On the Relation between Directional Bands and Head Movements

HAN, H.L.;
Delft University of Technology, Delft, The Netherlands

[3PS1.09]
Preprint 3293

Presented at the 92nd Convention 1992 March 24–27 Vienna

AES

AN AUDIO ENGINEERING SOCIETY PREPRINT

On the Relation between Directional Bands and Head Movements

H.L. Han

Faculty of Applied Physics
DELFT UNIVERSITY OF TECHNOLOGY
P.O. Box 5046
2600GA Delft
The Netherlands

ABSTRACT

Above 3 kHz elevation and azimuth of sound sources are cued by steep spectral slopes due to pinna reflections. Other cues in this range, that provide front-back discrimination, are more in agreement with Blauert's idea of boosted bands.
Through inspection of head-related transfer functions it was found that there are azimuth-dependent level variations in the two directional bands below 3 kHz. Having only interaural level differences (ILDs), would reliable cuing be possible? This paper shows how ILD, ITD, head position vector and memory could interact to resolve front-back ambiguities. A simple experiment supplies evidence.
The new insights in conjunction with some other observations has led to a better understanding of front-back reversals that many listeners hear during dummy-head reproduction.

0 INTRODUCTION

Models of the human auditory system that are based on interaural differences have limited usefulness in understanding localization in 3-dimensional space. They are relevant only to listening to synthetic signals or ordinary stereo-recordings via headphones. In these situations sound sources are localized in the head on the interaural axis. This is called lateralization. How is localization achieved along the two other dimensions?

Although psychophysical experiments showed that the pinnae were essential for localization in the vertical sense, it was not understood how they encode the directional information. In order to shed more light in this matter, the author measured head-related impulse responses and transfer functions (HRTFs) of a KEMAR dummy head. This was the subject of a paper presented at the 90th AES Convention <1>. Its major conclusions can be summarized as follows.

1. The pinna encodes azimuth and elevation of a sound source in the spectrum of the first millisecond of the arriving sounds.
2. The most plausible way of extracting directional information from the internal spectrum is through feature detectors <2>.
3. There is a clear connection between pinna-based spectral features and some of Blauert's directional bands <3>.

As all directional bands except one deal with front-back discrimination, conclusion 3 implies that the pinnae also aid in localizing along the front-back dimension. In view of this observation it is natural to investigate how front-back cuing is accomplished below the working range of the pinnae. The present paper deals specifically with the two directional bands below 3 kHz.

Section 1 reviews prior work with special attention to the high-frequency directional bands. HRTFs are again examined in section 2, but now with the object of finding systematic variations in the lower-frequency directional bands. The question then arises how front-back discrimination is possible without a priori knowledge of the spectrum of the source. It will be shown that head movements are essential for unambiguous front-back discrimination. Some simple experiments using octave noise in a directional band, described in section 3, not only provide evidence for this but also demonstrate that memory is used to avoid ambiguities whenever the head does not move. Section 4 deals with front-back confusions in dummy-head stereo, and a possible way to alleviate the problem.
Finally, section 5 discusses some implications the present study has for work on interaural level difference (ILD).

1 REVIEW OF PRIOR WORK

1.1 Spectral features as localization cues

The first question addressed in our previous study <1> was whether processing of pinna cues takes place in the time domain or in the frequency domain. According to Batteau <4>, azimuth and elevation of a source are encoded by the time delays of pinna reflections. Our measurements using much finer spatial sampling showed that Batteau's model cannot work for all directions. Particularly sources with elevation angles greater than 30°

would not be accurately localized if cuing were based on time delays.

On the other hand, in the HRTFs we found features that systematically vary with the direction of the source. Fig. 1 maps these spectral features for the right ear of the KEMAR dummy head. The numbers indicate the frequencies of spectral minima, which for this specific "subject" usually follow the systematic shifts of the cuing slope along the frequency axis. This would not generally be the case (Shaw in <5>, pp. 30-41), but it is used here for want of a better method to indicate the position of spectral edges. Fig. 2 shows some examples of normalized HRTFs on which the feature chart is based (normalization is against 90° elevation).

The most often occurring spectral feature is a single or double notch, designated V and W respectively in fig. 1 (1). Systematic variation of notch depth (V↓) happens only above 10° EL (elevation). There is also no elevation-dependent variation of level (L↑) below 20° EL. When the source is on or below ear level, the only elevation cue is the left slope of a V- or W-notch. It shifts to lower frequencies when the source moves downward (↓V or ↓W). These spectral notches are due to pinna reflections, which are particularly strong when the source is relatively close to the head and on ground level. This is significant with respect to survival of early man.

As pinna reflections cause the comb filter minima, a priori knowledge of the spectrum of the source is not required. Therefore, movement of the steep spectral slopes along the frequency axis is a reliable way of cuing. Systematic level variations (L↑) due to the directional characteristics of the outer ear occur only for overhead sounds (around 90° EL, fig. 1). As intensity modulation originating from the source could bring about the same level variations, they would constitute a less reliable cue. The upper hemisphere may indeed be of lesser importance in terms of survival (as we can rely more on vision), yet this is where most work in directional hearing was performed (for a review see Blauert <6>).

In the top diagram of fig. 1, some bars have gaps with numbers indicating that the cuing slope does not move. In the vertical plane for 45° AZ (azimuth), the ambiguity between 40° and 30° EL is resolved by a level change (L↑). In the transversal plane (90° AZ) the V-bar has a gap between 50° and 30° EL. The slope is again stationary, but just in this range there is a peak in the HRTF moving from 9 to 8 kHz. This is a secondary cue, designated A2. For 135° AZ the ambiguity between 40° and 20° EL is resolved by a secondary notch (V2) starting at 30° EL. So in short, it can be said that each direction is represented by a unique set of features.

The lower diagram in fig. 1 shows the azimuthal dependence of spectral features for a clockwise moving source. As can be expected, the major feature in the anterior and posterior sector is level variation. The ILD we measured had systematic behavior in a somewhat wider range than given in the diagram.

In the lateral sectors interaural differences hardly vary with azimuth. A more reliable cue here is a spectral notch with a right hand edge shifting to higher frequencies for a clockwise movement of the source. The working range of this cue is 40° - 180° AZ.

In the frontal sector there is a pinna cue that differs from all others (1). It is a W-notch, of which the lower frequency minimum ascends between -40° and 0° AZ (fig. 2), causing the left slope to move to higher frequencies. Then the left minimum moves up again between 0° and 30° AZ.

It was pointed out in <1> that simple neural assemblies are capable of detecting spectral features. Feature detectors are known to exist in the visual system since the early sixties <2>. It would be hard to find a more plausible way of extracting pinna-based information.

1.2 The relation between spectral features and directional bands

In determining the directional bands, Blauert <3> used 1/3-octave bands of noise. The subjects, receiving the stimuli via headphones or loudspeakers in an anechoic room, were requested to indicate the apparent direction of the sound: in the frontal (v), posterior (h), or overhead (o) sector of the median plane. Fig. 3 shows the relative frequencies of the judgements v, h, and o. The directional bands (shown on top of the diagram) were defined by Blauert as "those bands in which the absolute majority of the population of observers gives one of the judgements more frequently than both of the others together".

Blauert then determined for 10 subjects the SPL at the ear-canal entrance when stimulated by a front loudspeaker minus the SPL when stimulated by a rear loudspeaker at the same distance. Fig. 4 shows the average front-back difference and the confidence interval. The boosted bands on top of the diagram were defined as "those bands, in which the majority of the population of observers has a higher SPL at the eardrum when stimulated from the front (fr) resp. from the rear (re) than when stimulated from the opposite direction".

Blauert could thus show that with respect to their position on the frequency axis, there is a close correspondence between directional bands and boosted bands.

In his experimental setup the sound sensations always appeared in the median plane, as there was no interaural time difference (ITD). Several years later, Hebrank & Wright <7> performed similar experiments with filtered white noise presented via a loudspeaker in the median plane. They not only used bandpass filtering, which in their case was 1/12 octave, but also low-pass, high-pass, and bandstop filtering.

H&W reported that a front sensation could be induced in 5 different ways. Let us first consider 3 of these, which are related to the same directional band:

(f1) 3.9 - 8.0 kHz low-pass cutoffs induce a front sensation. Increasing the cutoff frequency within this range is perceived as an increase in the elevation angle from 0° to 60°.
(f2) 4.0 - 7.2 kHz bandpass filtering is perceived in front with an elevation of 60°.
(f3) 7.4 - 10.8 kHz notches induce "frontness". Increasing the center frequency of the notch causes the perceived elevation to increase from 0° to 60°.

