would involve modifying some basic property of the NLD.

Amplitude non-linearities on various time scales As presented here, a point has been made that the level-*dependent* nature of the clipper is beneficial to subjective quality.

However, on several earlier occasions it has been emphasized that a level-*independent* NLD is the best for BWE applications (NLDs should be homogeneous systems). Thus, there appears to be a contradiction. The resolution of this (apparent) contradiction is the fact that the notion 'level dependent' can be viewed on different time scales. The clipper with varying l_c (Eqn. 2.33) is designed with the aim to provide a level-dependent clipping on a small time scale, that of a single note. If the entire level of the audio signal changes on a larger time scale, the clipping level will adjust appropriately. Therefore, the level-dependent clipper is level independent (linear in amplitude) if one considers a 'large-enough' temporal window. In conclusion, a more precise statement might be that BWE algorithms should be homogeneous on a 'large' time scale, in which 'large' means considerably longer than a typical musical note. On a 'small' time scale (approximately the duration of a tone), the algorithm may be non-linear in amplitude.

2.3.2.5 Discussion of Non-linear Devices

All of the NLDs discussed previously in this section have distinct advantages and disadvantages. For each NLD, an analysis was presented of spectral, temporal, and intermodulation distortion characteristics, and a summary of these is shown in Table 2.1. Apart from such an objective point of view, a subjective rating is ultimately more important; of course the objective analysis helps to understand the subjective impressions.

Subjective experiments will be discussed in Sec. 2.5, in which the rectifying and integrating NLDs were included. The result of that experiment was that both these NLDs were rated approximately equal, with a slight advantage for the rectifying NLD, which may seem surprising given the better spectral characteristic of the integrator. The clipping NLD was not included in this test as it was, at the time of the experiment, not fully developed. Subsequent subjective evaluations have usually favoured the clipping NLD over all others, although no formal experiments have been conducted to confirm this. Much more on the subjective evaluation of low-frequency psychoacoustic BWE will follow in Sec. 2.5.

Table 2.1 Summary of objective features of the various NLDs. The last row describes a frequency-tracking NLD, to be discussed in Sec. 2.4

Characteristic	Amp.-linear	Spectral	Temporal	Interm. dist.
Multiplier	No, needs signal level scaling	Flexible	Good	Variable, depends on harmonic number
Rectifier	Yes	Even harmonics (pitch doubling)	Good	Moderate-poor depends on input
Integrator	Yes	All harmonics	Poor, slow attack/decay	Good
Clipper	Long-time: yes Short-time: no	Odd harmonics, is ambiguous without f_0	Good	Very good
Freq. track.	Yes	Flexible	Good	Excellent

2.3.3 FILTERING

Whichever NLD is used in the low-frequency psychoacoustic BWE system (Figs. I.2 and 2.4), it must be supplied with an appropriate input signal. Also, its output usually needs some filtering to yield a pleasant timbre. The characteristics of these two filters will be discussed in this section. Also, for low-frequency psychoacoustic BWE, it is important that the filters are linear phase, as will be demonstrated in Sec. 2.3.3.3.

2.3.3.1 Filter 1

Filter 1 precedes the NLD and its function is to pass only those frequencies that need to be enhanced. The use of this filter is one of the essential differences between the use of controlled distortion for low-frequency psychoacoustic BWE applications versus uncontrolled distortion as may occur in amplifiers or loudspeakers. In the discussions of various NLDs in Sec. 2.3.2, we found that introducing more than one frequency component to the NLD leads to intermodulation distortion, which should be avoided. Thus, the bandwidth of filter 1 should not be too large. If necessary, filter 1 could be replaced by a filterbank, spanning the same frequency range as the original filter, with each filter connected to an identical NLD, the outputs of which will be summed at the end. In such an arrangement, each filter has a very narrow bandwidth, and intermodulation distortion will be minimized, at the expense of increased algorithmic complexity. However, on the basis of our experience, the use of one single filter does not cause excessive intermodulation distortion, and therefore the use of the just-mentioned filterbank does not seem necessary.

Filter 1 should not pass frequencies above the low-frequency cut-off, f_l, of the loudspeaker, as these components should be adequately reproduced without processing. Therefore, the upper limit of filter 1 will be at most f_l. In most applications, this value will vary somewhere between 70 and 200 Hz. In principle, the lower limit of filter 1 should be approximately 20 Hz, as this is around the minimum audible frequency. But, if the upper limit of filter 1 is very high, it may be better to increase this lower limit somewhat; a bandwidth of two octaves should suffice. Note that the lower limit of musical pitch lies around 40 Hz (Guttman and Pruzansky [102]), so it might be argued that including frequency components below this limit is of questionable value. Nevertheless, frequencies below 40 Hz do occur in music (albeit very rarely), and as it is the aim of low-frequency psychoacoustic BWE to enhance bass perception, we will advocate the use of 20 Hz as the lower limit. If the limiting frequency is set even lower, then any energy below 20 Hz (if present in the audio signal for whatever reason), will be reproduced at the correct fundamental frequency, which, being so low, is heard as an amplitude modulation instead of a unified low-pitch percept. The effect on artificially generated tones of frequency lower than 20 Hz does not sound good, and therefore, the lower cut-off frequency of filter 1 should not be lower than 20 Hz.

The order of the filter does not seem to have too great an effect on quality. Low- and high-pass flanks of second order (-12 dB per octave) seem to be sufficient for adequate separation of the bass frequencies. Alternatively, a stopband attenuation of 20 dB will suffice. For the passband ripple, a value of 1 dB seems good enough; it is hard to perceive any deleterious effects of this ripple if one is presented only with the BWE-processed signals.

The phase response of filter 1 should be linear, the reason for which will be demonstrated in Sec. 2.3.3.3.

2.3.3.2 Filter 2

The second filter, placed behind the NLD, filters the harmonics spectrum such that a reasonable timbre results. This is necessary as the timbre of the harmonics directly out of the NLD usually sounds too 'sharp', which is caused by the harmonic amplitudes being too large. By low-pass filtering, a more pleasant timbre can be achieved. Again, a second-order filter (12 dB per octave) usually suffices. Note that a low-pass flank of −12 dB per octave does not mean that successive harmonics will be attenuated by 12 dB relative to each other. Because the fundamental frequency of the filtered signal is usually quite low, there may be several harmonics present in a single octave at the low-pass flank of the filter. This is illustrated in Fig. 2.20.

The fundamental frequency is usually attenuated by filter 2, because it is either not desired in the output signal, and if it is, it is available directly from the original audio signal. Thus, filter 2 employs a high-pass flank, of moderate order, with a cut-off frequency that is (roughly) equal to the higher frequency limit of filter 1. Thus, filter 2 has a bandpass characteristic, with a bandwidth of about 1 to 1.5 octaves wide.

Filter 2 is preferentially implemented as a linear-phase filter, for reasons that will be explained in the next section.

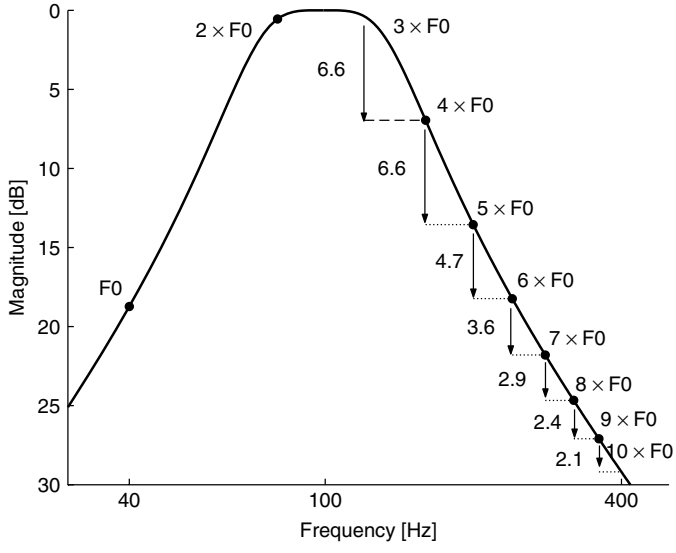

Figure 2.20 A second-order Butterworth bandpass filter with cut-off frequencies of 70 and 140 Hz. A harmonics signal with 40-Hz fundamental has harmonics as indicated by the filled circles: the filter attenuation in dB of successive harmonics is indicated on the right flank of the filter. Note that this attenuation depends both on the filter order *and* on the fundamental frequency value

2.3.3.3 Linear versus Non-linear Phase Filters

The topic of using linear-phase versus non-linear-phase filters in audio processing can sometimes be a controversial one. Although there is little scientific evidence that people are sensitive to phase distortion – excluding some special cases – some would claim that linear-phase systems sound far better than their non-linear-phase counterparts. The issue of linear or non-linear phase for low-frequency psychoacoustic BWE filtering can be analysed objectively, and the conclusion is that it *is* better to employ linear-phase filtering, for reasons to be explained next.

Because in low-frequency psychoacoustic BWE the filter bandwidths and cut-off frequencies are usually orders of magnitude smaller than the system sample rate, IIR implementations are more efficient computationally than a direct FIR implementation. For a modest sample rate of 10 kHz, a frequency resolution of 10 Hz (which would be required to design a filter with a passband of about 100 Hz) would necessitate 1000 taps. In contrast, an IIR filter of ten or less coefficients will probably achieve the desired requirements as well. A drawback of IIR filters is that the phase is non-linear. Lower-order FIR filters are possible if the signal is downsampled before NLD processing. Because NLD implementations are computationally trivial (rectification, clipping, integration), there is not much to gain from downsampling from a computational point of view, and the required anti-alias filters will probably negate the advantage of processing at a lower sample rate. Another option to use FIR filters at reduced complexity is through frequency warping (Härmä *et al.* [104]). With this technique, it is possible to trade high-frequency resolution for low-frequency resolution, which would allow lower-order FIR filters to be used. The concept has not been evaluated for low-frequency psychoacoustic BWE, though.

To be explicit, the problems with non-linear-phase filters in low-frequency psychoacoustic BWE are:

- Interference of synthetic harmonics signal with other signal components. The output of the NLD consists of harmonics, and sometimes the fundamental frequency component as well. After filtering by filter 2, these are added back to the main signal. Because the main signal also contains the fundamental, and possibly some harmonics, interference will occur. As the phase relationships of the original fundamental and its harmonics and the synthetic BWE signal are impossible to predict a priori, the nature of this interference (constructive or destructive) is unknown. We can examine this issue to some degree by using a method devised by C. Polisset (private communication), in which we use a pure-tone input and compute the steady-state output signal for a BWE algorithm. If we then compare the energy of this output signal to the input signal, it will be apparent if interference occurs. We will denote this frequency ratio by $h(f, g)$, akin to a transfer function, with frequency f and harmonics gain g as parameters. Note that $h(f, g)$ is not a transfer function in the common sense of the term, because BWE systems are not linear (and sometimes not time-invariant either, in the case of level-dependent clipping). We have

$$h(f, g) = \frac{[\sin(2\pi f t) + g\phi(\sin(2\pi f t))]_{\text{rms}}}{[\sin(2\pi f t)]_{\text{rms}}}. \tag{2.35}$$

The function ϕ indicates the BWE processing. We can also compute $h(f, g)$ for the ratio of rms value of harmonics signal (without adding the input signal) to the rms value of the input signal. By means of example, we use a BWE algorithm with elliptic IIR filters (FIL1 from 20–70 Hz, both flanks of second order; FIL2 from 70–140 Hz, both flanks of second order) and a clipping non-linearity at 50% clipping level. The results are shown in Fig. 2.21. Panels a and c show values of $h(f, g)$, with harmonics-to-input energy ratio in panel a and harmonics + input-to-input energy ratio in panel c. Note that significant gain is obtained in a narrow frequency interval, and that destructive interference occurs in panel c for frequency values slightly below 100 Hz. This would mean that the output energy of the BWE system would actually be smaller than the input energy. In contrast, panels b and d show $h(f, g)$ in the same conditions, but using linear-phase implementations of the same filter characteristics (to be discussed below). Note that gain is obtained over a much broader frequency interval, and that variations of h as functions of f and g are much smoother than in panels a and c. The general features shown in the four panels of Fig. 2.21 are not dependent on the use of elliptic filters or the specific frequency bands used.

