The Amie Project – Part 1

This project began with a request from Skywalker to build a small, near field monitor that could be used in a small space to mix movie sound that would translate well to a larger space, the movie theatre. In order to explain the design of this new near field monitor, we need to define the criteria of measurements we are using. The frequency response (FR) of the speaker is all of the tones the speaker can reproduce and the phase (time) relationship of the tones to each other. The FR is made up of the magnitude and phase response.

In an editing room, the speakers would be placed one to two meters away from the sound mixer. For a 5.1 system, there would be five speakers placed around the mixing area, three in front of the mixer and two in back. The speaker would need to be able to reproduce frequencies (tones) from 42 Hz to 20 kHz. Using pink noise, each speaker would need to produce 96 dBc sound pressure level (SPL) continuously with 108 dBc peaks, with low distortion at the listening position. We will refer to this as the cinema spec. The near field monitors that we tested could only meet this specification at one meter. We would need 6 dB more output in order to meet the cinema specification at two meters.

Skywalker requested that a horn/waveguide implementation be explored to see if it would improve translation to typical horn/compression driver setups. In the cinema world, horn loudspeakers developed in the 1930s needed high gain to provide the high sound levels from low power drivers to fill the movie theatre. These horns produced high amounts of harmonic distortions. Modern waveguides using modern transducers have minimal harmonic distortion. Waveguides have the advantage over small direct radiator high frequency drivers, in that you have precise control over where the sound field goes. Small domes mounted on a baffle board have a very wide sound polar pattern.

If we take a speaker outside and mount it on a very tall pole above the ground, and mount a microphone one meter on axis to the speaker, we have what we call the free field measurement. If we send tones to the speaker and measure the sound pressure level of each tone the speaker can reproduce, we can plot the magnitude response. All we know at this stage is the speaker can reproduce notes (tones), but we don't know anything about the relationship of the notes to each other. That is, have some of the notes been delayed compared to other notes? Phase shift is difficult to measure and interpret. For example, the phase response of almost all loudspeakers doesn't fit into the 360˚ (±180˚) window, making it difficult to comprehend the meaning without more tests.

Let's look at another way to measure the speaker. Instead of looking at how the speaker responds to tones, let's see how the speaker would respond to a very short time pulse, called the impulse response (IR). (See Addendums beginning on page 6 for more information on the properties of the IR and how it can be measured.) This pulse contains all the frequencies. In order to make the pulse useful in the real

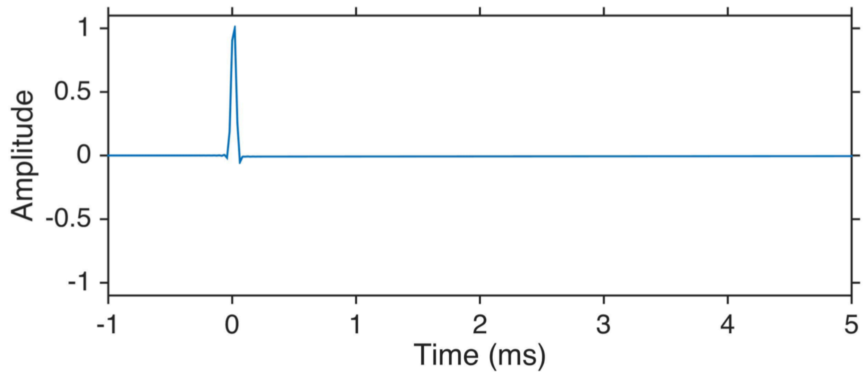

Figure 1. Ideal impulse response of a speaker band limited from 20 to 20000 Hz.

world, we will filter the pulse with a 20 Hz high pass filter to rid it of very low frequencies, and a low pass filter at 20000 Hz to remove frequencies above 20000 Hz. Both filters are 2nd order. This pulse would look like Fig. 1.

If we use this pulse to test the speaker, we will see any time distortion in the form of time smear. This measurement can be acquired by using a time-based measurement system, such as an oscilloscope. In Fig. 2, we are looking at a system that has a flat magnitude response but some low frequency phase shift.

