

(19)



Europäisches Patentamt

European Patent Office

Office européen des brevets



(11)

**EP 0 959 644 A2**

(12)

**EUROPEAN PATENT APPLICATION**

(43) Date of publication:

**24.11.1999 Bulletin 1999/47**(51) Int Cl.<sup>6</sup>: **H04S 3/00**(21) Application number: **99303966.8**(22) Date of filing: **21.05.1999**

(84) Designated Contracting States:

**AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU  
MC NL PT SE**

Designated Extension States:

**AL LT LV MK RO SI**(30) Priority: **22.05.1998 GB 9811054**(71) Applicant: **CENTRAL RESEARCH****LABORATORIES LIMITED****Hayes, Middlesex, UB3 1HH (GB)**

(72) Inventors:

- **Sibbald, Alastair**  
**Maidenhead, Berkshire SL6 1XL (GB)**
- **Nackvi, Fawad**  
**Southall, Middlesex UB2 5PR (GB)**

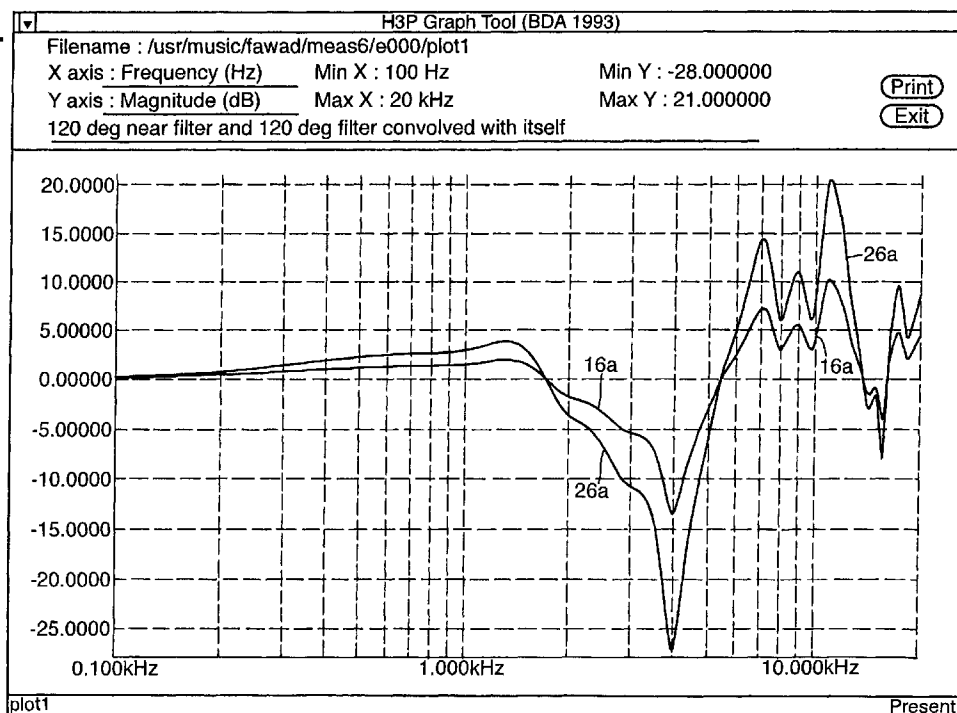
(74) Representative: **Sharp, Alan Cooper et al****QED I.P. Services Limited,****Dawley Road****Hayes, Middlesex UB3 1HH (GB)**(54) **Method of modifying a filter for implementing a head-related transfer function**

(57) A head-related transfer function (HRTF) is used to place a virtual sound source at a particular position in 3D space. The absence of modelled sound reflections from the virtual sound source, and the presence of unwanted real reflections from real sound sources, can impair the effectiveness of the positioning of the virtual sound source. The invention describes a method of

modifying a filter for implementing an HRTF (13), whereby the spectral profile of the HRTF is exaggerated by convolving the near-ear (16a, 18a) and/or far-ear (16b, 18b) transfer function with itself. This results in more effective placement of virtual sound images in 3D space, giving improved realism of 3D effects.

The invention is of particular use in the virtualisation of multi-channel surround sound systems.

Fig.7.



## Description

**[0001]** This invention relates to a method of modifying a filter for implementing a head-related transfer function (HRTF) for use in the reproduction of three-dimensional (3D) sound.

**[0002]** The processing of binaural (two channel or stereo) audio signals to produce highly realistic 3D sound images is well known. One method is described in International Patent Application No. WO-A1-9422278, and is known as the Sensaura™ system. This system is based on recordings made using a so-called "artificial head" microphone system, and the recordings are subsequently processed digitally. The use of the artificial head ensures that natural 3D sound cues - which the brain uses to determine the position of sound sources in 3D space - are incorporated into the stereo recordings. 3D sound cues are introduced naturally by the head and ears when we listen to sounds in real life, and they include the following characteristics: inter-aural amplitude difference (IAD), inter-aural time delay (ITD), and spectral shaping by the outer ear.

**[0003]** By electronically synthesising these natural acoustic processes, it is possible to create "virtual" sound sources for headphone and loudspeaker reproduction. To set the position of a single channel virtual sound source in a plural channel system, separate audio filters for the left and right channels of the audio signal, together with a relative time delay, introduce the above mentioned characteristics. The filters used, and the time delay introduced, depend on the desired position of the virtual sound. The characteristics themselves are initially determined by measurement of an appropriate head-related transfer function (HRTF). The HRTF characterises the modifications which an audio signal undergoes on its path from a point in space, at a defined direction and distance from a listener, to the eardrums of the listener. An HRTF comprises a left-ear transfer function, a right-ear transfer function, and an inter-aural time delay. A block diagram of the synthesis of a virtual sound source is shown in Figure 3.

**[0004]** When a pair of audio signals incorporating such 3D sound cues are introduced efficiently into the ears of the listener, by headphones for example, then he or she perceives a virtual sound source to be located at the associated position in 3D space. However, if the processed signals are not conveyed directly and efficiently into the ears of the listener, then the full 3D effects will not be perceived. For example, when listening to sounds via conventional stereo loudspeakers, the left ear hears a little of the right loudspeaker signal, and vice versa. This is known as transaural crosstalk. By cancelling out transaural crosstalk, full 3D effects can be enjoyed via loudspeakers remote from the listener. Transaural crosstalk from each of the loudspeakers may be cancelled by creating appropriate crosstalk cancellation signals from the opposite loudspeaker. Crosstalk cancellation signals are equal in magnitude and inverted

(opposite in polarity) with respect to the transaural crosstalk signals. A system for performing transaural crosstalk cancellation is discussed in the published International Patent Application No. WO-A1-9515069.

