(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
17 July 2014 (17.07.2014)

W I P O | P C T

(10) International Publication Number
WO 2014/108677 Al

(51) International Patent Classification:
G11B 20/10 (2006.01)

(21) International Application Number:
PCT/GB2014/050040

(22) International Filing Date:
7 January 2014 (07.01.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
1300309.0 8 January 2013 (08.01.2013) GB

(71) Applicant: **MERIDIAN AUDIO LIMITED** [GB/GB]; Latham Road, Huntingdon, Cambridgeshire PE29 6YE (GB).

(72) Inventor; and
(71) Applicant : **CRAVEN, Peter, Graham** [GB/GB]; 43 Bournemouth Road, London, Greater London SW19 3AR (GB).

(72) Inventor: **STUART, John, Robert;** 21 Storeys Way, Cambridge, Cambridgeshire CB3 0DP (GB).

(74) Agent: **GILL JENNINGS & EVERY LLP;** The Broadgate Tower, 20 Primrose Street, London, Greater London EC2A 2ES (GB).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available):* AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available):* ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

(54) **Title:** DIGITAL ENCAPSULATION OF AUDIO SIGNALS

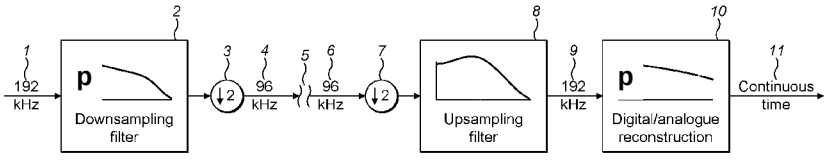

FIG. 3

(57) **Abstract:** Encoding and decoding systems are described for the provision of high quality digital representations of audio signals with particular attention to the correct perceptual rendering of fast transients at modest sample rates. This is achieved by optimising downsampling and upsampling filters to minimise the length of the impulse response while adequately attenuating alias products that have been found perceptually harmful.

DIGITAL ENCAPSULATION OF AUDIO SIGNALS

Field of the Invention

The invention relates to the provision of high quality digital representations of audio signals.

Background to the Invention

In the thirty years since the introduction of the Compact Disc (CD), the general public has come to accept "CD-quality" as the norm for digital audio. Meanwhile, two types of argument have raged in audio circles. One centres around the proposition that the 16 bits resolution and 44.1 kHz sampling rate of the CD are wasteful of data and that the equivalent sound can be conveyed by a more compact lossy-compressed format such as MP3 or AAC. The other takes the diametrically opposing view, asserting that the resolution and sampling rate of the CD are inadequate and that audibly better results are obtained using, for example, 24 bits and a sampling rate of 96kHz, a specification commonly abbreviated to 96/24.

If 44kHz is indeed not considered good enough, the question arises as to whether 96kHz is the answer or whether 192kHz or even 384kHz should be the sampling rate for 'ultimate' quality. Many audiophiles assert that 192kHz does indeed sound better than 96kHz.

Historically, the transition from a continuous-time representation of an analogue waveform to a sampled digital representation has been justified by the sampling theorem (www.en.wikipedia.org/wiki/Sampling_theorem), which states that a continuous-time waveform containing only frequencies up to a maximum f_{max} can be reconstructed *exactly* from a sampled representation having $2xf_{max}$ samples per second. Therefore, the continuous-time waveform is first filtered by a bandlimiting 'antialias' filter in order to remove frequencies above f_{max} that would otherwise be 'aliassed' by the sampling process and be reproduced as images below f_{max}. Following standard communications practice, the bandlimiting antialias filter usually approximates a flat frequency response up to f_{max} so the frequency response graph has the appearance of a 'brickwall'. The same applies

to a reconstruction filter used to regenerate a continuous waveform from the sampled representation.

According to this methodology the process of sampling and subsequent reconstruction is exactly equivalent to a time-invariant linear filtering process that removes frequencies above f_{max} and makes little or no change to frequencies significantly lower than f_{max}. It is therefore hard to understand that sampling at 192kHz can sound better than sampling at 96kHz, since the only difference would be the presence or absence of frequencies above about 40kHz, which exceeds the conventional human hearing range of 20Hz to 20kHz by a factor two.

To reconcile this discrepancy there have been suggestions that a frequency-domain description may not be adequate and that the time-domain response of the brickwall antialias and reconstruction filters should instead be examined. Time-domain and frequency-domain descriptions of a system response are related by the Fourier Transform, from which it follows that the sharp transition of a 'brickwall' filter inevitably results in an extended time response.

The term 'Nyquist frequency' refers to a frequency of half the sampling rate, for example 48kHz when sampling at 96kHz. The sampling theorem tells us that unambiguous reconstruction is possible up to the Nyquist frequency and it would be normal communications practice to place a 'brickwall' antialias filter just below the Nyquist, as shown in figure 1, in order to provide good reconstruction over as large a bandwidth as possible. Unfortunately the impulse response of this filter is considerably extended in time, as shown in figure 2A.

Since the Nyquist frequency of 48kHz is considerably above the conventional audio limit of 20kHz, there is room for a much wider 'transition band' such as 20kHz-48kHz, as shown by the dotted line in figure 1. Such a wider and more 'gentle' transition can be achieved by cascading the brickwall filter with another lowpass filter, a process known as "Apodising" and described in Craven, P.G., "Antialias Filters and System Transient Response at High Sample Rates" J. Audio Eng. Soc. Volume 52 Issue 3 pp. 216-242; March 2004. Figure 2B shows a considerably more compact impulse response corresponding to the wider transition band.

Another prior-art improvement, discussed in section 3.6 of the 2004 paper by Craven, is the replacement of a linear phase filter by a minimum-phase filter having in an asymmetric impulse response with no pre-ring. However the post-ring is thereby increased and the first 'main lobe' of the impulse response becomes broad. Listening tests indicate that minimum-phase filtering provides an audible improvement but does not provide completely transparent reproduction of audio at a sampling rate of 96kHz or less.

Thus, the route to higher quality has hitherto been seen as increasing the sampling frequency until no further improvement can be heard. Given the practical constraints of file sizes and bitrates for transmission, the prospects for interesting the public at large in high resolution sound have appeared bleak. What is needed is a methodology for high-quality digital sound recording and reproduction based on observed characteristics of the human ear rather than on conventional communications theory.

Summary of the Invention

According to a first aspect of the present invention, there is provided a system comprising an encoder and a decoder for conveying the sound of an audio capture, wherein the encoder is adapted to furnish a digital audio signal at a transmission sample rate from a signal representing the audio capture, and the decoder is adapted to receive the digital audio signal and furnish a reconstructed signal,

wherein the encoder comprises a downsampler adapted to receive the signal representing the audio capture at a first sample rate which is a multiple of the transmission sample rate and to downsample the signal to furnish the digital audio signal; and,

wherein an impulse response of the encoder and decoder in combination has a largest peak, and is characterised by a contiguous time region having an extent not greater than 6 sample periods of the transmission sample rate outside of which the absolute value of the impulse response does not exceed 10% of said largest peak.

Listening tests have indicated that shorter impulse responses are almost always better, and in most cases it has proved possible to design a filter that does not have significant responses extending beyond 6 sample periods.

Preferably, the downsampler comprises a decimation filter specified at the first sample rate, wherein the asymmetric component of response of the decimation filter is characterised by an attenuation of at least 32dB at frequencies that would alias to the range 0-7 kHz on decimation.

The range 0-7kHz is the range where the ear is most sensitive. The amount of attenuation required varies greatly according to the spectrum of the signal to be encoded in the vicinity of its Nyquist frequency, and may signals will require more than 32dB of attenuation.

It is further preferred that the impulse response of the decimation filter has a largest peak, and that the asymmetric component of the response is characterised by a contiguous time region having an extent not greater than 6 sample periods of the transmission sample rate outside of which the absolute value of the impulse response does not exceed 10% of said largest peak.

In some embodiments the encoder comprises an Infinite Impulse Response (MR) filter having a pole, and the decoder comprises a filter having a zero whose z-plane position coincides with that of the pole, the effect of which is thereby cancelled in the reconstructed signal.

In other embodiments the decoder comprises an Infinite Impulse Response (MR) filter having a pole, and the encoder comprises a filter having a zero whose z-plane position coincides with that of the pole, the effect of which is thereby cancelled in the reconstructed signal.

Preferably, the decoder comprises a filter having a response which rises in a region surrounding the Nyquist frequency corresponding to the transmission sample rate and the encoder comprises a filter having a response that falls in said region, thereby reducing downward aliasing in the encoder of frequencies above the Nyquist frequency to frequencies below the Nyquist frequency without

compromising the total system frequency response or impulse response. This feature is particularly helpful in cases where the original signal has a steeply rising noise spectrum.

In preferred embodiments the transmission sample rate is selected from one of 88.2kHz and 96kHz and the first sample rate is selected from one of 176.4kHz, 192kHz, 352.8kHz and 384kHz, these being standardised sample rates at which the invention has been found to be audibly beneficial.

Although the invention operates with contiguous time region having an extent not greater than 6 sample periods of the transmission sample rate, in some embodiments the extent of this contiguous time region is advantageously no greater than 5 period, 4 periods or even 3 periods of the transmission sample rate. It has been found on some signals that these shorter impulse responses are audibly even more beneficial than embodiments with an impulse response lasting 6 periods.

It is preferred that the impulse response rises monotonically to the largest peak, thus avoiding pre-responses which are known to be audibly deleterious.

According to a second aspect of the present invention, there is provided a method of furnishing a digital audio signal for transmission at a transmission sample rate by reducing the sample rate required to convey the sound of captured audio, the method comprising the steps of:

filtering a representation of the captured audio having a first sample rate that is a multiple of the transmission sample rate using a decimation filter specified at the first sample rate; and,

decimating the filtered representation to furnish the digital audio signal,

wherein an impulse response of the decimation filter has a largest peak and an asymmetric component of response of the decimation filter is characterised by

(i) an attenuation of at least 32dB at frequencies that would alias to the range 0-7 kHz on decimation; and,

(ii) a contiguous time region having an extent not greater than 6 sample periods of the transmission sample rate outside of which the

absolute value of the impulse response does not exceed 10% of the largest peak.

The invention thus provides adequate rejection of undesirable alias products, and of any ringing near the Nyquist frequency of the representation at the first sample rate, while not extending the system impulse response more than necessary.

In some embodiments the method further comprises the steps of analysing a spectrum of the captured audio, and choosing the decimation filter responsively to the analysed spectrum. The method may then further comprise the step of furnishing information relating to the choice of decimation filter for use by a decoder. In that way both the decimation filter and a corresponding reconstruction filter in a decoder can be optimally matched to the noise spectrum or other characteristics of the signal to be conveyed.

In preferred embodiments the transmission sample rate is selected from one of 88.2kHz and 96kHz and the first sample rate is selected from one of 176.4kHz, 192kHz, 352.8kHz and 384kHz, these being standardised sample rates at which the invention has been found to be audibly beneficial.

Although the invention operates with contiguous time region having an extent not greater than 6 sample periods of the transmission sample rate, in some embodiments the extent of this contiguous time region is advantageously no greater than 5 period, 4 periods or even 3 periods of the transmission sample rate. It has been found on some signals that these shorter impulse responses are audibly even more beneficial than embodiments with an impulse response lasting 6 periods.

It is preferred that the impulse response rises monotonically to the largest peak, thus avoiding pre-responses which are known to be audibly deleterious.

According to a third aspect of the present invention, a data carrier comprises a digital audio signal furnished by performing the method of the aspect aspect.

According to a fourth aspect of the present invention, an encoder for an audio stream is adapted to furnish a digital audio signal using the method of the second aspect.

In preferred embodiments the encoder comprises a flattening filter having a symmetrical response about the transmission Nyquist frequency. Preferably, the flattening filter has a pole.

According to a fifth aspect of the present invention, there is provided a system for conveying the sound of an audio capture, the system comprising an encoder and a decoder,

wherein the encoder is adapted to receive a signal representing the audio capture and to furnish a digital audio signal, and the decoder is adapted to receive the digital audio signal and furnish a reconstructed signal, and

wherein the encoder and decoder have impulse responses which, when combined, produce a total system impulse response that is more compact than an impulse response of the encoder alone.

Preferably, the decoder comprises a filter having a z-plane zero whose position coincides with that of a pole in the response of the encoder.

Preferably, the decoder comprises a filter chosen in dependence on information received from the encoder.

In some embodiments it is preferred that an impulse response of the encoder and decoder in combination has a largest peak, and is characterised by a contiguous time region having an extent not greater than 6 sample periods of the transmission sample rate outside of which the absolute value of the averaged impulse response does not exceed 10% of said largest peak.

According to a sixth aspect of the present invention, there is provided an encoder adapted to furnish a digital audio signal at a transmission sample rate from a signal representing an audio capture, the encoder comprising a downsampling filter having an asymmetric component of response equal to the asymmetric component of response of a filter whose frequency response has a double zero at

each frequency that will alias to zero frequency and has a slope at the transmission Nyquist frequency more positive than minus thirteen decibels per octave.

It is preferred that the encoder comprises a flattening filter having a symmetrical response about the transmission Nyquist frequency. Preferably, the flattening filter has a pole. It is further preferred that the transmission frequency is 44.1 kHz and the encoder's frequency response droop does not exceed 1dB at 20kHz.

According to a seventh aspect of the present invention, there is provided a system comprising an encoder and a decoder for conveying the sound of an audio capture, wherein the encoder is adapted to furnish a digital audio signal at a transmission sample rate from a signal representing the audio capture, and the decoder is adapted to receive the digital audio signal and furnish a reconstructed signal,

wherein the encoder comprises a downsampler adapted to receive the signal representing the audio capture at a first sample rate which a multiple of the transmission sample rate and to downsample the signal to furnish the digital audio signal; and,

wherein the encoder comprises an Infinite Impulse Response (MR) filter having a pole, and the decoder comprises a filter having a zero whose z-plane position coincides with that of the pole, the effect of which is thereby cancelled in the reconstructed signal.

Preferably, an impulse response of the encoder and decoder in combination has a largest peak, and is characterised by a contiguous time region having an extent not greater than 6 sample periods of the transmission sample rate outside of which the absolute value of the averaged impulse response does not exceed 10% of said largest peak.

According to an eighth aspect of the present invention, there is provided an encoder adapted to furnish a digital audio signal at a transmission sample rate from a signal representing an audio capture, the encoder comprising a downsampling filter adapted to receive the signal representing the audio capture at a first sample rate which a multiple of the transmission sample rate and to

downsample the signal to furnish the digital audio signal, wherein the encoder is adapted to analyse a spectrum of the captured audio and select the downsampling filter responsively to the analysed spectrum.

Preferably, the selected downsampling filter has a steeper attenuation response at the transmission Nyquist frequency if the analysed spectrum is rising rapidly at the transmission Nyquist frequency.

