IMPROVING PERCEIVED BASS AND RECONSTRUCTION OF HIGH FREQUENCIES FOR BAND LIMITED SIGNALS

Ronald M. Aarts and Erik Larsen and Daniël Schobben

Philips Research Laboratories
Eindhoven, The Netherlands
daniel.schobben@philips.com, ronald.m.aarts@philips.com

ABSTRACT

Bandwidth extenstion methods are required in systems that playback bandlimited signals (e.g. telephone, MP3) or that are not capable of reproducing signals with a large bandwidth (small or inexpensive loudspeakers). The first part of this paper deals with reproducing low pitched signals through small loudspeakers. This work is based on psychoacoustic phenomena to invoke a deep bass impression using only higher frequency components. The second part of this paper deals with high-frequency bandwidth extension of music and speech. In this method, no additional information on the original wide band content is required. It aims at providing a more pleasant and brighter sound impression.

Part I - Reproducing low pitched signals through small loudspeakers[1]

Ever since the invention of the electrodynamic loudspeaker, there has been a need for greater acoustical output, especially at low frequencies. From a manufacturer's point of view, since a long time it has been desirable to reduce the size of the loudspeaker (and cabinet). These two demands are physically contradictory. It is the aim of part I of the paper to offer options to evoke the illusion of a higher low-frequency response of the loudspeaker, while the power radiated by the loudspeaker at those low frequencies remains the same, or is even lower. This is feasible by exploiting certain psychoacoustic phenomena. The required non-linear signal processing is studied for a number of specific implementations. An elaborate analysis of the outcome of a listening test, aimed at assessing the subjective evaluation of the presented system is also presented.

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. Needless to say, the just mentioned competitive market also dictates 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.

A traditional and conceptually very simple method to increase the perceived sound level in the lower part of the audible spectrum (below the loudspeaker's resonance frequency, which is usually the lower limit) is to amplify the low frequency part of the audio spectrum, by a fixed or dynamic (depending on signal amplitude and/or reproduction level) amount. A special system that purposefully drives a loudspeaker below resonance is the 'ELF' system, described in Long and Wickersham [2,3]. From an efficiency point of view these methods are unfavourable, but an even more serious problem is the high cone excursion at low frequencies (quadrupling for every octave down in frequency). For very low frequencies, the mechanical limits of the loudspeaker will limit the stroke the cone can make, leading to distortion and possibly loudspeaker overload. Thus, physically increasing the radiated sound pressure level means forcing the loudspeaker to radiate sound in a frequency range for which it is not equipped. It may be better to prevent this completely, by methods outlined below. In the process we shall discover several advantages of these methods.

Because the radiation characteristics of (small) loudspeakers are at the core of the topic discussed in this paper, we shall review these characteristics in terms of the loudspeakers parameters. These characteristics are discussed in greater detail in the Appendix of of [1] and in e.g. Beranek [4], Olson [5] or Borwick [6] for extensive reviews).

In the normal operating range of the loudspeaker (above resonance and below the transition frequency, where the wavelength becomes roughly equal to the cone diameter), the efficiency η is proportional to

$$\eta \propto \frac{S^2}{m_t^2}, \tag{1}$$

where S is the cone area and m_t the total moving mass. So a high efficiency requires a large cone area and a small moving mass, which is difficult to be realised simultaneously. The resonance frequency ω_0 of the loudspeaker, which is just a simple mass-spring system, equals

$$\omega_0 = \sqrt{\frac{k_t}{m_t}}, \tag{2}$$

where k_t is the total spring constant (cone suspension and influence of enclosed air volume), and is in part inversely proportional

[1] This part of the paper is adapted from [1].

to the cabinet volume. Any attempt to lower the resonance frequency by increasing m_t will decrease the efficiency, unless cabinet size and/or cone area are increased as well, but this is not possible for a small loudspeaker. More intricate cabinet designs employing (multiple) ports or passive radiators behave somewhat differently in the low-frequency range, but do not overcome the basic problems of good low-frequency reproduction. It is precisely Eqns. 1 and 2 that, for a small loudspeaker, prevent a good and extended low-frequency response; good in the sense of a high radiated sound pressure level and a high efficiency. Obviously, all the qualifications of the previous sentence are relative, but they can be seen in the light of the applications mentioned above and what is expected by a 'demanding' listener.

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). In Sec. 2 we review these two effects, which appear to be very suitable effects for our purpose of enhancing bass perception using small loudspeakers. Basically, this works by some simple non-linear (but controlled) processing, replacing very low frequencies in the audio signal by higher frequencies. These will still have the same perceived pitch of the original, using the psychoacoustic effects previously mentioned. Such effects also occured in transistor radios, where undesired non-linearities gave rise to a distorted sound. However, the method that we now propose uses non-linearities in a controlled manner, and restricted to only the lowest frequencies, such that the effect is to our benefit. Without any information about the signal processing employed, we can immediately infer a number of advantages that such a scheme shall provide:

1. High radiated sound pressure level, because of increased efficiency and decreased cone excursion. Furthermore, at higher frequencies the auditory system is more sensitive, which will also contribute to increased loudness.
2. Less power consumption, because of increased efficiency. This can be very important for portable applications.
3. Less disturbance in neighbouring areas, because of increased absorption of high frequencies in structures.

We propose a signal processing scheme, consisting of a few basic and efficient operations, to achieve such a psychoacoustic bass enhancement, see Fig. 1. This method was already introduced by

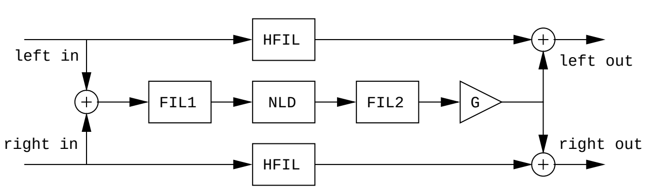

Figure 1: *Signal processing for psychoacoustic bass enhancement. The input signal is summed and filtered to obtain the bass portion. Then harmonics are created and added to left and right output signals. In the direct path a high-pass filter is implemented.*

Aarts *et al.* [7–10]. The concept of this processing was already mentioned: to prevent the radiation of very low frequencies, by instead radiating higher frequency components, leading to the same pitch. Preventing radiation of very low frequencies is done by high-pass filtering the direct signal path through filters HFIL, see Fig. 1. To compensate for this, the input signals are summed and the low-frequency part is obtained through filtering by FIL1. The next step is to create higher harmonics of this bass portion, by processing through a non-linear device (NLD). These harmonics are filtered once again by FIL2, to obtain a suitable spectrum, amplified with a gain of G and added to left and right output signals. The resulting output signals now do not contain low-frequency elements, but do have the low pitch associated with the input (given that the input signals contain low frequencies). Note that the processing as outlined above may also be performed for each channel seperately, if so desired. As low-frequency content is usually identical in left and right signal channels, this is usually not necessary. Furthermore, localization for low frequencies is known to be very poor, so a change in low-frequency balance should not be detected. In Sec. 3 we will discuss in more detail the important aspects of the processing scheme proposed in Fig. 1.

As is apparent now, we are focusing on time domain instead of frequency domain methods. In the frequency domain we would have the familiar problems such as connection of consecutive output frames, spectral leakage for frequencies which are not harmonic to the DFT window and non-stationarity of the input signal in any one frame. By focusing on the time domain we circumvent these problems. Furthermore, in the time domain we can make implementations which are much more efficient computationally, and we can even devise simple analog circuitry to do the processing.

About 50 years ago, a circuit based on the same concept was announced in Radio Electronics [11]. The circuit was used in a radio receiver and used the non-linearity of the tubes to create odd harmonics of the low-frequency part of the spectrum. The concept has also been discussed by Ben-Tzur and Colloms [12]. However, the focus of that discussion is on achieving equal loudness of the output signal containing the synthetic bass, with respect to the input signal. We will comment some on this aspect in Sec. 2, but in this paper we shall focus more on creating the synthetic bass itself.

