

Binaural Techniques for Music Reproduction

David Griesinger

Lexicon, 100 Beaver Street, Waltham, MA 02140

Presented at the 8th International conference of the AES, May 3-6 Washington, DC 1990

[With comments from 3/9/2009 in red]

INTRODUCTION

Binaural recording and signal processing are generating boundless enthusiasm in the audio press these days, potentially offering perfect surround from only two loudspeakers, incredible earphone reproduction, etc. Yet the physics of binaural hearing and the enormous differences in ear shape between different individuals present possibly insurmountable barriers to these goals. This paper will review the principals of binaural hearing, and use the results of our own research and that of many others to describe just how high these barriers are. We will then show a few ways they can be bypassed or worked around. Our own research goals at Lexicon are binaural recording techniques which are at least as effective for two channel loudspeaker reproduction as standard miking, improved performance from loudspeaker stereo, and headphone equalization which allows the full benefits of binaural recording to be enjoyed by a large fraction of interested listeners.

LOCALIZATION WITH BINAURAL HEARING

First lets review what scientists in the field know already:

1. People perceive sound distance and direction only through cues present in the sound pressures at the two eardrums. There are no magic bone conduction or body conduction effects.
2. The influence of the shadowing of the torso, head, and pinnae on the frequency spectrum perceived at the eardrums is both profound and a strong function of sound direction.
3. The frequency effects of the pinnae are radically different between different individuals, and between the two ears of a single individual. Pinnae response curves are as unique as fingerprints.
[I now believe that pinna responses are more similar than different. However there are large variations in the shape and dimensions of ear canals, and these differences provide more variance in the overall transfer function from the external soundfield to the eardrum than the pinna.]
4. Direction in the horizontal plane is almost entirely determined through pressure DIFFERENCES between the two ears, both amplitude and time. These differences tend to be similar between individuals.
5. Direction in the vertical (medial) plane is determined through comparing the perceived frequency spectrum of a sound source to previous experience with such sources at known directions. Thus it is not in general possible to determine the height of a sound you have never heard before, or to determine the timbre of a sound from an unknown direction. The timbre cues used to determine height for one person may bear little resemblance to those of another. [On the contrary, it is easy to determine the azimuth and elevation of a sound you have never heard before. A bird you have never heard is as easy to spot as one you have often heard. Nearly all natural sounds have a relatively simple spectral shape above 2000Hz, where most of the useful HRTF variance occurs. We fit the observed spectra to fixed spectral templates]

to find elevations, and this process happens within milliseconds. Detecting timbre takes more time – fractions of a second. I think it is rare that we are unable to determine at least a rough direction for a sound, and given enough time, we can get a good idea of the timbre.]

6. The ability to localize sound without visual reinforcement varies widely among individuals. If you measure the frequency response variations with angle of the individuals who localize poorly you find their pinnae have much more uniform response than individuals who localize well. Some people have difficulty telling if a sound source is in front or behind them.

7. Frequency response differences between individuals are maximum in the forward direction.

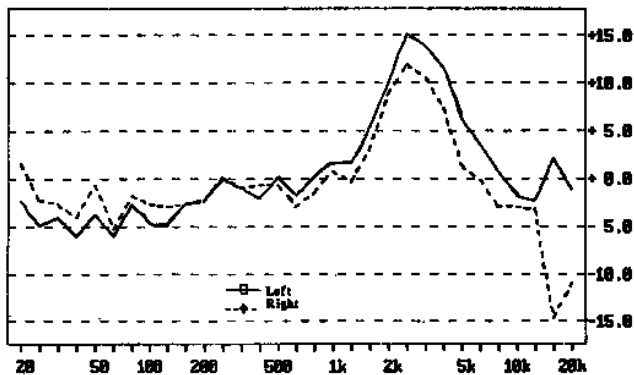


Figure 1. Forward Pinnae Response, DG

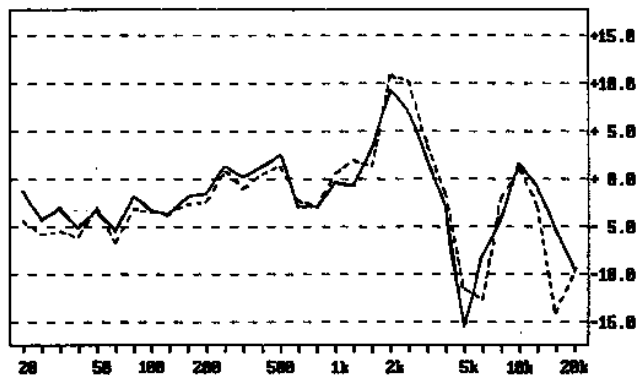


Figure 2. Forward Pinnae Response, FC

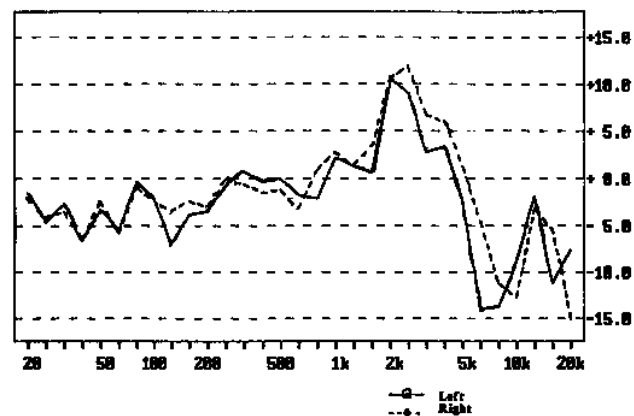


Figure 3. Forward Pinnae Response, PM

Figures 1,2, and 3 show some response curves measured for forward incidence of several subjects. A probe microphone was placed as close to the eardrum as possible. Note the differences between the left and right ear of each subject, and the large variation between individuals. In figure 1 notice the left ear of this individual has more gain at 2kHz than the right. This ear will be more susceptible to hearing damage. An earphone placed on this ear does not yield this extra gain, and so the treble response from earphones will always be weighted to the right. Such differences between the right and left ears are very common among individuals, and show no clear pattern in our data. Lets see why these differences exist (Kuhn).

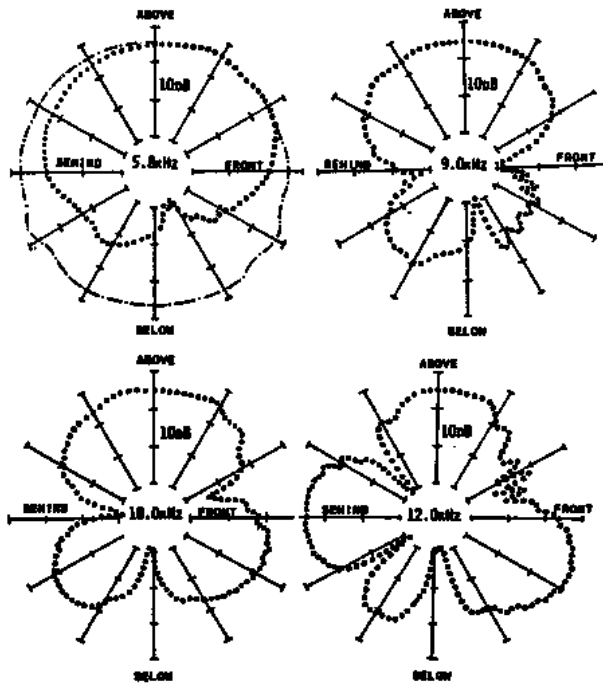


Figure 4. Vertical median plane directivity measured at the coupler microphone of KEMAR (without torso). Male adult pinna. “_.” shows the directivity of the head alone at 5.8kHz, pinna replaced by a flat plate. – From Kuhn.