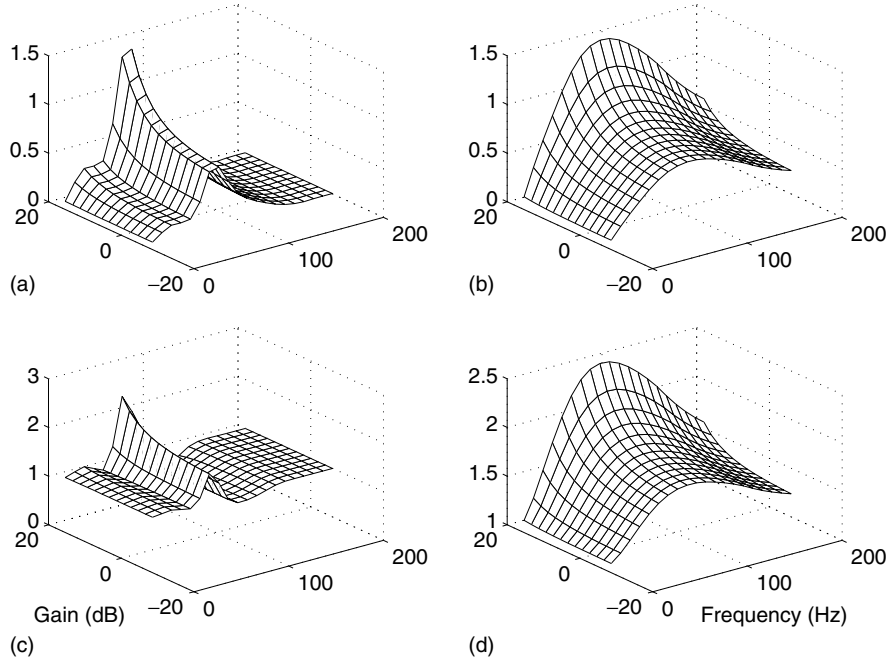

Figure 2.21 (a): h (see Eqn. 2.35) considering BWE with *non-linear* phase filters; input frequency and harmonics gain are variables. The ratio of harmonics-to-input energy is shown as the third dimension. (b): same, but for a *linear* phase filter with the same spectral specifications, and otherwise similar BWE processing. (c): h as in (a), but the value shown is of harmonics+input-to-input energy. (d): same, but for a *linear* phase filter. Note that in both cases the linear-phase filter implementations of the BWE processing give much smoother characteristics

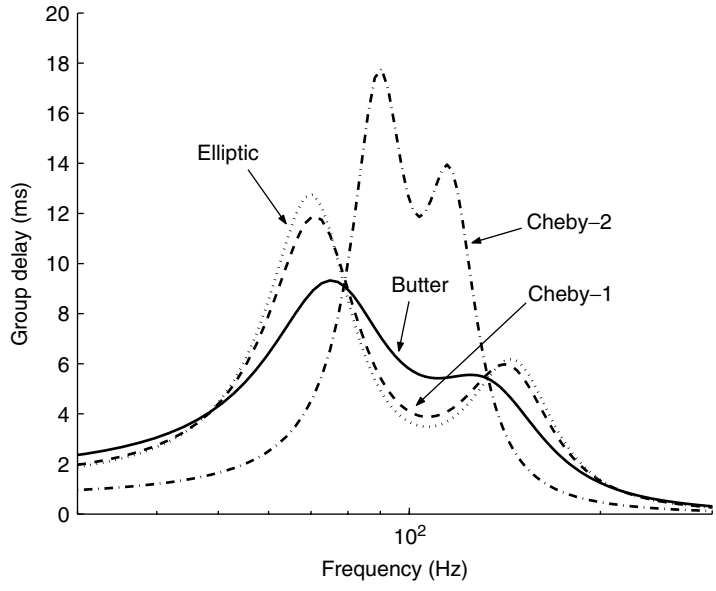

Figure 2.22 Group delay of four IIR filters (Butterworth, Chebyshev type 1, Chebyshev type 2, and elliptic, with bandpass 70–140 Hz, second-order low-pass and high-pass, $f_s = 44.1$ kHz). Every filter shows large variations in group delay

- Frequency-dependent delay of harmonics signal. Because the group delay of a filter is equal to the derivate of the phase response, a non-linear-phase filter will have a non-constant group delay. This means that different frequency components will be delayed by different amounts of time. For low-frequency psychoacoustic BWE applications, this variation in delay can be significant even for successive harmonics. To see this, we computed group delay for four common IIR filter types in Fig. 2.22, where all filters were bandpass filters with cut-off frequencies of 70 and 150 Hz, and both flanks were of second order. The total group delay varies with filter type, but reaches 10 to 15 ms in the passband of the filters, and almost 20 ms for the Chebyshev type 2 filter; this is around 1–2 signal periods for the frequencies of interest. The *variation* in group delay with frequency, being more important, is around 5 ms for most filter types. This may seem a small amount, but results from a study by Zera and Green [304] indicate that such delays may be audible. They investigated the audibility of onset asynchronies of various harmonics of a multi-tone complex, for a variety of onset times, and found that delays of 2 ms are audible. Also note that the BWE algorithm uses two filters, and, thus, group delay variations will be approximately twice the value as indicated in Fig. 2.22: in the order of 10 ms. They also found that thresholds for offset asynchronies are significantly larger than for the onset asynchronies. Thus, we may expect that the delay variations caused by non-linear-phase filters in BWE are audible at the onset of tones with a fast attack. Another effect that may be important is that common onset of individual frequency components is a strong grouping cue (Bregman [38]). Conversely, an asynchrony in onset across frequency may cause segregation of some harmonics

from the bulk of the harmonic complex, causing two (or more) tones to be heard in the BWE-processed signal. We have some anecdotal evidence that this indeed occurs, as musically trained listeners have sometimes commented that bass tones sound delayed with respect to the rest of the music, when listening to low-frequency psychoacoustic BWE processing (with IIR filters).

Low-frequency psychoacoustic BWE sounds better with linear-phase filters than with non-linear-phase filters, and it is plausible that the two reasons discussed above are responsible for this. It is not clear how much either effect (interference and group delay variations) deteriorates the quality by itself, but in any case, there is sufficient motivation to use linear-phase filtering in low-frequency psychoacoustic BWE algorithms. A useful method of implementing linear-phase IIR filters was devised by Powell and Chau [213]. Basically, the method involves a double filtering of the signal; once in 'forward' time and once in 'backward' time (reversing the order of the samples). The backward-time filtering exactly compensates the frequency-dependent delay of the forward-time filtering; the magnitude characteristic imposed on the signal is the square of the filter when applied once, but this can be accounted for in the design of the filter. In a real-time system, finite blocks of data must be used, and for a good block connection, the overlap-add method (OLA) is used (Allen [18]). In this way, the low computational complexity of the IIR filter is maintained, although the OLA requires each sample to be effectively processed four times. Still, this will be much more efficient than a direct FIR implementation.

The linear-phase filtering by filter 1 and filter 2 should be accompanied by a corresponding delay of the main signal in the unprocessed signal branch (Fig. 2.4). If the main signal is high-pass filtered (to remove frequencies below f_1), this is best done with a linear-phase filter as well. The net effect is then a delay of the entire signal.

2.3.4 GAIN OF HARMONICS SIGNAL

The final step before adding the generated harmonics signal back to the main signal is scaling. In Sec. 2.2.3, three points were noted:

- Frequency dependence of loudness: This implies that loudness of equal-level (sound pressure level) harmonics will be higher than that of the fundamental.
- Frequency dependence of 'loudness growth': This implies that equal variations in sound pressure level will lead to smaller loudness variations for the harmonics than for the fundamental.
- 'Tone-lengthening' effect: The combination of increased loudness and better loudspeaker response at frequencies of harmonics (vs fundamental) leads to tones that sound longer when BWE-processed.

In Sec. 2.3.4.2, a method is presented to adaptively vary the gain of the harmonics signal, based on the equal-loudness contours. In Sec. 2.3.4.3, a method that varies the gain according to the total output level of the BWE signal is presented.

2.3.4.1 Fixed Gain

The simplest solution is to simply ignore the loudness variations with frequency, and apply a fixed gain to the harmonics signal. Loudness variations for various frequencies

are generally not huge, and, therefore, a fixed gain can be a suitable solution for a simple low-frequency psychoacoustic BWE system. The gain value will depend on the NLD used and the characteristics of the loudspeaker. From a manufacturer's point of view, maximum bass loudness is usually desired. The gain value can be set as high as is desired, the only limitation being cone excursion and power-handling capacity of the loudspeaker (Sec. 1.3.2).

2.3.4.2 Frequency-adaptive Gain

Another method is due to Gan *et al.* [83] ('Virtual Bass'). They consider that the 'transfer function' of pressure amplitude to loudness (SPL to phones) is similar to that of a downwards expander, with a frequency-dependent expansion ratio E_r. In other words, if the level at the input is lower by x dB, the loudness will decrease by $E_r(f) \times x$ dB. For example, for 40 Hz the expansion ratio is 1.52, while for 100 Hz the expansion ratio is 1.24. A further assumption is that the expansion ratio is nearly independent of absolute loudness in the range 20–80 phones, for frequencies of 110–1000 Hz[2]. With respect to log-frequency, a simple relationship is found to estimate the frequency-dependent expansion rate $\hat{E}_r$, as

$$\hat{E}_r = -0.103 \ln f + 1.71, \quad f > 100 \text{ Hz}, \tag{2.36}$$

where it is acknowledged that for frequencies below 100 Hz this approximation underestimates the actual expansion ratio. Suppose now that harmonics of a fundamental frequency f_0 are generated, and consider the nth harmonic. The 'harmonics expansion ratio' $H\hat{E}_r(f_0, n)$ to be used for this harmonic is then given by

$$H\hat{E}_r(f_0, n) = \frac{\hat{E}_r(f_0)}{\hat{E}_r(nf_0)}, \tag{2.37}$$

and specifies the expansion ratio of the nth harmonic *relative* to the fundamental. For $40 < f_0 < 100$ Hz, the expansion ratio for higher harmonics is fairly independent of f_0 and is given as in Table 2.2. The 'Virtual Bass' algorithm uses a modulation technique (the original reference does not detail this procedure) to generate the individual higher harmonics, and therefore it is possible to apply the appropriate expansion ratio to each harmonic. Observing that the expansion ratio for all the harmonics are roughly 1.10, a simplification could be to apply this expansion ratio to the entire harmonics signal. In a structure as in Fig. 2.4, such an expansion may be achieved by first estimating the envelope $\tilde{f}(t)$ of the BWE harmonics signal f and scaling according to

$$f(t) = (\tilde{f})^{1.1}(t) \times f(t). \tag{2.38}$$

Envelope of a signal can be estimated by low-pass filtering the absolute value of the signal. Figure 2.23 illustrates how such an approach can be incorporated in to the general BWE structure.

[2] It is also implicitly assumed that the equal-loudness-level contours that were measured for pure tones can be used to assess loudness growth of complex tones, which is unlikely to be entirely valid.

Table 2.2 Harmonics expansion ratio as proposed by Gan *et al.* [83], using Eqn. 2.37. These values are valid for the range of fundamental frequencies Gan *et al.* considered (40–100 Hz)

Harmonic n	2	3	4	5
$HEr(f_0, n)$	1.06	1.10	1.13	1.15

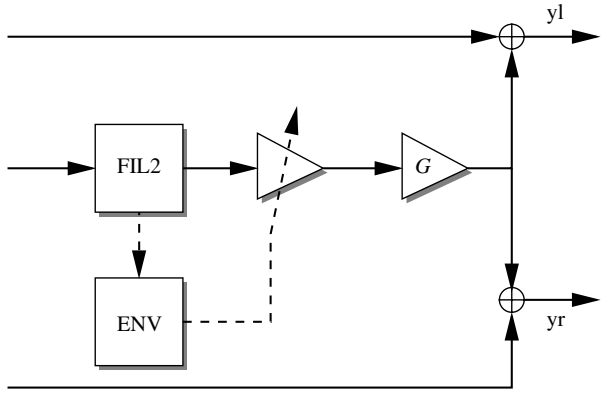

Figure 2.23 Part of a low-frequency psychoacoustic BWE algorithm, compare to Fig. 2.4. The dashed lines indicate a feedforward loop from the output of filter 2, which estimates the envelope of the harmonics signal. The envelope is then expanded by a factor of 1.1, and this value is used to scale the harmonics signal. The final scale factor G is used to bring the entire harmonics signal to an appropriate level

2.3.4.3 Output-level-adaptive Gain

In low-frequency psychoacoustic BWE, the scaled harmonics signal is added back to the main signal (Fig. 2.4) and applied to the loudspeaker terminals. The BWE processing emphasizes frequencies above the loudspeaker cut-off frequency f_l relative to frequencies below f_l, but still care must be taken to avoid overloading the loudspeaker. One could of course implement a fixed scaling of the harmonics signal such that at high reproduction levels distortion is avoided, but this may compromise performance at lower reproduction levels. Especially if a large bass response is desired, a high gain for the harmonics signal should be used at low reproduction levels, as audibility rapidly decreases at low sound pressure levels. Both loudspeaker protection and better matching of human audibility can be achieved by controlling the gain of the harmonics signal in response to the level of the output signal, as in Fig. 2.24. The feedback loop will ensure that at intermediate and low output levels, the gain is at its maximum value, but if the output level is high, the gain is adjusted appropriately. The decay time of the gain control signal must be very small, such that distortion is prevented when a sudden loud sound is reproduced. The gain should increase so slowly as to be inaudible, that is, over a period of several seconds.