**[0005]** When listening to a real sound source in an ordinary environment (e.g. a living room), the first sound that the listener hears is termed the "direct" sound (so called because it travels directly to the ears). The direct sound is soon followed by the first reflections from the floor, ceiling and walls, some milliseconds later (or tens of milliseconds, depending on the dimensions of the room). The first reflections are themselves reflected back again to the listener from other boundaries, and these sound waves are termed secondary reflections, or second-order reflections. This process continues until the sound energy has been totally absorbed by the boundaries of the environment, and by the air itself. The reflections which follow the first few reflections soon begin to overlap each other, becoming complex and scattered, and are termed the reverberant sound.

**[0006]** Because the placing of a virtual sound source using HRTF filters uses a considerable amount of computational effort, it is common to simulate only the direct sound, and not the reflections. Consequently, the resulting virtual sound is anechoic, that is, it lacks the reflected components. This can be a disadvantage, as such reflected components can help the brain determine distance and reinforce spatial effects.

**[0007]** A further limitation in conventional 3D sound reproduction is that when reproducing virtual sounds via loudspeakers, the sounds originating from the loudspeakers themselves may be reflected from surfaces such as walls, floor, ceiling, and furniture. These sound reflections may conflict with the virtual sound image, especially if the virtual sound image is placed behind the listener. This is because sound reflections from room boundaries close to the loudspeaker "overwhelm" the 3D cue arising from spectral shaping by the outer ear, and so the inter-aural time delay (ITD) cue predominates. This causes the virtual sound source to flip from the required rearward position to a position in front of the listener which shares the same ITD value.

**[0008]** It can be concluded that the absence of synthesised sound reflections in the virtual image, in addition to the presence of real reflections from room boundaries, can impair the effectiveness of positioning the virtual sound source.

**[0009]** An example illustrating this point is the virtualisation of rear surround speakers for the Dolby AC-3 5.1 system. Dolby and AC-3 are trademarks of Dolby Laboratories Inc. An audio system incorporating the AC-3 compression standard provides for multi-channel digital surround sound. AC-3 5.1 gives separate audio channels for left, right, and centre speakers in front of the listening position, two rear surround speakers, and a sub-woofer positioned according to the listener's preference. A typical loudspeaker configuration for the AC-3 system is shown in Figure 4.

**[0010]** Figures 1 and 2 show a co-ordinate system used for the following description. The convention chosen here for referring to azimuth angles is that they are measured from the frontal pole  $P$  towards the rear pole  $P'$ , with positive values of azimuth on the right-hand side of the listener and negative values on the left-hand side. Rear pole  $P'$  is at an azimuth of  $+180^\circ$  (and  $-180^\circ$ ). Angles of elevation are measured directly upwards (or downwards, for negative angles) from the origin at the centre of the head of the listener relative to the horizontal plane.

**[0011]** The preferred positions of the rear surround speakers in the AC-3 system are  $\pm 120^\circ$  azimuth and  $0^\circ$  elevation. Therefore, the use of a  $+120^\circ$ , and a  $-120^\circ$ , HRTF is required. However, the characteristics of the  $+120^\circ$  and  $-120^\circ$  HRTF are very similar to those of the  $+60^\circ$  and  $-60^\circ$  HRTF: the inter-aural time delays for both HRTFs are identical ( $522 \mu\text{s}$ ). Consequently, when attempts are made to create a virtual sound source at  $+120^\circ$  (or  $-120^\circ$ ), the presence of unwanted reflections from room boundaries adjacent the loudspeakers, in addition to the absence of virtual reflections from the virtual sound source, causes the image to flip to the  $+60^\circ$  (or  $-60^\circ$ ) position. Thus sounds placed at an azimuth of  $+120^\circ$  (or  $-120^\circ$ ) appear to be in front of the listener at  $+60^\circ$  (or  $-60^\circ$ ), and the illusion of the surround sound effect is disturbed.

**[0012]** An aim of the present invention is to provide more effective virtual sound source placement in three dimensions, particularly, but not exclusively, for virtual sound sources placed behind a listener, by modification of the characteristics of a filter for implementing a head-related transfer function.

**[0013]** According to a first aspect of the invention there is provided a method of modifying the characteristics of a filter for implementing a head-related transfer function (HRTF), the HRTF including a near-ear transfer function and a far-ear transfer function, the method comprising increasing the magnitude of the amplitude of the near-ear transfer function and/or far-ear transfer function over a range of frequencies to give an exaggerated near-ear transfer function and/or an exaggerated far-ear transfer function, the amount of the increase at a given frequency being a function of the amplitude of the corresponding transfer function or functions at the given frequency, thereby forming a filter which implements an HRTF having an exaggerated near-ear transfer function and/or an exaggerated far-ear transfer function.

**[0014]** Preferably the magnitude of the amplitude of the near-ear transfer function, and/or the far-ear transfer function, is increased by convolving the transfer function with itself.

**[0015]** The amplitude of the exaggerated near-ear transfer function and/or the amplitude of the exaggerated far-ear transfer function may be limited over a range of frequencies above a threshold value. The threshold value may be, for example, 6 kHz.

**[0016]** The amplitude of the exaggerated near-ear

transfer function and/or the amplitude of the exaggerated far-ear transfer function may be adjusted so that the amplitude of the exaggerated near-ear transfer function and the amplitude of the exaggerated far-ear transfer function tend to the same value at frequencies below, for example, 100 Hz.

**[0017]** According to another aspect of the invention, there is provided a filter modified using the aforementioned method. Preferably the modified filter is used for implementing an HRTF, the HRTF having an amplitude response characteristic curve substantially as shown in plot B of Figure 8.

**[0018]** The filter may also include crosstalk cancellation means. The filter may be used in a multi-channel surround sound system, or a multi-channel encoding system.

**[0019]** Preferably the modified filter for implementing an HRTF places a virtual sound source at positions behind a listener. For AC-3, or other, surround sound systems, preferably the virtual sound sources are placed at azimuths of  $\pm 120^\circ$  and elevations of  $0^\circ$  relative to a listener. For different applications such as AC-3, or other, mastering (or encoding) applications, preferably the virtual sound source is placed at an elevation of  $\pm 90^\circ$  relative to a listener. Preferably the modified filter is a finite impulse response filter.

**[0020]** According to another aspect of the invention, there is provided a sound recording or transmission made using a modified filter implementing an HRTF.

**[0021]** According to a further aspect of the invention, there is provided a signal processed using a modified filter implementing an HRTF.

**[0022]** The invention will now be described, by way of example only, with reference to the accompanying Figures, in which:-

Figure 1 shows the head of a listener within a reference sphere, and a co-ordinate system;

Figure 2 shows the position of a sound-source on the reference sphere with respect to the listener;

Figure 3 shows a schematic representation of the conventional method for creating a virtual sound source;

Figure 4 shows a schematic representation of a typical Dolby AC-3 surround sound system configuration;

Figure 5 shows a graph of  $120^\circ$  near-ear and far-ear transfer functions;

Figure 6 shows a graph of  $60^\circ$  near-ear and far-ear transfer functions;

Figure 7 shows a graph of a  $120^\circ$  near-ear transfer function, and the  $120^\circ$  near-ear transfer function convolved with itself, according to the invention;

Figure 8 shows a graph of a  $120^\circ$  near-ear transfer function convolved with itself and a high frequency limited version of the same, according to the invention;

Figure 9 shows a graph of near-ear transfer func-

tions for positions directly above the listener and directly below the listener; and

Figure 10 shows a graph of modified near-ear transfer functions for positions directly above the listener and directly below the listener, according to the invention.

