Introduction

I.1 BANDWIDTH DEFINED

The word 'bandwidth' can apply to different situations. The IEEE Standard Dictionary of Electrical and Electronics Terms [134] gives for the most relevant cases:

Definition 1 Bandwidth of a continuous frequency band: *The difference between the limiting frequencies.*

Definition 2 Bandwidth of a waveform: *The least frequency interval outside of which the power spectrum of a time-varying quantity is everywhere less than some specified fraction of its value at a reference frequency.*

Definition 3 Bandwidth of a signal transmission system: *The range of frequencies within which performance, with respect to some characteristic, falls within specific limits (usually 3 dB less than the reference or maximum value).*

Two of the above definitions show that what exactly the bandwidth of a signal or transmission system is will depend on a more or less arbitrary choice. For example, in Def. 3 the standard notion of '3 dB below the reference' is indicated. This problem arises because the power spectrum of a signal does not terminate abruptly, at least not for physically realizable signals. An extensive study of bandwidth and the relations between time limiting and frequency limiting was conducted by Slepian, Landau, and Pollack, and published in a series of landmark papers [250, 154, 155]; a short overview of this work is presented by Slepian [249].

'Bandwidth extension' (BWE) indicates the process of increasing the bandwidth of a signal. In the context of this book, BWE is usually achieved by a signal-processing algorithm. Sometimes, explicit use is made of the properties of the auditory system, and the signal processing is done in such a way as to let the actual BWE take place in the auditory system itself. The use of BWE implies that at some point bandwidth reduction has taken place, which is the opposite of bandwidth extension. Examples of where bandwidth reduction occurs are telephony, perceptual audio coding (at low bit rates), and sound reproduction with non-ideal transducers; these examples will be further explored in the various chapters of this book, and solutions in terms of BWE algorithms are also presented.

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

I.2 HISTORIC OVERVIEW

BWE methods are required because the systems they operate with are somehow sub-optimal, and usually so by design. For example, loudspeakers can be built such that they properly reproduce the entire audible frequency spectrum, down to 20 Hz; but such systems would be very expensive and also very bulky. As another example, digital storage and transmission of audio can be done without loss of information, but at a rather high bit rate. To achieve higher storage (coding) efficiency such that more audio can be stored with the same amount of bits, information has to be discarded. Telephony is another example where economic constraints led to the design of a transmission system that had the smallest bandwidth that could be used, while ensuring good speech intelligibility at the price of a markedly reduced quality.

The process of being ever more economical is still going on, but at the same time people demand the highest possible sound quality. Here, we briefly look at this from a historical perspective.

I.2.1 ELECTROACOUSTIC TRANSDUCERS

Loudspeakers have been around for a long time, but the practicality of loudspeakers is still limited. In 1954, Hunt [114] pointed out prosaically

> '*Electroacoustics is as old and as familiar as thunder and lightning, but the knowledge that is the power to control such modes of energy conversion is still a fresh conquest of science not yet fully consolidated.'*

This is still true today, and in fact one of the reasons for the need for BWE is the limited bandwidth of transducers, in particular, electroacoustic transducers (devices to convert electric energy into acoustic energy, or vice versa). Especially, low frequencies are difficult to reproduce efficiently. We can classify electroacoustic transducers in the following five categories:

- *Electrodynamic*: Movement is produced because of a current flow in a wire located in a fixed magnetic field. Most drivers in audio and TV sets are of this type, and will be discussed in greater detail in Sec. 1.3.2.
- *Electrostatic*: Movement is produced because of a force between two or more electrodes with a (high) voltage difference. Condenser microphones are of this type; for loudspeakers, they are mainly for Hi-Fi use.
- *Magnetic*: Movement is produced by attraction of metal due to an electromagnet. This is very common for doorbells and hearing aids, but it is not very much in use for loudspeakers.
- *Magnetostriction*: Movement is due to the magnetostriction effect – an effect arising in a variety of ferromagnetic materials whereby magnetic polarization gives rise to elastic strain, and vice versa.
- *Piezoelectric*: Movement is produced because of the direct and converse piezoelectric effect – an effect arising in a variety of non-conducting crystals whereby dielectric polarization gives rise to elastic strain, and vice versa. One usually sees this type of

loudspeakers for high-frequency units (tweeters) only, since with a low voltage only small movements can be achieved.

