

Test Listening Genelec's Aural ID

by Matt Foglia CAS

A number of years ago, the good folks from Smyth-Research (Smyth-Research .com) came to Middle Tennessee State University, where I am a professor. They brought their A8 Realiser in for us to demo. The purpose of this hardware box was to allow the perception of listening over loudspeakers (up to eight channels) while actually listening via headphones. Peter Damski CAS wrote an article in the Fall 2011 *CAS Quarterly* about the unit and included a quote from me describing the representation as “mind numbingly accurate.”

What did Smyth-Research do so accurately? First off, they took a personalized HRTF (Head-Related Transfer Function) of the individual in the room they were listening in. They did this by placing small microphones in the listener's ear while seated in the listening position. Each speaker emitted test signals that were picked up by the microphones in the listener's ears and took into account the HRTF of the listener. (Our setup at the time was 5.1, though the new A16 Realiser can handle 16 tracks for immersive representation.) The unit took into account the room acoustics, speaker placement, and listener's HRTF relative to that space with precision strong enough to make you question whether the headphones were on or off your head.

While the Realiser is extremely impressive as a hardware solution, portability is often the reason we're mixing on headphones in the first place. As such, software solutions are very desirable and even more possible given the processing power of personal computers these days. Some companies have tried to emulate loudspeaker listening through headphones with the use of plugins. Two that come to mind are CanOpener by Goodhertz (goodhertz.co) and Nx Virtual Mix Room by Waves (waves.com).

CanOpener (get it?), which is limited to stereo, uses a

crossfeed algorithm that allows some of the left-channel signal to feed the right side of the headphone and vice versa. It also allows for adjustment of “speaker” angles, eq, and room sound. The user can alter these based on taste to try and match a listening environment they're familiar with.

Nx is more sophisticated, allowing for up to a 7.1 emulation. Waves also allows you to input your head circumference and inter-aural arc (distance from ear to ear around the back of your head) to gain a closer representation of how the user's head may affect perception. Adjustments to eq, angles, and room can also be made. Like the Realiser, Nx even offers head-tracking to add a level of realism to the listening experience as you move your head.

So, what is Genelec trying to do with Aural ID? What they're not trying to do is give you what these plugins and hardware do.

What they are doing is creating the fingerprint equivalent of how your head and auditory system affect the way you hear.

Given that the company has spent 40-plus years developing and improving the accuracy of their loudspeaker monitors, accurate headphone monitoring was a natural step.



Smyth-Research A8 Realiser

Here are a couple of questions I was looking to answer when researching Aural ID. What do you need to do on your end? What do they do with what you give them? What do you get? And what do you do with it? Here's a look into my experience.

CAPTURING | In order to create an accurate HRTF, one would, traditionally, have to put special microphones in their ear canals (not dissimilar to how the Realiser captures this) and record test signals in an anechoic chamber to remove the effects a room has on direct sound. The subject would have to be extremely still during the process so the orientation is consistent. Given the unlikelihood of people actually doing this, Genelec created a very clever approach for capturing this information using readily available technology.

The technology needed is a video camera and a person to operate the camera. For me, it was my iPhone and my teenage son Michael. If you want to illicit a strange reaction, tell your kid, "Hey (insert said kid's name), I need you to video my head and pay particular attention to my ears." While not the cinematography debut he was looking for, Michael recorded my head by walking slowly around me as I sat on a stool in our garage (see photos on the next page). Genelec provides a good video example of how to do this with enough detail for the video to be effective. It's very easy to record this too quickly. Slow and steady are your friends, though you may have to introduce the concept to your kid. You then put a ruler up to your ear and take a picture to show its length and to provide a

measurement reference for the folks at Genelec (again, insert odd reaction from kid). The video and the ruler photos of your ear are then uploaded to Genelec's site.

PROCESSING | Back in Finland, Genelec's team uses software to create a 3D model of the head recorded in the video and seen in the photos. Once this is created, they use an acoustical simulation to calculate the acoustical field. The process measures the equivalent of 836 different orientations. Think about that for a second: 836. This allows for hundreds of aural origination points (e.g., monitors) to be represented in the HRTF.