**[0023]** In a first embodiment, a filter implementing an HRTF (12), shown in Figure 3, is modified to provide improved positioning of a virtual sound source. In particular, an HRTF (12) placing a virtual sound source at an azimuth of  $+120^\circ$  and elevation of  $0^\circ$  is described. Similarly, an HRTF of azimuth angle  $60^\circ$  and elevation  $0^\circ$  will be referred to as a  $60^\circ$  HRTF. The method described may also be applied to the  $-120^\circ$ , or indeed any, HRTF.

**[0024]** Figure 5 shows the near-ear amplitude response (16a) of a  $120^\circ$  HRTF, and the far-ear amplitude response (16b) of the same function. Here, near-ear corresponds to the ear of a listener which is nearest to the virtual sound source, and far-ear is the ear furthest away from the virtual sound source. At positions where the sound source is located at identical distances from the left and right ears, the near-ear (16a) and far-ear responses (16b) are identical. The HRTF (12) therefore comprises a near-ear transfer function (16a), a far-ear transfer function (16b), and an inter-aural time delay.

**[0025]** Figure 6 shows the near-ear amplitude response (18a), and the far-ear amplitude response (18b), of a  $60^\circ$  HRTF. It can be seen that the general form of the far-ear data (16a and 18b) for both plots is similar. However, the near-ear data (16a) of Figure 5 exhibits some differences from the near-ear data (18a) of Figure 6. It should be noted that, in this example, differences in the far-ear responses (16b, 18b) are not as obvious to the brain as differences in the near-ear responses (16a, 18a). This is because the far-ear response (16b, 18b) is generally associated with less energy than the near-ear response (16a, 18a).

**[0026]** By inspection of the graphs of Figures 5 and 6, it can be seen that the prime difference between the  $120^\circ$  HRTF and the  $60^\circ$  HRTF appears to be the near-ear amplitude responses (16a, 18a). However, this difference is not large enough for the brain to be able to distinguish the  $120^\circ$  near-ear response (16a) from the  $60^\circ$  near-ear response (18a) in the presence of real reflections, and the absence of virtual reflections. The invention overcomes this deficiency by exaggerating the spectral features of the near-ear amplitude response (16a, 18a) to provide more spectral information to the listener's brain.

**[0027]** However, the best means of providing more spectral information is not immediately apparent. One may, for example, select a particular spectral feature of the HRTF data (a peak, or a trough, say), and increase its magnitude. Unfortunately, there is no way of knowing whether any particular spectral feature (or combination thereof) is important or not to the brain for the purpose

of identifying the location of a sound. Also, there is the difficulty of merging such an exaggerated feature with the remainder of the spectral response. Finally, it would not be possible to automate such a process for application to an entire library of HRTFs (12), as such a library may contain more than a thousand HRTF pairs.

**[0028]** Accordingly, the first embodiment of the present invention provides a method of creating more pronounced spectral data by increasing the magnitude of the amplitude of the near-ear function (16a, 18a) over a range of frequencies. The amount of the increase at a given frequency is a function of the amplitude of the near-ear function (16a, 18a) at the given frequency. In this particular example, for the  $120^\circ$  HRTF, the near-ear function (16a) is convolved with itself. This results in an exaggerated near-ear function (26a), as shown in Figure 7, with an increase in the magnitude of peaks and troughs, at all frequencies. In particular, it can be seen from Figure 7 that the magnitude of the trough at 4 kHz in the unmodified function has been increased. A filter may then be designed to implement an HRTF having an exaggerated near-ear function (26a). Hereinafter, a near-ear function and a far-ear function which have undergone any one of a number of processing steps according to the method described herein, are known as exaggerated near-ear and far-ear functions, respectively.

**[0029]** It is required that the magnitudes of the near-ear and far-ear amplitudes at low frequencies are similar. Therefore, it is necessary to set the overall gain factor of the modified function so as to align its low frequency response to match that of the corresponding unmodified function. Figure 7 shows the near-ear transfer function (16a) of the  $120^\circ$  HRTF (12a), convolved with itself (26a), and its overall gain adjusted for low frequency alignment of the modified and unmodified functions.

**[0030]** When an audio signal is processed by a modified filter which implements the exaggerated  $120^\circ$  HRTF, the virtual sound source appears to be located at  $+120^\circ$ , and not at  $+60^\circ$  as can occur with the unmodified filter which implements the original  $120^\circ$  HRTF.

**[0031]** In order to vary the subtlety of the 3D effects, the size of the increase in magnitude of the amplitude of the near-ear function may be varied. For example, if the near-ear transfer function is convolved with itself, the amplitude values of the transfer function are squared at a given frequency. If, however, the amplitudes of the transfer function are raised to the power 3, the resulting modified function will have more exaggerated features, and the 3D effects will be enhanced further. This may be appropriate for use in computer games, for example. Alternatively, the amplitude values of the transfer function may be raised to the power 1.5. This results in more subtle effects, and may be used advantageously, for example, for classical music recordings.

**[0032]** The high-frequency components of the exaggerated near-ear function can be limited, typically by appropriate design of the filters used for the signal

processing. In this example, frequencies of more than 10 kHz are limited. This is shown in Figure 8, plot B. However, the point at which the high frequencies are limited may vary from 10 kHz. For example, it may be desirable to reduce high frequency components above 6 kHz, or above 20 kHz.

**[0033]** Limitation, or attenuation, of high frequencies may be carried out for the following reasons: For 3D sound conveyed via loudspeakers remote from the listener's ears, high-frequency information cannot, in practice, be crosstalk cancelled effectively. We can therefore attenuate the high frequencies with little effect on the apparent placement of the virtual sounds. This is discussed in our co-pending UK Patent Application No. GB 9805534.6.

**[0034]** When listening to sounds via loudspeakers, high frequencies are attenuated more than low frequencies along the pathway from the loudspeakers to the listener's head. However, when listening to sounds via headphones (where crosstalk cancellation is not required), high frequencies are not attenuated along the pathway from the headphones to the ears of a listener, due to the proximity of the headphones to the ears. Thus more high frequency sound is presented to the ears than would be so via loudspeakers. This may result in the virtual sound image appearing to be close to the listener's head. For this reason, a reduction in high frequencies is desirable for headphone reproduction to enable the virtual sound image to appear "out-of-the-head".

**[0035]** Modified filters which implement the exaggerated HRTFs may be used in many applications. Examples of these applications will now be described.

