On the use of impulse responses
in high-end 3D audio virtual reality systems

Prof Guy-Bart Stan, Department of Bioengineering, ©Imperial College London

Acoustical spaces as LTI systems

Under the assumption that acoustical spaces (rooms, lecture theatres, etc.) are LTI systems, their **impulse response (IR)** for given positions of the sound emitting source and receiver/listener within these acoustical spaces constitutes their fundamental characteristic (1 IR per (source, receiver) pair within a given acoustical space). Using this signature, i.e. the IR(s) of the acoustical space, it is possible to give the impression to a listener that he/she is listening to an acoustic signal (sound, speech, song) as if he/she was (moving) in a different acoustical space. In what follows, we outline the principles through which you can immerse the listener in a synthetic 3D acoustical environment, as you would for vision-based 3D virtual reality.

How can we measure the Impulse Response of an acoustical system (rooms, loudspeakers and binaural impulse responses)?

The brain is easier to trick visually than acoustically. The slightest differences in the signal listened to can make a huge difference in terms of the 3D audio effect perceived by the listener. This is why, if the end goal is 3D audio virtual reality, it is important to be able to measure the impulse response of an acoustical system with high accuracy.

- To measure an IR, we can use:
  - **Impulse signals**, e.g. Popping balloons or gun shots
    However there are various issues associated with the use of impulse-like sounds to measure the impulse response of an acoustical space or a transducer (e.g. a loudspeaker):
    - The impulse-type sound produced is not very reproducible from one experiment to the next.
    - All energy is concentrated in a very short amount of time, which can either be dangerous to the ear that receives this impulse sound wave — or to the transducer (e.g. loudspeaker) that needs to produce the impulse-type sound.
    - Loudspeakers cannot produce a perfect impulse.
  - **Broad spectrum signals**, e.g. MLS or Sinesweep, and corresponding deconvolution techniques based on the convolution of the time response to these broad spectrum signals with their "inverse (a.k.a "deconvolution") filter".
- I have developed algorithms to precisely and efficiently measure the IR of acoustical spaces or acoustical systems and transducers (louspeaker + soundcard + microphone): JAES paper, 2002 on my webpage: `Comparison of Different Impulse Response Measurement Techniques, G.-B. Stan, J.-J. Embrechts, D. Archambeau, Journal of the Audio Engineering Society, Volume 50 (2002), n°4, pp. 249-262`.
- The IR of an acoustical space is the result of the various reflections on the walls and objects of the room by the sound waves emitted from the sound source. Because of these reflections, several sound waves arrive to your ears with different times of arrival depending on the path they took from the source to your ears.

- To measure the IR of an auditorium or an opera house you would thus typically place the source (loudspeaker) at the position of the speaker and the microphone at one or several positions in the audience. You would then capture the typical IR for this (source, receiver) pair. Each (source, receiver) pair yields a different impulse response for the acoustical space. Therefore, in general when the impulse response of an acoustical space is measured the loudspeaker is placed at the "average position" of the source (e.g. the speaker or orchestra), while the microphone is placed at the "average position" of the listeners.

True 3D audio virtual reality systems

```
(1) Impulse response of acoustical space * (anechoic) signal
=> "colouring" the (anechoic) signal to make it sound
as if it was played/recorded in that specific acoustical space

(2) Binaural Impulse Responses * (anechoic) signal
=> "spatialisation" of the (anechoic) signal to make it sound
as if it was coming from a specific direction
```

where * denotes the convolution operation.

(1) & (2) combined = 3D audio virtual reality

What are Binaural Impulse Responses (BIR)?

BIR consist of 2 Impulse Responses (1 for the left ear and 1 for the right ear) per position of the source around the head of the listener.

How does our brain use Binaural Impulse Responses to automatically infer the direction of incidence of sounds?

- The filtering done by the shape of your ears and the ear canal to any sound is specific to the angle of incidence of the sound when it reaches your ears.
- Your brain "knows" these various filters (it has a sort of filter bank) and thus can use this information to infer the angle of incidence of the sound perceived (the brain performs some form of deconvolution for this from its knowledge of the BIR associated with various sound incidence angles).
- BIR are measured by inserting tiny microphones in your ear canal or using a "dummy head" whose shape and ears are moulded to represent an "average listener".
- A BIR bank (1 BIR for each desired direction around the head of the listener) can then be combined with head-tracking systems so that when the head of the listener moves in a simulated 3D environment this can be taken into account by the 3D audio virtual reality system to "spatialise" the sound accordingly.

Examples of situations and applications where Binaural Impulse Responses are used

- Cocktail Party Effect (google it if you want to know more).
- Jet fighter pilots headphones use this technology to simultaneously emit spatialised sounds carrying different information:
  - weather on the right
  - communication with other pilots on the left
  - enemy alerts "spatialised" to reflect the position of the enemy around the jet fighter plane in real-time
- Consumer electronics products:
  - Some headphone manufacturers are starting to exploit binaural impulse responses to create personalised "perfect" sound through smart headphones that automatically measure your binaural

impulse responses and, based on them, create individualised equalisation filters to give you the best sound personalised for your ears (e.g. Nura headphones).

– Other headphone manufacturers use binaural impulse responses combined with head tracking to create real-time and immersive 3D audio virtual reality environments for gaming and sound spatialisation applications.