It is preferred that the encoder is adapted to transmit information identifying the selected downsampling filter to a decoder as metadata.

In preferred embodiments the encoder comprises a flattening filter having a symmetrical response about the transmission Nyquist frequency. Preferably, the flattening filter has a pole.

According to an ninth aspect of the present invention, there is provided a decoder for receiving a digital audio signal at a transmission sample rate and furnishing an output audio signal, wherein the decoder comprises a filter having an amplitude response which increases with frequency in a frequency region adjoining the Nyquist frequency corresponding to the transmission sample rate.

This feature is necessary in order to optimise a signal-to-alias ratio for frequencies near the Nyquist frequency in cases where the representation at the higher sample rate shows a strongly rising spectrum at the said Nyquist frequency and where it is desired to minimise phase distortion over the conventional audio band 0-20kHz.

Preferably, the filter has an amplitude response of at least +2dB at the Nyquist frequency corresponding to the transmission sample rate, relative to the response at DC. In general, a rising decoder response can be advantageous in allowing an encoder to provide adequate alias attenuation while providing a flat frequency response in the audio range and not lengthening the total system impulse response, and while the decoder response should eventually fall, it is generally still somewhat elevated at the said Nyquist frequency.

In some embodiments it is preferred that the filter has a response chosen in dependence on information received from an encoder. This allows the encoder to choose the filtering optimally on a case-by-case basis.

As will be appreciated by those skilled in the art, various methods are disclosed for optimising the sound of the reconstructed signal and in particular for controlling decimation aliases without lengthening the total impulse response of the system in an undesirable manner.

Advantageously, filters are selected responsively to the characteristics of the source material. Likewise, different filter implementations such as all-zero, all-pole and polyphase may be employed as appropriate for each situation. Further variations and embellishments will become apparent to the skilled person in light of this disclosure.

Brief Description of the Drawings

Examples of the present invention will be described in detail with reference to the accompanying drawings, in which:

Figure 1 shows a known (continuous) 'brickwall' antialias filter response for use with 96kHz sampling, and (dotted) an apodised filter response;

Figures 2A and 2B show known impulse responses corresponding to linear phase filters having the frequency responses shown in Figure 1;

Figure 3 shows a system for transmitting an audio signal at a reduced sample rate, with subsequent reconstruction to continuous time.

Figure 4 shows the response of a (½, 1, ½) reconstruction filter, normalised for unity gain at DC;

Figure 5A shows the frequency response of an unflattened downsampling filter.

Figure 5B shows the frequency response of a downsampling filter incorporating flattening;

Figure 6 shows the response of a reconstruction filter including upsampling to continuous time and a third-order correction for the passband droop of Figure 5A;

Figure 7 shows the total system impulse response when the filters of Figure 4 and Figure 5B are combined with further upsampling to continuous time;

Figure 8 shows the spectrum of two commercial recordings having a strongly rising ultrasonic response.

Figure 9 shows the response of a flattening filter symmetrical about 48kHz for use with the downsampling filter of Figure 5B;

Figure 10 shows (lower curve) the response of the downsampling filter of Figure 5A and (upper curve) the response after flattening using the symmetrical flattener of Figure 9;

Figure 11 shows a linear B-spline sampling kernel;

Figure 12A illustrates impulse reconstruction at 88.2kHz from 44.1kHz infra-red encoded samples aligned with even samples of an original 88.2kHz stream.

Figure 12B illustrates impulse reconstruction at 88.2kHz from 44.1kHz infra-red encoded samples aligned with odd samples of an original 88.2kHz stream.

Figure 13A shows the response of a downsampling filter having zeroes to provide strong attenuation near 60kHz;

Figure 13B shows the response of an upsamping filter having poles to cancel the effect on total response of the zeroes in the filter of Figure 13A; and,

Figure 13C shows the end-to-end response from combining the responses of figure 13A, figure 13B and an assumed external droop.

Detailed Description

The present invention may be implemented in a number of different ways according to the system being used. The following describes some example implementations with reference to the figures.

Axioms

Most adult listeners are unable to hear isolated sinewaves above 20kHz and it has hitherto often been assumed that this implies that frequency components of a signal above 20kHz are also unimportant. Recent experience indicates that this assumption, though plausible by analogy with linear-system theory, is incorrect.

Current understanding of human hearing is very incomplete. In order to make progress we have therefore relied on hypotheses that have been only partially or indirectly verified. The invention will thus be explained on the basis of the following hypotheses:

- The ear does not behave as a linear system
- As well as analysing tones in the frequency domain, the ear also analyses transients in the time domain. This may be the dominant mechanism in the ultrasonic region.
- "Ringing" of filters used for antialiassing and reconstruction is undesirable, even if in the high ultrasonic range 40kHz-100kHz.
- Aliassing of frequencies above 48kHz to frequencies below 48kHz is not catastrophic to sound quality, provided the aliased products do not fall within the conventionally audible range 0-20kHz.
- A pre-ring is usually more of a problem than a post-ring, but both are bad.
- It seems best if the temporal extent of the total system impulse response can be minimised.

Regarding the last of these points, the "total system" is intended to include the analogue-to-digital and digital-to-analogue converters, as well as the entire digital chain in between. Ideally, one might include the transducer responses too, but these are considered outside the scope of this document.

Sampling and Aliassing

A continuous time signal can be viewed as a limiting case of a sampled signal as the sample rate tends to infinity. At this point we are not concerned whether an original signal is analogue, and therefore presumably continuous in time, or whether it is digital, and therefore already sampled. When we talk about resampling, we mean sampling a notional continuous-time signal that is represented by the original samples.

A frequency-domain description of sampling or resampling is that the original frequency components are present in the resampled signal, but are accompanied

by multiple images analogous to the 'sidebands' that are created in amplitude modulation. Thus, an original 45kHz tone creates an image at 51kHz, if resampled at 96kHz, the 51kHz being the lower sideband of modulation by 96kHz. It may be more intuitive to think of all frequencies as being 'mirrored' around the Nyquist frequency of 48kHz; thus 51kHz is the mirror image of 45 kHz, and equally an original 51kHz tone will be mirrored down to 45kHz in the resampled signal.

If a transmission channel involves several resamplings at different rates, images of the original spectrum will accumulate and there is every possibility that an audio tone will be mirrored upward by one resampling and then down by a subsequent resampling, landing within the audible range but at a different frequency from the original. It is to prevent this that 'correct' communications practice teaches that antialias and reconstruction filters should be used at each stage so that all images are suppressed. If this is done, resamplings may be cascaded arbitrarily without build-up of artefacts, the limitation being merely that the frequency range is limited to that which can be handled by the lowest sample rate in the chain.

However, we take the view that filters that would be considered correct in communications engineering are not audibly satisfactory, at least not at sample rates that are currently practical for mass distribution. We accept that aliasing may take place and are proposing to balance aliasing against 'time-smear' of transients due to the lengthening of the system's impulse response caused by filtering.

Thus, unlike in traditional practice, aliasing is not completely removed and will build up on each resampling of the signal. Hence, multiple resamplings to arbitrary rates are not undertaken without penalty and it is best if the signal is always represented at a sample rate that is an integer multiple of the rate that will be used for distribution. For example, analogue-to-digital conversion at 192kHz followed by distribution at 96kHz is fine, and conversion at 384kHz may be better still, depending on the wideband noise characteristics of the converter.

Following distribution, the consumer's playback equipment also needs to be designed so as not to introduce long filter responses, and indeed the encoding and decoding specifications should preferably be designed together to give certainty of the total system response.

Downsampling from 192kHz for 96kHz Distribution

We consider the problem of taking a signal that has already been digitised at 192kHz, downsampling the signal to 96kHz for transmission and then upsampling back to 192kHz on reception. It is understood that the principles described here apply to storage as well as transmission, and the word 'transmission' encompasses both storage and transmission.

Referring to the system shown in figure 3, the input signal 1 at a sampling rate such as 192kHz is passed to a downsampling filter 2 and thence to a decimator 3 to produce a signal 4 at a lower sampling rate such as 96kHz. After passing through the transmission or storage device 5, the 96kHz signal 6 is upsampled 7 and filtered 8 to furnish the partially reconstructed signal 9, at a sampling rate such as 192kHz.

The main focus of this document is the method of producing the partially reconstructed signal 9, but we also note that further reconstruction 10 is needed to furnish a continuous-time analogue signal 11. The object of the invention is to make the sound of signal 11 as close as possible to the sound of an analogue signal that was digitised to furnish the input signal 1. This does not necessarily imply that signal 9 should be as close as possible in an engineering sense to signal 1. Moreover, the further reconstruction 10 may have a frequency response droop which can, if desired, be allowed for in the design of the filters 2 and 8.

Figure 3 shows the filter 2 and downsampler 3 as separate entities but it will sometimes be more efficient to combine them, for example in a polyphase implementation. Similarly the upsampler 7 and filter 8 may not exist as separately identifiable functional units.

Downsampling uses decimation, in this case discarding alternate samples from the 192kHz signal, while upsampling uses padding, in this case inserting a zero

sample between each consecutive pair of 96kHz samples and also multiplying by 2 in order to maintain the same response to low frequencies. On downsampling, frequencies above the 'foldover' frequency of 48kHz will be mirrored to corresponding images below the foldover frequency. On upsampling, frequencies below the foldover frequency will be mirrored to corresponding frequencies above the foldover frequency. Thus, upsampling and downsampling create upward aliased products and downward aliased products, which can be controlled by an upsampling filter prior to decimation and a downsampling filter following the padding. The upsampling and downsampling filters are specified at the original sampling frequency of 192kHz.

If the aliased products are ignored, the total response is the combination of the responses of the upsampling and downsampling filters. In the time domain, this combination is a convolution.

We have found that good results are obtained by designing upsampling and downsampling filters such that the total response is that of a Finite Impulse Response (FIR) filter of minimal length. In the z-transform domain, zeroes can be introduced into each of these filters to suppress undesirable responses. In particular, it is likely that each filter will have one or more transfer function zeroes near z=-1 in order to suppress signals near the Nyquist frequency of 96kHz. In downsampling without filtering, such signals would alias to audio frequencies, including frequencies below 10kHz where the ear is most sensitive. Conversely, if upsampling is performed by padding without filtering, large low frequency signal content will create large image energy near 96kHz which, whether or not of audible consequence, may place unacceptable demands on the slew-rate capabilities of subsequent electronics, and possibly also burn out loudspeaker tweeters.

FIR filters whose zeroes are all close to the Nyquist will not, by themselves, cause overshoot or ringing: the impulse response will be unipolar and reasonably compact. However a $(1 + z^{-1})$ factor implemented at 192kHz introduces a frequency response droop of 0.47dB at 20kHz. This would be considered only marginally acceptable in professional digital audio equipment, and if we need several such factors, say five or more, the passband droop and resulting dulling

of the sound certainly becomes unacceptable. Accordingly, a correction or "flattening" filter is needed, as will be discussed shortly.

Upsampling from 96kHz for Playback

It is usual for reconstruction to a continuous-time signal to be performed using a sequence of '2χ' stages. I.e., the sampling rate is typically doubled at each stage and a conversion from digital to analogue is performed when the sampling rate has reached 384kHz or higher. We shall concentrate firstly on the first and most critical stage: that of upsampling from 96kHz to 192kHz.

At the heart of this upsampling is an operation, conceptual or physical, of zero-padding the stream of 96kHz samples to produce the 192kHz stream. That is, we generate a 192kHz signal whose samples are alternately a sample from the 96kHz signal and zero.

Zero-padding creates upward aliased products having the same amplitude as the frequencies that were aliased. In the current context, these products are all above 48kHz and one might assume that they will be inaudible. However the signal will generally have high amplitudes at low audio frequencies, which implies high-level alias products at frequencies near 96kHz. As already noted, these alias products need to be controlled in order to not to impose excessive slew-rate demands on subsequent electronics and risk the burn-out of loudspeaker tweeters. The purpose of an upsampling or reconstruction filter is to provide this control, and it will be seen that strong attenuation near 96kHz is the prime requirement.

The simplest reconstruction filter that we consider satisfactory for 96kHz to 192kHz reconstruction is a 3-tap FIR filter having taps (½, 1, ½) implemented at the 192kHz rate. Its normalised response is shown in figure 4. This filter has two z-plane zeroes at z=-1 , corresponding to the Nyquist frequency of 96kHz. These zeroes provide attenuation near 96kHz which may or may not be sufficient so further near-Nyquist zeroes may be required. The (½, 1, ½) filter also introduces a droop of 0.95dB at 20kHz, or 1.13dB if operated at 176.4kHz, which will need to be corrected.

Passband Flattening

Since the system includes a downsampler, correction to flatten a frequency response that droops towards the top of the conventional 0-20kHz audio range could be provided either at the original sample rate or the downsampled rate, but to provide the shortest end-to-end impulse response on the upsampled output the flattening should be performed at the higher sample rate, such as 192kHz. This still leaves choice about where the correction is performed:

a. The encoder (downsampler) and decoder (upsampler) each incorporates a correction for its own droop

b. The encoder provides correction for itself and for the decoder

c. The decoder provides correction for itself and for the encoder

d. Arbitrary distribution of correction between encoder and decoder.

Option (a) may be convenient in practice since the resulting downsampled stream will have a flat frequency response and can be played without a special decoder, However the resulting combined of "end-to-end" impulse response of encoder and decoder is then likely to be longer than when a single corrector corrector is designed for the total droop.

Options (b) and (c) may provide the same end-to-end impulse response, and so may option (d) if a single corrector to the total response is generated, factorised ad the factors distributed. However although the end-to-end responses may be the same, putting the flattening filter in the encoder prior to downsampling generally increases downward aliassing in the encoder, and listening tests have tended to favour putting the flattening filter in the decoder after upsampling, even though upward aliases are thereby intensified.

As for the design of the correction filter, the skilled person will be aware that in the case of a linear-phase droop, a linear-phase correction filter can be obtained by expanding the reciprocal of the z-transform of the droop as a power series in the neighbourhood of $z=1$. This total response can thereby be made maximally flat to any desired order by adjusting the order of the power-series expansion. In the present context however a minimum-phase correction filter is preferred in

order to avoid pre-responses. To this end, the droop is first convolved with its own time reverse to produce a symmetrical filter and above procedure applied. This will result in a linear-phase corrector which provides twice the correction, in decibel terms, needed for the original droop. The linear-phase corrector is then factorised into quadratic and linear polynomials in z, half of the factors being minimum-phase and half being maximum-phase. The minimum-phase factors are selected and combined and normalised to unity DC gain to provide the final correction filter. This methodology was illustrated in section 3.6 of the above-mentioned 2004 paper by Craven, building on the work of Wilkinson (Wilkinson, R.H., "High-fidelity finite-impulse-response filters with optimal stopbands". IEE Proc-G Vol. 120, no. 2, pp. 264-272: 1991 April).

