Buenos Aires – 5 to 9 September, 2016
Acoustics for the 21st Century…

PROCEEDINGS of the 22nd International Congress on Acoustics

Challenges and Solutions in Acoustical Measurement and Design: Paper ICA2016-381

Playback of non-individual binaural recordings without head tracking, and its potential for archiving and analyzing concert hall acoustics

David Griesinger [(a)]

[(a)] David Griesinger Acoustics, United States, dgriesinger@verizon.net

Abstract

Visually blind, rapid A/B comparisons are essential for studying auditory perceptions of all kinds, but are difficult to achieve in concert halls. Accurate binaural recording potentially allows instant A/B comparisons between halls and seats, and could verify that a laboratory simulation works as intended. But to be useful a standardized binaural recording needs to be accurately reproduced for a large variety of people. We have developed a computer-based loudness matching application that can quickly and non-invasively achieve individual headphone equalization with a variety of phones. When these phones are used to reproduce a free-field equalized microphone timbre is precisely preserved, and no head tracking is needed. The result is an uncanny impression of being exactly in a particular seat in a given hall. Halls and seats can be A/B compared with both live music and impulse responses from a virtual orchestra. These impulse responses can be manipulated to test the effects of different reflections, stage conditions, and reverberation times. By binaurally recording the sound in a particular seat a loudspeaker simulation of the same seat can be verified by taking the headphones on and off.

Keywords: Concert Halls, binaural technique, head tracking, individual equalization

Playback of non-individual binaural recordings without head tracking, and its potential for archiving and analyzing concert hall acoustics.

1 Introduction

It is well known in the audio field that it is nearly impossible to tease out differences in sound quality without instant A/B comparisons. To do this for concert halls requires that the sound in different halls and seats be exactly reproduced in a laboratory, with the ability to instantly switch from one sound to another. Binaural technology promises to provide such a technique, but it is currently considered both inaccurate and too cumbersome for practical use. The goal of this preprint is to show why and how the addition of individual headphone equalization can restore the promise of binaural technology as a simple and robust method for recording concert hall sounds, and for enabling the analysis of the acoustical properties that enhance them.

2 A brief history

Binaural technology for concert hall research has illustrious and almost successful origins in the work of Manfred Schroeder. [1] Schroeder understood that if the sound pressure at the eardrum of a listener could be precisely duplicated at a later time, the sonic experience would be recreated. Optimally this requires that the sound pressure at a subjects eardrums should be recorded, and then played back with a method that creates the identical pressure at the eardrums of the same subject. The author has developed discrete, comfortable probe microphones that do exactly this. He can attest to the success of the method through a great many recordings of good and bad concert seats around the world. The image is always frontal and needs no head tracking. [6]

Experiments at IRCAM have also shown that this can be done, and the result is startlingly successful. Instruments are correctly localized without head tracking, and the experience is convincingly real. [3] At IRCAM live sound was recorded from two probe microphones almost touching the eardrums. A head clamp was required, so head tracking was impossible, and not necessary. Schroeder utilized a dummy head microphone developed by Mellert [2] that duplicated the averaged eardrum pressure from a number of test subjects. To reproduce the sound pressures recorded by this microphone Schroeder utilized a crosstalk cancellation system that was individually created to perfectly reproduce the pressure at the eardrums of a particular listener. The same method was used at IRCAM. The listener’s head was again put in a clamp, and the same steel probes were used to measure the response of two loudspeakers in front. The crosstalk and the frequency response was then mathematically adjusted, precisely re-creating the eardrum pressure of the recording. Live music was reproduced without head tracking.

Schroeder was interested in manipulating impulse responses to find the effects of measureable parameters such as Reverberation Time (RT) and Clarity (C80) on preference. But any attempt to push a whole field in a new direction is fraught with serious errors, and his was no exception.

He required loudspeakers as sound sources, and anechoic recordings of music. He chose to use stereo recordings of orchestral music reproduced by two speakers on stage.

He was ahead of his time. Later work by many researchers, including Richard Campbell in Massachusetts and Kimio Hamasaki in Japan has showed that to convincingly reproduce the sound of an ensemble of musicians a separate loudspeaker or loudspeaker array is necessary for each musician.