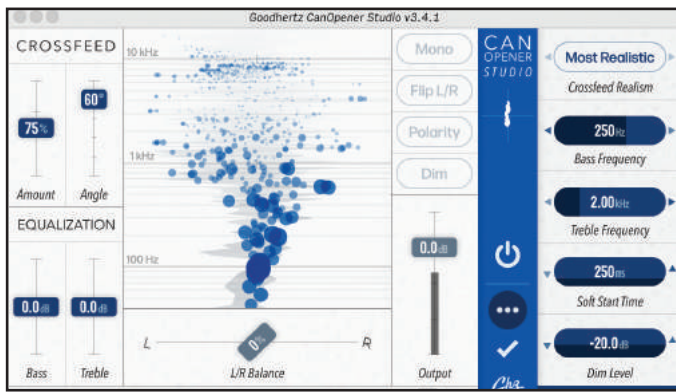
DELIVERY | Once Genelec has completed their processing, you are sent .SOFA files at three different sampling frequencies: 44.1 kHz, 48 kHz, and 96 kHz. The file to be used is based on your session's sampling rate. If you're unfamiliar with the .SOFA format, it stands for "Spatially Oriented Format for Acoustics" and is a format standardized by the AES to store acoustical data such as HRTF and binaural or directional room impulse responses. You can read more about the file at www.SOFaconventions.org. With Aural ID, this file contains your HRTF.

INCORPORATION | Once I received the file, I was curious to hear what it would do. To use the file, you need a plugin or a program that can read a .SOFA file. Pro Tools does



www.WBSound.com

©2019 Warner Bros. Entertainment Inc.



Goodhertz CanOpener

Waves Nx Virtual Mix Room



Getting ready to record a video of my head.

depth added to the source material relative to just listening on headphones. Normally, the spatial content on headphones can be very extreme at times but engage the Aural ID and the imaging narrows as if to incorporate Interaural Time Differences. For kicks, I adjusted the azimuth to 45 degrees and the imaging became closer to what the headphones naturally

reproduced. Phasing became noticeable with some speech on occasion. For one show I sampled, the narration's vertical position seemed to elevate slightly higher with the Aural ID engaged compared to when bypassed. When I tested some of my music reference songs, I could really hear how hard panned instruments and vocals were compared to when I bypassed the Aural ID.

I took a number of breaks over the days I experimented with Aural ID. If I came back to a session and listened first with Aural ID engaged and then bypassed, I preferred the Aural ID characteristics. If I first listened with it bypassed, I would initially prefer it that way. I'm looking forward to trying it out with surround and immersive sources.



Top: Exciting shot of my left ear. Bottom: When the only ruler you can find is a pink one from your daughter's elementary school days.

CONCLUSION | My initial thought when I decided to look into Aural ID was that I'd find something that could more accurately replicate mixing on monitors. However, by itself, that is not the point of the Aural ID as it does not include room characteristics. If you were to also incorporate an impulse response of a specific

Matt Foglia (CAS Quarterly) Journalist Case		
Name	Date Modified	Size
Matt_Foglia48.sofa	Jun 11, 2019 at 6:27 AM	13.7 MB
Matt_Foglia96.sofa	Jun 11, 2019 at 6:27 AM	27.4 MB
Matt_Foglia441.sofa	Jun 11, 2019 at 6:27 AM	13.7 MB

My .SOFA files at three different sampling frequencies.

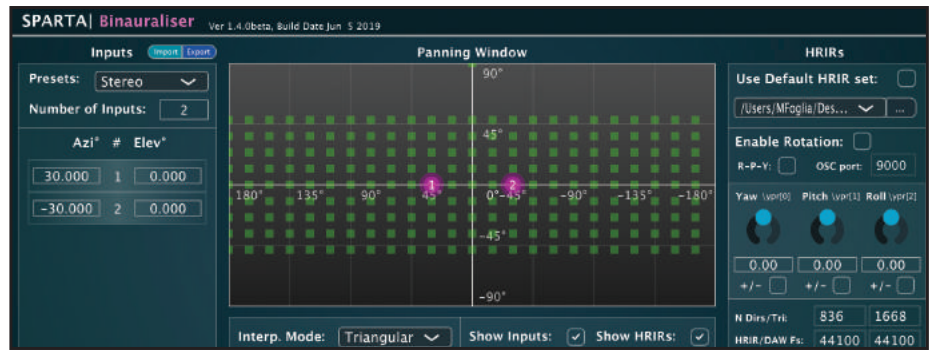
not natively do this, so I used one of the recommended plugins Genelec listed in their literature, the free SPARTA/COMPASS Binauraliser (how's that for a name?) plugin from Aalto University (research.spa.aalto.fi/projects/sparta_vsts/). Unfortunately, this is not available in AAX format, so I also used DDMF Metaplugin (ddmf.eu) which can convert AU and VST format plugins to AAX. This is a very useful tool given that these are common plugin formats but aren't supported by Pro Tools.