The electronic network (crossover) used for Fig. 2 is made up of a 1000 Hz 2nd order high pass network feeding a perfect (no phase shift) high frequency driver and a 1000 Hz 2nd order low pass network feeding a perfect low frequency driver through a 180˚ phase shift network. We get this result when the signals are summed together at the microphone. If you don't include the 180˚ phase shift network, you have a magnitude loss at 1000 Hz. This was pointed out in the 1960s, and a simple method was suggested to accomplish this 180˚ phase shift. All one has to do is reverse the polarity of one of the drivers. That is, the (+) terminal on the driver is now driven by the (–) lead from the network, and the (-) terminal by the (+) network. This method of correcting the magnitude response was adopted by many loudspeaker companies and is still in widespread use today.

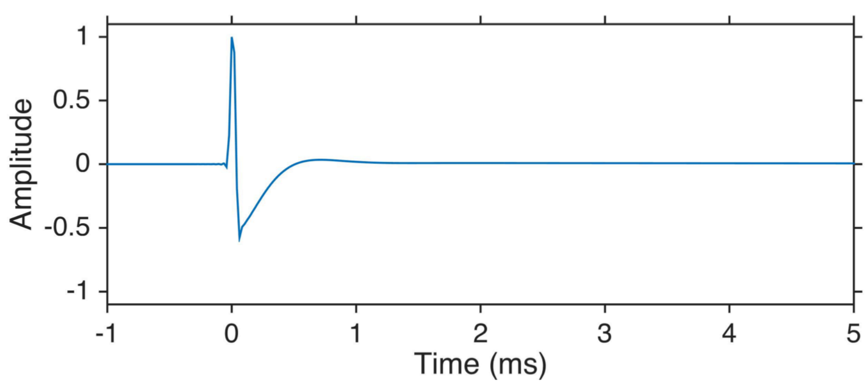

Figure 2. Ideal impulse response of a speaker with mid-frequency phase shift.

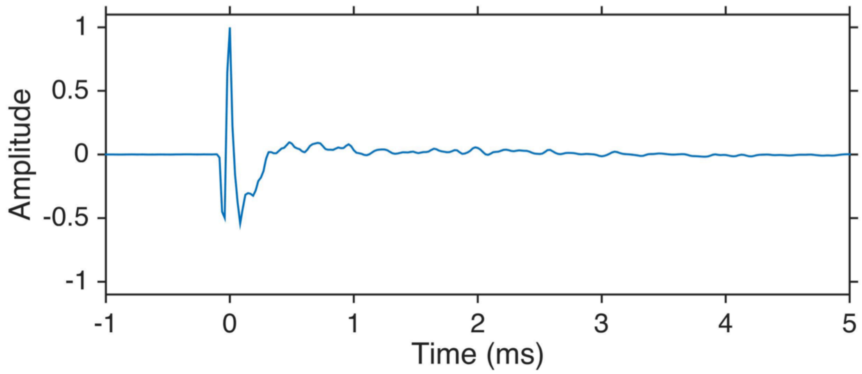

Figure 3. The impulse response of a modern 2-way near field loudspeaker, measured in the free field.

Fig. 3 shows what it is like to have the crossover network from Fig. 2 driving real world transducers. The real world transducer consists of a small low frequency driver and a small dome high frequency driver mounted above the low driver, whereby one of the drivers has reversed polarity. This causes a phase wrap (±180˚) at 2500 Hz in the phase response. The IR response using the above method is shown in Fig. 3. Because the transducers are not perfect, we see more time smear compared to Fig. 2. The HF driver has some phase shift, causing the plus and minus pressure peaks at the beginning of the time response. The larger negative dip is caused by the polarity-reversed low driver.

In Fig. 4 we have another classic 2-way phase inverted driver design; however, there is more phase shift in the high driver compared to the above speaker example.

The question is, do these different impulse responses sound different? As long as the magnitude is flat, this time smear is correctable. In the past, it was corrected by recording the impulse from the speaker on a tape recorder and then playing the tape backwards through the loudspeaker. This somewhat restores the impulse response to a single pulse. This “corrected” impulse sounds different than the original time smeared impulse. What is even more interesting is how the stereo image is affected by time smear.