Blauert's v2-band is between approx. 3 - 6 kHz (fig. 3). All these kinds of filtered noise have one spectral slope in common. It is the slope that falls with increasing frequency. If there is no energy below this descending slope, the perceived elevation is high (f2). If the energy is high at lower frequencies, the elevation is cued by the position of the descending slope on the frequency axis (f1 and f3).

Our results are in agreement with the findings of H&W In fig. 1 notch frequencies run from 6 to 9.5 kHz with increasing elevat in the frontal median plane. Seen as low-pass slopes, the cutoffs move om 4.5 to 6 kHz (fig. 12A in <1>). As to cuing of elevation the agreement is even better on the other vertical planes of fig. 1. For KEMAR the frontal median plane is somewhat problematic, as the slope of the V-notch does not move between 20° and 80° EL. This will be further examined in section 4, which deals with front-back reversals. The HRTFs of Shaw (in <5>, pp. 34-35, also fig. 15 in <1>), which were measured with the source in the sagittal plane of an ear canal entrance, show for all 10 subjects a systematic shift of the descending slope when the source moves from -15° to 60° EL.

According to H&W an overhead sensation could be induced by 3 modes of filtering:

(a1) 10.3 kHz low-pass,
(a2) 8.1 - 9.1 kHz bandpass, and
(a3) 12.0 - 17.8 kHz notch.

These 3 cases and Blauert's o-band at 8 kHz have again a descending slope in common. Interestingly, a1, a2, and a3 can be seen as end points of the trends in f1, f2, and f3 respectively.

Fig. 6 (bottom part) shows the unnormalized HRTF of KEMAR for 90° EL. The N-curve is for the pinna in normal condition, the F+C curve for the pinna with its cavities filled with absorbent material <1>. Above 3 kHz the N-curve has the characteristics of a low-pass filter with a low-Q resonance peak at 9 kHz. (The small dip at 11 kHz is not due to a pinna reflection, as filling the cavities with cotton does not reduce it. It is caused by a 3/4 λ antiresonance of the ear canal). In view of these findings it may be conjectured that a descending spectral slope starting at 8 or 9 kHz activates an edge detector for overhead sensations.

Mehrgardt & Mellert <8> examined the dependence of HRTFs on source elevation for certain frequencies. They reported that only frequencies around 8 kHz have a significant boost of up to 14 dB for 90° EL. If this were the only cue for overhead sounds, intensity modulation of a source at 0° AZ would make it seemingly move up and down in the median plane. To distinguish between a source-induced and a location-induced change in the spectrum, an additional cue is required. The steep spectral edges introduced by the pinnae can make localization independent of level variations.

Continuing our review of H&W's article, the two remaining modes of filtering that elicit a front sensation are:

(f4) 13.2 - 15.3 kHz high-pass: the perceived elevation is 0°, and
(f5) 14.5 kHz bandpass: source appears somewhat elevated at 30°.

The common denominator of these two cases and Blauert's rudimentary v3-band at 16 kHz (2) is an ascending slope with increasing frequency. In fig. 2 we find a deep notch at 16 kHz if the source is between 90° and 180° (AZ). However, no such notch exists in the frontal quadrant. Front-back cuing must therefore be based on relative level in a frequency band. Frontal sources do not activate a detector for a near-vertical edge but a detector for a horizontal feature. There is no cuing of elevation angle here, the only question this feature detector has to answer is: is the source in front or behind?

In <1> we examined the boosted-band concept for directions outside the median plane tha ave the same ITD. We determined the difference HRTF for the intersec is of a cone of confusion with the horizontal plane. (Front and rear directions are thus symmetrical with respect to the interaural axis). These differences, front HRTF minus rear HRTF, are shown here again in fig. 5 for various azimuthal angles. It was noted in <1> that our results support the v2-band around 4 kHz and the rudimentary v3-band at 16 kHz. The peaks in fig. 5 around these frequencies are due to notches in the rear HRTFs. As we can see in fig. 2, for 90° AZ there is a notch at 3.75 kHz. Moving beyond 90°, it deepens and shifts to higher frequencies. Upon reaching 180° AZ, the frequency of the minimum is 7.5 kHz. The HRTFs in the frontal quadrant (0° - 90° AZ) however, show no substantial troughs between 3.75 and 7.5 kHz. This is in agreement with the bias towards the front direction that a v-band has.

The absence of the 3.75 - 7.5 kHz troughs in the front HRTFs may serve another useful purpose. Examining the interaural HRTFs in fig. 7 (right minus left), we find that the curves between -40° and +40° AZ are running more or less parallel to each other, while in the posterior sector between +140° and -140° AZ they intersect because of the deep troughs. Localization on the basis of ILD must therefore be more accurate in the frontal sector. ILD cuing in the posterior sector would still be possible, but in a limited frequency range.

It was found by H&W that 10.0 kHz high-pass, and to a lesser extent, 10.2 - 12.8 kHz bandpass noise is perceived behind. Both types of noise and the h2-band share the same ascending slope to 10 kHz. Examining fig. 2, we find that there is such a slope for all angles between -30° and 180° AZ. Fig. 5 shows that sources between 150° and 90° bring about a higher level in the 10 - 16 kHz range than sources between 30° and 90° AZ. Detection of rear sources would thus depend more on the boosted re-band than on the presence of an ascending edge. KEMAR has a boosted fr-band if the source is on the median plane or close to it (see section 4 for further comment).

1.3 Conclusions

It has been particularly rewarding to look for correlations between psychophysical data and features of HRTFs. The main conclusion to be drawn is that there are basically two kinds of pinna cues. One cue is a near-vertical spectral edge that moves along the frequency axis depending on elevation or azimuth of the source. Cues of the second kind, which work on signal level in a frequency band, do not specify an angle, but indicate whether the source is in front or behind, or fulfill a secondary function, such as resolving an ambiguity. This classification only partially coincides with the division according to directional bands. The reasons are that in Blauert's work percepts are limited to the median plane, and boundaries between v, o, and h are set arbitrarily (at -15° and 45° EL). One consequence is that the first kind of cue starts as a v- and ends as an o-band; another that a cue of the second kind operates in the same v-band.

2 DIRECTIONAL BANDS BELOW 3 kHz

2.1 Directional cues in the HRTFs

Now that we have clarified the relation between directional bands and pinna cues (3), it is natural to ask how cuing is performed below the working range of the pinna. For the lower frequencies useful raw data have been published. The collection of HRTFs made by Mehrgardt & Mellert <8> is one of the most extensive. Each of their curves is the mean of 20 subjects. Although they applied structural averaging, the notches in the high frequencies have become much less deep. It would not be a good idea to attempt finding pinna cues after the averaging process has obliterated fine details. An even more fundamental objection against averaging is that the feature detectors adapt to the individual HRTFs, not the average HRTFs. They must be adaptive in order to accomodate for growth.

Below 3 kHz, where the HRTFs do not depend on the pinna, averaging is allowable. In fig. 8 we have regrouped the HRTFs of M&M in order to show systematic variations in the lower-frequency directional bands. According to fig. 3 the v1-band is between 300 - 600 Hz and the h1-band between 700 - 1800 Hz. With some allowance in the boundaries of these bands, we can draw the diagram of fig. 9. It is expressed in terms of level variation (L↑, L↓) in a frequency band, although between -54° and 18° AZ it can be viewed as shifting of a spectral slope along the frequency axis. In the frontal and posterior sectors the direction of the change in level is in agreement with the clockwise movement of the source. Curiously, between -126° and -54° AZ on the contralateral side, the level variation in the h1-band is systematic but "against the grain". It probably is due to head diffraction. This cue is the only one that works on the contralateral side. The fact that it works at all in a lateral sector is in itself remarkable, since interaural differences are expected to be unreliable here.

Fig. 10 again examines the boosted-band concept not only for the median plane. The HRTFs are shown pairwise so as to show the difference between front and rear (in the horizontal plane) of source directions that are symmetrical with respect to the interaural axis. Within the h1-band it is particularly the range between 800 - 1500 Hz that a rear source gets a boost relative to a front source.

As to the v1-band, the difference between 0° and 180° AZ is in agreement with Blauert's data (fig. 4). It is maximal at 300 Hz. However, at +18° and -18° AZ it has practically vanished, and further from the median plane there is erratic behavior. In other words, outside the median plane there is no evidence for the existence of an fr-boosted band below 600 Hz. It can be hypothesized that the perceptual bias to the frontal direction in the v1-band does not develop by learning but is genetically determined.

2.2 The median plane

The HRTFs for the median plane in <8>, regrouped here in fig. 11, show several small peaks, which the authors attribute to reflections from the shoulder and knees of the sitting subject. These small fluctuations aside, the curves are practically independent of elevation angle above 45°. This is plausible as the upper part of the head can be approximated by a sphere.