Through many informal listening tests and demonstrations of the system it was soon obvious that our method is well appreciated. To show this more formally, and also to be able to compare various implementations amongst one another in a reliable way, a listening test was conducted. The description of this test and the presentation of the results will be found in Sec. 4

The key element of the processing scheme shown in Fig. 1: the non-linear device, where the harmonics are created. To analyze what exactly happens to the signal spectrum here is a difficult task. Because of the non-linear behaviour, the operation performed on the signal cannot be represented by a transfer function; every single frequency component in the output spectrum depends on all frequency components of the input spectrum. However, we have devised a compact formulation for representing the output spectrum in terms of the input spectrum, for all possible signals, in continuous and discrete time.

2. LOW PITCH IN THE ABSENCE OF LOW FREQUENCIES

There are several options to increase the perception of low pitch based on psychoacoustic events. In this section we discuss three such options.

2.1. Frequency doubling

In previous studies [8,13–15] a frequency doubler was introduced, whereby signals in the low frequency band also appear one octave higher. It can be considered as one of the options for increased low pitch perception. The mentioned method, shown in Fig. 2-a has extreme simplicity as an advantage. By means of a simple non-linear element, e.g. a diode, frequencies between, say, $1/2f_0$ and f_0, the frequency region where the loudspeaker does not radiate sufficient power, are shifted to $f_0 - 2f_0$. A drawback is that the pitch has been changed. Furthermore, impulsive sounds with a high low frequency content are seriously distorted. Nevertheless, informal listening tests have shown that this frequency doubling can be an improvement, and an implementation according to this concept is also included in the listening test described in Sec. 4.

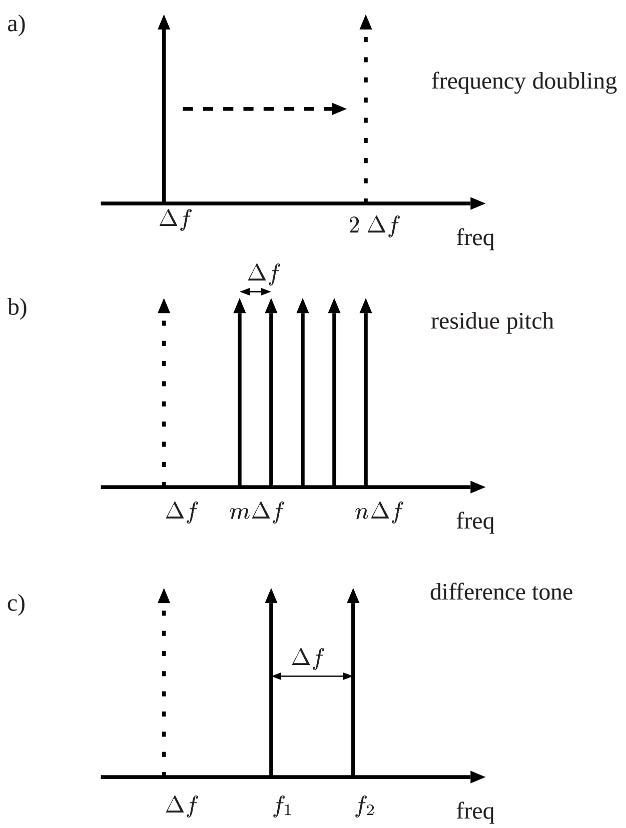

Figure 2: *Possible options for psychoacoustic bass enhancement. The dotted frequency component denotes the perceived pitch (but is not necessarily acoustically radiated).*

2.2. Virtual pitch

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". For a pure tone, where the fundamental frequency corresponds to the frequency of the tone, the pitch is unambiguous and— if we neglect the influence of sound level on pitch— one can identify pitch with the frequency of the pure tone. For a complex 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, see Fig. 2-b, rather than by a single sinusoid, is called a *residue*. In Fig. 2-b the fundamental frequency is missing, yet will still be perceived as a residue pitch, which in this case is also called *virtual pitch*. The psychoacoustic phenomenon responsible for this effect is the 'missing fundamental' effect. There is long history of investigations into pitch perception, also regarding virtual pitch. Famous are the experiments of Seebeck in 1843, and the controversy of him with Ohm; see Plomp [16] for a historical review. There is a vast amount of literature on this topic, just a few interesting references are e.g. [17–21]. As the frequency of a pure tone decreases to very low values, say under 100 Hz, the pitch becomes more difficult to determine. This is also true for the missing fundamental effect, and because our algorithm is aimed at this very low frequency range, we need psychoacoustic data regarding the perception of virtual pitch for this range. Unfortunately, only sparse data is available. The work of Ritsma [22, 23] investigates the existence region of the tonal residue, for frequencies above 200 Hz.

2.3. Difference tone

We will illustrate how difference tones are generated by means of an example, involving organ pipes. As in any other instrument, the frequency reproduced by an organ pipe has a direct (inverse) relationship with the size of the pipe. Thus, for very low notes, very long pipes are required. Now, if there is not enough space (in a church) for a pipe long enough to produce such very low notes, one can combine two higher notes to get a similar perceptual effect. The cause is that non-linearities of the auditory system, within the cochlea, produce the difference tone of these two higher notes. This principle was — according to Helmholtz [24] — discovered in 1745 by Sorge, a German organist; the tones thus achieved are often known as Tartini's tones. Since the end of the sixteenth century, many organs include a stop (the "$5\frac{1}{3}$-foot fifth") composed of pipes sounding a fifth higher than the pitch of the actual note as played from the musical score. The purpose is to stimulate or reinforce the bass one octave below the pitch of the actual note (that is, to reinforce the 16-foot sound of the organ). Of older use, according to Roederer [25], is the use of the $10\frac{2}{3}$-foot fifth in the pedals, which in combination with 16-foot stops, evokes the 32-foot bass. Roederer attributes this to residue pitch (Fig. 2-b), however, this is doubtful, since the effect of residue pitch decreases very fast for low frequencies. Consequently, it is probably due to difference tones, see Fig. 2-c. This is illustrated, as an example, in Table 1, showing which frequencies are obtained for the mentioned pipes. The acoustical bass concept for organs has also been considered by Terhardt and Seewann [26]; see also Terhardt [27], which includes a concise overview of various aspects of pitch perception.

2.4. Considerations on loudness effects

Loudness perception depends strongly on the frequency of the stimulus that is presented to the ear(s). This is illustrated in Fig. 3, showing the equal loudness-level contours for steady-state pure tones. For the bass frequency range, there are two important observations. Firstly, the hearing threshold increases sharply for low

Table 1: Example of the acoustical bass of an organ. The length of the two organ pipes in feet (m) and their corresponding individual frequencies. The perceived pitch is Δf, which would require a very long pipe.

f	L		freq. [Hz]	key
Δf	$32'$	(9.75 m)	16.35	C_0
f_1	$16'$	(4.9 m)	32.70	C_1
f_2	$10\,\frac{2}{3}'$	(3.3 m)	49.05	G_1

frequencies. Secondly, the level contours lie close together, implying that small changes in sound level lead to large changes in loudness level.

