PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Accurate reproduction of binaural recordings through individual headphone equalization and time domain crosstalk cancellation

David GRIESINGER[1]

[1] David Griesinger Acoustics, Cambridge, MA, USA

ABSTRACT

Accessing the acoustic quality of spaces of all sizes depends on methods that can instantly and precisely compare the sound of different seats and spaces. Complex systems using many loudspeakers have been developed that hopefully achieve this goal. Binaural technology offers a simpler solution. The sound pressure at a listener's eardrums can be measured with probe microphones, and reproduced with headphones or speakers calibrated at the eardrums. When carefully done the scene is precisely reproduced. But due to the variability of ear canal resonances such recordings and playbacks are highly individual. In this paper we present methods that are non-individual. Our recordings from a dummy head or from the eardrums are equalized to be essentially frequency linear to sound sources in front, giving the recording head the frontal frequency response of studio microphones. The recordings are then played back either through headphones equalized at the eardrum to match the response of a frontal source, or with a simple, non-individual crosstalk cancelling system. We equalize headphones at an individual's eardrums using equal loudness measurements, and generate crosstalk cancellation with a non-individual algorithm in the time domain. Software apps and plug-ins that enable headphone equalization and crosstalk cancellation on computers and cellphones are now available.

Keywords: Binaural, Headphones, Crosstalk

1. INTRODUCTION

This paper concerns methods for binaurally recording and later precisely reproducing the sound in a hall or room. The methods provide a simple and inexpensive way to instantly compare a particular seat in a hall to other seats, and to seats in other halls or rooms.

We have developed software that makes it easy to record and play back binaural signals. We are porting this software to a variety of platforms, including Windows computers, Apple computers, Android cellphones and IOS cellphones. The software reproduces binaural recordings with impressive realism, either through individually equalized headphones or through a pair of speakers in front of the listener.

Individual equalization at the eardrums is essential when headphones are used for critical listening, and especially important for reproducing binaural recordings. The pinna, concha, and ear canals form a resonant horn that can increases the sound pressure at the eardrum by more than 20dB, and these resonances are very different for different individuals. We find that the eardrum pressure at the eardrums from a frontal source can be different between individuals by more than 10dB for frequencies between 500Hz and 8kHz. We use a non-invasive equal loudness test to match the eardrum pressure at the listener's eardrums to the pressure from a frequency linear frontal loudspeaker. The difference in the sound is dramatic. Sound from any source is perceived as frontal and with the correct frequency balance. The equalization is different for each individual, and different for each headphone tested.

Listening to recorded music through equalized headphones is a revelation. Sound is perceived outside the head and frontal without head tracking. The timbre is accurate. Playing binaural recordings that have been frontally equalized with individually equalized headphones is even more amazing. The sense of being in the recorded space is overwhelmingly real.

Binaural rendition of binaural measurement data allows the sonic consequence of adding or altering individual reflections to be instantly heard. Hall measurement systems such as the ones by Lokki and

[1] dhgriesinger@gmail.com

Neil can be used in this way, but their playback systems are complex, and their data is not simple to modify. But binaural data – in combination with data from a spherical microphone – allows an experimenter to alter reflections with ease.

Our software also provides a means of reproducing binaural recordings without individual equalization through crosstalk cancellation. Crosstalk cancellation or transaural reproduction is not new. Our implementation uses an adjustable time delay and a set of semi-recursive time domain filters. The filters can be tweaked by the user if desired, but in most cases the default settings provide at least 10dB of crosstalk attenuation at all frequencies. The reproduction of binaural recordings is almost as good as using individually equalized headphones. The impression of being in the venue is strong.

A crosstalk playback system can be setup for only a few hundred dollars. Our software then allows acousticians and students to accurately hear different seats in a venue, and with our small library of binaural recordings compare them to examples of good and poor seats in halls around the world. A loudspeaker crosstalk cancelling system can be setup for only a few hundred dollars, and occupies very little space. The goal of this paper is to make such comparisons both easy and common.

2. IMPROVING ACOUSTICS WITH BINAURAL TECHNOLOGY

2.1 Difficulties with Sound Quality Assessments

There are inherent difficulties in determining the success of acoustical designs. Primary among these is the brevity of sonic memory. Toole found that results were not consistent if the switching from one speaker to another took longer than a few seconds.