Figure 4 shows the polar response of KEMAR at several frequencies with a male adult pinna and without a torso. At 5kHz we get a notch from beneath the dummy. At 9kHz there is a notch below the dummy, but also one in front. At 10kHz this notch has moved to about 20 degrees above the horizontal plane, and at 12kHz there are a series of complicated notches at 30 degrees elevation. Clearly the brain could determine the height of a broad band sound source by detecting the frequency of these notches. Figure 5 shows the same type of data a different way, and graphs five different pinnae. Note the general rotation of the notches is easy to see, but note also that the differences between the pinnae are maximal from the front, and minimal from above! When the frequency response of a sound source does not match the one expected from the front the brain tends to interpret it as coming from above or inside the head. Notice also how all the pinnae have a maximum in treble response about +50 degrees from the front above the horizontal plane. Such a rising treble response is typical of all ears. [Better – When the frequency spectrum at the eardrum does not match one of the learned spectral templates the brain is unable to determine the location of the sound, and we perceive it as a default location inside and at the top of the head.]

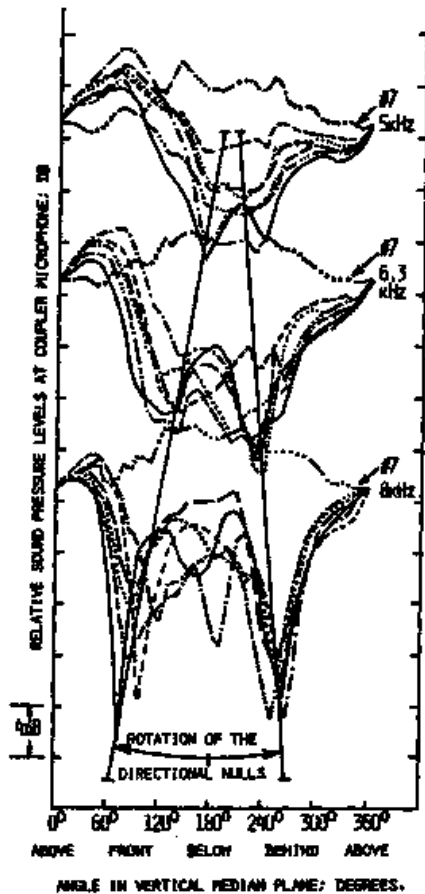


Figure 5. Vertical, median plane response for five pinnae and small KEMAR's head without torso. Curve #7 is for head alone where pina was replaced by a flat plate. Note the rotation of nulls in front and rear. – From Kuhn.

Figure 6 shows how these rotating notches can be created with a very simple model. (from Butler and Beldndiuk, 1977) Figure 7 shows the resonances and anti resonances in a typical ear which combine to give the response we measure. Note there are a lot of them. (from Platte via Blauert)

It is obviously to our advantage as a species to be able to discriminate the elevation of a frontal sound as accurately as possible. To do this you want the response in that direction to be as complex as possible, and to change as quickly as possible with direction. Couple this needed complexity with the genes responsible for human facial differences, and you have an audio engineer's nightmare.

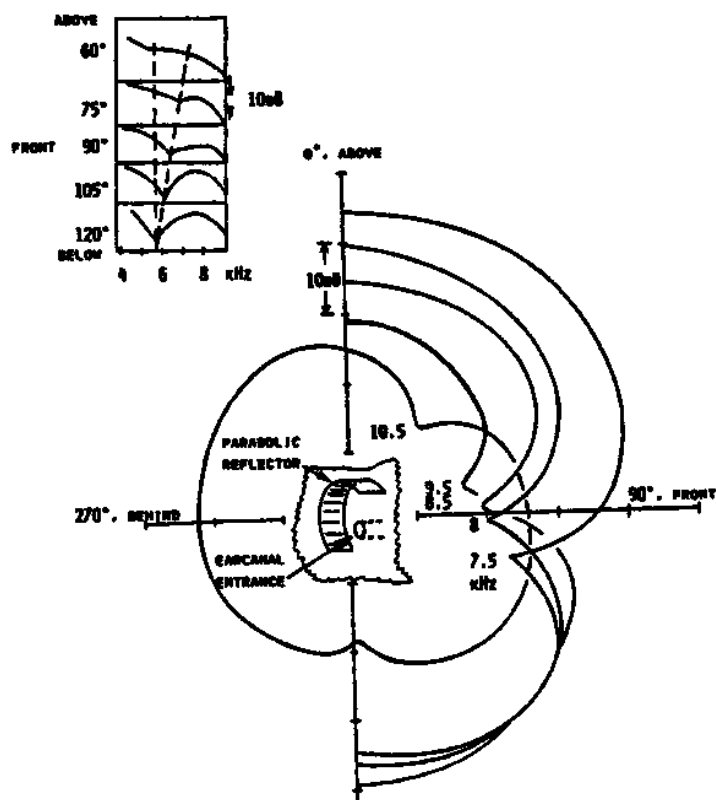


Figure 6: Rotation of null in the vertical median plane directivity measured at coupler microphone for a model reflector. The insert shows the rotation of the null in the spectrum at the ear canal entrance. — From Butler and Belendiuk, 1977

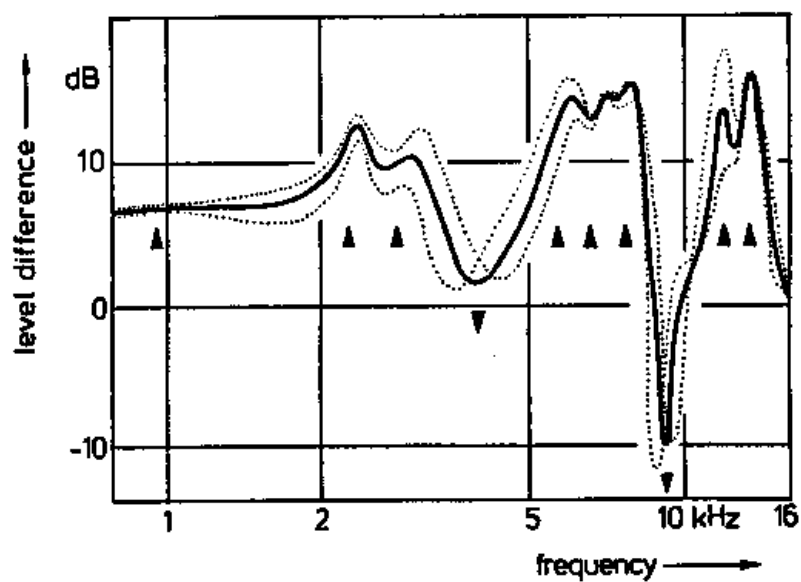


Figure 7 – Two individual, smoothed external-ear transfer functions and structurally averaged curve. The resonances and antiresonances are indicated. – From Platte 1979 by way of Blauert.