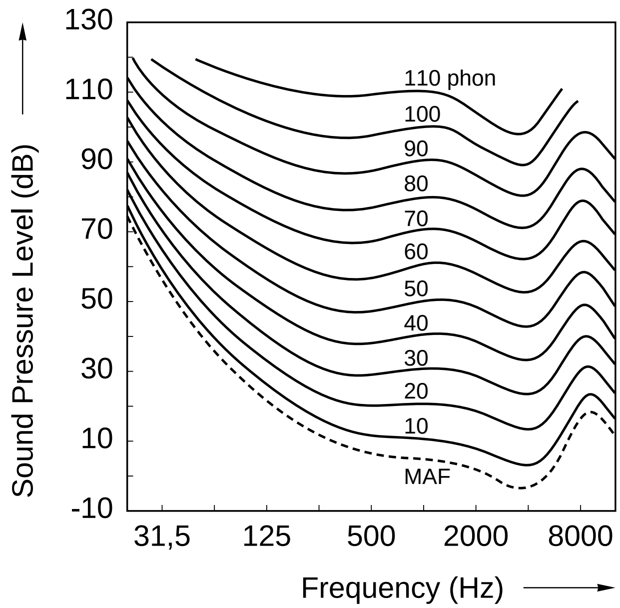

Figure 3: *Equal-loudness level contours for pure tones (binaural free-field listening, frontal incidence), from [39] Fig. 1.*

Relating these properties of the auditory system to the processing scheme of Fig. 1, we can make a few *a priori* observations on loudness effects we can expect to occur. Firstly, because low frequencies are replaced by higher frequencies, the perceived loudness (of the total signal, but in particular of the bass range) will increase, due to the lower hearing threshold. This is beneficial, as it is our goal to maximize the perceived low-frequency reproduction level. Although no formal verification of this assumption has been made, it is quite apparent from informal listening that the loudness does increase, even if the harmonics signal has equal amplitude as the original bass signal. Secondly, the 'dynamic loudness range' of the perceived bass frequencies will be decreased, due to the greater separation of the equal loudness-level contours at higher frequencies. The effect might be opposed, or reversed, by introducing a dynamics processing of the harmonics signal in Fig. 1, and Ben-Tzur and Colloms [12] have proposed such, based on the contours of Fig. 3.

Another psychoacoustic effect that we need to consider is a timbral change of the bass perception when we replace the original bass by harmonics. Although with our proposed method, the pitch should remain identical, the timbre or sound color will be altered after processing. Another cause of a timbral change might be the interference of the generated harmonics with frequency components in the original signal. The phase relationship between these original components and the synthesized harmonics are not clear *a priori* and therefore the effect this interference will have on the perception is hard to predict.

3. GENERATING HARMONICS

3.1. Processing scheme

As previously stated, Fig. 1 presents the general processing scheme that we propose for psychoacoustic bass enhancement. In this section, we will go into more detail regarding the choice of the filters 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. Therefore, emphasis will mostly lie on how to choose the filters in relation to the loudspeaker that is used. In the Appendix of [1] a complete design for a specific reproduction set is presented.

3.1.1. High-pass filter - `HFIL`

The high-pass filter, in the direct signal path (both left and right for a stereo signal), is used to prevent low-frequency reproduction by the loudspeaker. The primary purpose is to limit cone excursion. Since, if this cone excursion is well controlled, we will prevent distortion when large input signals are used, or a high output volume is selected. The maximum cone excursion that will be attained depends not only on the input signal or volume level, but also on the gain of the harmonics signal that the system adds, and the frequency range of `FIL1`. The larger this range, or the higher the gain, the larger the cone excursion will become. Thus, the attenuation rate of the high-pass filter must be tuned to suit these parameters (it is our experience that a fourth order IIR filter is certainly sufficient). In less demanding cases the filter might be completely omitted, however. The cut-off frequency of the high-pass filter can be equal to the resonance (or cut-off) frequency of the loudspeaker, because obviously, below this frequency the louspeaker's efficiency decreases drastically. In particular, we should choose this cut-off frequency equal to the cut-off frequency of the low-pass flank of band-pass filter `FIL1`, as we shall see in Sec. 3.1.2.

3.1.2. First filter - `FIL1`

This filter extracts frequencies from the left and right input signals of which harmonics will be created in the non-linear device. This filter usually is a band-pass filter. For the cut-off frequency of the low-pass flank of `FIL1`, we choose the cut-off frequency of the loudspeaker. This ensures that all frequencies which are inefficient to reproduce (i.e. below the cut-off frequency) will be passed to the non-linear device or harmonics generator. This cut-off frequency can also be chosen lower if desired— this will in general give a somewhat deeper bass impression of the output signal— at the expense of lower efficiency. The high-pass flank has a cut-off frequency of, say, 20 Hz. This serves merely to prevent DC signals from entering the non-linear device. This is necessary, because the output of the non-linear device depends on *all* components of its input. As a DC signal serves no purpose in a reproduction system, we need to ensure no DC signal enters the non-linear device. Only if the cut-off frequency of the low-pass flank lies substantially above 150 Hz, leading to a bandwidth of `FIL1` of over three octaves, should we choose the cut-off frequency of the high-pass flank higher. At a bandwidth of approximately three octaves or more, the input signal for the non-linear device contains a too wide frequency range, resulting in too much intermodulation distortion, which would become clearly audible in the reproduced output signal. In such cases the cut-off frequency of the high-pass flank of `FIL1` can be tuned to suit the loudspeaker. The low- and high-pass flanks are usually third or fourth order IIR.

3.1.3. Non-linear device - `NLD`

The non-linear device, or harmonics generator, 'shifts' signal components in a low frequency range to a higher frequency range. The pitch of the input signal is preserved, because the components in the higher frequency range are harmonics of the original components. As discussed in Sec. 2, the preservation of the original low pitch is due to the virtual pitch (or difference tone effect) of the harmonics signal. Because this element is a non-linear device, any single output component depends on all input components. Moreover, at the output, frequency components will be generated, which are not present at the input. This is a desired effect, since this is how the harmonics are obtained. However, it also leads to sum and difference components, which are not desired, for they are not harmonically related to the input signals.

There are many non-linear functions available to create harmonics, but for our purposes, they should be computationally simple (and/or achievable with analog circuit elements). Another desirable characteristic is amplitude linearity: the operation of the device does not depend on the amplitude of the input signal. For example, an input–output relationship that has a quadratic or higher polynomial element is in principal unsuitable. For such non-linearities, the 'amount' of non-linearity generally increases with increasing magnitude of the input signal. This will lead to weak harmonics for small input signals and strong harmonics for large input signals. For the auditory system, an opposite effect would be more desirable. At low levels, we need strong harmonics (strong non-linearity of the non-linear device), because the bass frequencies have a very small loudness at low levels. At higher levels, loudness increases sharply for bass frequencies, so we do not need very strong harmonics in such cases. In Secs. 3.2 and 3.3 some specific examples are presented.

3.1.4. Second filter (BP2)

This filter serves to extract a suitable spectrum from the output of the non-linear device. The cut-off frequency of the high-pass flank is chosen equal to the cut-off frequency of the low-pass flank of `FIL1`. The value of the cut-off frequency of the low-pass flank of `FIL2` depends somewhat on the choice of the non-linear device. If its value is too high, the harmonics spectrum is wide, and has a very sharp timbre, which does not sound very pleasant. Thus, usually, this value is chosen roughly one octave above the cut-off frequency of the high-pass flank. Depending on the implementation of the non-linear device, this value has to be tuned. Orders for the low- and high-pass flanks are again third or fourth order IIR.