The situation at lower elevation can best be appreciated from fig. 12, where each pair of HRTF shares the same elevation angle. The solid curve is for 0°, and the stippled curve for 180° AZ. Below 45° EL we find around 1 kHz the same re-boosted band as in the horizontal plane (fig. 10). The difference between front and rear is probably due to diffraction by the nose and other parts of the face.

Turning back to fig. 11, a possible elevation cue in the frontal median plane is a shallow dip moving from 700 Hz to 1.1 kHz when elevation angle decreases from 45° to 9°.

2.3 Ambiguities in localization

Below 3 kHz, front-back cuing is only based on level variation in a frequency band (figs. 8 and 9). While the steep edges in the high frequencies span 1/3 octave, the directional bands v1 and h1 are at least one octave wide (fig. 3). If localization were dependent only on HRTFs, it would be hard to tell apart level variations arising from (a) movement of the source, (b) movement of the head, and (c) spectral modulation of a stationary source (head also stationary).

Now let us assume that the source is emitting octave noise below 3 kHz, and that it is moving clockwise, for example from -10° to +10° AZ (fig. 13a). If the listener does not move his head, is there any reason why he should not hear the source moving from -170° to +170° AZ? The variation in ITD is identical, and the variation in ILD is very similar. If localization were only a matter of ITDs and HRTFs, the subject would not be able to distinguish front from rear in this situation.

Let us next examine case b. If the source is standing still in front, and the head turns counterclockwise (fig. 13b), one is easily tempted to think that it is not different from the former case, because the relative motions of head and source are equivalent. Yet the human perceptive system does not treat the two cases alike.

If the head turned ccw, the SPL at the right ear would increase for a front source and decrease for a rear source, and the time delay from source to right ear would decrease for a front source and increase for a rear source. In other words, by sensing the directions in which the ILD and the ITD vary when the head turns, the perceptive system is able to determine whether the source is in front or behind. The hidden assumption here is that the listener knows the position of his head relative to his environment. If he keeps his eyes closed during the experiment, he will be using his proprioceptive system.

Lackner <9> defines proprioception as: "sense of body position and configuration provided by receptor systems in the muscles, joints, and tendons of the body and the vestibular system". In order to be meaningful for an organism that moves about in his environment, data from the sensors of body position and movement should not only be integrated but also stored in memory. It is not often realized that sense organs and memory are inseparable. A simple experiment will prove the point. Let the subject be seated on a turntable, that at a given instant starts to rotate ccw. Assuming that he is memoryless, he would not be able to tell the difference between a stationary front source and a rear source moving ccw at twice the angular speed

of the turntable. In both cases the variations in ITD and ILD are identical. If the subject has unambiguous localizations, he must e memory. This experiment can actually be carried out in a slightly diff nt form by using a variant of experiment 1 as described in the next section.

Memoryless localization without ambiguities is possible if the sound field is sampled simultaneously at minimally 3 different points. For this reason alone it is most unlikely that binaural systems are memoryless. Memory would not only explain why we have unambiguous localization in case b, but also why we do not get front-back confusions the instant we cease moving our head. The simple experiments in section 3 demonstrate the effects of head movement and memory.

Before proceeding to these experiments, let us first examine case c. If for example, a source at 0° AZ is emitting low-frequency octave noise and modulating its intensity, it would not cause an apparent variation in azimuth because the ITD is constant.
In view of what has been said in §2.2, there is a possibility of an apparent movement in vertical sense. Gardner <10> reported that median plane localization was very poor when signals were low-pass filtered with a cutoff of 1 kHz, but memory and head movements were not given due regard in his study. This matter will remain inconclusive until further experiments.

3 EXPERIMENTS WITH FRONT-BACK REVERSAL

3.1 Experiment 1

Using the simple setup of fig. 14, it is possible to deliberately evoke a front to back reversal even in an ordinary, semi-reverberant room. In such an environment, the recommended distance between the front loudspeaker LS1 and the rear loudspeaker LS2 is 2 meter. A typical distance between the subject and LS2 is 60 cm, it should be reduced if the subject fails to get a false localization in step 4 below.

We used a pink noise generator and an Allison 2BR bandpass filter set between 800 and 1200 Hz. This is part of the h1-band, that according to fig. 10, is most significant in front-back discrimination. If such a filter is not available, an octave filter with a center frequency of 1 kHz can be used instead.

The procedure is as follows.
Step 1. In order to clear the memory, always start with a period of silence. The noise generator should be shut off.
Step 2. Ensure that LS2 is connected to the amplifier, then switch on the noise generator.
Step 3. Subject moves his head, preferably by rotation, so that he is positive about hearing LS2 behind.
Step 4. Subject turns his face to LS1, and keeps his head still. Then the noise signal is switched to LS1. Apparent location of LS1 will be behind LS2 (fig. 14b). While subject is still not moving his head, the conductor of the experiment may move LS1 perpendicular to the median plane. Subject has the impression that the source moves behind him. If the real source moves to the left side, the percept also moves to the left. This situation is similar to fig. 13a.
Step 5. LS1 is brought back to its original position, then the subject is

allowed to move his head. He will now hear LS1 in front.

Using narrow-ba noise, only one directional band is activated. If head movement is suppressed, the auditory system can be tricked to a false localization. This experiment proves that movements of the source alone cannot resolve the ambiguity (step 4). The subject continues to have a false localization as long as he holds his head still. This in itself is enough to falsify the contention that localization can be fully described by static HRTFs.

In a variant of this experiment, the subject is seated on a turntable. While he is not initiating any head or body movements, the turntable is set in motion by the conductor of the experiment. (In step 5).

3.2 Experiment 2

This experiment deviates from experiment 1 only in the beginning stimulus, yet the result is dramatically different. The same noise band is used.

Step 1. Silent period lasting a few minutes.
Step 2. After having connected LS1 to the amplifier, switch on the noise generator.
Step 3. Subject moves his head, so that he hears LS1 in front.
Step 4. Subject ceases movements, and facing LS1 continues to keep his head still. The signal is now switched to LS2.
Step 5. After a couple of seconds, switch back to LS1. The subject hears LS1 in front.

Now there is no front to back reversal. Once the true position of LS1 is stored in the subject's memory, it cannot be falsely localized even after a short interruption.

After performing this experiment, a long pause is required in order to have a front-back confusion again using the procedure of experiment 1.

3.3 Discussion

As former workers have silently assumed that localization is a memoryless process, we should be cautious in adopting their interpretations. Most statements in literature regarding the importance of head movements are not based on the proper experiments. Any experiment, in which sounds prior to its start are ignored or stimuli are presented in random order, is suspect. If a series of experiments were performed, consisting of n times experiment 1 and n times experiment 2 in random order with brief intervals between them, the outcome would most likely be that head movements have no effect on localization.

Front-back confusions are less likely to occur if stimuli are broadband. In the high frequencies, where pinna reflections provide information on elevation and azimuth, it may seem that memory can be dispensed with. However, some observed phenomena suggest that for some listeners the pinna cues fail in the median plane, and that memory is used to "fill the gap". This is the subject of the next section.

4 FRONT-BACK CONFUSIONS IN DUMMY-HEAD STEREO

4.1 Some observations on front-back confusions

When dummy-head stereo is demonstrated to a randomly selected group of people, a substantial part of it hears frontal sources behind the head. As we have seen above, localization depends on (associative) memory with inputs from at least four separate systems: (a) the binaural system that processes ITD cues, (b) the feature detectors, (c) the ILD system receiving information from horizontally oriented feature detectors, and (d) the proprioceptive system. In view of this it is not likely that the observed front-back confusions have only one cause. Formerly they used to be attributed to differences in pinna shape between the dummy head and the listener. It was only guessed that the absence of head movements in the dummy could play a significant role, and for a long time it remained a mere guess. There was no real need to thoroughly investigate this aspect because of an obvious reason. Listening to a dummy-head recording on tape, LP, or CD it would not make sense to require that the dummy head moves in the same way the listener's head moves. It is only recently that this attitude has changed somewhat, primarily through the advent of virtual reality. Having available head-position sensors to control the video, they could as well be used to control the HRTFs in the acoustic simulation. There is no doubt that virtual reality has a bright future, but it is not yet affordable for the average consumer. Until then, the best thing we could do is to arrive at a new understanding of dummy-head stereo.

It is our experience that localization problems in dummy-head stereo are most severe on or close to the median plane. There are a number of observations that are relevant in this context.

Observation 1. Butler and Belendiuk <11> made tape recordings with miniature microphones in the ear canals of 4 subjects. Sound sources were placed in the frontal median plane. Each subject listened to all recordings and was requested to locate the taped sounds. It turned out that most accurate performances were observed when listening to the tape made through the ears of subject 1, and least accurate performances were observed when locating sounds taped in the ear canals of subject 4. Only 2 listeners (subject 1 and subject 3) performed best on his or her own tape. B&B concluded: ".... some pinnae, in their role of transforming the spectra of the sound field, provide more adequate cues for MSP localization than do others". (MSP = median sagittal plane).