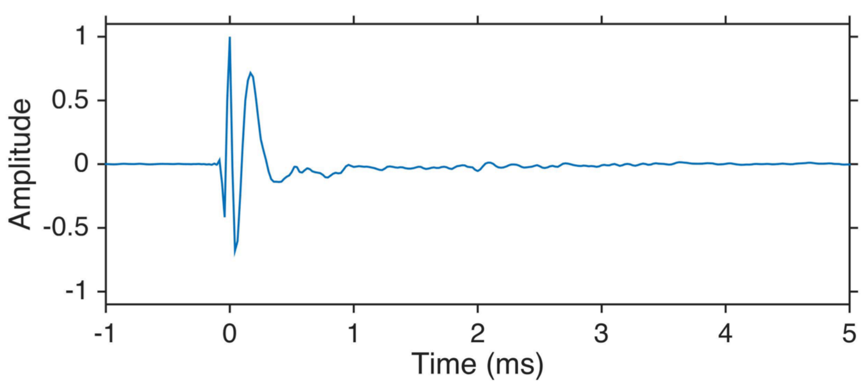

Figure 4. The impulse response of a popular 2-way near field monitor, measured in the free field.

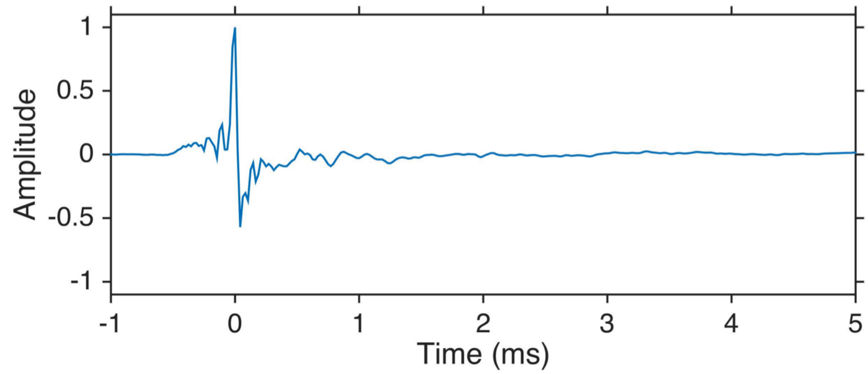

Figure 5. The impuse response of the HD-1 monitor speaker, measured in the free field.

In 1990 Meyer Sound introduced the HD-1, a small, 2-way, self-powered linear loudspeaker. This speaker was developed for research measurements where we needed a sound source that would have a flat frequency response (in order to test our new Source Independent Measurement system, aka SIM), where the magnitude is flat from 32-18000 Hz and the phase response would not have any phase wraps between 40 Hz to 18 kHz. This was achieved by using all pass filters (frequency selective delay networks), which make the IR more symmetrical (Fig. 5). We also flattened the phase response between 125 to 16000 Hz to be within ±45˚. The imaging of the HD-1 was an improvement over the classic phase-inverted design.

In order to achieve the imaging required for Amie, we needed to get all the high frequencies concentrated into a single positive pulse. The low frequencies of a small near field speaker are delayed compared to the higher frequencies. To completely time-correct the speaker to 40 Hz can take up to 100 ms of digital processing. This may be too long a delay for cinema mixing. What we decided to do is to follow the phase correction of our Acheron Cinema loudspeakers, which is phase corrected to 120 Hz. This way the small near field should sound the same as the larger Acheron speaker (Fig. 6). The amplitude ripple is 20 dB below the amplitude peak.

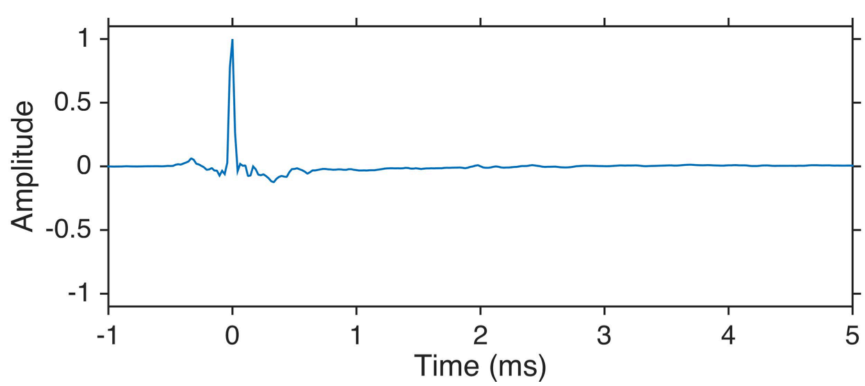

Figure 6. The impulse response of the Amie 2-way monitor, measured in the free field.

