Audio Engineering Society
Convention Paper 5921
Presented at the 115th Convention
2003 October 10–13 New York, NY, USA

A unified approach to low- and high-frequency bandwidth extension

Ronald M. Aarts[1], Erik Larsen[2], and Okke Ouweltjes[1]

[1] *Philips Research Labs, 5656 AA Eindhoven, The Netherlands*

[2] *University of Illinois at Urbana-Champaign, Beckman Institute for Adv. Sci. and Techn., Urbana, IL, 61801, USA*

Correspondence should be addressed to Erik Larsen (elarsen@uiuc.edu)

ABSTRACT

Extending the bandwidth of an audio signal may be useful at the low or high end of the frequency spectrum, depending on the application. Also, the actual bandwidth extension algorithm may rely entirely on psychoacoustic effects or may create a physical extension of the signal spectrum. We have developed a common framework for all these problems, and from this framework derived algorithms that address diverse applications in audio signal processing for bandwidth extension. Specifically, we describe algorithms for bandwidth extension applied to enhancing reproduction of bandlimited signals (at the low or high end of the frequency spectrum), and for enhancing reproduction over small loudspeakers.

1. INTRODUCTION

Bandwidth limitation of audio signals may occur in a variety of circumstances, such as telephony, perceptual audio coding, or due to transducer limitations. For each specific case, a dedicated signal processing algorithm can be used to enhance the reproduction of the audio signal, depending on the nature of the

bandlimitation (low- or high-frequency), and on the statistics of the reproduced signals (speech or audio). Such signal processing algorithms are termed *bandwidth extension.* It is the main purpose of this paper to identify the various kinds of bandwidth extension algorithms, and for a subset of these to present a set of requirements and a general signal processing framework for implementation. The second half of the paper will discuss three specific bandwidth extension algorithms (two of which are based on prior published material). We will not report subjective evaluations of the algorithms here, as this has either been reported previously or will be reported at a later stage.

We classify bandwidth extension algorithms according to their frequency range, as mentioned above, but also as either *blind* or using *a priori* information about the 'missing' frequency components (non-blind). A final classification that can be made is whether the bandwidth extension method evokes a purely psychoacoustic extension of the frequency spectrum (and no physical or measurable extension is present), or whether the signal's spectrum is physically extended. In conclusion, bandwidth extension algorithms may be classified in a (at least) three-dimensional space with dimensions: low/high frequency, blind/non-blind, psychoacoustic/physical; this leads to eight different classes of algorithms.

It appears that of these eight classes of algorithms, quite a few share similar requirements, and can be dealt with in similar ways. An exception is the dimension blind/non-blind, where in either category quite different approaches will need to be taken. An algorithm using *a priori* information about the missing parts of the frequency spectrum can be designed in several ways. There may be an initial training phase, where the algorithm parameters are 'tuned' or initialized with typical full bandwidth and bandwidth limited signals, such that afterwards the algorithm may 'predict' what the full bandwidth signal should look like based on the examples in the training phase. Another example is the 'Spectral Band Replication' technique (e.g. Gröschel *et al.* [1]), used in perceptual audio codecs, where high-frequency parts of the audio signal are derived from baseband, aided by auxiliary data that was generated by the coder. We will not further consider such algorithms here.

We restrict our discussion of bandwidth extension algorithms to time domain methods. In the frequency domain there are problems such as spectral leakage for frequencies that are not harmonic to the DFT window, connection of output frames, and non-stationarity of the input signal within frames. By using time domain methods we can prevent these problems. We shall see that efficient implementations are possible.

A blind algorithm can only use statistical information about the expected signals. This calls for a more general approach than non-blind algorithms would employ, and in the following we show how this generality can be exploited to cast the various bandwidth extension categories into a generalized signal processing framework. The main focus of the paper will be on Secs. 2 and 3. In Sec. 2 we make clear by various examples the necessity of bandwidth extension, and show how the statistics of speech and audio signals can be used to formulate and design bandwidth extension algorithms. Sec. 3 then continues to categorize various bandwidth extension methods, and we define a set of requirements for these methods. A signal processing framework is set up which can be used to design bandwidth extension algorithms for many applications. The remainder of the paper is devoted to a discussion of the various bandwidth extension methods. These are Secs. 4 (low-frequency extension, new material), 5 (psychoacoustic low-frequency extension, discussed in Larsen and Aarts [2]), and 6 (high-frequency extension, material from Larsen *et al.* [3]).

2. STATISTICS OF AUDIO AND SPEECH

2.1. Introduction

For bandwidth extension methods it is important to have knowledge of the spectrum of the signal. Because audio is not stationary, the spectrum varies from moment to moment, and the spectrogram is useful to visualize this spectro-temporal behavior of speech and music. First we will consider speech, and then music. For both cases we will make plausible that certain bandwidth limitations can be overcome by suitable processing.

2.2. Statistics of speech

An important parameter for speech is the fundamental frequency. Furui [4] presents a statistical analysis of temporal variations in fundamental frequency of conversational speech for individual speakers, which indicates that the mean and standard deviation for a female voice are roughly twice those for a male voice. This is shown in Fig. 1. The fundamental frequency distributed over speakers on a logarithmic frequency scale can be approximated by two normal distributions which correspond to the male and female voice, respectively. The mean and standard deviation for a male voice are 125 and 20.5 Hz, respectively, whereas those for a female voice are twice larger.

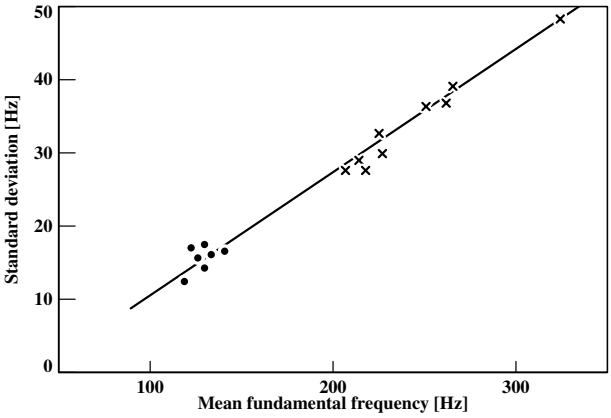

Fig. 1: Mean and standard deviation of the temporal variation of the fundamental frequency during conversational speech for various speakers (Fig. 2.11 from [4]). The full circles are male and the crosses female speakers.

In order to gain insight in the long-term average of speech spectra, six speech fragments of utterances of various speakers of both sexes were measured. Fig. 2 shows the power spectra of tracks 49–54 of the SQAM disk [5]. To parameterize the spectra in the plot we derived the following heuristic formula

$$\|H(f)\| \approx \frac{(\frac{f}{120})^6}{1+(\frac{f}{120})^6}\,\frac{1}{1+\frac{f}{1000}}\,, \qquad (1)$$

were f is the frequency. Byrne *et al.* [6] have shown that there is not much difference in the long-term speech spectra of different languages. The first factor in the product of Eq. (1) denotes the highpass behavior and the second one the lowpass behavior.