**[0036]** In the AC-3 surround sound listening format, there is provision for 6 loudspeakers: front left, centre, front right, surround left (rear), surround right (rear), and a non-directional sub-woofer. During the sound mixing process (wherein the sound is encoded for the AC-3 format), a sound engineer can "pan" sounds from one position to another by varying the relative loudness of the sound being fed to the various loudspeakers. For example, a sound source may be panned from the front right speaker to the rear left speaker, and the sound would appear to the listener to move from the front right speaker to the rear left speaker through him or herself. However, it may be required for some applications that a sound is panned over the head of the listener, or underneath the listener. For example, it might be required to move the sound of a helicopter from the front right speaker over the head of the listener, and then to the front left speaker. With present panning systems this would not be possible as the apparent positions of virtual sounds are restricted to the horizontal plane. By the use of an exaggerated "height" filter, it is possible to introduce height elements into the system.

**[0037]** For example, an exaggerated "overhead" (that is, where elevation=90°) HRTF may be produced via the method described in the first embodiment of the invention, and used as a "height" filter for surround sound

mastering (or encoding) applications. This would enable panning from the front of a listener, to behind the listener, passing over the top of the listener's head. An exaggerated "below" (for example, elevation=-90°) HRTF may also be produced to make a "depression" filter, and could be used to enable panning from a position in front of a listener, passing underneath the listener, to a position behind the listener. This approach enables the conventional sound format to extend into the third dimension without any changes in the user's hardware, and without any change in format, bandwidth and the like.

**[0038]** The method of the invention may also be used in conjunction with vertical balance adjustment. Vertical balance adjustment is described in published International Patent Application, No. WO-A1-9517799.

**[0039]** A set of digital filters may be produced which implement an entire exaggerated HRTF library. This may be appropriate for applications such as PC games, where 3D effects with great spectral impact are more important than optimal tonal quality.

**[0040]** A sound recording or a transmission such as, for example, via wire based or wireless telegraphy, may be made by using modified filters which implement the exaggerated HRTFs.

**[0041]** Variation may be made to the aforementioned embodiments without departing from the scope of the invention. For example, the method of the invention may be applied to the far-ear transfer function (16b,18b), or to both the near-ear transfer function (16a,18a) and the far-ear transfer function (16b,18b).

## Claims

1. A method of modifying the characteristics of a filter for implementing a head-related transfer function (HRTF), the HRTF (12) including a near-ear transfer function (16a) and a far-ear transfer function (16b), the method comprising increasing the magnitude of the amplitude of the near-ear transfer function and/or far-ear transfer function over a range of frequencies to give an exaggerated near-ear transfer function (26a) and/or an exaggerated far-ear transfer function, the amount of the increase at a given frequency being a function of the amplitude of the corresponding transfer function or functions at the given frequency, thereby forming a filter which implements an HRTF having an exaggerated near-ear transfer function (26a) and/or an exaggerated far-ear transfer function.
2. A method according to claim 1 wherein the magnitude of the amplitude of the near-ear transfer function is increased by convolving the near-ear transfer function (16a) with itself.
3. A method according to claims 1 or 2 wherein the magnitude of the amplitude of the far-ear transfer

function is increased by convolving the far-ear transfer function (16b) with itself.

filter as claimed in any of claims 7 to 16.

4. A method according to any preceding claim wherein the amplitude of the exaggerated near-ear transfer function (26a) and/or the amplitude of the exaggerated far-ear transfer function is limited over a range of frequencies above a threshold value. 5
5. A method according to claim 4 wherein the threshold value is 6 kHz. 10
6. A method according to any preceding claim wherein the amplitude of the exaggerated near-ear transfer function (26a) and/or the amplitude of the exaggerated far-ear transfer function is adjusted so that the amplitude of the exaggerated near-ear transfer function (26a) and the amplitude of the exaggerated far-ear transfer function tend to the same value at frequencies below 100 Hz. 15  
20
7. A filter modified using the method as claimed in any of claims 1 to 6.
8. A filter according to claim 7 for implementing an HRTF, wherein the HRTF has an amplitude response characteristic curve substantially as shown in plot B of Figure 8. 25
9. A filter according to claim 7 including transaural crosstalk cancellation means. 30
10. A filter according to claim 7 wherein the filter places a virtual sound source at positions behind the preferred position of a listener in use. 35
11. A filter according to claim 7 wherein the filter places a virtual sound source at an azimuth of  $\pm 120^\circ$  and an elevation of  $0^\circ$  relative to the preferred position of a listener in use. 40
12. A filter according to claim 7 wherein the filter places a virtual sound source at an elevation of  $\pm 90^\circ$  relative to the preferred position of a listener in use. 45
13. A filter according to claim 11 for use in a multi-channel surround sound system.
14. A filter according to claim 13 wherein a multi-channel audio signal is converted to a binaural signal. 50
15. A filter according to claim 12 for use in a multi-channel encoding system.
16. A filter according to claims 7 to 15 wherein the filter is a finite impulse response filter. 55
17. A sound recording or transmission made using the

18. A signal processed using the filter claimed in any of claims 7 to 16.

Fig.1.

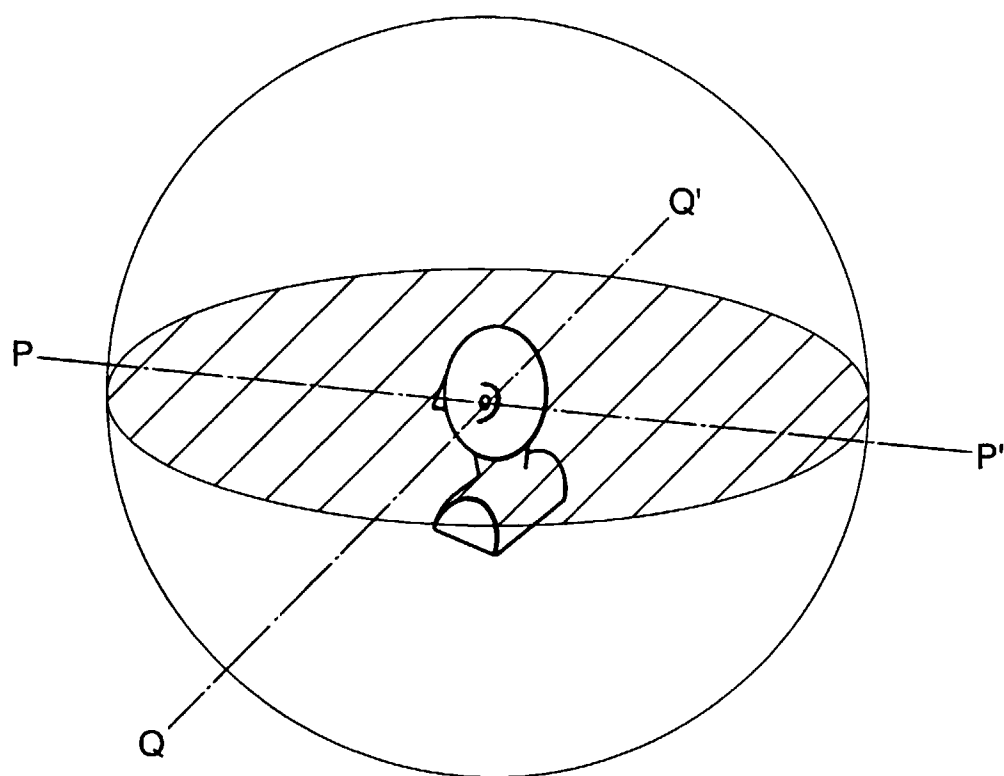


Fig.2.