The effect of the correction filter is not only to flatten the passband but also to increase the near-Nyquist response of the encoder in case (b) or of the decoder in case (c), or potentially both in case (d), the increase probably requiring the introduction of further zeroes near z=-1 in order to achieve a desired near-Nyquist attenuation specification. The further zeroes will require an increase in the strength of the correction filer. Thus, the zeroes that attenuate near Nyquist and passband correction filter need to be adjusted together until a satisfactory result is obtained.

Total System Response

If fed with a zero-padded 96kHz signal, the output of a 3-tap reconstruction filter having taps (½, 1, ½) implemented at the 192kHz rate is a 192kHz stream in which each even-numbered sample has the same value as its corresponding 96kHz sample and each odd-numbered sample has a value equal to the average of its two neighbouring even-numbered samples. If now multistage reconstruction to continuous time similarly uses a 3-tap (½, 1, ½) reconstruction filter at each stage, the result will be equivalent to linear interpolation between consecutive 96kHz samples.

In the frequency domain, the response of such a multistage reconstruction is the square of a sine function:

$$\left(\operatorname{sinc}\left(\frac{\pi f}{96\ kHz}\right)\right)^2$$

where f is frequency and $\mathrm{sinc}(x) = \frac{\sin(x)}{x}$.

The passband droop may be approximated by a quadratic in f.

$$1 - \frac{\pi^2 (f796kHz)^2}{3} \approx 1 - 3.290//96kHz^2,$$

which implies a response of -1.34dB at 20kHz if reconstructing from 96kHz, or -1.61dB at 20kHz if reconstructing from 88.2kHz.

Reconstructed thus, the slew rate of the continuous time the slew rate of the continuous-time signal is never greater than that implied by the 96kHz samples. Nevertheless, it will have small discontinuities of gradient. Viewed on a sufficiently small time scale, this is not possible electrically, let alone acoustically. It is outside our scope to consider the analogue processing in detail, but we note that an impulse response that is everywhere positive must, unless it is a Dirac delta function, have some frequency response droop. We prefer not to require the use of an analogue 'peaking' filter to produce a flat overall response since the shortest overall impulse response is likely to be obtained if all passband correction is applied at a single point. We therefore prefer that the digital passband flattening should have some allowance for analogue droop.

Nevertheless, the more droop that is corrected, the less compact is the downsampling filter. In the filters presented here we have therefore compensated for the $sinc()^2$ droop for assumed multistage reconstruction from a 192kHz stream to continuous time, with a further margin to allow for a small droop, amounting to 0.162dB at 20kHz, in subsequent analogue processing. This margin would allow for an analogue system having a strictly nonnegative impulse response of rectangular shape and extent 5µs, or alternatively a Gaussian-like response with standard deviation approximately 3µs.

Figure 5A shows the response of a 6-tap downsampling filter designed according to these principles having a near-Nyquist attenuation of 72dB and z-transform response:

$$0.0633 - 0.2321\, z^{-1} + 0.3434\, z^{-2} + 0.2544\, z^{-3} + 0.0934\, z^{-4} + 0.0134\, z^{-5}$$

If paired with the previously discussed 3-tap upsampling filter having response (½ + z-1 + ½ z -2), we find that a 4-tap correction filter:

$$4.3132 - 5.3770 z^{-1} + 2.4788 z^{-2} - 0.4151 z^{-3}$$

will correct the total droop from the downsampling filter and the 3-tap upsampling filter, to provide an end-to-end response flat within 0.1 dB at 20kHz, including the effect of analogue droop as discussed above. If this correction filter is folded with the downsampling filter, the combined encoding filter has z-transform:

$$0.27289 + -\frac{0.66093}{z} + \frac{0.39002}{z^2} - \frac{0.20014}{z^3} - \frac{0.20992}{z^4} + \frac{0.04329}{z^5} + \frac{0.05411}{z^6} - \frac{0.00563}{z^7} - \frac{0.00555}{z^8}$$

and the response shown in figure 5B, which rises above 20kHz in order to pre-correct the droop from the subsequent upsampling and reconstruction.

Alternatively, the correction can be folded with the upsampling filter *(½ + z~1* + ½ z ~2) whose response is shown in figure 4 to produce a decoding filter having the response shown in figure 6 and the z-transform:

$$2.1566 - 0.5319 z^{-7} + 0.7076 z^{-2} - 1.6566 z^{-3} + 1.0319 z^{-4} - 0.2076 z^{-5}$$

In this case it is the decoder that has a rising response, to correct the droop from the 6-tap encoding filter having the response of figure 5A. Listening tests have indicated that this 9-tap downsampling filter has a distinct superiority relative to longer filters and we have deduced that shorter filters are preferable generally.

Of greater significance however is the total response when the downsampler, upsampler and assumed analogue response are combined. Figure 7 shows the impulse response from the downsampler, a multi-stage upsampler as proposed above and an analogue system having a rectangular impulse response of width 5µs. With no threshold applied, the total extent of the response is 13 samples or 67.7ps, but with a threshold of -40dB or 1% of the maximum, the absolute value of the response exceeds the threshold only in a region of extent 49.5ps, i.e. 9.5 samples at the 192kHz rate or 4.75 samples at the transmission sample rate of 96kHz. Similarly, with a threshold of-20dB or 10% of the maximum, the absolute value of the response exceeds the threshold only in a region of extent 32.2ps, i.e. 6.2 samples at the 192kHz rate or 3.1 samples at the transmission sample rate of 96kHz. Thus, it is safe to say that the temporal extent of this filter does not

exceed 4 sample periods of the transmission sample rate. When other criteria are tightened, the impulse response may need to be somewhat longer, but in nearly all reasonable cases it is possible to achieve an impulse response of length not exceeding 6 sample periods at the transmission sample rate.

An encoder and decoder combination incorporating the downsampling and upsampling filters described above and with the total system response shown in figure 7 has been found to produce audibly good results on available 192kHz recordings. Indeed the decoded signal has sometimes sounded better than conventional playback of the 192kHz stream without downsampling, a result that could be attributed to the attenuation by the downsampling filter of any ringing near 96kHz already present in the 192kHz stream.

Alias Trading Based on Noise Spectrum Analysis

Much commercial source material has a noise floor that rises in the ultrasonic region because of the behaviour of analogue-to-digital converters and noise shapers. For example, the spectrum of a commercially available 176.4kHz transcription of the Dave Brubeck Quartet's "Take 5", shown as the upper trace in figure 8, reveals a noise floor that increases by 42dB between 33kHz and 55kHz, these frequencies being equidistant from the foldover frequency of 44.1kHz when downsampled. If there were no filtering before decimation, the resulting 88.2kHz stream would have noise at 33kHz composed almost entirely of noise aliased from 55kHz and would thereby have a spectral density some 42dB higher than in the 175.4kHz presentation of the recording.

The downsampling filter of figure 5B, if operated at 176.4kHz instead of 192kHz, would provides gain of +2.3dB and -6.7dB at 33kHz and 55kHz respectively, a difference of 9dB. Downsampling "Take 5" with this filter, components aliased from 55kHz would still dominate original 33kHz components by 33dB. The alternative downsampling filter of figure 5A provides 16.8dB discrimination between these two frequencies, resulting in aliased components 25dB higher than the original components. For this is a somewhat exceptional case, filters (to be described) having still larger discrimination might be preferable; nevertheless the filter of figure 5A has been found satisfactory in many cases, and to provide better audible results than the filter of figure 5B. Thus placing the correction filter

in the decoder, as in option (c) discussed earlier, seems preferable to placing it in the encoder, option (b).

The above discussion has concentrated on downward aliased signal components, but it should be noted that putting the correction filter in the decoder will have the effect of boosting upward aliased components. It is a matter of trading downward aliasing against upward aliasing, and for downsampling from 192kHz to 96kHz, or from 176.4kHz to 88.2kHz it seems audibly better to reduce downward aliasing even if upward aliasing thereby increased.

There is no established criterion for how much aliased components should be reduced relative to original components, but a criterion may be derived based on balancing phase distortion in the audio band against total noise. We assume that the total response should be minimum-phase in order to avoid pre-responses. The flattening filter is always designed to give an total amplitude response flat to fourth order but Bode's phase-shift theorems tell us that when ultrasonic attenuation is introduced, phase distortion is inevitable in a minimum-phase system. When the phase response is expanded as a series in frequency, only odd powers are present. The linear term is irrelevant since it is equivalent to a time delay, hence the cubic term is dominant. If now additional attenuation δg decibels is introduced over a frequency interval δf centred on frequency f, we can deduce from Bode's theorems that the resulting addition to the cubic term in the phase response will be proportional to $\delta g.\delta f/f^4$. From the inverse fourth power dependence on/we can deduce that for lowest total noise consistent with a given phase distortion and a given end-to-end frequency response, the upward and downward aliassing should be balanced so that the ratio of the original noise power to the aliased noise power is equal to the inverse fourth power of the ratio of the two frequencies involved.

In the case of downsampling to 96kHz, this criterion implies that the noise spectral density at 36kHz that results from original 60kHz noise should be 8.9dB below the noise spectral density at 36kHz in the original 192kHz sampled signal. Also, at the foldover frequency of 48kHz, the spectrum of the noise after filtering by the downsampling filter should optimally have a slope of -12dB/8ve. It follows that the slope of the downsampling filter of figure 5A is not sufficient in the case of

"Take 5" according to this criterion, and a downsampling filter with a steeper slope near 48kHz is indicated if this criterion is considered relevant. "Take 5" is somewhat exceptional but the spectrum of "Brothers in Arms" by "Dire Straits", also shown in figure 8, also has a high slope near the foldover frequency.

Flattening the downsampled signal

As discussed, aliasing considerations often suggest that that the downsampling filter be not flattened, flattening being postponed to a subsequent upsampler. The transmitted signal will thereby not have a flat frequency response, which may be a disadvantage for interoperability with legacy equipment that does not flatten.

A way to avoid the disadvantage without affecting the alias property of the downsampler is to flatten using a filter with a response such as shown in figure 9 that is symmetrical about the transmission Nyquist frequency, i.e. half the transmission sample frequency. The transmission Nyquist frequency is 48kHz if downsampling from 192kHz to 96kHz, giving the unflattened and flattened downsampling responses are shown in figure 10.

The reason that the disadvantage is avoided is that the 'legacy flattener' is a symmetrical filter that treats each frequency and its alias image equally. The two frequencies are boosted or cut in the same ratio so the ratio of upward to downward aliasing in a subsequent decimation is not affected.

The response shown in figure 9 is in fact the response of the filter:

$$\frac{1.660575124}{1 + 0.6108508622z^{-2} + 0.04972426151z^{-4}}$$

which is minimum-phase all-pole and contains only even powers of z. Filtering with this filter prior to decimation-by-2 is equivalent to filtering the decimated stream using the all-pole filter:

$$\frac{1.660575124}{1 + 0.6108508622z^{-7} + 0.04972426151z^{-2}}$$

which is a process that can be reversed in a decoder, for example by applying a corresponding inverse filter:

$$.6022009998\,(1 + 0.6108508622\zeta^{-1} + 0.04972426151z^{-2})$$

to the received decimated signal prior to upsampling. Thus, z-plane poles in the encoding filter are cancelled by zeroes in the decoder. In the time domain, any ringing caused by the legacy flattener in the encoder is quenched by the corresponding 'legacy unflattening' in the decoder, and this is one of the ways in which the total impulse response of the combination of encoder and decoder is more compact than that of the encoder alone.

After upsampling, a decoder can apply a psychoacoustically optimal flattener at the higher sample rate, just as if there were no legacy flattener. It is thus completely transparent that that the decimated signal has been flattened and then unflattened again.

The 'legacy unflattener' can alternatively be implemented after usampling, using:

$$.6022009998\ (1 + 0.6108508622z^{-2} + 0.04972426151z^{-4})$$

at the higher sampling rate. As this is an FIR filter, it may well be convenient to merge it with the upsampling filter and the end-to-end flattener. In this case the legacy unflattener may not be a separately identifiable functional unit. Thus, for both the legacy flattener and the legacy unflattener there is the option of implementation at the transmission sample rate or at the higher sample rate, in the latter case using a filter whose response is symmetrical about the transmission Nyquist frequency. In this document these two implementation mechods are considered equivalent and a reference to just one of them may be taken to include the other. Moreover if implemented at the higher rate the flattener or unflattener may be merged with other filtering, though its presence may be deduced if the z-transform of, respectively, the total decimation filtering or the total reconstruction filtering has z-transform factors that contain powers of z^n only where n is the decimation or interpolation ratio.

It is not required that the legacy flattener be all-pole: it could be FIR or a general MR filter provided its response is symmetrical about the transmission Nyquist frequency. For example the FIR filter:

$$1.444183138 - 0.5512608378\ z^{-1} + 0.1190498978z^{-2} - 0.01197219763\ z^{-3}$$

could be applied after decimation in an encoder and its inverse prior to upsampling in a decoder, this third-order FIR filter being similarly effective to the

second-order all-pole filter of figure 9 in flattening the transmitted signal. In this case the decoder would have poles that cancel zeroes in the encoder. This FIR flattener could alternatively be implemented prior to decimation using:

$$1.444\,183\,138 - 0.551\,260\,837\,8z^{-2} + 0.119\,049\,897\,8z^{-4} - 0.011\,972\,197\,63z^{-6}$$

and in this form it could be merged with the downsampling filter and so not be identifiable as a separate functional unit.

While the legacy flattener has here been explained in the context of a 2:1 downsampling, the same principles apply in the case of an n:1 downsampling, where the legacy flattening and unflattening may be performed at the transmission sample rate using a general minimum-phase filter and its inverse, or it may be performed at the higher sample rate using a filter containing powers of z" only. In both cases the legacy flattener has a decibel response that is symmetrical about the transmission Nyquist.

Having noted that an invertible symmetrical filter applied at the original sample rate makes no difference to the alias characteristics of the filtering and that its effect can be reversed completely in a decoder, it follows that in comparing the suitability of one candidate downsampling filter with another, symmetrical differences in the decibel response are irrelevant. Hence we decompose the decibel response $dB(f)$ of a given filter into a symmetric component:

$$\frac{dB(f) + dB(fs_{trans} - f)}{2}$$

and an asymmetric component:

$$\frac{dB(f) - dB(fs_{trans} - f)}{2}$$

where f is frequency, fs_{trans} is the transmission sampling frequency, and a comparing between two downsampling filters we concentrate on the asymmetric component, leaving the symmetric component to be adjusted if necessary in a decoder.

Infra-red coding

We refer to the paper by Dragotti P.L., Vetterli M. and Blu T.: "Sampling Moments and Reconstructing Signals of Finite Rate of Innovation: Shannon Meets Strang-Fix", IEEE Transactions on Signal Processing, Vol. 55, No. 5, May 2007. Section III A of this paper considers a signal consisting of a stream of Dirac pulses having arbitrary locations and amplitudes, and the question is asked of what sampling kernels can be used so that the locations and amplitudes of the Dirac pulses may be deduced unambiguously from a uniformly sampled representation of the signal.