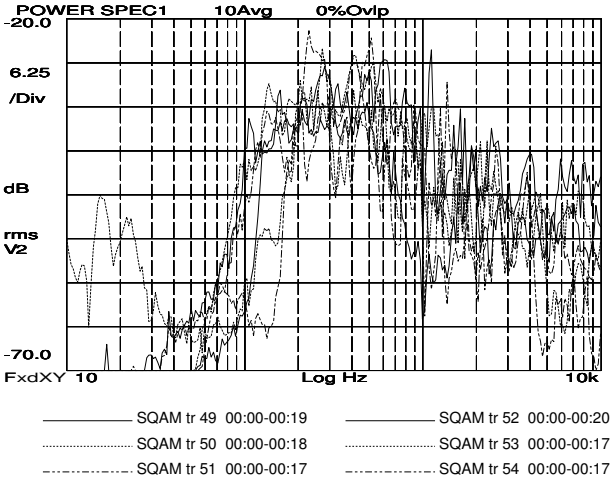

Fig. 2: Spectra of long time average of of six speech fragments of utterances of various speakers. Tracks 49–54 of the SQAM disk [5].

It shows that there is a steep slope at low frequencies and a rather modest slope at high frequencies. The lowest pitch of the male voice is about 120 Hz (see Fig. 1) which corresponds very well with the high pass frequency in Eq. 1, and Fig. 2. The telephone range is 300–3400 Hz, which clearly is not sufficient to pass the lowest male fundamental frequencies, and also not most female fundamental frequencies. A bandwidth extension algorithm that can recreate the fundamental (and possibly the lower harmonics) of speech signals (as will be outlined in Sec. 4) should be able to create a more natural sounding speech for telephony.

Because the speech spectrum changes over time, it is necessary to compute spectra at frequent intervals and display the changing spectra. A spectrogram is shown for a particular speech sample in Fig. 3 (2nd panel); the pitch of the voice is time varying, as can be seen in the lower panel of Fig. 3. A bandwidth extension algorithm must be able to follow these pitch changes and change its output frequencies accordingly. Fig. 4 shows a time plot and spectrogram (top two panels) of the same speech utterance, and an 8 kHz lowpass filtered version thereof, which could occur in perceptual audio coders at very high compression rates; the telephone channel would lowpass filter the signal at 3.4 kHz. A high-frequency band-

width extension algorithm (as will be outlined in Sec. 6) can resynthesize an octave of high frequency components, as shown in the lower panel of Fig. 4; note similarities and differences with respect to the original spectrogram.

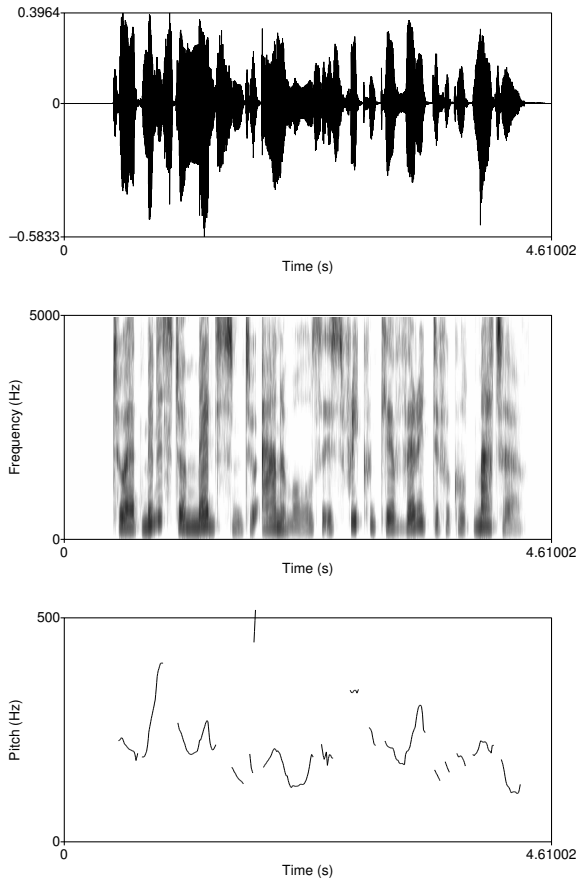

Fig. 3: Time signal (upper panel) and spectrogram (middle panel) of the first 4 s of a female voice utterance 'To administer medicine to animals is frequently a very difficult matter' (track 49 of the SQAM disk [5]). The middle panel shows the spectrogram in grayscale (dark tone indicates high energy). The lower panel shows the pitch of that utterance, determined by a pitch tracker ('Praat' [7]).

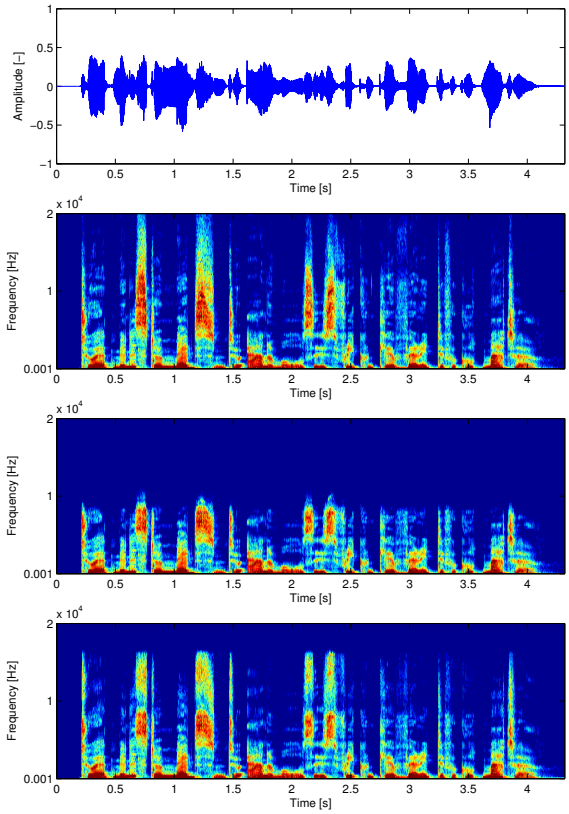

Fig. 4: Time signal (upper panel) and spectrogram (second panel) of the first 4 s of a female voice utterance 'To administer medicine to animals is frequently a very difficult matter' (track 49 of the SQAM disk [5]). The third panel shows the spectrogram of the same signal, but filtered by a 3rd order Chebyshev lowpass filter at 8 kHz. The fourth panel shows the signal after high-frequency bandwidth extension as will be discussed in Sec. 6. These spectrograms are color coded, where blue–red is low–high energy.