Remarks:

- When, during the demo in the lecture theatre, we play the convolved signal `Desired_IR * (anechoic) signal`, there are actually other additional convolutions happening: the convolution of the emitted signal with the impulse response of the room (the lecture theatre) in which the signal is emitted, and also the convolution of this resulting signal with the impulse response of the loudspeaker that is emitting the signal in the room:
  `signal received by each student = Desired_IR * (anechoic_)signal * Room_IR * Loudspeaker_IR`
  (remember that `*` denotes the convolution operation and that **`Room_IR`** is specific to a (source, receiver) pair and thus depends on where you (i.e. the receiver) are in the room (as the position of the loudspeaker is typically fixed)).

- **Room equalisation**: If the IR of the room, **`Room_IR`**, in which the signal is going to be played is known (e.g. it was measured) and **if the "inverse" of this IR can be computed**, then this inverse can, in principle, be used to eliminate the effect of the room IR. This is called room equalisation.
  `signal received by listener = (anechoic_)signal * inverse(Room_IR) * Room_IR (* Loudspeaker_IR) = (anechoic_)signal (* Loudspeaker_IR)` as, ideally, **`inverse(Room_IR)(t) * Room_IR(t)`**$=\delta(t)$, where $\delta(t)$ is the Dirac-delta.
  On the top of this you can add the sound "colouration" of any room for which the IR was measured, called hereafter **`Desired_IR`**. To do this, the signal emitted by the loudspeaker needs to be: **`signal to send to the loudspeaker = Desired_IR * (anechoic_)signal * inverse(Room_IR)`**
  When this signal is emitted in the room that is characterised by the impulse response **`Room_IR`**, the effect for the listener is "equivalent" to removing the acoustical colouration (or acoustical signature) of the room in which you are listening to the signal, **`Room_IR`**, and replacing it by the acoustical colouration (acoustical signature) of the room that you want to acoustically simulate, **`Desired_IR`** since: **`signal received by listener = Desired_IR * (anechoic_)signal * inverse(Room_IR) * Room_IR (* Loudspeaker_IR) = Desired_IR * (anechoic_)signal (* Loudspeaker_IR)`**. This should allow you, in principle, to listen to any sound signal as if it was played in the "desired room", e.g. cathedral, lecture theatre, opera house, etc.

  In practice, it is extremely difficult to compute a useful inverse filter **`inverse(Room_IR)`** with which to cancel the effect of the room in which the signal is played. This is because a room impulse response is the result of a multitude of reflections on the walls and objects in the room, and the slightest perturbation modifies the response, making it extremely hard to compute a "usable" inverse filter **`inverse(Room_IR)`**, i.e. a filter such that **`inverse(Room_IR)(t) * Room_IR(t) =`** $\delta(t)$.

  Because of this difficulty in computing a useful **`inverse(Room_IR)`**, this step is typically not performed and instead the sound emitted in the room is simply **`Desired_IR * (anechoic_)signal`**. Remember though that this means that the signal you are listening to is actually:
  `signal received = Desired_IR * (anechoic_)signal * Room_IR (* Loudspeaker_IR)`

  Note that in rooms that have good acoustical properties, i.e. they don't really colour the sound too much or have been designed specifically to allow faithful reproduction of the sounds that are emitted into them (e.g. sound/music studios, or, even better, anechoic rooms), there is little need to try to equalise their impulse response using signal processing.

- **Loudspeakers equalisation**: If the signal is played through a loudspeaker, for which the impulse response measured in an anechoic chamber is **`Loudspeaker_IR`**, the same equalisation method (based on pre-convolution of the signal with the inverse impulse response of the loudspeaker) can be used to remove the acoustical signature of the loudspeaker. This is feasible in practice and leads to: **`signal emitted = (anechoic_)signal * inverse(Loudspeaker_IR)`** which means that: **`signal received = (anechoic_)signal * inverse(Loudspeaker_IR) * Loudspeaker_IR * Room_IR = (anechoic_)signal * Room_IR`**

- **Headphones equalisation**: Again, equalisation can be used to compensate the IR of headphones: you would first measure the impulse responses (left and right) of the headphones and use these impulse responses to compute the (left and right) inverse impulse responses of the headphones. These inverse impulse responses can then be convolved with the signal to be sent to the headphones so as to eliminate the acoustical signature of the headphones themselves. This is feasible in practice.
- **The convolution operation can be algorithmically costly**. This is why, typically, convolution is performed in the frequency domain by multiplying the Fourier transforms of the two signals that one wants to convolve. This is most of the time more economical as there are very fast algorithms for computing the Fourier transform of a time domain signal (Fast Fourier Transform (FFT)) and the inverse Fourier transform of a frequency domain signal (Inverse Fast Fourier Transform (IFFT)).
  The cost of `real(IFFT(FFT(signal_1) .* FFT(signal_2)))` is typically lower than `conv(signal_1,signal_2)`.
- **Audio software that add sound effects (reverb, etc.)** rarely do it through the convolution process explained above as this requires quite a lot of computation (even when convolution is performed through a multiplication in the frequency domain). The result obtained from these audio software is ok but is far from the almost perfect results obtained by the convolution processes outlined above. Since the brain is not easily tricked acoustically the convolution processes described above are essential to create truly immersive 3D audio virtual reality systems.
- Now that powerful computers and Digital Signal Processors are becoming common and inexpensive, various companies have started using the convolution process described above (convolution with room IR and BIR) to colour and spatialise sounds in **immersive 3D audio simulators and games**, and for high-end audio equipment (e.g. headphones such as Nuraphone and Nuraloop, and high-end loudspeakers).
- If you want to try yourself some binaural recordings in real-life settings, I recommend to have a listen at some of those presented on this youtube channel (please use headphones or earphones when listening to binaural recordings): The Binaural Guy