The Mix of Bombay Velvet

The reason I decided to write a blog and subsequently make this small book out of it is important for me. There is nothing special in the way Justin and I have done the mix, yet this mix was very special for us. We have approached the mix based on the many discussions and talks we had over time with Kunal Sharma, who took care of the Location recording and the Sound Design of the film. The main thought behind documenting this mix is because I personally feel that the art of mixing and what happens in the mix theater should not be a secret or a dark art. I would not be where I was if I wasn't taught by the many senior stalwarts before me. The ones whom I had the opportunity to assist have always taught me the approach and methodology to a mix. There is a lot of discipline that one will get to in a mix. But one doesn't need to constantly reinvent the wheel. There are a lot of students in the field of sound and a lot of them who want to know what it is that happens. This is an invitation to witness that. We need and can make do with great talent and ideas. They should not be limited to a select few who have the luck to be in such a place. One never knows from where ideas, and talent comes. They only need to be guided. This is also a chronicle of how things are thought through in a mix and why and how some techniques are created and employed. Also, let's keep it real. A mix can never always be perfect in the beginning. That's the beauty of experimenting. One can fail, and learn from that or learn something completely new.

-Sreejesh Nair

Table of Contents

The Story

Justin and I have handled the mix over 28 days and culminating to quite a few hours of work. This is a project I had been waiting for more than a year and there was a bit of nervousness. But the best thing is I am working with friends and there is no bigger comfort than this. The design skills of Kunal Sharma is not something I need to think about especially after working on a movie like Gangs of Wasseypur. Kunal had spoken to me about this project while I was on Gangs. I had said yes and from then it was a long wait. A wait that was not empty but a good one because we spoke of the many things we wanted to do and experiment.

Kunal was sure that he was going to do this project on 96kHz samplerate. He had done Udaan and Lootera on that way and was confident of all the technicalities and aware of the limitations we would have during the mix. But then I suggested that we do this in Dolby Atmos. This was at a time in 2012 when Atmos was very new. Yet, we were sure we wanted to do it in this, as we were very excited by what that format would offer us. Yet, this was not attempted in 96kHz. We agreed to look at it as it comes. (Not a very professional approach I know!!).

It was at the same time that I decided for a change in career and move to another company and Justin took my place in my company. This was also exciting as Justin is a very clever mixer and is one of the few Disney approved film mixers in India. We decided on our team of mix then itself. And from then, started the conversations and ideas, some of which made it, some didn't.

The thought behind Dolby Atmos

At this point, I think it would be useful to talk about the format. This would make it easier for you to follow what I write in terms of techniques later and how the movie is mixed. This will also help understand why we decided to do this film in this format, as that was a very important decision to make.

Before Dolby Atmos

Traditionally, mixes were done in 5.1 for a very long time. (I am not going into Mono mixes as that is mostly irrelevant today unless a creative technique.) This means that the speakers we had were Left, Center, Right, Left Surround, Right Surround and Low Frequency Elements. This is usually abbreviated as L, C, R, Ls, Rs, LFE. Within Pro Tools, the Pan position is as shown in the figure below.

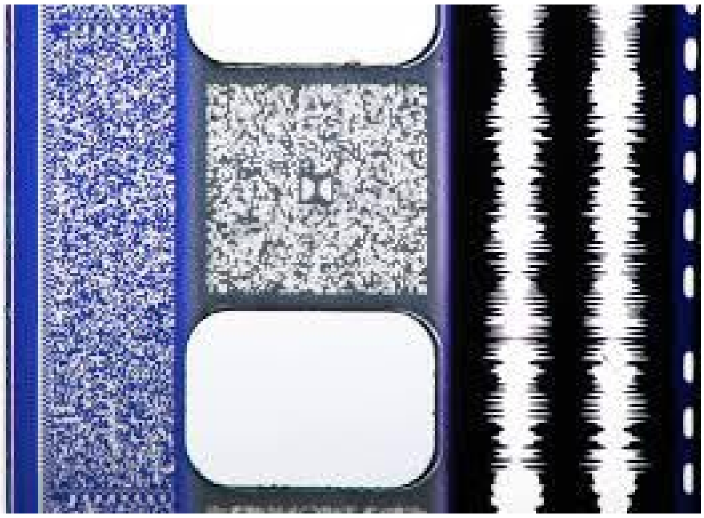
5.1 Sound Field

You can see the speaker positions in the above figure. The little dot is what we use to position the sound in the field. But in the theater, this is not as simple as this. The Surrounds are actually an array of speakers. So, although in the above picture the surrounds are a point source, in the theater, this is actually not. One of the things to get your head around initially is also the position and the sound in the theater. For example, if you pan the dot to the speaker in the Right Surround, the sound in the theater would come from the array on the right and not exactly the back.

