IMPROVING THE LABORATORY REPRODUCTION OF CLARITY, PROXIMITY, AND LOCALIZATION

D Griesinger David Griesinger Acoustics, Cambridge, Massachusetts, USA

1 INTRODUCTION

Progress in acoustic research depends on laboratory methods that accurately reproduce a measured sound field. But to be useful any such technique must be verified by direct A/B comparison to an accurate reference, and such a reference is lacking. In our research the perception of proximity and he ability to localize and separate individual instruments is primarily dependent on the phase alignment of harmonics above 1500Hz. This alignment is disturbed by reflections from all directions. Thus it is critical that these phases be reproduced correctly in the laboratory, along with the level and direction of reflections that might disturb them. Many techniques, such as 5.1 surround, Ambisonics and wavefield synthesis have been tried, but in our opinion all of these have difficulty reproducing the perception of proximity, even before reflections are added. High order Ambisonics is in principle capable of reproducing a sound field exactly, but implementing it is difficult and expensive. The author has yet to hear a reproduction of a known hall with a second or third order system that is believable.

The recent recording and reproduction method developed in Finland by Tapio Lokki and his colleagues is capable of reproducing a convincing sense of proximity, as well as believable localization and hall sound. In our opinion the system works well enough to be very useful for acoustic research, and is already producing exciting results. The system sounds good and sounds real, but does it exactly reproduce the sounds we hear in a particular seat in a particular hall? How can we tell? We need a method that allows a rapid A/B comparison between the “real” hall sound and the reproduced one. With current technology there is no machine that can prove the issue. It requires trained listeners. But if the switch between “real” and “synthesized” can be made rapidly enough trained listeners can do this with precision.

Binaural recording and reproduction can supply such a reference. In principle, and in practice, recording the sound pressure at a subject’s eardrums and then reproducing that pressure exactly does reproduce the sonic impression of a scene. (Verifying a reproduction system would require taking headphones on and off, but this can be done rapidly with training.) This paper presents the problems with current practice in binaural recording and reproduction, and inexpensive methods that can overcome them. With individual headphone equalization and careful attention to detail external frontal localization without head tracking can be achieved for an individual listener, and acceptable judgments of proximity, clarity, and localization for most non-individually optimized listeners. The technique is portable, discrete, and startlingly realistic.

This preprint is intended as an addendum to another preprint in this conference titled “The effects of early reflections on proximity, localization and loudness”. The reader is advised to read that one first. In this preprint we will discuss binaural recording and reproduction in more detail. Background on the subject, and an explanation of the meaning of “proximity” can be found in the other preprint.

2 BINAURAL RECORDING AND DUMMY HEADS

In the previous preprint, “The effects of early reflections on proximity, localization and loudness,” the author describes how impulse response measurements from a dummy head can be manipulated to auralize concert halls. In this preprint we are interested in recording and reproducing actual sounds. The best way to binaurally record natural sounds of any type is through probe microphones on your own eardrums. The best known examples of this technique used steel probe microphones and required clamping the head of the subject. But a better way has been hiding in plain sight. In the

1980s Meade Killion of Etymotic briefly sold light probe microphones with silicon tubes, expressly for recording sounds from the eardrums. More recently the author found that technicians that fit hearing aids often use miniature probe microphones with soft, flexible tubes to match the eardrum pressure from a frontal loudspeaker to the frequency response of the aid.

We have developed our own version of such probes, but in our version the majority of the length of the tube is made of relatively stiff PVC. Only the tip is super soft. Figure one shows an example of a probe microphone with a very soft silicon tip that can be made in about an hour. Instructions for how to do it are on the author's web-page. In practice the microphone needs to be calibrated to be useful for obtaining impulse response data. Instructions for doing so are also on the site. In practice the length of the tube is adjusted so the tip rests comfortably just next to the eardrum.

Recordings made with the probe are played back through headphones. To make this work properly we record a sine sweep from the headphones to the eardrums with the same probes. Inverting the response made this way and equalizing the recording with it reproduces the recorded pressure at the eardrums exactly.