Each of these classes has its benefits and specific applications areas, but none can yield an overall desirable performance. Then there are some other, more exotic methods of sound transduction (which are not much in use), like:

- Laser loudspeakers, using the photoacoustic effect (Westervelt and Larson [296]).
- 'Audio spotlight', using interfering ultrasonic sound rays (Yoneyama and Fujimoto [300], Pompei [211]) to make a narrow beam of audible sound from a small acoustic source.
- 'Singing display', which is based on electrostatic forces between the plates of an LCD display (description in Chapter 8).
- Flame loudspeaker (Gander [84]) and Ionophone (Russell [230]), using pyroacoustic transduction.

After the discovery of electromagnetism in 1802 and Reiss' telephone in 1860, it was Bell, on 10 March 1876, who uttered the famous words 'Mr. Watson, come here, I want to see you!', in the first successful electromagnetic transmission of speech. Not long after Bell's invention of the telephone, Charles Cuttris and Jerome Redding [55] (see also Hunt [114]) filed a US patent application describing what appears as the first moving-coil electroacoustic transducer. Various principles were explored, but it took until 1925 before the loudspeaker came to its full growth, due to the work of Rice and Kellog [223]. This year is generally considered as the birth of the modern loudspeaker. An intimidating array of books, research papers and patents has been devoted to the science and technology of transducers since 1925, a few of which are Beranek [28], Borwick [36], Gander [84], Geddes and Lee [85], Hunt [114], McLachlan [172], and Olson [192]. Although a tremendous amount of energy has been devoted to increasing the performance of transducers, Chapter 4 presents, as a special case of a BWE system, an unusual loudspeaker design that has the curious property that it has a very high efficiency at one (low) frequency only. This frequency is then used to reproduce most of the low bass of the audio signal, together with appropriate signal processing.

I.2.2 SOUND QUALITY

There is an ever-continuing desire to increase sound quality (which often competes with economic constraints). From 1925 to 1926, the Edison Company sponsored 'tone tests', recitals in which phonographic 're-creations' of musicians, as reproduced by the Edison 'Diamond Disc Phonograph', were compared directly to live performances by those same musicians (Thompson [269]). In auditoriums and concert halls across the US, curious crowds gathered to engage in a very public kind of critical listening. Today, we can hardly believe that these re-creations were indistinguishable from the original. But two things can be observed: (1) Those gatherings can be considered as the start of the 'A/B' listening test, and (2) that most people increase their demands (or, perhaps, change their expectations) for quality as soon as they get used to a certain quality level. After the introduction of loudspeakers in the early 1920s, there was a demand for more bass and more volume (Read and Welch [221, p239]), and this demand has never gone away.

As electrical engineering advanced over the years, and especially with the advent of digital technology, sound enhancement became possible through electronic means; nowadays, audio engineering relies heavily on signal-processing techniques. BWE is one of the methods that can be used to enhance the quality of sound, which is especially attractive in areas such as consumer electronics. In this market, sound quality is often sub-optimal because of economic constraints on the size and cost of components. Most manufacturers want to produce as cheaply as possible, yet retain a high subjective quality. Nonetheless, quality does suffer, and in many cases a bandwidth reduction results. Electronic means (such as BWE systems) are comparatively cheap and flexible, and play an ever-larger role in determining the sound quality of audio systems. Chapter 2 presents several signal-processing methods that allow a small loudspeaker to be used for reproducing a wide low-frequency bandwidth, and is thus a prime example of how signal processing can be used to circumvent physical/acoustical difficulties. Bandwidth reduction due to audio compression can be (partially) negated by BWE algorithms discussed in Chapter 5, and in Chapter 6, we show how speech quality can be improved using the existing narrow-band telephone network.