Two speakers on stage do not sound like an orchestra. The spread of sound that a string section provides is due to many independent sound sources playing a once. More important is that a recording with multiple instruments playing similar music through the same speaker is incapable of reproducing proximity, the acoustic sense of closeness between source and listener that Lokki [9] is finding is a primary predictor of preference. Schroeder’s experiments all lacked the immediacy that proximity provides. His subjects had to concentrate on aspects of the sound of the hall in the absence of the sound of the music, which is very unnatural. But his use of binaural technology was accurate.

3 Spikofski, Theile, and Möller

Schroeder’s head clamps were cumbersome and impractical. But shortly after his work DIN standard 45-619 (June 1975) showed a practical, although tedious, method of achieving individual headphone equalization without probe microphones. The method is simple. A third-octave noise band is played through a loudspeaker. A subject rapidly puts a pair of headphones on and off while adjusting a third octave equalizer in the headphone channel until the headphones and the speaker are equally loud. The procedure is repeated until all third octave bands have been adjusted. This method adjusts the headphones to reproduce the complex ear canal resonances of a particular listener, just as Schroeder’s method did. Recordings from a dummy microphone equalized to be nearly frequency linear from the front should then be perceived with accurate timbre, and a binaural recording from such a head should be realistically perceived. In the author’s experience the method works well. The author’s dummy microphones and his eardrum recordings, equalized to match a frontal loudspeaker, sound natural on loudspeakers, and extraordinary on equalized headphones.

But DIN 45-619 was considered impractical. Several groups worked to find ways of equalizing headphones that would work for anyone. Almost all this work made the implicit assumption that the major problem with matching headphones to individuals lay in the directional components of the pinna. Theile and Spikofski introduced the concept of “diffuse field equalization” in an effort to remove the effects of directional dependence. [3], [4]. This work was expanded by Möller and Hammershøi, who exhaustively tested the frequency transfer from various points near the ear canal to the eardrum. [5] They divided the response into two components, a directionally dependent part and a directionally independent part. They found that when the directionally dependent component is measured with a blocked ear canal the response is less individually variable. But their data shows that there is enormous individual variation in the directionally independent component. See figure 13 in reference [5].

They make the argument – as does Theile – that it is the directionally dependent component that is important. The directionally independent component is either unimportant to localization, or is unlikely to be altered by the headphone. We find evolution has not favoured this argument. Evolution has found it quicker and more accurate to match the timbre detected *at the eardrum* to a set of relatively fixed reference patterns. If the ear canal resonances are altered by a headphone a match is unlikely to be found.

4 Ear canal resonances

Both the direction dependent and the direction independent timbre patterns are essential for perception of elevation. If the spectrum at the eardrum of a listener from a headphone does not match the timbre that individuals experience with live sound, the ear and brain system will be unable to determine elevation. In our view, in-head localization will result. And it does. The large variation between individuals of the direction independent component of the headphone response effectively prohibits a universally accurate headphone equalization. The fact that nearly any object that changes the impedance at the entrance of the ear canal will change the directionally independent component of the eardrum pressure makes the problem even more difficult.

Möller's graphs and our own data show that the concha and ear canal form a resonant horn, which boosts the sound pressure at the eardrum (for the author) of up to 18dB. The boost is evolutionarily important, as most of the information in speech and the environment lies at frequencies above 500Hz, and most of the sound pressure lies below that frequency. The pressure boost keeps the high-frequency detection area of the basilar membrane from being overwhelmed by low frequency sound. The horn also matches the impedance of the eardrum to the air outside the head. Without it our ears would be much less sensitive. But the tuning of this horn depends on its length, diameter, volume, the flair and volume of the concha, and the eardrum impedance. These are all highly individual.

Any such resonant system is also easily perturbed by changes in impedance at the opening. In our experience all headphones perturb it to some degree, and differently for different individuals.

5 Headphone equalization through loudness matching

ISO 226:2003 specifies methods to measure equal loudness curves for individuals. Plane waves of sine tones at different frequencies are presented from a frequency linear loudspeaker to a subject, alternating with a reference tone at 1000Hz. The subject adjusts the level of the tone under test until it is perceived equally loud as the reference. For most listeners the result is repeatable to +- 1 decibel. We adapted the method to equalize headphones. A subject sits in front of a frequency linear loudspeaker that alternates between tones or noise bands at a test frequency and a reference frequency of 500Hz. They adjust a 1/3 octave equalizer until the test signals match the loudness of the reference. The equalization that results is their individual equal loudness response. They then put on the headphones under test and find an equal loudness curve for that headphone. The difference between the two loudness curves is the desired headphone equalization. They can then listen to pink noise or music of their choice through the equalization they have just found. The timbre of pink noise and commercial recordings is accurate, and my binaural recordings can be startlingly real.