2.3. Statistics of music

More than 70 years ago Sivian *et al.* [8] performed a pioneering study of musical spectra using live

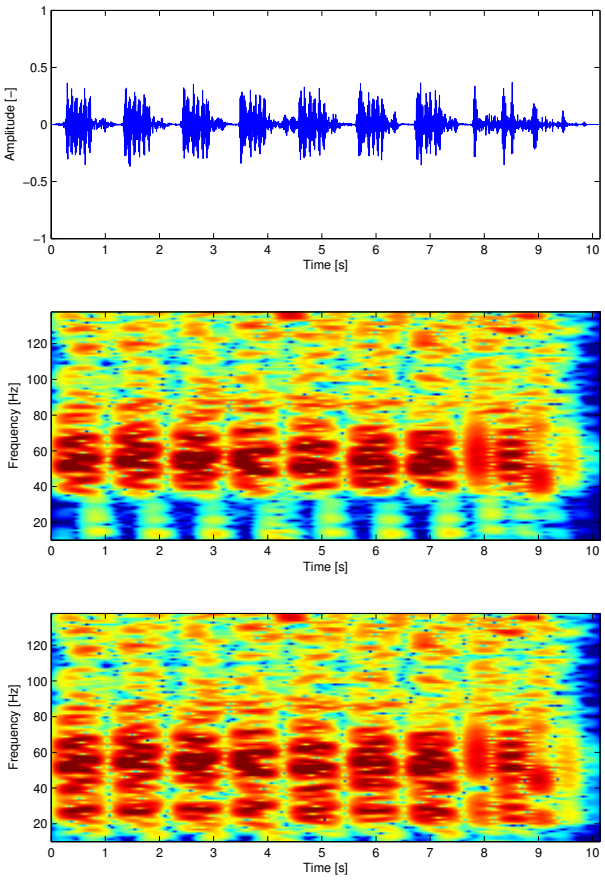

Fig. 5: The top two panels show time plot and spectrogram (only low frequencies) of 'One' by Metallica. The lowest frequencies occur around 40 Hz. A low-frequency bandwidth extension algorithm extends this low-frequency limit to 20 Hz, as shown in the lower panel. These spectrograms are color coded, where blue–red is low–high energy.

musicians and—for that time—innovative electronic measuring equipment. Shortly after the introduction of the CD, this study was repeated by Greiner and Eggers [9] by using modern digital equipment and modern source material, at that time CDs. The result of both studies was a series of graphs showing for each instrument or ensemble the spectral amplitude distribution of the performed musical passage. In general the spectrum has a bandpass characteristic, the exact shape of which is determined by the music and the instrument. As in speech, the fundamental frequency (pitch) is time varying. A complicating factor is that various instruments may be playing together.

An example is shown in Fig. 5, where the variable time-frequency characteristic of metal music is shown ('One' by Metallica). The middle panel shows a spectrogram (frequencies 0 – 140 Hz) of the original performance. The energy in the signal extends down to about 40 Hz. By using low-frequency bandwidth extension, we can extend this lower limit to about 20 Hz (lower panel of Fig. 5), which requires a subwoofer of excellent quality for correct reproduction. The resulting synthetic frequencies have similar spectro-temporal characteristics as the original low frequencies, and will add 'feeling' to the music.

Another study (Fielder and Benjamin [10]) was con-

ducted to establish design criteria for the performance of subwoofers to be used for the reproduction of music in the home. The focus on subwoofers was motivated by the fact that low frequencies play an important part in the musical experience. A first conclusion of that study was that recordings with audible bass below 30 Hz are relatively rare. Second, these very low frequencies were generated by pipe organs, synthesizers, or special effects and environmental noises. Other instruments, such as bass guitar, bass viol, tympani, or bass drum, produce relatively little output below 40 Hz, although they may have very high levels at or above that frequency. Fielder and Benjamin [10] gave an example that for an average listening room of 68 m^3, the required acoustic power for reproduction is 0.0316 W, giving an SPL of 97 dB, which requires a volume displacement of 0.685 l at 20 Hz. This requires an excursion of 13.5 mm for a 10 in (0.25 m) woofer. These are extraordinary requirements, and very hard to fulfil in practice. An alternative is to use psychoacoustic low-frequency bandwidth extension, where frequencies that are too low to reproduce are shifted to higher frequencies, in such a way that the pitch percept remains the same. If we consider the lower panel of Fig. 5 as the original signal, we could think of such a bandwidth extension as shifting the frequency band 20 – 40 Hz to above 40 Hz. The result would look somewhat like the middle panel of Fig. 5, with increased energy above 40 Hz. We will outline such bandwidth extension methods in Sec. 5.

3. THE FRAMEWORK

3.1. Bandwidth extension categories

In the Introduction the eight categories of bandwidth extension were discussed. If, for the moment, we do not consider algorithms that use *a priori* information, this reduces to four categories. These categories are arranged in matrix form in Fig. 6, were the columns indicate either low- or high-frequency extension, and the rows indicate psychoacoustic or physical bandwidth extension. Each of the four graphs indicates the power spectrum of an audio signal. The arrow indicates the action of the bandwidth extension 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 next obviously needs to be done in a special way; this is the topic of Secs. 4, 5, and 6. The four indicated categories of bandwidth extension can be seen in the following light:

1. Low-frequency (physical) bandwidth extension: 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 'effect'. The loudspeaker will need to have an extended low-frequency response to reproduce the synthesized low frequencies. We discuss these methods in Sec. 4, all of which is new material. Previous investigations are mainly reported in the patent literature, some recent examples are Grob-Da Veiga [11] and Oda [12]. Most of these methods deal with specific applications; no detailed studies of a more general nature are known.

2. Low-frequency (psychoacoustic) bandwidth extension: the lowest frequency components of the signal cannot be reproduced by the loudspeaker, and are shifted to above the loudspeaker's low cutoff frequency. This must be done in such a way as to preserve the correct pitch and timbre of the low frequencies. Several implementations of such algorithms exist, and are described in e.g. Tan *et al.* [13], Gan *et al.* [14], and Larsen and Aarts [2] (the latter including a subjective quality study). Griffiths [15] studied the effect as it may unintentionally occur for small loudspeakers that are driven into saturation, causing distortion products. Sec. 5 reviews this category of algorithms.

3. High-frequency (physical) bandwidth extension: 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 'effect'. The loudspeaker will need to have an extended high-frequency response to reproduce the synthesized high frequencies. Such algorithms have been previously described in

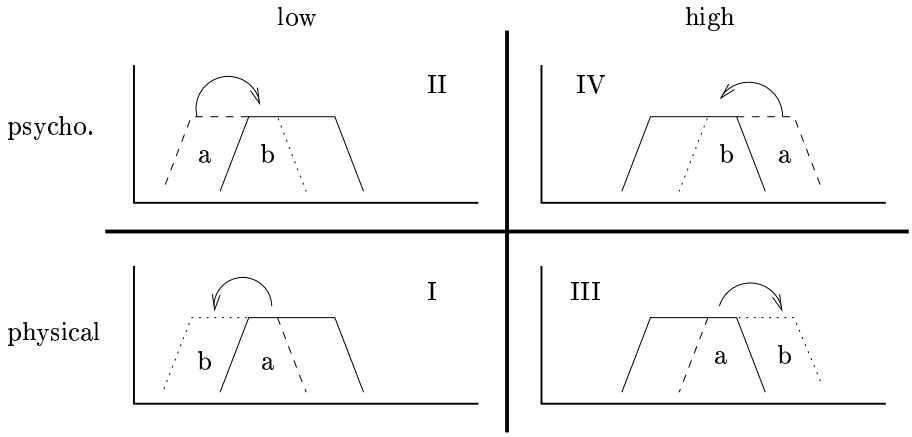

Fig. 6: Four categories of bandwidth extension, along the dimension psychoacoustic/physical (rows) and low/high frequency (columns). In each case, the arrow indicates 'shifting' of energy from frequency region 'a' to frequency region 'b'.