This was a print and a digital format and the sound field was the same in DTS too. Then followed a digital format called 7.1. This was not available in the print format and so was very popular with Digital Cinema. This is different from the SDDS format as it splits the surrounds into 4 zones, namely Left Side Surround, Left Rear Surround, Right Side Surround, and Right Rear Surround. These are abbreviated as Lss, Rss, Lsr, Rsr.

7.1 Sound Field

This actually was a format adapted from the 7.1 HD made initially for broadcast and home entertainment. Now, as mixers we quickly realized that if the pans were bought beyond the middle like that crosses the side surrounds, that would mean the sound would be taken off from the stage (L, C, R). The problem this bought about was that usually in the theaters, the numbers of speakers for the rear surround are lesser than the side surrounds. (They are calibrated to be exact in terms of loudness though.) But for the audience, this would take attention away from the front and would be direct sounds to the ears. So, some amount of getting used to was required for us although that wasn't really a long period.



Dolby Digital on Print. You can see the Double D logo in the middle.

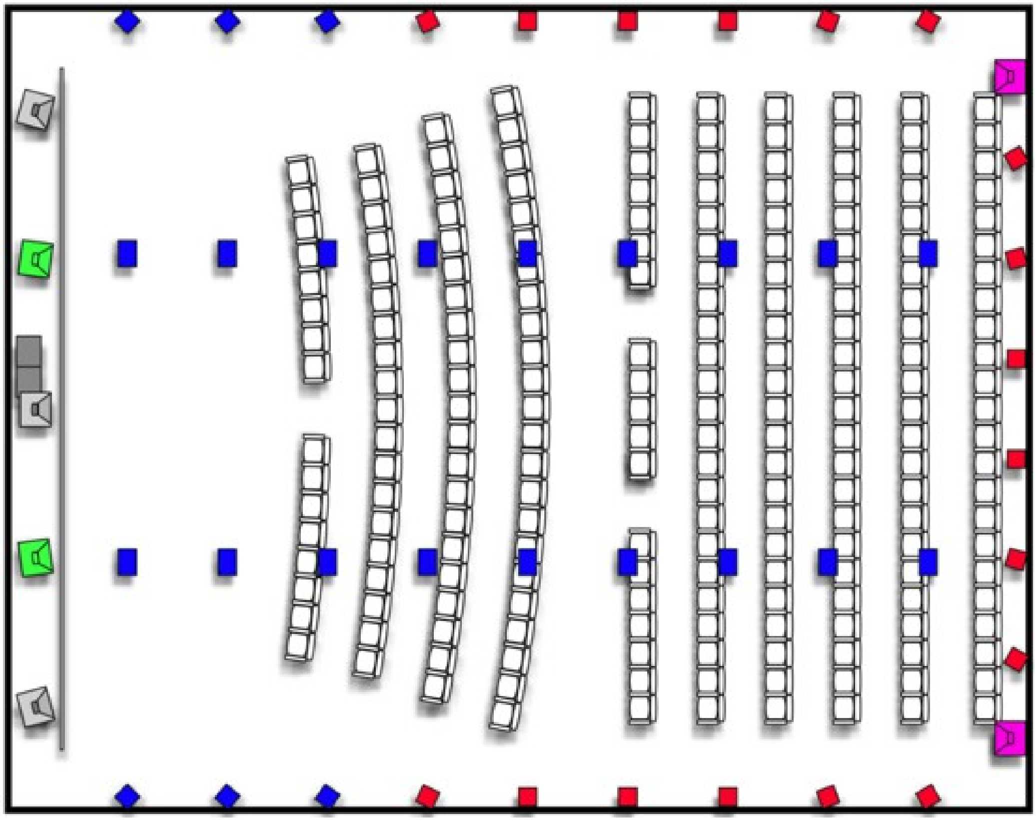
Dolby Atmos

One of the challenges in the above configuration was that the screen speakers were calibrated to 85 dB SPL and the surrounds to 82.