3.2. Full wave rectifier

Let us know focus on the non-linear device (NLD) of Fig. 1. A very simple non-linear device is a full wave rectifier, i.e. a device which gives the absolute value of the input. For a sine-wave input, only even harmonics are generated. So for a pure tone of frequency f_0, we get at the output frequency components of $2f_0$, $4f_0$, $6f_0$ etc. (and a non-interesting DC component as well), with a decay of 12 dB / octave. Expressions for the output spectrum in terms of the input spectrum for arbitrary periodic inputs are given in the Appendix of [1]. Fig. 4 shows the input (solid line) and output (dashed line) signals for this situation. The output spectrum is indicated in Fig. 5. The fundamental component of the out-

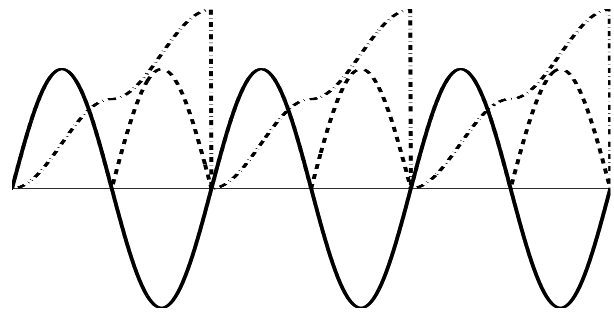

Figure 4: *The solid line represents a sine wave input. The dashed line is obtained by full wave rectification, the dash-dotted line by full wave integration.*

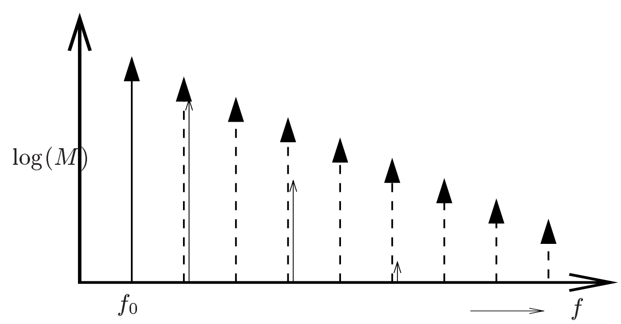

Figure 5: *The solid arrow at f_0 represents the input spectrum. The arrows at $2f_0$, $4f_0$, etc., represent the output spectrum for the full wave rectifier, the dashed arrows at $2f_0$, $3f_0$, etc., represent the output spectrum for the full wave integrator.*

put is $2f_0$, which is not equal to the original pitch. Therefore, this non-linear device cannot be an ideal element for our purpose, but nonetheless, due to its simplicity, it may be an option. Informal listening tests have indicated that in many cases the bass enhancement obtained using a rectifier is an enhancement over the original.

3.3. Full wave integrator

To create a deeper bass impression, corresponding to the pitch of the original fundamental, we must create a harmonics signal containing all (odd and even) harmonics. This is possible by using a 'full wave integrator'. This non-linear device integrates the absolute value of the input, and the output is reset to zero when the input has a zero crossing with a positive slope. For a sine-wave input the situation is shown in Fig. 4, the full wave integrator output is shown as the dash-dotted signal. Fig. 5 shows the input and output spectra: the output contains all harmonics which decay relatively slowly.

To explore what happens when the input signal is not a single frequency, but a more realistic (musical) signal, consider that the locations of the zero crossings are very important (as the output signal is reset to zero at every zero crossing with positive slope) and determine the fundamental frequency of the output signal, which should ideally be the same as the fundamental frequency of the input signal. If we consider a weakly stationary time series with spectral distribution $F(\omega)$, we can then write, as is shown in Kedem [28],

$$\cos \pi\gamma = \frac{\int_0^\pi \cos\omega \, \mathrm{d}F(\omega)}{\int_0^\pi \mathrm{d}F(\omega)}, \tag{3}$$

where γ is the expected zero crossing rate of the signal, i.e. the expected number of zero crossings in a certain interval (which is e.g. 2 per period for a sine). This is called the *zero crossing spectral representation* and it expresses the tendency of $\pi\gamma$ to be attracted to a specific frequency (band), if this frequency (band) is dominant in the signal. This is a well known empirical fact known as the *dominant frequency principle*. It shows us that we can be confident that if an input signal has a dominant frequency component, this will be reflected in its zero crossings. Consequently, this frequency component will be the fundamental frequency of the output of the full wave integrator. Exact expressions for the output as function of an arbitrary periodic input signal are given in the Appendix of [1].

4. SUBJECTIVE EVALUATION

In order to gain insight into the subjective appreciation of the psychoacoustic bass enhancement, a formal listening test was conducted. Although subjects may differ in what aspects they consider when determining 'best bass quality', this does have the benefit of obtaining an overall assessment, and is therefore deemed to be quite useful.

Besides this obvious reason for performing a listening test, there was another reason. At an early stage of this research, it was recognized that while for many repertoire the psychoacoustic bass enhancement gave very pleasant results, in some cases artefacts can occur, which could even lead to a degredation of bass quality. In most cases, such artefacts can be classified as bass sounds acquiring an unnatural timbre, or excessively strong or prolonged bass notes. This last case might not actually be an artefact, but merely the unusual effect of a seemingly very good bass reproduction on a small loudspeaker. When the same piece of music was played back over a good subwoofer, the same prolonged bass note could be observed. Two examples of such repertoire are "Hotel California" by The Eagles (which is one of the tracks used in the listening test), and "That's the way it is" by Phil Collins. What must be clear now, is that for some repertoire it is questionable if the applied processing will improve the subjective quality of the music; we call this 'difficult' repertoire. In the listening test described below we have mainly used such difficult repertoire.

4.1. Experiment description

4.1.1. Music selection

As the system is meant primarily to enhance music reproduction, it is an obvious choice to use some typical music fragments in the listening test. Using musical fragments makes the listening test quite difficult. Instead of testing for one specific auditory attribute like loudness, pitch or even speech intelligibility, music appreciation may be seen as a combination of multiple attributes. When testing a bass enhancement system, the result would almost certainly be influenced by the fact if the test subject likes music with a high bass quantity or not. It is also felt that appreciation of any kind of music depends to some extent on the mood of the person at the moment of listening.

First, the material to be used in the test was selected. Obviously, this should be music with sufficient deep bass. The genre of music to be used in the test was rock and pop – it is in these genres that bass enhancement systems find their most frequent application. Three of the four tracks used fall into the category of difficult repertoire, as described above. Only track 1 is not difficult repertoire. The following four tracks were used:

1. Bad, Michael Jackson. This track contains a typical pop bass line.
2. My Father's Eyes, Eric Clapton. A very deep and strong bass line accompanies the music here, which might soon sound too imposing when enhanced.
3. Hotel California (live version), The Eagles. In the first part of this track a bass drum is heard, with some ambient sounds. The fact that besides the bass drum, few other instruments are heard, could lead to an over-emphasis of the bass after processing.
4. Twist and Shout, Salt n' Peppa. The bass in this track is a tight bass beat, which is difficult to preserve after processing.

4.1.2. Music preparation

Each track will be processed in various ways, so as to create a number of different versions for each track. We use the following four processing methods:

1. No processing. The unprocessed version will be included as a reference for evaluations afterwards. However, the subjects will not be informed that one of the four versions is the original unprocessed fragment. Therefore, the subjects have no reference with which to compare the processed versions.
2. Linear bass boost. To compare the psychoacoustic processing with a traditional bass enhancement processing method.
3. Psychoacoustic bass enhancement with rectifier as harmonics generator.
4. Psychoacoustic bass enhancement with integrator as harmonics generator.

The details of how the processing for each of the four mentioned versions was achieved can be found in the Appendix of [1].

4.1.3. Listening test procedure

To ensure a fair and valid result, the experiment is performed randomized 'double blind', i.e. the order of presentation of the pairs is completely randomized (with respect to track order and versions of the track), and neither subject nor experiment leader knows which versions are being presented.

Fifteen subjects participated in the listening test, 11 male, 4 female. Most subjects were in the age group of 25-30, and had normal hearing, with varying degrees of experience in listening tests. Some of the subjects had no experience at all, while some may be termed experts. The subjects were also asked to indicate their general music preference (it was thought this might have some influence on their results). This revealed that pop and rock music was by far the most popular music category of the subjects. They were encouraged to jot down any comments they could have during or after the listening test.