Larsen *et al.* [3] and Aarts *et al.* [16]. These algorithms have been studied fairly extensively for applications to speech in telephony, as in e.g. Taori *et al.* [17], Chennoukh *et al.* [18], and references therein (some of these methods employ frequency domain algorithms). Sec. 6 will review this category of bandwidth extension.

4. High-frequency (psychacoustic) bandwidth extension: the highest frequency components of the signal cannot be reproduced by the loudspeaker, and are shifted to below the loudspeaker's high cutoff frequency. This must be done in such a way as to preserve the correct pitch and timbre of the high frequencies. To the authors' knowledge, there is no known effect that evokes a high frequency percept when only lower frequency components are present[1], and therefore this category of bandwidth extension has no known implementation. Beside applications in audio reproduction, a successful algorithm of this kind could have great potential for the hearing impaired, as most hearing impaired people have a sloping hearing loss that increases with frequency. We will not further consider this category in this communication.

[1]Except possibly edge pitch (Kohlrausch and Houtsma [19]), that can evoke a pitch slightly above the upper spectral edge of a complex tone. The effect does not seem useful for the kind of bandwidth extension proposed, though.

Bandwidth extension using *a priori* information can also be described as applied in one of the four preceding categories. However, their implementation will be very different from the algorithms that we shall discuss.

3.2. Perceptual considerations

All of the bandwidth extension algorithms derive from a part of the signal's spectrum another 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 *group* these two signals into the same auditory *stream*, yielding a single percept. Bregman [20] gives some clues as to what signal characteristics are important in this grouping decision: pitch, timbre, and amplitude envelope. If any one of these parameters differs 'too much' between the two signals, the signals will be *segregated* and heard as two individual streams. This would constitute a failure of the bandwidth extension 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. We will indicate the synthetic signal (output of bandwidth extension algorithm) by

s_s and the main signal by s_m. We have the following considerations:

Pitch: Signals s_s and s_m should have a similar tonal structure, i.e. a common fundamental f_0. If the signals are atonal (noise), then s_s and s_m should have similar moments (at least up to second order). We shall see that we can design efficient algorithms such that the pitch of s_s matches that of s_m.

Timbre: Timbre is a psychophysical parameter that is hard to define concisely. The current definition of the American Standards Association defines "that attribute of an auditory sensation in terms of which two sounds with similar pitch, loudness, and duration, presented under similar circumstances, may be distinguished", which is a better description of what timbre is *not* than what it *is*. Timbre is usually associated with the shape of the signal spectrum, although amplitude envelope also has an influence. In the bandwidth extension algorithms we can control timbre to some extent by correct design of various parts (to be discussed later).

Envelope: Similar envelopes for s_s and s_m are required for covarying loudness of both signals, and can be achieved by ensuring that the amplitudes of s_s and s_m are (nearly) linearly related. The bandwidth extension algorithms described later will all be linear in amplitude.

Because there is no hard data available on how 'close' or 'similar' the mentioned psychoacoustic parameters must be for two signals to be perceptually grouped as one, the design of bandwidth extension algorithms necessarily has some trial-and-error aspects.

3.3. Implementational considerations

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

1. Low computational complexity and low memory requirements.
2. Independent of signal format.
3. Applicable to music and speech.

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 (DSPs) is becoming more widespread in consumer electronics, such DSPs have many tasks to perform, of which any one may take up only a limited amount.

Independence of signal format means that the algorithm is applied to a PCM-like (or even analog) signal. Dependence on a certain coding or decoding scheme would limit the scope of the algorithm and is therefore to be avoided.

The third constraint implies that we cannot use a very detailed signal model. If application was limited to, say, speech only, a speech model could be used which would probably allow a more accurate bandwidth extension. Signals not well described by this model, such as music, would not give good results. To comply with the third constraint mentioned above, we can only use characteristics common to all, or at least most, possible speech and music signals.

These requirements force us to design algorithms that are of general nature, giving good results for a wide class of signals. This also implies that the algorithm will not give optimal results for any one particular signal.

3.4. Processing framework

Fig. 7 presents the signal processing framework that we propose for all categories of bandwidth extension described here, and covers most of the bandwidth extension methods described elsewhere. Elaborations are possible, and some of the processing steps may be implemented in a more sophisticated way than we shall reveal here, but the essential elements will always reduce to the framework we are about to describe. For a stereo or multichannel audio signal, each channel can be processed independently.

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

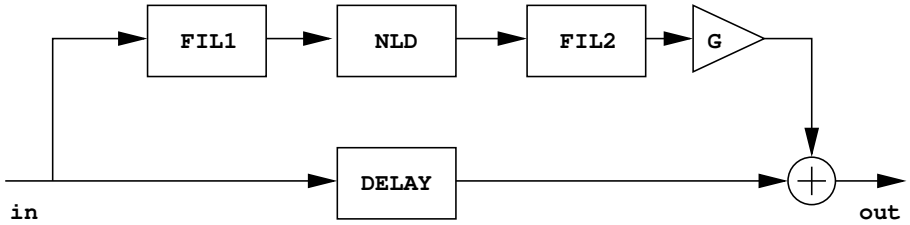

Fig. 7: Basic algorithm of the bandwidth extension framework, from Larsen *et al.* [3]. The top branch does the bandwidth extension, and is added back to the (possibly delayed) input signal.

the input signal with FIL1 to select a portion of the audio signal (indicated by letter the 'a' in Fig. 6). 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 by FIL2, to obtain a suitable spectrum and timbre (the signal now has frequencies in the range 'b' in Fig. 6). The resulting signal is amplified or attenuated as desired and mixed back with the main signal to form the ouput.

It is desirable that filters FIL1 and FIL2 are linear phase, as the addition of the synthetic and main signals at the last step of the processing can cause destructive interference if the relative phases of the two signals (with partially overlapping frequency spectra) are unspecified. Usually, FIR filters would be used to design linear phase filters, but for some bandwidth extension applications this will be very inefficient, as the bandpass regions can be a very small fraction of the sample rate, requiring inordinately high filter orders. Another option is to use linear phase implementations of IIR filters; efficient algorithms are described in Powell and Chau [21]. The delay accrued by the filters FIL1 and FIL2 can be matched by a delay in the main signal branch, such that both signals are added in phase.

Specific implementations for the NLD, which 'shifts' frequency components from low to high values, or vice versa, will be given in later sections covering specific bandwidth extension categories. They are based on generating harmonics or subharmonics of the signal passed by FIL1. Because naming (sub)harmonics can be ambiguous, we adhere to the following convention: a component at frequency kf_0 (integer k) is said to be the k-th harmonic of f_0. Thus, f_0 is its own first harmonic. A component at f_0/k is said to be the k-th subharmonic of f_0; thus, f_0 is its own first subharmonic.