Originally with Dolby Stereo the mono surrounds were set to 85dBc, to match each of the screen channels. With the introduction of split surrounds, each array was then calibrated to 82dBc due to average 3dB acoustic summation between the arrays, this means that a common signal sent to both channels will give an equivalent monitoring level of the mono surround for compatibility. With the introduction of Atmos, Dolby decided that Lt-Rt compatibility is less important (i.e. can be performed using fixed level offsets), and that it makes sense for all channels to monitor at the same level hence they made the change to 85dBc for each surround bed.

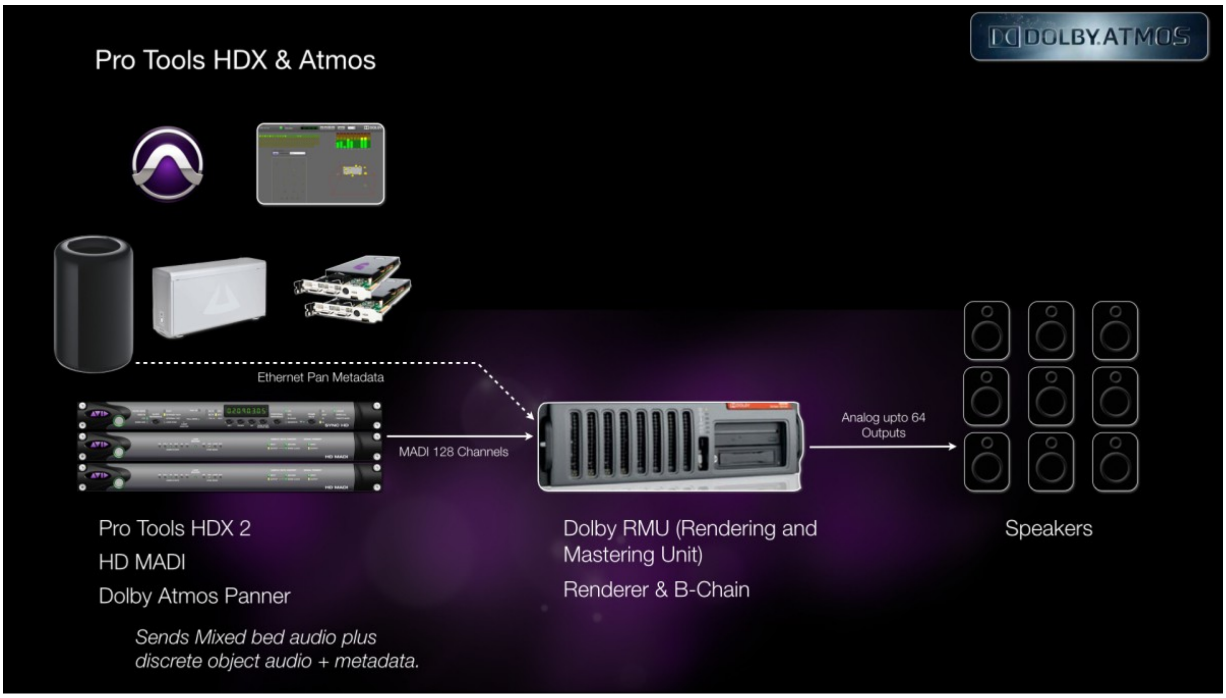
This means that each array in Atmos is now calibrated to 85 dB and also that each speaker in the surround is Full Range as they have additional Subwoofers in the surrounds that bass manage the low frequency in an intelligent way.

The real advantage with Atmos is 3D audio or Immersive audio. If you think about it, sound was always in 2 dimensions. You could represent it on a paper or a pan pot as above. Sound always was on a horizontal plane. Dolby Atmos made it possible to make sound 3D. They introduced an over head array of speakers. Thus, the whole sound format basically starts with Left, Center, Right, Left Side Surround, Right Side Surround, Left Rear Surround, Right Rear Surround, *Left Overhead*, *Right Overhead* and the LFE. This is called a 9.1 Array. (9 mains and one subwoofer)



