Fundamentals of Binaural Technology

Møller, Henrik

Published in:
Applied Acoustics

Publication date:
1992

Document Version
Accepted author manuscript, peer reviewed version

Applied Acoustics 36 (1992) 171–218

Fundamentals of Binaural Technology

Henrik Møller

Acoustics Laboratory, Aalborg University, Aalborg, Denmark

(Received 3 December 1991; revised version received 24 February 1992; accepted 3 March 1992)

ABSTRACT

This article reviews the fundamental ideas of the binaural recording technique. A model is given that describes the sound transmission from a source in a free field, through the external ear to the eardrum. It is shown that sound pressures recorded at any point in the ear canals—possibly even a few millimeters outside and even with a blocked ear canal—can be used for binaural recordings, since they include the full spatial information given to the ear.

The sound transmission from a headphone is also described. It is shown how the correct total transmission in a binaural system can be guaranteed by means of an electronic equalizing filter between the recording head and the headphone. The advantage of an open headphone is stated. It is shown that a certain degree of loudspeaker compatibility can be achieved, if the equalizer is divided into a recording side and a playback side. A method for true reproduction of binaural signals through loudspeakers is also described.

A number of topical and prospected applications of binaural technology are mentioned. Some of these utilize computer synthesis of binaural signals, a technique which is also described.

1 INTRODUCTION

The idea behind the binaural recording technique is as follows: The input to the hearing consists of two signals: sound pressures at each of the eardrums. If these are recorded in the ears of a listener and reproduced exactly as they were, then the complete auditive experience is assumed to be reproduced,

Applied Acoustics 0003-682X/92/$05.00

including timbre and spatial aspects. The term *binaural* recording refers to the fact that the *two* inputs to the hearing are reproduced correctly.

The recording may be made with small microphones placed in the ear canals of a human listener, but normally a copy of a human head is used. The copy has the shape of an average human head, including nose, orbits, pinnae and ear canals, and sometimes the head is even attached to a torso copy. Also the acoustical impedance of the ear drum is sometimes simulated. By accurately copying a human head it is ensured that sound waves reaching the head, undergo the same transmission on their way to the ear canals, as if they were reaching a real listener. A copy of the human head is called an *artificial head*, a *dummy head*, a *head simulator* or, in German, a *Kunstkopf*. These terms have inspired alternative terms for the binaural recording technique: *artificial head recording technique*, *dummy head technique* and *Kunstkopftechnik*. Also the expression *head-related technique* is used.

The playback is normally done with headphones, since this method ensures that sound picked up in one ear is only reproduced in that ear. Reproduction through loudspeakers would introduce an unwanted crosstalk, since sound from each of the loudspeakers would be heard with both ears.

The basic idea of the binaural recording technique is not new. Descriptions of the idea, of its applications and of details in the sound transmission from the recording head to the listener's eardrums have been found in the literature for more than 60 years. Examples are Refs 1–37.

Several artificial heads are commercially available from manufacturers such as Neumann (Berlin, Germany), Head Acoustics GmbH (Aachen, Germany), Brüel & Kjær (Nœrum, Denmark) and Knowles Electronics (Itasca, IL). Nevertheless, the binaural recording technique has not yet got the widespread use that might be expected. Among the reasons are the fact that binaural signals are intended for headphone reproduction, and the recording and broadcasting industries have not yet been prepared to make special recordings for this purpose, except for experimental issues. Another obstacle to use in broadcasting is lack of mono compatibility.

Investigations have shown, though, that proper equalization of the microphone output may guarantee preservation of timbre, even when signals are reproduced through an ordinary loudspeaker stereo set-up. There is some disagreement about the exact way of doing this, the main concepts being *free-field equalization* and *diffuse-field equalization*.[18,23,25,27,28,31,33,38–45,83] Of course, the spatial reproduction of the binaural technique is not obtained, but it is claimed that the quality is comparable to that of traditional intensity stereo recordings with respect to timbre and spatial characteristics.[42–43]

The most important impediment for success of the binaural technique, however, is presumably problems with frontal localization. Sound sources that were originally in the frontal hemisphere are often perceived as being behind the listener, or they appear closer to the listener than they were during the recording. Sometimes localization in the head occurs. One explanation of these observations is that human heads and pinnae have individual differences, and only recording with a listener's own pinna may guarantee proper frontal localization.[46–48]

Another explanation is the following: humans can use small head movements to distinguish between front and back. A right turn of the head will cause sound from frontal sources to arrive earlier to the left ear and later to the right ear. The opposite happens for sound sources behind. A binaural recording does not react on head movements, and this may explain the problems with front/back confusions.[49] However, experiments have shown that for broadband sound sources, humans are able to distinguish between front and back, even when their head is kept still.[50–51] Consequently, lack of proper response to head movements cannot solely explain the problems.

Even encumbered with the above-mentioned problems, the binaural technique is superior to other recording techniques. Properly used it gives a very realistic impression of being present during the recording, and people are often surprised with the authenticity.

A more widespread acceptance of headphones as means for reproduction—somewhat promoted by the concept of the Walkman and other kinds of personal stereo—is now contributing to a revival of the technique. Possibilities also exist for total restoration of the binaural signals, when reproduced through loudspeakers.[52–63]

Conducive to a revival of the technique are also new technological possibilities for computer synthesis of binaural signals. This research is often connected to projects on artificial environments.[45,61,64–68]

Hopefully, the new interest will inspire research that also helps to overcome the problems with frontal localization.

The main purpose of the present article is to give an overview of the methods that can be used for recording and playback of binaural signals. A new model for sound transmission to the eardrum is introduced, together with the associated notation. The model has already proven useful during work in our laboratory,[35–37,69] and it is the author's hope that others may also gain advantage from it.

In the above description of the binaural technique, it has been assumed that the recording is made at the position of the eardrum for the artificial head. Sound pressures recorded at other points in the external ear can also be regarded as inputs to the hearing, provided that there is an unambiguous transmission from this pressure to the pressure at the eardrum. In practice this means that the sound transmission from that other point to the eardrum must be independent of the direction and the distance to the sound source. Any set of left and right channel signals, recorded at points that fulfil this requirement, are called *binaural* signals. The sound transmission from a source in a free field, through the external ear to the eardrum, is described in a model given in Section 2.

The model involves a somewhat untraditional application of transmission-line theory, that makes the calculations simpler. An example illustrating this is given in Appendix A.

In the binaural technique, the importance of a properly equalized headphone is often overlooked. Much effort is spent on design of artificial recording heads, and then any headphone is used for listening. This is remarkable, since it is quite evident that the headphone, as well as the head used for the recording, contributes to the total sound transmission. The correct reproduction of the recorded sound pressure can only be guaranteed when certain characteristics of the headphone are known. Therefore, a model describing the sound transmission from this device to the eardrum is also developed. This model is given in Section 3.

Section 4 gives a description of the binaural technique based upon the models from Sections 2 and 3. Three possible recordings points are selected: (a) at the eardrum, (b) at the entrance to the ear canal and (c) at the entrance to the ear canal, but with the ear canal physically blocked. For each of these points it is shown how the correct reproduction can be obtained by the introduction of an electrical equalizing circuit. Section 4 concludes mentioning recording at other points in the ear canal, using miniature microphones. Such microphones are often used, since they are produced small enough to be inserted in the ear canal.

In the description it is assumed that the artificial head used during recording, and for calibration of the headphone, is a perfect copy of the listener present during the reproduction. Alternatively, the listener himself may be used instead of the artificial head.

Techniques that make binaural signals suitable for loudspeaker reproduction are covered in Sections 5 and 6. Section 5 covers equalizing methods that give the same tonal balance in loudspeakers, as if the recording were made with traditional microphones. However, the precise reproduction of the eardrum signals is not preserved, and the spatial reproduction is not superior to that of traditional stereo. Section 6 covers a more sophisticated method for loudspeaker reproduction that gives complete restoration of the two eardrum signals. Thus the outstanding spatial reproduction of the binaural technique is preserved. The method works only in anechoic surroundings, and the position of the head must be rather precise.

Until this point, the binaural signals are assumed to originate in a recording of an acoustical event that has taken place in real life. It is also possible to synthesize the signals on a computer. The possibility of computational creation of binaural signals is described in Section 7. Section 8 is a brief mention of some possible applications of binaural technology.

2 LISTENING IN A FREE FIELD

The listener in Fig. 1 is in a concert hall. He prefers the live concert to his stereo set at home—not only because of the atmosphere associated with live concerts, but also because the live concert gives him a true three-dimensional auditive experience. He can hear the direct sound from each of the instruments and the reflections coming from the sides and above, and thus create an 'image' of the orchestra, the concert hall and its acoustics. If it is a well-designed hall, the reflections contribute positively to the musical experience, and they help in localizing the instruments.

A precondition for the creation of an 'auditive image' of the orchestra and the room, is the ability of the human hearing to determine direction and distance to single sound sources. In the case of the concert hall, each instrument is a sound source that sends a direct sound to the listener. The instruments also send sound waves in other directions, waves that give rise to reflections. The reflections can be regarded as sound coming from additional and imaginary

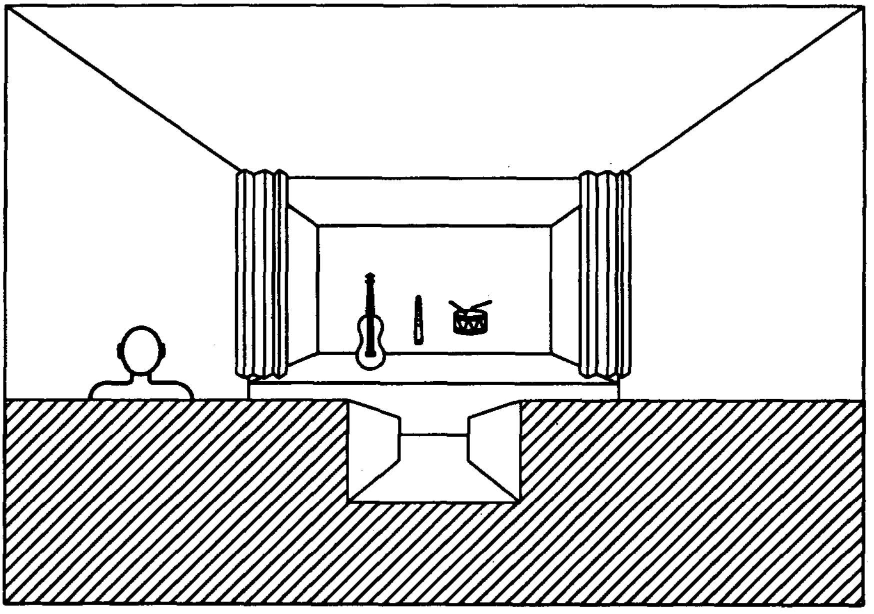

Fig. 1. A listener in the concert hall.

sound sources of which the direction and distance can be determined, for example from a mirror image model. If these sound sources were playing one at a time, the listener would—at least to some extent—be able to determine direction and distance to the 'source'. When the direct sound and the reflections are present all together, the listener gets the total spatial experience.

It is traditionally said that the hearing uses a number of cues in the determination of direction and distance to a sound source. Among the cues are

(1) coloration
(2) interaural time differences
(3) interaural phase differences
(4) interaural level differences

These cues are claimed to be responsible for the directional hearing in each of their 'domain'. For instance, in the horizontal plane low frequencies are said to be assessed by interaural phase differences, medium frequencies by interaural time differences and high frequencies by interaural level differences. Coloration is claimed to be responsible where no interaural differences exist, that is in the median plane. A thorough discussion of cues to directional hearing is given by Blauert.[13,32]

This way of splitting up the cues for the directional hearing is only justified on the basis of experiments with presentation of sophisticated artificial signals. A natural sound coming from a given direction will—on its way to the two ears—be exposed to two unique filterings, of which the spectral and time attributes cannot be separated. The fundamental idea in the present description will therefore be more general, namely:

A sound wave coming from a given direction and distance, results in two sound pressures, one at each eardrum. The transmissions are described in terms of two transfer functions that include any linear distortion, such as coloration and interaural time and spectral differences. The task of a binaural recording and playback system is to present the correct inputs to the hearing, that is to reproduce the eardrum signals correctly. In this connection, it is not important how the hearing extracts information from the eardrum signals about distance and direction.

Knowledge about the way in which the hearing extracts the distance and directional information may prove useful at the time, when the needed accuracy of the transmission is to be assessed. Or expressed in another way: if the eardrum signals cannot be reproduced 100% correctly, this knowledge may tell which aspects of the eardrum signals are most important to reproduce correctly.

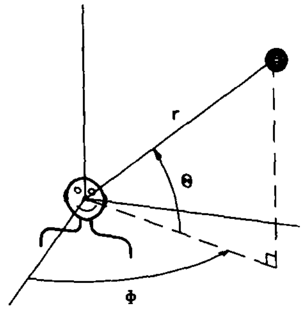

Fig. 2. Sound source and listener in a free field. Conventions for the variables indicating distance and direction are shown.

2.1 Transmission in a free field

A terminology will be introduced that describes the situation, when a sound wave hits a human.

In Fig. 2, a sound source radiates a sound wave in a free field. Somewhere the wave hits a listener. The distance from the listener to the source is denoted by r, and the angle of incidence is characterized by the azimuth ϕ and the elevation θ. ($\phi = 0°$, $\theta = 0°$) is the direction right in front of the listener. Positive values of ϕ are defined to characterize directions to the left of the listener, while positive values of θ indicate directions above the horizontal plane. The whole sphere is covered for $-\pi < \phi \leq \pi$ and $-\pi/2 \leq \theta \leq \pi/2$.

The notation which will be used is illustrated for one side in Fig. 3. Part (a) shows an anatomical sketch and part (b) is an analogue model. The sound pressure created at the input to the ear canal is denoted by p_3. In a certain frequency range, the canal acts as an acoustical transmission line, and the resulting pressure at the eardrum is denoted by p_4.

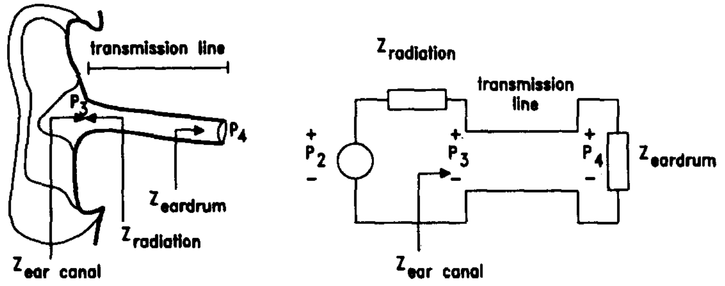

Fig. 3. Sound transmission through the external ear: (a) sketch of the anatomy and (b) an analogue model.