We consider that this question may be relevant to the reproduction of audio, in that many natural environmental sounds such as twigs snapping and it is by no means clear that a Fourier representation is appropriate for this type of signal. The linear B-spline kernel shown in figure 11 is the simplest polynomial kernel that will enable unambiguous reconstruction of the location and amplitude of a Dirac pulse. We have given the name "infra-red coding" to a downsampling specification based these ideas.

In downsampling, we start with a signal that is already sampled but the conceptual model is that this is a continuous time signal, in which the original samples are presented a sequence of Dirac pulses. The continuous time signal is convolved with a kernel and resampled at the rate of the downsampled signal. Referring to figure 11, the resampling instants are the integers 0, 1, 2, 3 etc while the original signal is presented, on a finer grid. Assuming that the original samples and resampling instants are aligned, then the continuous time convolution with the linear B-spline followed by resampling is equivalent to a discrete-time convolution with the following sequences prior to decimation:

$(1, 2, 1) / 4$	for decimation by 2
$(1, 2, 3, 2, 1) / 9$	for decimation by 3
$(1, 2, 3, 4, 3, 2, 1) / 16$	for decimation by 4
...	
$(1, 2, 3, 4, 5, 6, 7, 8, 7, 6, 5, 4, 3, 2, 1) / 64$	for decimation by 8.

These sequences are merely samplings at the original sampling rate of the B-spline kernel. Since the kernel has a temporal extent of two sample periods at

the downsampled rate, in all cases the downsampling filter will have a temporal extent not exceeding two sample periods at the downsampled rate.

Thus for decimation by 2 the downsampling filter would have z-transform (¼ +½ z^{-1}+1/4 z^{-2}). We have found that very satisfactory results can be obtained using this filter for downsampling in combination with the same filter, suitably scaled in amplitude, for upsampling, with also a suitable flattener, which can be placed after upsampling, or merged with the upsampler. For downsampling from 176.4kHz to 88.2kHz the combined downsampling and upsampling droop of 2.25dB @ 20kHz can be reduced to 0.12dB using a short flattener such as:

$$2.1451346747 - 1.4364916731 z^{-1} + 0.2913569984 z^{-2}$$ at 176.4kHz.

The total upsampling and downsampling response is then FIR with just 7 taps, hence a total temporal extent of six sample periods at the 176.4 sample rate or three sample periods at the downsampled rate. This is the shortest total filter response known to us that is often audibly satisfactory and maintains a flat response over 0-20kHz.

The infra-red prescription does not provide the strong rejection of downward aliasing considered desirable for signals with a strongly rising noise spectrum but there are many commercial recordings whose ultrasonic noise spectra are more nearly flat or are falling. With a downsampling ratio of 2:1 the slope of an infra-red downsampling filter is -9.5dB/8ve at the downsampled Nyquist frequency; with a ratio of 4:1 it is - 11.4 dB/8ve and in the limiting case of downsampling from continuous time it is -12dB/8ve. This compares with a slope of -22.7dB/8ve for the downsampling filter of figure 5A and for this type of source material the infra-red encoding specification may not be suitable.

An encoder for routine professional use should ideally attempt to determine the ultrasonic noise spectrum of material presented for encoding, for example by measuring the ultrasonic spectrum during a quiet passage, and thereby make an informed choice of the optimal downsampling and upsampling filter pair to reconstruct that particular recording. The choice then should be communicated as metadata to the corresponding decoder, which can then select the appropriate upsampling filter.

The above discussion has concentrated substantially on downsampling from a '4χ' sampling rate such as 192kHz or 176.4kHz to a '2χ' sampling rate such as 96kHz or 88.2kHz, but of commercial importance also is downsampling from a 4x or a 2x sampling rate to a 1x sampling rate such as 48kHz or 44.1 kHz. In fact the same 'infra-red' coefficients ¼ +½ z~1+1/4 z~2 as discussed above for use at higher sampling rates havealso been found to provide audibly good results when downsampling from 88.2kHz to 44.1 kHz. This is perhaps surprising as one might have expected that the ear would require greater rejection of downward aliased images of original frequencies at this lower sample rate, but repeated listening tests have confirmed that this does not seem to be the case. The same filter can be used for upsampling, combined with or followed by a flattener. At this lower sample rate, a flattener with more taps is needed, for example the filter:

$$4.0185 - 5.9764z^{-1} + 4.6929\zeta^{-2} - 2.4077\,\zeta^{-3} + 0.8436\zeta^{-4} - 0.1971\,\zeta^{-5} + 0.0279z^{-6} - 0.0018\zeta^{-7}$$

running at 88.2kHz, flattens the total response of downsampler and the upsampler to within 0.2dB at 20kHz and has found to be audibly satisfactory.

A flattener and unflattener pair can be provided as was described previously to allow compatibility with 44.1 kHz reproducing equipment. To provide a maximally flat response with a droop not exceeding 0.5dB at 20kHz, a nine-tap all-pole flattener implemented at 44.1 kHz is theoretically required:

$$\frac{1.2305}{1 + 0.2489\,z^{-1} - 0.0231\,z^{-2} + 0.0058\,z^{-3} - 0.0015\,z^{-4} + 0.0003\,z^{-5} - 0.0001\,z^{-6} + 0.8166\,10^{-5}z^{-7} - 0.7262\,10^{-6}z^{-8} + 0.3151\,10^{-7}z^{-9}}$$

though some of the later terms of the denominator here given could be deleted with minimal introduction of passband ripple. Either way, the expression here given can be inverted to provide a corresponding FIR unflattener. A high-resolution decoder would typically unflatten at 44.1 kHz, upsample to 88.2kHz and then flatten using an optimally-designed flattener at 88.2kHz such as the 7th order FIR flattener given above. In this case, the impuse response of the encoder and high-resolution decoder together has 12 nonzero taps, whereas the the encoder alone has an impulse response that continues longer, albeit at lower levels such as -40dB to -60dB.

One or both of the flattening and unflattening filters presented here for operation at the 44.1 kHz rate could be transformed as indicated previously to provide the same functionality when operated at 88.2kHz or a higher rate, if this is more convenient.

Reconstruction as described above to continuous time from a 44.1 kHz infra-red coding of an impulse presented as a single sample at time t=0 within an 88.2kHz stream is illustrated in figures 12A and 12B. In figure 12A the reconstruction is from 44.1 kHz samples, shown as diamonds, coincident in time with even samples of the 88.2kHz stream, whereas in figure 12B the reconstruction is from 44.1 kHz samples, shown as circles, coincident with odd samples of the 88.2kHz stream points. The horizontal axes is time t in units of 88kHz sample periods and the vertical axes shows amplitude raised to the power 0.21 , which provides visibility of small responses but also may have some plausibility according to neurophysiological models of human hearing which suggest that for short impulses, peripheral intensity is proportional to amplitude raised to the power 0.21 . The 44.1 kHz representations have been derived using the infra-red method as described above including flattening for compatibility with legacy equipment, while the two high-resolution reconstructions similarly use a legacy unflattener followed by infra-red reconstruction and a flattener implemented at 88.2kHz.

It will be noted that the 44kHz stream shows a time response that continues long after the high resolution reconstruction of the impulse has ceased, thus demonstrating the effectiveness of the pole-zero cancellation in providing an end-to-end response that is more compact than the response of the encoder alone.

Figures 12A and 12B also illustrate that the concept of an 'impulse response' needs to be defined more clearly when decimation is involved. In the case of decimation-by-2 the result is different for an impulse presented on an odd sample from that on an even sample. In this document we use the term 'impulse response' to refer to the *average* of the responses obtained in these two cases.

It will be appreciated that infra-red coding as described provides two z-plane zeroes at the sampling frequency of the downsampled signal, and in the case of a

downsampling ratio greater than 2, at all multiples of that frequency. This may be considered the defining feature of infra-red coding.

Suppression of downward aliasing

As noted, when encoding an item such as 'take 5", see figure 8, it may be desirable that the downsampling filter provide strong attenuation at frequencies such as 55kHz where the noise spectrum peaks. It would be natural to think of placing one or more z-plane zeroes to suppress energy near this frequency. To do so would however increase the total length of the end-to-end impulse response: firstly because each complex zero requires a further two taps on the downsampling filter, and secondly because a zero near 55kHz adds significantly to the total droop so a longer flattening filter will likely also be required.

With one caveat, the increase in length can be avoided using pole-zero cancellation: the complex zero in the encoder's filter is cancelled by a pole in the decoder. In one embodiment, a downsampling filter incorporating three such zeroes is paired with an upsampling filter having three corresponding poles. The resulting downsampling and upsampling filter responses are shown in figure 13A and figure 13B and the end-to-end response from combining these two filters with an assumed external droop is shown in figure 13C. For consistency with other graphs, these plots assume a sampling rate of 196kHz so the maximum attenuation is near 60kHz rather than 55kHz.

The caveat here is that although downward aliasing has been suppressed, upward aliasing has been increased. For use on tracks such as Take 5', the increased upward-aliased noise increase is well covered by the steeply-rising original noise. However signal components near 33kHz would also result in much larger aliases near 55kHz. It is thus arguably misleading simply to present an end-to-end frequency response that ignores aliased components; nevertheless it appears that the ear is relatively tolerant to the upward aliases provided the boost applied to the alias is not excessive.

The heavy boost of 38dB at 57kHz shown in figure 13B may seem at first unwise, but if a legacy flattener is used as described above then the decoder will

incorporate a legacy unflattener which will compensate most of this boost, so the decoder as a whole will not exhibit the boost.

Concluding Remarks

It is to be noted that some of the decoding responses described in this document have features that would normally be absent from reconstruction filters. These features include a response that is rising rather than falling at the half-Nyquist frequency of 44.kkHz or 48kHz, and a z-transform having one or more factors that are functions of even powers of z only, and thereby have individual responses that are symmetrical about the half-Nyquist frequency

In the time domain the total system impulse response may be shorter than the encoder's impulse response. In assessing whether this is the case, a threshold may need to be applied to exclude very small responses that are unlikely to be of audible significance. However there is little data on which to determine what the value of that threshold should be.

Claims

1. A system comprising an encoder and a decoder for conveying the sound of an audio capture, wherein the encoder is adapted to furnish a digital audio signal at a transmission sample rate from a signal representing the audio capture, and the decoder is adapted to receive the digital audio signal and furnish a reconstructed signal,

wherein the encoder comprises a downsampler adapted to receive the signal representing the audio capture at a first sample rate which is a multiple of the transmission sample rate and to downsample the signal to furnish the digital audio signal; and,

wherein an impulse response of the encoder and decoder in combination has a largest peak, and is characterised by a contiguous time region having an extent not greater than 6 sample periods of the transmission sample rate outside of which the absolute value of the impulse response does not exceed 10% of said largest peak.

2. A system according to claim 1, wherein the downsampler comprises a decimation filter specified at the first sample rate, wherein the asymmetric component of response of the decimation filter is characterised by an attenuation of at least 32dB at frequencies that would alias to the range 0-7 kHz on decimation.

3. A system according to claim 2, wherein the impulse response of the decimation filter has a largest peak, and the asymmetric component of the response is characterised by a contiguous time region having an extent not greater than 6 sample periods of the transmission sample rate outside of which the absolute value of the impulse response does not exceed 10% of said largest peak.

4. A system according to any one of claims 1 to 3, wherein the encoder comprises an Infinite Impulse Response (IIR) filter having a pole, and the decoder comprises a filter having a zero whose z-plane position coincides with that of the pole, the effect of which is thereby cancelled in the reconstructed signal.

5. A system according to any one of claims 1 to 3, wherein the decoder comprises an Infinite Impulse Response (IIR) filter having a pole, and the encoder comprises a filter having a zero whose z-plane position coincides with that of the pole, the effect of which is thereby cancelled in the reconstructed signal.

6. A system according to any preceding claim, wherein the decoder comprises a filter having a response which rises in a region surrounding the Nyquist frequency corresponding to the transmission sample rate and the encoder comprises a filter having a response that falls in said region, thereby reducing downward aliasing in the encoder of frequencies above the Nyquist frequency to frequencies below the Nyquist frequency.

7. A system according to any preceding claim, wherein the transmission sample rate is selected from one of 88.2kHz and 96kHz and the first sample rate is selected from one of 176.4kHz, 192kHz, 352.8kHz and 384kHz.

8. A system according to any preceding claim, wherein said contiguous time region has an extent not greater than 4 periods of the transmission sample rate.

9. A system according to any preceding claim, wherein the impulse response rises monotonically to the largest peak.

10. A method of furnishing a digital audio signal for transmission at a transmission sample rate by reducing the sample rate required to convey the sound of captured audio, the method comprising the steps of:

filtering a representation of the captured audio having a first sample rate that is a multiple of the transmission sample rate using a decimation filter specified at the first sample rate; and,

decimating the filtered representation to furnish the digital audio signal,

wherein an impulse response of the decimation filter has a largest peak and an asymmetric component of response of the decimation filter is characterised by

(i) an attenuation of at least 32dB at frequencies that would alias to the range 0-7 kHz on decimation; and,

(ii) a contiguous time region having an extent not greater than 6 sample periods of the transmission sample rate outside of which the absolute value of the impulse response does not exceed 10% of the largest peak.

11. A method according to claim 10, further comprising the step of establishing the representation of the captured audio at the first sample rate.

12. A method according to claim 10 or claim 11, further comprising the steps of:

analysing a spectrum of the captured audio; and,

choosing the decimation filter responsively to the analysed spectrum.

13. A method according to claim 12, further comprising the step of furnishing information relating to the choice of decimation filter for use by a decoder.

14. A method according to any one of claims 10 to 13, wherein the transmission sample rate is selected from one of 88.2kHz and 96kHz and the first sample rate is selected from one of 176.4kHz, 192kHz, 352.8kHz and 384kHz.

15. A method according to any one of claims 10 to 14, wherein said contiguous time region has an extent not greater than 4 periods of the transmission sample rate.

16. A method according to any one of claims 10 to 15, wherein the impulse response rises monotonically to the largest peak.

17. A data carrier comprising a digital audio signal furnished by performing the method according to any one of claims 10 to 16.

18. An encoder for an audio stream, wherein the encoder is adapted to furnish a digital audio signal using the method of any one of claims 10 to 18.

19. An encoder according to claim 18, comprising a flattening filter having a symmetrical response about the transmission Nyquist frequency.

20. An encoder according to claim 19, wherein the flattening filter has a pole.

21. A system for conveying the sound of an audio capture, the system comprising an encoder and a decoder,

wherein the encoder is adapted to receive a signal representing the audio capture and to furnish a digital audio signal, and the decoder is adapted to receive the digital audio signal and furnish a reconstructed signal, and

wherein the encoder and decoder have impulse responses which, when combined, produce a total system impulse response that is more compact than an impulse response of the encoder alone.