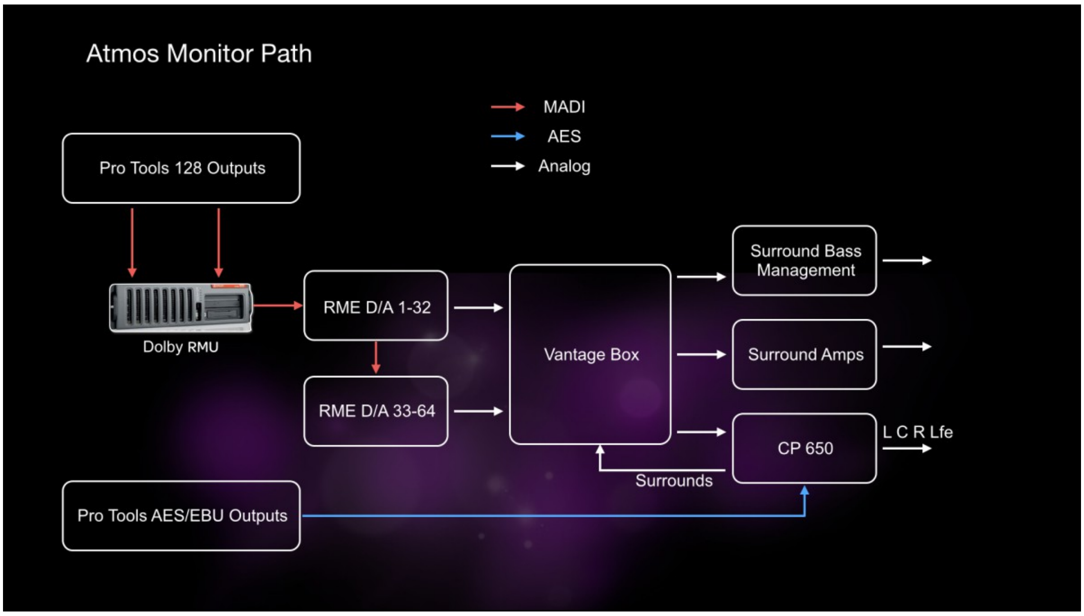
Speaker Placement in Dolby Atmos

The whole format of Atmos consists of 10 tracks of Beds (9.1) and 118 tracks of Objects that are fed into a unit from Dolby called the Rendering Mastering Unit or RMU. These require 2 MADI cards as there is a total of 128 inputs to it. This means the Pro Tools system would be an HDX2 system.



Signal Flow in Atmos

The output of the RMU is a total of 64. This means the maximum number of speakers that can be addressed in the theater is 64. So, input is 128, output is upto 64. That being said, the number of speakers in a theater depends on the size of the theater. They can go anywhere from 16 to 64. The beauty in this is that the format is scalable. Which means a sound placed 1/3rd between the screen and back wall will be 1/3rd from the screen to back wall irrespective of the size of the room. (Eg, if the length is 3 meters in the mix room, it will be at 1 meter. If the Hall it is playing back is 6 meters long, the sound will be reproduced at 2 meters, thus keeping the ratio.) The geography will be accurately maintained for anyone watching the movie.