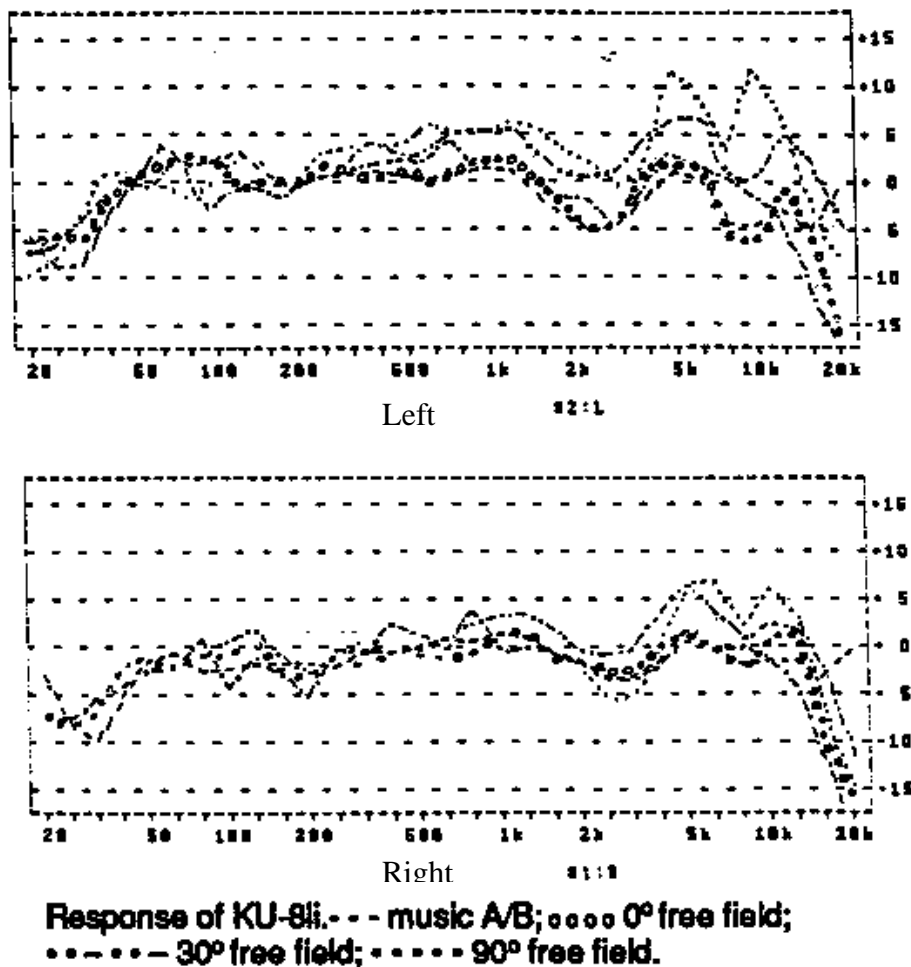


Figure 8 – Response of Neumann KU 81i dummy at various angles in the horizontal plane. (Griesinger)

LOUDSPEAKER STEREO

Nearly everyone expects the musicians to sound in front of them when they listen to stereo. Fortunately when the loudspeakers are in front the individual's own pinnae give the frequency response necessary for frontal localization. It is almost accidental that two stereo loudspeakers at ± 30 degrees also work. They do so only because for most individuals the frequency response is nearly constant as a sound source moves from zero to ± 30 degrees in the horizontal plane. Ordinary stereo is only capable of reproducing sounds which are perceived as coming from a line connecting the loudspeakers.

USING FREQUENCY RESPONSE TO CREATE AN ILLUSION OF HEIGHT

In some cases with loudspeakers we can use frequency response to pull images above the horizontal plane. If we give sounds a frequency contour equal to the difference between the frequency response of a

desired elevation and the response at the actual position of the loudspeaker we can create an image with the illusion of height. We need the specific responses of the individual listener to make this work well, but fortunately directions above the listener tend to have rather simple frequency response, and it is frequently not too difficult to achieve a height illusion which works for many people. Making a source appear to descend below the loudspeaker is more difficult. Motion in the vertical plane is only possible for relatively broad band sources with a spectrum which is familiar to the listener, especially sources which move from an expected position to a new one.

IT ONLY WORKS IN NEAR FIELD OR A WELL DAMPED ROOM

Phantom images outside the line connecting the speakers are only possible if the room is not reflective. The hearing mechanism needs a relatively pure spectrum from the direction of the loudspeaker in order to have enough information to detect a change. Thus the playback room should be absorbent and very carefully set up.

WIDENING THE HORIZONTAL PLANE WITH CROSSTALK CANCELLATION

We can extend the localization in the horizontal plane beyond the ± 30 degree spread of the loudspeakers by increasing the interaural time and level differences at the ears of the listener through interaural crosstalk cancellation. Ideally we should also make an additional correction to the equalization, since the 90 degree response in the horizontal plane is slightly different than the frontal response. (see figure 8.to see curves for 90 degrees on the Neumann Dummy) In my experience spectral cues are incapable of causing side localization without crosstalk elimination. Interaural crosstalk elimination works well without spectral cues, but confines the listener to a very small area. Spectral cues spread the effective area only slightly. Interaural crosstalk elimination works well for a correctly positioned listener, and creates the most realistic illusion of being in an original sound field that I have heard without custom equalized headphones. [In Schroeder's original work with crosstalk cancellation the inverse filters for the crosstalk cancellation were calculated from measurements with probe microphones at the listener's eardrums. Thus the crosstalk filter not only removed the crosstalk between the two ears, it also removed the HRTF functions for forward localization. When the signal was played the forward HRTF of the dummy was perceived, without being further convolved with the listener's. With a simplified crosstalk cancellation based on a spherical head model with no pinna the reproduction of a dummy head recording is nowhere near as accurate. However a similar result can be obtained with individual equalization of headphones if measurements are made at the eardrum, and a flat frontal loudspeaker is used as a reference for both the calibration of the earphones and the calibration of the dummy head.]

LOW FREQUENCY L-R BOOST (SPATIAL EQ)

If we really try to increase the pressure differences between the two eardrums at low frequencies we must dramatically boost the difference signal between the two loudspeakers. All crosstalk elimination schemes are equivalent to a L-R boost at low frequencies. Such a boost is vital in loudspeaker reproduction of recordings made with a dummy head, which have little or no difference signal at low frequencies. The resulting boost simply raises their low frequency separation enough to match normal stereo recordings. Alan Blumlein realized the need for an L-R boost in his original 1931 work on stereo, where he started

with recordings made with closely spaced omni microphones. What Blumlein called "shuffling" is simply crosstalk elimination confined to low frequencies.

If ordinary intensity or widely spaced stereo recordings are played with such a boost there will be much too much low frequency energy. Thus when ordinary stereo recordings are played through crosstalk elimination the circuit must confine its action to mid and high frequencies.

The need for bass separation is quite controversial in Europe, where many engineers use closely spaced pressure microphones and find no problems with the lack of bass separation. To my ears these recordings can sound quite natural on earphones, but lack spaciousness on loudspeakers. Interestingly, carefully set-up symmetric playback rooms can reduce the audibility of difference signals at low frequencies, since the same room modes will be excited by each loudspeaker. In such a room an L-R bass boost is inaudible.

The lack of low frequency separation may not be noticeable in a completely symmetric playback room, but it is quite noticeable on a typical home system. To be really compatible with loudspeaker stereo recordings, binaural recordings should have a L-R boost applied. When played over earphones these recordings will sound too wide and spacious, just as ordinary recordings do. Ideally earphone amplifiers should remove the boost, which will make ordinary stereo recording sound better also.

One of the major requirements of any recording microphone is a signal which allows easy mixing of accent microphones. The low frequency L-R boost is important here too. Once the head has been equalized and spatially equalized it is possible to mix accents into the signal with pan pots in the usual way, since the low frequency width will then match the width of the panned images. This ability to mix allows the dummy to be used as a main microphone pick-up, where its ability to record the hall and maintain proper depth perspective can be optimally used. After spatial equalization accents can be added in a familiar way without spoiling the result either on headphones or speakers. As usual with accents the image from the accent tends to be brought closer than the original. If this effect is not desirable it can be corrected with an ambience simulator such as the Lexicon 480.