Their equalization settings for a 1/3 octave graphic equalizer are written as a .txt file, along with their equal loudness data. The app also creates a .wav file of an impulse response of their equalization that can be convolved with music to equalize that pair of headphones. We find it is additionally useful to have the subject balance the perceived left-right azimuth of the headphone tones. Not all ears are the same, and neither are headphone drivers. People with some mild hearing loss in one ear also find the balancing procedure very useful (including the author.) The user can select to use sine tones, noise bands, or filtered harmonic tones. They all give similar results.

Our procedure requires a reference loudspeaker, particularly for frequencies above 250Hz. We find lower frequencies can be assumed to be equally loud. With the help of a calibrated cell-phone analyser Inexpensive single driver speakers intended for computers can be equalized with the app. Such speakers, the cell phone app and a calibrated microphone can all be purchased inexpensively.

We have conducted experiments with the headphone app with the help of Ville Pulkki at Aalto University in Finland, and Jonas Braash at Rensselaer University in the US. Students familiar with sound recording find the procedure easy and fast. Older or more naive subjects take more time to get facile, but they all can do it. The results have been uniformly good. Almost everyone achieves frontal localization. Distance perception is highly variable. The perception of presence is clear. Sounds that have proximity sound close. But distance at which the subject perceives these sources depends more on expectation than any acoustic cue. We believe this is also true of natural hearing. Experiments with the app are on-going at Rensselaer.

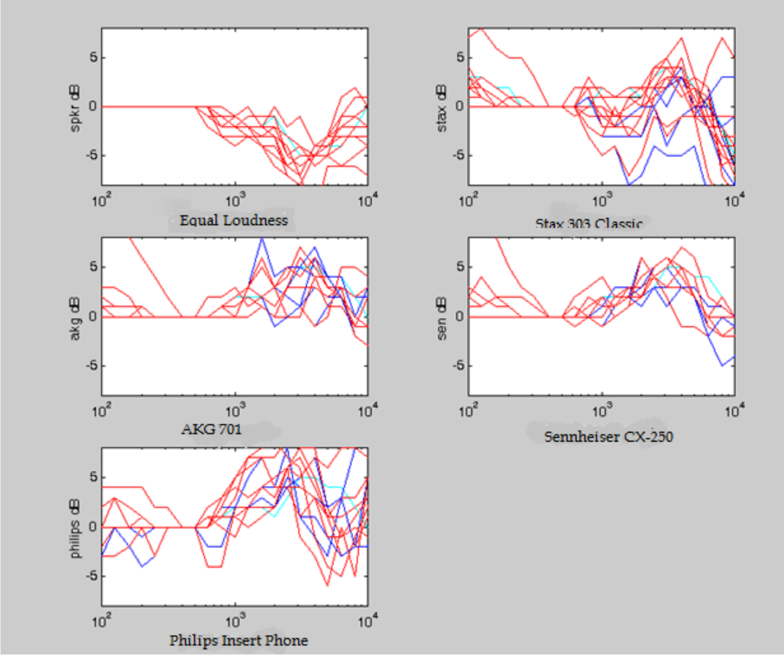
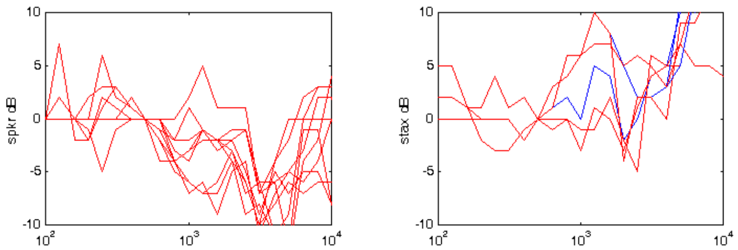


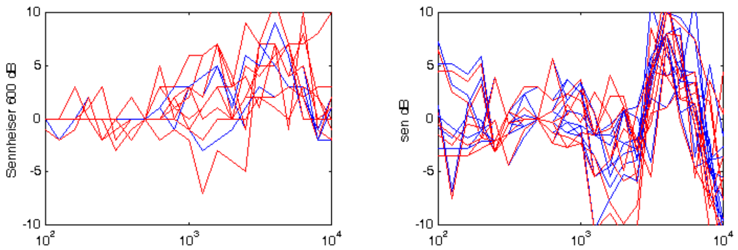
Figure 1: Data from Ville Pulkki's group in Aalto University of about 10 students. The first box on the left shows their equal loudness curves. Frequencies below 300Hz were not tested. The next

four boxes show the individual headphone correction curves for four different headphones. If the subjects needed a balance correction the left ear curve is plotted in blue.