The next part of the Amie design is the horizontal and vertical coverage of the HF waveguide. Since we were designing this monitor to create a sonic image similar to the image of Acheron loudspeakers, we needed to make sure that we avoided anything that would upset the projected stereo image. Small near field speakers have very wide coverage patterns and sit slightly above and on the back of the mixing desk. There is a lot of sound bouncing off the console's surface, causing echoes which will alter the sonic image. HD-1s are used to hear exactly what is on the recording, whereby the mix is sent to the horn-loaded 'mains' speaker system to listen to the sonic image. Amie was designed for translation to the larger horn-loaded cinema sound system; however, since it would be used in rooms that have no other systems, it also needed to be sonically transparent. That is, it should sound as close as possible to the HD-1.

Amie's horn was designed for an 80˚ horizontal coverage and a 50˚ vertical coverage. This narrow vertical coverage helps eliminate the sound bounce off the console surface above 1200 Hz. The high frequencies give us the most clues to the image that is created between the loudspeakers.

This is a precision near field speaker intended for listening to fine detail in the sonic mix. Designed to work from one to two meters, it makes for greater flexibility in the setup. Although this speaker was designed for cinema mixing, we have found that it makes for an excellent near field monitor for music recording.

The next paper will be about Amie and low frequency room gain.

Speakers measured in this paper: HD-1, Amie, Dynaudio BM6 mklll, Genelec 8040B.

Addendums

Addendum 1: How to Measure the Impulse Response of a System

Every linear system has a unique impulse response (IR) that can be measured in several ways, with advantages and disadvantages for each. Determining which to use is complicated by the fact that the descriptions are often mathematical and highly technical. This paper will take a different approach, describing the various methods informally and without equations. It turns out that the best way to measure the IR is not the most straightforward.

A system is anything that transmits sound information: it could be a physical loudspeaker, a single transducer, a DSP filter implemented in an FPGA, or a concert hall. A linear system doesn't have any compression, distortion (harmonic or otherwise), or other "artifacts"; all signals are affected the same way.[1]

Conceptually, an IR is how a system reacts to a very loud and very short click (an impulse). The benefit of knowing the IR is that it allows one to predict what any signal will sound like; it's a complete description of that system. However, the IR is challenging to measure correctly.

From this description, a natural and straightforward way to measure the IR is to put a loud impulse or "bang" into the system and measure the output with a microphone. The impulse used is often a starter's pistol or firecracker, or the recording of one. If the system is a loudspeaker in an anechoic chamber, this measures the "free space" IR. If the loudspeaker is in a room or concert hall, then the mic records the impulse response of the entire system, which is the loudspeaker together with the room. In both cases, the IR depends on where the sound source and the microphone are placed, and changes when either one is moved. For most rooms, when played back, the IR will sound like an initial bang (the direct sound), then smaller bangs (the first reflections from walls and ceilings), and finally the reverb "tail" (from reflections off multiple surfaces).

Counterintuitively, this is actually NOT the best method to measure the IR of any system. Real world imperfections get in the way, all related to the fact that an impulse is a lot of energy in a short time:

1) The impulse itself is hard to reproduce, with either a limited peak level or a longer duration than ideal. For an electrical or electroacoustic device, the level needs to be low enough so the peak isn't clipped.[2] For a starter pistol, the peak is lengthened by the amount of time it takes the explosive powder to begin burning; there is also distortion from the barrel and a poorly defined directional pattern.

---

[1] If the system isn't linear, it will have a family of impulse responses, one for each different way the system reacts.

[2] Another way of saying this is that the system needs to stay within its linear operating range.

2) The measurement is sensitive to background noise. All the energy arrives in a short time period and then slowly decays, while the background noise stays at a steady level throughout. This means that the sound level gets closer to the background noise over time – the signal-to-noise ratio decreases – resulting in a less accurate measurement.
3) It's impossible to repeat an impulse exactly. For instance, starter pistols can vary as much as 6 dB in level from shot to shot, between differences in shell size and unburnt powder from previous shots. The result is that one can't reduce noise by averaging multiple measurements together.