22. A system according to claim 21, wherein the decoder comprises a filter having a z-plane zero whose position coincides with that of a pole in the response of the encoder.

23. A system according to claim 21 or claim 22, wherein the decoder comprises a filter chosen in dependence on information received from the encoder.

24. A system according to any one of claims 21 to 23, wherein an impulse response of the encoder and decoder in combination has a largest peak, and is characterised by a contiguous time region having an extent not greater than 6 sample periods of the transmission sample rate outside of which the absolute value of the averaged impulse response does not exceed 10% of said largest peak.

25. An encoder adapted to furnish a digital audio signal at a transmission sample rate from a signal representing an audio capture, the encoder comprising a downsampling filter having an asymmetric component of response equal to the asymmetric component of response of a filter whose frequency response has a double zero at each frequency that will alias to zero frequency and has a slope at the transmission Nyquist frequency more positive than minus thirteen decibels per octave.

26. An encoder according to claim 25, comprising a flattening filter having a symmetrical response about the transmission Nyquist frequency.

27. An encoder according to claim 26, wherein the flattening filter has a pole.

28. An encoder according to claim 26 or claim 27, wherein the transmission frequency is 44.1 kHz and the encoder's frequency response droop does not exceed 1dB at 20kHz.

29. A system comprising an encoder and a decoder for conveying the sound of an audio capture, wherein the encoder is adapted to furnish a digital audio signal at a transmission sample rate from a signal representing the audio capture, and the decoder is adapted to receive the digital audio signal and furnish a reconstructed signal,

wherein the encoder comprises a downsampler adapted to receive the signal representing the audio capture at a first sample rate which a multiple of the transmission sample rate and to downsample the signal to furnish the digital audio signal; and,

wherein the encoder comprises an Infinite Impulse Response (MR) filter having a pole, and the decoder comprises a filter having a zero whose z-plane position coincides with that of the pole, the effect of which is thereby cancelled in the reconstructed signal.

30. A system according to claim 29, wherein an impulse response of the encoder and decoder in combination has a largest peak, and is characterised by a contiguous time region having an extent not greater than 6 sample periods of the transmission sample rate outside of which the absolute value of the averaged impulse response does not exceed 10% of said largest peak.

31. An encoder adapted to furnish a digital audio signal at a transmission sample rate from a signal representing an audio capture, the encoder comprising a downsampling filter adapted to receive the signal representing the audio capture at a first sample rate which a multiple of the transmission sample rate and to downsample the signal to furnish the digital audio signal, wherein the

encoder is adapted to analyse a spectrum of the captured audio and select the downsampling filter responsively to the analysed spectrum.

32. An encoder according to claim 31, wherein the selected downsampling filter has a steeper attenuation response at the transmission Nyquist frequency if the analysed spectrum is rising rapidly at the transmission Nyquist frequency.

33. An encoder according to claim 31 or claim 32, wherein the encoder is adapted to transmit information identifying the selected downsampling filter to a decoder as metadata.

34. An encoder according to any one of claims 31 to 33, comprising a flattening filter having a symmetrical response about the transmission Nyquist frequency.

35. An encoder according to claim 34, wherein the flattening filter has a pole.

36. A decoder for receiving a digital audio signal at a transmission sample rate and furnishing an output audio signal, wherein the decoder comprises a filter having an amplitude response which increases with frequency in a frequency region adjoining the Nyquist frequency corresponding to the transmission sample rate

37. A decoder according to claim 36, wherein the filter has an amplitude response of at least +2dB at the Nyquist frequency corresponding to the transmission sample rate, relative to the response at DC.

38. A method according to claim 36 or claim 37, wherein the filter response is determined in dependence on information received from an encoder.

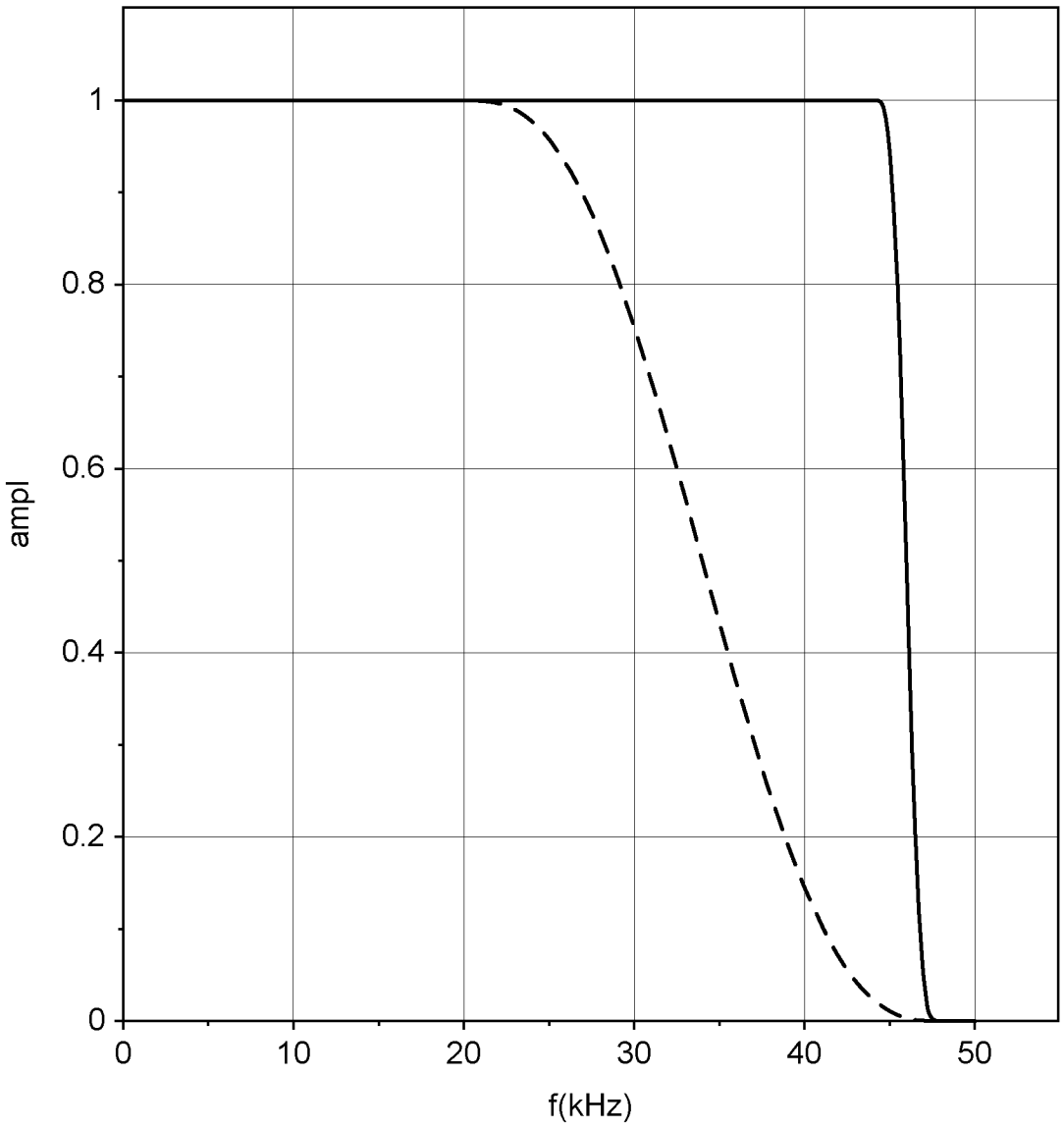

FIG. 1

(Prior art)

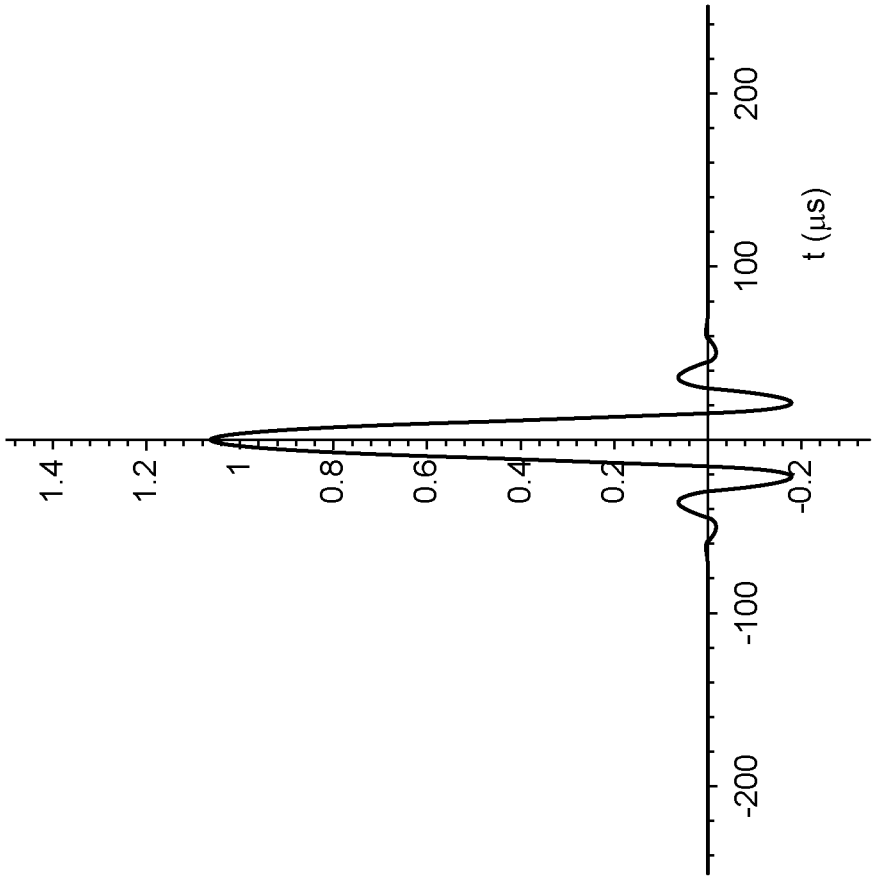

FIG. 2B
(Prior art)

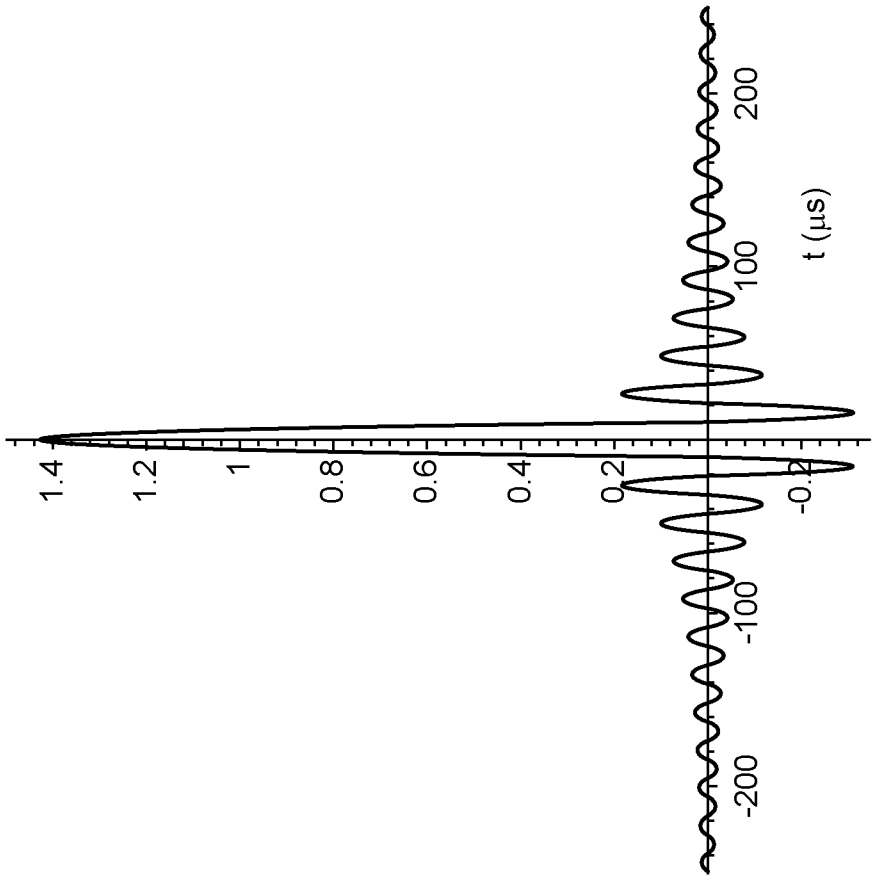

FIG. 2A
(Prior art)

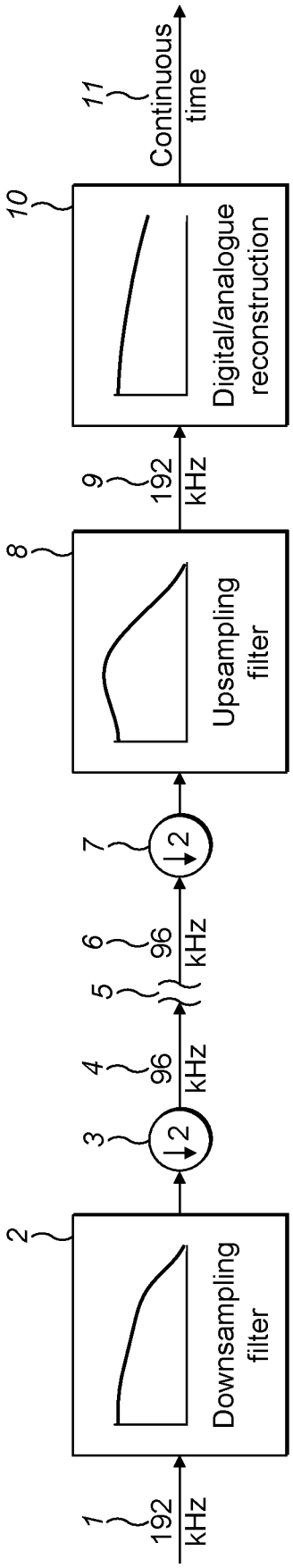
1
192
kHz
2
Downsampling filter
3
↓2
4
96
kHz
5
6
96
kHz
7
↓2
8
Upsampling filter
9
192
kHz
10
Digital/analogue reconstruction
11
Continuous time