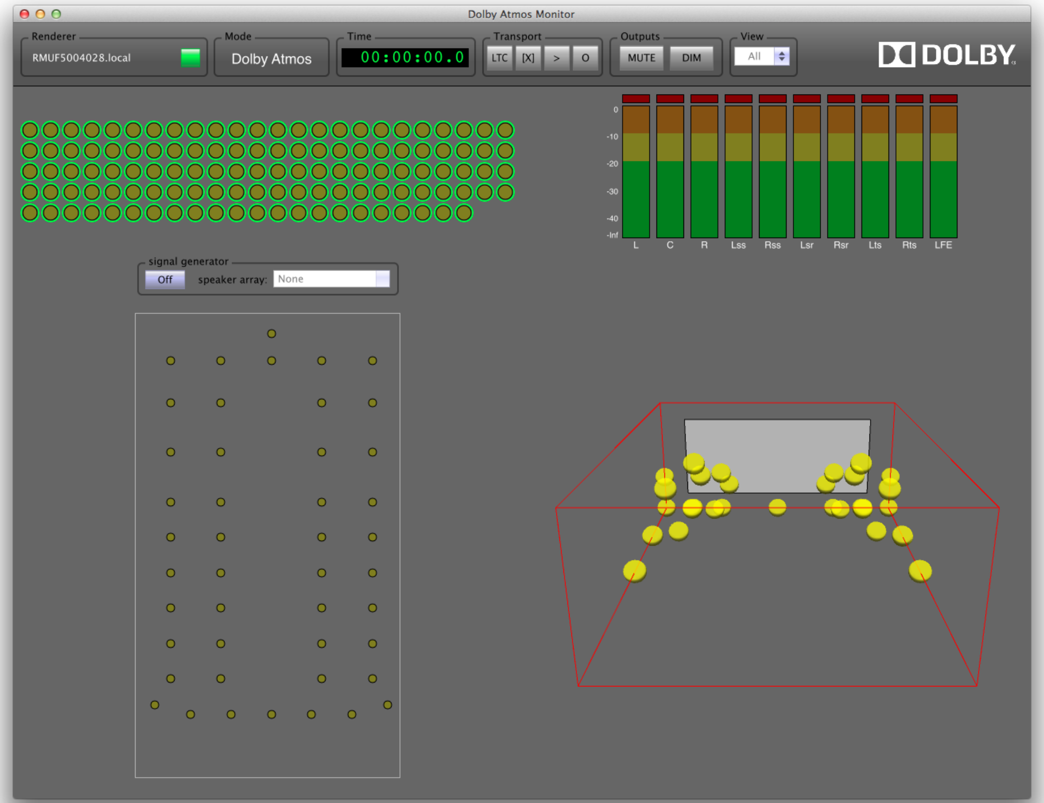
Atmos Monitor Path. The blue line is for traditional Mixes.

There is another concept that needs to be understood with Dolby Atmos.

About Bed and Objects

The 9.1 configuration I mentioned earlier is called a bed. A bed is essentially a part of the mix that is played through an array speaker. In music it would be something like a Pad or strings etc. In Effects, it will be ambience like the wind or general traffic etc. And in general Reverbs are beds too. Keep in mind that this is just to explain and as with sound, there are no rules!

The other very distinguishing feature is Objects. These are very specific point sources or sounds. For example a car pass or a single bird, etc. The specialty of an Object versus a bed is that the object if panned from front to the back along the right surround **will pan through each individual speaker in the path.** This means the precision offered in the pan or the resolution is extremely high. A proper mix of the Beds and Objects would make for a very immersive mix.

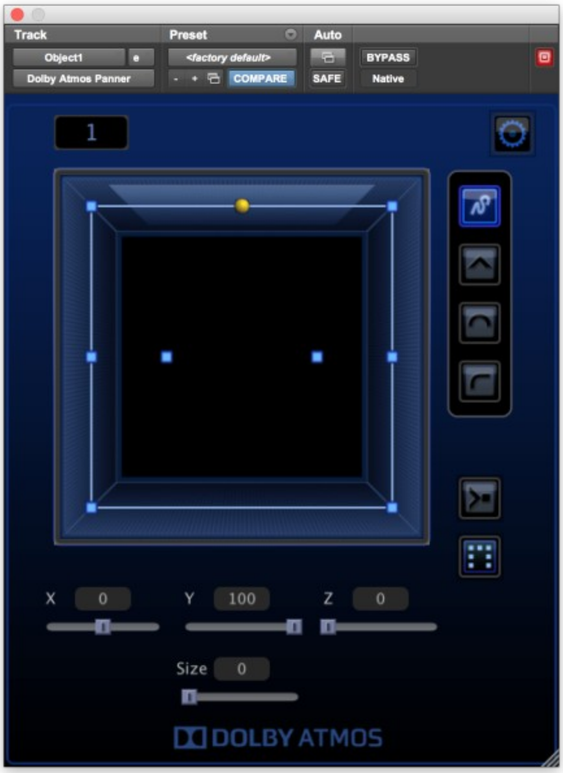


Atmos Monitor and Speaker placements

The objects can be used for pans and for static placements. So if for example, there is a restaurant scene and you want to place some cutlery sounds in a very specific speaker, you can do that too, thus making for a very realistic space. If you look at the above figure, all the speakers are part of the objects and beds, but the blue in the left surround and right surround are exclusive objects. So is the case for the two green speakers inside the screen, which can be called as Left Center and Right Center. This is present in Rajkamal, where we are mixing the movie.