Playing back such a compatible binaural signal on loudspeakers with crosstalk elimination (remember to remove the bass L-R boost before you connect Cooper's system, or use "Panorama Normal" on the Lexicon) is for most people the most accurate way to reproduce an original sound field. In a good playback room the illusion of height and depth in the forward hemisphere is marvelous, and sound extends all the way to the sides of the listener. I have not been successful in creating reliable images fully behind the listener, which is easy to do in earphones, but the accuracy of the front image is excellent. The issue of matching the forward response of the system to the listener's ears does not arise. [Although the front image is pleasant, it does not reliably give the impression of actually being in the original space. True accuracy in playback requires a near anechoic playback room, and Schroeder's method of removing the listener's own HRTF functions. Ideally the recording should then be made with probe microphones at the listener's own ears. This method is very convincing.]

TO SUMMARIZE:

Ordinary loudspeaker stereo works in part because it utilizes the listener's own pinnae in a fashion which just happens to be close to their natural function. [And ordinary stereo does not include forward HRTF functions, so only the listener's are active.] Under ideal listening conditions it is possible to use

frequency contouring and crosstalk elimination to move some images above and beyond the loudspeakers. The process works best when sources have familiar spectra, or with sources which move. However even under ideal conditions the process can be expected to work for some people and not for others. Movement beyond the loudspeakers in the vertical plane can be made to work in a much larger listening area than large movement in the horizontal plane, where the difference between the ear signals provides the primary cue.

EARPHONE REPRODUCTION OF BINAURAL AND STEREO

Alas the subject of earphones is more complex. Earphones are problematic not only for binaural reproduction, but even for ordinary stereo. Placing the earphone on the pinna profoundly alters the frequency response of the ear. If we wish to reproduce a sound source in front of the listener we have no choice but to accurately measure the transfer function between the sound source and the eardrum, measure the transfer function between the earphone and the eardrum, and correct the earphone response to match the natural spectrum. Correction curves developed this way are very different for different individuals. Work of this kind has been reported by many researchers, including Plenge, Blauert, Theile, and Wightman. With a lot of luck the procedure works. With careful measurement of both ears, and independent equalization of each channel, the earphone response can be compensated so that a sound source appears to be outside the listener's head and in the horizontal plane. Ordinary stereo played through these phones sounds very good. Timbre appears entirely natural, and balance and localization are also good. It is not sufficient to simply bypass the pinnae in playback by using transducers which couple directly to the ear canal. Such a transducer still must be equalized so the resulting response at the eardrum matches the measured frontal transfer function. [Note the insistence on eardrum measurement with an open ear canal. Both before and after this paper was written there was a shift from eardrum measurement to measurement of a blocked or partially blocked ear canals. Measurements made with blocked canals do not correctly account for the individual transfer functions of the ear canals, and especially do not account for the effects of these differences on the response of headphones. The author's recent data typically show ± 10 dB variations in frequency response for popular headphones depending on whether an ear drum or blocked ear canal was used. These differences are highly audible – and an equalization obtained from an eardrum measurement is clearly superior both in timbre and in localization.]

Our goals for earphones are: More accurate headphones, both for use by the recording engineer in making a recording, and by the consumer. This goal has several sub goals:

Earphones should provide:

- a. sharp localization of pan-potted sources when listening with headphones
- b. accurate right/left balances when using headphones to balance recordings
- c. no change in timbre of a pan potted source as the position is swept from right to left.
- d .. accurate (or at least plausible) timbre.
- e. Out of head localization.

In addition we want much more accurate frontal localization than current headphones provide, so binaural recordings have a chance of sounding realistic. To meet these goals some form of custom earphone equalization is unavoidable. Is such equalization possible?

The answer is a definite YES, but the measurement procedure is tricky, tedious, and not very repeatable. Furthermore, in spite of the very best efforts it is unsuccessful for some individuals. Lets see why: See figure 9.

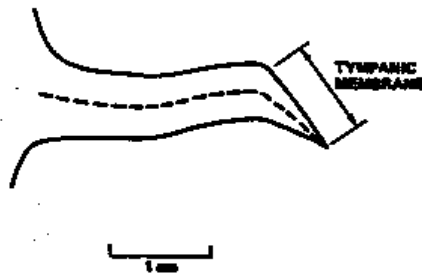


Figure 9 - Cross section through a typical ear canal. The dashed line is an approximation to the midline of the canal. – From Stevens et. al.

[Figure 9 is a gross simplification of a real ear canal, which is more exponential than cylindrical, and bends and twists in a highly individual fashion. My recent work has shown that a headphone equalization made at my eardrums works to create frontal localization for binaural recordings also at my eardrums for at least 50% of the people who try it. This can be raised to about 75% if there are some adjustments to the treble balance. Recent work has shown that an individual calibration of a pair of headphones can be obtained without a probe microphone through loudness matching.]

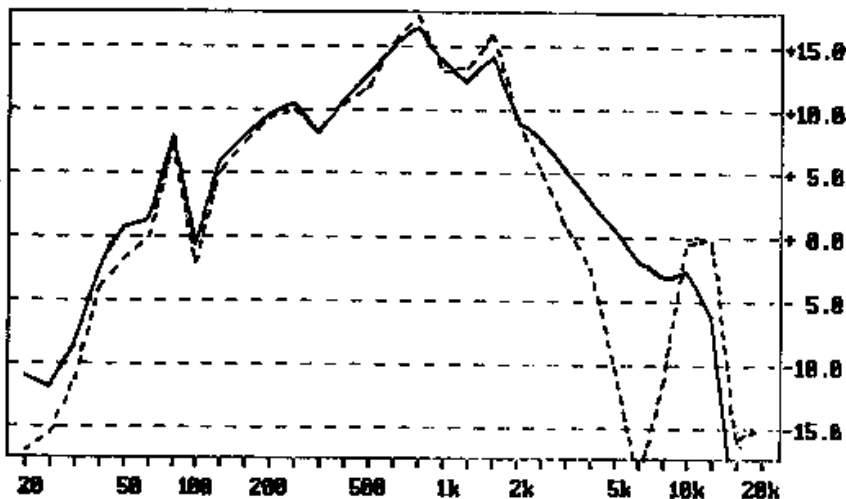


Figure 10: Two uncorrected right transfer functions of the same ear. Solid line with the probe against the eardrum, dashed with the probe pulled away, but still well within the ear canal.

Notice in figure 9 that the ear drum is at an angle relative to the ear canal. The actual angle varies greatly from individual to individual. A very different pressure will be measured if the probe tube moves from the bottom to the top.

Figure 10 shows that as the probe moves closer and further from the drum the response also changes wildly. Furthermore, the way it changes depends on whether the pinna is open or if it is closed by an earphone. It is also nearly impossible to keep the probe from moving when an earphone is put on or taken

off, which is why measurement at the eardrum (where pressure is relatively constant) is helpful. With my equipment the measured frequency correction above 10kHz is nearly always wrong. (see Stevens et al.) With luck we get repeatable data below about 8kHz, and when this is equalized correctly the listener can usually successfully manually adjust the top 3 bands of a 1/3 octave equalizer while listening noise or to familiar music. See figure 11, 12 correction curves for two people.