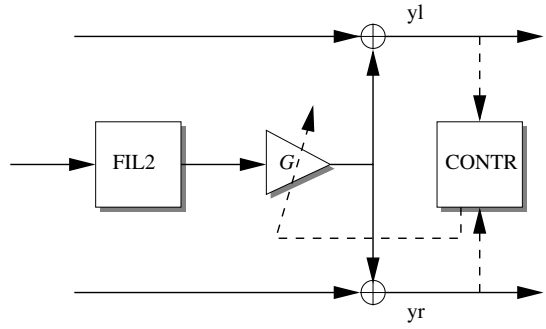

Figure 2.24 Part of a low-frequency psychoacoustic BWE algorithm, compare to Fig. 2.4. The dashed lines indicate the feedback loop from the output to a control unit that modifies the gain of the harmonics signal, such that at low and intermediate output levels the gain is maximum, but gradually decreased as the output level becomes high

2.4 LOW-FREQUENCY PSYCHOACOUSTIC BANDWIDTH EXTENSION WITH FREQUENCY TRACKING

2.4.1 NON-LINEAR DEVICE

The BWE algorithm as outlined previously does not discriminate between tonal and atonal (noise-like) signals. This is because any signal with frequency components within the passband of FIL1 will be processed by the NLD. Occasionally, this can result in annoying artifacts, if noise-like signals are processed, such as may occur in music and speech. This problem could be prevented if BWE processing is only applied to periodic signals. The scheme presented in Fig. 2.25 achieves this goal. One of the attractive features of the algorithm is that it does not explicitly decide if the signal has a tonal or noise characteristic. Rather, BWE processing is implicitly faded out when noise-like signals are present.

The algorithm will be explained following Fig. 2.25. The first step is to estimate the dominant frequency ω_0 in the input signal $x(t)$ (the box labeled FT), where $x(t)$ is obtained by filtering the full-bandwidth input signal, such that only low-frequency components are retained. This frequency estimation is carried out by a recursive frequency-tracking algorithm, which updates at each new sample according to

$$\hat{r}_k = \hat{r}_{k-1} + x_{k-1}\gamma[x_k + x_{k-2} - 2x_{k-1}\hat{r}_{k-1}], \tag{2.39}$$

where $\hat{r}_k = \cos(\omega_0(k)T_s)$, T_s being the sample time, gives the frequency estimate. The frequency-tracking algorithm will be discussed in more detail later. The frequency estimate is used to control a harmonics generator (box labeled HG), which generates a harmonics signal $x_h(t)$ as

$$x_h(t) = \sum_{k=M}^{N} A_k \sin(k\omega_0 t), \tag{2.40}$$

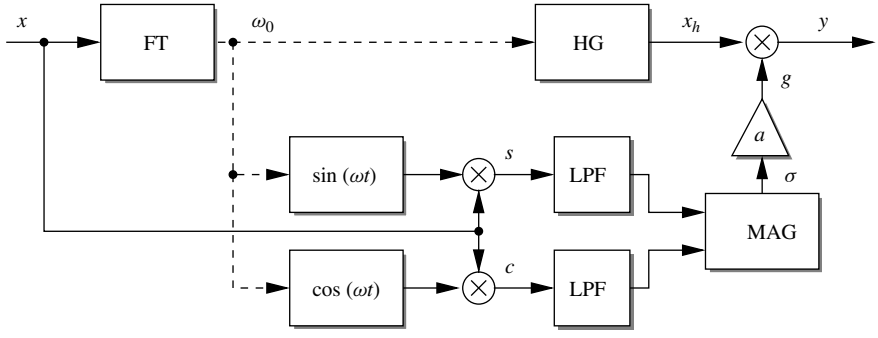

Figure 2.25 Part of a low-frequency psychoacoustic BWE scheme using frequency tracking. FT is a frequency tracker, LPFs are low-pass filters, HG is a harmonics generator, and a is a scaling factor. The output signal $y(t)$ contains harmonics of the strongest frequency component contained in $x(t)$, but only if $x(t)$ is periodic

where M and N equal the minimum and maximum harmonic numbers that are desired. These could be determined by ω_0 or simply be constant, for example, $M = 2$ and $N = 5$. Note that this method of harmonics generation prevents intermodulation distortion, which is an advantage over the NLDs discussed previously in this chapter. Also, the amplitudes of the harmonics can be arbitrarily chosen to produce a desired timbre. It would even be possible to adapt the amplitudes depending on the input signal, although this possibility has not yet been further investigated.

Next, the signal is scaled by a gain control signal $g(t)$, resulting in the output signal $y(t)$. To see how $g(t)$ is determined by $x(t)$, we first assume that $x(t) = A_0 \sin(\omega_0 t + \phi)$. We also assume that the frequency tracker (FT) correctly estimates the frequency of this signal. As shown in Fig. 2.25, the following signals are then generated

$$s(t) = A_0 \sin(\omega_0 t + \phi) \sin(\omega_0 t), \tag{2.41}$$

$$c(t) = A_0 \sin(\omega_0 t + \phi) \cos(\omega_0 t), \tag{2.42}$$

which can also be written as

$$s(t) = \frac{A_0}{2}[\cos\phi - \cos(2\,\omega_0 t + \phi)], \tag{2.43}$$

$$c(t) = \frac{A_0}{2}[\sin\phi + \sin(2\,\omega_0 t + \phi)]. \tag{2.44}$$

After averaging (by low-pass filtering), we get

$$\overline{s}(t) = \frac{A_0}{2}\cos\phi, \tag{2.45}$$

$$\overline{c}(t) = \frac{A_0}{2}\sin\phi. \tag{2.46}$$

where $\overline{s}(t)$ and $\overline{c}(t)$ are slowly time-varying signals. Taking the square root of the sum of squares, this becomes

$$\sigma(t) = \sqrt{\overline{s}^2(t) + \overline{c}^2(t)} = \frac{A_0}{2}. \tag{2.47}$$

The control signal g is then obtained as

$$g(t) = a\sigma(t) \tag{2.48}$$

if, for example $a = 2$, then $g(t) = A_0$. Thus, we see that for a sinusoidal input signal $y(t) = x_h(t)$, that is, the output signal has maximum amplitude. Now if $x(t)$ is a noise signal, the averaged (low-pass) $s(t)$ and $c(t)$ will tend to zero if the averaging time is sufficiently long. For intermediate cases, the gain control signal $g(t)$ will be scaled $x_h(t)$ down to an appropriately lower amplitude. It thus appears that the gain control signal varies between 0 (noise inputs) and 1 (sinusoidal input), with a gradual transition between these two extremes, depending on the periodicity of the input signal. In practice, $x(t)$ may contain multiple sinusoids and/or a sinusoid in the presence of noise. Section 2.4.2.3 shows how the initial frequency estimation is affected by such signals.

In conclusion, this alternate NLD generates a harmonics signal without intermodulation distortion, but only for periodic input signals. For noisy input signals, the output tends to zero.

2.4.2 FREQUENCY TRACKING

Here, we will elaborate on the frequency-tracking algorithm utilized in the NLD of the previous section. There is a vast literature regarding frequency tracking, owing to the many applications in, for example, astronomy, acoustics, and communications; see e.g. Quinn and Hannan [214] and Tichavsky and Nehorai [271] for a comparative study of four adaptive frequency trackers. Recently, an algorithm was devised for rapid power-line frequency monitoring (Adelson [15]), on the basis of a number of formulas presented in Adelson [14]. Most of the algorithms presented in the book of Quinn and Hannan are complex and not very suitable for real-time implementation, while for BWE algorithms, we (as usual) strive for maximum computational efficiency, by avoiding divisions, trigonometric operations such as FFTs – which also necessitate the use of buffers – and the like.

Here, we will develop an efficient frequency-tracking procedure, which uses only a few arithmetic operations, and is insensitive to the initial state of the algorithm parameters. We also analyse the convergence behaviour of the algorithm for stationary input signals, and the dynamic behaviour if there is a transition to another stationary state, the latter being considered important to assess the tracking abilities for realistic signals. The following derivations and analyses were also published in Aarts [5], and in analogous form previously for another application (cross-correlation tracking) in Aarts *et al.* [8]. In slightly modified form, the algorithm can also be used to track amplitude instead of frequency.

We shall show in Sec. 2.4.2.1 that the recursion

$$\hat{r}_k = \hat{r}_{k-1} + x_{k-1}\gamma[x_k + x_{k-2} - 2x_{k-1}\hat{r}_{k-1}], \tag{2.49}$$

estimates to a good approximation the frequency of a signal given by

$$r_k = \cos(\omega_0(k)T_s), \tag{2.50}$$

where $\omega_0(k)$ is the frequency of the input signal x to be determined, k is the time index, and $f_s = 1/T_s$ is the sampling frequency. The parameter γ determines the convergence speed, and hence determines the tracking behaviour of $\hat{r}$, but not the actual value of $\lim_{k\to\infty} \hat{r}_k$ in the stationary case. Equation 2.49 is the basis for our approach of recursively tracking the frequency. In Sec. 2.4.2.2, we shall analyse the solution of Eqn. 2.49, starting from an initial value $\hat{r}_0$ at $k = 0$, when $\gamma \downarrow 0$, and we shall indicate conditions under which

$$\lim_{\gamma\downarrow 0}[\lim_{k\to\infty} \hat{r}_k] = \cos(\omega_0 T_s). \tag{2.51}$$

The analysis is similar to that of an algorithm (Aarts *et al.* [8]) to track correlation coefficients, and can be facilitated considerably by switching from difference equations, as in Eqn. 2.49, to differential equations.

In Sec. 2.4.2.3, we consider the case of a sinusoidal input signal x, and we compute explicitly the left-hand side of Eqn. 2.51 for the solution of Eqn. 2.49. It turns out that the recursion Eqn. 2.49 yields the correct value r for the left-hand side of Eqn. 2.51.

2.4.2.1 Derivation of Tracking Formulas

Here, we consider r as defined in Eqn. 2.50, and we show that r satisfies to a good approximation (when γ is small) the recursion in Eqn. 2.49.

We start with Adelson's [15] Eqn. 1

$$r = \frac{\sum_{j=1}^{n-1} x_j(x_{j-1} + x_{j+1})}{2\sum_{j=1}^{n-1} x_j^2}. \tag{2.52}$$

In order to make this formula suitable for tracking purposes, it is modified into

$$r_k = \frac{\sum_{j=1}^{n-1} x_{k-j}(x_{k-j-1} + x_{k-j+1})}{2\sum_{j=1}^{n-1} x_{k-j}^2}. \tag{2.53}$$

Now r_k depends on $n-1$ samples from the past, and the current sample x_k. However, it is not optimal for tracking purposes, since it suffers from the fact that it requires many operations and may lead to numerical difficulties in the case of a small denominator in Eqn. 2.53. Therefore, a second modification is made by using – instead of a rectangular window and an averaging over $2n$ $x_i x_{i+1}$ products – an exponential window. In order to minimize the number of operations, we select $n = 2$. Now, we define

$$r_k = \frac{S_n}{S_d}, \tag{2.54}$$

where

$$S_n(k) = \sum_{l=0}^{\infty} c\, \mathrm{e}^{-\eta l} x_{k-l-1}(x_{k-l} + x_{k-l-2}), \tag{2.55}$$

$$S_d(k) = \sum_{l=0}^{\infty} 2c\, \mathrm{e}^{-\eta l} x_{k-l-1}^2, \tag{2.56}$$

$$c = 1 - \mathrm{e}^{-\eta}, \tag{2.57}$$

with η is a small but positive number that should be adjusted to the particular circumstances for which tracking of the frequency is required. We now show that r of Eqs. 2.54–2.57 satisfies to a good approximation the recursion in Eqn. 2.49. To this end, we note that

$$S_n(k) = \mathrm{e}^{-\eta} S_n(k-1) + c x_{k-1}(x_k + x_{k-2}), \tag{2.58}$$

and

$$S_d(k) = \mathrm{e}^{-\eta} S_d(k-1) + 2c x_{k-1}^2. \tag{2.59}$$

Hence, from the definition in Eqn. 2.54,

$$r(k) = \frac{S_n(k-1) + c\ \mathrm{e}^{\eta} x_{k-1}(x_k + x_{k-2})}{S_d(k-1) + 2c\ \mathrm{e}^{\eta} x_{k-1}^2}. \tag{2.60}$$

Since we consider small values of η, $c = 1 - \mathrm{e}^{-\eta}$ is small as well. Expanding the right-hand side of Eqn. 2.60 in powers of c and retaining only the constant and the linear term, we get after some calculations

$$r(k) = r(k-1) + \frac{c\ \mathrm{e}^{\eta}}{S_d(k-1)} x_{k-1}[x_k + x_{k-2} - 2r(k-1)x_{k-1}] + O(c^2). \tag{2.61}$$

Then, deleting the $O(c^2)$ term, we obtain the recursion in Eqn. 2.49 when we identify

$$x_{\mathrm{rms}}^2 = S_d(k), \tag{2.62}$$

for a sufficiently large k, and

$$\gamma = \frac{c\ \mathrm{e}^{\eta}}{x_{\mathrm{rms}}^2}, \tag{2.63}$$

which is a constant for a stationary signal $x(t)$.

