

**THE SCIENCE OF SOUND**

Ivor Taylor, Technical and Finance Director at Grand Central Recording Studios (GCRS), offers his thoughts...

The Science of Sound

As someone who has a forty-nine-year career in audio production, I'm not going to surprise anyone by saying that I believe sound to be of utmost importance – and some of my recent reading has drawn me to the conclusion that scientifically, sound comes first.

A friend of mine sent me a link (see bottom of article) which reports that scientists have discovered that when you change where your gaze is looking, your eardrums make a low frequency sound of about 30Hz and they do this 10 thousandth of a second before your eyes actually move. To make this even more bizarre when you shift your gaze to the left both of your eardrums flex left and vice a versa when you look right.

The idea that your eardrums move on their own when you change your gaze and make a very low frequency-sound was strange enough, but that they do this in synchronisation just before your eyes moved seemed so far out there, that at first it seemed to be nonsense to me.

In a brilliant piece of work, scientists placed miniature microphones down the ear canals of 16 humans and a couple of monkeys until they were adjacent to the test subjects ear drums. When the subject's eyes moved, their eardrums made a low frequency sound. With eye tracking they could track and time the movement of the subject's eyes and then correlate this to when their ear drum made the sound. The end result of this research was that your eardrums make this sound 10 milliseconds before your eyes change gaze. Why is this happening?

No one has yet found out why, so from here on in this is pure conjecture on my part as to what is happening, and not scientifically tested facts.

You hear things before you see them – audio is processed quicker than vision. A crude measure is that the process of hearing a sound is twice as fast as seeing a light flash. However, when a door slams the sound of the door occurs exactly at the



same time as we see the door shut. That's the physical reality but the linked video clearly shows that the brain processes sound and vision at different rates. (Checkout this great David Eagleman/BBC video <https://youtu.be/C8k-lrJldw?t=1160>)

This video shows us that your sensing of reality is slightly delayed and that the sensing of sound is delayed less than the sensing of vision. However, as you experience both sound and vision as happening at the same time, your brain must be time shifting your perception of one, or both of these relative to each other.

Your brain isn't 'looking' at your visual data stream or 'listening' to your audio data stream. It is analysing these and from the analysis it constructs the spatial world you 'see' and the audio world you 'hear'. These are cognitive constructs, not videos or audio soundtracks. There is no cinema inside your head.

So this time alignment trick is not aligning sound to a continuous stream of data from your eyes, but aligning the sound to a cognitively constructed picture which your brain constructs and then updates every 40ms or so.

An easy assumption is that processing audio and visual data always takes the same amount of time – no matter what you are hearing or seeing. In the world of digital post production, that is not true as it is in the neurological wonder of your brain. Engineers use different processing algorithms (usually called plugins) depending on what they are trying to achieve. For example, the time delay produced by an audio reverberation plugin will be different from that of a compressor limiter. This is hidden from the engineer by the computers systems, so he or she never has to think about it.

When you shift your gaze, you do not know what you will see until your gaze has settled. Your gaze shifts, so you reset your cognitive analysis to analyse your audio and visual data to find the unexpected. The neurological analysis engine in your brain then readjusts and tries to extract as much information as it can, until you shift your gaze again, and another reset of your cognitive analysis occurs and so on.

An example would be on safari in tiger country. Your companion is amiably chatting away. You are hearing her and at the same time improving the intelligibility of what she is saying by unconsciously reading her lips. The guide suddenly makes a hand signal to stop, and points at a strand of grass that is moving. You look at the moving grass. Everything is quiet apart from the slight sound of the grass being pushed down by something. Is it reasonable to assume that at this instant your brain will not be trying to hear a tiger speak but desperately trying to correlate these grass sounds with the tiger's feet that are just starting to emerge into clear view.

You will change your visual processing from lip reading to visual processing for camouflaged animals in the undergrowth and your audio processing from optimised for speech to optimised for the sound of rustling grass. For me it seems reasonable to assume that the time taken for the neurological processing of audio and video data will vary depending on what you are hearing or seeing just as it does in the



world of digital audio processing. These neurological processing changes will have differing time delays.

The obvious concern is that, with these differing time delays, a person's audio and visual perception might go out of time synchronisation. But for your brain to make the best cognitive 'guesses', both audio and video constructs need to be time coherent or said simply 'in sync'.

So why do your eardrums make a low frequency sound?

I always assumed there must be a nerve process connecting the audio and visual systems to allow constant time sync, but with the eardrums sounds happening 10ms before you shift your gaze I'm not so sure. Any such neurological process would have to take a sync point from the cognitive visual process, mark that as being a point in time, and then find the matching point in time on the audio stream. How would it know what point in the audio stream exactly corresponds to the visual cognitive construct it is continuously generating.

A simpler way might be if you had in your head the equivalent of what is used in post production, a sound that occurred at a specific point in time along with the visual cue for the sound being made – bring on the Slate Board with its 'clapper' attachment. Shut the clapper in front of the camera and you get a sound (the sync plop) perfectly in sync with one frame of the film.

So I think that the eardrums sounds embed a 'unique' low frequency sound onto your heard audio data whenever you shift your gaze. It marks the start of the next analysis sequence for the brain as it assembles the visual cognitive construct. The brain assembles the first 'frame' of the cognitive construct of the scene you are looking at, and it knows that it should present that to your higher cognitive processes aligned with the audio data marked by the eardrum noise or 'sync plop'.

Why do you not hear this low frequency sound?

When you hear a sound, your ears are sensing small and rapid changes from high to low pressure in the air around your ears. For low frequency sound, both ears always sense almost identical changes in pressure. As the pressure from the low frequency sound wave increases both eardrums flex inwards into the centre of your head, and when the pressure decreases both eardrums flex outwards. That's how we hear low frequency sounds.

Something strange and different happens when you shift your gaze to the left and your eardrums make the 30Hz sound. Your left eardrum moves inwards to the centre of your head and your right eardrum outwards away from the centre of your head and then vice a versa when you shift your gaze to the right.

That's fundamentally different and in the audio world we would describe these movements as being 'out of phase' relative to the centre of your head. If you add out of phase signals together they sum to zero.



The result of this difference is that the audio data signal produced by the eardrum sound is unique from all of the other low frequency sounds which surround us all the time. That allows the brain to perceive this as a specific unique point in time in the audio data and when the left and right audio signals are added to together they add to nothing so you do not hear them. How your brain actually accomplishes this adding together and removal of the 30hz sound from your audio perception is another matter.

I'm sure I have committed many major sins in how I think the brain processes vision and sound. Please forgive me, I'm just a humble audio engineer.

Your sync plop generator is truly unique.