The primacy of vision in human perception is also problematic. If you can see the performers you are sure you are also hearing them precisely. If you close your eyes for a few minutes the sound can change dramatically. In my experience very few acousticians test their designs by listening with their eyes closed or averted. Lokki and others are finding that without a visual image a primary predictor of preference is the sonic perception that sounds are “Proximate” - sharply localized and perceived as close to the listener. This judgement can only be made in the absence of a visual image.

Typical courses in acoustics are deaf to these issues. Students are taught how to calculate reverberation time, how to measure the standard acoustic parameters, and perhaps how to make acoustic models. They are not taught to listen and make an accurate description of what they hear. There is no opportunity to test the validity of current acoustic mythology. Binaural methods can provide this vital opportunity.

The author started to work on acoustics after many years as a recording engineer and as a designer of electronic equipment that allowed me to add or alter reflections in real time. Recording requires meeting stringent requirements for clarity, balance, and the sense of the hall. But I am still learning new things from listening to real halls with the techniques in this paper.

2.2 Difficulties with Current Binaural Hearing Research

There is a quote attributed to Einstein that is pertinent to the field of sound perception and acoustic design: “In every field of inquiry, it is true that all things should be made as simple as possible – but no simpler.” The columnist Sydney Harris added in 1964: “And for every problem that is muddled by over-complexity, a dozen are muddled by over-simplifying.” Research into hearing and acoustics fell into this trap.

The urge to simplify has done particular damage to research on the mechanisms human hearing uses to determine the timbre and localization of sound sources. It is widely and correctly believed that localization in the horizontal plane is largely – but not entirely – determined by interaural level differences (ILDs) and interaural time differences (ITDs). But there is evidence that differences in timbre due to the shadowing of the head are also important. I worked with an excellent sound engineer that suddenly lost all hearing in his left ear. Within a few months his brain had adapted to using timbre cues from his remaining ear to determine azimuth. His ability was diminished, but it still worked.

It is also widely believed that cues for vertical localization are based primarily on timbre, but we believe that the separation into horizontal localization and vertical localization is misleading. A source in space generates ILDs, ITDs, and timbre cues simultaneously. The brain has evolved to use all this information to form a multidimensional matrix which takes all of it in and comes up with a perceived direction of the sound.

ILDs and ITDs are binaural perceptions, depending largely on perceived *differences* between the two ears. But timbre is an overall gestalt. It depends on the absolute strength of each frequency band in

both ears simultaneously. The ear and brain are constantly rebuilding the matrix of frequency spectra the two ears perceive as we go through life, trying to make the auditory perception of direction match the perception from the eyes and other senses.

These spectral maps can only work with the data that the ears provide – and the only place the data is captured is at the eardrum. The spectrum at the entrance to the eardrum is meaningless to human perception of the timbre, or spectrum of sound.

The concha and the ear canal form a highly resonant structure – a horn that focuses high frequency energy strongly on the eardrum. If we eliminate or alter these resonances the timbre of external signals becomes unrecognizable to the brain. If all humans had the same resonances the problem might be solvable, but they do not. The ear canal resonances are highly individual.

Figure 4 from a paper by Hammershøi and Møller (1) plots the frequency difference between the concha and the eardrum for twelve subjects. Although the resonances are similar between 1kHz and 7kHz, where most sources have sufficient energy, they have differences of 10dB or more between each individual as well as differences in the sharpness of the resonance. Damping these resonances with headphones or frequency domain crosstalk cancellation makes the determination of source direction by timbre impossible.

The paper by Hammershøi goes on to plot the large effects on these resonances when different headphones are worn. In spite of the clear message both these figures, the paper concludes that the best place to measure sound for testing any kind of hearing effect or the quality of headphones is at the entrance to the ear canal, as if there were some magical way the brain could perceive the spectrum at this point.

They imply that we do not have to consider individual variations in the eardrum spectrum when we are studying binaural hearing, or listening to headphones. We find the conclusion untenable. We strongly believe that to reproduce any sound with headphones, and not just binaural recordings, with the correct timbre and localization we must take individual ear canal resonances into account. If we want to reproduce *binaural* recordings with headphones it is mandatory to match the playback system to the individual at the eardrums.