Observation 2. Asano, Suzuki, and Sone <12> investigated median plane localization through DSP simulations. Two subjects listened either through his own HRTFs or through the other's. HRTFs could be simplified by reducing the poles and zero in the approximation. Localization errors were consistently lower when using the HRTFs of subject 1 in the simulation (fig. 5 in <12>). Errors in front-rear judgement increased with simplification of the HRTFs, but they were significantly less when listening through the HRTFs of subject 1 (table I in <12>).

Observation 3. A sound source moving in a circle around a dummy head is often perceived as moving in the trajectory sketched in fig. 15. The shape of this trajectory suggests that the problem area is the frontal median plane. Back to front reversals are very rare if the dummy is in a reverberant environment.

Observation 4. It has been observed in <1> that the fr-boosted band at 5 kHz (fig. 4) practically vanished in KEMAR for the difference between 0° and 180° AZ, fig. 5. The difference HRTFs in fig. 5 show a large peak at 4.5 kHz for 40° AZ. It shifts to higher frequencies and decreases in height as the source moves closer to the median plane. At 10° the peak is still visibly present at 7 kHz, but at 0° it has become insignificant. As intersubject variations can be quite large, it is possible that KEMAR happens to have "bad" ears, while the majority of Blauert's subjects by chance had "good" ears.

Further evidence that KEMAR's pinnae are not so good can be found in the upper feature chart of fig. 1. The only cue between 20° and 80° EL in the frontal median plane is depth variation of a V-notch. The minimum remains at 9.5 kHz, and there is very little shift of the left edge. Contrary to what prior investigators may have thought, it is more efficient for neural assemblies to detect one edge of a spectral notch than the depth of a minimum. The last case would require two edge detectors. As the edges are very steep close to the minimum, it is doubtful that notch-depth cuing is as reliable as edge movement along the frequency axis. Outside the median plane, depth variation is confined to a 20° range and supplemented by another cue (fig. 1, 45° and 90° AZ).

Another failure of KEMAR concerns the h2-band. As reported on p. 6, there is an fr-boosted band instead of re- if the source is between 0° - 20° AZ.

Poulsen <13> reported that, particularly in the lower frequencies, the Neumann KU80 head has fr- and re-boosted bands reversed with respect to Blauert's data (fig. 4).

A possible cause of the anomaly in the frontal median plane is diffraction through the facial features. In his measurements at the ear canal entrance, Shaw could avoid diffraction by using a special source generating a progressive wave at grazing incidence (p. 19 in <5>). Therefore the HRTFs of all his 10 subjects clearly show the pinna cue of the first kind (pp. 34-35 in <5>, also fig. 15 in <1>).

4.2 Hypothesis

The writing of this paper produced a multitude of ideas that await further testing. Measurement of the HRTFs of various subjects, to be performed in the near future, will hopefully verify the hypotheses presented in this section. In other words, the present hypotheses merely serve the function of giving direction to future research.

It occurred to us that facial diffraction may in itself not be enough to account for the fact that there are "good" ears and "bad" ears. Why are there people who can unfailingly localize frontal sources in dummy-head stereo? Would facial diffraction not make them all bad? Griesinger's article <14> suggested the possible nature of a second factor: the angle between the pinna and the sagittal plane. It made us perform a simple experiment using the setup of fig. 14 in a semi-reverberant room. We fed octave noise having a center frequency of 4 kHz to LS1. While listening to this frontal source, we pressed the pinnae flat against the head with a finger. This created the impression of the source moving to behind the head. What actually happened was that the frontal sensation just ceased, and localization was hardly possible. When on the other hand, the pinnae were pressed forward thus

increasing the angle with the sagittal plane, the frontal sensation remained but the source is perceived as louder and at a shorter distance.

Examining a randomly selected group of people, it would not be surprising to find a substantial variation in the pinna angle. Yet we all have learned to assign proper distances to sound sources. This happened during our earliest years with the aid of visual information. Therefore, listeners having a greater pinna angle than the dummy head will, as Griesinger said, find frontal sources too far away. We would add that some of these listeners will have a front to back reversal. Listeners who have a smaller pinna angle will perceive the sources as too close, but still in front.

Returning to our experiment with the setup of fig. 14, it was equally instructive to listen to the rear loudspeaker LS2. Pressing the pinnae flat against the head resulted in in-head-localization (IHL) and an apparent increase in pitch. Pressing the pinnae forward did not alter the pitch, but created a diffuse, nonlocalizable image. On a higher SPL or at a reduced distance to LS2, there was also IHL in the latter situation.

These effects are in agreement with the fact that the most often perceived faults in dummy-head reproduction are front to back reversal and IHL. Now the question arises: why do they rarely occur in normal circumstances, i.e. when listening directly to original sound sources? An obvious answer is that those who have "bad" ears use their associative memory to compensate for their "handicap". Unlike the dummy-head situation, this memory receives proper information during head movements so that it is able to "fill the gap".

4.3 Improving dummy-head stereo

How to build a better dummy head? The course to be taken has now become rather obvious. Instead of trying to determine the average anthropometric pinna, find a good localizer and copy his ears. The methods used by Butler and Belendiuk <11> or Asano et al <12> can be the basis for finding a set of pinnae that does not cause front-back reversals in most listeners.

Alternatively, correlations could be determined between good localization performance and HRTF features. Once the useful features have been identified, selection of "good" ears is a matter of objective measurements.

5 IMPLICATIONS FOR MODELING THE AUDITORY SYSTEM

It was more or less silently assumed that the feature detectors work monaurally. If feature detection took place after determining the interaural differences, it would indeed be more difficult to identify the features originating from the left ear and those from the right ear. Left-right ambiguities would be particularly undesirable in the learning phase during infancy, when the feature detectors are adapting to the pinnae. In view of these considerations, it is fairly certain that monaural feature detectors precede the ILD processor.

For the same reason, it is not likely that ILDs are processed simultaneously with ITDs as proposed by Lindemann <15>. There is another, even greater objection against such processing: the auditory system would then not be able to distinguish front from rear during head movements.

It is interesting in this context that van Keulen et al <16> arrived at the same conclusions via a completely different route. Through psychophysical experiments using acial stimuli presented via headphones, these authors found that there are two primal images: a time image based on ITD, and an intensity image based on ILD. If these two images are lateralized close to each other, they will fuse to one lateralized image. Conflicting observations of previous investigators could be reconciled, as time-intensity trading turned out to apply to the fused image but not to the primal images. In the model that has these characteristics, ITDs and ILDs are processed separately, and monaural channels precede the ILD processor.

A block diagram of the auditory system that incorporates these insights is shown in fig. 16. In the visual system neural cells are known to exist that fire only when, for example, a pattern moves from right to left. Similar sensors of motional direction could well be used to compare the variations in the ITD, the ILD, and the proprioceptive vector for front-rear discrimination during head movements.

The associative memory can be represented as an energy landscape <17>. The inputs define a point in the landscape, from which as it were a ball is let loose. It will fall in the nearest minimum, where the coordinates of some specific direction are stored.

6 CONCLUSIONS

Examination of HRTFs revealed that, in the lower frequency directional bands v1 (300 - 600 Hz) and h1 (700 - 1800 Hz), azimuth angle is encoded as level variations.

In these bands front-back ambiguities are resolved during head movements by comparing variations in the ITD, the ILD, and the proprioceptive vector.

Associative memory is used to stabilize localization when the head does not move, and when there is a failure in pinna cuing.

Some experiments and observed phenomena suggest that for a part of the population pinna cuing works poorly in the frontal median plane. It has no consequences in normal listening situations, but during dummy-head reproduction with headphones it will result in front-back reversals because the stationary dummy does not disambiguate.

7 NOTES

(1) One error in fig. 14 of <1>, of which fig. 1 here is a revised version, is the lowest bar for $\alpha = 0°$ (α = azimuth, θ = elevation). It should be placed in the V-column. This eliminates the only remaining ambiguity (between -20° and -30° EL) as noted between parentheses in §2.4 of <1>.
Another oversight was the W-cue in the frontal sector of the horizontal plane, bottom chart of fig. 1.

(2) The v3-band did not reach the required level of significance to be classified as a directional band, but Blauert was confident that it would in tests with a greater number of subjects.

(3) Hebrank and Wright <7> unified the effects of high-pass, low-pass, band-pass, and notch filtering through an analogy of Mach bands. Although this

idea is not invalidated, we consider edge detection to be the principal factor in the unification.

8 REFERENCES

<1> H.L. Han, "Measuring a dummy head in search of pinna cues", 90th AES Convention (1991), preprint 3066.