Handling Objects

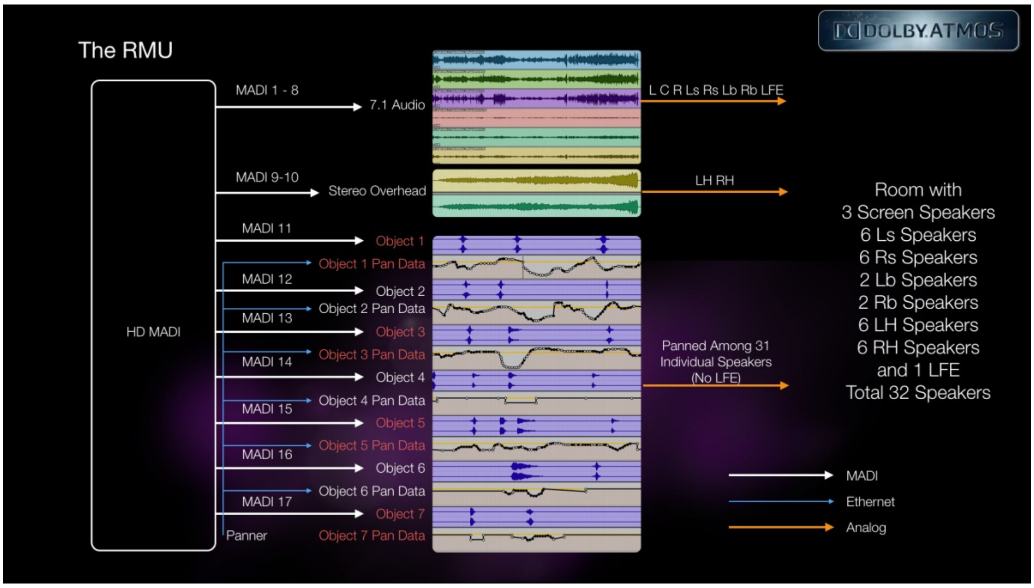
Now, we know that the main difference between Atmos and traditional mixing is the resolution in the Pans; there is an obvious question. How are they panned? This is where the Atmos Panner plugin comes in play.



Dolby Atmos Panner

If you look at the panner above, you will see the squares that represent the areas in the theater. One very important thing to note is that they are not shown as speakers but rather as indications of the borders of the room. This means that this format allows you to mix to Space rather than Speakers and this is a very important distinction. So while mixing it is much more easy to create a real word space. The two dots in the middle of the square are the overhead section. If you see, there is a number designated as 1. This means this panner is controlling object number 1, which comes from MADI no 11. (The first 10 are beds remember?). How is this transmitted?

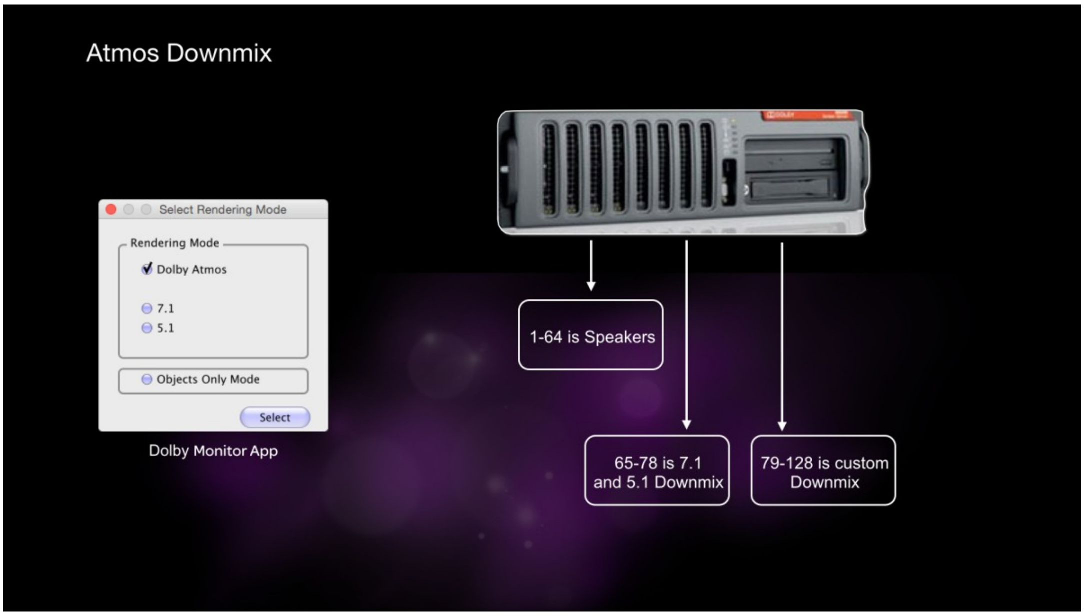
The Pro Tools system is connected to the RMU via an Ethernet cable. This pan data is sent to the RMU as Metadata and it associates this with the track in it is recording. There are a few options on the panner itself that I won't go into a lot detail as that is related to panning. The main parameters are X, Y and Z which are self-explanatory on the position. The Size is kind of like Divergence in Pro Tools. The below figure is like a Pro Tools version of how the RMU records Data. This is just for explanation, although the RMU does it in a far more sophisticated manner.