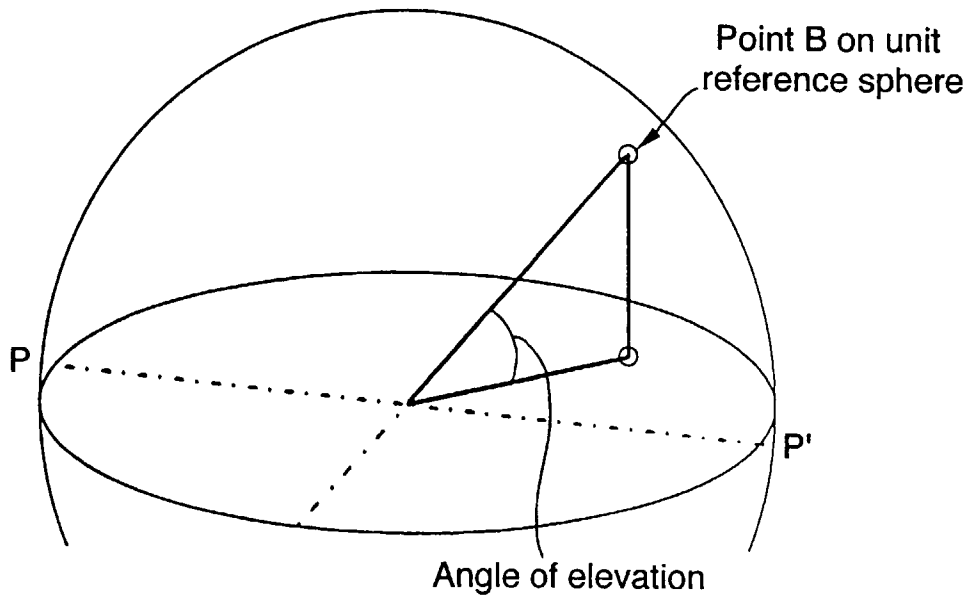
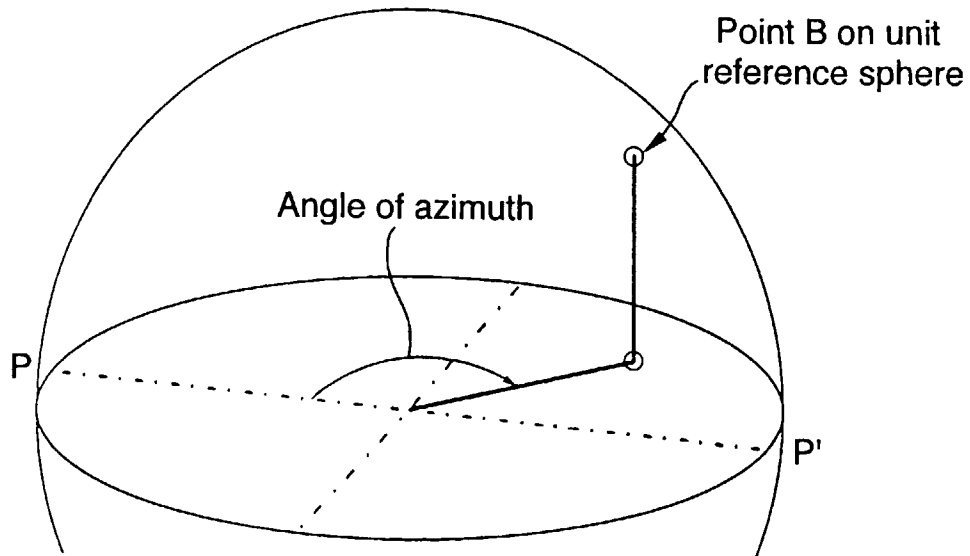




Fig.3.

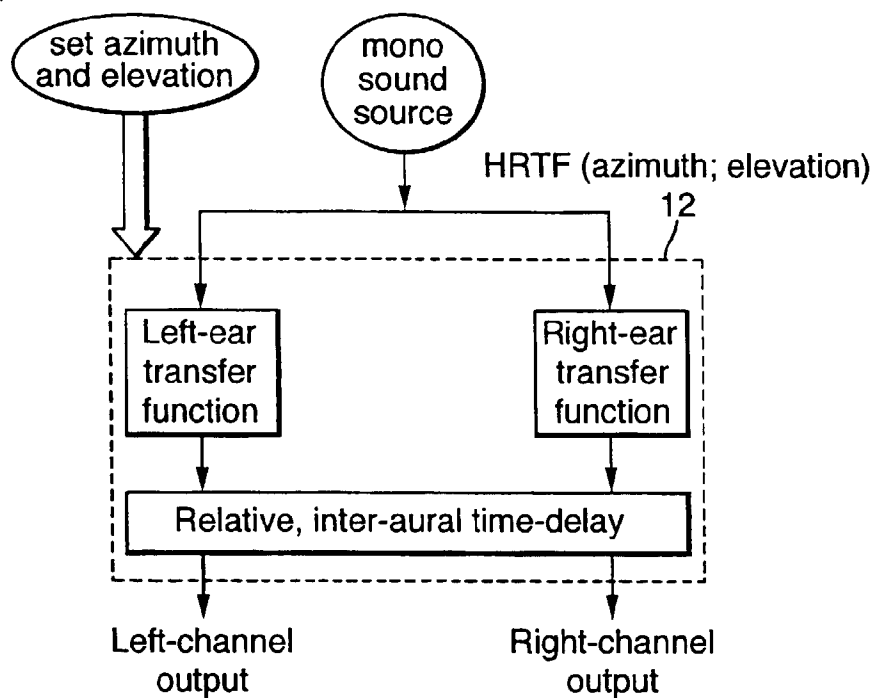


Fig.4.