4.2. Listening test results

In Fig. 6 we present a boxplot of the results obtained from the listening test as described previously. The distribution of scores

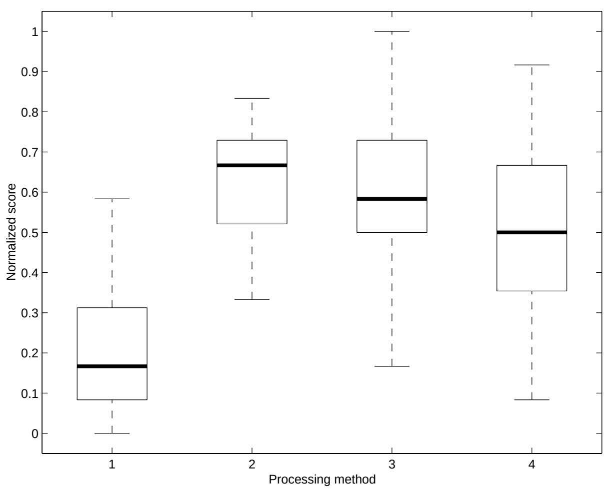

Figure 6: *Distribution of normalized scores per processing method, averaged over all four tracks. Processing methods: 1. Unprocessed, 2. Linear bass boost, 3. Psychoacoustic with rectifier, 4. Psychacoustic with integrator.*

for the four versions is indicated, where we have summed the scores over all four tracks. The thick horizontal line indicates the median value for each version, and the bottom and top of the box indicate the 25% and 75% quartiles of the distribution. The whiskers extending from the bottom and top of the box show the remaining spread in results. Those unfamiliar with the boxplot display method are referred to Tukey [30] for a comprehensive treatment. It appears obvious that all three bass enhancement methods are ranked significantly higher than the unprocessed version. Although the median values for the three bass systems do differ, the spread in the ranking per system is quite large. The spread in results for the two psychoacoustic systems is larger than for the linear bass boost.

4.2.1. Analysis by multi-dimensional scaling

The overall result of the experiment does not give us insight into how the individual responses lead to this result. It will be interesting to investigate if there is any structure in the distribution of subject responses. For this we can use multi-dimensional scaling, or in short MDS.

MDS is a technique to assess similarity between objects, especially useful if the attribute(s) of the objects is (are) vague or difficult to judge. Through some kind of procedure one determines the *proximities* between the objects, which can represent either similarity or dissimilarity, and the output will be a spatial map, where the objects are represented as locations in an n-dimensional space. In case the proximities are dissimilarities, like objects will be positioned close together, and unlike objects far apart. An important parameter of the acquired mapping is the stress value, which indicates how well the distances in the mapping correspond to the true distances (proximities). A comprehensive treatment of MDS can be found in Kruskal and Wish [31], while Schiffman *et al.* [32] present a concise introduction to MDS, including a glossary of commonly used terminology and a treatment of several programs that can be used for MDS calculations.

In our case, the objects to analyze are the subject responses, and as proximities we have taken the dissimilarity between any two subjects. Fig. 7 presents the two dimensional mapping of the fifteen subjects.

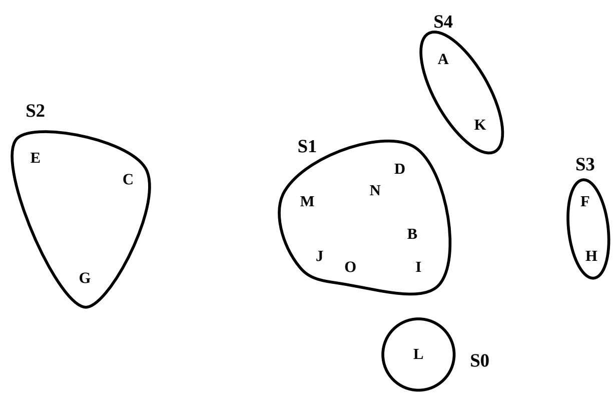

Figure 7: *Two dimensional scaling result for the pooled subject responses. Stress value is 0.0216. Note the five subgroups indicated by S0 – S4.*

We observe that the fifteen subjects are distributed in a number of subgroups, which have been indicated by numbers in Fig. 7. The largest subgroup, S1, containing seven subjects, may be considered as having the most typical responses in this test. The second largest subgroup is indicated by S2, and these are subjects C, E and G, which all have a small preference for the two versions of psychoacoustic bass enhancement. Moving on to subgroup S3, including subjects F and H; most obvious in their results is the fact that as only two subjects of all fifteen, they have never preferred an unprocessed version over another version. Furthermore, they prefer the two types of psychoacoustic bass enhancement very clearly over a linear bass enhancement, which is not the case for the "average" subgroup S1. From these three subgroups it appears that the position along the horizontal dimension indicates an appreciation of psychoacoustic bass enhancement. Subjects oriented towards the left have a clear preference for linear bass enhancement, subjects on the right clearly prefer psychoacoustic bass enhancement. Those in the middle have no clear preference. Also, since the amount of bass attainable with versions three and four is greater than with version two, the horizontal dimension may represent a more general preference for bass. From left to right along the horizontal dimension, the preferences of the subjects appear to shift to those versions which contain a louder bass. Three subjects remain: A, K and L. For subject L, subgroup S5, we observe that the score for version three is much higher than typical, while for subjects A and K (subgroup S4) version four seems to have a higher preference. So along the vertical dimension it appears that we can observe an indication of which kind of psychoacoustic bass enhancement is preferred (since versions three and four each use a different non-linear device to generate higher harmonics). It may be noted that the spread along the vertical dimension is smaller than the spread along the horizontal dimension. From this we might conclude that the associated attribute is less important for the subject responses.

5. CONCLUSIONS

In this paper we have proposed a psychoacoustically based signal processing system to enhance the perceived bass response of a loudspeaker below its cut-off frequency. The main concept of this system is to replace very low frequency components by their harmonics, through controlled non-linear processing. The resulting harmonics yield the same (virtual) pitch as the original signal, due to the missing fundamental effect. Beneficial characteristics of the system include:

- (Very) low radiation of energy below loudspeaker cut-off frequency.
- Less headroom required compared to a traditional (linear) bass boost, for a comparable bass enhancement effect.
- Computationally very efficient, simple circuit in case of analog design.
- Power efficient.
- Tuneable to any kind and size of loudspeaker.

Some drawbacks are:

- Intermodulation distortion can lead to audible artefacts. Careful tuning of filters will be beneficial.
- The added harmonics interfere with frequency components in the original audio, which in some cases may alter the timbre of the signal to some extent.

A listening test has been performed, with fifteen subjects participating. The psychoacoustic bass enhancement system in two different versions was compared against a system with a linear bass boost. Unprocessed music was included as a reference. Both psychoacoustic systems as well as the linear bass boost system were appreciated significantly higher than the unprocessed audio. A large variation in subject responses has been observed. This has been further analyzed by means of multi-dimensional scaling The visualization method has indicated a number groups, out of the total subject pool, which have similar in-group preferences. Personal preference for any kind of processing appears to be highly individual, where some subjects prefer no processing at all.

6. ACKNOWLEDGEMENTS

The authors would like to thank their colleagues Catherine Polisset and Erik van der Tol for their parts in the work presented in this paper. We would also like to acknowledge Guido Janssen for providing great help in deriving the expressions for the non-linearities and Jan Engel for contributions in the evaluation of the listening test results. We greatly appreciate the participation of all people in the listening test.