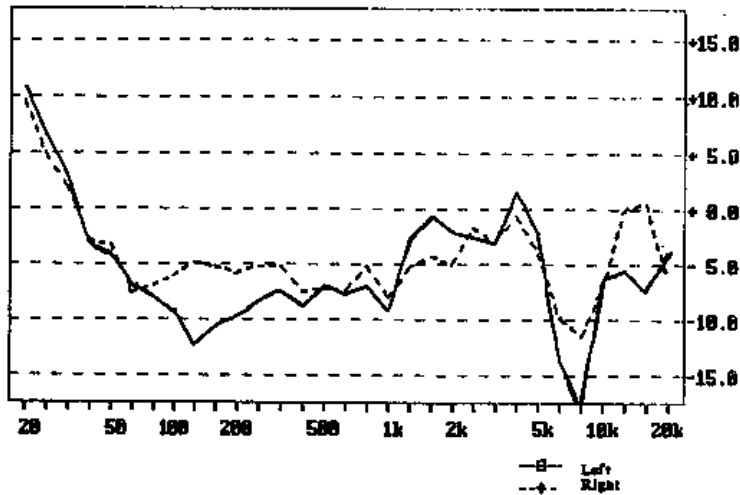


Figure 11. Correction curves for AKG 240 headphones on PM

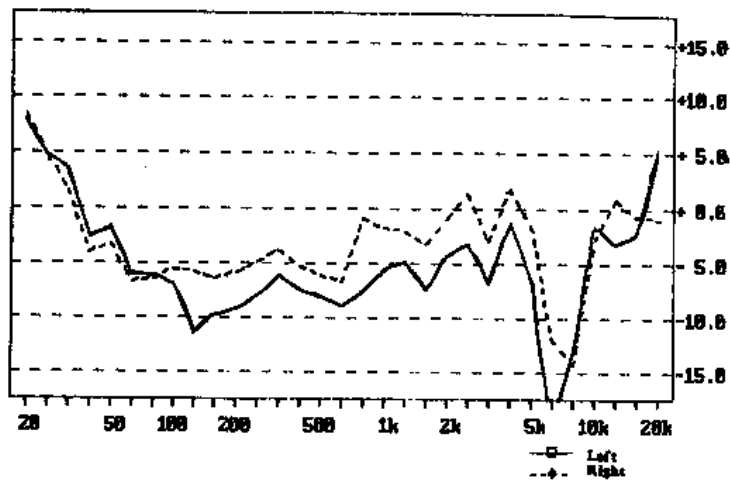


Figure 12. Correction curves for AKG 240 headphones on MC

The equalization curves developed for one individual usually have a passing resemblance to those for another individual, especially when they are measured to 1/3 octave. The resemblance is only apparent. I have yet to find any individual who does not prefer their own correction curves to anyone else's. Usually another person's equalization is incapable of providing frontal localization.

Spikofski and Theile at the IRT in Munich have developed a standard for earphone equalization. Under this standard earphones are equalized so the ear canal pressure with the phones (averaged over a large

number of individuals) matches the diffuse field response of the same individuals. This may be the best compromise if one earphone equalization must be imposed on all people. The difference in timbre and localization between this and custom equalization is dramatic. The only encouraging point in all this is that one can grow accustomed to a given earphone response through repeated listening. Many engineers report that after long use they are able to effectively balance with the oddest earphones. Hearing is marvelously adaptive. However in my own experience this adaptation is misleading. One is convinced of the reality of the listening experience, but balances made with mismatched earphones are not correct when later heard over loudspeakers. [Alas – a diffuse field response is not identical to a response obtained with no HRTF functions. There is an identifiable sound to a diffuse field, which puts it immediately outside the head and at some distance. For dummy head recording and headphone equalization it does not matter whether you use a diffuse or frontal source as a reference, so long as you use the same source for calibrating the dummy head and for calibrating headphones. However when ordinary music is played through diffuse field equalized headphones (assuming they are measured at the eardrum) the timbre is considerably too bright. If the headphones were equalized using a blocked ear canal their response is probably not predictable – and either choice of reference will give equally poor results.]

TO SUMMARIZE:

Reproducing binaural sound through earphones has a built in requirement that the earphones be matched in some way to the listener. When this is not done binaural reproduction through headphones is basically a sham. However after long use motivated listeners can convince themselves that binaural recordings work. For such listeners, it probably does. However the reality they experience may not be the one which was actually present in the recording situation.

DIFFICULTY OF MEASUREMENT

As we mentioned before it is not at all easy to repeatably measure the frequency response of a person's external ear and ear canal, especially if you want the position of a probe microphone in the ear canal not to change when earphones are placed on the head. In fact, Klaus Genuit says the differences between his anthropomorphic dummy and his simplified dummy are within the differences measured with the same test subject on two different days. The best measurements I have seen have been made by Wightman. Deep lucite earmolds of his subjects were made, and then the plastic in the ear canals was drilled until it was paper thin. [I cannot imagine making such an insert for my ear canals without drastically altering their shape. I don't doubt they get reproducible results, but I have trouble believing the results reflect the actual sound pressure at the unmodified ear.] The microphone probe was affixed to the end of the earmolds, right next to the eardrum. Repeatability of these measurements is very good, and when the data are used in Scott Foster's computer the results are excellent. Our current data are only marginally repeatable, even with the frequency averaging of the 1/3 octave measurements and the deliberate spatial averaging. They also usually suffer from a consistent over boosting of the high frequencies. Figure 13 shows two correction curves for the same person made at two different times.

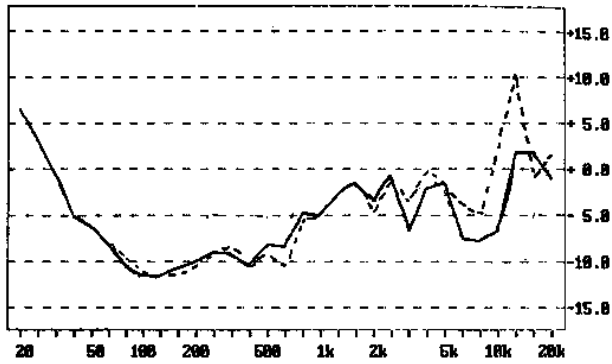


Figure 13. Correction curve for AKG 240 headphones on DG left ear on two separate measurements. Solid curve 10:00am, dashes 3:30pm.

ACCURACY NEEDED

The degree of accuracy needed in the response appears to depend on the listener. Some listeners appear to achieve out of head localization (OHL) with a simple 2/3 octave magnitude equalization of the system, and some listeners never achieve OHL even when the exact impulse response is duplicated digitally, both amplitude and phase. (This comment is from Elizabeth Wenzel.) [Wenzel may not have been using eardrum measurements. If binaural recordings are made at the subject's eardrums, and then are played back through earphones equalized at the subject's eardrums they are almost guaranteed to be successfully perceived. However, some subjects may be intensely visually oriented. Without a visual cue they may be unable to externalize sound. I suspect these people are rare. If they had their eyes open during the recording, and re-visualize the scene I suspect they will also hear it correctly.] At this point my best guess is that the ability to achieve OHL depends not only on the ears of listeners, but also on their ability or willingness to suspend the evidence of their other senses. In the absence of visual cues external localization may be impossible for some people. In this regard the ability of a binaural system to track head movements may be helpful. Scott Foster has developed a system called the Convolvotron (using Wightman's ear data) which is capable of generating a binaural representation of a moving monaural source. The position of the source accurately tracks the movements of the listener's head. This system works surprisingly well, especially when the transfer functions of the system match those of the listener. Bo Gering in Toronto has also developed such a system, which works surprisingly well for me. But even with these systems not all listeners are satisfied.