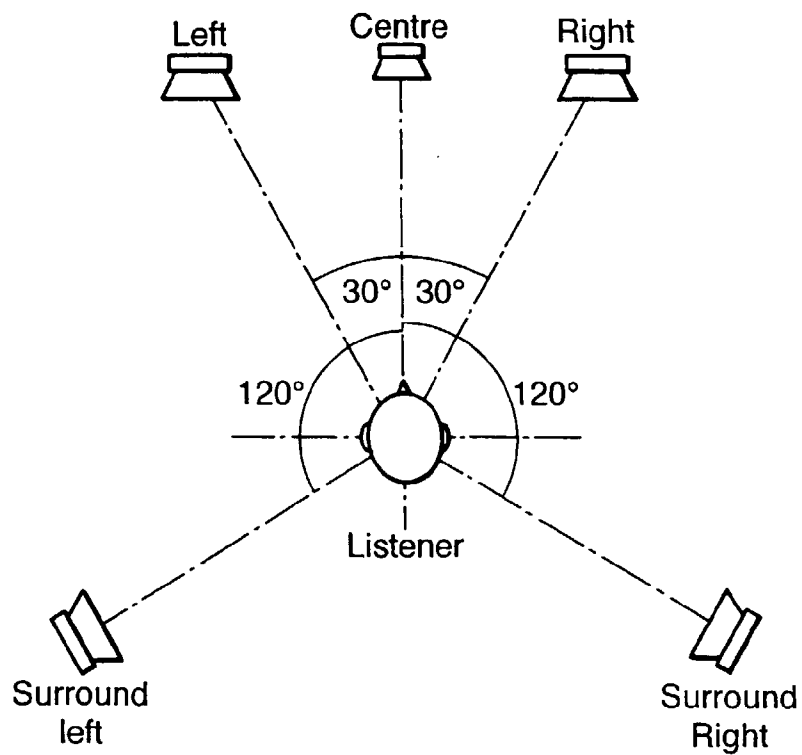
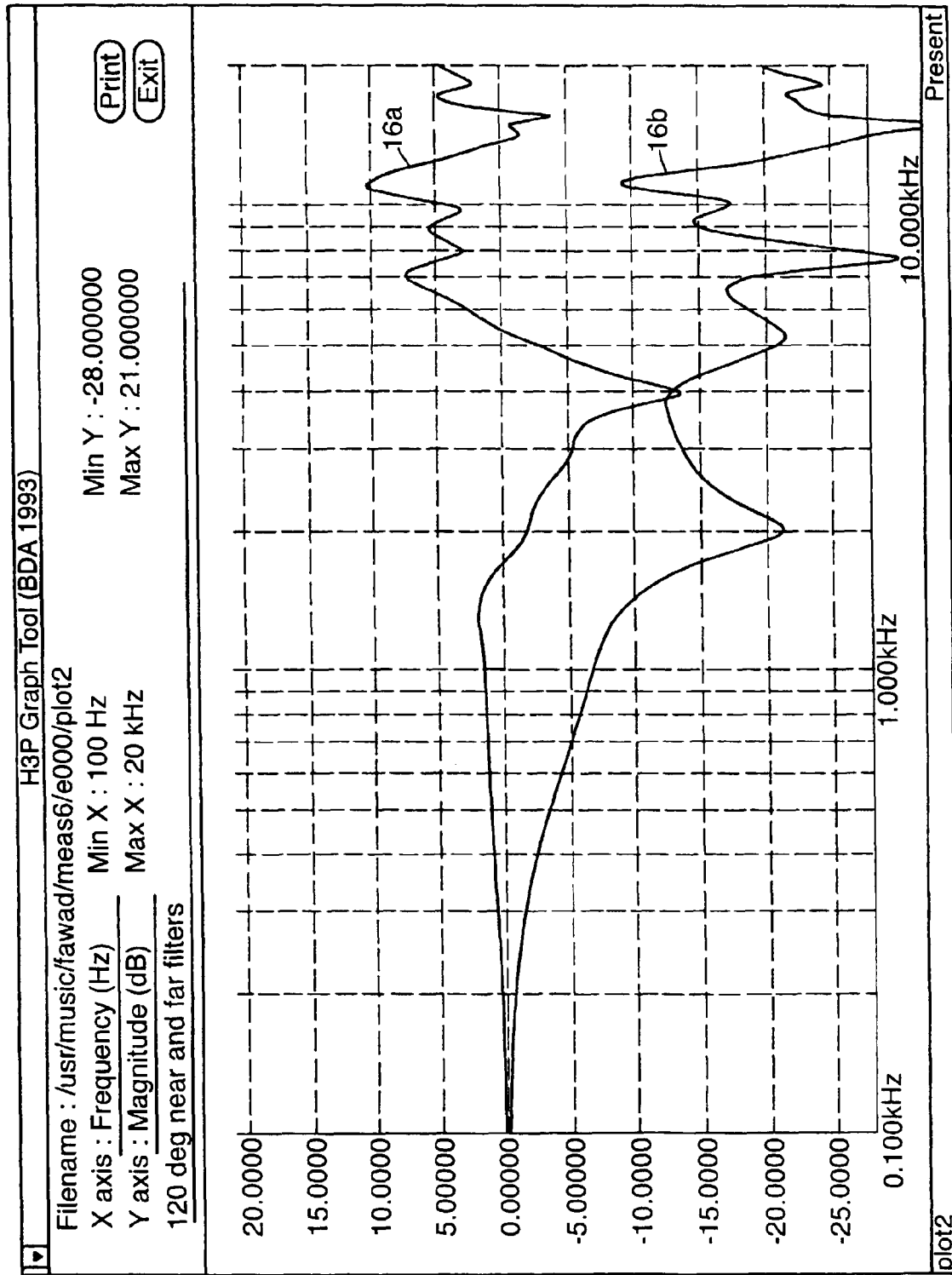


Fig. 5.



