Audio Engineering Society
Convention Paper 5627
Presented at the 112th Convention
2002 May 10–13 Munich, Germany

Efficient high-frequency bandwidth extension of music and speech

Erik Larsen[1], Ronald M. Aarts[1], and Michael Danessis[2]

[1] *Philips Research, Prof. Holstlaan 4, 5656 AA Eindhoven, The Netherlands*

[2] *The University of Salford, Salford, Greater Manchester, M5 4WT, United Kingdom*

Correspondence should be addressed to Erik Larsen (`erik.larsen@philips.com`)

ABSTRACT

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.

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 [1] 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.* [2] 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.

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.
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 [3]. 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.

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 [4], 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 [5]. 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 multi-layer perceptron network which has been trained in an initialization phase. The higher frequency band is synthesized using non-linear 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 [6]. 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.

ALGORITHM DESCRIPTION

The processing scheme for efficient high-frequency bandwidth extension is based on techniques proposed by Larsen and Aarts [7] 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

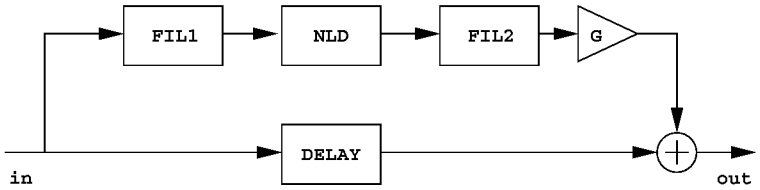

Fig. 1: High-frequency bandwidth extension.

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

Processing details

Fig. 1 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. 1, 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. 1. 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 [7], through which it is possible to analyze the strength of harmonic and inharmonic components.

Preferably the filters `FIL1` and `FIL2` in Fig. 1 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 [8] 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.

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 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. 2, 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 [8]). `FIL1` is a second order elliptic filter, `FIL2` is a second order

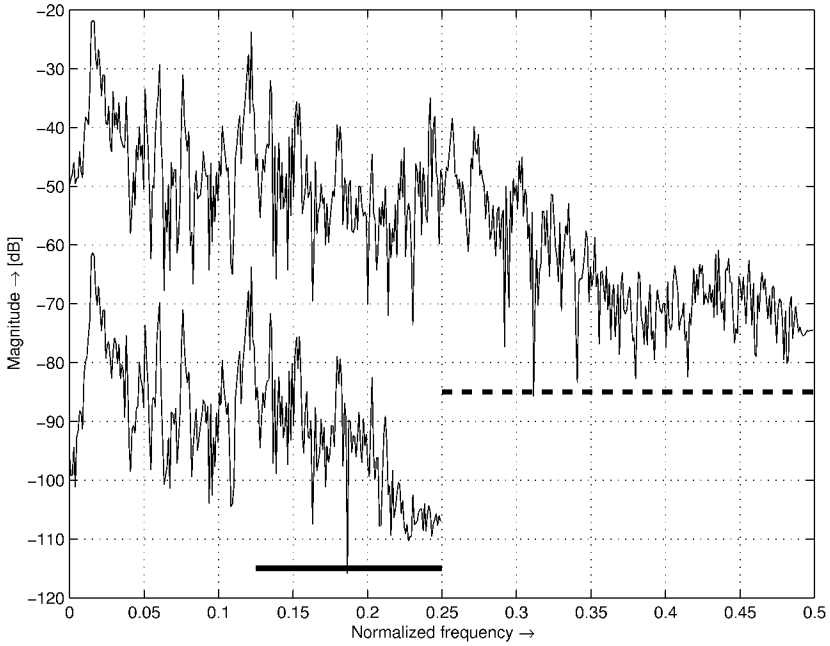

Fig. 3: Input and bandwidth extended spectra. The horizontal lines indicate the passbands of `FIL1` and `FIL2`. The input spectrum is offset by 40 dB.

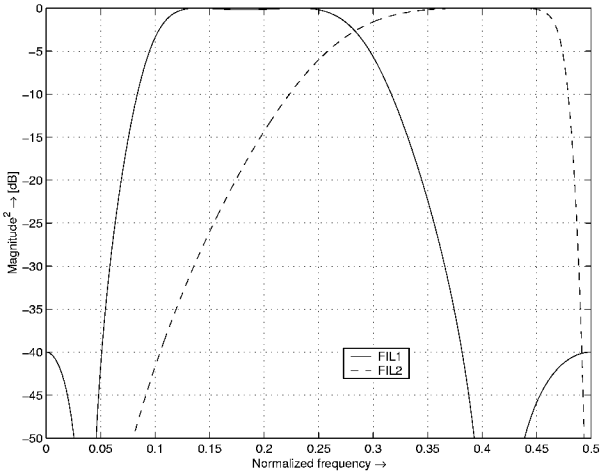

Fig. 2: Squared filter magnitudes, `FIL1` and `FIL2`.

Butterworth filter. The value of G, the harmonics scaling factor (see Fig. 1), 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. 3 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. 4 displays the average hearing loss (with respect to threshold of hearing, see ISO [9]) 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.

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

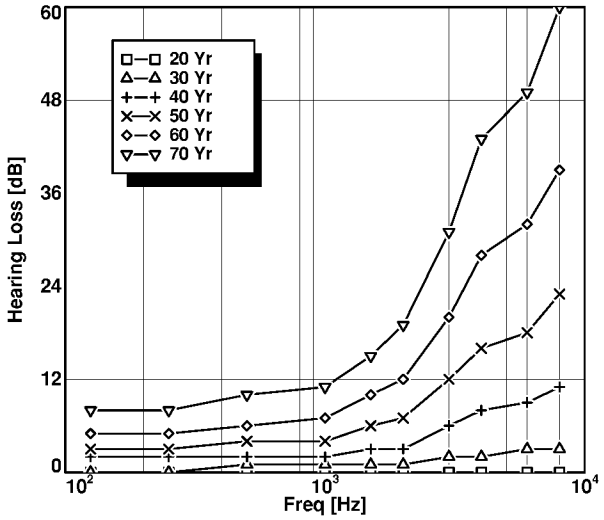

Fig. 4: Hearing loss (with respect to threshold of hearing) for a group of otologically normal males of various age. For each age 50% of the group has a higher hearing loss and 50% has a lower hearing loss.

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.

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.

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.

ACKNOWLEDGEMENT

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

REFERENCES

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

[2] 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.

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

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

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

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

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

[8] 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.

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