Part II - Efficient high-frequency bandwidth extension of music and speech[2]

The use of perceptually based (lossy) audio codecs, like MPEG 1 – layer 3 ('mp3'), has become very popular in the last few years. However, at very high compression rates the perceptual quality of the signal is degraded, which is mainly exhibited as a loss of high frequencies. We propose an efficient algorithm for extending the bandwidth of an audio signal, with the goal to create a more natural sound. This is done by adding an extra octave at the high frequency part of the spectrum. The algorithm uses a non-linearity to generate the extended octave, and can be applied to music as well as speech. This also enables application to fixed or mobile communication systems.

7. BACKGROUND

Often it is desirable to extend the bandwidth of an audio (music or speech) 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 ratex, etc. For speech, it has been established by the ITU [40] that wide-band speech (50 – 7000 Hz) is preferred over narrow-band speech (300 – 3400 Hz). For music the perceptual difference may be even larger, although a formal test has not been done. In this paper we shall consider only high-frequency bandwidth extension; Aarts *et al.* [41] describe an algorithm for low-frequency bandwidth extension, with a focus on speech applications. In the following we shall address more precisely the objectives and constraints of our research, followed by a description of the proposed bandwidth extension algorithm.

7.1. Objectives and constraints

The objective is to extend the bandwidth of the reproduced sound by synthesizing and adding additional high frequency components to the received low-bandwidth audio signal, complying with the following constraints:

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

[2] This part of the paper is adapted from [48].

4. No *a priori* knowledge about the missing high frequencies.

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 model to derive high-frequency information from low-frequency information. If the application was limited to speech only, a speech model could be used, which would allow an accurate reconstruction of the original wide-band speech signal. For a music signal, which is not well described by such a speech model, the extension obtained would not be correct however. Accurate reconstruction of a special signal class, say speech, could also be achieved by a training phase, where narrow-band and wide-band speech signals are both available to the system. It would then be possible to initialize a set of parameters which after completion of the training phase, would allow estimates of the wide-band signal to be made for a given input narrow-band signal, using the parameters obtained in the training phase. However, in order 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.

The fourth constraint implies the algorithm is 'blind', i.e. there is no information available to the system to aid the high-frequency reconstruction process. It must be clear that this constraint prevents a perfect reconstruction of the original full-bandwidth signal. The motivation of designing a blind system, which in principle performs worse than a non-blind system, is to be able to apply the system on any existing audio source.

The constraints mentioned above prevent a (near)-perfect reconstruction of the original full-bandwidth signal. The bandwidth-extended signal may be expected to deviate considerably from the original full-bandwidth signal. Therefore, we can restate the objective of the algorithm accordingly: to create a signal with an extended bandwidth, that sounds more pleasant than the received low-bandwidth signal.

High-frequency bandwidth extension has also been proposed as a means to enhance virtual acoustic systems, as described in Dempsey [42]. It is argued that a wider bandwidth signal would give more localization cues and therefore a listener would be better able to position the virtual acoustic source.

7.2. Other methods

A simple way to increase high-frequency content is linear filtering, i.e. equalization. Only in cases where the signal-to-noise ratio of the high frequencies is sufficiently high this can be an option, but in the remainder we shall assume this not to be the case. In other situations the sample frequency may be low, such that the Nyquist frequency is well below the signal's natural bandwidth (e.g. in telephony, where the Nyquist frequency is 4 kHz). In such cases, first an upsampler and low-pass filter will be used, followed by bandwidth extension. Because the extra octave obtained by upsampling contains aliased components of the baseband signal, equalization is not an option.

An advanced method for obtaining a wide-band signal from a narrow-band signal is described in Liljeryd [43], which forms the basis for the 'mp3pro' method. Here, 'spectral band replication' is used as means for bandwidth extension. The lower frequencies are coded in a known way (in this case according to the old 'mp3' method); the higher frequencies are derived from these lower frequencies. However, of the four constraints mentioned earlier, only one is valid for the method of Liljeryd, namely it is applicable to music and speech. The other three constraints are not met. Firstly, since 'mp3pro' is a codec with a specific signal format, the decoding part only works on a correctly encoded signal. Furthermore, the bandwidth extension is aided by information embedded in the encoded bitstream by the encoder, which makes the system non-blind. The advantage of this non-blind, special encoding is that the final signal perceptually matches the original full-bandwidth signal very closely. It is therefore possible to obtain a higher audio quality than possible with the method described in this paper. Finally, the computational complexity of the complete system (encoder and decoder) is considerable.

A method optimized for creating a wide-band speech signal from a narrow-band speech signal is described in Valin and Lefebvre [44]. Both the additional low band (50 – 300 Hz) as well as the high band (3400 – 7000 Hz) are derived from the narrow-band signal. The lower band is synthesized by means of a sinusoidal oscillator, the fundamental frequency of which is determined by a pitch tracker and the amplitudes of which are determined by a multilayer perceptron network which has been trained in an initialization phase. The higher frequency band is synthesized using nonlinear processing and LPC filtering. The coefficients of the LPC filter are transmitted along with the speech signal, making the method non-blind. It is reported that the high-frequency band is perceptually very close to the original, whereas the low-frequency band has audible artefacts. Several variations on this processing scheme are described in the literature.

A simple narrow-band to wide-band speech converter is described in Yasukawa [45]. The method described has low complexity, uses no training and is also independent of signal format. It can be seen as a 'special purpose' (speech-only) implementation of the algorithm presented below. The subjective performance is reported to be good.

8. ALGORITHM DESCRIPTION

The processing scheme for efficient high-frequency bandwidth extension is based on techniques proposed by Larsen and Aarts [1] for low-frequency bandwidth extension on small loudspeakers. There, harmonics are generated of very low-frequency signal components which can not be reproduced on a small loudspeaker. The generated harmonics will give the listener a pitch perception corresponding to the original very low-frequency components, even if these very low-frequency components are not present in the reproduced signal any more. The extension is therefore only perceptual. This psychoacoustic phenomenon is known as the 'missing fundamental'. For the present case of high-frequency bandwidth extension, a similar processing scheme can be used. Harmonics will be generated from a part of the input signal spectrum, which will then be used to extend the input signal's bandwidth. The extension thus obtained will most typically be one octave, although more or less would be possible. In contrast to the low-frequency bandwidth extension mentioned before, for the high-frequency case no psychoacoustic effect is used to create the additional high-frequency

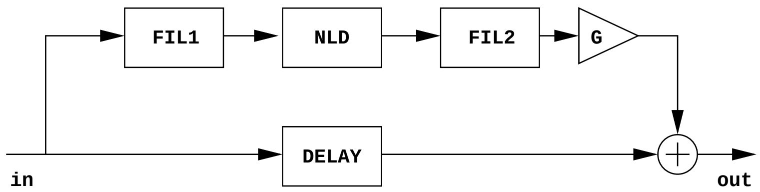

Figure 8: High-frequency bandwidth extension.

percept; instead, there is a measurable extension of the signal's spectrum.

8.1. Processing details

Fig. 8 displays 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 following processing steps are taken:

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 - $2f_u$.
4. Scaling by gain G.
5. Addition to the (delayed) input signal. This delay is used to compensate for delays occurring due to the filtering in the previous steps.

As may be deduced from the above, the high-frequency limit of the output signal now equals $2f_u$, double 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. 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. 8, otherwise it is not possible to extend the signal's spectrum by a complete octave.
The non-linear device `NLD` is the element that creates the additional high frequencies to the output spectrum. 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, because the system should add the same amount of harmonics to the signal, independent of signal level. A full-wave rectifier has both these characteristics and is therefore highly suitable for use as non-linear device in the scheme of Fig. 8. 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.