Throughout the present article, time-varying signals are denoted by small letters, and for simplicity the variation with time t is not explicitly indicated. Thus p_3 is short for $p_3(t)$. The same signals can be given in the frequency domain, where capital letters are used. The frequency domain representation of p_3 is P_3, which is short for $P_3(f)$, where f is the frequency.

The transmission line analogy is valid for frequencies where the wavelength λ is much larger than the diameter of the ear canal. If an ear-canal diameter of 8 mm is inserted, then

$$\lambda \gg 8 \times 10^{-3}\,\mathrm{m} \tag{1}$$

$$f \ll \frac{340\,\mathrm{m/s}}{8 \times 10^{-3}\,\mathrm{m}} = 42{\cdot}5\,\mathrm{kHz} \tag{2}$$

Normally, an upper limit is used where the diameter is a quarter of a wavelength. This means that the model is supposed to be adequate up to about 10 kHz.

It may be assumed that the diameter of the ear canal is the same at all points from the entrance to the eardrum. In that case, the characteristics of the transmission line are constant. However, this assumption is not necessary for the following description, since a transmission line with properties varying with position is also acceptable.

Now, if Thevenin's theorem is used at the entrance to the ear canal, the source can be split into two parts: the 'open circuit' pressure p_2, and the radiation impedance $Z_{\mathrm{radiation}}$. The open-circuit pressure is the pressure that would exist at the position of p_3, when no volume velocity ('current') runs through $Z_{\mathrm{radiation}}$. Zero volume velocity can only be guaranteed if the ear canal is physically blocked.

Therefore, p_2 does *not* exist physically in the listening situation. For measurement purposes, it can be found as the sound pressure at the entrance to the ear canal, when the ear canal is blocked, for instance with an ear plug. Naturally, p_3 and p_4 do not exist at this time then.

It is possible to find p_3 from a pressure division of p_2 between the radiation impedance and the impedance of the ear canal, $Z_{\mathrm{ear\,canal}}$. Thus

$$[P_3/P_2] = \frac{Z_{\mathrm{ear\,canal}}}{Z_{\mathrm{ear\,canal}} + Z_{\mathrm{radiation}}} \tag{3}$$

2.2 Definitions of transfer functions

The model is going to lead to a description in terms of transmission characteristics from a point source to the various pressures indicated. If the sound source is a loudspeaker, the voltage at the loudspeaker terminals,

$e_{\text{loudspeaker}}$, can be used as the 'origin'. This is practical in a measurement situation, but for characterization of the sound transmission from a given direction and distance it is unsuitable, since it also includes the radiation characteristics of the loudspeaker.

The reference traditionally chosen—and also in this article—is the sound pressure p_1 at a position right in the middle of the head, but with the listener absent. This choice causes some complications in the instrumentation, since p_1 is not present at the same time as the other sound pressures, but it makes the results more general and easier to compare. How this problem is overcome in measurements is described in Section 2.5.

If we turn to the frequency domain, transfer functions can be introduced. (Of course, we might stay in the time domain and use impulse responses, but this would involve convolutions rather than multiplications, and the notation would be a lot more complicated.) The following transfer functions are important:

$$[P_4/P_1] = \frac{\text{sound pressure at the eardrum}}{\text{sound pressure in the middle of the head with the listener absent}} \tag{4}$$

$$[P_3/P_1] = \frac{\text{sound pressure at the entrance to the open ear canal}}{\text{sound pressure in the middle of the head with the listener absent}} \tag{5}$$

$$[P_2/P_1] = \frac{\text{sound pressure at the entrance to the blocked ear canal}}{\text{sound pressure in the middle of the head with the listener absent}} \tag{6}$$

These transfer functions are dependent upon angle of incidence (ϕ and θ) and distance to the sound source (r). If r is reasonably large (probably two meters or more), then the incident wave turns into a nearly plane wave, and the above transfer functions become almost independent of r. In the following sections, r may be omitted, if it is assumed that $r \geq 2\,\text{m}$, approximately.

Definitions corresponding to eqns (4), (5) and (6) were given for the first time by Blauert,[13] and he introduced the name *free-field transfer function* (FFTF). Other names seen in the literature are *outer ear transfer function*, *external ear transfer function*, *pinna response*, *ear-canal transfer function*, *directional transfer function* (DTF), *head-related transfer function* (HRTF or HTF), *head transfer function* (HTF), *pressure transformation from the free field to the eardrum* and *free-field-to-eardrum transfer function*. The last two refer explicitly to $[P_4/P_1]$ (eqn (4)).

Unfortunately, some of the terms refer to less precise definitions than the ones given above. In any case, it is extremely important to know which of the three sound pressures p_2, p_3 or p_4 they refer to (or a sound pressure at another position).

In this article, the term *head-related transfer function* (HTF) will be used for any of the three transfer functions in eqns (4), (5) and (6). If the transfer function is given in the time domain, the term *head-related impulse response* (HIR) is used.

Two other terms are to be given: *monaural transfer function* and *interaural transfer function*. They were also introduced by Blauert,[13] and are defined in the following (where P_i refers to any of the pressures P_2, P_3 or P_4):

$$\text{monaural transfer function} = \frac{[P_i/P_1](r, \phi, \theta)}{[P_i/P_1](\phi = 0°, \theta = 0°)} \tag{7}$$

$$\begin{aligned}\text{interaural transfer function} &= \frac{[P_i/P_1]_{\text{ear opposite sound source}}}{[P_i/P_1]_{\text{ear facing sound source}}}(r,\phi,\theta)\\ &= \left[\frac{P_{i,\text{ ear opposite sound source}}}{P_{i,\text{ ear facing sound source}}}\right](r, \phi, \theta)\end{aligned} \tag{8}$$

A popular, verbal description of the monaural transfer function, is that it tells how the transmission from a given direction 'deviates' from the transmission from the front. The interaural transfer function describes the difference in transmission to the ears at the two sides of the head.

In the denominator of the monaural transfer function, r is not mentioned, since for reference, sound waves from a distant source are used.

The reference for the monaural transfer function may be chosen differently from that of eqn (7). A useful choice is to refer to the transmission that would be in a diffuse sound field. A diffuse sound field is characterized by sound arriving randomly from all directions. Therefore, the reference should be an average of the transmission from all directions.

The way the averaging is done deserves a few comments. In principle, averaging may be done in the time domain or in the frequency domain. A simple averaging in the time domain makes no sense, because of the variation in arrival time with azimuth angle. And as a diffuse sound field is characterized by sound waves with random phase, averaging in the frequency domain is only meaningful on a power basis. Then it is only possible to determine the magnitude of $[P_i/P_1]$ (average of all angles).

$$\begin{matrix}\text{monaural transfer function}\\ \text{referenced to diffuse field}\end{matrix} = \frac{[P_i/P_1](r, \phi, \theta)}{|[P_i/P_1](\text{average of all angles})|} \tag{9}$$

The monaural transfer function referenced to diffuse field (eqn (9)) has the advantage compared to the traditional monaural transfer function (eqn (7)) that it is less dependent upon the transfer function from a single direction.

Problems with possible zeros in the reference transfer function are also overcome, since they hardly exist in the average.

The description above applies to the transmission into each ear. If symmetry is assumed, then

$$[P_{i,\,\mathrm{left}}/P_1](r,\phi,\theta) = [P_{i,\,\mathrm{right}}/P_1](r,-\phi,\theta] \tag{10}$$

It is important to understand that the transfer functions defined in this section give a complete description of the sound transmission, including diffraction around the head, reflections from shoulders, reflections in the ear canal, etc. It may be a little tricky to see, for instance, that the reflections from the eardrum are included in a simple pressure ratio as in eqn (3). The explanation is, of course, that the impedances $Z_{\mathrm{ear\,canal}}$ and $Z_{\mathrm{radiation}}$ are frequency-dependent variables given in amplitude and phase for a wide frequency range. This is in contrast to traditional transmission-line calculations, where normally steady-state conditions are assumed at a specific frequency. If the pressure ratio in eqn (3) is inversely Fourier-transformed, the reflections will be identifiable in the time domain. An example of this is given in Appendix A.

At this place it is also important to note that the head-related transfer functions include all the cues to the directional hearing mentioned in the introduction to Section 2, being directional-dependent coloration and interaural differences. Be also aware that some of the head-related transfer functions are non-causal, since the sound arrives earlier to the ear than to the middle of the head, if the ear is closer to the sound source.

2.3 Splitting up the transmission

$Z_{\mathrm{radiation}}$ is the radiation impedance being looked out into from the entrance to the ear canal. This impedance is independent of source location. Then the propagation from p_2 through p_3 to p_4 must be independent of source location. This means that all of the sound pressures p_2, p_3 and p_4 include the full spatial information given to the ear. Thus, any of these sound pressures will be suitable for binaural recording, if only the reproduction is adapted, so that the 'final product'—sound pressure at the eardrum—is correct. The same applies to any sound pressure along the ear canal. How the adaptation should be made is described in Section 4.

Let the transmission from a sound source to sound pressure at the eardrum be split up in the following way:

$$[P_4/P_1] = [P_4/P_3]\cdot[P_3/P_2]\cdot[P_2/P_1] \tag{11}$$

In addition to a possible dependence upon distance and direction to the sound source, these transfer functions are assumed dependent upon the anatomy of the listener involved. A list of the involved transfer functions is given in the following, where it is indicated which variables each of them depend upon ('side' means left or right ear).

$$
\begin{aligned}
&[P_2/P_1](\text{subject}, \text{side}, r, \phi, \theta) \\
&[P_3/P_2](\text{subject}, \text{side}) \qquad (12) \\
&[P_4/P_3](\text{subject}, \text{side})
\end{aligned}
$$

The transformation from a three-dimensional sound field to two binaural signals is described by the transfer functions $[P_2/P_1]$ (one for each side). If this transformation is very different for different humans, then each listener must have his own recordings. This will restrict the applications of the binaural technique very much.

Individual variations in the two remaining parts of the transmission are one-dimensional, and—if needed—they can be compensated for by individual adaptation during playback. A recording with as little individual information as possible is obtained by recording outside a blocked ear canal (p_2).

2.4 Where does the one-dimensional transmission start?

When the expression 'entrance to the ear canal' is used, it is assumed that from this point the propagation into the ear canal is one-dimensional and thus independent of direction and distance to the sound source. Whether the one-dimensional propagation starts immediately at the entrance, or it is necessary to go a few millimeters into the ear canal, is not clear. From a theoretical point of view, the axial modes will attenuate with distance from the entrance, and most authors mention a point 0–4 mm from the ear-canal entrance. However, only sparse experimental verification is available, when specific deviations from a directionally independent transmission are considered.

Mehrgardt and Mellert[70] used a point 2 mm inside the ear canal, and reported very briefly dependence upon the direction, if a point outside this was chosen. A different result was obtained in a recent study carried out at our laboratory.[69]

In this study, transfer functions were measured from various points in the ear canal to the eardrum. The measurements were made with probe microphones in the ear canals of four humans. The probes had an outer diameter of 1·5 mm, and did not disturb the sound field in the ear canal. The sound was transmitted from three different directions in a free sound field (left ear, $\phi = 0°$, $90°$ and $180°$, $\theta = 0°$).

The results from a typical subject are shown in Fig. 4. It is seen that the transmission in the two lower lines of the figure is highly dependent upon the direction of sound incidence. The differences exist for frequencies above approximately 4 kHz. The four upper graphs in Fig. 4 show no or very little dependence on direction. This means that the one-dimensional transmission has already started at a point located a few millimeters outside the ear canal.

This conclusion has been based on measurements covering only three directions. Unpublished measurements at our laboratory of the pressure ratio $[P_3/P_2]$ at the entrance to the ear canal with five directions of sound incidence ($\theta = 0°$, $\phi = 0°, 90°, 180°, 270°$ and $\theta = 90°$) show the same absence of directional dependence.

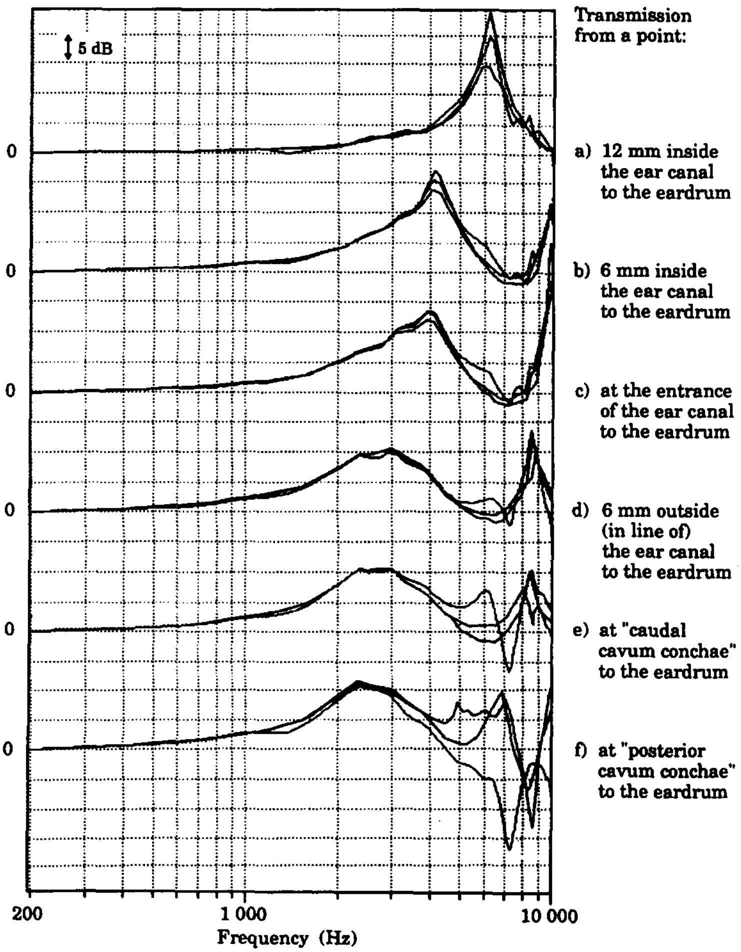

Fig. 4. Magnitude of transfer function from different points in the ear canal to the eardrum on a typical subject. Each line shows measurements with sound from three directions.