Inserting Metaplugin across my stereo program master and then selecting the SPARTA Binauraliser, I imported my .SOFA file that matched the sampling rate of my session. As you can see in the snapshot of the Binauraliser, the default azimuth is 30 degrees and -30 degrees. As you'd guess, this is the recommended "ideal" equilateral triangle listening angle when working in stereo. While I worked in stereo, Binauraliser allows 64 independently controlled channels and the Aural ID .SOFA file allows you to have tons so that you can replicate nearly all monitoring formats, as mentioned earlier. For all sources, you can adjust angles and many other attributes to your liking. Doing so for stereo, however, will yield results that are not in line with typical speaker positioning.

REACTION | It was interesting to engage and bypass the Aural ID since I didn't have a room IR file that I was trying to replicate listening in. There is definitely some

space (such as your dub stage) and process that IR with your .SOFA file, then you could replicate how you would hear things in that specific space. While this could be great, it is not the sole purpose of Aural ID, as you will learn if you read the following interview with Genelec's Director of R&D, Dr. Aki Mäkitvirta.

Using SPARTA Binauraliser to read the .SOFA file.



AFTER EXPLORING MY AURAL ID, I HAD A COUPLE FOLLOW-UP QUESTIONS. ONCE WE COORDINATED OUR SCHEDULES (GENELEC'S HQ IN FINLAND IS SEVEN HOURS AHEAD OF MY CENTRAL STANDARD TIME IN NASHVILLE), I SPENT SOME SKYPE TIME WITH DR. AKI MÄKIVIRTA, GENELEC'S DIRECTOR OF R&D.

WHO DO YOU SEE AS YOUR MAIN AUDIENCE FOR AURAL ID?

There is an application area that is particularly a good fit for this technology and that is game engineering. Games contain synthesized room effects and immersive audio content. Using a personal HRTF contained in Aural ID for these applications is a really straightforward way of increasing the precision and productivity of sound engineering work.

The second application would be engineering of binaural audio content. The personal HRTF in the Aural ID offers, again, a more reliable presentation and can increase productivity. In the case of binaural audio, the sound is presented using headphones. Today, audio is increasingly recorded using higher order Ambisonic microphone technologies. Aural ID suits perfectly to monitoring such recordings binaurally and offers improved reliability.

HOW ABOUT USING AURAL ID FOR TRADITIONAL MIXING?

Your thinking is in the direction of directly replacing loudspeaker-based conventional monitoring with headphone-



based monitoring. A lot of work is going on at the moment around the world to develop different aspects of such technology. A significant part of this work has been to reduce or remove the unit-to-unit variations of [different] headphones' sound characters, as this is usually the first problem that pops up. There is also other software that is essentially claiming to be able to create a number of virtual loudspeaker sources around the listener. Some software also offers a head tracker to update the HRTF in real time to create an illusion of the sound reproduction remaining stationary in the room when the head turns. Once these allow the use of personal HRTF data, the quality and reliability of these methods can increase.

HOW CAN USERS CREATE AN ACCURATE SPACE TO PUT THEIR AURAL ID "IN"?

At the moment, there are several companies that offer technology to record 3D impulse responses in rooms. Such technology can be used to capture the room influence in a way that can be used in rendering the room over headphones. Each early reflection has a certain level, timing relative to direct audio, and orientation relative to the listener. The information in Aural ID enables the presentation of this information correctly for the headphone listener. Naturally, a set of early reflections is unique to each audio source (loudspeaker) in the room. The number of early reflections to handle is significant, but technology suitable for handling this has been developed, for example, in the area of Ambisonic reverberation software.

HOW WOULD YOU SEQUENCE PLUGINS TO ACCURATELY REPRODUCE LISTENING IN A ROOM IF YOU HAD AN IR AND AN AURAL ID?

If we think of the correct order of arranging the signal processing steps, first we need the simulations of the real sources (loudspeakers) we want to listen, then we need the room-related early reflections generated for all these sources, and finally, we render all this audio information for headphones. The rendering uses Aural ID-contained information about the HRTF for all the directions of arrivals relative to the listener. Aural ID is a complete presentation of the HRTF for each ear individually, and it is calculated in 836 different orientations for each ear, covering the whole sphere around the listener fairly densely. •