We observe at this point that we have obtained the recursion in Eqn. 2.49 by applying certain approximations (as in Eqn. 2.62) and neglecting higher-order terms. Therefore, it is not immediately obvious that the actual r of Eqn. 2.50 and the solution of $\hat{r}$ of the recursion in Eqn. 2.49 have the same value, in particular for large k. However, next we shall show that $\hat{r}$ shares some important properties with the veridical r.

2.4.2.2 Analysis of the Solution of the Basic Recursion

Now we consider the basic recursion in Eqn. 2.49, and we analyse its solution $\hat{r}(k)$, given an initial value $\hat{r}_0$ at $k = 0$, when $\gamma \downarrow 0$. It is convenient to introduce the new variables

$$\beta_k = 2x_{k-1}^2, \tag{2.64}$$

and

$$\delta_k = x_{k-1}(x_k + x_{k-2}), \tag{2.65}$$

to remain compatible with the notation in Aarts *et al.* [8] and Aarts [5]. Thus, we shall consider the recursion in Eqn. 2.49, which we rewrite as

$$\hat{r}(k) = (1 - \gamma\beta_k)\hat{r}(k-1) + \gamma\delta_k \tag{2.66}$$

for $k = 1, 2, \ldots$ with γ a small positive number and δ_k, β_k bounded sequences with $0 \leq \beta_k \leq 1$.

In Aarts *et al.* [8], it was shown how to obtain the limiting behaviour of $\hat{r}(k)$ as $k \to \infty$ when $\gamma > 0$ is small. This was done under an assumption (slightly stronger than required) that the mean values (denoted by $M[.]$)

$$\begin{aligned} b_0(\gamma) &= M\left[\frac{-1}{\gamma}\log(1-\gamma\beta_k)\right] \\ &= \lim_{K\to\infty}\frac{1}{K}\sum_{l=1}^{K}\frac{-1}{\gamma}\log(1-\gamma\beta_l), \\ \mathrm{d}_0 &= M[\delta_k] = \lim_{K\to\infty}\frac{1}{K}\sum_{l=1}^{K}\delta_l \end{aligned} \tag{2.67}$$

for the discrete-time case and

$$\begin{aligned} b_0 &= M[\beta(t)] = \lim_{T\to\infty}\frac{1}{T}\int_0^T \beta(s)\,\mathrm{d}s, \\ \mathrm{d}_0 &= M[\delta(t)] = \lim_{T\to\infty}\frac{1}{T}\int_0^T \delta(s)\,\mathrm{d}s, \end{aligned} \tag{2.68}$$

for the corresponding continuous-time case, exist.

Since $b_0(\gamma) \to b_0$ as $\gamma \downarrow 0$, it was shown that

$$\lim_{\gamma\downarrow 0}\left[\lim_{k\to\infty}\hat{r}(k)\right] = \frac{M[\delta_k]}{M[\beta_k]} = \frac{\mathrm{d}_0}{b_0} = \frac{M[\delta(t)]}{M[\beta(t)]}, \tag{2.69}$$

and for any number $b < b_0(\gamma)$

$$\hat{r}(k) = \frac{\mathrm{d}_0}{b_0(\gamma T_\mathrm{s})} + O(\mathrm{e}^{-\gamma b k T_\mathrm{s}}), \quad k \geq 0. \tag{2.70}$$

This shows that the time constant τ, that is, the time for the exponential term to drop to e^{-1} of its original value, for the tracking behaviour is given by

$$\tau = \frac{T_s}{\gamma b_0(\gamma T_s)}. \tag{2.71}$$

We finally observe that $b_0(\gamma) \to b_0$ as $\gamma \downarrow 0$. In the next section, we shall work this out for sinusoidal input signals.

2.4.2.3 Sinusoidal Input Signals

In this section, we test the algorithms derived in Sec. 2.4.2.1, and analysed in Sec. 2.4.2.2, with respect to their steady-state behaviour, for sinusoidal input signals. Hence we take

$$x_k = A_0 \sin(\omega_0 k T_s + \phi), \tag{2.72}$$

with arbitrary A_0 and ϕ. Calculating δ and β with Eqs. 2.65–2.64, and using Eqn. 2.69, it is easy to find

$$\lim_{\gamma \downarrow 0} \left[\lim_{k \to \infty} \hat{r}(k) \right] = \cos \omega_0 T_s; \tag{2.73}$$

compare with Eqn. 2.50. This limit obviously does not depend on A_0, nor on ϕ. If Eqn. 2.72 and $r_{k-1} = \cos \omega_0 T_s$ are substituted into Eqn. 2.49, then we get $r_k = r_{k-1}$, independent of γ, indicating that r has converged to a constant value. Using Eqn. 2.71 and $b_0 = A_0^2$, it appears that the time constant τ_d of the tracking behaviour is equal to

$$\tau_d = T_s/(\gamma A_0^2). \tag{2.74}$$

Now consider the case that the signal x consists of two sinusoids (with unequal frequencies), where the latter can represent a disturbance signal or as harmonic distortion of the first sinusoid. Thus,

$$x_k = A_0 \sin(\omega_0 k) + A_1 \sin(\omega_1 k), \tag{2.75}$$

and by using Eqns. 2.64, 2.65, and 2.69 we get

$$\lim_{\gamma \downarrow 0} \left[\lim_{k \to \infty} \hat{r}(k) \right] = \frac{A_0^2 \cos \omega_0 T_s + A_1^2 \cos \omega_1 T_s}{A_0^2 + A_1^2}. \tag{2.76}$$

Alternatively, consider the case that the signal x consists of a sinusoid with additional noise $n(k)$ (with autocorrelation function $R_n(k)$). Then

$$x_k = A_0 \sin(\omega_0 k) + n(k), \tag{2.77}$$

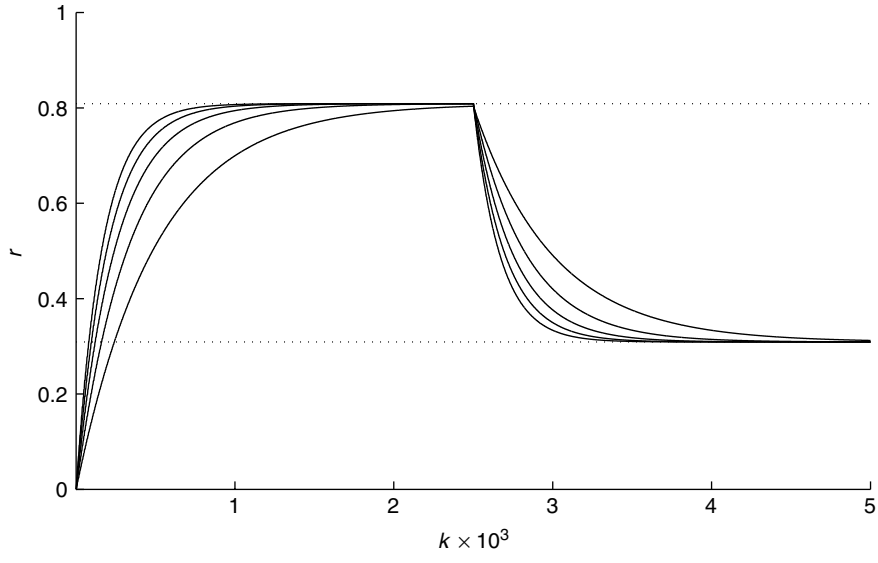

Figure 2.26 Step response of Eqn. 2.49 for a sinusoidal input signal with amplitude $A_0 = 1$, $\hat{r}_0 = 0$, and $\gamma = 2 \cdot 10^{-3}$–$6 \cdot 10^{-3}$ in three increments, and making a step from $\omega_0 T_s = \pi/5$ ($r = 0.81$) to $\omega_0 T_s = 2\pi/5$ ($r = 0.31$). The dotted lines are the final values given by Eqn. 2.73

and we get

$$\lim_{\gamma \downarrow 0}[\lim_{k \to \infty} \hat{r}(k)] = \frac{{A_0}^2 \cos \omega_0 T_s + 2R_n(T_s)}{{A_0}^2 + 2R_n(0)}. \tag{2.78}$$

Equation 2.78 shows that if $R_n(T_s)$ and $R_n(0)$ are known, or can be estimated, the estimation of $\hat{r}$ can be easily improved.

To demonstrate the tracking behaviour of Eqn. 2.49, in Fig. 2.26 the step response is plotted for a sinusoidal input signal, making a change in frequency, for various values of γ. It appears that the time constants correspond well with the values predicted by Eqn. 2.74. The values of γ used in Fig. 2.26 are just for illustration purposes, and in practice they could be much larger. To obtain stability, we need $|1 - \beta_k \gamma| < 1$. Practical values for sinusoidal input signals are $0 < A_0 \gamma < 0.5$.

Using the same procedure as for tracking frequency, we can track the amplitude A_0 of the input signal as well. To that end, β and δ are modified into

$$\beta'_k = 1 - r_k^2, \tag{2.79}$$

and

$$\delta'_k = x_{k-1}^2 - x_k x_{k-2}. \tag{2.80}$$

Using Eqn. 2.66, we get

$$A(k) = (1 - \gamma\beta'_k)A(k-1) + \gamma\delta'_k, \tag{2.81}$$

and, finally, we get $A_0 = \sqrt{A(k)}$.

2.5 SUBJECTIVE PERFORMANCE OF LOW-FREQUENCY PSYCHOACOUSTIC BANDWIDTH EXTENSION ALGORITHMS

There is little published data on the subjective quality of low-frequency psychoacoustic BWE systems, although from informal listening it is known that well-designed systems can yield good-quality sound. Here, we shall discuss the results of a formal listening test, parts of which was also published in Larsen and Aarts [156]. First we present the results from two other studies.

2.5.1 'VIRTUAL BASS'

Performance of 'Virtual Bass' system In Gan *et al.* [83], a low-frequency psychoacoustic BWE algorithm is presented, called 'Virtual Bass'. They also report results from human subject testing, which was performed in the following manner. Ten naive subjects were used in an age group of 24 to 35 years old. Three test signals were employed: a sequence of gunshots, a bass guitar soundtrack, and classical music. Three loudspeakers were used: a high-quality two-way monitor (5" cone; 68–20,000 Hz; 30 W), a multi-media speaker (3" cone; 110–15,000 Hz; 6 W), and a flat-panel speaker (150–20,000 Hz; 3.6 W). The authors did not mention whether the subjects could see the speaker that was being used; if so, this might have biased the results. For each signal that was tested, subjects first heard the unprocessed signal as a reference (presumably using the monitor speaker), followed by the processed signal using either the 'Virtual Bass' algorithm or a commercially available system (unspecified). The subjects were then asked to grade both bass quality and signal impairment on a five-grade scale as in Table 2.3. Bass quality was judged as good for both the 'Virtual Bass' and the commercial bass system, with a slight advantage for the 'Virtual Bass' system. The impairment ratings were a bit lower for the 'Virtual Bass' than for the commercial bass system, however. The average impairment was 3.67 for 'Virtual Bass' and 4 for the commercial bass system. The impairment of the 'Virtual

Table 2.3 Five-grade quality and impairment scale used by Gan *et al.* [83]

Grade	Quality	Impairment
5	Excellent	Imperceptible
4	Good	Perceptible, not annoying
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very annoying

Bass' signals were reported as a humming pitch, attributed to the method of harmonics generation (modulating function). These artefacts were more audible in the better-quality monitor speakers than in the other two reproduction systems. This is favourable considering that low-frequency psychoacoustic BWE will typically not be used in good-quality speakers with extended low-frequency response.