Observations similar to the above were reported by Middlebrooks *et al.* from an investigation including 356 source directions.[71] They found directional dependence in the transmission from the cavum concha to the ear canal and from a point 10 mm lateral of tragus (entirely free of the external ear) to the ear canal. Within the ear canal, including the point of entrance, the transmission was not dependent on direction. Middlebrooks *et al.* did not include a point similar to the point in Fig. 4, line (d).[71]

The consequence of this is that the physical location of a point, where full directional information is present, may be chosen anywhere from the eardrum to the entrance of the ear canal. Possibly, even points a few millimeters outside the ear canal and in line with it may be used.

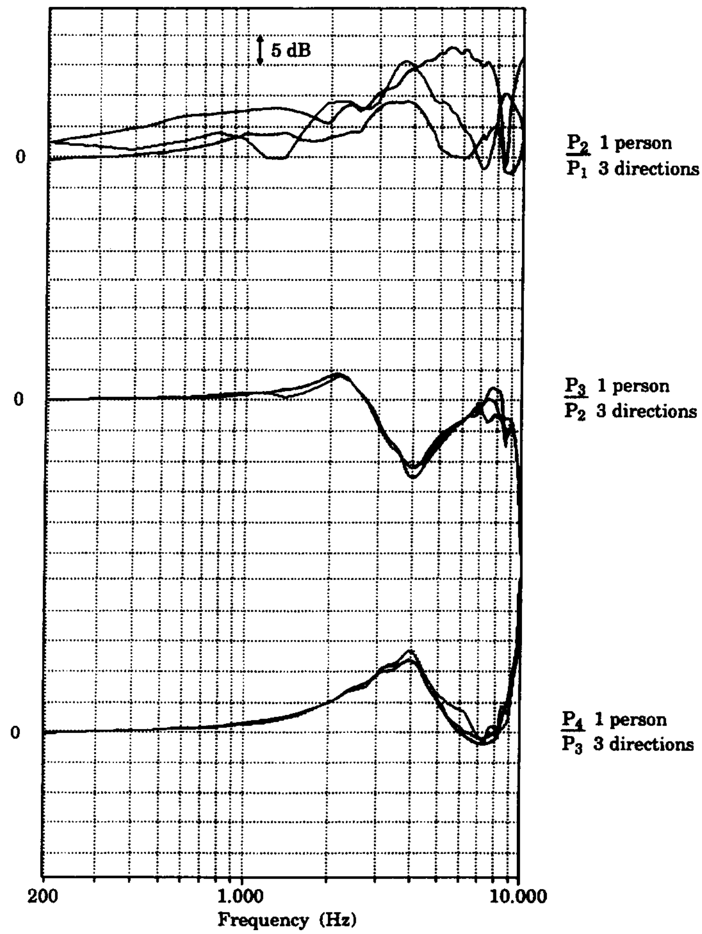

Fig. 5. Magnitude of transfer functions for one subject, one ear. Each line shows measurements from three different directions.

2.5 Measurement of head-related transfer functions

None of the head-related transfer functions can be measured directly, since the two sound pressures involved do not exist physically at the same time. However, it is possible to introduce $[P_1/E_{\text{loudspeaker}}]$, this being the loudspeaker frequency response measured in a position in the middle of the head, but with the listener absent. Then the head-related transfer functions can be found as transfer functions from voltage at the loudspeaker terminals to the appropriate sound pressure, denoted by p_i, divided by the loudspeaker response.

$$[P_i/P_1] = \frac{[P_i/E_{\text{loudspeaker}}]}{[P_1/E_{\text{loudspeaker}}]} \tag{13}$$

Because of their physical size, normal microphones cannot be used for measurement of the involved sound pressures, except for p_1. Miniature microphones or probe microphones are needed. This implies some practical problems. Miniature microphones often have a low sensitivity and may be difficult to calibrate. Also, probe microphones may have low sensitivity, and as an additional disadvantage, they often have a non-flat frequency response because of standing waves in the probe tube.

At high frequencies (above approximately 10 kHz) it may also be a problem that probe and miniature microphones measure the sound pressure at one specific point. It would probably be better to measure the sound pressure integrated over a cross-section of the ear canal—just like the hearing does itself at the eardrum.

Examples of measurements of the sound transmission to the human ear canal, are given in Fig. 5 (measured in our laboratory, experiments further reported by Hammershøi *et al.*)[35]

3 LISTENING WITH HEADPHONES

Binaural signals are normally reproduced with headphones. This guarantees that each channel is sent to only one ear. It also makes the reproduction free from reflections from the listening room. If loudspeakers were used, crosstalk and reflections from the listening room would disturb the reproduction. This section gives the theory needed to describe the playback of binaural recordings through headphones.

3.1 Transfer functions for a headphone

The reproduction through a headphone is illustrated in Fig. 6. The transmission from voltage at the headphone terminals, $e_{\text{headphone}}$, can be

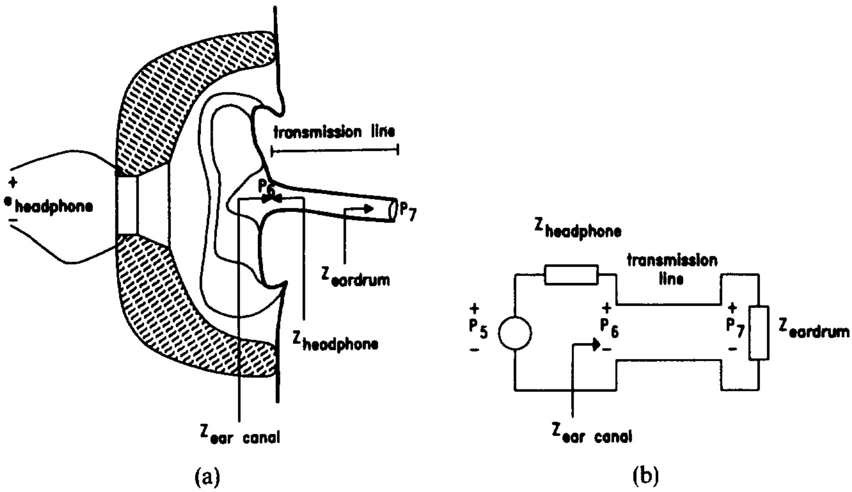

Fig. 6. Sound transmission through the external ear, when the source is a headphone: (a) sketch of the anatomy and (b) an analogue model.

described with some analogy to the description in Section 2 of the transmission from the free field. Here, p_7 denotes the pressure at the eardrum, p_6 the pressure at the entrance to the ear canal and p_5 the 'open circuit' pressure at the entrance. $Z_{\text{headphone}}$ is the impedance being looked out into from the entrance, including a possible volume enclosed by the headphone, and the mechanical and electrical side of the headphone, transferred to the acoustical side.

The following *headphone transfer functions* are important:

$$[P_7/E_{\text{headphone}}] = \frac{\text{sound pressure at the eardrum}}{\text{voltage at the headphone terminals}} \tag{14}$$

$$[P_6/E_{\text{headphone}}] = \frac{\text{sound pressure at the entrance to the open ear canal}}{\text{voltage at the headphone terminals}} \tag{15}$$

$$[P_5/E_{\text{headphone}}] = \frac{\text{sound pressure at the entrance to the blocked ear canal}}{\text{voltage at the headphone terminals}} \tag{16}$$

Just like in the free field, a set of transfer functions exists for either side. If symmetry is assumed, then

$$[P_{i,\text{left}}/E_{\text{headphone}}] = [P_{i,\text{right}}/E_{\text{headphone}}] \tag{17}$$

where i refers to 5, 6 or 7.

3.2 Splitting up the transmission

As for the free-field transmission, the transmission can be split up:

$$[P_7/E_{\text{headphone}}] = [P_7/P_6] \cdot [P_6/P_5] \cdot [P_5/E_{\text{headphone}}] \tag{18}$$

The expected dependencies are:

$$[P_5/E_{\text{headphone}}](\text{headphone, subject, side}) \tag{19}$$

$$[P_6/P_5](\text{headphone, subject, side}) \tag{20}$$

$$[P_7/P_6](\text{subject, side}) \tag{21}$$

3.3 Measurement of headphone transfer functions

The headphone transfer functions can be measured directly. The input is the voltage at the headphone terminals, and the response is the sound pressure at some point in the external ear.

Like in the free field, miniature or probe microphones must be used in order not to disturb the sound field in the confined space in the ear. None of the transfer functions include any directional or distance dependence, but of course they refer to a well-defined position of the headphone relative to the ear.

3.4 Similarities in free-field and headphone listening

Although two different descriptions have been given, part of the transmission is the same in the two cases.

The transmission from p_3 to p_4 depends entirely upon the ear canal and the termination of it, Z_{eardrum}. These components are exactly the same for the transmission from p_6 to p_7. Therefore,

$$[P_7/P_6] = [P_4/P_3] \tag{22}$$

The transmissions from p_2 to p_3 and from p_5 to p_6 consist of pressure divisions. An expression for $[P_3/P_2]$ was given in eqn (3), and $[P_6/P_5]$ can be expressed as

$$[P_6/P_5] = \frac{Z_{\text{ear canal}}}{Z_{\text{ear canal}} + Z_{\text{headphone}}} \tag{23}$$

In some of the following sections, the expression $([P_3/P_2]/[P_6/P_5])$ will appear. If the eqns (3) and (23) are combined, then

$$\frac{[P_3/P_2]}{[P_6/P_5]} = \frac{Z_{\text{ear canal}} + Z_{\text{headphone}}}{Z_{\text{ear canal}} + Z_{\text{radiation}}} \tag{24}$$

Simplifications can be made if the impedances $Z_{\text{radiation}}$ and $Z_{\text{headphone}}$ are equal, or if they are both small compared to $Z_{\text{ear canal}}$. A measurement of the pressure ratio $[P_3/P_2]$ (eqn (3)), is seen in Fig. 5. From the figure, it is seen that there is a considerable loading of $Z_{\text{radiation}}$ at some frequencies, and it is concluded that $Z_{\text{radiation}}$ is not in general small compared to $Z_{\text{ear canal}}$. Consequently, simplifications require that

$$Z_{\text{headphone}} \approx Z_{\text{radiation}} \tag{25}$$

At frequencies below 1 kHz, there is no loading of $Z_{\text{radiation}}$, so $Z_{\text{radiation}} \ll Z_{\text{ear canal}}$, and the requirement reduces to

$$Z_{\text{headphone}} \ll Z_{\text{ear canal}} \tag{26}$$

Headphones that fulfil these requirements to the acoustical impedance are called *open*. The term *open* can be explained like this: a headphone that does not disturb the radiation impedance as seen from the ear could be a relatively small unit positioned some distance from the ear. Such a unit would seem rather 'open'. Be aware that this definition may differ from the commercial use of the expression *open headphone*, where it probably only means that sounds from outside can be heard when wearing the headphone.

For an open headphone, eqn (24) becomes unity, and

$$[P_3/P_2] \approx [P_6/P_5] \tag{27}$$

Møller *et al.*[36] examined a number of headphones with respect to the criteria of being open.

4 BINAURAL TECHNIQUE

This section deals with different approaches to binaural technique. All methods include recording somewhere in the ear canals of a recording head, transmission through a filter, G, and playback through headphones. This is illustrated in Fig. 7.

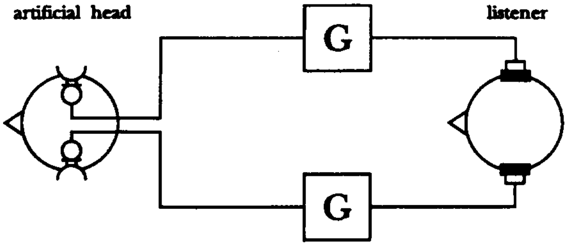

Fig. 7. Transmission in a binaural system.

The recording head may be artificial or human, although it is assumed that it has the same acoustical properties as the listener's head. Recording will be considered at three different recording points, where the sound pressure contains full spatial information. This gives three different methods:

(a) recording at the eardrum;
(b) recording at the entrance to the open ear canal;
(c) recording at the entrance to the blocked ear canal.

4.1 Method A: recording at the eardrum

The basic binaural technique includes recording of the pressure at the position of the eardrum and playback through a headphone. If the microphone has a transfer function M_1 (including sensitivity as well as frequency response), and the electrical transfer function from the microphone to the headphone is G_A, then the total transfer function from the sound field, p_1, to sound pressure at the eardrum, p_7, is

$$[P_4/P_1] \cdot M_1 \cdot G_A \cdot [P_7/E_{\text{headphone}}] \tag{28}$$

It is the goal that this should be:

$$[P_4/P_1] \tag{29}$$

(here, of course, the time delay from the recording event to the moment of reproduction is ignored). If expressions (28) and (29) are made equal, then

$$G_A = \frac{[P_4/P_1]}{[P_4/P_1] \cdot M_1 \cdot [P_7/E_{\text{headphone}}]} = \frac{1}{M_1 \cdot [P_7/E_{\text{headphone}}]} \tag{30}$$

Conclusion. The correct transfer function is obtained when the electronic circuit compensates for (1) the microphone sensitivity and (2) the headphone transfer function from the terminals to the sound pressure at the eardrum.

4.2 Method B: recording at the entrance to the open ear canal

As argued for in Section 2, the sound pressure at the entrance to the open ear canal includes full spatial information. it is therefore appropriate also to have a look at the situation of when this sound pressure is recorded.

The electrical gain is called G_B, and the total transfer function from the sound field to the eardrum is

$$[P_3/P_1] \cdot M_1 \cdot G_B \cdot [P_7/E_{\text{headphone}}] \tag{31}$$

while it should still be

$$[P_4/P_1] \tag{32}$$

This goal is reached, if expressions (31) and (32) are made equal, which gives

$$G_B = \frac{[P_4/P_1]}{[P_3/P_1]\cdot M_1 \cdot [P_7/E_{\text{headphone}}]} = \frac{[P_4/P_3]}{[P_7/P_6]}\cdot\frac{1}{M_1\cdot[P_6/E_{\text{headphone}}]} \tag{33}$$

The first term here is unity, since the transmission from pressure at the entrance to the open ear canal to the eardrum is the same, whatever the source (see eqn (22)). Then

$$G_B = \frac{1}{M_1 \cdot [P_6/E_{\text{headphone}}]} \tag{34}$$

Conclusion. The correct transfer function is obtained when the electronic circuit compensates for (1) the microphone sensitivity and (2) the headphone transfer function from the terminals to the sound pressure at the entrance to the open ear canal.

4.3 Method C: recording at the entrance to the blocked ear canal

The sound pressure at the entrance to the blocked ear canal also contains the complete spatial information.