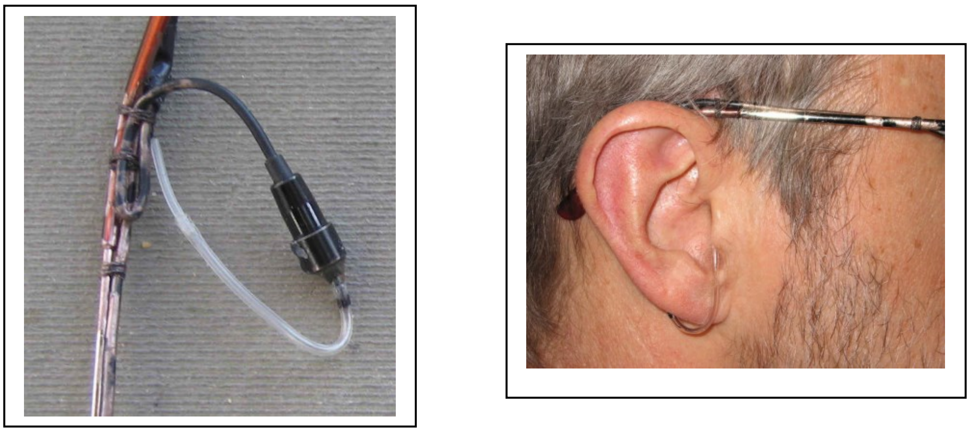

Figure 1: a miniature probe microphone constructed by the author mounted on the bow of a pair of glasses, and how the microphone is used in practice.

Figure 2: Calibrating the microphone and the headphone at the same time.

Ideally we would like the recordings to be equalized in such a way that they can be played on loudspeakers. To do this we need to know both the frequency calibration of each probe microphone and the frequency dependence of the sound pressure at the eardrum from a source in front of the person making the recording – his or her frontal HRTF. We prefer to determine both factors separately. To calibrate the microphones we tape the probe tips to a reference microphone and record a sweep with all three microphones at the same time. The probe responses are then inverted using the reference microphone as a standard. Matlab scripts and instructions for this procedure are on the author's web site.

Convolving the probe microphone output with the inverse found above makes the two probes match each other and the reference microphone exactly. We can now record the precise sound pressure at the eardrum to determine the frontal HRTF. Adding an inverse of that response to the probe equalization turns the human head into a microphone with a flat frequency response on-axis. We do not attempt to correct all the features of the front-to-eardrum HRTF. We leave the dips and notches above 6kHz alone, as they help the brain determine the vertical location of a sound source, and are important for creating a frontal sound image.

To do this we record a sweep from a calibrated frontal loudspeaker with the probes in our own ears, as close to the eardrum as possible. It is painless for the probe tip to touch the eardrum, and you can both hear it and feel it when it does. Pulling it back a millimeter or so prevents it from eventually itching, and is recommended. The frequency response obtained from this sweep is what we call the “expectation”, the sound pressure that we expect to hear from a sound source in front of us. If we invert this response, and equalize our recordings with both this and the inverse of the probe calibration, we can play our binaural recordings through loudspeakers. The recordings typically sound very good. It is in fact amazing how closely they resemble commercial CD recordings of similar ensembles.

The fact is that the human head and ear system is a very sophisticated microphone. A typical first-order microphone array, such as a soundfield microphone or an array like ORTF has much lower angular resolution that a human head. The combination of head shadowing which produces the interaural level difference (ILD) and the diffraction that produces the interaural time difference (ITD) gives the head a just noticeable difference (JND) for azimuth of about two degrees. The best a first-order microphone can do is about eight degrees.

As a consequence recording engineers need to put their microphones much closer to an orchestra than a human would prefer to sit. Binaural recordings from a good audience position, when correctly equalized and played back, can be stunning.

This is our standard equalization for our binaural recordings. If the world was perfect, these recordings would play back perfectly through any pair of headphones. The world is not perfect, as we will see.