2.3 Difficulties with Current Acoustic Theory

For nearly a century hearing has been studied using sine waves, clicks, and noise as test signals. By themselves none of these signals are useful for speech communication. Humans and other creatures encode information in the timbre of the upper harmonics from low frequency tones.

Powerful sounds can be generated by interrupting a flow of air at a distinct period with vocal cords, lips, or reeds. The pulses of air create a rich harmonic spectrum that can be seen on any spectrograph. But spectrograms do not show that the harmonics thus formed are phase-locked. Once in every period the harmonics are constrained to combine with each other to recreate the pulse that formed them.

Our neurology has evolved to detect these pulses. Using them humans can determine azimuth by ITD alone over the entire audible frequency range. (2) Even more important, filters tuned to the fundamentals of the pulses can separate the formants of individual speakers or musicians from each other and from noise. The author has been studying the acoustic consequences of these pulses and building models that can perform pitch-based source separation since 2004. (3) He is not alone. Recently papers are proliferating on the importance of "Periodicity" in separating vowels in complex environments. (4)

When we perceive the pulse waveforms of individual sound sources our attention is drawn to them. But acoustic reflections modify or eliminate the pulses that draw our attention, as well as making it impossible for us to choose which conversation we follow in a crowd, or separately hear three lines of a string quartet. Pulses in speech and music waveforms were not recognized or studied because research concentrated on spectrograms for speech, and research into hearing used sinewaves.

Barron and Marshall's classic experiment on early reflections used live subjects in a laboratory listening to *monaural* supposedly anechoic recordings of an orchestra from a *single* loudspeaker. There was no phase information for single instruments nor any azimuth information. They found that without artificially added early reflections the recording was sharply localized and did not sound like an orchestra. Adding lateral reflections widened the image. Strong lateral reflections have been considered essential to good concert acoustics ever since. Our experiments have shown the conclusion is false. Recent work by Lokki, (5) Neal, (6) uses a single speaker array for each instrument, makes measurements to each seat from each speaker, and auralizes the measurements with separate anechoic recordings. They are finding the vital importance of proximity for listener preference. Their systems

are expensive, complex, and beyond the resources of most acoustic classrooms. Binaural methods are simple, accurate, and cost-effective. We need to use them for research and for teaching.

2.4 Early Binaural Recording and Playback

Manfred Schroder (7) studied concert halls using a dummy head microphone developed by Mellert. (8) Although the head was anatomically accurate, it was equalized to match the *average* of the frequency response at the eardrums of *ten* subjects from a *laterally* placed loudspeaker. Playback was accomplished by crosstalk cancellation measured at the listener's ears from two frontal loudspeakers.

Schroder failed to recognize the importance of proximity because he used just two speakers to emulate an orchestra, playing a supposedly anechoic stereo recording of a number of musicians. It is unlikely that the phase information of the instruments was preserved.

Another consistent error hearing research and in binaural reproduction is ignorance of the variations in the ear canal resonances of subjects. Headphone reproduction almost always fails to produce a frontal image without head tracking, and head tracking does not correct timbre. Crosstalk cancellation calculated from probe measurements in the ear canals also often disturbs or eliminates these resonances.

Experiments at IRCAM as reported to me personally by Eckhard Kahle used probes at the eardrums to record live sound around a listener. The recordings were played back with frequency domain crosstalk cancellation with the same probes, also at the eardrum. Eckhard reports this worked very well – and it should. But the recordings are individual, and the system requires a head clamp.

2.5 Recent Work with Early Reflections using Binaural Techniques

The author has demonstrated the power of binaural technologies using a binaural data set we made in Boston Symphony Hall with Ning Xiang and Leo Beranek. While Ning and his students were setting up we recorded impulse responses in seven seats from a small loudspeaker near the conductor's podium. We recorded the impulse responses with one of my dummy microphones and a Soundfield microphone of my own design. The impulse responses were equalized using the early energy as a reference, and were auralized using Lokki's anechoic recordings of a Mozart aria. I had made binaural recordings of live concerts in some of the same seats. The sonic impression from the impulse response renderings were convincingly accurate.