It would be helpful to know the accuracy to which the spatial transfer function of a recording and a listener must match. It may be possible to obtain this information with a system such as the convolvotron. The convolvotron derives its continuous position changes by interpolating time responses stored in its memory. Clearly the intermediate transfer functions thus derived will not match the real ones very well, and yet the machine is amazingly effective. Averaging time responses in this way continuously alters both amplitude and phase in an interesting way. We do not know if this averaging is optimal, or if interpolating the magnitude response only might work as well.

Alexander Persterer from AKG in Vienna tells me that equalizations which include corrections for phase produce more precise localization than those which do not. However it is my guess that for satisfactory reproduction of music the equalization need not be very precise. I suspect that magnitude only equalization to within 1dB or so when averaged over 1/3 octave bands will be good enough. Phase

information may be chiefly useful to catch needlessly bizarre settings of the equalizer. Our current analysis system does not measure phase.

In my experience $2/3$ octave magnitude equalization works pretty well - and $1/3$ octave equalization is more than good enough. Equalization to higher accuracy than this is probably a bad idea, particularly for reproducing music from broad sources.

INDIVIDUAL VARIATION AND LEFT/RIGHT DIFFERENCES

Unfortunately different individuals have very different head/pinnae equalization curves, even when spatially averaged and measured to $1/3$ octave accuracy. There can also be differences between the left and right ears of a particular individual.

It appears to be a popular belief that left/right asymmetries are necessary or perhaps even critical for OHL. We have no data to support this contention, and several bits which contradict it. The newest Aachen head is entirely symmetric, and works well for some people, including myself when I listened through my own custom equalized phones. But we have noticed that reversing the channels of a standard binaural recording even after both channels have been separately $1/3$ octave equalized makes a difference to the sound of the recording. The difference is hard to describe. Reversing the channels does not destroy OHL for the author, but the correct channel assignment definitely sounds better.

There is a much more dramatic difference if the two channels of the recording microphone have not been separately equalized, or if a correctly equalized set of headphones is worn with the earphones reversed. Errors in balance and timbre are obvious. These balance and timbre errors persist even when ordinary stereo is played.

For example, If we play monaural pink noise through headphones we usually do not hear a sharp image in the middle of our nose. The image is usually spread out, since the transfer function from the earphone to the ear drum is different for each ear, and the resulting response does not match the transfer function of the same pinnae to a frontal sound source. The image can be narrowed (and the lateralization accuracy of the headphones can be considerably improved) by introducing a $2/3$ octave or a $1/3$ octave equalizer into each channel and changing the left/right balance of each frequency band until the noise image is narrow and exactly centered. The differences in equalization can be surprisingly large - more than 4dB for some bands. A simple equalization procedure of this type can greatly improve the usefulness of earphones in balancing a recording, and I highly recommend it to engineers. I will also make the observation that it is possible to make an image in the earphones sharper than it is in natural hearing, and that phones equalized in this way are more precise for balancing than loudspeakers.

TIMBRE AND FRONTAL LOCALIZATION WITH PINK NOISE

The apparent lateral position of the image can be easily adjusted by such a process, but it will probably appear to be either in the middle, above, or behind your head. It is much harder to make an image move to the front in the medial (up/down) plane. A pink noise signal through headphones usually sounds pretty bad in timbre, full of peaks and valleys in the sound. One might think forward localization would improve if you removed these timbre errors with further equalization. Unfortunately it is not possible to simply adjust each frequency band until they all appear to be equally loud, since what you are really doing in this

case is adjusting each band for equal masking. As Genuit has pointed out, masking depends strongly on the pinnae functions, and constant masking is not a good match to the frontal frequency response. In any case, if you think you know what pink noise should sound like when heard from the front you are probably fooling yourself.

A pink noise source which is acoustically flat does not sound like it has equal energy in each 1/3 octave band. It is exceedingly instructive to put a reference microphone in front of a good loudspeaker and equalize the speaker output for flat response. If you then listen to the loudspeaker with your head in the same place you will be very surprised by the timbre you hear. Not all the timbre problems will be due to the loudspeaker! The important point is that the timbre we hear from sound sources depends strongly on our past experiences with that sound source and its apparent direction. We have no way of knowing the real timbre of pink noise.

EQUALIZING HEADPHONES

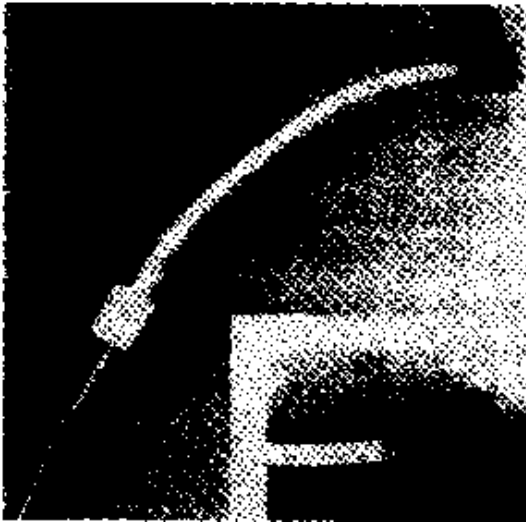


Figure 14: Probe microphone used by Spikofski. Ours was similar, including the open cell foam to keep the probe tube centered in the ear canal.

The procedure we use is relatively simple. First, we set up a high quality loudspeaker in a large well damped room, and measure its frontal pink noise response with a B&K 4133 microphone and a dbx RTA1 1/3 octave analyzer. We then attempt to measure the transfer functions (the frequency responses) between the loudspeaker and the eardrums of a test subject. The loudspeaker is moved around slightly during the test to give some spatial averaging, but is in front of the listener.

Our first measurements used a tiny microphone inserted in the ear canal. This method did not give us data which was accurate enough to yield either frontal localization or correct timbre. This technique might have worked if the microphone fully closed the ear canal - but the complex response of the microphone in the canal and the very large changes in that response if the microphone moved just a little bit masked the response of the pinnae we were trying to measure. We then glued a 1/16" OD thin plastic tube onto the microphone, and inserted the tube deep into the ear canal of the subject. The major frequency response

anomalies due to the tube were removed with equalization. The microphone is very similar to the one used by Spikofski. (Meade Killion makes an excellent (and gentle) probe microphone.)

The probe tube is inserted as deeply as possible, so it is right next to the ear drum. It needs to be as deep as possible, because otherwise the huge notches in the response curve from reflections off the eardrum will mask the notches due to the pinnae. [See figure 10] A frontal response curve is made and stored, with a slight spatial average. A pair of earphones is carefully put on the subject. Pink noise is played through the earphones, and another curve is made. The previously measured response of the loudspeaker is subtracted from the frontal response, and the earphone response is subtracted from the difference. The resulting curve is the equalization necessary to correct the phones. A 1/3 octave equalizer is adjusted to match this curve as close as possible. The subject then listens to pink noise and fine adjusts the relative balance of the bands for the sharpest image, and adjusts the treble balance if necessary.

When this equalization is finished normal stereo recordings usually appear to be outside the head and in the horizontal plane - at least after a short period of accommodation. As we pointed out before, custom earphone equalizations made this way are wonderful for balancing recordings. Timbre is also very good. You can take the headphones off and listen to loudspeakers with almost no change in apparent timbre - a first for me. As a recording tool individually equalized earphones are great.

DIFFERENT EARPHONES

We tried equalizing several different phones. It helps greatly to have a smooth response to start with. A pair of STAX SR3 on the ear phones sounded quite nice. A pair of AKG 240 phones needed little change, and were quite good. A pair of Sony MDR 50 phones sounded decent although a bit metallic. A pair of Sennheiser 222's (a closed design) needed a lot of EQ, but were OK once we were through. A pair of Koss ESP-10 electrostatics needed a bit of damping foam between the phone and the ear before they were useable, but after this modification they came out pretty well. All these phones sounded similar after equalization, and amazingly different before.