However, there's another way to estimate the IR of a system, and it depends on the fact that the IR can be related to the system's frequency response (FR) by the Inverse Fourier Transform. The frequency response describes how each frequency in the input signal is affected by the system in both magnitude and phase. With a good estimate of the FR, computing the IR is simple.

There are several robust methods to estimate the FR, all of which depend on the Fourier Transform and the Fast Fourier Transform (FFT). Because computers only became fast enough to calculate the FFT in the 1970s, IRs were used before then.

Even though this way of finding the IR is conceptually more complicated, the signals used to measure the FR have several practical advantages. They all have energy spread out over time, which is the opposite of an impulsive signal. This means:

1) The signals are easy to reproduce, since the system doesn't need to reproduce all frequencies, at maximum power, at once. Possible signals include pink noise with constant energy per octave, maximum length sequences, sine wave sweeps, or even music. Also, signals played back on a loudspeaker have a consistent directivity.
2) Noise sensitivity is greatly decreased; the signal is always much louder than the noise floor and the signal-to-noise ratio remains high. Uncontrolled noise from the HVAC, traffic, or people talking has less effect on the measurement.
3) The signal is reproducible. This means that it's possible to average multiple estimates of the FR together, which will further decrease noise levels. Other types of averaging, such as vector averaging, can reduce the noise even further.

There are many variations of methods to estimate FR, and several devices on the market that can compute it, such as Meyer Sound's SIM3 system. As long as the FR estimate is accurate, it can be used to calculate the IR.

In summary, the impulse response describes how a system will respond to any signal. It can be measured in two basic ways: while conceptually straightforward, measuring it with an impulsive signal results in a bad estimate because of practical limitations. Conversely, determining the IR from the frequency response has great practical advantages and produces a much better estimate, although less conceptually straightforward.

Addendum 2: Measuring Loudspeaker Impulse Response using SIM Delay Finder

One mode of the SIM 3 measurement system is called "Delay Finder." The main purpose of Delay Finder is to set sample accurate delay times to improve the FFT-based frequency response algorithms of the "Frequency Response" (FR) mode -- the main algorithms used for loudspeaker tuning in rooms.

However, the Delay Finder algorithm can be used to accurately measure the impulse response (IR) of a loudspeaker, if one understands the algorithm. The Delay Finder algorithm in SIM 3 uses an Inverse Fast Fourier Transform (IFFT) to estimate the IR of a loudspeaker system in a room. In order to create a clean (non-ringing) estimate of the IR, the FR is low pass filtered before the IFFT is applied.

All the figures in this paper were either measured in SIM 3 or calculated theoretically in MATLAB using the same algorithm. As a result, SIM 3 and MATLAB produce an identical IR, as shown in Fig. A for an ideal band-pass limited system.

As described in the previous addendum, there are different methods and algorithms to estimate the FR of a loudspeaker. The filtering method described above produces a high resolution estimate. It will differ slightly from time-based methods (i.e. a starter's pistol). For example, it slightly widens the peak of the IR, but has a much higher signal-to-noise ratio (dynamic range) than other filters or other time-based methods.

Since all the IR measurements in this paper were created or measured using the same algorithm, the results are directly comparable. When reproducing IRs in this paper, be aware of slight differences in the IR estimation algorithm used, and use the same algorithm for consistency.

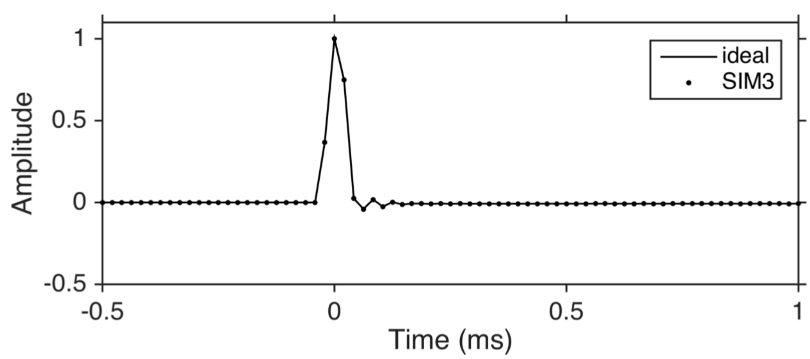

Figure A. Comparison of an ideal IR with filtering to an IR measured on a SIM 3 audio analyzer. The root-mean-square difference between measured and ideal is 4.2×10^{-3}.