I.3 BANDWIDTH EXTENSION FRAMEWORK

I.3.1 INTRODUCTION

Here we introduce BWE from a general point of view, adapted from Aarts *et al.* [10]. We focus on requirements that BWE algorithms need to comply with, both perceptual as well as implementational (in a broad sense; more specific treatments are given in the appropriate chapters). This overview is also useful because various BWE algorithms are very similar from a conceptual point of view, even though in detail they can be quite different. Keeping the general picture of BWE in mind makes it easier to find connections between the topics discussed in the various chapters.

An obvious way to categorize various BWE methods is based on the frequency range of interest. Some methods extend the low end of the audio spectrum, other methods extend the high end of the spectrum. The classifications ‘low’ and ‘high’ in this sense are relative to the remaining audio spectrum, and should not be considered in absolute sense. The second categorization is to realize ‘where’ the signal bandwidth is actually extended: in the auditory system or in the reproduced waveform. In other words, psychoacoustic or physical. These four categories are indicated in Table I.1 (and in Fig. I.1), together with references to the chapters where the various BWE categories are discussed. Finally, we

Table I.1 The four categories of BWE as function of frequency band and type, with reference to the chapters that cover that kind of BWE application. Chapter 1 covers background material, and Chapter 8 presents an overview of BWE patents

	Low band	High band
Psychoacoustic	Chapters 2, 7	N/A
Physical	Chapters 3, 4	Chapters 5, 6

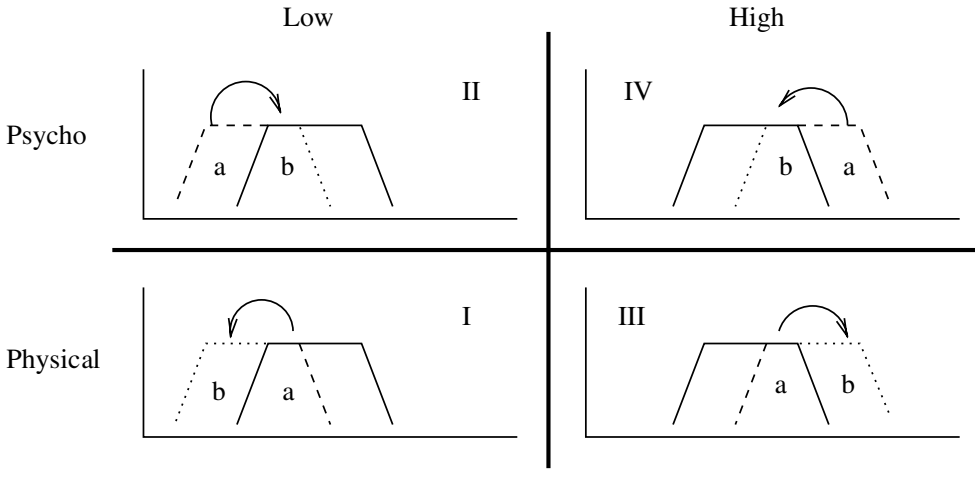

Figure I.1 The four categories of bandwidth extension. Low-frequency psychoacoustic BWE in the upper left panel, high-frequency psychoacoustic BWE in the upper right panel (which as yet has no practical implementation), low-frequency physical BWE in the lower left panel, and high-frequency BWE in the lower right panel. Energy from the dashed frequency range 'a' is shifted to the dotted frequency range 'b'. Adapted from Aarts *et al.* [10]