All of the NLDs treated in the following sections will (implicitly) determine the frequency of the incoming signal by its zero crossings. 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 NLD will likely 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, and this is known as the *dominant frequency principle*. This principle can be made explicit by the *zero crossing spectral representation* (Kedem [22]) as

$$\cos \pi\gamma = \frac{\int_0^\pi \cos\omega \, \mathrm{d}F(\omega)}{\int_0^\pi \mathrm{d}F(\omega)}, \qquad (2)$$

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 Eq. (2), 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, NLD will 'detect' this frequency by the zero crossings of the signal, and construct (sub)harmonics of this dominant frequency.

The non-linear element NLD of the processing framework should be amplitude linear. This will ensure that the shifted frequency components covary in proportion with the original signal, which is of obvious importance in audio applications. Thus we desire that the NLD satisfy the homogeneity property of

linear systems; it will not satisfy the superposition property however. Thus, in Vaidyanathan's [23] terminology, our NLDs should be homogenous systems.

4. PHYSICAL LOW-FREQUENCY EXTENSION

4.1. Introduction

If a loudspeaker has a wider bandwidth than the audio signal, all frequencies in the audio signal can be reproduced at their correct level, and this may be considered an ideal situation. Especially at the low-frequency end of the spectrum it is desirable to use a loudspeaker with an extended frequency response, as the bass portion of an audio signal has a large influence on perceived quality. We could increase the quality of the audio signal by extending its low-frequency spectrum to fully exploit the bandwidth of the loudspeaker, although if this increases the quality is ultimately subjective. On the other hand, the audio signal might have been band-limited at a prior stage, and in this case, bandwidth extension is very desirable (e.g. telephony). In both cases we can use bandwidth extension algorithms to extend the low-frequency end of the audio spectrum.

4.2. Perceptual requirements

If the bandwidth extension is used to compensate for bandwidth limitations in storage or transmission channels, the algorithm should output a signal that is as close as possible to the original audio signal. For example, a voice signal with a bandwidth of about 7 kHz will be bandlimited to a bandwidth of about 3 kHz when passed through a telephone network. The bandwidth limitation occurs at both high and low frequencies. A bandwidth extension algorithm should then resynthesize frequency components to yield the original 7 kHz bandwidth voice signal. This algorithm will probably incorporate knowledge about the channel limitations, and statistics of the commonly received audio signals. The high-frequency extension will be dealt with in Sec. 6, here we focus on low-frequency extension.

If the received signal has the full original bandwidth, but the transducer's frequency response extends below that bandwidth, we may want to extend the reproduced bandwidth. Reasons for doing this are that additional (very) low frequencies add 'feeling' to the music, and may increase perceived quality. Ultimately, this is a subjective evaluation. Applications are in the 'effects' category, and may include cinema, home theater, automotive and gaming. In this case, there is no clearly defined 'target' for the bandwidth extension, and we must consider what characteristics the added frequency components must have in order to yield a pleasing output signal. It is obvious that the added components must covary in amplitude/loudness with the rest of the signal in order to be perceived as integral part of it, i.e. to be grouped with the original stream. In terms of frequency content, it is plausible that a harmonic relationship with the rest of the signal is desirable. Both of these requirements were discussed previously in Sec. 3.2. How this is implemented is not unambiguous, however. Suppose that the received signal consists of a fundamental f_0 and harmonics $i \cdot f_0$ ($i = 2, 3, 4, \ldots$). One option is to add a component at $f_0/2$, the second subharmonic of f_0. The resulting signal will now have as fundamental $f_0/2$ and all even harmonics $k \cdot f_0/2$ ($k = 2, 4, 6, \ldots$). Another option is to add components at $k \cdot f_0/2$ ($k = 1, 3, 5, \ldots$), thus generating a sequence with fundamental at $f_0/2$ and including all its harmonics (even and odd). It will be more difficult to construct a sequence with fundamental at $f_0/2$ and only the odd harmonics (at $k \cdot f_0/2$ ($k = 3, 5, 7, \ldots$) because that would involve removing the original harmonics at $m \cdot f_0$ ($m = 1, 2, 3, \ldots$). In any case, it is not *a priori* obvious which of these options is preferable, and the choice may be signal dependent (and subjective as well). Alternatively, we may want to generate third or fourth subharmonics, i.e. $f_0/3$ or $f_0/4$, with or without accompanying harmonics. It is clear that there is potentially a lot of design freedom for this kind of bandwidth extension.

4.3. Algorithm

We focus on low-frequency bandwidth extension applied to signals which are not bandlimited: the synthesized frequencies in this case are intended as audio 'effects', as explained before. The low-frequency bandwidth extension method follows the functional description as shown in Fig. 7. A possible simplication for multichannel systems is that all the channels can be mixed prior to the processing and distributed back to the output channels with equal amplitude, as very low frequencies are hard to localize. A reduction in complexity can thus be achieved. The

synthesized low frequencies are added back to the main signal, such that the entire signal is reproduced by the loudspeaker. A variation is possible where the synthesized frequencies are reproduced by a seperate loudspeaker, e.g. a subwoofer. Several ways to create a subharmonic at half the input frequency are: rectification, clipping, and integration (see Fig. 8). These are a few examples of the many options that are available for creating subharmonics. Which option is chosen will depend on the desired subharmonics spectrum (even, odd, or all harmonics). Given an input frequency of f_0, the method of integration yields a complex output spectrum with fundamental $f_0/2$, and all (even and odd) of its associated harmonics $(f_0, 3f_0/2, 2f_0, \ldots)$. An example in pseudo-code of such an integrator is

```
DEFINE x[N], y[N]; /* (input, output) */
DEFINE x_prev, y_prev = 0;
DEFINE c_1, c_2 = 0; /* (counters) */

WHILE c_1 < N
{
    READ x[c_1];
    IF (x[c_1] >= 0) AND (x_prev < 0)
    {
      IF c_2
      {
         y[c_1] = 0;
      }
      ELSE
      {
         y[c_1] = ABS( x[c_1] ) + y_prev;
      }
      c_2 = !c_2;
    }
    ELSE
    {
      y[c_1] = ABS( x[c_1] ) + y_prev;
    }
    x_prev = x[c_1]; y_prev = y[c_1]; c_1++;
}
```

The period doubling occurs because the variable `c_2` keeps track of zero crossings. The other subharmonics generators can be implemented in similar fashion.

The filter characteristics of `FIL1` should be chosen such that the subharmonics signal will extend below the original signal bandwidth. Although this actually requires a signal-dependent filter characteristic, it is convenient to fix the filter coefficients to yield a representative filter characteristic. For example, to create subharmonics $f_0/2$ below 80 Hz, `FIL1` should be a bandpass filter with cut-off frequencies of 80 and 160 Hz: the subharmonics will then lie between 40 and 80 Hz. If the $f_0/3$ subharmonic is desired, cut-off frequencies of 80 and 240 Hz should be chosen, in which case the synthesized frequency components will lie between 27 and 80 Hz.