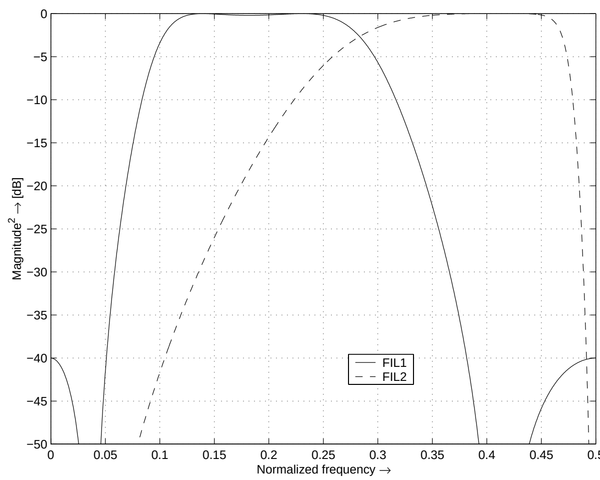

Figure 9: Squared filter magnitudes, `FIL1` and `FIL2`.

An analytic expression for the frequency spectrum of an arbitrary full-wave rectified signal is given in Larsen and Aarts [1], through which it is possible to analyze the strength of harmonic and inharmonic components.
Preferably the filters `FIL1` and `FIL2` in Fig. 8 are linear phase filters. If also an appropriate delay is used in the lower branch, the two signal branches will add exactly in phase. This has the advantage that transients in the input signal will remain compact in the output (because of the filters' constant group delay), which is beneficial for perceptual quality. Also, the lower and upper signal branch may have some spectral overlap. If in this overlapping region the signals from the upper and lower branch do not add in phase, interference may cause an amplitude modulation of the signal spectrum, which is undesirable. Therefore, filters `FIL1` and `FIL2` should be either FIR filters, or linear phase IIR filters (using time-forward and time-reversed filtering), which may be more efficient. Powell and Chau [46] present an efficient method for linear phase IIR filtering.
We focus on time instead of frequency domain implementations. In the frequency domain we would have familiar problems such as connection of consecutive output frames, spectral leakage for frequencies which are not harmonic to the DFT window and non-stationarity of the input signal in an input frame. In the time domain we prevent these problems.

8.2. Example

As an implementation example, consider a signal with a sample rate of f_s' and bandwidth $\frac{1}{2}f_s'$. To extend the bandwidth of this

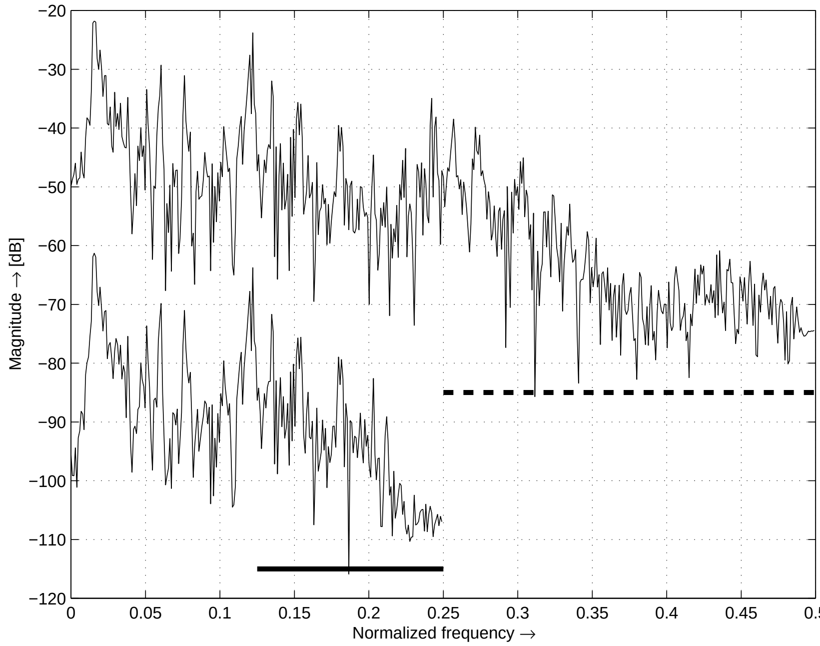

Figure 10: Input and bandwidth extended spectra. The horizontal lines indicate the passbands of `FIL1` and `FIL2`. The input spectrum is offset by 40 dB.

signal, first an upsampler is used with an upsample factor of 2, yielding a sample rate of $2f_s' = f_s$. The squared magnitude response of `FIL1` and `FIL2` are shown in Fig. 9, on a frequency axis normalized with respect to f_s. The squared magnitude is plotted, because each filter is used twice, once filtering in forward time and once in reversed time (see [46]). `FIL1` is a second order elliptic filter, `FIL2` is a second order Butterworth filter. The value of G, the harmonics scaling factor (see Fig. 8), is 0.5. The delay in the lower branch is chosen such that it exactly matches the added delay of `FIL1` and `FIL2`.

A 10 ms frame containing a musical signal is bandwidth extended according to the implementation described above. Fig. 10 shows both input (offset by 40 dB) and output spectra, on a frequency axis normalized with respect to f_s. The solid horizontal line indicates the passband of `FIL1` and the dashed horizontal line indicates the passband of `FIL2`.

Apart from this specific example, the filter characteristics can be adapted to any feasible frequency range and thus accommodate any input signal bandwidth. The algorithm can also be applied to any sound reproduction system by adapting the value of the harmonics gain G. The value of G also depends on the listener's preference, and on the listener's hearing threshold at high frequencies. This high-frequency hearing threshold is highly correlated with age. Fig. 11 displays the average hearing loss (with respect to threshold of hearing, see ISO [47]) for male subjects from 20 to 70 years of age. The average hearing loss for a 70 year old male at 8 kHz is 60 dB.

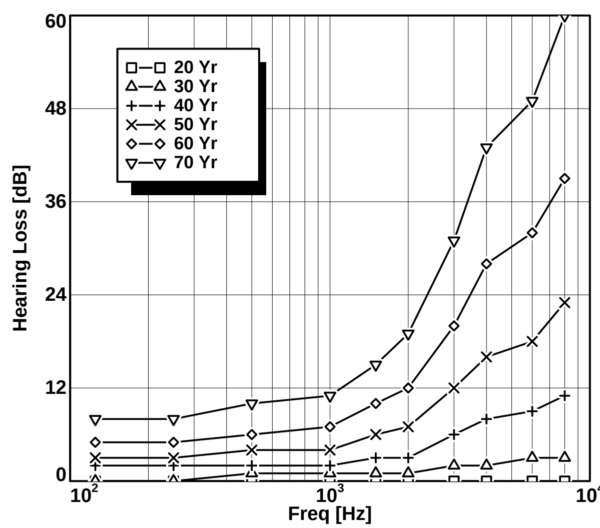

Figure 11: Hearing loss (with respect to threshold of hearing) for a group of otologically normal males of various age. For each age 50% of the group has a higher hearing loss and 50% has a lower hearing loss.

9. PERCEPTUAL EVALUATION

Subjective quality assessments have been made through informal listening. As mentioned at the beginning of this paper, the extended signal deviates considerably from the original full-bandwidth signal. But in practice the listener will not have the original full-bandwidth signal available, and therefore has no reference signal.
For speech (narrow-band to wide-band conversion) the extended signal is experienced to be pleasant. Extension of unvoiced fricatives, such as /s/, /f/ and /ch/ is not very good due to the relative low amount of energy these sounds contain in the narrow-band frequency range.
For musical signals the quality of the bandwidth extended signal depends somewhat on the original signal's bandwidth. To a certain extent the quality decreases as the input bandwidth decreases; the lowest useable input bandwidth being roughly 4 kHz. However, in most practical situations audio bandwidths smaller than 4 kHz do not occur. For input bandwidths of 8 kHz or larger the perceptual effect is very pleasant. For all bandwidths the signal's transients are most effected and improved by the system.
Since in many cases the quality of the output signal is perceived to be enhanced relative to the input signal, the bandwidth extension method presented in this paper offers a practical and feasible solution to the problem of high-frequency audio bandwidth extension.