The electrical gain is here called G_C, and the total transfer from the free sound field to the sound pressure at the eardrum is:

$$[P_2/P_1]\cdot M_1 \cdot G_C \cdot [P_7/E_{\text{headphone}}] \tag{35}$$

Again it should be

$$[P_4/P_1] \tag{36}$$

This is fulfilled if expressions (35) and (36) are equal, which gives

$$G_C = \frac{[P_4/P_1]}{[P_2/P_1]\cdot M_1 \cdot [P_7/E_{\text{headphone}}]} = \frac{[P_4/P_3]}{[P_7/P_6]}\cdot\frac{[P_3/P_2]}{[P_6/P_5]}\cdot\frac{1}{M_1\cdot[P_5/E_{\text{headphone}}]} \tag{37}$$

As above, the first term is unity, and the second can be exchanged with expression (24), so

$$G_C = \frac{[P_3/P_2]}{[P_6/P_5]}\cdot\frac{1}{M_1\cdot[P_5/E_{\text{headphone}}]} = \frac{Z_{\text{ear canal}} + Z_{\text{headphone}}}{Z_{\text{ear canal}} + Z_{\text{radiation}}}\cdot\frac{1}{M_1\cdot[P_5/E_{\text{headphone}}]} \tag{38}$$

Sometimes it is not possible or practical to measure the first term. Fortunately, according to eqn (27) this term is unity, if the headphone is ideally open as defined in eqns (25) and (26). If G_C^* is used to denote an approximate value for G_C—applicable when an open headphone is used—then

$$G_C \approx G_C^* = \frac{1}{M_1 \cdot [P_5/E_{\text{headphone}}]} \tag{39}$$

Conclusion. The correct transfer function is obtained when the electronic circuit compensates for (1) the microphone sensitivity, (2) the headphone transfer function from the terminals to the sound pressure at the entrance to the blocked ear canal and (3) the transmission difference caused by different acoustic source impedances in the two listening situations. When an ideally open headphone is used, compensation for (3) is not needed.

4.4 Comparison of the three methods

In the previous sections, it has been shown that all three methods can be adapted to give the correct transmission from the recorded sound field to the listener's eardrum, simply by choosing the proper equalization. This is not surprising, since the only difference between the methods is the one-dimensional transmission from open circuit pressure at the input of the ear canal to pressure at the eardrum.

All methods require recording at a point in the ear canal and determination of the headphone transfer function to the same point in the ear canal. The point is sometimes called the *reference point*. Determination of the headphone transfer function is called *calibration* of the headphone. The third method requires an extra correction to the equalization. This correction, though, disappears when an ideally open headphone is used.

If a human subject is used for recording and headphone calibration, the first method with recording at the eardrum might be considered inconvenient because of the practical problems involved in recording the sound here. In Method C the ear canal is blocked, and this method has the disadvantage that the listener is not able to hear the sound during recording.

For recording with an artificial head, Method C is attractive, since the head needs no ear canals. The method is particularly simple, if an ideally open headphone is used for reproduction.

It is not clear which method will give the best results, when differences occur between the recording head, the head used for headphone calibration and the head of the listener. Method C may seem immediately attractive, since the recorded signals contain as little individual information as possible. Only the transfer to p_2 is included, and less will not do.

4.5 Significance of the microphone used for headphone calibration

All expressions for G include the transfer function of the microphone M_1. Therefore, a calibration of this would be necessary. In the following, however, it is shown, how this calibration can be avoided in certain cases.

The calibration of the headphone has until now been described as a measurement of the transfer function from voltage at the headphone terminals, $e_{\text{headphone}}$, to sound pressure somewhere in the ear canal, p_5, p_6 or p_7. When this is done, it involves a microphone for measurement of the sound pressure. Let the transfer function (sensitivity and frequency response) of this microphone be M_2, its output $e_{\text{microphone}}$, and the sound pressure be generally called p_i. G_A, G_B and G_C are in general called G. The expressions from eqns (30), (34) and (39) (also (38), if the impedance term is omitted) will have the form:

$$\begin{aligned} G &= \frac{1}{M_1 \cdot [P_i/E_{\text{headphone}}]} \\ &= \frac{1}{M_1 \cdot [(E_{\text{microphone}}/M_2)/E_{\text{headphone}}]} \\ &= \frac{M_2}{M_1} \cdot \frac{E_{\text{headphone}}}{E_{\text{microphone}}} \end{aligned} \tag{40}$$

It is seen that the electrical gain, G, must equalize the transfer function measured from headphone voltage to output voltage of the microphone, as well as the ratio between the two microphones used for calibration and for recording.

Now, the smart reader will probably suggest that the same microphone be used for recording and calibration, so $M_1 = M_2$. In that case eqn (40) becomes

$$G = \frac{E_{\text{headphone}}}{E_{\text{microphone}}} \tag{41}$$

and the need for calibration of the microphones is avoided.

4.6 Other recording methods

Until now, it has been assumed that the recording could be made without disturbing the sound field with the microphone. It may be possible with an artificial head, where a microphone can be built into the head, but this possibility does not exist with a real head. For measurement purposes, probe microphones with very thin probe diameter will do the job, but their internal noise is too high for making recordings.

Very often, miniature microphones are used, a typical one having dimensions of 2 mm × 3 mm × 4 mm. If method C is used, microphones of this size can be used without disturbance, since they can be inserted in an earplug that blocks the ear canal.

If a miniature microphone is inserted in the open ear canal (Methods A and B), it will disturb the sound field. What this disturbance means for the final transfer function will be calculated in the following.

It is assumed that the microphone is inserted in the open ear canal, and that it measures the pressure at a point called x, somewhere between the p_3 and p_4 points. The calculations will take into account a possible disturbance of the sound transmission at both sides of the measuring point of the microphone.

The transfer function from free field to output of the microphone is

$$[P_x/P_1]^* \cdot M_1 \tag{42}$$

where the asterisk indicates a value with the microphone inserted. If calibration of the headphone is done with the same microphone at the same position as described in Section 4.5, then

$$G = \frac{1}{M_1 \cdot [P_x/E_{\text{headphone}}]^*} \tag{43}$$

The total transfer function to the eardrum will be

$$[P_x/P_1]^* \cdot M_1 \cdot \frac{1}{M_1 \cdot [P_x/E_{\text{headphone}}]^*} \cdot [P_7/E_{\text{headphone}}] = \frac{[P_x/P_1]^* \cdot [P_7/E_{\text{headphone}}]}{[P_x/E_{\text{headphone}}]^*} \tag{44}$$

This should have been

$$[P_4/P_1] \tag{45}$$

The error is found by dividing eqn (44) by (45).

$$\begin{aligned} \text{error} &= \frac{[P_x/P_1]^* \cdot [P_7/E_{\text{headphone}}]}{[P_4/P_1] \cdot [P_x/E_{\text{headphone}}]^*} \\ &= \frac{[P_x/P_3]^* \cdot [P_3/P_2]^* \cdot [P_2/P_1]^* \cdot [P_7/P_6] \cdot [P_6/P_5] \cdot [P_5/E_{\text{headphone}}]}{[P_4/P_3] \cdot [P_3/P_2] \cdot [P_2/P_1] \cdot [P_x/P_6]^* \cdot [P_6/P_5]^* \cdot [P_5/E_{\text{headphone}}]^*} \end{aligned} \tag{46}$$

$[P_2/P_1]$ and $[P_5/E_{\text{headphone}}]$ are the same with and without the microphone. The same arguments as in Section 3.4 can be used, so

$$[P_x/P_6]^* = [P_x/P_3]^* \tag{47}$$

Now, by also using eqns (22), (3) and (23), eqn (46) can be reduced to

$$\text{error} = \frac{[P_3/P_2]^*}{[P_3/P_2]} \cdot \frac{[P_6/P_5]}{[P_6/P_5]^*}$$

$$= \frac{Z^*_{\text{ear canal}}/(Z^*_{\text{ear canal}} + Z_{\text{radiation}})}{Z_{\text{ear canal}}/(Z_{\text{ear canal}} + Z_{\text{radiation}})} \cdot \frac{Z_{\text{ear canal}}/(Z_{\text{ear canal}} + Z_{\text{headphone}})}{Z^*_{\text{ear canal}}/(Z^*_{\text{ear canal}} + Z_{\text{headphone}})}$$

$$= \frac{Z_{\text{ear canal}} + Z_{\text{radiation}}}{Z^*_{\text{ear canal}} + Z_{\text{radiation}}} \cdot \frac{Z^*_{\text{ear canal}} + Z_{\text{headphone}}}{Z_{\text{ear canal}} + Z_{\text{headphone}}} \tag{48}$$

The error is only connected to the incorrect impedance match. If an open headphone is employed, eqn (25) can be used, and eqn (48) reduces to unity, so no error is then introduced.

If the head used for recording and calibration is artificial and has an incorrect ear canal, it leads to changes in the transmission similar to those given above. It can therefore be concluded that an incorrect ear canal is acceptable without special corrective means, if an open headphone is used for reproduction.

5 EQUALIZATION FOR LOUDSPEAKER COMPATIBILITY

In Section 4 it was shown how the correct transmission of sound from the artificial head to the listener's eardrums was achieved by the introduction of an electrical equalizing filter G. If the signals either before or after this filter are reproduced through a traditional stereo set-up with loudspeakers, a coloration will in general occur because of the non-flat frequency response from the original sound field to the voltage at these points.

However, it is possible to divide the G filter into two, as illustrated in Fig. 8. The filters are called G' and G'', and

$$G' \cdot G'' = G \tag{49}$$

The two filters should be chosen so that the frequency response would be flat from the free sound field to the dividing point between them. In this way signals from the dividing point can be played through loudspeakers without coloration.

As mentioned in Section 1, the way of dividing G has been the matter of some discussion. The two main concepts are free-field equalization and diffuse-field equalization. These terms will be described in the following.

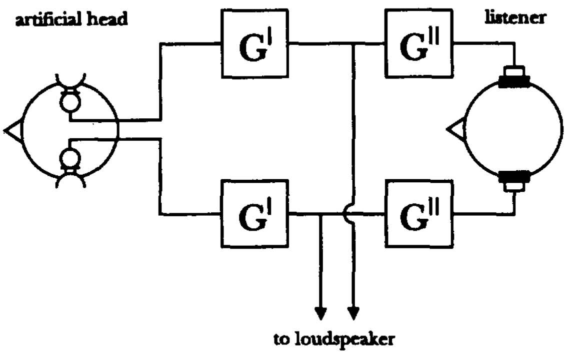

Fig. 8. Division of the G filter into G' and G''.

As the recording head has different transfer functions for different angles of sound incidence—that is the idea of binaural recording—a flat frequency response cannot be obtained for all angles. As most sound sources are in front of the recording head, it has been argued that a flat frequency response should be obtained for sound arriving from a sound source in the front. This way of choosing G' and G'' is called *free-field equalization.*

It has also been argued that the recording head is normally further away from the sound sources than the hall radius of the recording room. Then it is located in the diffuse part of the sound field, and a flat frequency response should be preferred for sound coming from random directions. This choice is called *diffuse-field equalization.*

5.1 Free-field equalization

In this case there should be a flat frequency response for a sound arriving from the front to voltage after filter G'. For method A this means

$$[P_4/P_1](\phi = 0°, \theta = 0°) \cdot M_1 \cdot G'_{\text{A free field}} = \text{constant} \tag{50}$$

where the constant determines the gain. It has the dimension voltage per pressure, giving the units V/Pa. $G'_{\text{A free field}}$ and $G''_{\text{A free field}}$ are easily found from eqn (50), by using eqn (30)

$$G'_{\text{A free field}} = \frac{\text{constant}}{M_1 \cdot [P_4/P_1](\phi = 0°, \theta = 0°)} \tag{51}$$

$$G''_{\text{A free field}} = \frac{G_{\text{A}}}{G'_{\text{A free field}}} = \frac{[P_4/P_1](\phi = 0°, \theta = 0°)}{[P_7/E_{\text{headphone}}] \cdot \text{constant}} \tag{52}$$

Similar expressions can be derived for recording in methods B and C (using eqns (34) and (38)).

$$G'_{\text{B free field}} = \frac{\text{constant}}{M_1 \cdot [P_3/P_1](\phi = 0°, \theta = 0°)} \tag{53}$$

$$G'_{\text{C free field}} = \frac{\text{constant}}{M_1 \cdot [P_2/P_1](\phi = 0°, \theta = 0°)} \tag{54}$$

$$G''_{\text{B free field}} = \frac{[P_3/P_1](\phi = 0°, \theta = 0°)}{[P_6/E_{\text{headphone}}] \cdot \text{constant}} \tag{55}$$

$$G''_{\text{C free field}} = \frac{Z_{\text{ear canal}} + Z_{\text{headphone}}}{Z_{\text{ear canal}} + Z_{\text{radiation}}} \cdot \frac{[P_2/P_1](\phi = 0°, \theta = 0°)}{[P_5/E_{\text{headphone}}] \cdot \text{constant}} \tag{56}$$

A head followed by the filter $G'_{\text{free field}}$ (or a head that is constructed so that $G'_{\text{free field}}$ becomes a constant) is called free-field-equalized. Usually this refers to an artificial head, but the expression can be used in connection with a real head as well. The transfer function for the free-field-equalized head with recording at the eardrum becomes

$$\begin{aligned} [P_4/P_1] \cdot M_1 \cdot G'_{\text{A free field}} &= \frac{[P_4/P_1] \cdot M_1 \cdot \text{constant}}{M_1 \cdot [P_4/P_1](\phi = 0°, \theta = 0°)} \\ &= \text{constant} \frac{[P_2/P_1](r, \phi, \theta)}{[P_2/P_1](\phi = 0°, \theta = 0°)} \end{aligned} \tag{57}$$

Exactly the same expression can be derived for

$$[P_3/P_1] \cdot M_1 \cdot G'_{\text{B free field}} \tag{58}$$

and

$$[P_2/P_1] \cdot M_1 \cdot G'_{\text{C free field}} \tag{59}$$

This indicates that the output of a free-field-equalized head is independent of the recording point in the ear canal. This is not unexpected. In Section 2.3 it was stated that the only difference between the pressure at the three points is a one-dimensional transmission. When the total transmission is equalized to give a flat frequency response for a certain direction in all methods, the output for other directions will also be the same.