<2> D.H. Hubel & T.N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex", J. Physiol., vol. 160, pp. 106-154 (1962 Jan.).

<3> J. Blauert, "Sound localization in the median plane", Acustica, vol. 22, pp. 205-213 (1969/1970).

<4> D.W. Batteau, "The role of the pinna in human localization", Proc. Roy. Soc. (London), vol. B168, pp. 158-180 (1967).

<5> R.W. Gatehouse (ed.), Localization of Sound: Theory and Applications (Amphora Press, Groton, 1982).

<6> J. Blauert, Spatial Hearing (The MIT Press, Cambridge, 1983).

<7> J. Hebrank & D. Wright, "Spectral cues used in the localization of sound sources on the median plane", J. Acoust. Soc. Am., vol. 56, pp. 1829-1834 (1974 Dec.).

<8> S. Mehrgardt & V. Mellert, "Transformation characteristics of the external human ear", J. Acoust. Soc. Am., vol. 61, pp. 1567-1576 (1977 June).

<9> J.R. Lackner, "Influence of posture on the spatial localization of sound", J. Audio Eng. Soc., vol. 31, pp. 650-661 (1983 Sep.).

<10> M.B. Gardner, "Some monaural and binaural facets of median plane localization", J. Acoust. Soc. Am., vol. 54, pp. 1489-1495 (1973 Dec.).

<11> R.A. Butler & K. Belendiuk, "Spectral cues utilized in the localization of sound in the median sagittal plane", J. Acoust. Soc. Am., vol. 61, pp. 1264-1269 (1977 May).

<12> F. Asano, Y. Suzuki & T. Sone, "Role of spectral cues in median plane localization", J. Acoust. Soc. Am., vol. 88, pp. 159-168 (1990 July).

<13> T. Poulsen, "Hörvergleich unterschiedlicher Kunstkopfsysteme", Rundfunktech. Mitt., vol. 22, pp. 211-214 (1978).

<14> D. Griesinger, "Binaural techniques for music reproduction", Proc. 8th Int. Conf. (AES, 1990) pp. 197-207.

<15> W. Lindemann, "The extension of binaural crosscorrelation modelling by a mechanism of lateral inhibition", J. Acoust. Soc. Am., vol. 74, S85(A) (1983).

<16> W. van Keulen, F.A. Bilsen & J. Raatgever, "Interaural intensity and time images revisited", Fortschritte der Akustik - DAGA '91. (Bochum, 1991) pp. 489-492.

<17> D.W. Tank & J.J. Hopfield, "Collective computation in neuronlike circuits", Scientific Am., vol. 257, pp. 104-114 (1987).

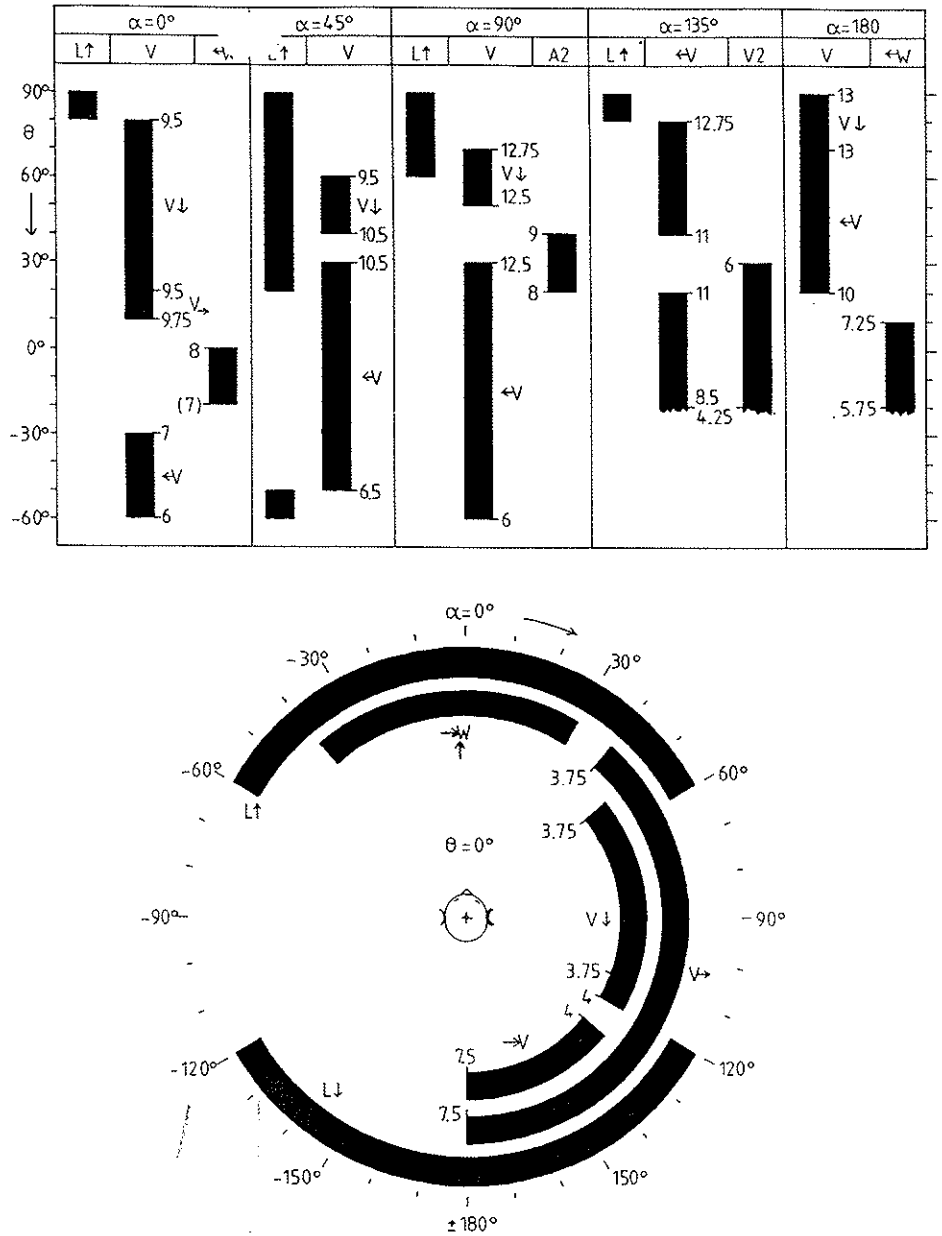

Fig. 1 Spectral features varying with decreasing elevation in vertical planes (top), and with increasing azimuth in the horizontal plane (bottom). Frequencies of minima are indicated in kHz.
L↑ = level increase; V↓ = deepening valley; ↔V = valley with left slope shift; ↔W = double valley with left slope shift; V2 = secondary valley; A2 = peak, secondary cue.