A particular caution when using these probe microphones. It is ESSENTIAL that the probe tips be as close as possible to the eardrum when you record or make a measurement. If you are not willing to put something so far into your ears you should not attempt to use these microphones. One person, after going to the work of making the microphones, pushed the probe tubes in “far enough” in their opinion and recorded a sweep. It looked OK at first, but when the probes are not close enough to the eardrum there are deep notches in the response usually between about 8kHz and 14kHz due to the sound reflecting off the eardrum. These notches are different every time you insert the microphones, and they are not invertible. That person gave up using the microphones at this point. If you see notches at these frequencies in your calibration data, sweeps, or in your binaural recordings, you must learn to push the microphone probe tubes in further. We push them in until they touch the eardrum, and then back them off a bit. Some people may find the experience too frightening to continue.

Hopefully you will learn to do it. If not, don’t give up. There is another way. You can record with a dummy head, and learn to equalize the recording as described in the other preprint to make the dummy microphone flat on-axis. A method for equalizing headphones without using the probe microphones will be discussed below.

3 NOISE IN PROBE MICROPHONES

Small Microphones are inherently noisy. The ¼” capsules of the microphones we use have a much higher noise floor than a studio microphone. Adding a probe tube adds tube resonances to the response, and equalizing these resonances away with an inverse filter raises the noise level at frequencies that have dips in the response. But the recordings we make are surprisingly quiet. The reason is the gain of the pinna, concha, and ear canal for frequencies above 1000Hz. Evolution has provided us with a built-in system that amplifies the sound pressure at the eardrum at the frequencies that contain the most information in speech and music. This extra gain compensates for the increased noise at these frequencies. There are still problems. Most of the probes we use have a resonance dip at about 12kHz. Inverting that makes a peak in the noise at that frequency. Fortunately the author is old enough that he does not hear that frequency. A younger user might be well-advised to use an adaptive filter (Wiener filter) like the one in Audition, to remove this noise peak.

4 DUMMY HEAD RECORDING

Our dummy heads are fitted with our own pinna, ear canals, and eardrum impedance. They are equalized to produce a flat response to a frontal source as described above to make recordings that can be played on loudspeakers, or played back on ideal headphones. But high-quality commercial dummy head microphones can also be used. The equalizations of all of them are different, although some, like KEMAR and HATS are quite close to what we need. But you should check them. The best way to correct them is to play pink noise from a speaker in front of them, or to record a sine-sweep from a known-flat loudspeaker in front. Both for the expectation test and for dummy head calibration we like to record each ear one at a time, with the head offset in front of the speaker such that the pinna under test is in line with the center of the loudspeaker axis. This emulates a plane wave from the front impinging on the pinna, and minimizes the effects of head shadowing from a frontal source. A recording of frequency sweeps or pink noise can often be equalized by inspection to a useful flat response with a parametric or a 31 channel graphic equalizer.

Even a recording of orchestral music can sometimes be used as a signal to equalize a dummy head recording with the help of real-time spectral analysis and careful listening. A big orchestra playing full-tilt has a reasonably uniform, although gently rolling off, spectrum.

5 HEADPHONE EQUALIZATION

The other preprint for this conference points out that a headphone equalization that works for all users does not exist. People have different sizes of pinna, concha, and ear canals, and these differences affect both the sound pressure at the eardrum and the expectation response from a frontal source to the eardrum. The impedance of the eardrum also affects the sound pressure from a headphone, and has a lot of variance. The eardrum impedance can also depend to some extent on outer hair cell damage. There can also be considerable difference between the properties of the right and left ears. One can find in the literature wishful thinking that a very open earphone design can minimize the particulars of an individual, but that is not the case in our measurements. Open headphones have some of the largest variation in response among individuals.

Some headphones seem to have less individual variation than others, but we have found no headphone that is equalized properly for the author that produces frontal localization for more than about 60 percent of the people who try it. We need individual headphone equalization to accurately hear concert hall sounds. As mentioned in the other preprint, DIN standard 45619 tells you how to do it, although it is not easy. We have developed another way, based on equal loudness measurements, and have written a computer app that automates the process.