Use of definition (7) in expression (57), shows that the transfer function of a free-field-equalized head is given by

$$\text{constant} \cdot [\text{monaural transfer function}] \tag{60}$$

5.2 Diffuse-field equalization

Here a flat frequency response is wanted, when the head is placed in a diffuse sound field. The filter $G'_{\text{diffuse field}}$ can be determined from

$$[P_4/P_1](\text{average of all angles}) \cdot M_1 \cdot G'_{\text{A diffuse field}} = \text{constant} \tag{61}$$

Then

$$G'_{\text{A diffuse field}} = \frac{\text{constant}}{M_1 \cdot [P_4/P_1](\text{average of all angles})} \tag{62}$$

$$G''_{\text{A diffuse field}} = \frac{G_{\text{A}}}{G'_{\text{A diffuse field}}} = \frac{[P_4/P_1](\text{average of all angles})}{[P_7/E_{\text{headphone}}] \cdot \text{constant}} \tag{63}$$

For methods B and C, the following expressions are derived

$$G'_{\text{B diffuse field}} = \frac{\text{constant}}{M_1 \cdot [P_3/P_1](\text{average of all angles})} \tag{64}$$

$$G'_{\text{C diffuse field}} = \frac{\text{constant}}{M_1 \cdot [P_2/P_1](\text{average of all angles})} \tag{65}$$

$$G''_{\text{B diffuse field}} = \frac{[P_3/P_1](\text{average of all angles})}{[P_6/E_{\text{headphone}}] \cdot \text{constant}} \tag{66}$$

$$G''_{\text{C diffuse field}} = \frac{Z_{\text{ear canal}} + Z_{\text{headphone}}}{Z_{\text{ear canal}} + Z_{\text{radiation}}} \cdot \frac{[P_2/P_1](\text{average of all angles})}{[P_5/E_{\text{headphone}}] \cdot \text{constant}} \tag{67}$$

A diffuse-field-equalized recording head is defined as a head followed by the filter $G'_{\text{diffuse field}}$ (or a head that is constructed so that $G'_{\text{diffuse field}}$ becomes a constant). The transfer function for the diffuse-field-equalized head with recording at the eardrum becomes

$$\begin{aligned}[P_4/P_1] \cdot M_1 \cdot G'_{\text{A diffuse field}} &= \frac{[P_4/P_1] \cdot M_1 \cdot \text{constant}}{M_1 \cdot [P_4/P_1](\text{average of all angles})} \\ &= \text{constant} \cdot \frac{[P_2/P_1](r, \phi, \theta)}{[P_2/P_1](\text{average of all angles})}\end{aligned} \tag{68}$$

Also, here the same expression can be derived for

$$[P_3/P_1] \cdot M_1 \cdot G'_{\text{B diffuse field}} \tag{69}$$

and

$$[P_2/P_1] \cdot M_1 \cdot G'_{\text{C diffuse field}} \tag{70}$$

So also diffuse-field-equalized heads perform the same, independent of the recording point in the ear canal.

Note that only the magnitude of $[P_i/P_1]$ (average of all angles) has a meaning (see the arguments in connection with eqn (9)). Then it is only possible to determine the magnitude of $G'_{\text{diffuse field}}$ and $G''_{\text{diffuse field}}$, and minimum phase realizations of the filters are normally used.

Use of definition (9) in expression (68), shows that the output of a diffuse-field-equalized head is

$$\text{output} = \text{constant} \cdot [\text{monaural transfer function referenced to diffuse field}] \tag{71}$$

5.3 Headphones used for traditional stereo recordings

A headphone connected with the filter $G''_{\text{free field}}$—or constructed so the filter becomes a constant—is called free-field-equalized. If calibration is done according to method A and eqn (52) is inserted, the transfer function to the eardrum will be

$$\begin{aligned} G''_{\text{A free field}} \cdot [P_7/E_{\text{headphone}}] &= \frac{[P_4/P_1](\phi = 0°, \theta = 0°)}{[P_7/E_{\text{headphone}}] \cdot \text{constant}} \cdot [P_7/E_{\text{headphone}}] \\ &= \frac{[P_4/P_1](\phi = 0°, \theta = 0°)}{\text{constant}} \end{aligned} \tag{72}$$

Exactly the same expression is obtained for

$$G''_{\text{B free field}} \cdot [P_7/E_{\text{headphone}}] \tag{73}$$

and

$$G''_{\text{C free field}} \cdot [P_7/E_{\text{headphone}}] \tag{74}$$

from methods B and C (eqns (55) and (56) are inserted and eqns (22) and (24) are used). This indicates that a headphone that is free-field-equalized according to one method, is also free-field-equalized according to the other methods. As for the free-field-equalized head, this is not unexpected.

A verbal description of eqn (72) is as follows: A free-field-equalized headphone transmits the signals to the eardrums in the same way as—in the free field—a sound wave from the front is transmitted, This makes a free-field-equalized headphone useful for playback of traditional stereo recordings, since this type of headphone simulates listening to the direct sound from a loudspeaker with an ideally flat free-field frequency response.

In a normal stereo set-up the angle to the loudspeaker is 30°, and therefore it may be argued that an azimuth angle of 30° should be used for free-field-equalization of headphones.

At this point it is worth mentioning that reproduction with headphones will be very different from reproduction through loudspeakers, even when the headphone is properly equalized. Normally internal localization will occur.[72-76] This is due to the lack of crosstalk and the lack of reflections from the listening room.

It has been argued that when listening to loudspeakers, the listener is usually in the reverberant part of the sound field in the listening room. Therefore, headphone equalization that simulates listening in a diffuse field may be more appropriate. This equalization is obtained with the headphone and the $G''_{\text{diffuse field}}$ filters. If method A is used, and eqn (63) is inserted, then

$$G''_{\text{A diffuse field}} \cdot [P_7/E_{\text{headphone}}] = \frac{[P_4/P_1](\text{average of all angles})}{\text{constant}} \tag{75}$$

The same result is found for Methods B and C (eqns (66) and (67) are inserted, and eqns (22) and (24) are used). So, as for free-field equalization, it does not matter at which recording point the calibration is made.

Free-field- and diffuse-field-equalized headphones are commercially available, often in a form where the headphone itself is claimed to have the correct characteristic, so the filter G'' is not needed. If a free-field- or a diffuse-field-equalized headphone is used in connection with a similarly equalized artificial head, both filters G' and G'' should be omitted. Correct total equalization is obtained, when the output from the artificial head is connected directly to the headphone (through a power amplifier).

There may be some uncertainty about the quality of the free-field or diffuse-field equalization of the headphones. It is the author's experience that headphones that are claimed to have the same type of equalization, but produced by different manufacturers, may be very different in timbre as well as measured frequency response.[36]

6 PROCESSING FOR TRUE REPRODUCTION WITH LOUDSPEAKERS

The good directional characteristics of an artificial-head recording are destroyed if it is reproduced through loudspeakers. This is due to the crosstalk, which is introduced in any free-field situation. Crosstalk means that the right speaker is heard not only with the right ear, but also with the left ear and vice versa.

However, it can be shown that it is possible to add an artificial crosstalk which cancels out the natural crosstalk. Systems that perform crosstalk cancellation on binaural systems are sometimes called *transaural systems* or—earlier—*TRADIS systems* (true reproduction of all directional information by stereophony). As mentioned in the introduction, they have been known for some years. An early version involved—among other tricks—playback of a magnetic tape in reverse direction. The principle in a more recent system is shown in the following.[63]

$P_{\text{record, left}}$ and $P_{\text{record, right}}$ denote the two sound pressures recorded at the reference points in the ear canals of the recording head. M_1 is the transfer function of the recording microphone. $E_{\text{loudspeaker, left}}$ and $E_{\text{loudspeaker, right}}$ are the signals presented to the loudspeaker terminals. $P_{\text{playbacks, left}}$ and $P_{\text{playback, right}}$ denote the sound pressure at the reference points in the listener's ear canals. The transfer functions from loudspeakers to ear canals are denoted with Hs as indicated in Fig. 9.

The transmission is expressed in the following set of equations:

$$\begin{aligned} P_{\text{playback, left}} &= H_{\text{left-left}} \cdot E_{\text{loudspeaker, left}} + H_{\text{right-left}} \cdot E_{\text{loudspeaker, right}} \\ P_{\text{playback, right}} &= H_{\text{left-right}} \cdot E_{\text{loudspeaker, left}} + H_{\text{right-right}} \cdot E_{\text{loudspeaker, right}} \end{aligned} \tag{76}$$

The wanted transmission is

$$\begin{aligned} P_{\text{playback, left}} &= P_{\text{record, left}} \\ P_{\text{playback, right}} &= P_{\text{record, right}} \end{aligned} \tag{77}$$

If eqns (76) and (77) are combined and solved with respect to the Es, then

$$\begin{aligned} E_{\text{loudspeaker, left}} &= \frac{H_{\text{right-right}} \cdot P_{\text{record, left}} - H_{\text{right-left}} \cdot P_{\text{record, right}}}{H_{\text{left-left}} \cdot H_{\text{right-right}} - H_{\text{left-right}} \cdot H_{\text{right-left}}} \\ E_{\text{loudspeaker, right}} &= \frac{H_{\text{left-left}} \cdot P_{\text{record, right}} - H_{\text{left-right}} \cdot P_{\text{record, left}}}{H_{\text{left-left}} \cdot H_{\text{right-right}} - H_{\text{left-right}} \cdot H_{\text{right-left}}} \end{aligned} \tag{78}$$

Now symmetry is assumed, and the free-field response of the loudspeakers $[P_1/E_{\text{loudspeaker}}]$ is introduced along with the head-related transfer functions to the ear at the same side as the loudspeaker, $[P_{\text{playback}}/P_1]_{\text{same side}}$, and to the ear at the opposite side, $[P_{\text{playback}}/P_1]_{\text{opposite side}}$. Then

$$\begin{aligned} H_{\text{left-left}} = H_{\text{right-right}} &= [P_1/E_{\text{loudspeaker}}] \cdot [P_{\text{playback}}/P_1]_{\text{same side}} \\ H_{\text{left-right}} = H_{\text{right-left}} &= [P_1/E_{\text{loudspeaker}}] \cdot [P_{\text{playback}}/P_1]_{\text{opposite side}} \end{aligned} \tag{79}$$

If these terms are entered into eqn (78), then some manipulation leads to the following equation for each of the sides

$$E_{\text{loudspeaker}} = \frac{P_{\text{record, same side}} - P_{\text{record, opposite side}} \dfrac{[P_{\text{playback}}/P_1]_{\text{opposite side}}}{[P_{\text{playback}}/P_1]_{\text{same side}}}}{[P_{\text{playback}}/P_1]_{\text{same side}} [P_1/E_{\text{loudspeaker}}] \left(1 - \dfrac{[P_{\text{playback}}/P_1]^2_{\text{opposite side}}}{[P_{\text{playback}}/P_1]^2_{\text{same side}}}\right)} \tag{80}$$

The term

$$\frac{[P_{\text{playback}}/P_1]_{\text{opposite side}}}{[P_{\text{playback}}/P_1]_{\text{same side}}} \tag{81}$$

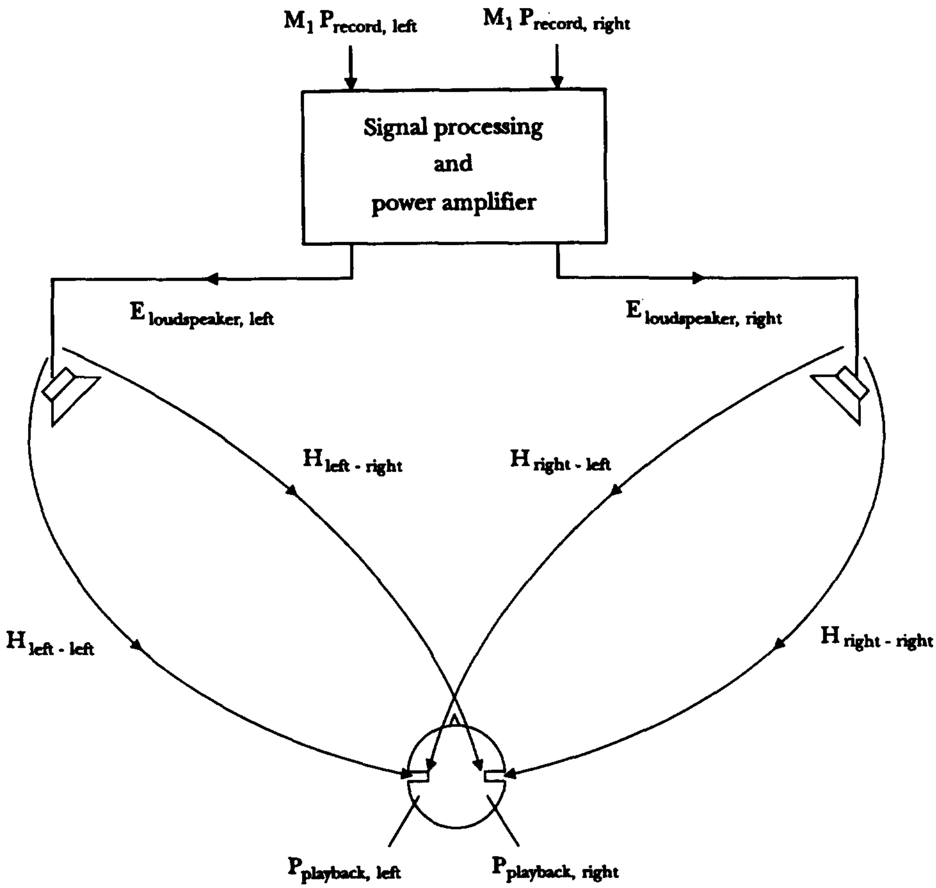

Fig. 9. Principal diagram showing the transmissions from recorded voltages to sound pressures in the ear canal.

is easily recognised as the *interaural transfer function* (see eqn (8)). In the following, it is denoted *ITF*.

$$[P_{\text{playback}}/P_1]_{\text{same side}} \cdot [P_1/E_{\text{loudspeaker}}] = [P_{\text{playback}}/E_{\text{loudspeaker}}]_{\text{same side}} \quad (82)$$

is the transfer function from a loudspeaker to the reference point in the ear canal at the side facing the loudspeaker. Now, eqn (80) can be written as

$$E_{\text{loudspeaker}} = \frac{P_{\text{record, same side}} - P_{\text{record, opposite side}} \cdot ITF}{[P_{\text{playback}}/E_{\text{loudspeaker}}]_{\text{same side}}(1 - ITF^2)} \quad (83)$$

This signal processing is shown in block-diagram form in Fig. 10. The blocks to the left perform an equalization of the recording microphone and the loudspeakers. The next blocks also perform an equalization, necessary because of coloration due to the crosstalk. At frequencies where the