`FIL2` should be designed in accordance with the frequency division of the input signal (e.g. above we used examples of division by 2 or 3, to yield either $f_0/2$ or $f_0/3$, for an input frequency of f_0). It also depends on the fact if *only* the subharmonic will be retained or if some of its harmonics are desired. If the frequency division factor is 2, and only the subharmonic will be retained, then (using the example of the previous paragraph) `FIL2` should be bandpass between 40 and 80 Hz. If at least one harmonic of $f_0/2$ should be included, then the bandpass becomes 40 to 160 Hz: if $f_0/2$ is 80 Hz, the second harmonic at 160 Hz is also included (note that for $f_0/2$ of 40 Hz, harmonics 2-5 are included). A lower limit of `FIL2` of about 40 Hz seems sensible from the point of view that 'musical' or 'melodic' pitch has a lower limit of about 30 – 40 Hz, as reported by Pressnitzer *et al.* [24], and references therein. Musical pitch is a sensation that can be used to identify certain musical intervals. Also, as stated before in Sec. 2, audio signals rarely have significant energy below 40 Hz. 'Regular' pitch has a lower limit of about 20 Hz, and around this frequency regions, sound can be felt if the level is high enough. Thus if these effects are desired, `FIL2` could be extended down to 20 Hz (going even lower might make sense if the derived signals are used to drive 'shakers', giving no audible but tactile sensations instead).

Thus, cut-off frequencies for `FIL1` and `FIL2` should be (k is frequency division factor, m is desired number of harmonic of f_0/k that should be included, f' is lower frequency limit of the input audio signal):

`FIL1`: $f' - kf'$ Hz,

`FIL2`: $f'/k - mf'$ Hz.

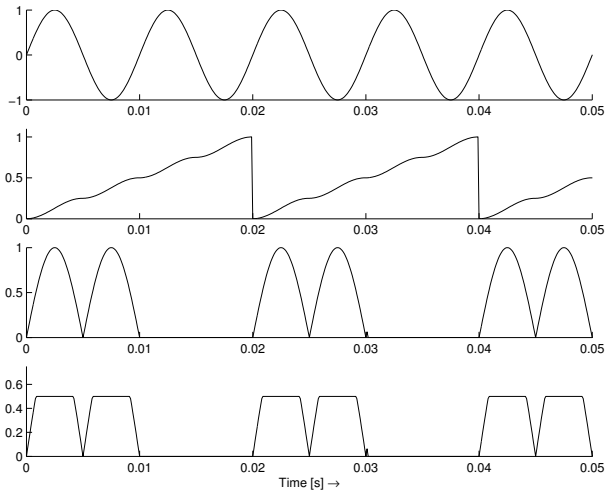

Fig. 8: Input and output signals for a specific implementation of a low-frequency bandwidth extension algorithm. The top panel is the sine wave input (100 Hz), the lower three panels are obtained by NLDs termed 'integrator', 'rectification', 'rectification–clipping'. Note that all the processed signals have a period twice as large as the input.

The high cutoff frequency of FIL1 does not necessarily need to be kf', but this makes sense as this will cause the highest frequency passed by FIL1 to shift down to f'. Filter order need not be high if IIR filters are used; an order of two for both high and low-pass flanks usually suffices. Note that linear phase implementation of these filters is desirable, as remarked in Sec. 3. Finally, the synthesized frequencies will be scaled to an appropriate level and mixed with the original input signal, or routed to a seperate loudspeaker.

Fig. 8 shows some examples of processing by NLD, where the input is the sine of the upper panel, and the output is any one of the three lower panels. The three NLD algorithms that were used to generate these outputs were integrator, rectification, and clipping. Also, Fig. 5 shows spectrograms of original and processed signals for a musical signal.

No formal listening test has been undertaken to assess the quality improvement obtained with these algorithms. However, numerous demonstrations and extensive listening by the authors indicate that this kind of bandwidth extension has great potential. Especially the fact that very low frequencies can add to the 'feel' of the music seems to be appreciated. This effect is most clearly noticeable in pop and rock music. Nonetheless, due to the freedom in designing the bandpass regions of FIL1 and FIL2, the method is scaleable to higher frequency ranges as well.

4.4. Other applications

Earlier in this section it was mentioned that low-frequency bandwidth extension can also be used to recover low frequencies that have been removed due to limitations in storage/transmission channels. In such applications the goal is to resynthesize as closely as possible the missing frequency components. Therefore there is less freedom in choosing the filter bandwidths and frequency division factor. The exact values for these parameters will depend on the application—the framework of Fig. 7 can be maintained, though. A confounding factor in these kinds of applications, e.g. telephony, is that it may not be unambiguous as to what frequency division factor to use. Recall from Sec. 2 that the male voice fundamental frequencies are distributed around 120 Hz, and around 240 Hz for the female voice. Furthermore, these values are time varying. The telephone channel lower cut-off frequency is 300 Hz, so the first harmonic to be passed by the network depends on the particular speaker. A pitch tracker can be used to determine the fundamental frequency, which then decides which frequency division factor to use in the NLD. For this, a very rough pitch estimate will suffice, and an efficient solution was proposed in Aarts *et al.* [25].

An interesting possibility is to choose the bandwidth of FIL2 such that the subharmonic f_0/k is *not* retained, but only its harmonics, which are then mixed back into the main signal stream. If the input signal is a harmonic complex with frequencies nf_0, $n = 1, 2, 3, \ldots$, the output signal will contain frequencies $nf_0/2$, $n = 2, 3, 4 \ldots$ if the particular implementation of the non-linear element generates all harmonics of $f_0/2$ (see earlier in this section). This harmonic complex evokes a residue pitch (explained in Sec. 5.2) at $f_0/2$, even though that particular frequency component is not present in the radiated signal. Effectively, we have created a low pitch percept for an audio signal that did not contain very low pitches, and reproduced the signal on a loudspeaker that is not capable of reproducing very low

frequencies! This technique is actually a different kind of bandwidth extension, called psychoacoustic bandwidth extension, and is the topic of the next section.

5. PSYCHOACOUSTIC LOW-FREQUENCY EXTENSION

5.1. Introduction

In many sound reproduction applications, it is not possible to use large loudspeakers, due to size and/or cost constraints. Typical applications are portable audio, multimedia, TV and public address systems, to name just a few. Hence, the devices are often of small size, and therefore the transducers are inherently small as well. At the same time we would like to obtain the highest possible audio quality of these products. However, probably the most well-known characteristic of small loudspeakers is a poor low frequency (bass) response. In practice, this means that a significant portion of the audio signal may not be reproduced (sufficiently) by the loudspeaker. For loudspeakers used in applications as mentioned above, reproduction below 100 Hz is usually negligible, while in some applications this lower limit can easily be as high as several hundred Hertz. The bass portion of an audio signal contributes significantly to the sound 'impact', and depending on the bass quality, the overall sound quality will shift up or down. Therefore a good low-frequency reproduction is essential.

Now, from psychoacoustic theory, we know that a pitch perception can occur at a frequency which is not contained in the audio signal. This is possible through non-linearities in the cochlea (difference tones), or a higher-level neural effect in the auditory system (virtual pitch). Thus through signal processing we can shift very low frequency components in the audio signal to higher frequencies, in such a way as to preserve the original pitch: psychoacoustic bandwidth extension.