FIG. 3

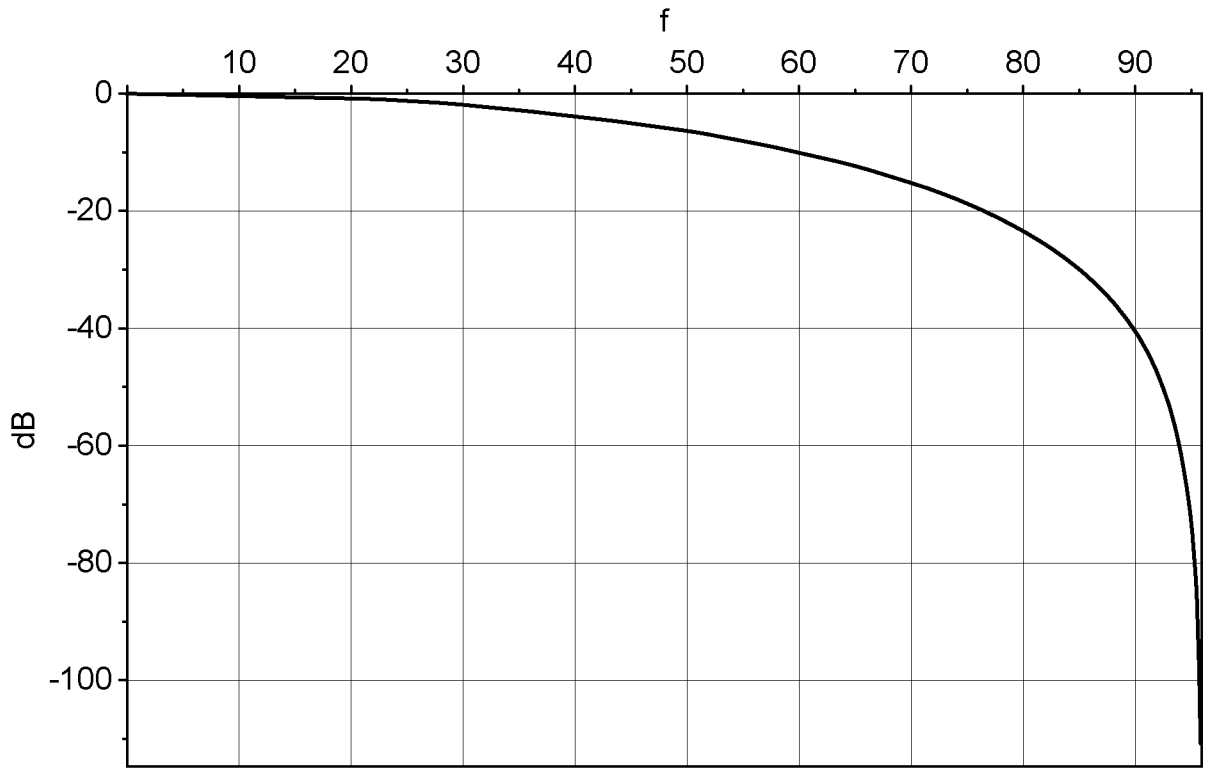

FIG. 4

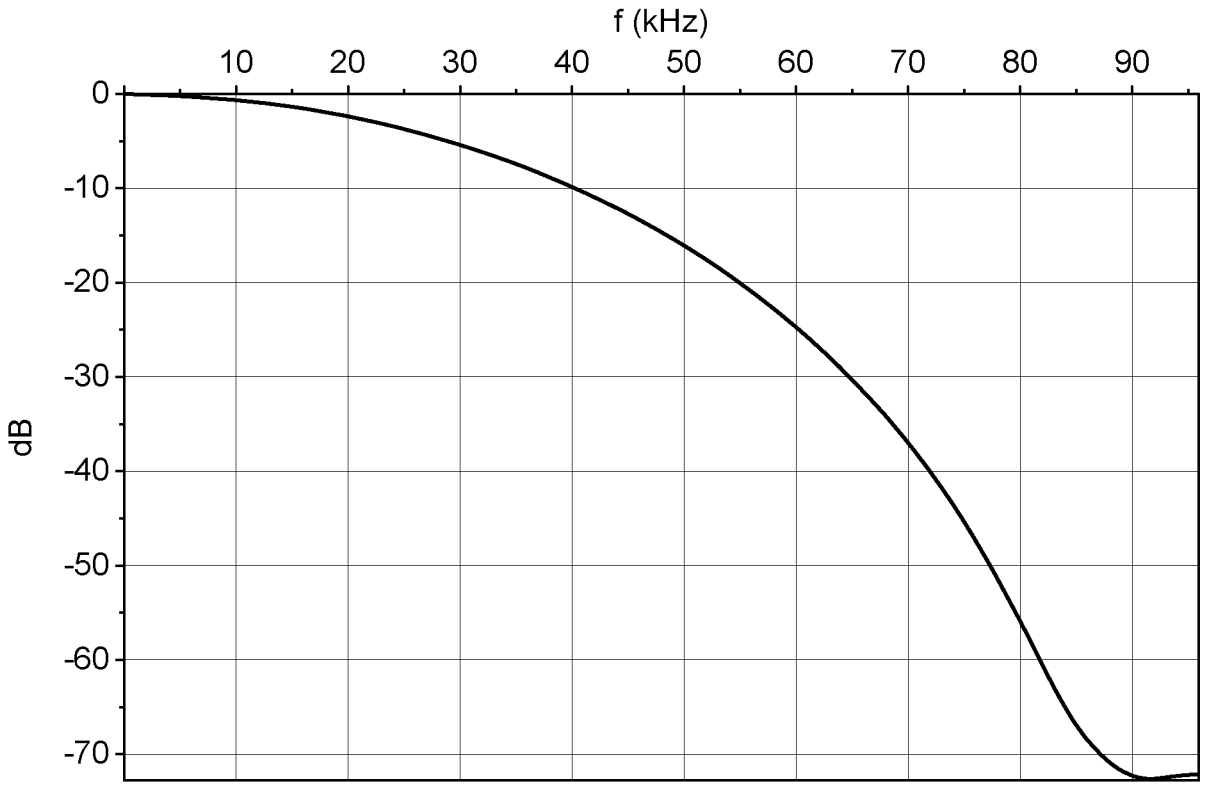
f (kHz)
10
20
30
40
50
60
70
80
90
0
-10
-20
-30
-40
-50
-60
-70
dB

FIG. 5A

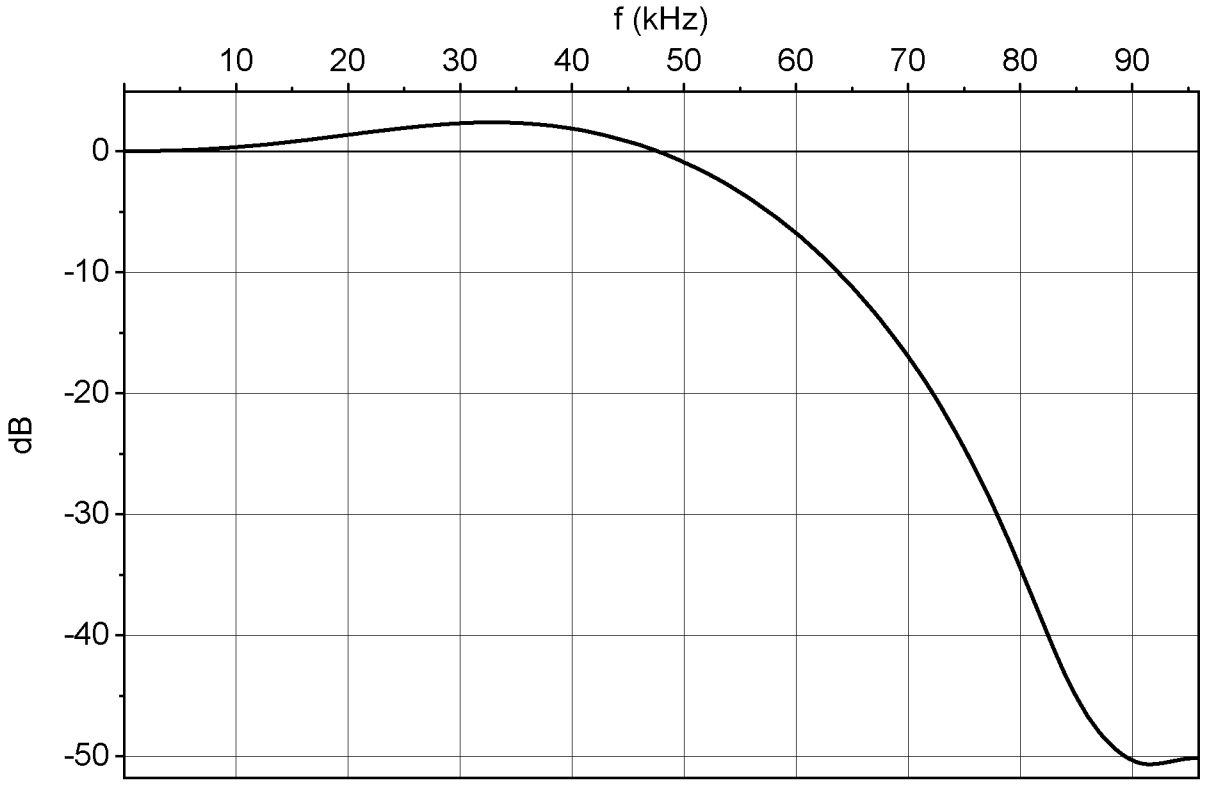
f (kHz)
10
20
30
40
50
60
70
80
90
0
-10
-20
-30
-40
-50
dB

FIG. 5B

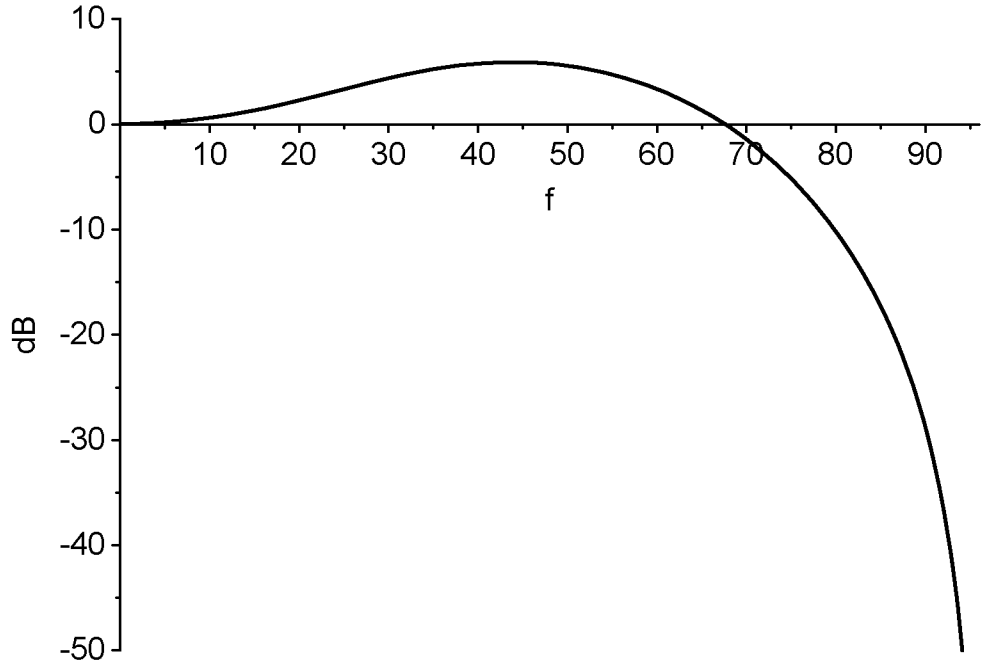
10
0
-10
-20
-30
-40
-50
dB
10
20
30
40
50
60
70
80
90
f

FIG. 6

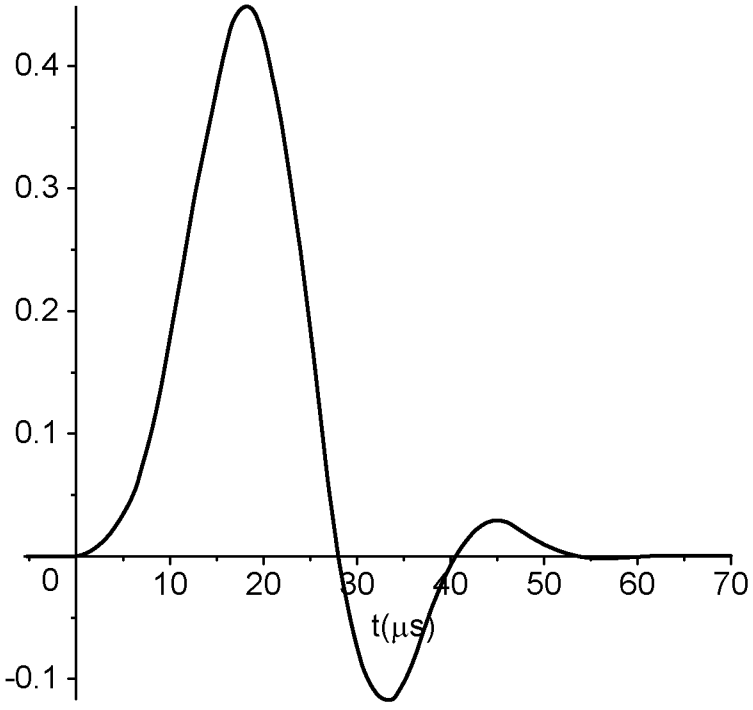
0.4
0.3
0.2
0.1
0
-0.1
10
20
30
40
50
60
70
t(μs)