can identify BWE algorithms that use a priori information about the desired frequency components that need to be resynthesized, and those that use no such information. The latter class is termed 'blind', the other 'non-blind'. Generally, psychoacoustic BWE algorithms are never blind, because it is not the signal that is bandlimited, but the transducer. For the low-frequency psychoacoustic case, the application area is small loudspeakers, which cannot radiate very low frequency components. These low frequencies are present in the received signal, but have to be modified in such a way that the loudspeaker can reproduce a signal that has the same pitch percept, yet does not physically contain very low frequencies. High-frequency psychoacoustic BWE does not exist; it would require an algorithm that can yield a 'bright' (in terms of timbre) sound percept without reproducing the required high-frequency components. No known psychoacoustic effect can do this, and therefore this technology does not exist[1]. Physical BWE methods can be either blind or non-blind, because in these cases it is the signal that is bandlimited, not the transducer. The BWE algorithm has to resynthesize the missing frequency components from the narrow-band input signal. This can be done quite well with a priori information about the high frequencies, but it is also possible without such information (although usually at a lower quality). In this book, we almost exclusively deal with blind BWE systems, the exception being a kind of high-frequency BWE for audio, discussed in Sec. 5.5. It appears that the four classes of blind algorithms (of which three have practical applications), all

[1] Therefore, we simply use the term high-frequency BWE for what is in terms of the categorization of Table I.1 properly called high-frequency physical BWE.

have similar requirements, and can be implemented in a broadly similar fashion (although varying greatly on a more detailed level).

Blind algorithms can only use statistical information about the expected signals. This calls for a more general approach than what non-blind algorithms would require, and in the following text, we show how this generality can be exploited to cast the various BWE categories into a generalized signal-processing framework. A signal-processing framework is developed, which can be used to design BWE algorithms for many applications.

I.3.2 THE FRAMEWORK

I.3.2.1 Bandwidth Extension Categories

The four categories of BWE were already presented, and have been arranged in matrix form in Fig. I.1, where the columns indicate either low- or high-frequency extension, and the rows indicate psychoacoustic or physical BWE type. Each of the four graphs indicates a stylized power spectrum of an audio signal. The arrow indicates the action of the BWE algorithm: energy from the dashed frequency range 'a' is shifted to the dotted frequency range 'b'. Such 'shifting' of energy from one frequency range to the other obviously needs to be done in a special way; this will be elaborately discussed in the remainder of the book. The four indicated categories of BWE have the following characteristics:

1. *Low-frequency physical BWE category*: The lowest frequency components of the signal are used to extend the lower end of the signal's spectrum. Such an algorithm can be used if the low-frequency bandwidth of the signal has been reduced in storage or transmission; alternatively, the algorithm can be used for audio enhancement purposes, even if no prior bandwidth reduction had taken place. The loudspeaker will need to have an extended low-frequency response to reproduce the synthesized low frequencies.
2. *Low-frequency psychoacoustic BWE category*: The lowest frequency components of the signal cannot be reproduced by the loudspeaker, and are shifted to above the loudspeaker's low cut-off frequency. This must be done in such a way as to preserve the correct pitch and loudness of the low frequencies (and timbre as well, but this is not entirely possible).
3. *High-frequency BWE category*: The highest frequency components of the signal are used to extend the higher end of the signal's spectrum. Such an algorithm can be used if the high-frequency bandwidth of the signal has been reduced in storage or transmission; alternatively, the algorithm can be used for audio enhancement, even if no prior bandwidth reduction had taken place. The loudspeaker will need to have an extended high-frequency response to reproduce the synthesized high frequencies (which is usually not a problem).
4. *High-frequency (psychoacoustic) BWE category*: the highest frequency components of the signal cannot be reproduced by the loudspeaker, and are shifted to below the loudspeaker's high cut-off frequency. This must be done in such a way as to preserve the correct pitch, timbre, and loudness of the high frequencies. However, there is no known psychoacoustic effect that evokes a bright timbre percept when only lower-frequency components are present. Therefore, this category of BWE has no known implementation.