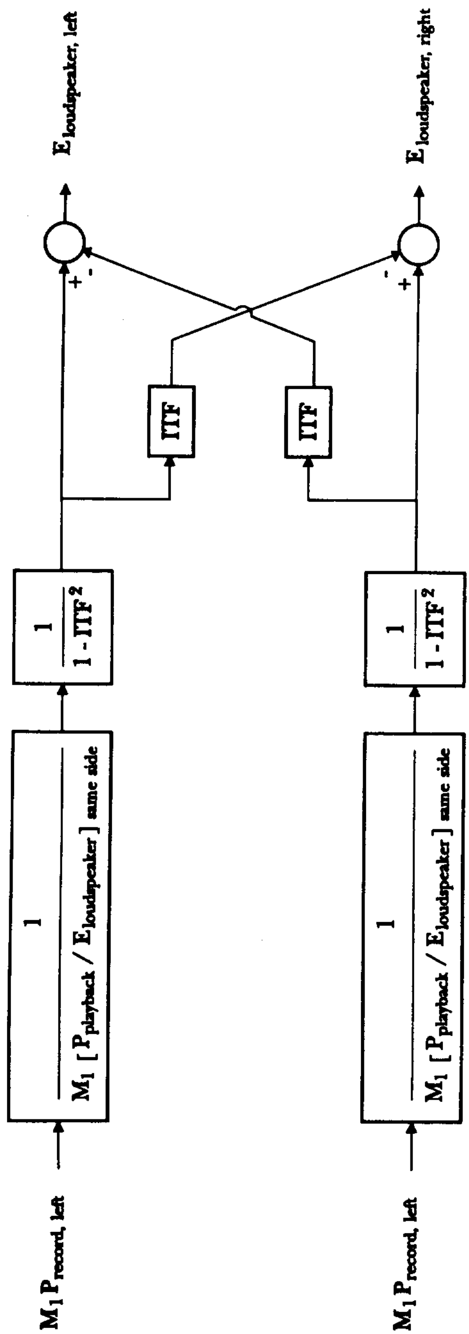

Fig. 10. The signal processing of eqn (83) shown in block diagram form.

crosstalk is low, this block is almost unity. The real suppressing of crosstalk is carried out by the two cross-coupled blocks.

The transfer function from voltage at loudspeaker terminals to sound pressure at the reference point in the listener's ear canal includes a delay. Naturally, this delay should also be accepted for the whole system, otherwise a non-causal filter would be required. A number of other practical considerations must be taken into account in a realization. Among these are problems at low frequencies, where the interaural transfer function approaches unity, and special care must be taken not to require a division by zero in the block containing $1/(1-ITF^2)$.

Examples of transfer functions, measured on a Neumann KU80 artificial head, are given in Fig. 11, and the corresponding interaural impulse response

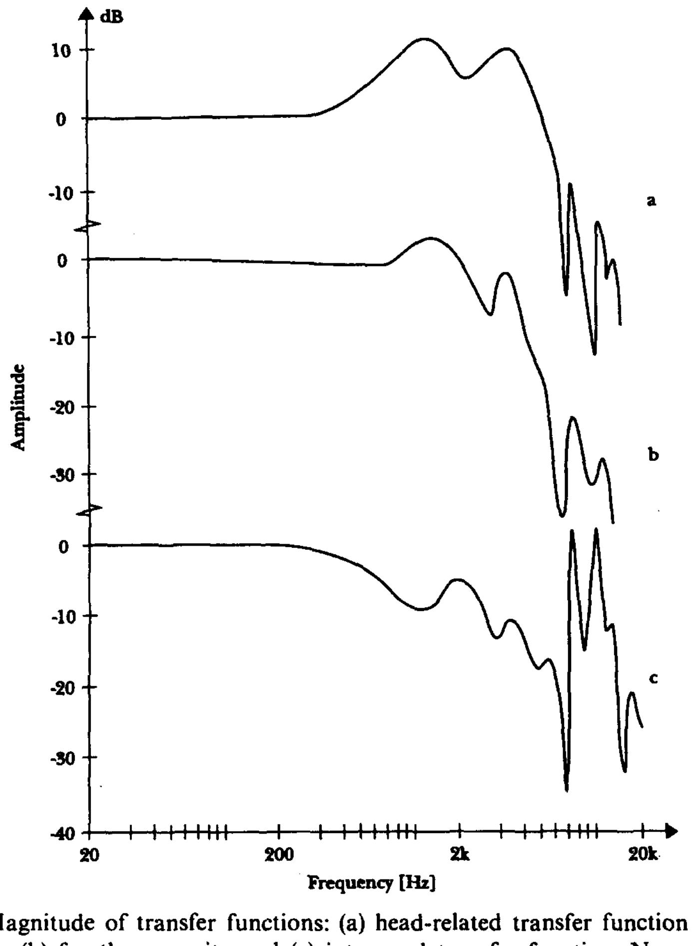

Fig. 11. Magnitude of transfer functions: (a) head-related transfer function for nearest loudspeaker, (b) for the opposite and (c) interaural transfer function. Neumann KU80, $\phi = 45°$.

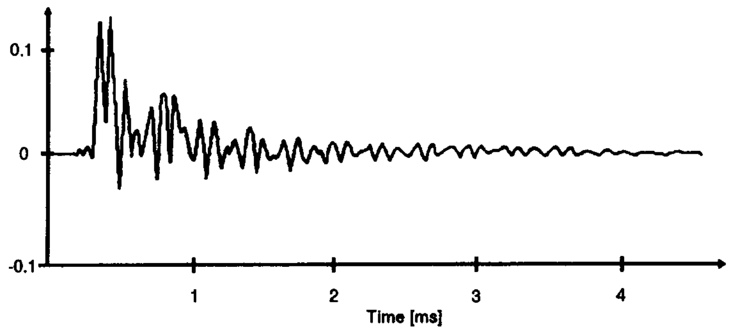

Fig. 12. Interaural impulse response. Neumann KU80, $\phi = 45°$, $\theta = 0°$.

in Fig. 12.[63] It is seen that the impulse response is only a few milliseconds long, and a finite impulse response filter is a possible means of implementation.

A crosstalk cancellation system has been in use at our laboratory for some years, and commercial systems are also available. Basically, they only work in a free field, which means that an anechoic room is needed. The position of the listener's head is also somewhat critical. Systems that account properly for reflections are still at the experimental stage.

7 COMPUTER SYNTHESIS OF BINAURAL SIGNALS

In the previous sections it has been assumed that the binaural signals originate in a recording of an acoustical event that takes place in real life. The binaural technique is used to transmit the event to listeners who are not present at the location, or to store it for later reproduction.

It is also possible to synthesize binaural signals on a digital computer and thus simulate that a sound has been played in a room and has been recorded by an artificial head. The room need not even exist in real life. Inputs to this kind of simulation are:

(1) Information on the sound transmission from source to listening point in the room.
(2) Information on the head-related transfer functions of the head being simulated.
(3) A recording of a 'dry' sound source, for example made in an anechoic room.

The output is the two channels of a binaural signal. How this is created will be described in principle in the following.

The information on the sound transmission is computed in a program of the ray-tracing or the mirror-image type.[77–81] For each of the transmission paths from sound source to listener—including the direct path—the impulse response of the transmission must be given together with information on the angle of incidence at the listener's position. For transmission path i, the information may be given in the following form

$$r_i, \phi_i, \theta_i, a_i(t) \tag{84}$$

where r_i is the distance traversed by the sound wave, ϕ_i and θ_i specify the angle of incidence to the listener and $a_i(t)$ is the impulse response of the transmission path. Here, $a_i(t)$ incorporates the delay corresponding to the propagation length, and includes the distance attenuation and filtering due to the reflective properties of the surfaces that the signal has passed. It also includes a possible non-uniform frequency response of the source in the direction of the transmitted wave, and the air damping.

Here, i starts with number 0 for the direct sound path and continues—at least theoretically—to infinity. In practice, a finite number of transmission paths are calculated, and in the following N is used as the upper limit.

The *head-related impulse responses* (HIRs) from the free field (p_1) to each of the two recording points (p_2, p_3 or p_4) are described as

$$\begin{aligned} &b_{\mathrm{left},r_i,\phi_i,\theta_i}(t) \\ &b_{\mathrm{right},r_i,\phi_i,\theta_i}(t) \end{aligned} \tag{85}$$

The contribution of each of the transmission paths to the transmission from the simulated sound source to the simulated recording points is found by convolution

$$\begin{aligned} &a_i(t) * b_{\mathrm{left},r_i,\phi_i,\theta_i}(t) \\ &a_i(t) * b_{\mathrm{right},r_i,\phi_i,\theta_i}(t) \end{aligned} \tag{86}$$

From this it is possible to calculate the total transmission from sound source to the simulated recording points

$$\begin{aligned} h_{\mathrm{left}}(t) &= \sum_{i=0}^{N} a_i(t) * b_{\mathrm{left},r_i,\phi_i,\theta_i}(t) \\ h_{\mathrm{right}}(t) &= \sum_{i=0}^{N} a_i(t) * b_{\mathrm{right},r_i,\phi_i,\theta_i}(t) \end{aligned} \tag{87}$$

This pair of functions is called the *binaural room impulse response.*

The binaural signals $e_{\mathrm{left}}(t)$ and $e_{\mathrm{right}}(t)$ (*voltage*) are obtained by convolution of the recorded dry sound signal $s(t)$ (*voltage*) with the binaural room impulse responses.

$$e_{\text{left}}(t) = h_{\text{left}}(t) * s(t) \qquad (88)$$
$$e_{\text{right}}(t) = h_{\text{right}}(t) * s(t)$$

If the room exists in real life, the binaural room impulse response in eqn (87) can be found by measurement rather than by calculation.

For implementation in a digital system, impulse responses $a(t)$, $b(t)$ and $h(t)$ should be replaced by sequences $a(n)$, $b(n)$, $h(n)$ that are samples from the impulse responses, and the signal $s(t)$ should be replaced by $s(n)$ that are samples from the signal.

8 APPLICATIONS OF BINAURAL TECHNOLOGY

Let us assume that the equalizing, transmission, etc., have been done properly, and that the artificial recording head is a true copy of the listener's own head. Then it is characteristic of the binaural recording technique that the sound presented to the listener's eardrums is exactly the same as would have been found, if the listener had been present during the recording. This makes the binaural recording technique superior to other recording techniques. It gives the most valid representation of the original sound, not only with respect to timbre, but also in relation to spatial aspects.

The applications are divided into (1) applications that involve recording with an artificial head, (2) applications that use computer-simulated binaural signals and (3) others.

8.1 Recording with an artificial head

The most obvious application of the binaural technique is the mere recording of an acoustical event, i.e. a concert or a radio drama, through an artificial head and playback through headphones. This is the normal use in connection with recording and broadcasting. And, of course, there are many possibilities for use in film and video production, not least connected to 'special effects'. But in addition to that, there are many other possibilities for utilizing the binaural technique itself and the know-how associated with it.

Noise evaluation

For noise evaluation it is now well known that A-weighted sound levels or other objective measures are often insufficient to characterize a noise. Subjective assessments have become more common, although they are difficult to carry out. Normally a number of subjects must listen to many kinds of noise, and some psychometric scaling technique is used. It is very impractical to move the subjects around to the real noise sources, so a recording of the noise is usually played back in a laboratory. A simple playback of an ordinary recording through loudspeakers does not give the subjects the feeling of 'being in the noise'.

With the binaural recording technique, the correct sound impression, including directional qualities, can be presented in the laboratory. Thus, many sounds can be presented to subjects in an easy way, and the psychological assessment scales can be properly calibrated. It is also possible to edit or filter the recording, so that the sound represents the noise after some kind of noise reduction. Different means of noise abatement may then be assessed without carrying it out physically.

Comparison of concert halls

It is normally quite difficult to compare the acoustics of different concert halls, because the human memory for sound impressions is very short—and not long enough to get from one hall to another. Binaural recordings made in different halls—if possible with the same orchestra—make it possible to make direct comparisons between two halls.

Teleconferences and recording of meetings

Normally, participants in meetings have no problems in listening to one particular speaker, even when several people speak at the same time. This is due to the well known *cocktail-party effect.* If a meeting is recorded or transmitted with traditional technique, this effect is lost, and it may be difficult to hear what a particular speaker says. Binaural transmission preserves the cocktail party effect and re-establishes our own 'selectivity tool'. This is useful for transmission of meetings and discussions. It may also prove useful for a secretary, when reporting a recorded meeting.

Assessment of speech in rooms

Several new techniques have been developed for objective characterization of the intelligibility and quality of speech in rooms. Binaural recordings will constitute a means for subjective evaluation.

Assessment of room impact on loudspeaker sound

The quality of the sound from a loudspeaker is very much influenced by the room in which the loudspeaker is used. When evaluating a loudspeaker, it is therefore necessary to listen to it in many rooms and include many positions, not only of the loudspeaker but also of the listener. The practical arrangements involved, combined with the short time humans can remember a sound impression, make such an evaluation almost impossible.

Binaural recordings made of the loudspeaker in different rooms and with different loudspeaker and listener positions may solve the problems. It makes it easy to present the sound to many people—and it is possible to change room and position just by means of a switch.

Control of public address systems
In theaters and during concerts with electronic amplification, the mixing console must be adjusted throughout the performance. The console may be located in a control room or in the concert hall. The control room location has the drawback that the operator does not properly hear the sound that he is adjusting. Setting up the mixer in the concert hall has an economic consequence—the operator and the equipment occupy several of the best seats. This leads to a loss of income that is often substantial, and the operators working place is inconveniently 'public'. An artificial head in the hall and transmission of the signals to the control room may solve the problem.

8.2 Simulated binaural signals—artificial environment

The operation of the head—artificial or human—is to transfer sound from a given direction and distance into two signals. If direction and distance to the sound source are known, and the transfer functions for this direction and distance to both signals are known, then the transfer functions can be simulated with electronic filters. The filters may be implemented in analogue or digital circuitry. With sufficient knowledge about the transfer functions in various directions and distances, a sound source can electronically be put anywhere in space. A very important application of binaural technology is therefore within the area of *artificial environment*. The simulation process was described in Section 7.

Binaural mixing console
A binaural mixing console is an electronic device that contains filters which—on the basis of a dry monophonic signal—creates a binaural signal, corresponding to a given direction and distance. It should be possible to adjust direction and distance while listening to the sound, just as with a pan-potmeter in a normal mixing console. Of course, it should be possible to mix more sound sources at the same time and in different positions.