IMPROVEMENTS NEEDED TO OUR TECHNIQUE

We have been using a dbx RTA-1 as the analyzer. The infinite integration time and the comparison mode in this analyzer are especially nice. The measurements are fast, but somewhat higher resolution than 1/3 octave would be useful, as would a computer interface. We hope to soon put together some digital analysis equipment which would allow measurement of both amplitude and phase. We have been using a RANE midi programmable 1/3 octave equalizer which allows instant switching between the curves of different individuals. This feature is quite useful. An interesting feature of these experiments is that the equipment needed is not particularly expensive or uncommon. All you really need is a little patience.

INDIVIDUAL DIFFERENCES

How different are the equalization curves needed for different individuals? We have not measured enough people to know, but they appear to be quite different. About 40% of the listeners who have tried earphones equalized for my ears report they work pretty well, at least after about 10 minutes of listening. The rest are sure they do not. Sound sources which are supposed to sound in front remain stubbornly

above or behind the head. We developed individual curves for several of these subjects, and all reported frontal localization and OHL except one. The individual who did not get OHL still liked the timbre.

What can be done to match different people better? It might be possible to make a collection of response functions, and let a listener select two which work for them. A large number might be required. There are no duplicates among the 5 or so curves I have obtained so far. Each individual greatly prefers their own. Selecting curves might also be difficult. In my experience the elevation of a music image typically does not change as equalization is varied. The mind latches on to a previous position too strongly, and any change is interpreted as a change in timbre, not position. It is not possible to just listen to music or noise and try to get it to sound correct by adjusting the timbre. I have tried this for hours, and it is hopeless. I have to remove the phones and let my ears rest for sometimes as long as several hours before I can be sure I am hearing a new medial position. When the equalization is not very accurate several listeners have reported that by closing their eyes and strongly imagining the musicians in front of them the image does move to the correct position after a few minutes. This training is not permanent. Sources return to their incorrect positions after a few hours rest. It appears that by far the best way to select curves for a particular listener is to be able to measure the ear response of that listener. Such measurements are not as hopeless as it may seem. The probe microphone made by Mead Killion is so light it does not hurt to push it against your eardrum, and the equipment needed to automate the process could be included in a digitally based earphone driver. Hopefully the future will bring a practical product which can do the job. Even if for reasons of liability insurance the measurement must be performed by a licensed hearing technician many of us would probably give it a try.

Until that day it is enormously frustrating as a recording engineer to have a playback method capable of superb results, but which prevents my friends from hearing the latest results without running them through a tedious and tricky procedure. [Recently we developed an automated system that allows an individual to equalize a pair of headphones in about 7 minutes through loudness matching of 1/3 octave noise bands, using 500Hz as a constant reference. They first find an equal loudness curve for a frontal loudspeaker, similar to the method used for the ISO equal loudness curves. They then perform the same experiment wearing headphones. You subtract the two curves to find the proper equalization for that individual, and the results are perceived as very good.]

DUMMY HEADS AND THE UNIVERSAL DUMMY

After this discussion it is probably obvious that the pinnae response of a dummy head should accurately match that of the listener, and we should all make casts of our crania. True enough. But there is hope. The forward direction has the most complex and unique frequency spectrum. The illusion of sources in other directions is far less difficult to achieve. If we insist that a simplified but adequate equalization for the forward direction be contained in the earphones, the response of the dummy head in the forward direction should be as flat as possible. The response from above the dummy should be the difference between the over the head response and the frontal response of the listener.

For any given listener this difference is just as complex as the frontal response, but I suspect that this complexity need not be exactly reproduced. In my experiments gross errors in equalization frequently cause an apparent above the head localization, and I suspect this direction can easily be simulated from a smooth alteration of an average frontal response. The same is likely to be true of sources behind the head. Precise localization of these directions is not necessary for full enjoyment of music.

Thus if our goal is to make recordings which will satisfy a large number of people, and also be capable of the highest quality playback on loudspeakers, we must think of separately equalizing BOTH the dummy and the headphones for a simplified flat frontal response. Ideally we should also spatially equalize the dummy by boosting the L-R signal at low frequencies, which will improve loudspeaker playback and permits mixing accent microphones. The headphone amplifier should correct for the inherent LR boost when either compatible binaural or ordinary stereo is played, and match the equalization of the earphones to the pinnae of the listener.

The universal dummy should do the following:

1. Have completely smooth frontal response.
 2. Vary the treble response smoothly in the horizontal plane - increasing treble to a maximum of about +5dB at 90 degrees in the horizontal plane.
 3. Duplicate the interaural time differences of a natural dummy for sources in the horizontal plane.
 4. Increase the response at 8kHz smoothly as sources move up in the vertical plane, reaching a maximum at about +75 degrees.
 5. Roll treble response smoothly but steeply down as sources move behind the head.
 6. Simulate the effect of a torso in the downward direction by sharply attenuating treble as well as upper midrange in that direction.
 7. Be non controversial in shape, preferably a transparent sphere or cylinder about 20cm in diameter.
- This last requirement may in fact be the most important.

I have been using a modified Neumann KU-81i dummy for about 2 years now, both for recording sessions and for concert work. It is not beloved of musicians, conductors, or stage managers. I have been forbidden to use it in many instances. It is disturbing and distracting to both the musicians and the audience to have a vaguely anthropomorphic head hanging by a string behind the conductor. It is also just plain too large. We are working on a simplified dummy built with the present pinnae and microphones mounted on a U shaped curved sheet of clear plastic. Current results have been quite favorable. The "Kunstkopfgeist" appears to localize vertically as well as the more anthropomorphic head, and has smoother horizontal localization in front. The Geistkopf has a forehead, but no chin, and the rear of the head is open. Visually it is a great improvement, and appears to be acceptable to those in charge.

We do not know how to make an ideal head at present. For the time being we will continue to use cast pinnae equalized for roughly flat forward response. We select a pair of pinnae for the dummy head which do not have wide and deep notches in the horizontal forward direction, and equalize the response of each ear independently for flat forward 1/3 octave response with a spatial averaging of ± 10 degrees. The dummy is quite effective with loudspeakers or custom equalized earphones. Sound sources in front of the dummy appear to be in front, rear sources are behind, and height is adequately reproduced. Two major problems remain. The timbre in front is not really flat enough. For broad sources such as chorus or orchestra the timbre can be very good, but with a soloist the particular response of the pinnae in that direction can sometimes be heard on loudspeakers. The major remaining problem is the apparent depth of sound sources. If you look around you will see that your colleagues have heads of all different diameters, and the degree of shadowing of the pinnae by the front of the head varies. The angle the ears stick out away from the head varies a lot.

We find that this angle is vital in reproducing the distance of objects in front of the listener. If we give our standard dummy an average ear angle, people whose ears are closer to the head will perceive all frontal sound sources as too close. People who have their ears further out will find the sources too far away. We have to live with this error for the time being.

Note that my suggestion for equalizing both the dummy and earphones to a spatially averaged forward response is very similar to the work of Theile, except he recommends using the diffuse field response as a standard. Clearly if the pinnae on the dummy match the listener's pinnae exactly it doesn't matter which standard we choose. However, the compatibility of the recording with loudspeaker reproduction is better with the forward equalization (ref - 00), and additionally when the earphones are custom equalized to the forward direction ordinary stereo frequently appears to be forward, well balanced, and outside the head. These are powerful arguments for our method. However I have found that for many pinnae the diffuse field response and the 1/3 octave spatially averaged frontal response are quite close. This is true both of our current dummy (which uses slightly modified casts of my own ears) and the Neumann KU8H. With luck Theile and I can avoid an argument on this point.