I.3.2.2 Perceptual Considerations

All of the BWE algorithms derive from one part of a signal's spectrum, a second signal in a different frequency range, which is then added to the input signal. The sum of these two signals should blend together to form an enhanced version of the original. The analysis by the auditory system should therefore group these two signals into the same stream, yielding a single percept. Bregman [38] gives some clues as to what signal characteristics are important in this grouping decision: pitch, timbre, and temporal modulation. If any one of these parameters differs 'too much' between the two signals, the signals will be segregated and be heard as two separate streams. This would constitute a failure of the BWE algorithm. Therefore, we must ensure that all the said signal characteristics of the synthetic signal remain as similar as possible to those of the original signal. On the other hand, a slight dissimilarity between, say, the pitches of two signals, can be 'overcome' by strong similarity in temporal modulation. Indicating the synthetic signal (output of BWE algorithm) by $y(t)$, and the input signal by $x(t)$, we have the following considerations:

Pitch: $x(t)$ and $y(t)$ should have a similar tonal structure, that is, a common fundamental frequency f_0. If the signals are atonal (noise), then $x(t)$ and $y(t)$ should have similar moments (at least up to second order). We shall see that we can design efficient algorithms such that the pitch of $y(t)$ matches that of $x(t)$.

Timbre: Timbre is usually associated with the spectral envelope, although temporal envelope and spectral phase also have an influence. In the BWE algorithms, we can control timbre to some extent by the correct design of filters.

Loudness: Similar temporal modulations for $y(t)$ and $x(t)$ are required for covarying loudness of both signals, and can be achieved by ensuring that the amplitudes of $y(t)$ and $x(t)$ are (nearly) proportional. The BWE algorithms described later on will usually be linear in amplitude, so this is automatically taken care of.

Because there is little objective data available on how 'close' or 'similar' the mentioned psychoacoustic parameters must be for $x(t)$ and $y(t)$ to be grouped, especially for realistic audio signals, the tolerance of BWE algorithms for these grouping and segregation effects is, to some degree, a matter of trial and error.

I.3.2.3 Implementational Considerations

Besides perceptual constraints, there are some constraints on the implementation of the algorithms. These are not necessarily exclusive to BWE methods, but to most signal-processing algorithms for use in consumer electronic applications. These constraints are:

1. Low computational complexity and low memory requirements.
2. Independence of signal format.
3. Applicable to music, speech and, preferably, both.

The first constraint is important for the algorithm to be a feasible solution for consumer devices, which typically have very limited resources. Although the use of digital signal

processors is becoming more widespread in consumer electronics, many signal-processing features are usually implemented together, and each one may only use a fraction of the total computing power.

Independence of signal format means that the algorithm is applied to a PCM-like signal. Dependence on a certain coding or decoding scheme would limit the scope of the algorithm. Of course, in some cases, this can lead to higher-quality output signals, as for the SBR technology in high-frequency BWE application, discussed in Chapter 5.

The third constraint determines if we can use a detailed signal model or not. If the application is limited to speech only, a speech model that allows a more accurate BWE (Chapter 6) can be used. Signals not well described by the speech model, such as music, would not give good results with this algorithm. Thus, if the nature of the processed signals is unclear, a general algorithm must be used. Specialized BWE algorithms for music would be more difficult to devise than those for speech, because the statistics of musical signals depend heavily on the instrument being used; also, typically many instruments are active simultaneously, whereas in speech applications the signal can be assumed to derive from one sound source only. If we can decide for a given signal whether it is music or speech, which can be done with a speech-music discriminator (Aarts and Toonen Dekkers [6]), we can use BWE for music with strategies as discussed in Chapter 5, and BWE for speech with strategies as discussed in Chapter 6.