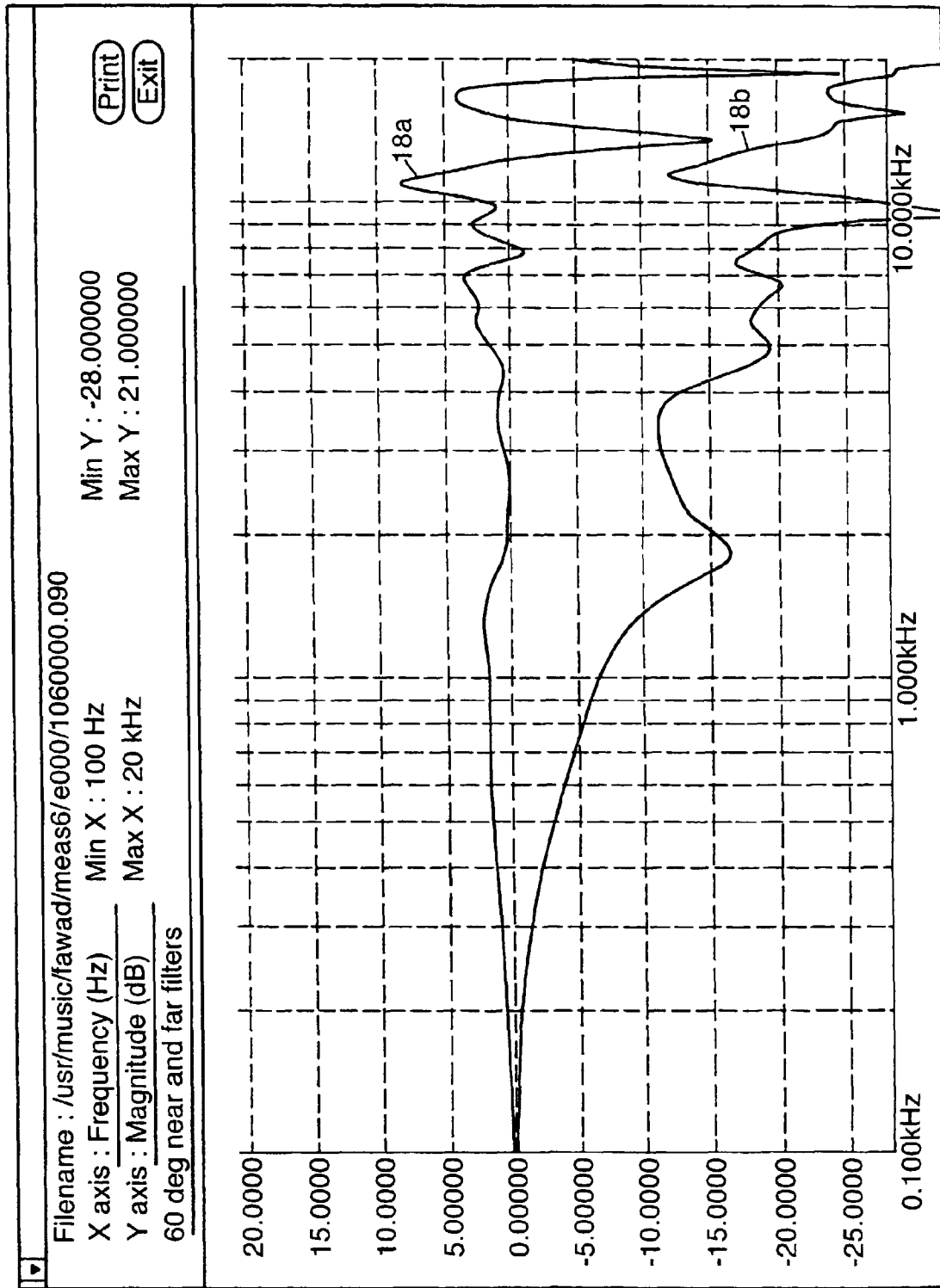


Fig. 6.

Fig. 7.

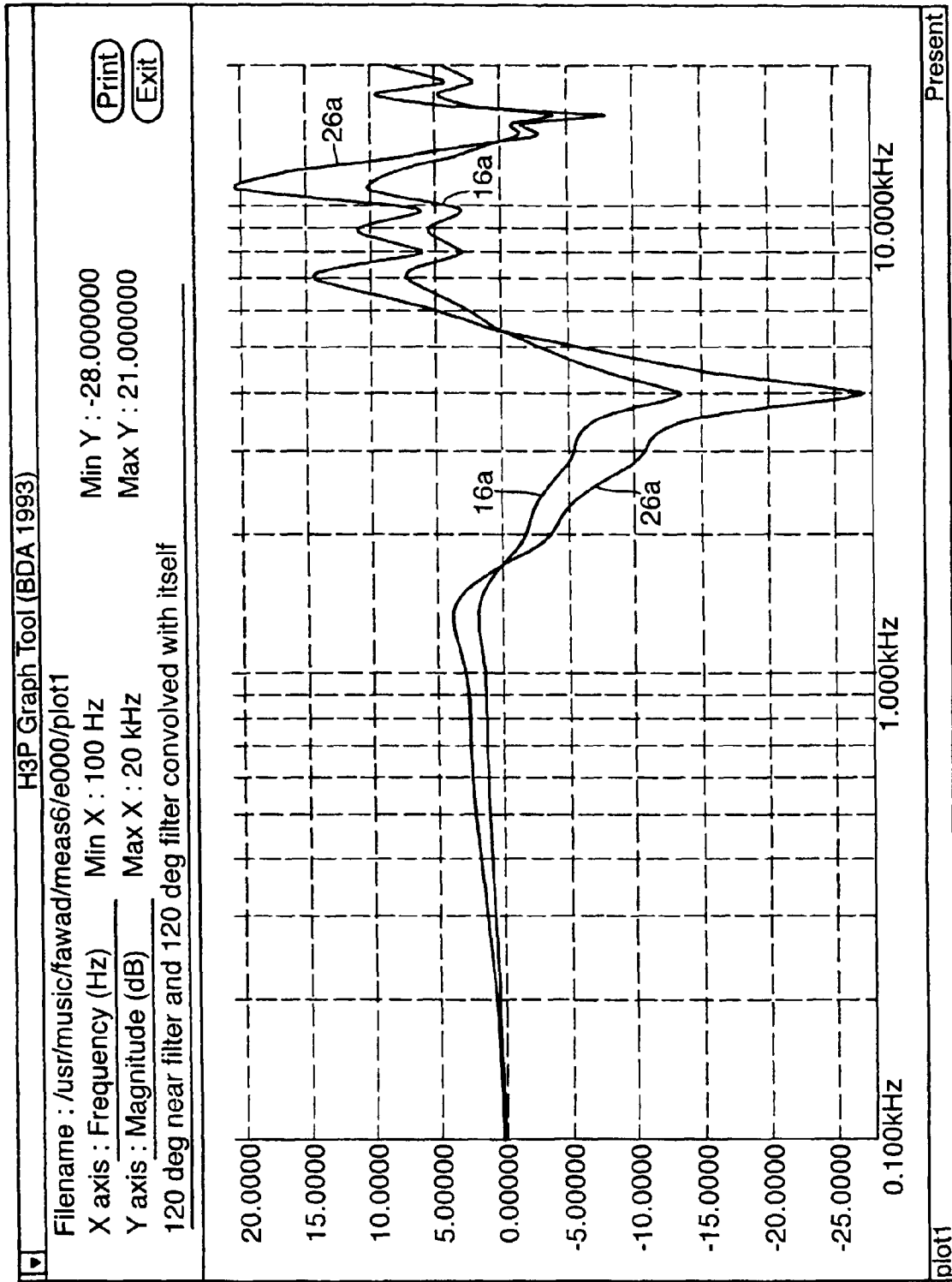
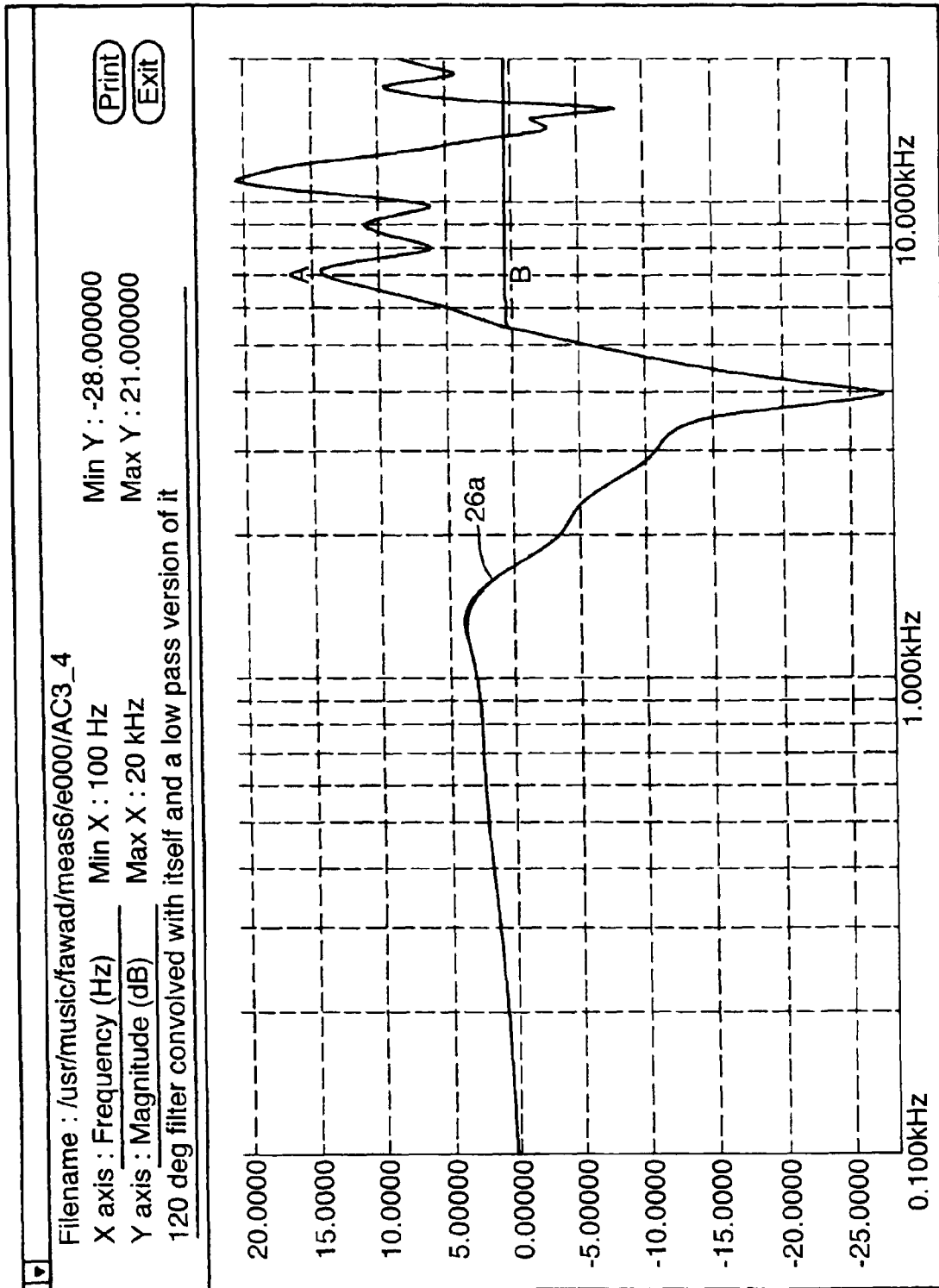