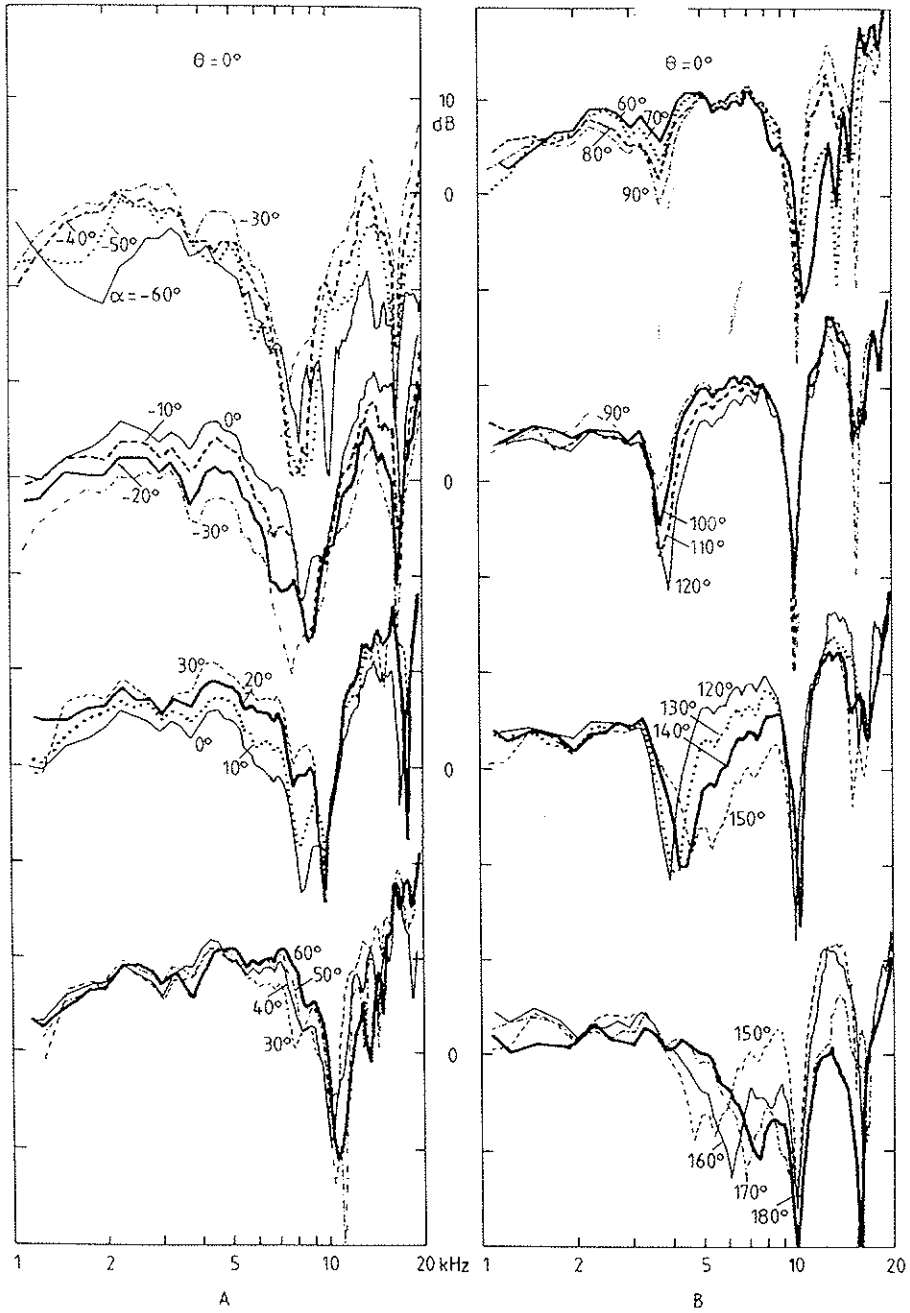

Fig. 2 Normalized transfer functions of KEMAR's right ear simulator. Source in the horizontal plane.

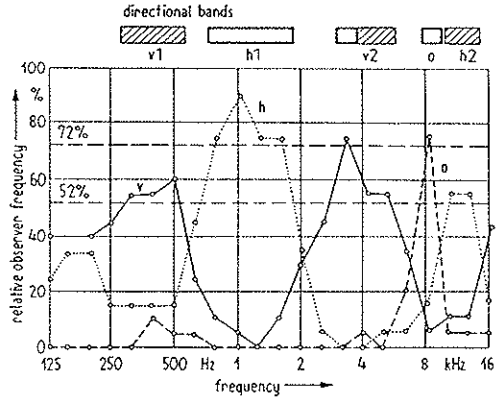

Fig. 3 Relative frequencies of judgements "v" (front), "h" (behind), and "o" (above) for 1/3-octave bands of noise. The directional bands are shown on top of the diagram. (See text for definition. Bordered: at 90% level of significance, shaded: most likely). From Blauert <3>.

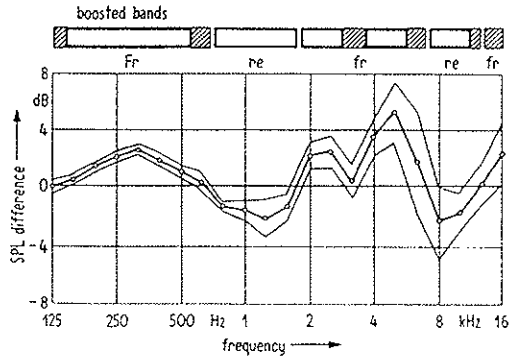

Fig. 4 SPL at ear canal entrance with source in front minus SPL with source behind the subject. The boosted bands are shown on top. (See text). From Blauert <3>.

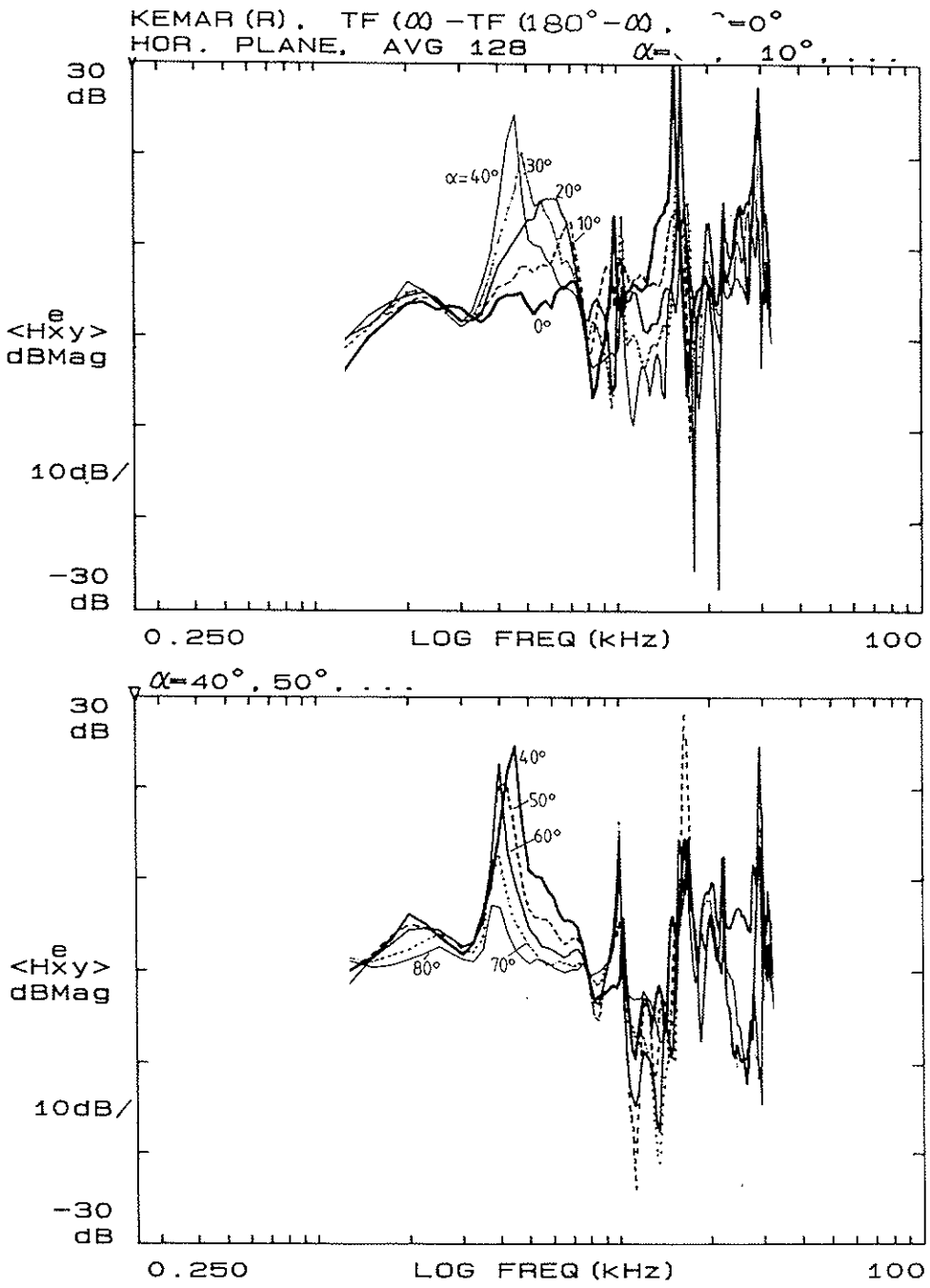

Fig. 5 Difference between front and rear on the cone-of-confusion intersections with the horizontal plane.