One seat near the right side wall was distinctly poorer than the others. Using a Matlab script I modified the binaural impulse response to extract the sound from that reflection and render it as a separate file. Deleting the reflection improved the sound dramatically. A Power point on the author's website has audible examples of this test. (9) In the other seats the sidewall reflections were either inaudible or detrimental. One of the best seats sonically was near the center of the first balcony 38 meters from the stage. There are no first order reflections from the orchestra to that seat, as the sidewall reflections are blocked by the first balcony at the sides. If you move about seven meters to either side of the center this is no longer true, and the sound is quite noticeably poorer.

2.6 Recording Binaurally

The key to making a binaural recording is to record from either a real head (your own) or a reasonably anthropomorphic dummy head microphone that has been equalized to be frequency linear from the front. The author has been recording with his own head in venues all over the world using discrete probe microphones that almost touch his eardrums. There is a YouTube video which describes how to make these probe microphones. (10)

Microphones are available from several manufacturers that sit in the concha. I have not tested them extensively. They are difficult to hide. I previously recorded binaurally with small microphones taped to my eyeglass bows just above the pinna. There is a some loss in the reality of the playback compared to the probe microphones at the eardrums, but these recordings are still useful.

I personally have used the Neumann KU-81 and the Head Acoustics HSU III. These heads sound quite different if you simply record with them, but once they have been carefully equalized to be frequency flat to a frontal source up to about 7kHz the sound impression is similar, and similar to my recordings with probe microphones. I have three dummy microphones. Two have accurate castings of my pinna and ear canals. One of them has a resistance tube for modeling the eardrum impedance. It was made to calibrate headphones before I learned that it only worked for me. One of the dummy's uses the pinna and equalized microphones from the Neumann KU-81. All these heads sound similar when frontally equalized. The key to making an accurate non-individual recording is to de-individualize your dummy or eardrum microphones by equalizing them flat from the front. Do not

trust the manufacturer's specifications. I have never found two heads from different manufacturers that match each other or match the frequency response of a carefully equalized frontal loudspeaker.

The equalization need not be perfect at high frequencies. For most dummy heads a 1/3 octave graphic equalizer will give good results. However, if you are using probes on your eardrums I find a parametric equalizer gives better results, because the main concha and ear canal resonances need two overlapping parametric filters to properly equalize. I do not attempt to remove the notches in the frequency response which appear in my own ears or my dummy heads between 7kHz and 10kHz. You need these notches for realistic playback, and equalizing them away makes a nasty peak in the background noise.

2.7 A Frequency Reference is Necessary

Our software app includes a stereo 32 band 1/3rd octave Q=5 graphic equalizer, a pink noise generator, and a generator of sharply filtered 1/3 octave noise bands. You can use the app to equalize your dummy microphone to a frontal source or to equalize headphones, but you need a loudspeaker and a means of measuring its frequency response. Both headphone equalization and our crosstalk cancellation require a speaker or speakers that are frequency flat on-axis, and you need a way to measure this. Most if not all student laboratories can provide a real-time analyzer with a calibrated microphone. If one is not available there are now inexpensive real-time analyzer apps for cell phones. Dayton Audio sells an inexpensive calibrated microphone for cell phone analyzers, the iMM6 for about $20. (11)

Using the iMM6 with the Android app "Audio Tool" is easy and very successful. The "Audio Tool" app for IOS phones also works but is more difficult to match to the microphone. There is a set of instructions at the link in reference 11 for how to do it. There are instructions for using our app to equalize the reference loudspeaker and the dummy head on my website. Many students interested in acoustics will already know how to do it.

We are in the process of finding a way to eliminate the need for a real time analyzer to provide a reference for equalizing the frontal loudspeaker. The app software generates readable text files for all the data that it stores. We might crowd-source response correction curves for a number of small loudspeakers. We are also looking at using the small MEMS microphone on Apple earbuds as a reference for equalizing the loudspeaker. This might be done automatically. Stay tuned.