The binaural mixing console can be improved and include also some of the first and most important reflections. The reflections should be delayed according to the geometrical conditions from a typical concert hall, and they should be filtered according to their correct angle of incidence. The mixer may also add some diffuse reverberation.

Binaural room simulator

If the concept of a binaural mixer is developed even further, it turns into a binaural room simulator. In this device, the filters that are introduced in the path from the dry monophonic source signal have the exact transfer functions from a source position to each of the two ears of a listener in a *specific position* and in a *specific hall.* These transfer functions are called *binaural room impulse responses* (be sure not to confuse this with *room impulse responses,* which refer to the transmission from source to sound pressure at a given point—without the head). With a binaural room simulator it is possible to listen to the same sound source playing in different concert halls, and to compare different source and listener positions in the same hall.

The distinction between a binaural mixing console and a binaural room simulator is not precise. If a mixer includes reflections and reverberation, it approaches a room simulator, and if a room simulator allows more than one source position at a time, it turns into a mixing console. Characteristics of a mixer are the many channels and the possibility of easy and on-line change of source positions. A room simulator is characterized by an accurate simulation of a specific room and a less easy access to changes in the room (or to change to another room for comparison).

Tool in acoustic design

The binaural room impulse responses, which are used in the binaural room simulator, can be found by measurement. They can also be found by calculation in a room acoustics simulation program of the ray-tracing or mirror-image type. If such a program is used, it is possible to listen to a concert hall before it is built. It is possible to subjectively assess the sound with different room shapes and different surfaces, and it is possible to compare it with existing halls.

It is also possible to evaluate the room acoustics simulation programs. If the program is used for an existing hall, convolution of a dry source signal is possible with the computed binaural impulse room response as well as the measured binaural room impulse response. In this way, subjective comparisons can be made.

For use in a room simulator, exact impulse responses must be known, and everything must be based on time signals including phase information from every single signal path. But normally, room-acoustics simulation programs calculate room-acoustic parameters such as reverberation time, early decay time, reflectograms, etc., on an energy basis. The strategy of the programs must therefore be changed slightly before the programs can be used to create inputs to a binaural room simulator.

In this connection, it is a complication that the reflective properties of materials are normally known from the energy absorption. It may be necessary to re-measure materials by methods that give the correct information either directly in the time domain as a *reflection impulse response* or—in the frequency domain—as the reflection in amplitude and phase as a function of frequency. Methods that offer this are various gating and impulse averaging techniques, maximum length sequence (MLS) technique and time delay spectrometry (TDS).

Simulated loudspeaker reproduction

When traditional stereo signals, such as X/Y and A/B signals, are reproduced through headphones, an in-the-head localization results. This is due to the lack of crosstalk and the lack of reflections from a listening room. Based on knowledge about the transfer functions to the opposite ear in a free field, crosstalk can be added to the original stereo signal. Also, reflections corresponding to a normal listening room can be added. At present it is not known, precisely how much is needed to get a correct out-of-the-head localization.

This application may be relevant for the consumer market, but it is especially attractive for control-room simulation. In all areas of sound production, large and expensive control rooms are used. And quite often the producer wants to listen to the recording in various surroundings and using various equipment, since he does not know what equipment will be used by the consumer. The binaural technique may compete with the control rooms or be a supplement to them.

Simulation of public address systems

A combination of a room acoustic simulator and a simulator of loudspeaker reproduction can be used to assess public address systems for use in auditoriums, theaters, airports, railway stations, at stadiums, etc. The value of this at the projection stage is obvious.

8.3 Other applications

Binaural hearing aids

It is a well-known fact that hearing-impaired persons have great difficulty following conversations with more than a few people involved. If they suffer from a conductive hearing loss, it is theoretically possible to correct their hearing completely with a hearing aid. If the hearing aid picks up sound at a point that includes full directional information and the correct equalization is made, then the directional hearing is re-established, and the cocktail-party effect is expected to work.

A practical way of making such a hearing aid is to plug the ear canal and put the microphone outside the plug. The position of the microphone should be so close to the entrance of the ear canal that the sound propagation from that point is one-dimensional. The electronics and the earphone can be included in the plug. Such a device could be denoted a *binaural hearing aid* (also if one ear is normal and only one hearing aid is used).

It is not clear whether a binaural hearing aid will improve directional hearing in the case of an non-linear, neural hearing loss with recruitment.

Binaural noise suppression

It has already been mentioned that the hearing is able to extract information on the signal from a specific sound source, even when other and possibly louder sound sources are present. The effect is called the cocktail-party effect and relies on neural processing in the auditory system. If a proper model of this is described, it can be implemented in a computer, which can perform noise suppression on the basis of signals recorded with an artificial head.

9 CONCLUDING REMARKS

The present article has described the basic theory of binaural recording and playback as well as the theory of computer generation of binaural signals. Also a number of topical and prospected applications have been described.

It has been shown that any point in the ear canal—and possibly even a point a few millimeters outside—can be used for recording, since sound pressure here includes the full spatial information given to the ear. It is also acceptable to record outside a blocked ear canal.

It is shown that correct overall transmission in a binaural system can be guaranteed if an electronic equalizing filter is introduced between the recording head and the headphone. The filter should equalize the characteristic of the recording microphone and the headphone transfer function measured at the point in the ear canal, where the recording is made.

The equalizing filter should include extra terms, if recording is made outside a blocked ear canal, or if the ear canal of the head used for recording and calibration is different from that of the listener. The extra terms are not required, when an open headphone is used for reproduction.

The description has concentrated on 'how it should work'. It should not be forgotten, though, that at present the binaural technique still suffers from some problems. Most obvious is the poor frontal localization, but also problems about standardization of recording points, equilization, compatibility, interface to room simulations, significance of individual variations and other things deserve still some attention.

However, binaural technology has greatly improved since it was invented.

Many qualified research groups are working on it, and the range of applications has broadened so much that the final breakthrough of the technique must be very near.

ACKNOWLEDGEMENTS

The author wants to thank his colleagues for their advice and encouragement. Of special value were a number of discussions with Dorte Hammershøi, her help in the verification of the models and her fruitful criticism during the preparation of this article.

REFERENCES

1. Firestone, F. A., The phase difference and amplitude ratio at the ears due to a source of pure tones. *J. Acoust. Soc. Amer.*, **2** (1930) 260–70.
2 Eichhorst, O., Zur Frühgeschichte der stereophonischen Übertragung. *Frequenz*, **13** (1959) 273–7.
3. Nordlund, B. & Liden, G., An artificial head. *Acta Oto-laryngol.*, **56** (1963) 493–9.
4. Schirmer, W., Die Veränderung der Wahrnehmbarkeitsschwelle eines künstlichen Rückwurfes bei kopfbezüglicher stereophoner Übertragung. *Hochfrequenztech. u. Elektroakustik*, **75** (1966) 115–23.
5. Schirmer, W., Die Unterscheidbarkeit von Hörerplätzen mittels kopfbezüglicher stereophoner und monophoner Übertragung. *Hochfrequenztech. u. Elektroakustik*, **75** (1966) 181–4.
6. Torick, E. L., di Mattia, A., Rosenheck, A. J., Abbagnaro, L. A. & Bauer, B. B., An electric dummy for acoustical testing. *J. Audio Engng Soc.*, **16**(4) (1968) 397–403.
7. Damaske, P. & Wagener, B., Richtungshörversuche über einen nachgebildeten Kopf. *Acustica*, **21** (1969) 30–5.
8. Kürer, R., Plenge, G. & Wilkens, H., Correct spatial sound perception rendered by a special 2-channel recording method. Paper presented at Audio Engng Soc., New York, NY, 1969.
9. Wilkens, H., Subjektive Ermittlung der Richtcharakteristik des Kopfes und einer kopfbezogenen Aufname und Wiedergabeanordnung. In *Gemeinschaftstagung für Akustik und Schwingungstechnik, Berlin 1970*, ed. T. Tarnoczy. VDI-Verlag, Düsseldorf, Germany, 1971, pp. 407–10.
10. Wilkens, H., Beurteilung von Raumeindrücken verschiedener Hörerplätze mittels kopfbezogener Stereophonie. In *Proc., 7th Int. Congr. on Acoustics, Budapest*. Akademiai Kiado, Budapest, Hungary, 1971, 24 S 5.
11. Wilkens, H., Kopfbezügliche Stereophone, ein Hilfsmittel für Vergleich und Beurteilung verschiedener Raumeindrücke. *Acustica*, **26** (1972) 213–21.
12. Mellert, V., Construction of a dummy head after new measurements of the threshold of hearing. *J. Acoust. Soc. Amer.*, **51** (1972) 1359–61.
13. Blauert, J., *Räumliches Hören*. S. Hirzel Verlag, Stuttgart, Germany, 1974.

14. Blauert, J., Vergleich unterschiedlicher Systeme zur originalgetreuen elektroakustischen Übertragung. *Rundfunktech. Mitt.*, **18** (1974) 222–7.
15. Kuhl, W. & Plantz, R., Kopfbezogene Stereophonie und andere Arten der Schallübertragung im Vergleich mit dem natürlichen Hören. *Rundfunktech. Mitt.*, **19** (1975) 120–32.
16. Platte, H.-J., Laws, P. & vom Hövel, H., Anordnung zur genauen Reproduktion von Ohrsignalen. In *Fortschritte der Akustik, DAGA '75, Weinheim.* Physik Verlag, Weinheim, Germany, 1975, pp. 361–4.
17. Laws, P. & Platte, H.-J., Spezielle Experimente zur kopfbezogenen Stereophonie. In *Fortschritte der Akustik. DAGA '75, Weinheim.* Physik Verlag, Weinheim, Germany, 1975, pp. 365–8.
18. Mellert, V., Verbesserte Schallfeldabbildung mit einem neuen Kunstkopf. In *Fortschritte der Akustik, DAGA '75, Weinheim.* Physik Verlag, Weinheim, Germany, 1975, pp. 433–6.
19. Platte, H.-J. & Laws, P., Die Vorne-Ortung bei der kopfbezogenen Stereophonie. *Radio Mentor Electronic*, **42** (1976) 97–100.
20. Laws, P., Blauert, J. & Platte, H.-J., Anmerkungen zur stereophonen, kopfbezogenen Übertragungstechnik. *Acustica*, **36** (1976/77) 45–7.
21. Laws, P. & Platte, H.-J., Ein spezielles Konzept zur Realisierung eines Kunstkopfes für die kopfbezogene stereophone Aufnahmetechnik. *Rundfunktech. Mitt.*, **22** (1978) 28–31.
22. Gottlob, D., Anwendung der kopfbezogenen Stereophonie in der akustischen Forschung. *Rundfunktech. Mitt.*, **22** (1978) 214–16.
23. Kleiner, M., Problems in the design and use of 'dummy heads'. *Acustica*, **41** (1978) 183–93.
24. Blauert, J., Mellert, V., Platte, H.-J., Laws, P., Hudde, H., Scherer, P., Poulsen, T., Gottlob, D. & Plenge, G., Wissenschaftliche Grundlagen der kopfbezogenen Stereophonie. *Rundfunktech. Mitt.*, **22** (1978) 195–218.
25. Fukudome, K., The Thevenin acoustic impedance and pressure of dummy heads. In *Proc. 10th Int. Congr. on Acoustics, Sydney.* Publications Unit, Sydney, Australia, 1980.
26. Schöne, P., Der Signalstörabstand bei Kunstköpfen. In *Fortschritte der Akustik, DAGA '80*, ed. E. Zwicker. VDE-Verlag, Berlin, Germany, 1980, pp. 835–8.
27. Schöne, P., Zur Nutzung des Realisierungspielraumes in der kopfbezogenen Stereofonie. *Rundfunktech. Mitt.*, **24** (1980) 1–11.
28. Hudde, H. & Schröter, J., The equalization of artificial heads without exact replication of the eardrum impedance. *Acustica*, **44** (1980) 302–7.
29. Hudde, H. & Schröter, J., Verbesserungen am Neumann-Kunstkopfsystem *Rundfunktech. Mitt.*, **25** (1981) 1–6.
30. Wollher, H., Mikrophonankopplung an das Aussenohr eines neuen Kunstkopfes. *Rundfunktech. Mitt.*, **25** (1981) 141–5.
31. Genuit, K., Untersuchungen zum Einsatz eines neuartigen rauscharmen zweikanaligen Aufnahmesystems mit hoher Dynamik im Rundfunkbereich. In *Proc. 6th NTG Hörrundfunktagung (Mannheim).* VDE-Verlag, Berlin, Germany, 1982, pp. 247–54.
32. Blauert, J., *Spatial Hearing.* The MIT Press, Cambridge, MA, 1983.
33. Griesinger, D., Equalization and spatial equalization of dummy-head recordings for loudspeaker reproduction. *J. Audio Engng Soc.*, **37**(1/2) (1989) 20–8.
34. Vorländer, M., Freifeld- und Diffusfeld-Übertragungsmasse von natürlichen Köpfen und von Kunstköpfen. *Acustica*, **74** (1991) 192–200.