The app sends a sequence of noise signals to a loudspeaker that alternate between a reference 1/3 octave band at 500Hz, and a test 1/3 octave band at another frequency. The user compares the level of the two bands with a sound level meter, and adjusts the 1/3 octave graphic equalizer built into the app to until the test noise and the reference noise have the same sound pressure. A sound level meter is useful for this procedure, but in a pinch you might be able to use a smartphone. This procedure equalizes the loudspeaker to flat response.

The user then generates their personal equal loudness curve by doing the same thing while listening to the loudspeaker. This might seem difficult at first but most people can do it with a little practice to an accuracy of +-1dB. They then do the same thing while wearing the headphones under test. When testing the headphones the subjects are asked to not only make the two noise bands equally loud, they are asked to balance the test band so that it is perceived directly in front of them. The issue of balance is quite important, because it ends up compensating for differences in the two ears that do not show up when listening to a live frontal sound source.

The test for balance ends up compensating not only for differences in the conchas and ear canals of a subject, but also to some degree compensating for some types of hearing loss. The net result is that sometimes subject will hear a bit better through the headphones than they do with their own ears.

When the test is over the app allows subjects to listen to music of their choice through the newly equalized headphones, and switch the new equalization on and off at will. They can also listen to pink noise, either decorrelated or in mono. Mono pink noise is very revealing. When properly balanced the image is sharply localized in front. With un-equalized phones the mono noise image is more widely spread out. If the mono image is not perceived as sharply localized we often ask the subject to go back to the noise bands and try again to improve the balance. They usually succeed. They can also compare the pink noise through the headphones to the pink noise through the equalized speaker. If they two are not close we ask the subject to go through test noises again. If they are familiar with recording technique they can play the pink noise and adjust the equalizer manually with broadband pink noise. Some subjects – including the author – find this helpful.

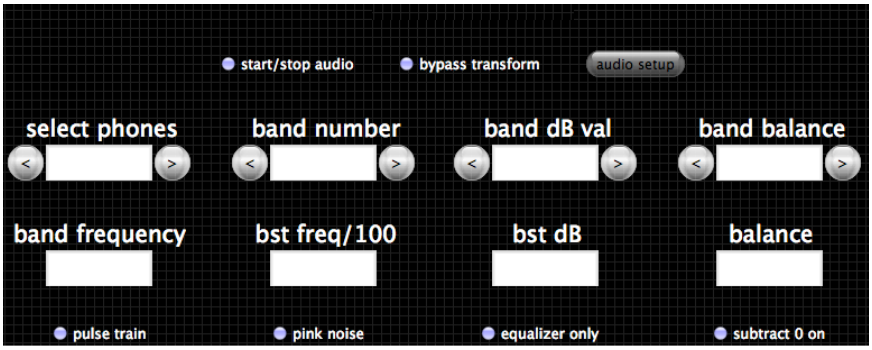

Figure 4: a draft of a possible user interface for a new version of the headphone app.

6 LISTENING

We have been conducting experiments with the current app, which is a Windows program that uses an ASIO interface in combination with a two-channel or multichannel external sound device. The app allows a subject to test up to five different headphones. The app stores the equalizations needed for each phone, and outputs .txt files with the subject’s equal loudness curves and the headphone equalizations. The curves, for a group of subjects in Finland, are shown in figure 3.