2.8 Using Our App to Equalize Headphones To an Individual.

Our app was written primarily to make it possible to equalize headphones to an individual by using equal loudness measurements to match the response to the eardrums from a pair of headphones to the eardrum pressure from frontal frequency flat source. Detailed instructions for how to do this are on my website, and are demonstrated on a YouTube video. (12)

When phones are individually equalized the difference in the sound is enormous. Music, both binaural and stereo, is frontal and beautiful. Surprisingly perhaps the best sounding headphones are not the expensive circumaural models favored by high end audio buffs. Some little earbuds such as the ones sold by Apple for iPhones are smooth and free of complicated resonances. After equalization they can sound very good indeed. Earbuds if they fit well also produce the same frequency response each time you insert them in your ears. Apple buds do not fit me well. I coated a pair with a thin layer of silicone glue, which keeps them from falling out. I use them frequently.

2.9 Crosstalk Cancellation – Transaural Playback

Reproducing binaural recordings with crosstalk or transaural methods has a long history. A summary of many methods can be found here: (13) The Bose corporation has designed a system similar to ours (Bose Audition) for reproducing binaural renderings from models. The system is large, proprietary, and expensive. Our system is portable, cost-effective, and works well for reproducing binaural renderings and binaural recordings of actual halls.

Our crosstalk cancelling system requires two speakers at head-height in a roughly equilateral triangle with the head. Many small powered two-way loudspeakers can work well. The ones I have been demonstrating are relatively inexpensive Audio Engine A2+ speakers. They are attached to a pair of NEEWER camera tripods. After equalization for flat frontal response with a real-time analyzer and our app and they sound very good, and play surprisingly loud.

2.10 Crosstalk Cancellation Design

Our Crosstalk cancellation is accomplished with a semi-recursive filter. It resembles in many ways

the “Panorama” circuit in the Lexicon processors, but it is specifically designed to work in the near field, and the filter is more complex. The crosstalk amplitude and delay can be adjusted with controls in the app to fit an individual head.

Many researchers use crosstalk cancelling systems that place probe microphones near or in the ear to measure the crosstalk at the two ears. The measurements are then inverted in the frequency domain to precisely cancel the crosstalk. When the matrix that results from the measurement can be inverted the cancelation is nearly perfect. But it is not clear where to put the test probe microphones. Schroder’s paper says he put them at the eardrums of the listener. If so, the cancellation system would eliminate not only the crosstalk, but also the ear canal resonances that are essential for hearing localization by timbre. The system used at IRCAM definitely used the eardrums as the reference. In their case the recording was also made from the same probes, so the recording had the ear canal resonances the cancellation would remove. The system worked well, but both the recording and the playback were individual.

We are interested in making a system that is non-individual. We want to eliminate the crosstalk at a point near the entrance of the ear canal without modifying the individual listener’s own ear canal resonances. It is possible to do this with a much simpler crosstalk circuit than matrix inversion.

To make it we measured the crosstalk from our loudspeaker from the left speaker to the right ear, and from the right speaker to the right ear. We then compared the two measurements to find both the delay and the frequency response that is needed to precisely cancel the crosstalk from the left ear to the right ear by sending a signal from the right speaker to the right ear. We then designed a filter that gives the needed frequency response and amplitudes.

The Lexicon Panorama program approximated the needed filter with a simple single pole, single zero shelving filter, adjustable from about -6db to -10dB over a frequency range of about 300Hz to 2kHz. In our current configuration we found that using two such shelving filters spaced out in frequency worked better than just one.

Crosstalk needs to be recursive. The correction signal sent from the right speaker to the right ear to cancel the crosstalk from the left speaker will also travel to the left ear. There needs to be a signal that will cancel that crosstalk too. The goal can be accomplished by sending the cancelling signal from the right ear through an additional delay and filter to the left speaker. And then again to the right, and so on.

In many crosstalk circuits the recursion is automatically performed by summing the output of the first delay and filter with the opposite channel input and with a negative sign. Many ways to do this are shown in reference 11. But we found that filtering the signal once and sending it back around changed the high frequency amplitude and phase unacceptably. The circuit works better if we send the signals to a second set of delays and filters before recirculating it. The result is slightly more complex, but more accurate and more stable.