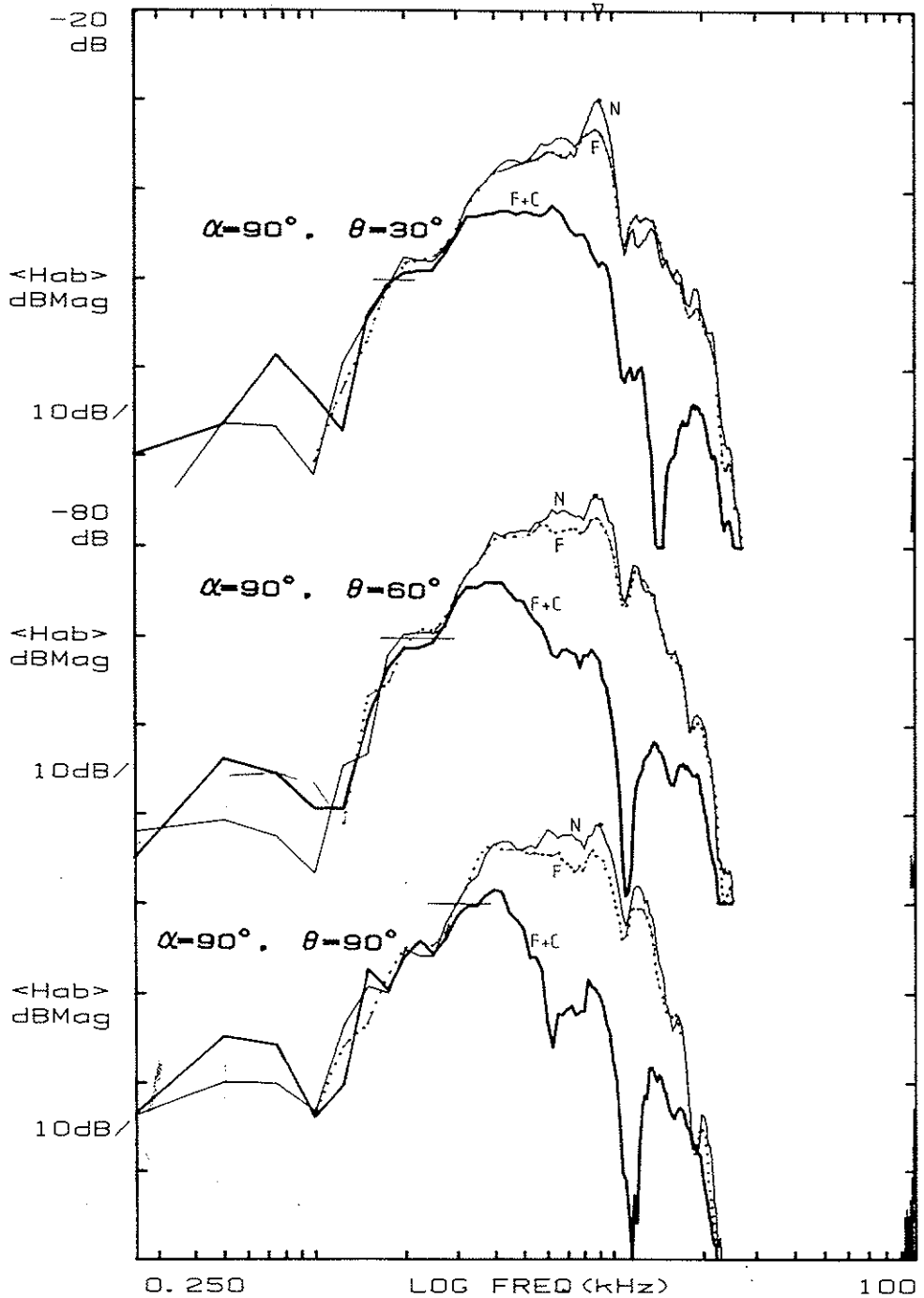

Fig. 6 HRTFs measured at KEMAR's right ear canal entrance. α = azimuth, θ = elevation.

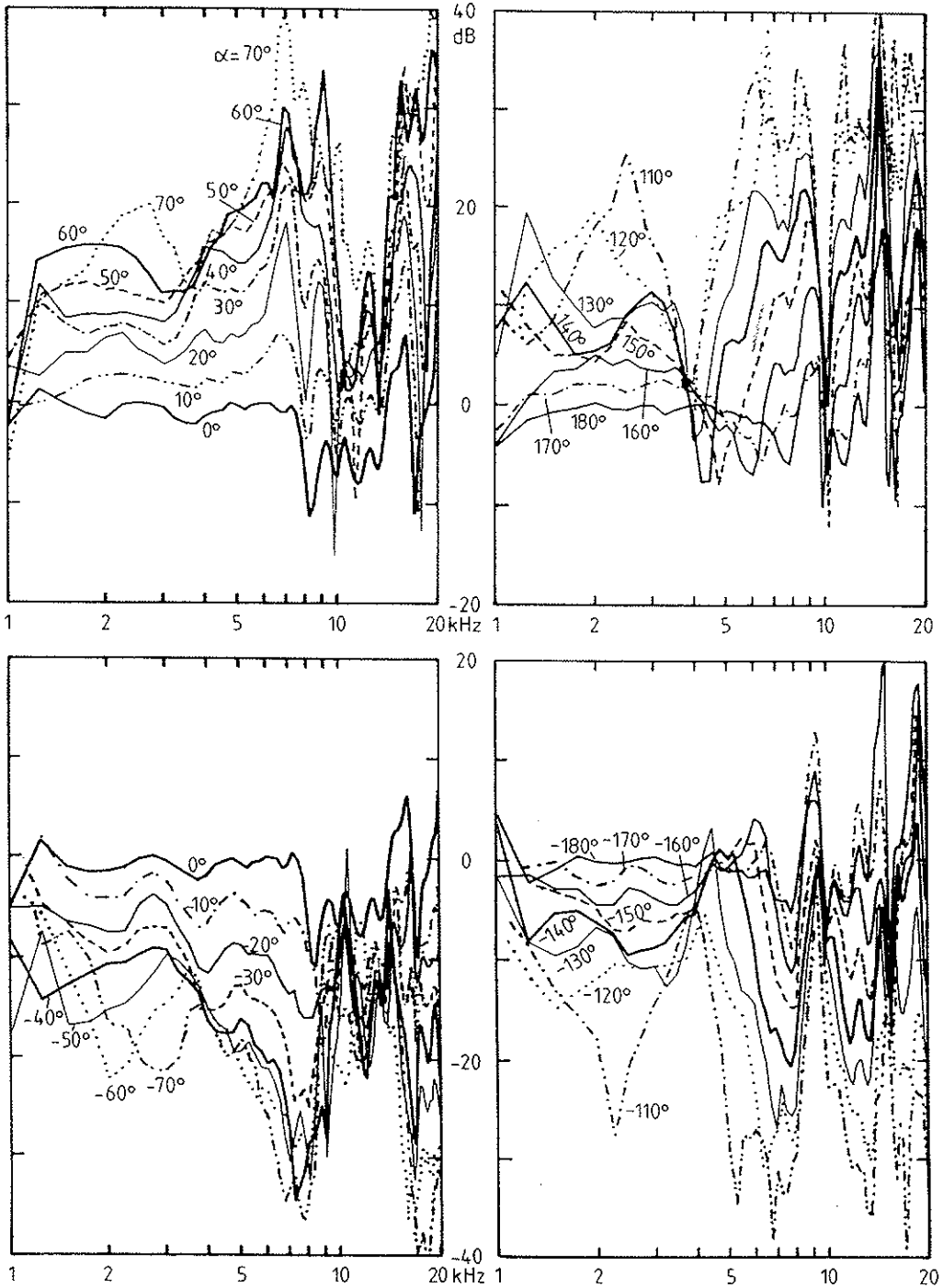

Fig. 7 Interaural HRTFs for various azimuthal angles on the horizontal plane. Measured on KEMAR.

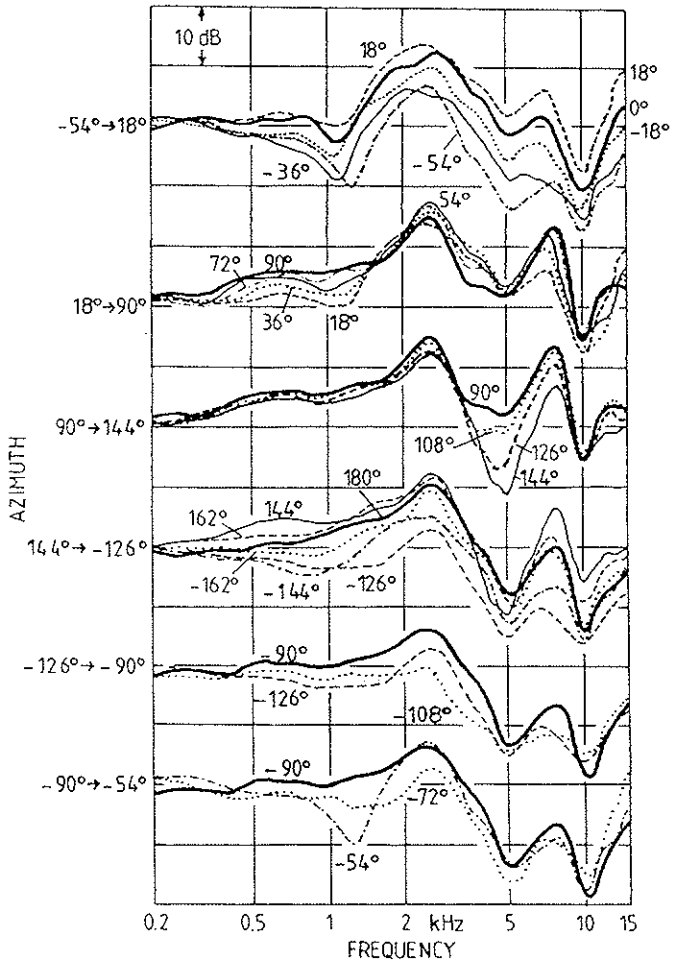

Fig. 8 HRTFs in the horizontal plane showing variations in the lower frequency directional bands v1 and h1. Data from Mehrgardt & Mellert <8>.