As shown in Sec. 3, Fig. 7 presents the general processing scheme for bandwidth extension. We can also use such a scheme for psychoacoustic bandwidth extension, and we will describe the required filter characteristics and the non-linear device. As the system is 'merely' based on a psychoacoustic model of pitch perception, and uses loudspeaker characteristics in a very general sense (it is only assumed that reproducing lower frequencies is less efficient than reproducing higher frequencies), the method can be employed for any kind and/or size of loudspeaker.

5.2. Low pitch in the absence of low frequencies

Pitch is a subjective, psychophysical quantity. According to the American Standards Association pitch is "that attribute of an auditory sensation in terms of which sounds may be ordered on a scale extending from low to high". According to Moore [26], there are various ways how the pitch of a pure tone depends on its frequency. One can obtain a pitch–frequency relation by various methods, the classical result being the mel scale. It has an arbitrary pitch reference of 1000 mel at a frequency of 1000 Hz. A tone that sounds, on average, twice as high receives a value of 2000 mel, whereas a tone that sounds only half as high has a pitch of 500 mel. Although the mel scale suggests that the pitch of a pure tone is simply determined by its frequency, the perceived pitch also depends on some other factors, one being intensity. If one measures for a group of subjects how, on average, the pitch of a pure tone changes with the tone's intensity, one typically finds that (1) for tones below 1000 Hz the pitch decreases with increasing intensity (about 15%), (2) for tones between 1000 and 2000 Hz the pitch remains rather constant, and (3) for tones above 2000 Hz the pitch rises with increasing intensity (about 20%). This effect varies considerably between listeners and also depends on the duration of the tone. It is not immediately obvious what this implies for psychoacoustic low-frequency bandwidth extension applied to typical musical signals, where tones have different durations, envelopes and intensities. Therefore, we pretend that only the frequency of a pure tone determines its pitch. For a complex tone (which is more common in music than a pure tone), consisting of more than one frequency, the situation is more complicated. Pitch should then be measured by psychophysical experiments. A pitch that is produced by a set of frequency components, rather than by a single sinusoid, is called a *residue*. Even if the fundamental frequency is missing, it will still be perceived as a residue pitch, which in this case is sometimes called *virtual pitch*, because the frequency corresponding to the pitch is absent.

There is a vast literature on pitch perception and residue pitch (Bilsen and Ritsma [27], de Boer [28], Houtsma and Goldstein [29], Schouten [30, 31]).

Non-linearities in the cochlea can also generate pitch percepts that are lower than the frequency components received by the ear (Plomp [32]; Goldstein [33]). At high levels, the cochlea creates difference (or combination) tones, and if there are frequency components at f_0 and $3f_0/2$, a difference tone at $f_0/2$ will be generated; the effect only occurs at high levels though.

5.3. Algorithm

The psychoacoustic bandwidth extension algorithm described here can be be explained in terms of the signal processing framework of Fig. 7. The filter `FIL1` selects those frequency components in the audio that lie beneath the loudspeaker's lower cutoff frequency f_c; this value can be taken as the high-frequency cutoff of the filter. The low-frequency cutoff can be taken as 20 Hz, or at most three octaves below f_c. A larger bandwidth of the filter may result in excessive intermodulation distortion in the non-linear device that follows `FIL1`. In the non-linear device harmonics of the frequencies passed by `FIL1` are generated. There are several options to implement this function, such as rectification, clipping, or integration. Rectification of the input signal leads to frequency doubling, and thus doubles the perceived pitch. Although the original pitch is lost, the increased sound output may lead to increased quality. A better choice is a clipper, which will need to follow the input signal level such that low and high level signals are clipped at the same relative level. Clipping produces odd harmonics, and the perceptual effect is good. Waveforms for the clipper and rectifier are analogous to those in Fig. 8 (only the output is not set to zero for every other input period, as in that figure), modified so that the period of the output signal is the same as the period of the input signal. An integrator, analogous to the one described in Sec. 4 (but resetting once per period instead of once per two periods of the input signal), yields a very strong low pitch percept, as all harmonics of the input are reproduced. The filter `FIL2` is used to shape the frequency spectrum generated by the non-linear device. The lower cutoff of this filter is usually set to f_c (which was the high cutoff of `FIL1`), and the high cutoff of `FIL2` should be about an octave above this value. The resulting harmonics signal can be amplified and mixed back with the main signal. A variable gain can be used to prevent loudspeaker saturation at very high output levels.

6. HIGH-FREQUENCY EXTENSION

6.1. Introduction

Often it is desirable to extend the high-frequency bandwidth of an audio signal. This may be because at some point during the transmission from source to receiver the signal's bandwidth has been decreased; examples are telephone communication, perceptual coding at very high compression rates, etc. For speech, it has been established by the ITU [34] that wide-band speech (50 – 7000 Hz) is preferred over narrow-band speech (300 – 3400 Hz). For the case of music, a recent investigation by Zielinski *et al.* [35] showed that for a wide variety of repertoire, a wide bandwidth is perceptually more important than a realistic spatial impression. Thus there is ample motivation to design algorithms for high-frequency bandwidth extension; both for speech and audio. We continue to present the proposed algorithm.

6.2. Algorithm

The framework of Fig. 7 displays once again the proposed processing scheme. There are two signal branches, the lower of which passes the input signal unprocessed (possibly delayed). The spectrum extension takes place in the upper branch. The details of the processing steps specific to this category of bandwidth extension are:

1. Filtering by `FIL1`. Here, the highest octave present in the signal is extracted, say $\frac{1}{2}f_u - f_u$, where f_u is the upper frequency limit of the input signal.
2. Processing by `NLD`, the non-linear device. Here, harmonics are created. The first harmonic, which is just the fundamental, is in the frequency range $\frac{1}{2}f_u - f_u$; the second harmonic is in the frequency range $f_u - 2f_u$, the third harmonic is in the range $2f_u - 3f_u$, etc.
3. Filtering by `FIL2`. Here, the desired part of the complete harmonics signal is extracted. Typically, this will be the range of the second harmonic, thus f_u - $2f_u$.

As may be deduced from the above, the high-frequency limit of the output signal now equals $2f_u$, twice that of the input signal.

Depending on the application, the filters `FIL1` and `FIL2` may be fixed, or signal dependent. If the bandwidth of the incoming signal is not known *a priori*, bandwidth detection must be used, which may be used to adapt the filter characteristics (this can occur for example in Internet radio). Such bandwidth detection may be based simply upon detecting the input signal's sample rate f_s and assuming the bandwidth to be $\frac{1}{2}f_s$. Alternatively, a more complex bandwidth detection means may be used. If for a given sample rate f_s we have that $f_u > \frac{1}{4}f_s$, an upsampler must be used before the processing scheme of Fig. 7, otherwise it is not possible to extend the signal's spectrum by a complete octave.