Equal Loudness curves

Stax 303 Classic



Sennheiser 600

Sennheiser 250

Figure 2: Similar data from students at Rensselaer University in Troy New York.

Our goal was to have a group test a fixed set of headphones, in order to see if the extent of variation depended on the headphone design. We were looking for a headphone design that was more independent of individuals than others. As can be seen in the figures, we did not find one.

The data are interesting primarily in the great deal of variation that they show. All the headphones need correction of at least 4 to 5dB, and no headphone type stands out as less individually variable than the others. Informally we find the on-ear types such as the Sennheiser CX-250 or CX-250 II to be slightly more uniform, in part because they are more repeatable in frequency response when they are put on and off. We believe this data confirms our hypothesis that individual headphone equalization is necessary if timbre and localization is to be accurately perceived.

6 Accuracy versus preference

Not every subject prefers the individual equalization we find with this method, although the great majority (especially the young students) do. We test the equalization by playing pink noise through the calibrated speaker used for the equal loudness test, and having them listen to the same noise through the headphones. If the two are not perceived to have the same timbre we ask them to re-do some of the frequency bands. Eventually they find the timbres to be nearly the same. But they may not like it. Pink noise is not perceived as frequency flat. The large boost in the sound pressure at the eardrum that corresponds to the dip in the equal loudness curve at ~3kHz is audible, and some subjects might prefer a headphone equalization closer to equal-loudness. But it is not natural, and it does not result in frontal, out of head localization. It also occurs that a subject very familiar with a particular headphone is unwilling to initially accept the equalized phone, which seems midrange-heavy. But any doubt by a particular subject in our experimental result disappears when we play one of our binaural recordings of a great performance in a great hall. There is almost always a sense of "being there" and it is quite difficult to get them to turn it off.

Toole [7] has done numerous blind preference tests with loudspeakers, and finds that invariably with music recorded with frequency flat microphones that the loudspeakers with the most linear frequency responses both on-axis and off-axis are preferred by both expert and naive listeners.

7 Future work

Marko Hiipakka developed two techniques which could simplify individual headphone equalization. [8] With the help of a miniature intensity probe built with a Microfloan™ velocity sensor and a tiny omnidirectional microphone he was able to infer the pressure at the eardrum within one decibel from the pressure and velocity of the air at the entrance to the ear canal. In principle this could automate the detection of the frequency transfer function from a frontal loudspeaker to the eardrum without actually entering the ear canal. A calibrated reference loudspeaker is still needed.

He also developed a method using an insert headphone and a tiny omnidirectional microphone of determining the eardrum pressure from an insert headphone fitted with a tiny microphone. He combined data from the output impedance of the headphone and the pressure detected at the output of the phone to calculate the eardrum pressure. Combining the data from both the Microfloan measurement and the insert phone measurement enables that particular insert phone to be individually equalized.

This author does not know if the Microfloan method would still work in the presence of a close fitting headphone at the concha. If this is so, then other types of earphones might be automatically equalized. But, as for our app, a reference loudspeaker would still be needed.

8 Conclusions

Our experiments with individual headphone equalization are on-going. The author lacks access to legions of eager students, and the current academic structure discourages experimentation with new procedures and methodologies. In my life experience, any attempt to move a field in a different direction will be full of unexpected results. New experiments are unlikely to work, and the same for a second and third try. Students and professors seem to prefer more reliable paths to publishable papers. So publishable conclusions on individual headphone equalization are hard to come by. It is widely believed to be either unnecessary or impractical.

But it is simple to prove that it works. You need only try it. We believe that the binaural examples we provide on our web-page speak for themselves, but only if someone will listen. We believe that individual equalization of headphones through loudness matching results in accurate timbre, and usually with frontal localization without head tracking. I encourage sceptical readers to email me. I am happy to provide the app and some binaural examples to anyone who wishes to hear it for themselves. An additional benefit is that nearly anything you play through individually equalized headphones sounds more natural.

With all this in mind, we offer some personal conclusions from this work, which the author has be pursuing for at least thirty years.