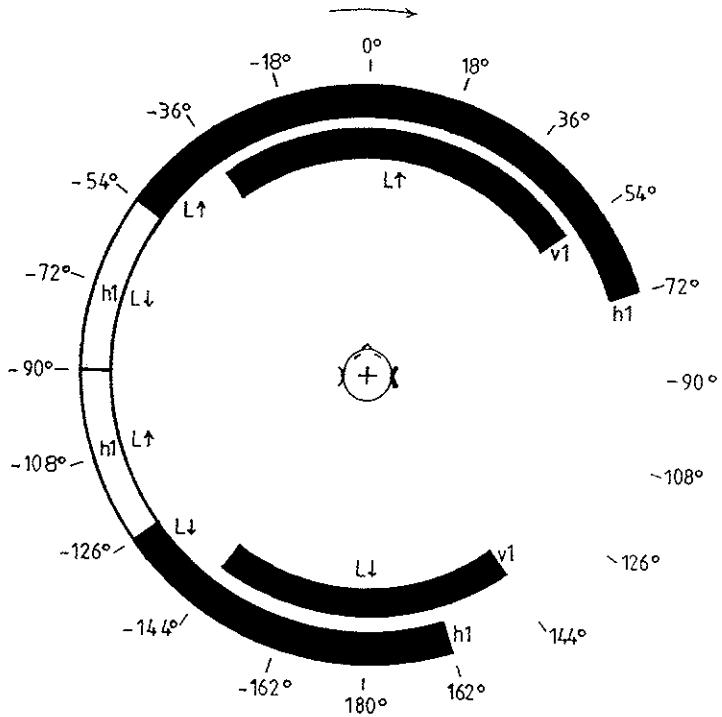

Fig. 9 Level variations in the directional bands v1 and h1 for a clockwise moving source. Derived from fig. 8.

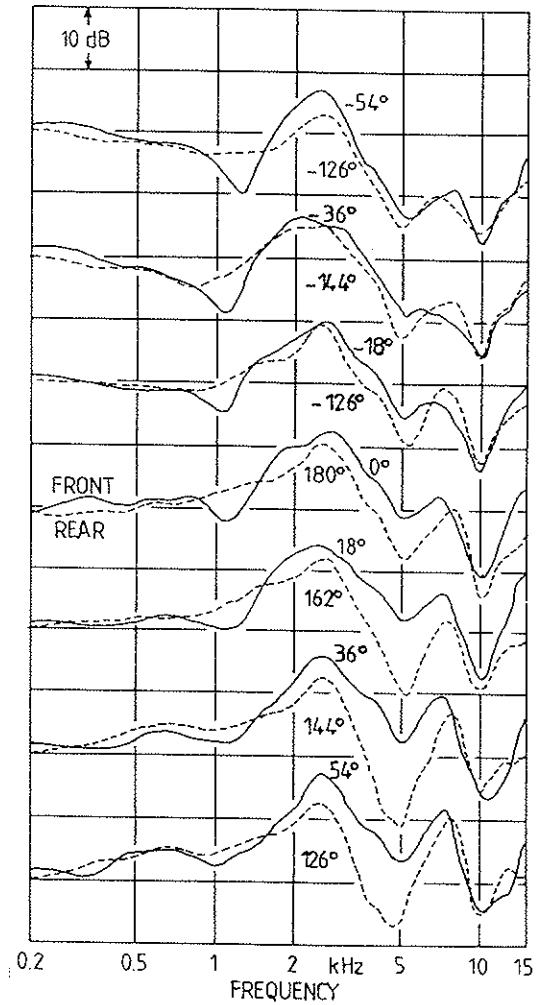

Fig. 10 Same HRTFs as in fig. 8, but shown pairwise for directions that are symmetrical with respect to the interaural axis.

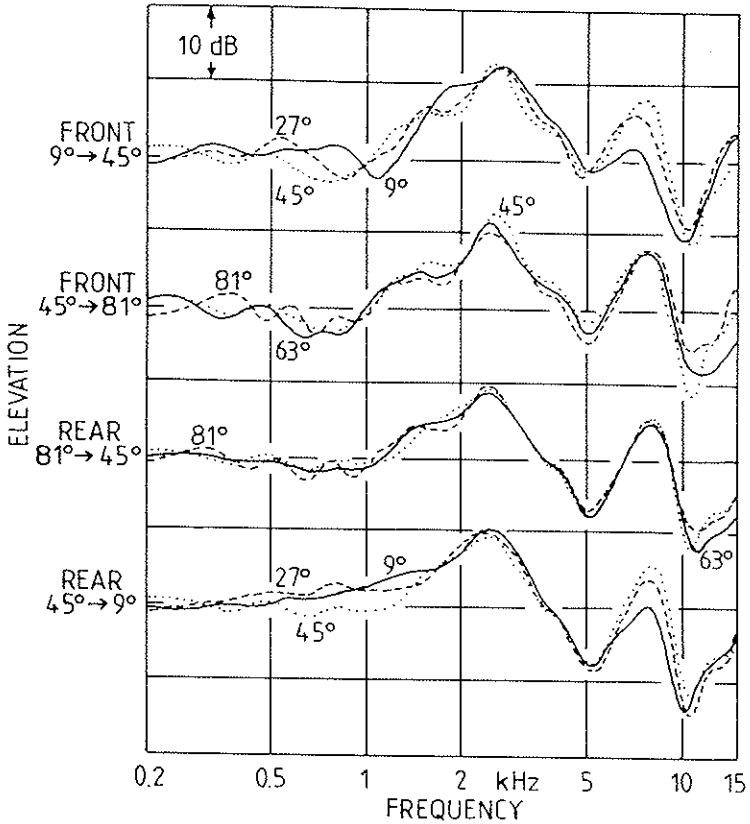

Fig. 11 HRTFs in the median plane. From Mehrgardt & Mellert <8>.

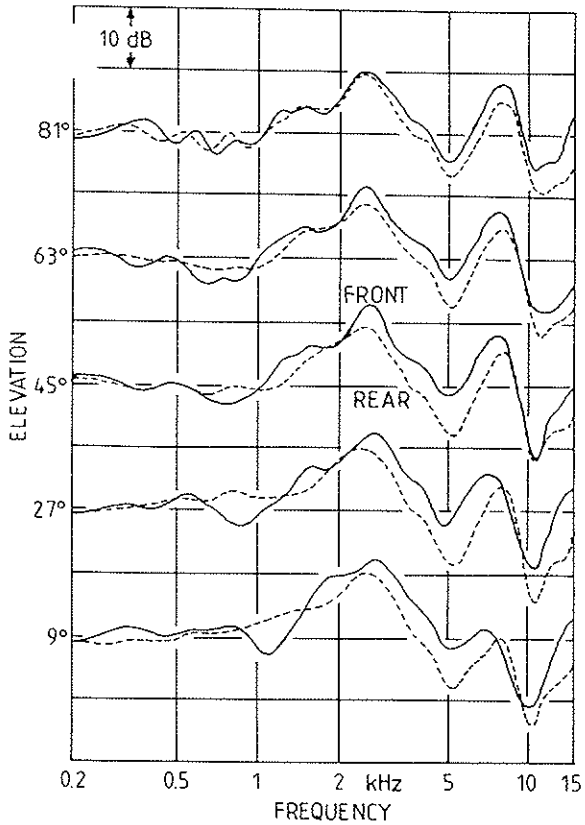

Fig. 12 HRTFs in the median plane for directions that are symmetrical with respect to the vertical axis.

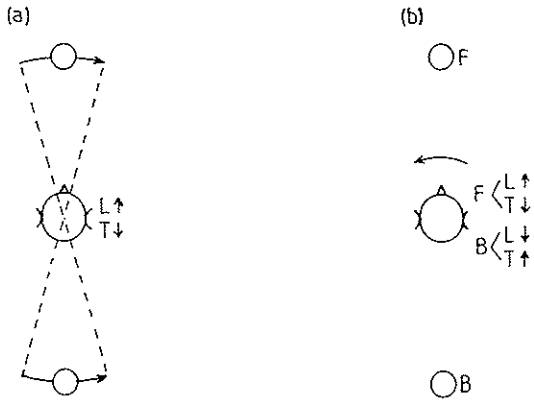

Fig. 13 (a) Moving source, head stationary. (b) Moving head, source stationary. L = level, T = delay from source to ear.