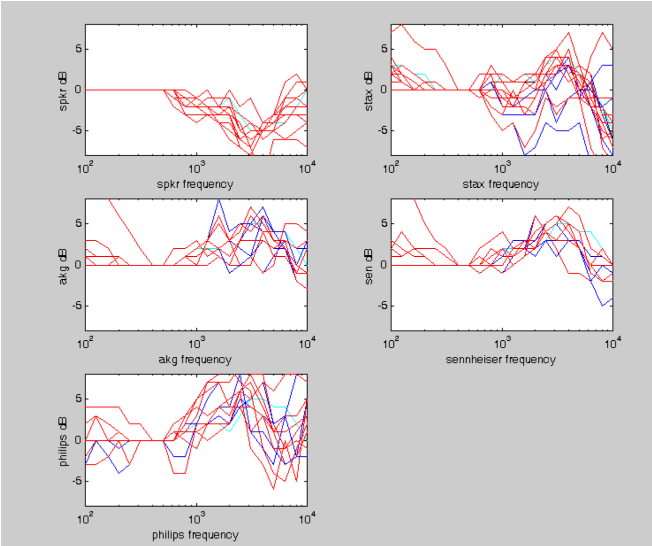

Figure 3: Top left – the equal loudness curves for the subjects in Finland. The other curves are for different headphones. An electrostatic phone by Stax, an AKG 701, a Sennheiser 250 noise-cancelling phone, and a Philips insert phone. Note that the curves are significantly different.

In the experiment that generated figure 3 we did not attempt to measure an equal loudness curve for the loudspeaker below 500Hz. The loudspeaker used a single small driver, which produced a soundfield with no interference between the LF and HF drivers. We currently use a small loudspeaker with very little interference in the front that is equalized to be flat to 120Hz. If the app is used with a common monitor speaker such as the Genelec 1029 or equivalent they are advised to have the subject sit at least three feet from the loudspeaker, with their ears exactly vertically positioned at the midpoint between the LF and HF drivers. Otherwise interference between the drivers can produce unreliable results.

It can be seen immediately in figure 3 that except for the Sennheiser on-ear phone the other headphones require quite different equalizations for each listener. In the author's opinion this is the reason head tracking is assumed necessary for headphone reproduction of binaural recordings. In our experience head tracking is only a crutch that gives the illusion of a sound image. The timbre of the image does not match the original, and sound quality judgments either of acoustics or the frequency balance of a commercial recording will not be accurate. Frontal localization and out of head perception requires accurate individual equalization. For reproduction of binaural recordings head tracking is not needed. Individual equalization of headphones is necessary.

The end result is amazing to all our recent subjects. Switching the equalization on makes a startling improvement to the sound. Just listening to FM radio – or these days YouTube performances – is a new and exciting experience. We are working on a version of the app that does not require the ASIO interface, and our subjects are eagerly awaiting it. It may be finished by the time of this conference. We are willing to make the current app available to other researchers. Send an email to the author.

The big advantage of the app for the purposes of this preprint is that it opens the possibility of verifying other methods of reproducing acoustic fields in the laboratory. If actual sound fields are recorded with a properly equalized dummy head at the same time measurements are made for later reproduction, these recordings can be A/B compared to the synthesized sound. We will finally be able to tell if the system is really working – and be able to fix it if it is not.

7 CONCLUSIONS

The graphs of headphone response functions in figure 3 show the variation between individuals for the frequency response of four different headphones. The variations are highly significant – yet the subjects were very happy with the sound of their individual equalizations.

For the purposes of this paper we are more interested in whether or not the procedure allows them to hear a binaural recording of a room or hall with accuracy. This question is currently under investigation by another group of students. My personal experience has been that the equalization procedure is successful. For me using the app is more effective in making a sharp central image, and a rock-solid sonic impression of a binaurally recorded performance than measuring the eardrum pressure with probe microphones and inverting it to equalize the headphones. Vertical localization also improves, at least for my personal recordings. Sometimes when playing a binaural recording of a fine performance I become mesmerized, unable for a time to pull myself out of it. I have seen this happen to others, to their surprise. These recordings can be very engaging.

It is almost too easy during a concert to become fascinated by watching the players or the conductor, or following some complex issue of the orchestration. Hearing a binaural snapshot of a really good performance insists that you concentrate on the sound itself. In a good seat in a great hall the sound is stunning – better than even the best stereo recordings. We believe this procedure promises to be an effective standard by which a loudspeaker-based hall synthesizer can be verified.