FIG. 7

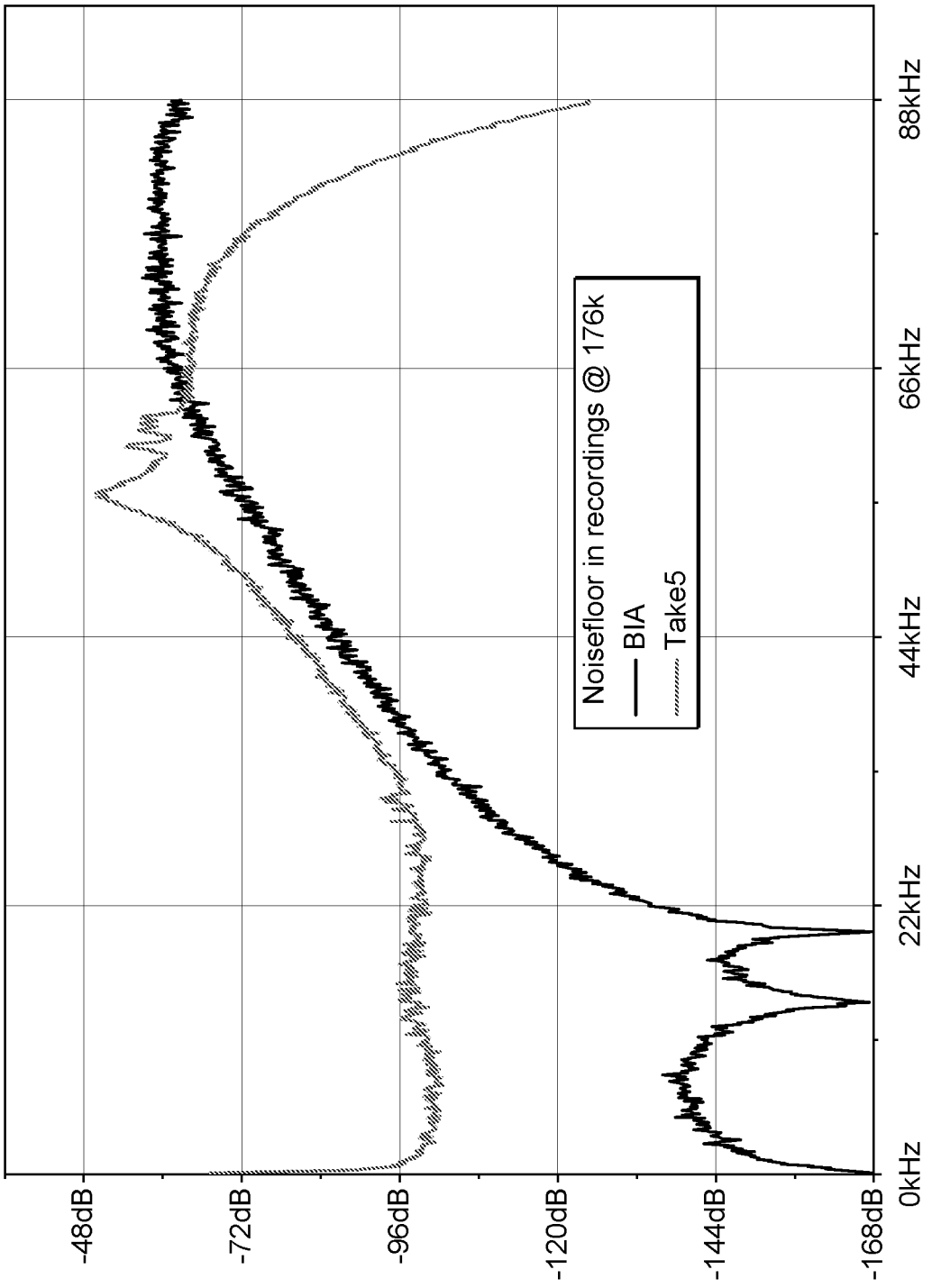
Noisefloor in recordings @ 176k
BIA
Take5
-48dB
-72dB
-96dB
-120dB
-144dB
-168dB
0kHz
22kHz
44kHz
66kHz
88kHz

FIG. 8

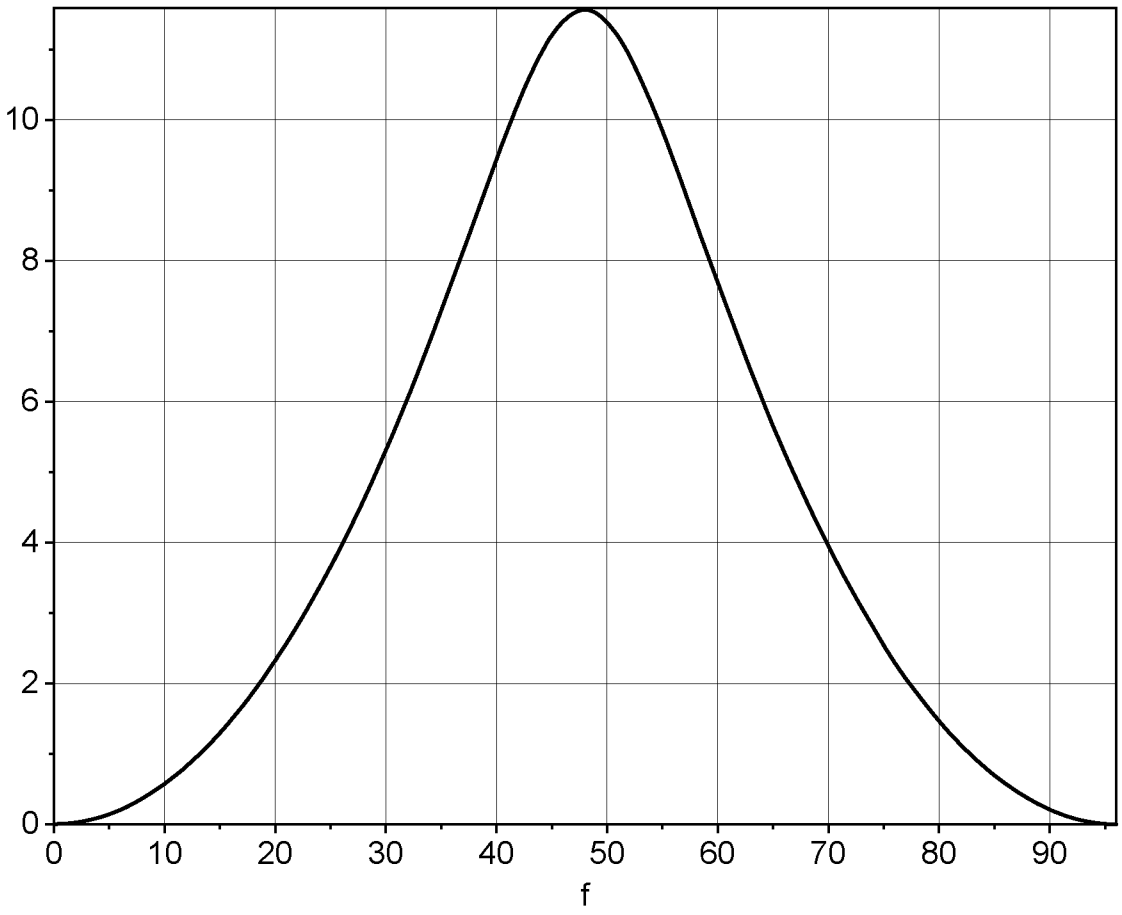

FIG. 9

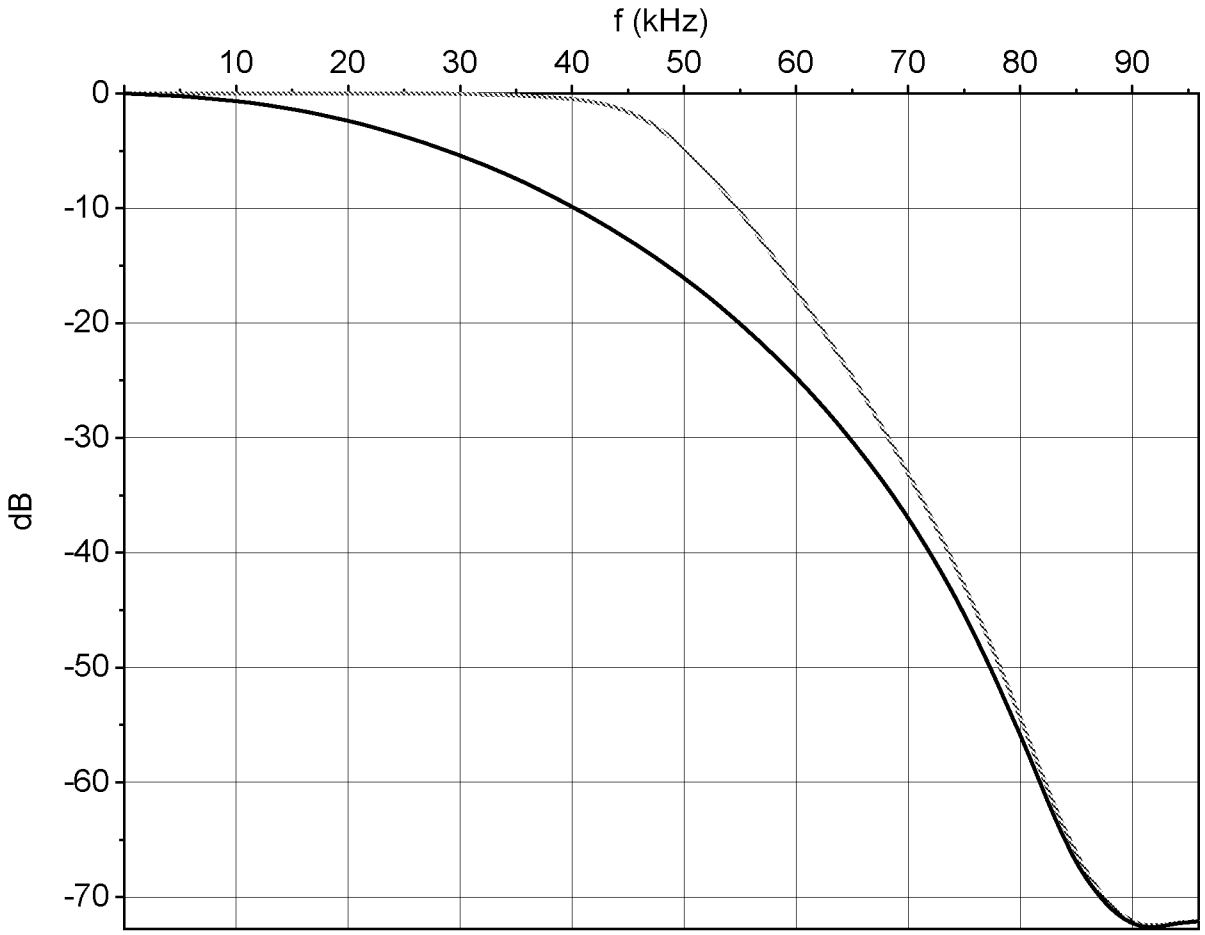
f (kHz)
10
20
30
40
50
60
70
80
90
0
-10
-20
-30
-40
-50
-60
-70
dB