Cross-talk cancellation application Tan *et al.* [261] use the 'Virtual Bass' algorithm of Gan *et al.* [83] in a cross-talk cancellation method. In cross-talk cancellation, the objective is to eliminate sound from the left loudspeaker reaching the right ear and sound from the right loudspeaker reaching the left ear. This is important for virtual audio applications using loudspeakers. The use of headphones in virtual audio applications would not require the use of cross-talk cancellation.

Tan *et al.* argue that the low interaural level difference (ILD) for low-frequency sounds makes cross-talk cancellation difficult, because it requires inversion of an ill-conditioned matrix. Even if cross-talk cancellation is possible, the required boosting of low frequencies will cause problems in the loudspeakers because of large cone excursion and power-handling capacity. They propose to circumvent these problems by using a low-frequency psychoacoustic BWE system to replace very low frequencies by higher harmonics, for which it is easier to cancel the cross-talk. A subjective test was performed with two different signals, at two different phantom source azimuths (45 and 90° relative to straight ahead); ten subjects were used. The quality was determined by ranking on a five-grade scale (different from the one used in Table 2.3). The phantom sources at 90° received slightly higher scores (about 0.4–0.5 points). We performed a t-test for the two samples (different azimuths) of each signal and found no significant difference between the means at the 10% significance level, however. In fact, the reference condition, which consisted of a cross-talk system without the 'Virtual Bass' processing, did not differ significantly at the 5% significance level from any of the conditions tested with the 'Virtual Bass' system (only one of the signals at 90° azimuth had a significantly different mean at the 10% level), as was determined by t-tests for each condition.

2.5.2 'ULTRA BASS'

In Larsen and Aarts [156], a discussion was presented on the results of a listening test that was conducted to assess the subjective quality of two low-frequency psychoacoustic BWE systems ('Ultra Bass'). Here, some of this discussion is repeated, together with some new analysis.

The experiment had three objectives:

1. To rank order preference for unprocessed, linearly amplified (bass only), and BWE-processed musical signals.
2. To evaluate if preferences vary per subject.
3. To evaluate if preferences vary per repertoire.

Algorithms tested In the following text, we will refer to four different algorithms, which are as follows:

1. Unprocessed signal, which was included as reference against which the processed signals would be compared.
2. Linear amplification, which is considered to yield 'baseline' performance for bass enhancement. The quality of the two BWE systems should at least match but preferably exceed the quality of the linear system.
3. Low-frequency psychoacoustic BWE system with rectifier as NLD.
4. Low-frequency psychoacoustic BWE system with integrator as NLD. This and the previous algorithm were chosen because from informal listening it was observed that the quality of the processed signals is quite different for the two cases (which is not surprising given the analysis in Secs. 2.3.2.2 and 2.3.2.3).

The linear amplification was done with commercial sound-editing software, using a graphic EQ in 1/2-octave bands. The amplification was 6 dB (44 Hz), 9 dB (62.5 Hz), 9 dB (88 Hz), and 6 dB (125 Hz); these values were chosen to give maximum bass boost without creating audible distortion at the reproduction level used in the experiment. The processing was identical for the two BWE systems, except for the implementation of the NLD. Filter 1 was implemented as a second-order Chebyshev-type I IIR filter; passband ripple was 1 dB, and the passband was 20–70 Hz. Filter 2 was implemented as a third-order elliptic IIR filter (also non-linear phase), passband ripple of 3 dB, stopband attenuation of 30 dB and passband of 70–140 Hz. The gain value for the harmonics signal was fixed at 15 dB, for both BWE systems. The signal in the main path was not processed (no high-pass filter, no delay). The implementation of the BWE systems as used in the test is now known to be suboptimal; particularly, the use of non-linear-phase IIR filters would be avoided in favour of using linear-phase filters.

Music selection, signal generation, and reproduction Music was selected according to genre and an a priori evaluation of subjective quality. Genres were pop and rock, and subjective quality criteria were that the bass content of the signals should be 'difficult' to reproduce well on a small loudspeaker system. This approach was taken so that the obtained results would indicate performance of some of the most demanding signals. Excerpts ($\approx$10 s duration) from the following four tracks were used:

1. 'Bad' by Michael Jackson. This track contains a typical pop bass line, which was known to give good subjective performance after BWE processing. It was included to contrast the other, more demanding, signals.
2. 'My Father's Eyes' by Eric Clapton. A very deep and strong bass line accompanies the music on this track, which may sound too imposing if the reproduced bass is not well balanced.
3. 'Hotel California' by The Eagles (live version). The excerpt was from the start of the track, which consists of a bass drum only (and some audience noise). This makes it easier to focus on the bass quality. The difficulty in reproduction lies in the low frequency and very fast attack of the drum. Also, the decay is very gradual and should not sound unnatural.
4. 'Twist and Shout' by Salt n' Peppa. In this track, the bass follows a tight beat, the main difficulty being to preserve the tight temporal envelope.

These four signals were processed by each of the three algorithms as described previously, and the test included the unprocessed signals as well. Prior to processing, all four test signals were scaled to obtain approximately equal loudness.

Reproduction was on a commercially available medium-sized Hi-Fi system. The low cut-off frequency was about 140 Hz. Listeners were seated at a distance of about 1 m in the median plane between two loudspeakers. Reproduction was at a comfortable listening level, and was fixed prior to the start of the experiment, being the same for all subjects.

Human subjects Fifteen unpaid volunteers (eleven male, four female) participated in the experiment. The age range was 25–30 years old. All had self-reported good hearing, varying degrees of experience in listening tests, and varying degrees of interest in music. Subjects were asked to indicate their preferred genres of music, which were pop and rock.

Experimental procedure A direct ranking of the various processed signals would be difficult, and the paired comparison paradigm was chosen because it is known to yield good results when used to compare several perceptually close signals (David [56]). Thus, a pair of signals (same repertoire, different processing) would be presented, and listeners were instructed to choose the version with the best bass quality. Although this allows the possibility that different listeners use different criteria in their selection, this was done to obtain general preference ratings; furthermore, one of the objectives was to find out if there would be differences in preferences among subjects. Instructing listeners to choose on the basis of the 'best bass quality' should meet both these objectives. Subjects could listen to the pair of the signals as long as was required to make a selection. Because each repertoire had four versions, six pairs were presented to the listener. After the six presentations, the next repertoire was used, until all four repertoire were completed. There were no repetitions, as in most cases the signal pair presented on any trial differed enough to be distinguishable, and we did not expect listeners to change their preference during the course of the experiment. Some listeners had prior exposure to signals processed by the BWE system.

The responses were recorded in preference matrices $\mathcal{P}$ (one for each repertoire); $\mathcal{P}$ is an anti-symmetric 4×4 matrix with elements $p_{ij} = \{0, 1\}$, a 1 indicating that the column element i is preferred over the row element j, and vice versa. The diagonal elements are not used. The preference matrix can be summarized in a score vector $\mathbf{s}$, which is a column vector, the elements of which are the sum of the rows of $\mathcal{P}$. The ranking of the different algorithms then follows directly from $\mathbf{s}$.

Results Table 2.4 gives the score vectors for all subjects, for each repertoire (numbered as indicated previously). Also, the number of circular triads is shown (CT) (Levelt *et al.* [159]). A circular triad occurs if, for example, version 2 is preferred over 3, 3 is preferred over 4, and 4 is preferred over 2; this indicates an inconsistency in the subject responses. A high value for CT probably indicates that the task is confusing, and would necessitate caution in interpretation of the results. As Table 2.4 shows, for most subjects CT is zero or one, which is normal.

For a preliminary analysis of the results, we plotted the normalized score of each algorithm, averaged over the four repertoire, for each subject; see Fig. 2.27. The normalized

Table 2.4 Score vectors obtained in the listening test. The fifteen subjects are labeled A–O. CT indicates the number of circular triads

	v	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	Total
Bad	1	0	0	0	0	3	0	2	0	0	1	0	0	1	0	0	7
	2	2	2	3	1	2	1	3	1	2	2	2	2	3	1	1	28
	3	1	2	2	2	1	3	0	3	2	2	2	3	0	3	3	26
	4	3	2	1	3	0	2	1	2	2	1	2	1	2	2	2	26
Eyes	1	1	0	3	1	3	0	3	0	0	0	0	0	2	1	2	15
	2	1	2	1	2	1	1	2	1	3	2	1	2	2	2	3	26
	3	1	2	0	1	0	2	1	3	1	2	2	2	0	0	1	18
	4	3	2	2	2	2	3	0	2	2	2	3	2	2	3	0	30
Hotel	1	0	1	2	1	2	0	1	0	1	0	2	1	1	1	1	14
	2	1	2	3	3	3	1	2	1	2	2	0	2	2	3	3	30
	3	2	3	1	2	1	3	3	3	3	2	3	3	3	2	1	35
	4	3	0	0	0	0	2	0	2	0	2	1	0	0	0	1	11
Twist	1	0	0	2	0	2	0	2	0	0	2	1	0	0	0	0	9
	2	2	2	3	1	3	1	3	1	1	3	1	2	1	2	1	27
	3	2	1	0	2	0	3	1	3	3	1	1	3	3	2	3	27
	4	2	3	1	3	1	2	0	2	2	0	3	1	2	2	2	26
Total	1	1	1	7	2	10	0	8	0	1	3	3	1	4	2	3	45
	2	6	8	10	7	9	4	10	4	8	9	4	8	8	8	8	111
	3	6	8	3	7	2	11	5	12	9	7	8	11	6	7	8	106
	4	11	7	4	8	3	9	1	8	6	5	9	4	6	7	5	93
CT		2	2	0	1	0	0	0	0	1	3	2	1	1	1	1	–

score was obtained as the sum of corresponding elements of the subject's four score vectors, for example, element 1 for the unprocessed version of each signal, divided by 12. In this way, the normalized score varies between 0 and 1. The subjects have been divided into two groups, A (subjects 3, 5, and 7) and B (all others); later on, we will motivate this division. For now we merely notice that, for each algorithm, the mean score assigned by groups A and B is different. Group A rates the unprocessed and linearly amplified signals higher than both BWE-processed signals; for group B, all three processing algorithms have scored approximately the same, while the unprocessed version gets a low score. Table 2.5 gives the mean normalized score for each algorithm, for both groups as well as overall (mean over groups). On the basis of Table 2.5, for group A the rank order of the algorithms would be: (1) linear amplification, (2) no processing, (3) BWE with rectifier, and (4) BWE with integrator. For group B, the rank order would be: (1) BWE with rectifier, (2) BWE with integrator, (3) linear amplification, and (4) no processing.