Fig. 8.



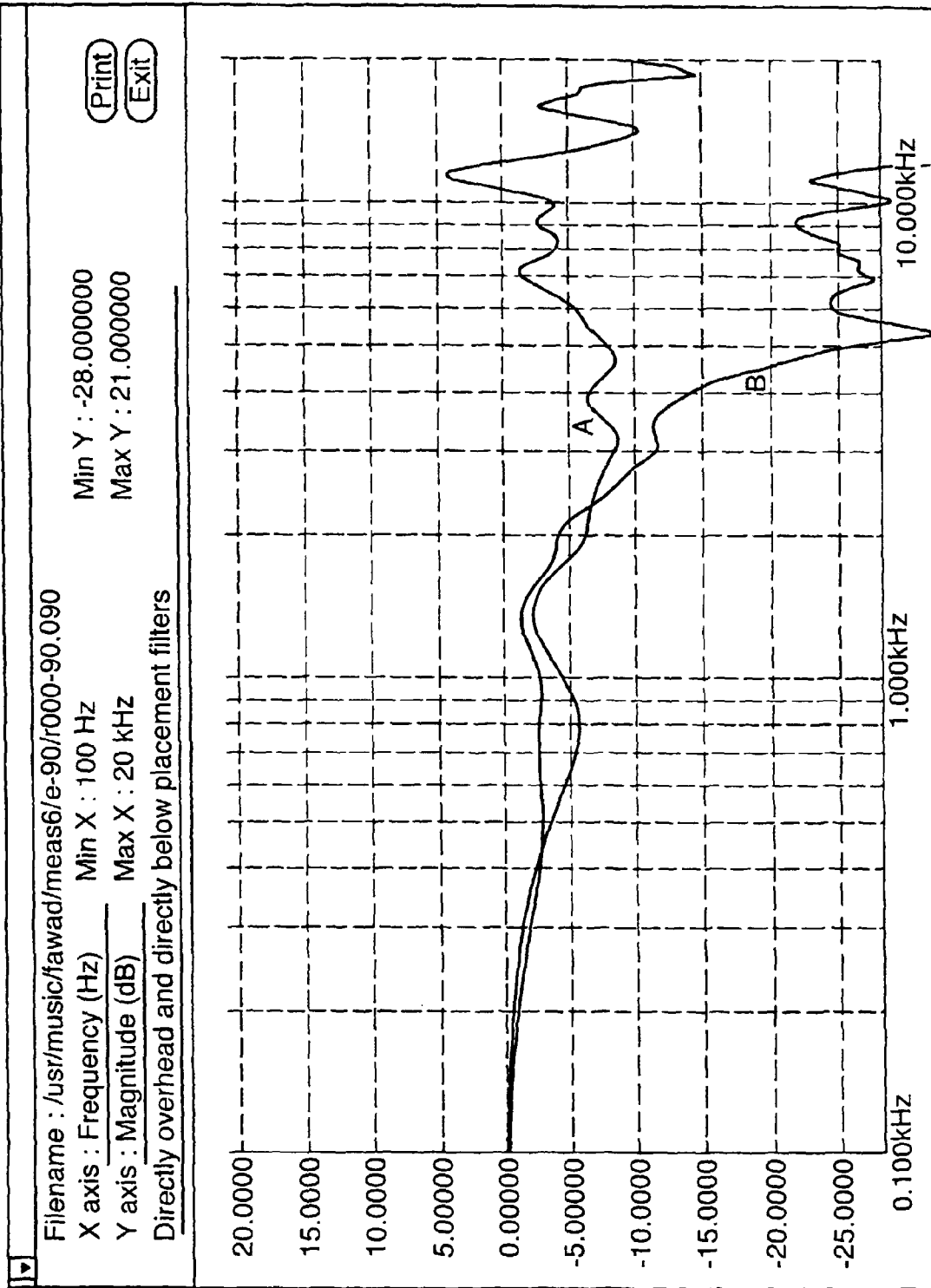


Fig. 10.