1. Two loudspeakers on stage do not an orchestra make. In general, reproduction of individual instruments with multiple loudspeakers destroys the phase coherence above 1000Hz that is vital to proximity. Spatial aliasing from Wave Field Synthesis is severe above 1500Hz, and an accurate third-order Ambisonic system is extremely difficult to make. To electronically create an orchestra on stage you need a separate speaker and anechoic track for each instrument. To reproduce the sound in a laboratory you either need individually equalized headphone playback of a binaural recording, or a recording and playback system similar to Lokki's. [10]
2. The oversimplification of the sound sources used by Schroeder and others to interpret the acoustic measures enshrined in ISO 3382 has led to more confusion than success. We need to start over with more realistic sources and playback systems, paying particular attention to the perception of proximity, and the effect proximity has on the perception of the hall. Lokki is making progress in this field, but binaural technology with individual headphone equalization is simpler, less expensive, and possibly more accurate. It should play an important part.
3. We have shown in another preprint for this conference that with the help of binaural technology the sound of an ensemble on the stage of a hall can be created from a single binaural measurement. A binaural impulse response at a particular seat can be manipulated to create at least six different azimuths. Convolving the result with Lokki's anechoic recordings and listening with individually equalized headphones can be convincing. The effects of different reflections can then be studied.
4. The search for a universal equalization for headphones has not produced any headphone design that reliably results in frontal, out of head localization. Such an equalization probably does not exist.
5. The presumed need for head motion for accurate localization is contrary to ordinary experience, and would be strongly discouraged by evolution. While there are occasional front-back errors without head motion these are rare. If head motion is required for frontal localization with a non-individually equalized headphone we believe the headphone timbre is sufficiently incorrect that it will create errors of judgement both for acoustic research and for balancing a recording. Adding head tracking to an incorrectly equalized headphone only makes the errors in judgement more convincing. It does not correct problems with timbre.
6. In theory a headphone with zero impedance at the concha might excite the same ear canal resonances as a free-field loudspeaker. But such a headphone would excite the listener's own HRTF for a source at 90 degrees azimuth. This HRTF would lie on top of any HRTF from a dummy microphone, and the result would not be accurate.

7. We need much more experimentation with individual headphone equalization. The author had hoped to have a commercial cell-phone version of the app by the time of this conference, but this has been more difficult than expected. Anyone interested in working with this app should email the author. He is happy to send you a version you can use, particularly if you are willing to publish or send him your data. Currently the app requires a Windows computer and a two-channel ASIO compliant audio interface. It is not particularly user friendly, but has additional features – such as a means of detecting sensory-neural damage – that might be interesting.

References

[1] Schroeder, M. R.; Gottlob, D.; Siebrasse, K. F. Comparative study of European concert halls: correlation of subjective preference with geometric and acoustic parameters. JASA 56, number 4, 1974, 1195-1201

[2] Mellert, V. Construction of a Dummy Head after New Measurements of Thresholds of Hearing. JASA 51 number 4, 1971, 1359-1361

[3] Theile, G.; On the standardization of the Frequency Response of High-Quality Studio Headphones. JASA (34) 12 1986 pp 956-969.

[4] Spikofski, G. The diffuse field transfer function of studio-quality headphones. EBU Review – Technical 229 1988 pp 111-126.

[5] Hammershøi, D.; Möller, H.; Sound transmission to and within the human ear canal. JASA vol. 100, No. 1, July 1996

[6] Griesinger, D. Binaural hearing and headphones. A power point presentation originally written for a convention of the Verband des Deutches Tonmiesters. It is available on the author's web-site. Binaural Hearing, Ear Canals, and Headphones"

[7] Toole, F. The Acoustics and Psychoacoustics of Loudspeakers and rooms. Routledge 2007

[8] Hiipakka, M. Estimating pressure at the eardrum for binaural reproduction. A thesis from Aalto University. http://lib.tkk.fi/Diss/2012/isbn9789526048659/isbn9789526048659.pdf

[9] Lokki, T., Pätynen, J., Kuusinen, A., & Tervo, S. (2012). Disentangling preference ratings of concert hall acoustics using subjective sensory profiles. *The Journal of the Acoustical Society of America, 132,* 3148-3161

[10] Pätynen, J., & Lokki, T. (2011). Evaluation of Concert Hall Auralization with Virtual Symphony Orchestra. *Building Acoustics,* **18**, 349-366