Discussion Division of subjects in two groups (A and B) can be made plausible by visualizing the subjects' responses with multidimensional scaling (MDS), see App. A. Fig. 2.28 shows the resultant two-dimensional mapping obtained using as proximities the Euclidian distances between score vectors (which are four dimensional). Subjects have been divided into five clusters, S0–S4. The previously mentioned group A corresponds to

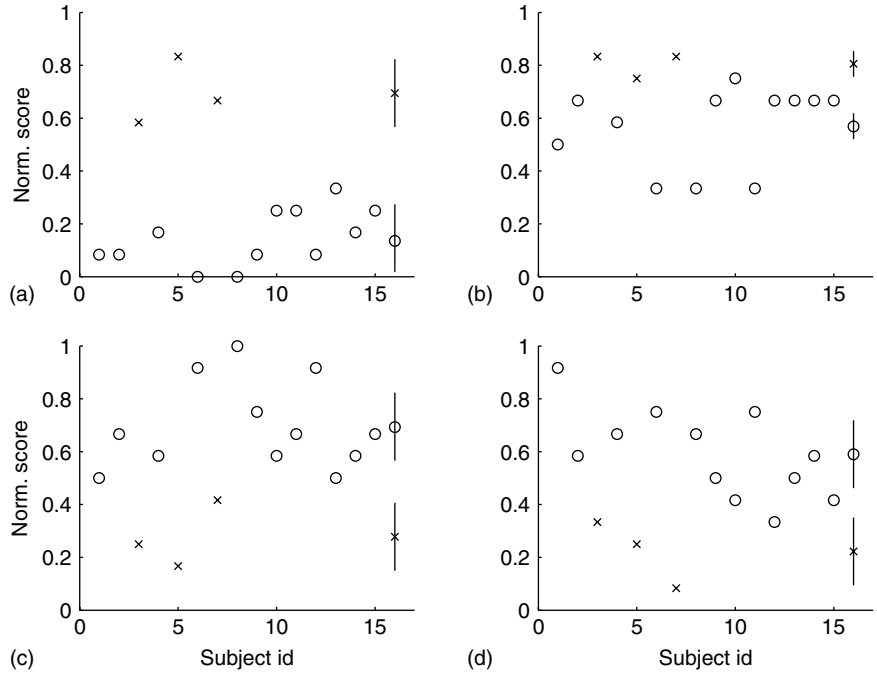

Figure 2.27 Normalized scores for each algorithm (a: unprocessed, b: linear gain, c: BWE with rectifier, and d: BWE with integrator), for each subject. The subjects were divided into two groups and plotted with different symbols (o, x). At the right-hand side of each graph, the means and standard deviation of each subject group is indicated

Table 2.5 Normalized score vectors for the four algorithms, which may be used for subjective quality ranking. Groups A and B are different subject groups, as defined in the text. The 'overall' numbers are weighted averages of the two group values

	A	B	Overall
1	0.69	0.15	0.26
2	0.81	0.57	0.62
3	0.28	0.69	0.61
4	0.22	0.59	0.52

cluster S2 of Fig. 2.28; group B corresponds to the other four clusters. The division into groups A and B is now obvious, as Fig. 2.28 shows that the MDS maps subjects in group A (cluster S2) far away from all the other subjects. By inspecting the subject's individual responses from Table 2.4, we can interpret the MDS dimensions. The horizontal dimension seems to indicate preference for BWE processing, with higher preference towards the right-hand side. The vertical dimension seems to indicate preference for NLD type, with integrator towards the top and rectifier towards the bottom. In Larsen and Aarts [156],

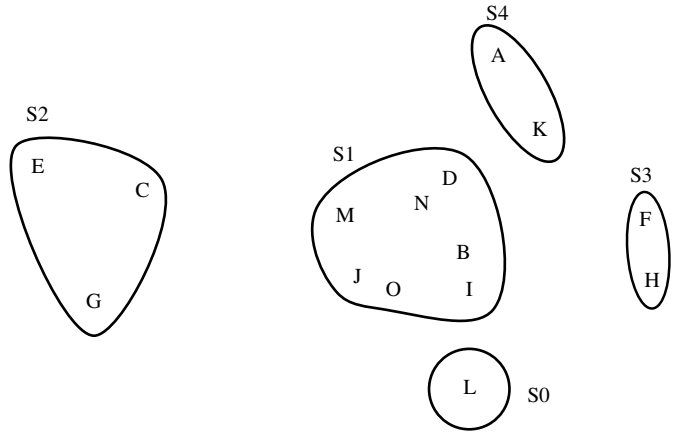

Figure 2.28 Two-dimensional scaling of subject (A–O) preferences of a subjective comparison between various bass enhancement systems. Interpretation of the two dimensions is made in the text. Subjects were grouped in five clusters. Group A mentioned in the text, and in Fig. 2.27 and Table 2.5 corresponds to cluster S2 here. Group B corresponds to the other four clusters. From Larsen and Aarts [156]

results were further analysed with biplots (Gabriel [81]), which showed that the two BWE systems were judged most similar, and the linear system versus BWE with integrator were judged most dissimilar. Conclusion It appears that there is no consistent judgement from the whole subject group regarding preference for a particular processing type. Out of 15 subjects, 3 preferred a linear bass enhancement system, while the other 12 preferred low-frequency psychoacoustic BWE processing. Within these 12 subjects, there was no clear preference for a rectifier or integrator as NLD, although the rectifier did receive a somewhat higher average appreciation. On the basis of this experiment and the response of all subjects taken as a whole, the main conclusion is that low-frequency psychoacoustic BWE can perform at least as well as linear systems. More recent developments in low-frequency psychoacoustic BWE methods, such as adaptive clipping (Sec. 2.3.2.4) or frequency tracking (Sec. 2.4) have shown superior performance in informal evaluations and may show a more conclusive benefit to linear bass enhancement systems in formal listening tests.

2.6 SPECTRAL CHARACTERISTICS OF NON-LINEAR DEVICES

In Sec. 2.3.2, intermodulation characteristics of non-linear devices were analysed. For the rectifying and integrating NLDs, expressions were given for the Fourier series coefficients of processed signals, given the Fourier series coefficients of the input signals. Sections 2.6.1 and 2.6.2 will present the full derivation of these expressions, originally published in Larsen and Aarts [156], and were largely due to A.J.E.M. Janssen. Discrete-time expressions are given in Sec. 2.6.3, and in Sec. 2.6.4 the Fourier series coefficients of a clipped sinusoid are given.

Consider a real periodic signal $f(t)$ of period $T_0 = 1/f_0$ and assume that

- $$f(t_i) = 0, \quad i = -1, 0, \ldots, N, \tag{2.82}$$
$$f(t) \neq 0, \quad t \neq t_i, \tag{2.83}$$
thereby defining the zeros of $f(t)$ during a one-period interval; t_{-1} is defined as the beginning of the period, and t_N as the end of the period (which is identical to t_{-1} of the next period). There are $N \geq 1$ zero crossing in between t_{-1} and t_N, and $t_N - t_{-1} = T_0$. We use the shorthand notation $\mathbf{t} = (t_{-1}, \ldots, t_N)^T$. Figure 2.29 illustrates the above notation.
- $f(t)$ changes sign at every t_i, which implies that N is odd.
- $f'(t_{-1}) > 0$.
- $f(t)$ is sufficiently smooth such that its Fourier coefficients a_n decay at a rate of at least $1/n^2$. This will be satisfied if, for instance, $f(t)$ is at least twice continuously differentiable.

We have for $f(t)$ the Fourier series representation

$$f(t) = \sum_{n=-\infty}^{\infty} a_n e^{i2\pi f_0 n t}, \tag{2.84}$$

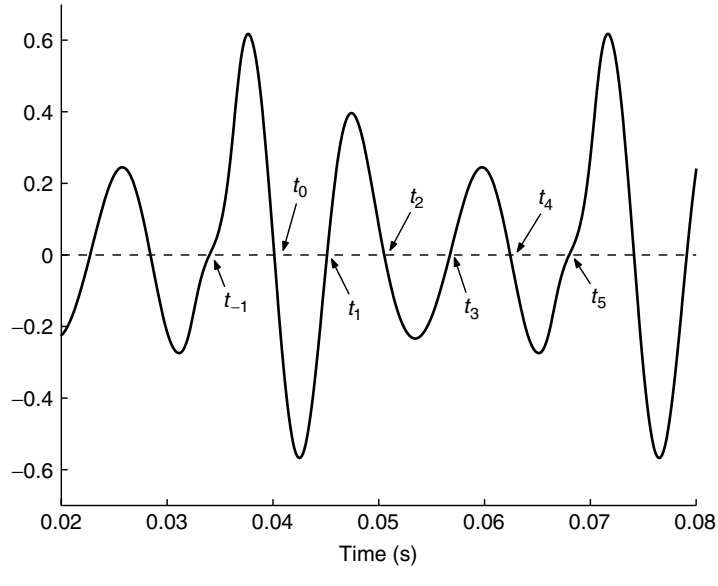

Figure 2.29 A signal with a period of 34 ms ($f_0 \approx 30$ Hz) has several zero crossings in the periodicity interval. The zero crossings for one period are indicated using the notation of this section: t_{-1} indicates the start of the period, and t_5 indicates the end of the period. There are five intermediate zero crossings. Note that $t_5 \equiv t_{-1}$ of the next period

where $a_n = a^*_{-n}$, because $f(t)$ is real. The output signal of the rectifying NLD is denoted by $g(t)$, which has the Fourier series representation

$$g(t) = \sum_{n=-\infty}^{\infty} b_n \mathrm{e}^{i2\pi f_0 nt}, \tag{2.85}$$

with of course also $b_n = b^*_{-n}$. The objective is to express the b_n in terms of the a_n.

2.6.1 OUTPUT SPECTRUM OF A RECTIFIER

On the periodicity interval $[t_{-1},\ t_{\mathrm{N}})$, the function $g(t)$ is given by

$$g(t) = |f(t)|, \tag{2.86}$$

which can also be written as

$$|f(t)| = f(t)h(t;\ \mathbf{t}) \tag{2.87}$$

where

$$h(t;\ \mathbf{t}) = \begin{cases} 1 & \text{for} \quad t_{-1} \le t < t_0, \\ -1 & \text{for} \quad t_0 \le t < t_1, \\ \vdots & \qquad\quad \vdots \\ (-1)^N = -1 & \text{for} \quad t_{\mathrm{N}-1} \le t < t_{\mathrm{N}}. \end{cases} \tag{2.88}$$

Let d_n be the Fourier coefficients of $h(t)$, so

$$\mathrm{d}_0 = 1 + 2f_0 \sum_{k=0}^{N} (-1)^k t_k, \tag{2.89}$$

$$\mathrm{d}_n = -\frac{1}{i\pi n} \sum_{k=0}^{N} (-1)^k \mathrm{e}^{-i2\pi f_0 n t_k}, \quad n \neq 0. \tag{2.90}$$

From the foregoing

$$g(t) = \sum_{n=-\infty}^{\infty} a_n \mathrm{e}^{i2\pi f_0 nt} \times \sum_{m=-\infty}^{\infty} \mathrm{d}_m \mathrm{e}^{i2\pi f_0 mt} = \sum_{k=-\infty}^{\infty} \mathrm{e}^{i2\pi f_0 kt} \times \sum_{n+m=k} a_n\, \mathrm{d}_m, \tag{2.91}$$

and therefore

$$b_k = \sum_{n=-\infty}^{\infty} a_n\, \mathrm{d}_{k-n}. \tag{2.92}$$

Combining all the previous results

$$b_k = \left(1 + 2f_0 \sum_{m=0}^{N} (-1)^m t_m\right) a_k - \sum_{n \neq k} \frac{a_n}{i\pi(k-n)} \sum_{m=0}^{N} (-1)^m \mathrm{e}^{i2\pi f_0 (n-k) t_m}. \tag{2.93}$$

Having expressed the b_k in terms of the a_k (also using the locations of the zeros of $f(t)$), the problem is solved in principle. However, the right-hand side of Eqn. 2.93 exhibits a decay of the b_k roughly as $1/k$, while the form of $g(t)$ suggests that there should be a decay like $1/k^2$, due to the triangular singularities at the t_i. This decay of the b_k can be made explicit by properly using the condition stated in Eqns. 2.82 and 2.83. Accordingly,

$$\sum_{n=-\infty}^{\infty} a_n \mathrm{e}^{i2\pi f_0 n t_m} = 0. \tag{2.94}$$

Then the series at the far right-hand side of equation 2.93, for $k \neq 0$, becomes

$$\begin{aligned}
\sum_{n \neq k} \frac{a_n}{i\pi(k-n)} \sum_{m=0}^{N} (-1)^m \mathrm{e}^{i2\pi f_0 (n-k) t_m} &= \sum_{n \neq k} \frac{a_n}{i\pi} \left(\frac{1}{k-n} - \frac{1}{k} + \frac{1}{k}\right) \sum_{m=0}^{N} (-1)^m \mathrm{e}^{i2\pi f_0 (n-k) t_m} \\
&= \sum_{n \neq k} \frac{n a_n}{i\pi k (k-n)} \sum_{m=0}^{N} (-1)^m \mathrm{e}^{i2\pi f_0 (n-k) t_m} + \\
&\quad \frac{1}{i\pi k} \sum_{n \neq k} a_n \sum_{m=0}^{N} (-1)^m \mathrm{e}^{i2\pi f_0 (n-k) t_m}.
\end{aligned} \tag{2.95}$$

And also

$$\begin{aligned}
\sum_{n \neq k} a_n \mathrm{e}^{i2\pi f_0 (n-k) t_m} &= -a_k + \sum_{n=-\infty}^{\infty} a_n \mathrm{e}^{i2\pi f_0 (n-k) t_m} \\
&= -a_k + \mathrm{e}^{-i2\pi f_0 k t_m} \sum_{n=-\infty}^{\infty} a_n \mathrm{e}^{i2\pi f_0 n t_m} \\
&= -a_k.
\end{aligned} \tag{2.96}$$

Hence for $k \neq 0$, the second term of Eqn. 2.95 vanishes, and

$$b_0 = f_0 \int_{t_{-1}}^{t_N} |f(t)| \, \mathrm{d}t, \tag{2.97}$$

$$b_k = \left(1 - 2f_0 \sum_{m=0}^{N} (-1)^m t_m\right) a_k - \sum_{n \neq k} \frac{n a_n}{i\pi k (k-n)} \sum_{m=0}^{N} (-1)^m \mathrm{e}^{i2\pi f_0 (n-k) t_m}. \tag{2.98}$$