THE KUGELFLACHENMIKROFON (KFM)

Theile and the IRT suggested a microphone which would give the treble response changes and time relationships of a dummy head, but which would have flat frequency response in the forward direction. This microphone is now being produced by Schoeps. Unfortunately the response of the microphone from above, below, and behind is not at all like that of a dummy. We have experimented with our own design of such a mike, and find that indeed the timbre in the forward direction is very good. However both on earphones and loudspeakers the reverberation does not spread up and away from the music as much as a true dummy. Our version has a substantial roll-off to the rear of the head. In the version by Schoeps there is no roll-off in the frequency response of reverberation picked up from the rear of the microphone, which will further reinforce the tendency for the reverberation to occupy the same direction as the music. The continued popularity of the Neumann M50 in classical recording indicates that a rear rolloff is often useful.

Other alternate dummies also exist, such as the one offered by Crown. This may also be capable of good recordings when used with care. At least it definitely rolls off the treble of sounds from the rear. This mike does not reproduce directional cues for sounds above and below the head, and has an unusual forward directional pattern.

OUT OF HEAD LOCALIZATION (OHL) THROUGH REFLECTIONS

Several researchers have reported increased OHL when side wall reflections of a listening room are deliberately recorded or introduced after the recording with electronic delays. The reflections should have equalization appropriate for their apparent direction. In the experience of this author such reflections are indeed helpful, but accurate frontal equalization is more important. If the equalization or impulse response of the recording/earphone system is not well matched to the listener, the results are poor.

RESEARCH BY OTHERS

Many groups are working on the problems of OHL and dummy head recording. Klaus Genuit and the Head Acoustics people in Aachen came (some years ago) to very similar conclusions to mine. I believe

both their heads are essentially free-field equalized to the forward direction, although I do not know if they use spatial averaging. The results with their first head are quite good. They also equalize a pair of STAX Lambda phones to the forward direction (on someone's ears!). When I tried these phones some of their recordings localized to the front, and some did not. Their first head sounds very good - especially when compared to the unmodified Neumann. I have heard their new simplified head for only a few minutes, and cannot judge its quality. The Head Acoustics recordings are quite good on loudspeakers. They also have a computer based binaural simulator.

Dr. Alex Persterer and the AKG group in Vienna has developed a computer to make pinnae related recordings. They use ear data generated by Blauert and Lehnert in Bochum. When I heard the computer in Hamburg I did not get any frontal localization. In fact it was this lack of frontal localization with 3 different pinnae functions which got me interested in studying this problem. The person who demonstrated the computer to me said he got frontal localization if he installed his own ear data, and not otherwise.

Plenge and Theile at the IRT in Munich continue to be interested in this problem. Their original design for the diffuse field equalized head has merit, even though the Neumann KU-81i head does not perfectly match the prototype they developed. Their diffuse field equalized earphone standard may be the next best thing to individual earphone equalization, although I am convinced that averaging the response of many individuals eliminates much of the advantages of ear related equalization. My preference for forward referenced equalization over diffuse field may in fact be partly a personal bias. My ears appear to be less notchy in front than some. A fellow engineer always tilts his head forward for critical listening. He says he is aware of a deep notch in the response at 8kHz! Clearly you want to avoid such problems when you equalize earphones, even if some individuals will perceive ordinary stereo as coming from an elevated direction.

The convolvotron developed by Scott Foster, Elizabeth Wenzel, and Fred Wightman is amazing. I got decent forward localization even without head tracking. The same is true of the Digidesign based system of Bo Gering in Toronto. [I have come to believe that the effectiveness of head movement in establishing frontal localization is not useful. It allows even gross errors in timber to be perceived as natural. If such a system is used for critical balancing of a recording, or for evaluating acoustics the results will be compromised. I believe that if the dummy/headphone system is not capable of creating frontal localization without head tracking it should not be used for research.]

Soren Nielsen has been doing simulations at Alborg University in Denmark which are successful. He finds adding "hyper realistic" room reflections to the basic sounds is helpful in establishing frontal directions and OHL. The work of Kendall and Martins of Auris in Illinois seems exciting. They seem to be primarily interested in loudspeaker stereo, and particularly film and video. They have a demo tape which seems to give a successful height illusion for many listeners, along with increased lateralization when the listener is on the center line between the speakers.

Q-Sound in Calgary has produced a commercial for Coke which gives greatly increased lateralization when heard on the center line between the loudspeakers. There is no sense of height in this commercial for me, and the lateral illusion collapses to normal stereo when you move one foot from the center line.

CONCLUSIONS

Although duplicating some aspects of natural binaural hearing would seem to be of great advantage in many applications of audio engineering, the effectiveness of the results depends to an alarming degree on the physical characteristics of the listener. Loudspeaker reproduction of binaural stereo without eardrum measured correction for the listener's frontal HRTFs can seem effective, but it is not precise enough for acoustic research, and it requires the listener to confine themselves to a limited space. Individual equalization of headphones through measurements at the eardrum are more promising. Perhaps the most interesting conclusion from our research is that earphones are at least as individual as shoes. Attempting to make one size fit all has severely limited the realism we all expect them to reproduce. Hopefully progress in finding how to match earphones to users will be forthcoming, along with a truly loudspeaker compatible dummy head microphone.

REFERENCES:

1. Blauert, "Spatial Hearing" (MIT Press, Cambridge, MA 1983)
2. D. Griesinger "Equalization and Spatial Equalization of Dummy-Head Recordings for Loudspeaker Reproduction" J. Audio Eng. Soc. Vol. 37 #1/2 1989 Jan/Feb
3. G. E Kuhn "Physical Acoustics and Measurements Pertaining to Directional Hearing" (in "Directional Hearing" Springer-Verlag 1987)
4. EL. Wightman, D.J. Kistler, and M.E. Perkins "A New Approach to the Study of Human Sound Localization" (in "Directional Hearing" Springer-Verlag 1987)
5. G. Spikofski "The Diffuse-field probe transfer function of studio-quality headphones" EBU Review - Technical #229 June 1988
6. G. Plenge "Über das Problem der Im-Kopf-Lokalisation" Acoustica #26 Vol 5 1972 p 241-252
7. G. Plenge "On the differences between localization and lateralization" J. Acoust. Soc Am. Vol. 56, #3, Sept 1974 pp 944-951
8. K. N. Stevens, R. Berkowitz, G. Kidd Jr, D. M. Green "Calibration of ear canals for audiometry at high frequencies" J. Acoust Soc Am 81(2) Feb 1987 pp 470-493
9. G. Theile "On the Standardization of the Frequency Response of High-Quality Studio Headphones" Audio Engineering Society Journal vol 34 #12 (1986) pp 956-969
10. G. Theile "Das Kugelflächenmikrofon" Presented at the 14th Tonmeistertagung Munich 1984 - Bericht der 14. Tonmeistertagung pp 277-293. Bildungswerk des Verbandes Deutscher Tonmeister - Masurenallee 8-14, 1000 Berlin 19.
11. G. Kendall, W. Martins "The Evaluation of Digital Filters for Controlling Perceived Direction in Stereo Loudspeaker Reproduction" presented at the 81st convention of the AES Nov. 1986
12. K. Genuit "Investigation and Simulation of Vehicle Noise Using the Binaural Measurement Technique" Noise and Vibration Conference 1987 Traverse City, MI April 28-30 (This paper was supplied to me by Jaffe Acoustics)