9.1. Evaluation with a masking model

As was mentioned before, the bandwidth extension is obtained by use of a non-linear element. This non-linearity generates harmonics of a part of the input signal spectrum. However, intermodulation distortion will also occur, which will lead to frequency components not harmonic to the input signal's spectrum being generated. Given that these inharmonic components have sufficient amplitude, they will become audible, leading to dissonance.
By using a pscychoacoustic masking model it would be possible to gain insight into the audibility of the intermodulation distortion components. Through this it should be possible to determine if the artefacts sometimes occurring in the bandwidth extended signals are due to these intermodulation distortion components, or if they have a different origin. As observed above, the artefacts become stronger as the input signal's bandwidth decreases. Using artificial as well as real-life signals in combination with a masking model, we might determine if this is caused by properties of the auditory system or by the frequency-dependent statistics of real-life signals. These issues are a current research topic.

10. CONCLUSIONS

We have presented a method for efficient high-frequency bandwidth extension of music and speech signals. This method complies with the following constraints:

1. Low computational complexity: the largest part of the computational burden consists of two low-order IIR filters, and is therefore quite small.
2. Independent of signal format. No special encoding or decoding is required.
3. Applicable to music and speech. Perceptual evaluations have shown that the method enhances the quality of music and speech signals in most cases for input signal bandwidth of 4 kHz and larger.
4. No *a priori* knowledge about the missing high frequencies. The only assumption is that the high frequencies are harmonically related to the lower frequencies.

Furthermore, the algorithm can easily be adapted to work on any input signal bandwidth by merely changing the two filter characteristics and possibly the harmonics gain value.
The presented method offers a practical and feasible solution for high-frequency audio bandwidth extension. Because of its flexible design many applications are possible.

11. ACKNOWLEDGEMENT

The authors acknowledge Jo Smeets and Derk Reefman for stimulating discussions and pleasant cooperation.

12. REFERENCES

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

[2] E.M. Long and R.J. Wickersham. Method and apparatus for operating a loudspeaker below resonant frequency, 1984. US 4,481,662. Filing year 1982.

[3] E.M. Long and R.J. Wickersham. Guide to ELF systems. A new era in bass reproduction, 1997. `http://www.bagend.com`.

[4] L.L. Beranek. *Acoustics*. American Institute of Physics, New York, 1954. 1986 reprint for the Acoust. Soc. Am.

[5] H.F. Olson. *Acoustical Engineering*. Van Nostrand, Princeton, New Yersey, 1957.

[6] J. Borwick, editor. *Loudspeaker and headphone handbook*. Butterworths, London, 1988.

[7] R.M. Aarts. *25 jaar AES in Nederland*, chapter 17, pages 158–159. Gebotekst, Zoetermeer, The Netherlands, 1999.

[8] R.M. Aarts and S.P. Straetemans. Circuit, audio system and method for processing signals, and a harmonics generator, Aug 29, 2000. US patent 6,111,960.

[9] G.F.M. De Poortere, C. Polisset, and R.M. Aarts. Ultra Bass, Oct 17, 2000. US patent 6,134,330.

[10] E. Larsen and R.M. Aarts. Perceiving low pitch through small loudspeakers. In *preprint of the 108th AES Convention Paris*. Audio Eng. Soc., February 2000.

[11] Sonora model RCU-208. A synthetic bass note circuit. *Radio Electronics*, 37:37, 1948.

[12] D. Ben-Tzur and M. Colloms. The effect of MaxxBass psychoacoustic bass enhancement on loudspeaker design. In *preprint 4892(F3) of the 106th AES Convention Munich*. Audio Eng. Soc., May 1999.

[13] W. Schott. Low frequency audio doubling and mixing circuit, 1991. European patent EP0546619.

[14] M. Oda. Low frequency audio conversion circuit, 1997. US patent 5,668,885.

[15] T. Unemura. Audio circuit, 1998. US patent 5,771,296.

[16] R. Plomp. *Experiments on tone perception*. PhD thesis, University of Utrecht, 1966.

[17] 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.

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

[19] 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.

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

[21] 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.

[22] R.J. Ritsma. Existence region of the tonal residue I. *J. Acoust. Soc. Am.*, 34(9):1224–1229, 1962.

[23] R.J. Ritsma. Existence region of the tonal residue II. *J. Acoust. Soc. Am.*, 35(8):1241–1245, 1963.

[24] H. Helmholtz. *Die Lehre von den Tonempfindungen*. Friedr. Vieweg & Sohn, Dover, 1913. The first printing: 1863, English translation of the 4th German edition of 1877.

[25] J.G. Roederer. *Introduction to physics and psychophysics of music*. Springer, New York, 1975.

[26] E. Terhardt and M. Seewan. Der 'akustische Bass' von Orgeln. In *Fortschritte der Akustik DAGA84*, pages 885–888, Darmstadt, 1999. DeGA.

[27] E. Terhardt. Perception of auditory pitch. http://www.mmk.e-technik.tu-muenchen.de//persons/ter.html.

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

[29] H.A. David. *The method of paired comparisons*. Ch. Griffin, London, 2nd edition, 1988.

[30] J.W. Tukey. *Exploratory data analysis*. Addison-Wesley Publishing Company, Reading, MA, 1977.

[31] J.B. Kruskal and M. Wish. *Multidimensional scaling, Quantitative applications in the social sciences*. Sage Publ., 1978.

[32] S.S. Schiffman, M.L. Reynolds, and F.W. Young. *Introduction to multidimensional scaling*. Academic Pres, New York, 1983.

[33] R.M. Aarts. A new method for the design of crossover filters. *J. Audio Eng. Soc.*, 37(6):445–454, 1989.

[34] J.B. Kruskal, F.W. Young, and J.B. Seery. *How to use KYST, a very flexible program to do multidimensional scaling and unfolding*. Bell Laboratories, Murray Hill, N.J., 1973.

[35] J.B. Kruskal. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1):1–27, 1964.

[36] K.R. Gabriel. The biplot graphical display of matrices with application to principal component analysis. *Biometrika*, 58:453–467, 1971.

[37] Syntrillium. Syntrillium Software Corporation. http://www.syntrillium.com.

[38] K.R. Gabriel and C.L. Odoroff. Biplots in biomedical research. *Statistics in Medicine*, 9:469–485, 1990.

[39] International standard ISO 226-1987(E), 1987. Acoustics – normal equal-loudness level contours.

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

[41] R.M. Aarts, J. Smeets, and P.C.W. Sommen. Bandwidth extension of narrow-band speech, Proceedings of 2nd IEEE Benelux Signal Processing Symposium (SPS-2000), Hilvarenbeek, March 23–24, 2000.

[42] M.J. Dempsey. Method for introducing harmonics into an audio stream for improving three dimensional audio positioning, 2001. US 6,215,879.

[43] L.G. Liljeryd. Source coding enhancement using spectral-band replication, 1998. Application WO 98/57436.

[44] J.-M. Valin and R. Lefebvre. Bandwidth extension of narrowband speech for low bit-rate wideband coding. In *IEEE Speech Coding Workshop*, 2000.

[45] H. Yasukawa. Signal restoration of broad band speech using nonlinear processing. In *proceedings of EUSIPCO*, pages 987–990, 1996.

[46] 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.

[47] International standard ISO 7029-1984E, Acoustics – Threshold of hearing by air conduction as a function of age and sex for otologically normal persons, 1984.

[48] E. Larsen, R.M. Aarts and M. Danessis. Efficient high-frequency bandwidth extension of music and speech, *112th AES Convention*, Munich, Germany, May 2002.