As the object is to add only the next highest octave to the input spectrum, a non-linear device that generates mainly the second harmonic is preferred. Also, amplitude linearity is desirable. A full-wave rectifier has both these characteristics and is therefore highly suitable for use as non-linear device in the scheme of Fig. 7 as applied to high-frequency bandwidth extension. A side-effect of non-linear processing is that beside harmonic frequencies also intermodulation distortion is introduced. In some situations this can give rise to audible artefacts. An analytic expression for the frequency spectrum of an arbitrary full-wave rectified signal is given in Larsen and Aarts [2], through which it is possible to analyze the strength of harmonic and inharmonic components.

7. CONCLUSIONS

We presented an analysis of bandwidth extension methods, leading to a discrimination along three dimensions: low/high frequency, physical/psychoacoustic based extension, blind/non-blind. Interesting applications were shown to exist for all coordinates along these dimensions, and a signal processing framework has been defined which has been used to design specific bandwidth extension algorithms (three were presented, of which one new). This framework has been validated by a statistical analysis of various kinds of audio signals.

8. REFERENCES

[1] A. Gröschel, M. Schug, M. Beer, and F. Henn. Enhancing audio coding efficiency of MPEG Layer-2 with Spectral Band Replication for DigitalRadio (DAB) in a backwards compatible way. In *Proceedings of the AES 114th Convention, Amsterdam.* AES, 2003.

[2] E. Larsen and R.M. Aarts. Reproducing low-pitched signals through small loudspeakers. *J. Audio Eng. Soc.*, 50(3):147–164, 2002.

[3] E. Larsen, R.M. Aarts, and M. Danessis. Efficient high-frequency bandwidth extension of music and speech. In *proceedings of the 112th AES Convention (Munich, Germany).* Audio Eng. Soc., May 2002.

[4] S. Furui. *Digital speech processing, synthesis and recognition.* M. Dekker, 1989.

[5] Sound Quality Assessment Material, (recordings for subjective tests). European Broadcasting Union, 1988. no. 422 204-2.

[6] D. Byrne *et al.* An international comparison of long-term average speech spectra. *J. Acoust. Soc. Am.*, 96(4):2108–2120, 1994.

[7] P. Boersma and D. Weenink. http://www.fon.hum.uva.nl/praat/, Univ. of Amsterdam, The Netherlands, Retrieved Jul. 2003.

[8] L.J. Sivian, H.K. Dunn, and S.D. White. Absolute amplitudes and spectra of certain musical instruments and orchestras. *J. Acoust. Soc. Am.*, 2(3):330–371, January 1931.

[9] R.A. Greiner and J. Eggers. The spectral amplitude distribution of selected compact discs. *J. Audio Eng. Soc.*, 37(4):246–275, April 1989.

[10] L.D. Fielder and E.M. Benjamin. Subwoofer performance for accurate reproduction of music. *J. Audio Eng. Soc.*, 36(6):443–456, June 1988.

[11] M. Grob-Da Veiga. String instrument sound enhancing method and apparatus. USPTO 5,218,160, Jun. 8, 1993.

[12] M. Oda. Music tone pitch shift apparatus. USPTO 5,131,042, Jul. 14, 1992.

[13] S.-E. Tan, W.-S. Gan, C.-W. Toh, and J. Yang. Application of virtual bass in audio cross-talk cancellation. *IEEE Elec. Letters*, 36(17):1500–1501, 2000.

[14] W.S. Gan, S.M. Kuo, and C.W. Toh. Virtual bass for home entertainment, multimedia PC, game station and portable audio systems. *IEEE Trans. Cons. Elec.*, 47(4):787–793, 2001.

[15] J.D. Griffiths. Apparent bass and nonlinear distortion. *IRE Trans. Audio*, 9(4):117–121, 1961.

[16] R.M. Aarts, E. Larsen, and D.W.E. Schobben. Improving perceived bass and reconstruction of high frequencies for band limited signals. In *Proc. first IEEE Benelux workshop on model based processing and coding of audio (MPCA-2002, Louvain, Belgium)*, pages 59–71. IEEE, Nov. 2002.

[17] R. Taori, R.J. Sluijter, and A.J. Gerrits. Hi-BIN: An alternative approach to wideband speech coding. In *ICASSP*, pages 1157–1160. IEEE, 2000.

[18] S. Chennoukh, A.J. Gerrits, G. Miet, and R.J. Sluijter. Speech enhancement via frequency bandwidth extension using line spectral frequencies. In *ICASSP*, pages 665–668. IEEE, 2001.

[19] A. Kohlrausch and A.J.M. Houtsma. Pitch related to spectral edges of broadband signals. *Phil. Trans. R. Soc. Lond. B*, 336:81–88, 1992.

[20] A.S. Bregman. *Auditory scene analysis.* MIT Press, 1990.

[21] S.R. Powell and P.M. Chau. A technique for realizing linear phase IIR filters. *IEEE Trans. on Sign. Proc.*, 39(11):2425–2435, 1991.

[22] B. Kedem. *Time series analysis by higher order crossings.* IEEE Press, New York, 1994.

[23] P.P. Vaidyanathan. Homogeneous Time-Invariant Systems. *IEEE Signal Proc. Letters*, 6(4):76–77, 1999.

[24] D. Pressnitzer, R.D. Patterson, and K. Krumbholz. The lower limit of melodic pitch. *J. Acoust. Soc. Am.*, 109(5):2074–2084, 2003.

[25] R.M. Aarts, J. Smeets, and P.C.W. Sommen. Bandwidth extension of narrow-band speech. In *Proceedings of 2nd IEEE Benelux Signal Processing Symposium (SPS-2000, Hilvarenbeek, The Netherlands).* IEEE, 2000.

[26] B.C.J. Moore. *Handbook of perception and cognition: Hearing.* Academic Press New York, 1995.

[27] F.A. Bilsen and R.J. Ritsma. Some parameters influencing the perceptibility of pitch. *J. Acoust. Soc. Am.*, 47(2 (Part 2)):469–475, 1970.

[28] E. de Boer. *On the 'residue' in hearing.* PhD thesis, University of Amsterdam, 1956.

[29] A.J.M. Houtsma and J.L. Goldstein. The central origin of the pitch of complex tones: Evidence from musical interval recognition. *J. Acoust. Soc. Am.*, 51(2 (Part 2)):520–529, 1972.

[30] J.F. Schouten. The perception of pitch. *Philips Technical Review*, 5(10):286, 1940.

[31] J.F. Schouten, R.J. Ritsma, and B. Lopes Cardozo. Pitch of the residue. *J. Acoust. Soc. Am.*, 34(8 (Part 2)):1418–1424, 1962.

[32] R. Plomp. Detectability threshold for combination tones. *J. Acoust. Soc. Am.*, 37(6):1110–1123, 1965.

[33] J.L. Goldstein. Auditory nonlinearity. *J. Acoust. Soc. Am.*, 41(3):676–689, 1967.

[34] Study group 12. Paired comparison test of wideband and narrow-band telephony, 1993. ITU COM 12-9-E.

[35] S.K. Zielinski, F. Rumsey, and S. Bech. Effects of bandwidth limitation on audio quality in consumer multichannel audiovisual delivery systems. *J. Audio Eng. Soc.*, 51(6):475–501, 2003.