35. Hammershøi, D., Møller, H., Sørensen, M. F. & Larsen, K. A., Head-related transfer functions: Measurements on 40 human subjects. Presented at the 92nd Convention of the Audio Engineering Society, Vienna, Austria, 1992.
36. Møller, H., Hammershøi, D., Hundebøll, J. V. & Jensen, C. B., Transfer characteristics of headphones: Measurements on 40 human subjects. Presented at the 92nd Convention of the Audio Engineering Society, Vienna, Austria, 1992.
37. Hammershøi, D. & Møller, H., Artificial heads for free field recording; how well do they simulate real heads? To be presented at the 14th International Congress on Acoustics, Beijing, China, 1992.
38. Weber, R. & Mellert, V., Ein Kunstkopf mit ebenem Frequenzgang. In *Forschritte der Akustik, DAGA '78*, ed. J. Blauert & H. Baule. VDE-Verlag, Berlin, Germany, 1978, pp. 645–8.
39. Killion, M. C., Equalization filter for eardrum-pressure recording using a KEMAR manikin. *J. Audio Eng. Soc.*, **27**(1–2) (1979) 13–16.
40. Tsujimoto, K., Equalization in artificial-head recording for loudspeaker reproduction. In *Proc. 10th Int. Congr. on Acoustics, Sydney*. The Publications Unit, Sydney, Australia, 1980.
41. Gotoh, T., Kimura, Y. & Sakamoto, N., A proposal of normalization for binaural recording. 70th Convention, Audio Engng, New York, 1981, preprint 1811 (B-2).
42. (a) Theile, G., Zur Kompatibilität von Kunstkopfsignalen mit intensitätsstereophonen Signalen bei Lautsprecherwiedergabe: Die Richtungabbildung. *Rundfunktech. Mitt.*, **25** (1981) 67–73. (b) Theile, G., Zur Kompatabilität von Kunstkopfsignalen mit intensitätsstereophonen Signalen bei Lautsprecherwiedergabe: Die Klangfarbe. *Rundfunktech. Mitt.*, **25** (1981) 146–54.
43. Theile, G., Zur Theorie der optimalen Wiedergabe von stereophonen Signalen über Lautsprecher und Kopfhörer. *Rundfunktech. Mitt.*, **25** (1981) 155–69.
44. Russotti, J. S., Santoro, T. P. & Haskell, G. B., Proposed technique for earphone calibration. *J. Audio Eng. Soc.*, **36**(9) (1985) 643–50.
45. Gierlich, H. W. & Genuit, K., Processing artificial-head recordings. *Features, AES*, **37**(1/2) (1989) 34–9.
46. Butler, R. A. & Belendiuk, K., Spectral cues in the localization of sound in the median sagittal plane. *J. Accoust. Soc. Amer.*, **61** (1977) 1264–9.
47. Morimoto, M. & Ando, Y., Localization in the median plane of sound sources simulated by a digital computer. In *Proc. 8th Int. Congr. on Acoustics*, ed. A. Lara-Saenz, Vol. 1. Socidad Espanola de Acustica, Madrid, Spain, 1977, p. 371.
48. Morimoto, M. & Ando, Y., Simulation of sound localization. Localization of sound: Theory and application, ed. R. W. Gatehouse. The Amphora Press, Groton, CT, 1982, 85–9.
49. Boerger, G., Laws, P. & Blauert, J., Stereophone Kopfhörerwidergabe mit Steuerung bestimmter Übertragungsfaktoren durch Kopfdrehbewegungen. *Acustica*, **39** (1977) 22–6.
50. Blauert, J., Untersuchungen zum Richtungshören in der Medianebene bei fixiertem. Kopf. Doctoral Dissertation, Technische Hochschule, Aachen, Germany, 1969.
51. Nielsen, S. H., *Distance perception in hearing*. Aalborg University Press, ISBN: 87 7307 447-0, 1991.
52. Bauer, B. B., Stereophonic earphones and binaural loudspeakers. *J. Audio Engng Soc.*, **9**(2) (1961) 148–51.

53. Atal, B. S. & Schroeder, M. R. Nachahmung der Raumakustik durch Elektronenrechner. *Gravesaner Blätter*, **27/28** (1966) 124–37.
54. Damaske, P. & Mellert, V., Ein Verfahren zur richtungstreuen Schallabbildung des oberen Halbraumes über zwei Lautsprecher. *Acustica*, **22** (1969/70) 153–62.
55. Damaske, P. & Mellert, V., Zur richtungstreuen stereophonen Zweikanalübertragung. *Acustica*, **24** (1971) 222–5.
56. Damaske, P., Head-related two-channel stereophony with loudspeaker reproduction. *J. Acoust. Soc. Amer.*, **50** (1971) 1109–15.
57. Damaske, P., Richtungstreue Schallabbildung über zwei Lautsprecher. In *Gemeinschaftstagung für Akustik und Schwingungstechnik, DAGA '71, Berlin.* VDI-Verlag, Düsseldorf, Germany, 1971, pp. 403–6.
58. Wilkens, H., Plenge, G. & Kürer, R., Wiedergabe von kopfbezogenen stereophonen Signalen durch Lautsprecher. Paper presented at Convention '71, Audio Eng. Soc., Cologne, Germany, 1971.
59. Sakamoto, N., Gotoh, T., Kogure, T., Shimbo, M. & Clegg, A., On the advanced stereophonic reproducing system 'ambience stereo', In *60th Convention, Audio Eng. Soc., Los Angeles,* preprint 1361 (G-3). Audio Engineering Society, New York, NY, 1978.
60. Schöne, P., Ein Beitrag zur Kompabilität raumbezogener und kopfbezogener Stereophonie. *Acustica*, **47** (1981) 170–5.
61. Cooper, D. H. & Bauck, J. L., Prospects for transaural recording. *J. Audio Engng Soc.*, **37**(1/2) (1989) 3–19.
62. Griesinger, D., Theory and design of a digital audio signal processor for home use. *Features, AES*, **37**(1/2) (1989) 42–50.
63. Møller, H., Reproduction of artificial-head recordings through loudspeakers. *J. Audio Engng Soc.*, **37**(1/2) (1989) 30–3.
64. Goerlich, H. W. & Genuit, K., Aufbau und Anwendung eines elektronischen Kunstkopfes. In *Proc. 13th Tonmeistertagung (Munich).* 1984, pp. 103–10.
65. Gierlich, H. W. & Genuit, K., Entwurf eines mikroprocessorgesteurten Aussenohrsimulators. In *Fortschritte der Akustik DAGA '84, Darmstadt*, ed. J. Franz. DPG, Bad Honnef, Germany, 1984, pp. 671–4.
66. Foster, S. H., Wenzel, E. M. & Taylor, R. M., Real time synthesis of complex acoustic environments. In *Proc. IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics, Mohonk, NY*, ed. P. M. Peterson. Bell Communications Research, Morristown, NJ, 1991.
67. Xiang, N. & Blauert, J., Binaural auralization by using hybrid scaled-down modelling. Paper presented at Int. Symp. on Computer Modelling and Prediction of Objective and Subjective Properties of Sound Fields in Rooms, Aug. 1991, Copenhagen, Denmark, and Gothenburg, Sweden.
68. Lehnert, H. & Blauert, J., Virtual Auditory Environment. In *5th Int. Conf. on Advanced Robotics (IEEE/ICAR), Pisa, Italy*, 1991, pp. 211–16.
69. Hammershøi, D. & Møller, H., Free-field sound transmission to the external ear; a model and some measurements. In *Fortschritte der Akustik, DAGA '91, Bochum*, ed. J. Blauert & E. Paulus. DPG, Bad Honnef, Germany, 1991, pp. 473–6.
70. Mehrgardt, S. & Mellert, V., Transformation characteristics of the external human ear. *J. Acoust. Soc. Amer.*, **61**(6) (1977) 1567–76.
71. Middlebrooks, J. C., Makous, J. C. & Green, D. M., Directional sensitivity of sound-pressure levels in the human ear canal. *J. Acoust. Soc. Amer.*, **86**(1) (1989) 89–108.

72. Schirmer, W., Zur Deutung der Übertragungsfehler bei kopfbezüglicher Stereophonie. *Acustica*, **18** (1966) 228–33.
73. Toole, F. E., In-head localization of acoustic images. *J. Acoust. Soc. Amer.*, **48** (1970) 943–9.
74. Plenge, G., Ein Beitrag zur Erklärung der Im-Kopf-Lokalisation. In *Gemeinschaftstagung für Akustik und Schwingungstechnik, Berlin 1970*, VDI-Verlag, Düsseldorf, Germany, 1970, pp. 411–16.
75. Plenge, G., Über das Problem der Im-Kopf-Lokalisation. *Acustica*, **26** (1972) 241–52.
76. Rubak, P., Headphone signal processing system for out-of-the head localization. *Audio Engng Soc., Paris*, Preprint 3063 (K-5), 1991.
77. Vian, J. P. & van Maercke, D., Calculation of the room impulse response using a ray-tracing method. In *Proc. ICA Symp. on Acoustics and Theatre Planning for the Performing Arts, Vancouver*. Canadian Acoustical Association, Ottawa, Canada, 1986, pp. 74–8.
78. Vorländer, M., Simulation of the transient and steady state sound propagation in rooms using a new combined ray tracing/image-source algorithm. *J. Acoust. Soc. Amer.*, **86** (1989) 172–8.
79. Naylor, G. M., ODEON—Another hybrid room acoustical model. Paper presented at Int. Symp. Computer Modelling and Prediction of Objective and Subjective Properties of Sound Fields in Rooms, Copenhagen, Denmark, and Gothenburg, Sweden, August 1991.
80. Lehnert, H., Strahlverfolgungsverfahren (Ray-Tracing) mit punktförmigen Quellen und Empfängern sowie idealen Strahlen. In *Fortschritte der Akustik, DAGA '91, Bochum*, ed. J. Blauert & E. Paulus. DPG, Bad Honnef, Germany, 1991, pp. 633–6.
81. Lehnert, H., Berechnung von langen Raumimpulsantworten durch ein schnelles Strahlverfolgungsverfahren (Ray-Tracing). In *Fortschritte der Akustik, DAGA '91, Bochum*, ed. J. Blauert & E. Paulus. DPG, Bad Honnef, Germany, 1991, pp. 637–40.
82. Fukudome, K., Equalization for the dummy-head–headphone system capable of reproducing true directional information. *J. Acoust. Soc. Japan (E)*, **1** (1980) 59–67.
83. Mellert, V., Die Normung Kopfkezogener Stereo-aufnahmen und ihre Wiedergabe über Lautsprecher. *Fernseh-Kinotech.*, **30** (1976) 86–8.

APPENDIX

In Section 2.2, difficulties were mentioned in appreciating that, for instance, a pressure ratio can include reflections in the ear canal. The reason is that the impedances are frequency-dependent variables, given in amplitude and phase for a wide frequency range. This is in contrast to traditional transmission-line calculations, where normally steady-state conditions at a specific frequency are assumed. This appendix shows an example of calculations on a transmission line model.

The ear canal is modelled by a transmission line with a characteristic

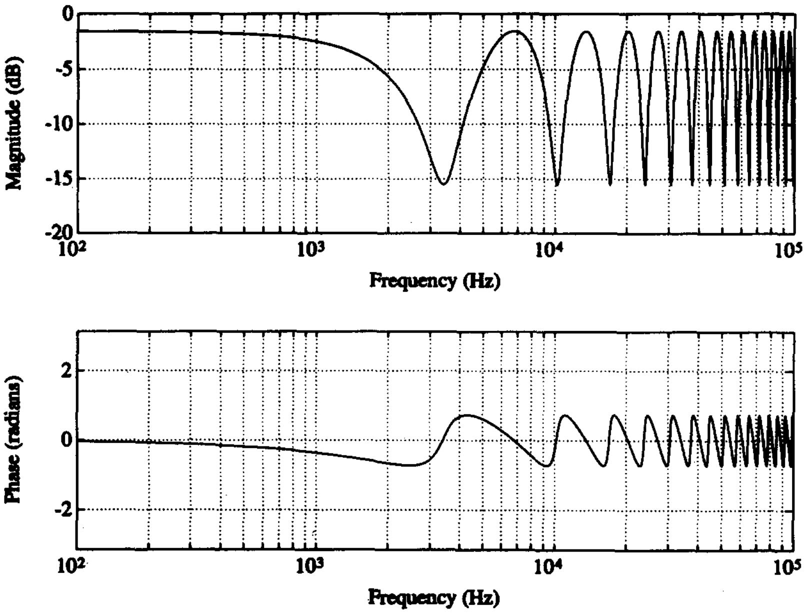

Fig. A1. $[P_3/P_2]$ calculated for values given in eqn (A4).

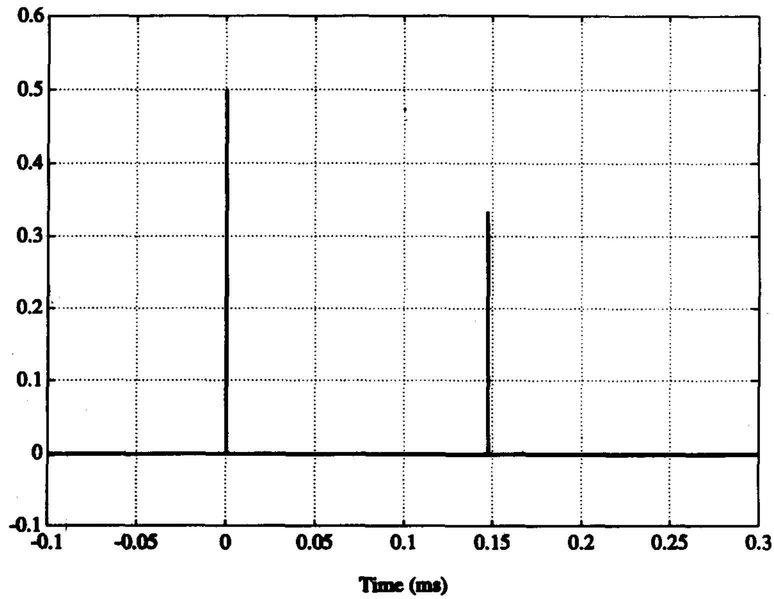

Fig. A2. Impulse response p_3/p_2, calculated as the inverse Fourier transform of $[P_3/P_2]$, Fig. A1. Values inserted from eqn (A4).

impedance Z_0, a propagation constant γ and a length l; the speed of sound is c. From traditional transmission line theory, it is known that

$$Z_{\text{ear canal}} = Z_0 \cdot \frac{1+\rho_{\text{L}}\,\mathrm{e}^{-2\gamma l}}{1-\rho_{\text{L}}\,\mathrm{e}^{-2\gamma l}} \tag{A1}$$

where

$$\rho_{\text{L}} = \frac{Z_{\text{eardrum}} - Z_0}{Z_{\text{eardrum}} + Z_0} \tag{A2}$$

The pressure division at the entrance to the ear canal can be found (eqn (3)) as

$$[P_3/P_2] = \frac{Z_{\text{ear canal}}}{Z_{\text{ear canal}} + Z_{\text{radiation}}} \tag{A3}$$

The following are inserted (the values serve as an example and not as an attempt to suggest real values):

$$\begin{aligned} Z_{\text{radiation}} &= Z_0 \\ Z_{\text{eardrum}} &= 5 \cdot Z_0 \\ \gamma &= j \cdot \frac{2\pi \cdot f}{c} \\ c &= 340\,\text{m/s} \\ l &= 25 \cdot 10^{-3}\,\text{m} \end{aligned} \tag{A4}$$

The result of a calculation of eqn (A4) for a wide frequency range is shown in Fig. A1.

In this case, a $\frac{2}{3}$ reflection is expected at the eardrum, while the reflected wave—as a consequence of the impedance match with the radiation impedance—is not reflected again at the entrance. This is exactly what is seen, if the transfer function of Fig. A1 is inverse Fourier transformed (see Fig. A2).