The right-hand side of Eqn. 2.98 does exhibit the correct $1/k^2$ behaviour that is expected from the b_k's for large k. More precisely, assuming that $a_k = 0$ for large k, this becomes (for large k)

$$\sum_{n\neq k} \frac{na_n}{i\pi k(k-n)} \sum_{m=0}^{N} (-1)^m \mathrm{e}^{i2\pi f_0(n-k)t_m} \approx \frac{1}{k^2} \sum_{n=-\infty}^{\infty} \frac{na_n}{i\pi} \sum_{m=0}^{N} (-1)^m \mathrm{e}^{i2\pi f_0(n-k)t_m}. \quad (2.99)$$

Since

$$f'(t) = \sum_{n=-\infty}^{\infty} i2\pi f_0 n a_n \mathrm{e}^{i2\pi f_0 nt}, \quad (2.100)$$

this can be written as

$$\begin{aligned} &\sum_{n\neq k} \frac{na_n}{i\pi k(k-n)} \sum_{m=0}^{N} (-1)^m \mathrm{e}^{i2\pi f_0(n-k)t_m} \\ &\quad \approx \frac{1}{k^2} \frac{1}{i\pi} \frac{1}{i2\pi f_0} \sum_{n=-\infty}^{\infty} i2\pi f_0 n a_n \sum_{m=0}^{N} (-1)^m \mathrm{e}^{i2\pi f_0(n-k)t_m} \\ &\quad = -\frac{1}{2\pi^2 f_0 k^2} \sum_{m=0}^{N} (-1)^m f'(t_m) \mathrm{e}^{-i2\pi f_0 k t_m}. \end{aligned} \quad (2.101)$$

Thus, if $a_k = 0$ for large k, then for large k

$$b_k \sim \frac{1}{2\pi^2 f_0 k^2} \sum_{m=0}^{N} (-1)^m f'(t_m) \mathrm{e}^{-i2\pi f_0 k t_m}. \quad (2.102)$$

It appears that the spectrum of $g(t) \equiv |f(t)|$ at high frequencies is mainly determined by the slope of $f(t)$ at its zero crossings.

2.6.2 OUTPUT SPECTRUM OF INTEGRATOR

Now we consider the integrating NLD; under the same assumptions as in Sec. 2.6.1, we get on the periodicity interval $[t_{-1},\ t_{\mathrm{N}})$

$$g(t) = \begin{cases} \int_{t_{-1}}^{t} |f(s)|\,\mathrm{d}s, & t_{-1} \le t < t_1, \\ -\alpha_0 + \int_{t_{-1}}^{t} |f(s)|\,\mathrm{d}s, & t_1 \le t < t_3, \\ \vdots & \vdots \\ -[\alpha_0 + \ \ldots\ + \alpha_{(N-1)/2}] + \int_{t_{-1}}^{t} |f(s)|\,\mathrm{d}s, & t_{z-2} \le t < t_z. \end{cases} \quad (2.103)$$

The α_i are the 'jumps' of $g(t)$ at the 'resetting moments', and are given by ($k \in \mathbb{Z}$)

$$\begin{aligned}
-\alpha_0 &= -\int_{t_{-1}}^{t_1} |f(s)|\,\mathrm{ds} \quad \text{at time } t = k/f_0 + t_1,\\
-\alpha_1 &= -\int_{t_1}^{t_3} |f(s)|\,\mathrm{ds} \quad \text{at time } t = k/f_0 + t_3,\\
&\vdots \quad \vdots\\
-\alpha_{(N-1)/2} &= -\int_{t_{z-2}}^{t_z} |f(s)|\,\mathrm{ds} \quad \text{at time } t = k/f_0 + t_{\mathrm{N}}.
\end{aligned} \tag{2.104}$$

The above notation is illustrated in Fig. 2.30. For $t \neq t_{-1},\ t_1,\ t_3, \ldots t_z$, we have $g'(t) = |f(t)|$, thus

$$g'(t) = |f(t)| - \sum_{m=0}^{(N-1)/2} \alpha_m \sum_{n=-\infty}^{\infty} \delta(t - n/F_0 - t_{2m+1}). \tag{2.105}$$

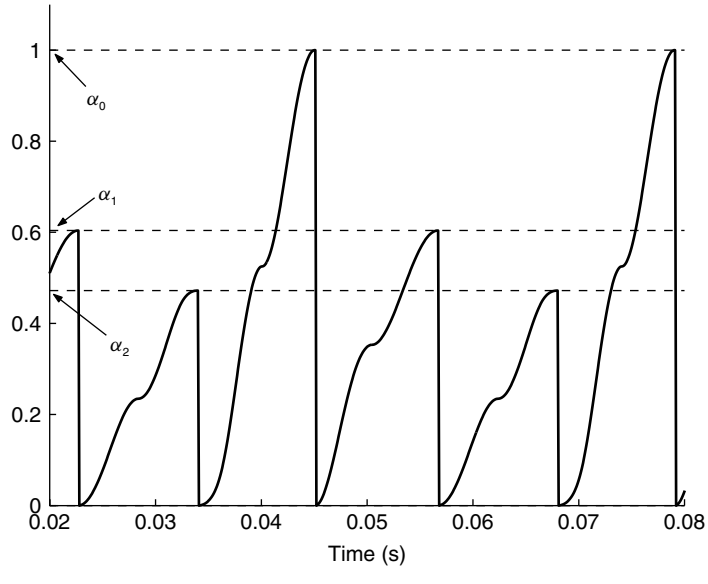

Figure 2.30 A signal with a period of 34 ms ($f_0 \approx 30$ Hz) has several zero crossings in the periodicity interval, leading to three 'resets' to 0 for one period of the output signal. The signal shown here is the result of applying the integrating non-linearity to the signal shown in Fig. 2.29. The magnitude of the resets, that is, the local maxima of the output signal, is indicated using the notation of this section

Denoting the Fourier coefficients of $|f(t)|$ by c_k, so that

$$|f(t)| = \sum_{k=-\infty}^{\infty} c_k \mathrm{e}^{i2\pi F_0 k t}, \tag{2.106}$$

and using $\sum_{n=-\infty}^{\infty} \delta(t - n/F_0 - t_{2m+1}) = \sum_{k=-\infty}^{\infty} \mathrm{e}^{i2\pi F_0 k(t-t_{2m+1})}$, we can write Eqn. 2.105 as

$$\sum_{k=-\infty}^{\infty} i2\pi F_0 k b_k \mathrm{e}^{i2\pi F_0 k t} = \sum_{k=-\infty}^{\infty} \left(c_k - \sum_{m=0}^{(N-1)/2} \alpha_m \mathrm{e}^{-i2\pi F_0 k t_{2m+1}} \right) \mathrm{e}^{i2\pi F_0 k t}, \tag{2.107}$$

so that

$$b_k = \frac{c_k - \sum_{m=0}^{(z-1)/2} \alpha_m \mathrm{e}^{-2\pi i \nu_0 k t_{2m+1}}}{2\pi i \nu_0 k} \quad k \neq 0. \tag{2.108}$$

The b_k show a decay of roughly $1/k$, which is what we expect owing to the discontinuities of $g(t)$ at t_{-1}, $t_1 \ldots t_{\mathrm{N}}$. The c_k can be found as the b_k of Eqn. 2.98. For $k = 0$ we get, with partial integration,

$$\begin{aligned} b_0 &= \int_{t_{-1}}^{t_{\mathrm{N}}} g(t)\,\mathrm{dt} \\ &= [\,tF(t)]_{t_{-1}}^{t_{\mathrm{N}}} - \int_{t_{-1}}^{t_{\mathrm{N}}} t \left(|f(t)| - \sum_{m=0}^{(z-1)/2} \alpha_m \sum_{n=-\infty}^{\infty} \delta(t - n/G_0 - t_{2m+1}) \right) \mathrm{d}t \\ &= -\int_{t_{-1}}^{t_{\mathrm{N}}} t|f(t)|\,\mathrm{d}t + \sum_{m=0}^{(N-1)/2} \alpha_m t_{2m+1}. \end{aligned} \tag{2.109}$$

2.6.3 OUTPUT SPECTRA IN DISCRETE TIME

For low-frequency psychoacoustic BWE applications, the frequencies of interest are orders of magnitude lower than the sample rate, such that continuous-time expressions, as we have used until now, are good approximations to the discrete-time expressions that actually should be used. However, for other BWE applications (notably high-frequency BWE treated in Chapters 5 and 6), the frequencies of interest can be in the same order of magnitude as the sample rate, and in such cases the proper discrete-time expressions must be used. These expressions can be developed along the same lines as the continuous-time expressions, and we therefore only give results, for the clipping and integrating non-linearity.

For this section, we use square brackets to index the variables, for example, as $f[n]$ instead of $f(t)$. We assume that $f[n]$ is periodic, with a period of N samples, sampled at a rate $f_{\mathrm{s}} = 1/\Delta t$. Zero crossings are defined by the sequence $x[n]$ from $f[n]$ as

$$x[n] = \begin{cases} 1 & \text{for } f[n] \geq 0, \\ 0 & \text{for } f[n] < 0. \end{cases} \tag{2.110}$$

We define $I[n]$ to be the *indicator* of the event $x[n] \neq x[n-1]$; if $x[n] \neq x[n-1]$, then $I[n] = 1$, else $I[n] = 0$. Now, we will define a zero crossing in $f[n]$ to occur for $n = n'$ if $I[n'] = 1$. Further assume that

•

$$I[n_{-1}] = I[n_0] = I[n_1] = \ldots = I[n_z] = 1, \tag{2.111}$$

$$I[n] = 0,\ n \neq n_{-1},\ n_0,\ n_1 \ldots n_z. \tag{2.112}$$

where all $n_{0,1\ldots z-1} \in (n_{-1}, n_z)$. Thus, $f[n]$ has z zero crossings in the interval (n_{-1}, n_z), and owing to the periodicity requirements on $f[n]$, z must be uneven; furthermore, $n_z - n_{-1} = N$ and $z \geq 1$.

- We choose $f[n_{-1} + 1] - f[n_{-1}] > 0$.

We have for $f[n]$ the Fourier series representation

$$f[n] = \frac{1}{N} \sum_{k=0}^{N-1} a[k] \mathrm{e}^{2\pi i k n/N}, \tag{2.113}$$

and because $f[n]$ is real we have $a[k] = a^*[-k]$. We consider the real periodic time series $F[n]$, derived by some non-linear operation from $f[n]$. We have for $F[n]$ the following Fourier series representation

$$F[n] = \frac{1}{N} \sum_{k=0}^{N-1} b[k] \mathrm{e}^{2\pi i k n/N}, \tag{2.114}$$

and again $b[k] = b^*[-k]$. Now the problem is again to express the $b[k]$ in the $a[k]$. In the limit that the sampling frequency tends to infinity, the discrete-time expressions are expected to equal the continuous-time expressions (this is indeed the case, as can be checked by setting $\lim_{\Delta t \downarrow 0}$ and replacing sums by integrals for the given discrete-time expressions). Note that due to the assumptions, specifically the assumption of periodicity, the derived results have limited applicability. This is because a periodic signal, when sampled, is only periodic if the sample rate and the signal's fundamental frequency have a greatest common divisor (GCD), in which we allow for non-integer arguments. If the GCD exists, it determines the periodicity of the sampled signal (N as mentioned above), which can thus be much longer than the periodicity of the continuous signal. For example, a 7-Hz pure tone (periodicity 0.144 s) sampled at 19 Hz (GCD is 1), has a periodicity of 1 s. In other cases, the periodicity interval can be extremely long (or non-existent) such that practical signals are not stationary within such time intervals. Nonetheless, the derived expressions have an academic use in that they can be used to assess output spectra for specifically chosen signals that have short periodicity. We may then expect that the conclusions from these output spectra can be used more generally.