FIG. 10

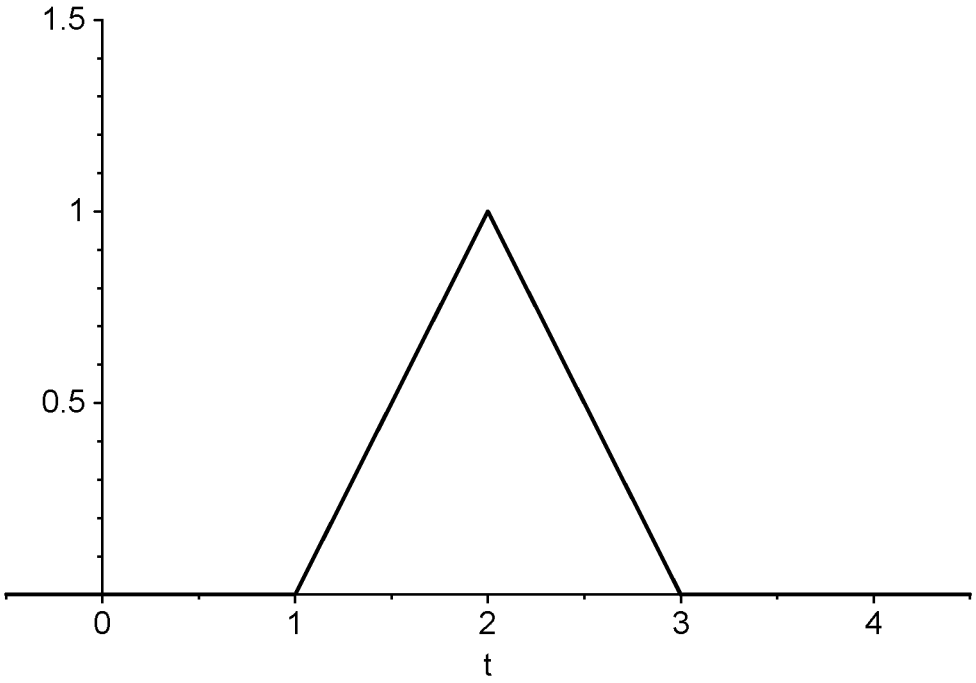
1.5
1
0.5
0
1
2
3
4
t

FIG. 11

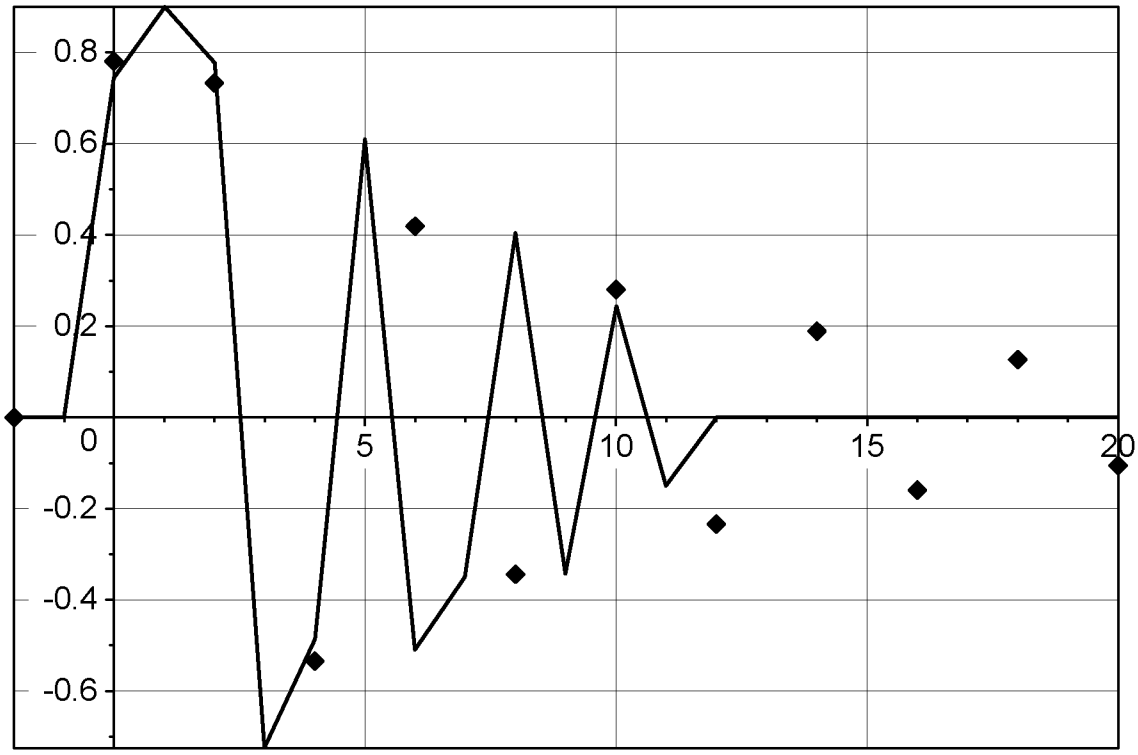

FIG. 12A

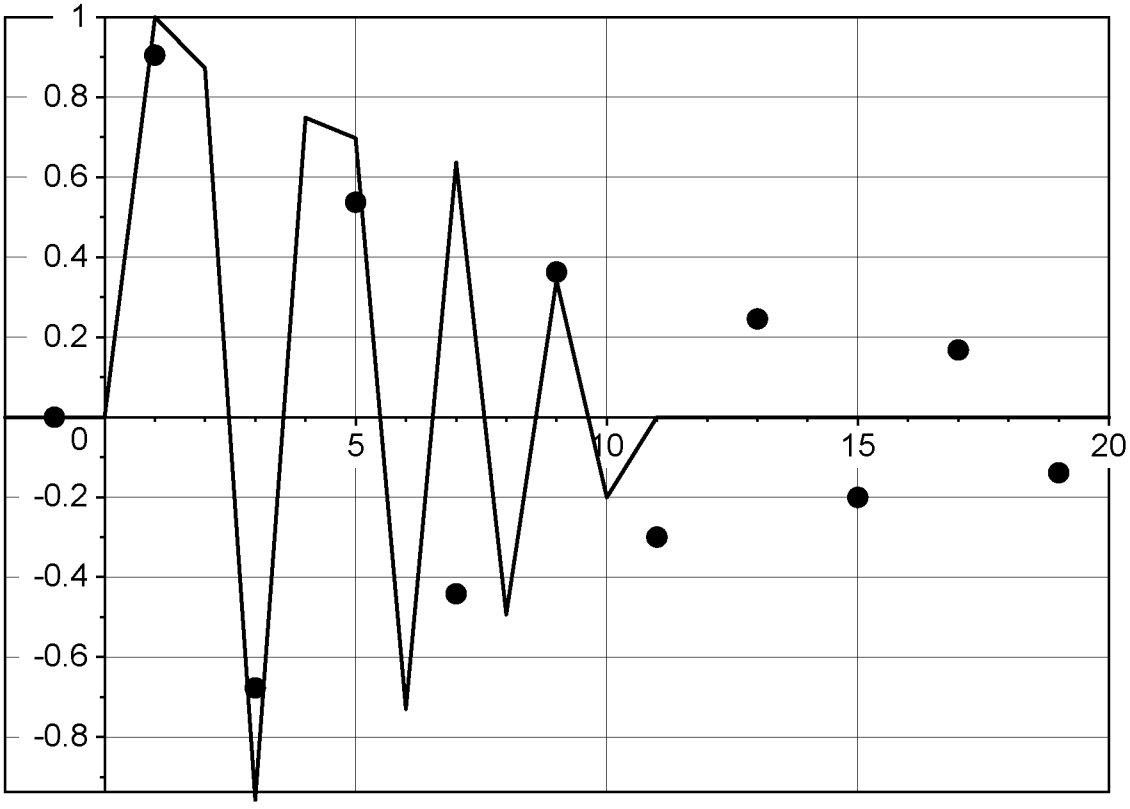

FIG. 12B

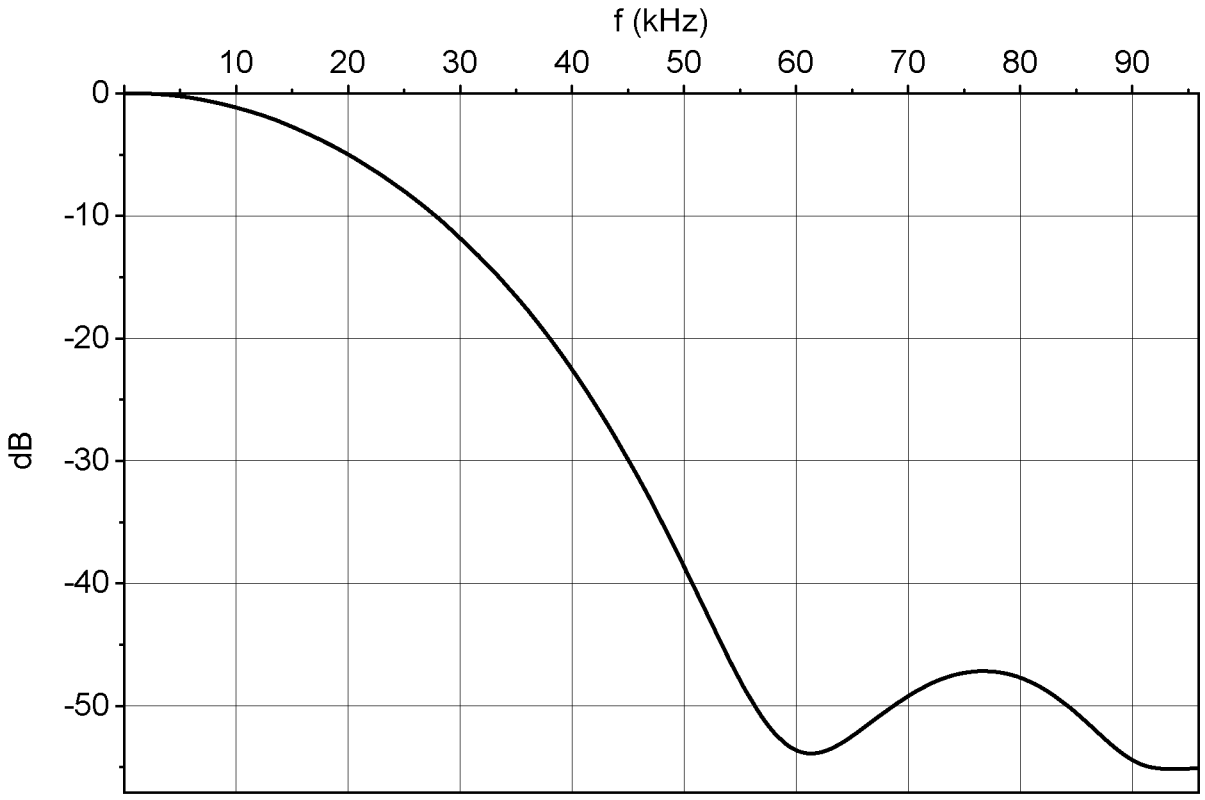

FIG. 13A

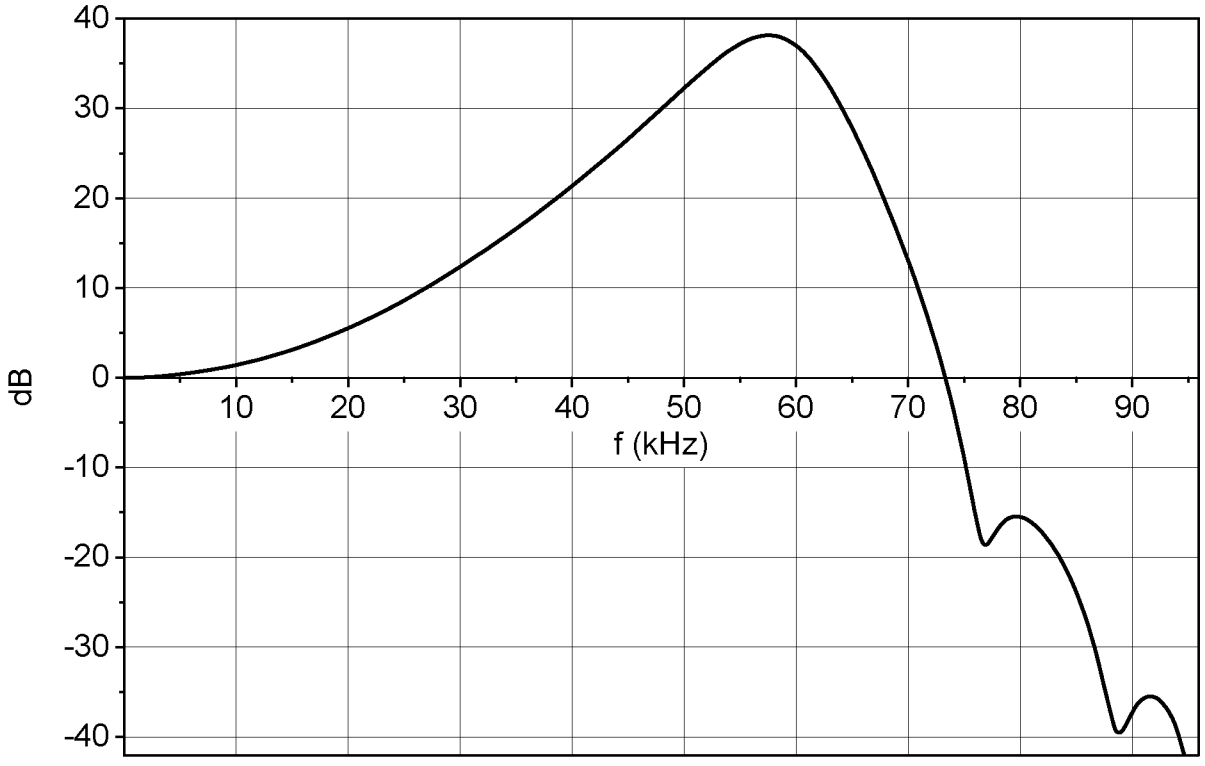

FIG. 13B

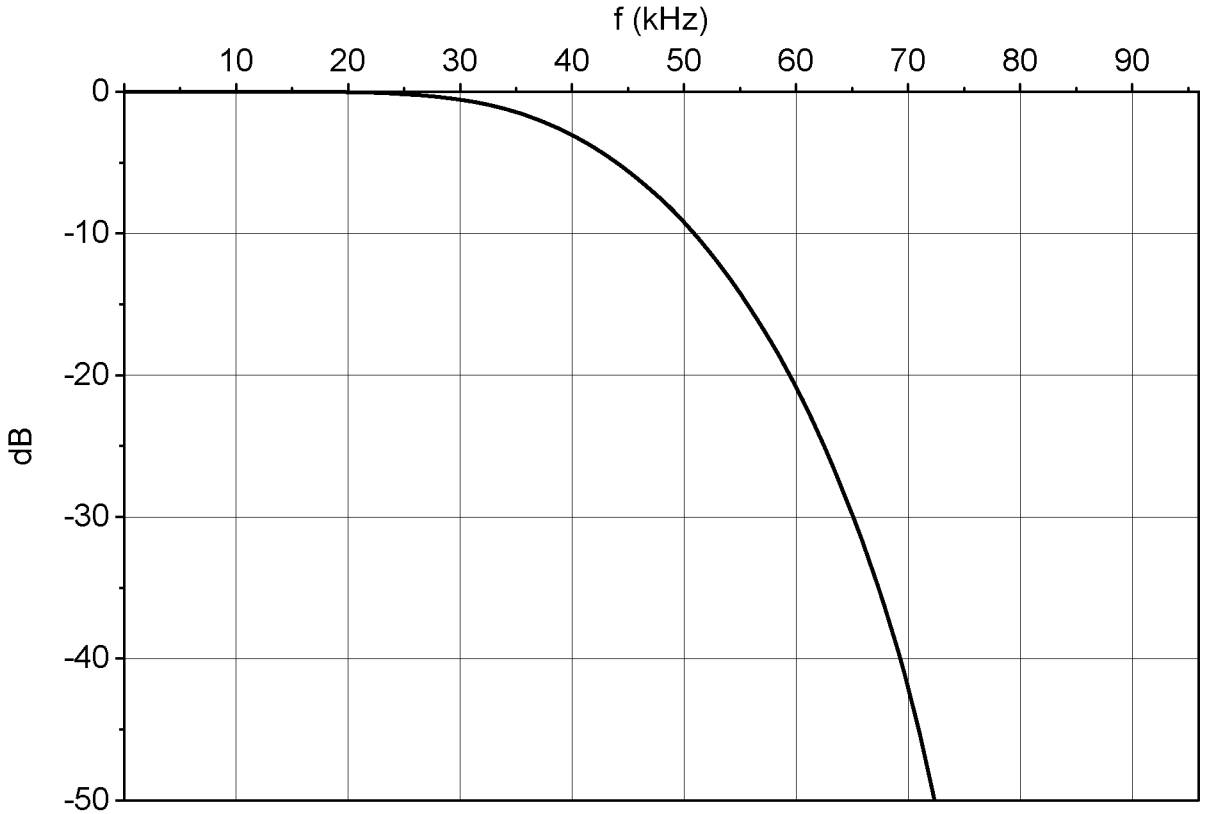

FIG. 13C

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2014/050040

A. CLASSIFICATION OF SUBJECT MATTER
INV. G11B20/10
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G11B G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 00/57549 A1 (PACIFIC MICROSONICS INC [US]) 28 September 2000 (2000-09-28) the whole document	1-38
Y	US 5 808 574 A (JOHNSON KEITH O [US] ET AL) 15 September 1998 (1998-09-15) column 10, line 60 - column 20, line 39; claims 1-13; figures 1-16 column 41, line 4 - line 36	1-38
Y	US 2009/027117 A1 (ANDERSEN JACK B [US] ET AL) 29 January 2009 (2009-01-29) paragraph [0018] - paragraph [0026]; claims 1-14; figures 1-36 paragraph [0088] - paragraph [0157]	1-38
A	EP 0 933 889 A1 (OLYMPUS OPTICAL CO [JP]) 4 August 1999 (1999-08-04) the whole document	1-38

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" documentwhich may throw doubts on priority claim(s) orwhich is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
9 April 2014	17/04/2014

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Durucan, Emrullah

1

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/GB2014/050040

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 0057549 A1	28-09-2000	AU 3903400 A	09-10-2000
		EP 1163721 A1	19-12-2001
		JP 4469090 B2	26-05-2010
		JP 2002540667 A	26-11-2002
		TW 457781 B	01-10-2001
		US 6337645 B1	08-01-2002
		WO 0057549 A1	28-09-2000
US 5808574 A	15-09-1998	AT 183328 T	15-08-1999
		AT 221690 T	15-08-2002
		AT 222019 T	15-08-2002
		AT 222396 T	15-08-2002
		AT 255267 T	15-12-2003
		AU 669114 B2	30-05-1996
		CA 2110182 A1	10-12-1992
		CA 2506118 A1	10-12-1992
		DE 69229786 D1	16-09-1999
		DE 69229786 T2	10-08-2000
		DE 69232713 D1	05-09-2002
		DE 69232713 T2	06-05-2004
		DE 69232729 D1	12-09-2002
		DE 69232729 T2	24-04-2003
		DE 69232734 D1	19-09-2002
		DE 69232734 T2	24-04-2003
		DE 69233256 D1	08-01-2004
		DE 69233256 T2	27-05-2004
		EP 0586565 A1	16-03-1994
		EP 0810599 A2	03-12-1997
		EP 0810600 A2	03-12-1997
		EP 0810601 A2	03-12-1997
		EP 0810602 A2	03-12-1997
		ES 2135408 T3	01-11-1999
		HK 1008368 A1	20-04-2000
		JP 3459417 B2	20-10-2003
		JP H06509201 A	13-10-1994
		US 5479168 A	26-12-1995
		US 5638074 A	10-06-1997
		US 5640161 A	17-06-1997
		US 5808574 A	15-09-1998
		US 5838274 A	17-11-1998
		US 5854600 A	29-12-1998
		US 5864311 A	26-01-1999
		US 5872531 A	16-02-1999
		WO 9222060 A1	10-12-1992
US 2009027117 A1	29-01-2009	NONE	
EP 0933889 A1	04-08-1999	EP 0933889 A1	04-08-1999
		JP H11215006 A	06-08-1999