I.3.2.4 Processing Framework

Figure I.2 presents the signal-processing framework that we propose for all categories of BWE described here, and covers most of the BWE methods described elsewhere. The general algorithm consists of two branches: one in which the input signal is merely delayed, and one in which the bandwidth extension takes place. This bandwidth extension is done by bandpass filtering the input signal (FIL1) to select a portion of the audio signal (indicated by the letter 'a' in Fig. I.1). This portion is then passed to a non-linear device (NLD), which 'shifts' the frequencies to a higher or lower region by a suitable non-linear operation, according to the particular application. Subsequently, the signal is bandpass filtered (FIL2), to obtain a suitable spectrum and timbre (the signal now has frequencies in the range 'b', as shown in Fig. I.1). The resulting signal is amplified or attenuated as desired and added back to the (delayed) input signal to form the output.

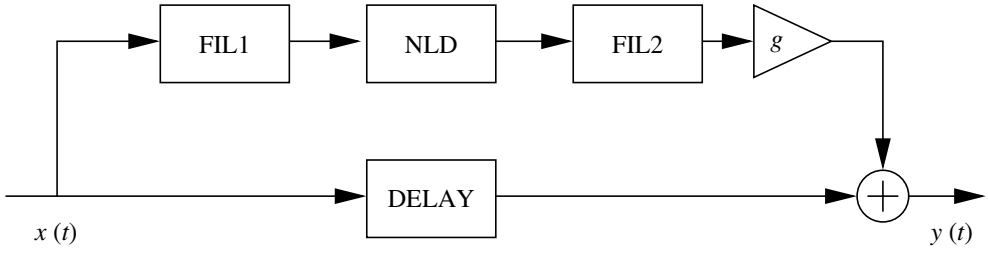

Figure I.2 General BWE framework, also shown as high-frequency BWE structure in Fig. 5.2. Similar structures are shown for low-frequency psychoacoustic BWE as Fig. 2.4 and for low-frequency physical BWE as Fig. 3.1

A somewhat different approach is taken in Secs. 2.4 and 3.3.2.4, where a frequency tracker is used and the harmonics signal is generated explicitly (not through non-linear processing of the filtered input). For speech BWE methods, discussed in Chapter 6, the same structure is used, but some of the processing steps are much more involved, in particular, the filtering by FIL2. This filter is derived adaptively from the input signal. For the non-blind BWE approach taken in Sec. 5.5, FIL2 is also adaptively determined, in this case by control information embedded in a coded audio bitstream. In Sec. 5.6, an extremely simplified algorithm (instantaneous compression) that only uses an NLD, without any filters, is used; this method is only suitable in particular circumstances and for particular signals.

Specific implementations for the NLD, which 'shifts' frequency components from low to high values (or vice versa), will be given in later chapters covering the specific BWE categories. They are based on generating harmonics or subharmonics of the signal passed by FIL1. Nearly all of the NLDs discussed in the following chapters implicitly determine the frequency of the incoming signal by its zero crossings (except the frequency tracker of Secs. 2.4 and 3.3.2.4). For a pure tone of frequency f_0, the situation is unambiguous, and there are $2f_0$ zero crossings per second. But because FIL1 is a bandpass filter of finite bandwidth, the signal going into the NLD will possibly contain more than one frequency component, which will disturb the zero crossing rate γ (number of zero crossing per second) of the signal. Fortunately, zero crossings appear to be quite robust in reflecting the dominant frequency of a signal, which is known as the 'dominant frequency principle'. This principle can be made explicit by the 'zero crossing spectral representation' (Kedem [141]) as

$$\cos(\pi\gamma) = \frac{\int_0^\pi \cos\omega \, \mathrm{d}F(\omega)}{\int_0^\pi \mathrm{d}F(\omega)}, \tag{I.1}$$

which holds for weakly stationary time series, with spectral distribution $F(\omega)$. For a pure tone of frequency $f_0 = \omega_0/2\pi$, we set $F(\omega) = \delta(\omega - \omega_0)$ in Eqn. I.1, and we find that $\cos(\pi\gamma) = \cos\omega_0$, which gives us the expected $\gamma = 2f_0$. If the signal passed by FIL1 has multiple frequency components, one of which is dominant, the NLD will 'detect' this frequency by the zero crossings of the signal, and construct a harmonics signal on the basis of this dominant frequency.