2.6.3.1 Rectifier

For the rectifier, $F[n]$ is given by

$$F[n] = |f[n]|, \tag{2.115}$$

which leads to

$$b[k] = \left(1 + \frac{2}{N}\sum_{m=0}^{z}(-1)^m n_m\right) a[k] + \tag{2.116}$$

$$\frac{1}{N}\sum_{n\neq k} a[n] \sum_{m=0}^{z}(-1)^m \mathrm{e}^{-\pi i (n_{m-1}+n_m-1)(k-n)/N} \frac{\sin \pi (n_m - n_{m-1})(k-n)/N}{\sin \pi (k-n)/N}.$$

2.6.3.2 Integrator

On the periodicity interval $[n_{-1}, n_z)$, we get

$$F[n] = \begin{cases} \Delta t \sum_{k=n_{-1}}^{n} |f[k]|, & n_{-1} \leq n < n_1, \\ -\Delta t \sum_{k=n_{-1}}^{n_1-1} |f[k]| + \Delta t \sum_{k=n_{-1}}^{n} |f[k]|, & n_1 \leq n < n_3, \\ \vdots & \vdots \\ -\Delta t \sum_{k=n_{-1}}^{n_{z-2}-1} |f[k]| + \Delta t \sum_{k=n_{-1}}^{n} |f[k]|, & t_{z-2} \leq t < t_z. \end{cases} \tag{2.117}$$

Defining α_m as

$$\alpha_m = \Delta t \sum_{k=n_{-1+2m}}^{n_{1+2m}-1} |f[k]|, \tag{2.118}$$

we have

$$b[k] = \Delta t \frac{c[k] - \sum_{m=0}^{(z-1)/2} \alpha_m \mathrm{e}^{-2\pi i k n_{2m+1}/N}}{1 - \mathrm{e}^{-2\pi i k/N}}, \quad k \neq 0. \tag{2.119}$$

The $c[k]$ can be found as the $b[k]$ of Eqn. 2.116. For $k = 0$ we get, with partial summation[3],

$$b[0] = \sum_{k=n-1}^{n_z-1} F[k]$$

[3] Let $\sum_{n=0}^{\infty} a[n]$ be a series of which $s[n]$ ($n \in \mathbb{N}$) is the sequence of partial sums, and let $b[n]$ be a sequence. Then $\forall\ n \in \mathbb{N}$ we have that

$$\sum_{n=N_1}^{N_2} a[n]b[n] = \sum_{n=N_1}^{N_2} s[n](b[n] - b[n+1]) + s[N_2]b[N_2+1] - s[N_1-1]b[N_1].$$

$$= \sum_{k=n-1}^{n_z-1} (k+1)\{-\Delta t|f[k+1]| + \Delta t \sum_{m=0}^{(z-1)/2} \alpha_m \sum_{l=-\infty}^{\infty} \delta[k+1-lN-n_{2m+1}]\} + \Delta t(n_z|f[n_z]| - n_{-1}|f[n_{-1}]|)$$

$$= -\Delta t \sum_{k=n-1}^{n_z-1} k|f[k]| + \Delta t \sum_{m=0}^{(z-1)/2} \alpha_m n_{2m+1}. \tag{2.120}$$

2.6.4 OUTPUT SPECTRUM OF CLIPPER

Consider a real-valued, 2π-periodic signal $f(x)$, given in Fourier series form as

$$f(x) = \sum_{n=-\infty}^{\infty} a_n e^{inx}, \tag{2.121}$$

with

$$a_n = \frac{1}{2\pi} \int_0^{2\pi} f(x) e^{-inx}\, dx, \tag{2.122}$$

where the complex Fourier coefficients a_n satisfy $a_{-n} = a_n^*$, $n \in \mathbb{Z}$. Furthermore, assume a number $l_c > 0 \ \in \mathbb{R}$ (the clipping level), and the set

$$\{x \in [0, 2\pi] \ | \ |f(x)| \le l_c\} \tag{2.123}$$

in the form

$$\bigcup_{k=1}^{K} [\alpha_k, \beta_k], \tag{2.124}$$

where the $[\alpha_k, \beta_k] \subset [0, 2\pi]$ are pairwise disjoint intervals, as shown in an example in Fig. 2.31.

Let

$$f_{l_c}(x) = \begin{cases} f(x), & |f(x)| \le l_c, \\ l_c, & f(x) \ge l_c, \\ -l_c, & f(x) \le -l_c, \end{cases} \tag{2.125}$$

which is the clipped version of f at clipping level l_c. We want to compute the Fourier coefficients

$$b_n = \frac{1}{2\pi} \int_0^{2\pi} f_{l_c}(x) e^{-inx}\, dx, \quad n \in \mathbb{Z}. \tag{2.126}$$

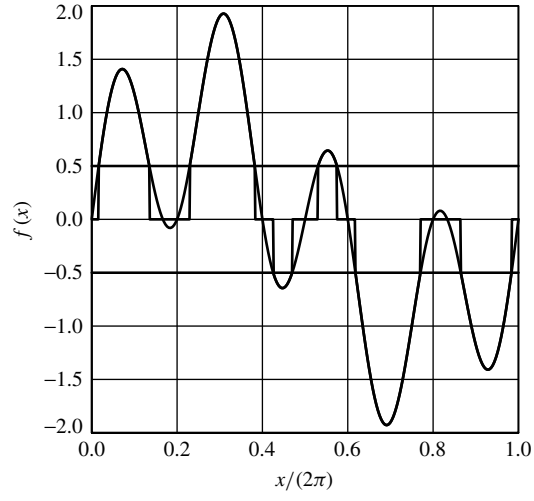

Figure 2.31 The function $f(x) = \sin(x) + \sin(4x)$ vs. x. The heavy portions of the x-axis indicate the intervals $[\alpha, \beta]$ where $|f(x)| \leq l_c$, where in this case $l_c = 0.5$

Note that

$$f'_{l_c}(x) = \sum_{n=-\infty}^{\infty} b_n e^{inx}, \tag{2.127}$$

hence

$$f'_{l_c}(x) = \sum_{n=-\infty}^{\infty} inb_n e^{inx}. \tag{2.128}$$

On the other hand, we have

$$f'_{l_c}(x) = \begin{cases} f'(x), |f(x)| < l_c, \\ 0, |f(x)| > l_c. \end{cases} \tag{2.129}$$

Therefore, using $f(x) = \sum_{m=-\infty}^{\infty} a_m e^{imx}$, we get

$$\begin{aligned} inb_n &= \frac{1}{2\pi} \int_0^{2\pi} f'_{l_c}(x) e^{-inx} \, dx = \frac{1}{2\pi} \sum_{k=1}^{K} \int_{\alpha_k}^{\beta_k} f'(x) e^{-inx} \, dx \\ &= \frac{1}{2\pi} \sum_{k=1}^{K} \int_{\alpha_k}^{\beta_k} \left(\sum_{m=-\infty}^{\infty} ima_m e^{imx} \right) e^{-inx} \, dx \\ &= \frac{1}{2\pi} \sum_{k=1}^{K} \sum_{m=-\infty}^{\infty} ima_m \int_{\alpha_k}^{\beta_k} e^{i(m-n)x} \, dx \end{aligned}$$

$$= \frac{1}{2\pi} \sum_{k=1}^{K} \sum_{m=-\infty}^{\infty} ima_m \frac{e^{i(m-n)\beta_k} - e^{i(m-n)\alpha_k}}{i(m-n)}. \tag{2.130}$$

Here, we have introduced the convention that for $\xi = 0$

$$\frac{e^{i\xi\beta} - e^{i\xi\alpha}}{i\xi} = \beta - \alpha, \tag{2.131}$$

which is correct in the limit and gives the same answer as treating the $m = n$ case separately in Eqn. 2.130. It thus follows that for $n \neq 0$

$$b_n = \frac{1}{2\pi n} \sum_{k=1}^{K} \sum_{m=-\infty}^{\infty} ma_m \frac{e^{i(m-n)\beta_k} - e^{i(m-n)\alpha_k}}{i(m-n)}. \tag{2.132}$$

For $n = 0$, we find more directly

$$b_0 = \frac{1}{2\pi} \int_0^{2\pi} f_{l_c}(x)\,dx = \frac{1}{2\pi} \sum_{k=1}^{K} \int_{\alpha_k}^{\beta_k} f(x)\,dx + \frac{l_c}{2\pi}|S_+| - \frac{l_c}{2\pi}|S_-|, \tag{2.133}$$

where $|S_+|$ and $|S_-|$ are the sizes of the sets

$$S_+ = \{x \in [0, 2\pi] \mid f(x) \geq l_c\}, \quad S_- = \{x \in [0, 2\pi] \mid f(x) \leq -l_c\}, \tag{2.134}$$

which should be available also. Note that the first number at the far right-hand side of Eqn. 2.133 can be expressed in terms of the a_n as

$$\begin{aligned} \frac{1}{2\pi} \sum_{k=1}^{K} \int_{\alpha_k}^{\beta_k} f(x)\,dx &= \frac{1}{2\pi} \sum_{k=1}^{K} \int_{\alpha_k}^{\beta_k} \sum_{m=-\infty}^{\infty} a_m e^{imx}\,dx \\ &= \frac{1}{2\pi} \sum_{k=1}^{K} \sum_{m=-\infty}^{\infty} a_m \frac{e^{im\beta_k} - e^{im\alpha_k}}{im}. \end{aligned} \tag{2.135}$$

Example Using the preceding method, we will calculate the Fourier coefficients of a clipped sine $\sin_{l_c}(x)$. Let

$$\begin{aligned} f(x) &= \sin x = \frac{e^{ix} - e^{-ix}}{2i}, \\ a_{\pm 1} &= \frac{\pm 1}{2i}, \text{ all other } a_m = 0, \\ l_c &\in (0, 1), \\ \alpha &= \arcsin l_c \in (0, \pi/2). \end{aligned} \tag{2.136}$$

$$\{x \in [0, 2\pi] \mid |f(x)| \leq a\} = [0, \alpha] \cup [\pi - \alpha, \pi + \alpha] \cup [2\pi - \alpha, 2\pi]. \tag{2.137}$$

Then we get

$$b_n = \frac{1}{2\pi n} \sum_{m=-\infty}^{\infty} m a_m \left\{ \frac{\mathrm{e}^{i(m-n)\alpha} - 1}{i(m-n)} + \frac{\mathrm{e}^{i(m-n)(\pi+\alpha)} - \mathrm{e}^{i(m-n)(\pi-\alpha)}}{i(m-n)} + \frac{1 - \mathrm{e}^{i(m-n)(2\pi-\alpha)}}{i(m-n)} \right\} \tag{2.138}$$

$$= \frac{1}{2\pi n} \sum_{m=-\infty}^{\infty} m a_m \left\{ \frac{\mathrm{e}^{i(m-n)\alpha} - \mathrm{e}^{-i(m-n)\alpha}}{i(m-n)} + (-1)^{m-n} \frac{\mathrm{e}^{i(m-n)\alpha} - \mathrm{e}^{-i(m-n)}}{i(m-n)} \right\} \tag{2.139}$$

$$= \frac{1}{2\pi n} \sum_{m=-\infty}^{\infty} m a_m . \frac{2\sin(m-n)\alpha}{m-n} (1 + (-1)^{m-n}). \tag{2.140}$$

Using Eqn. 2.136, we then get

$$b_n = \frac{1}{2\pi n} \left\{ \frac{\sin(1-n)\alpha}{i(1-n)} (1 + (-1)^{1-n}) + \frac{\sin(1-n)\alpha}{i(1-n)} (1 + (-1)^{-1-n}) \right\} \tag{2.141}$$

$$= \frac{1 + (-1)^{n-1}}{2\pi i n} \left(\frac{\sin(n-1)\alpha}{n-1} + \frac{\sin(n+1)\alpha}{n+1} \right), \tag{2.142}$$

and finally

$$b_n = \begin{cases} \frac{1}{\pi i(2\ell+1)} \left(\frac{\sin 2\ell\alpha}{2\ell} + \frac{\sin(2\ell+2)\alpha}{2\ell+2} \right), & n = 2\ell+1, \ell \in Z, \\ 0, & n = 2\ell, \ell \in Z. \end{cases} \tag{2.143}$$

Using $b_{-2\ell-1} = -b_{2\ell+1}$, we thus find

$$\sin_{\mathrm{lc}}(x) = \sum_{\ell=-\infty}^{\infty} b_{2\ell+1} \mathrm{e}^{i(2\ell+1)x} = \sum_{\ell=-\infty}^{\infty} (b_{2\ell+1} \mathrm{e}^{i(2\ell+1)x} + b_{-2\ell-1} \mathrm{e}^{-i(2\ell+1)x}) \tag{2.144}$$

$$= \sum_{\ell=-\infty}^{\infty} b_{2\ell+1} 2i \sin(2\ell+1)x = \frac{1}{\pi} \sum_{\ell=0}^{\infty} \left(\frac{\sin 2\ell\alpha}{\ell} + \frac{\sin 2(\ell+1)\alpha}{\ell+1} \right) \frac{\sin(2\ell+1)x}{2\ell+1}. \tag{2.145}$$