The app provides four broadband signals for adjusting the crosstalk: decorrelated pink noise, mono pink noise, noise in left channel only, and noise in right channel only. The decorrelated noise should sound wide and outside the speakers, and the left only or right only noises should be heard only in the proper ears. For most listeners a small adjustment to the delay gives attenuation to the opposite ear of at least 10dB. As a final refinement I added a five band user adjustable parametric equalizer to the crosstalk filters. A user can send the third octave noise bands available in the app to the left channel only, and listen at each frequency for the degree of cancellation in the right ear. If a particular band is not as effectively canceled as another the user can adjust the filer response with the equalizer. So far such individual adjustment has not been needed.

Subwoofers are not necessary for most binaural work. But we recently found two smallish subwoofers on ebay for under $100. Adding them to the system with a crossover of 90Hz and re-equalizing the combination made an impressive improvement to a recording from Boston Symphony Hall when the organ joined the orchestra.

2.11 Using Crosstalk on Stereo Music

The equalized small speakers and the crosstalk cancellation turn out to provide a wonderful way to listen to music of all types. But recordings intended for stereo playback often contain bass energy that a large L-R component. Binaural microphones record very little left minus right (L-R) information at low frequencies. One of the main functions of a crosstalk canceller is to increase the L-R level enough to re-create the ILD and ITD heard in the original venue. This increase in L-R is excessive on stereo music, where low frequencies may have an L-R signal almost as great as the L+R (mono) signal.

Playing stereo music through a binaural crosstalk canceller without a limiter can overload the loudspeakers and sound too spacey. I added a circuit to the app to limit the L-R excursion of the loudspeakers at frequencies below 300Hz when the L-R amplitude of the loudspeaker signals exceeds the amplitude of L+R amplitude. With the limiter bass is heard properly on all kinds of music. A similar circuit was present in “Panorama”.

I have recently used the small crosstalk system with the subwoofers to master stereo recordings and videos. The system is compact, the sound is carefully equalized and balanced, and the imaging is precise. The sound of recordings mastered this way is excellent when I play them in my studio.

Picture 1- A crosstalk system for binaural rendition and mastering

2.12 Conclusions

We have developed a series of software apps that enable a listener to match a pair of headphones to their own ear canals. The app is currently available at little or no cost for a wide variety of platforms, and we are gradually improving the user interfaces. When headphones are equalized at the eardrum to match the timbre of a frequency-flat frontal loudspeaker reproduction is dramatically improved for all music and for a wide variety of phones. When reproducing non-individual binaural recordings it becomes easy to compare different seats in a given hall, and to compare them to other venues. The crosstalk cancellation in the app is excellent for demonstrating or comparing binaural renderings or recordings without individual equalization. A subject can just sit down and listen. The crosstalk system is simple and inexpensive to assemble. We hope that the binaural techniques described in this paper will be widely used, both for acoustic research and for music listening.

REFERENCES

1. Hammershøi, D. Møller H. Sound transmission to and within the human ear canal. J. Acoust. Soc. Am. 100 (1), July 1996

2. Blauert, J. Spatial Hearing. MIT press 1983

3. Griesinger, D. Subjective aspects of Room Acoustics. Proc RADIS conference, Japan, 2004. www.davidgriesinger.com/RADIS_2004.pptx

4. Moore, B. C. J. The roles of temporal envelope and fine structure information in auditory perception. Acoust. Sci. & Tech. 40, 2 (2019)

5. Lokki, T., Patynen, 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

6. Neal, M.T.,Vigeant, M.C. A measurement database of US and European concert halls for realistic auralization and study of individual preference. Proc Institute of Acoustics Volume 40, 2018, 27-38

7. Schroeder, M. R, Gottlob, D, Siebrasse K. F. Comparative study of European concert halls. JASA 56, number 4, 1195-1201 (1974)

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

9. http://www.davidgriesinger.com/Accurate_reproduction_ASA_Boston.pptx

10. https://www.youtube.com/watch?v=2yYFND4lbAs

11. http://www.daytonaudio.com/index.php/imm-6-idevice-calibrated-measurement-microphone.html

12. https://www.youtube.com/watch?v=1VaKZr208Dc

13. http://www.fabiokaiser.eu/wordpress/wp-content/uploads/TransauralAudio_Report_FabioKaiser2011.pdf