RMU in Pro Tools Language

Backwards Compatibility

Not all theaters are Atmos. What happens in a regular theater? This is the ultimate strength of Atmos. The advanced algorithms in the RMU make a fantastic downmix of the Atmos mix. In addition, because we have the overheads and the objects, as mixers, we would do pans and balances much different and in my opinion more bold in pans and positioning. This gives a very different mix than if we had a traditional mix.



Atmos Downmix

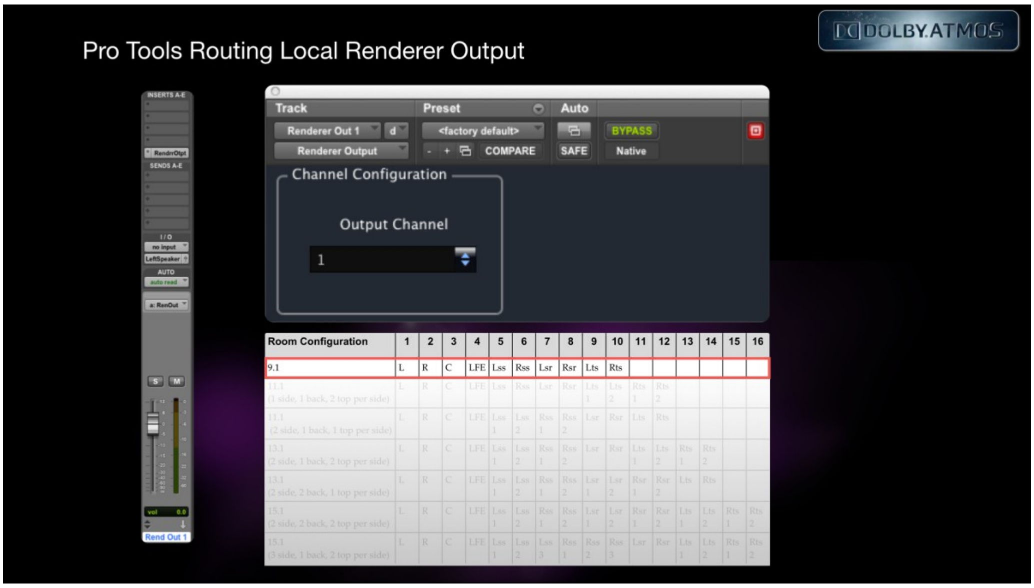
Sound Editing for Atmos

One fantastic tool that Dolby makes is something called the Local Renderer. What this is is essentially a software version of the RMU that runs in the background as a service with only its rendering capability. So, any sound editor or designer while doing the tracklay can make the tracks and pans using the Atmos Panner itself but without needing the RMU. (Remember, the RMU is a must for Mastering and in the Mix room.) After a preliminary discussion with the Mixer on the Object number that would be for FX, Ambience, Foley etc., the sound editor can then make a session that would match this in the mix. This is installed on the system where the sound edit is happening.

Basically, there are 2 plugins that are additionally installed while installing the Local Renderer. These are the Local Renderer Input and Local Renderer Output. The plugins do what is said. Send to the Renderer (input) and receive mix from the renderer (output). The Local renderer can support up to 16 speakers. So you would need an HD IO 16×16. Logically, this is how the Pro Tools session would look like in the sound editors room.

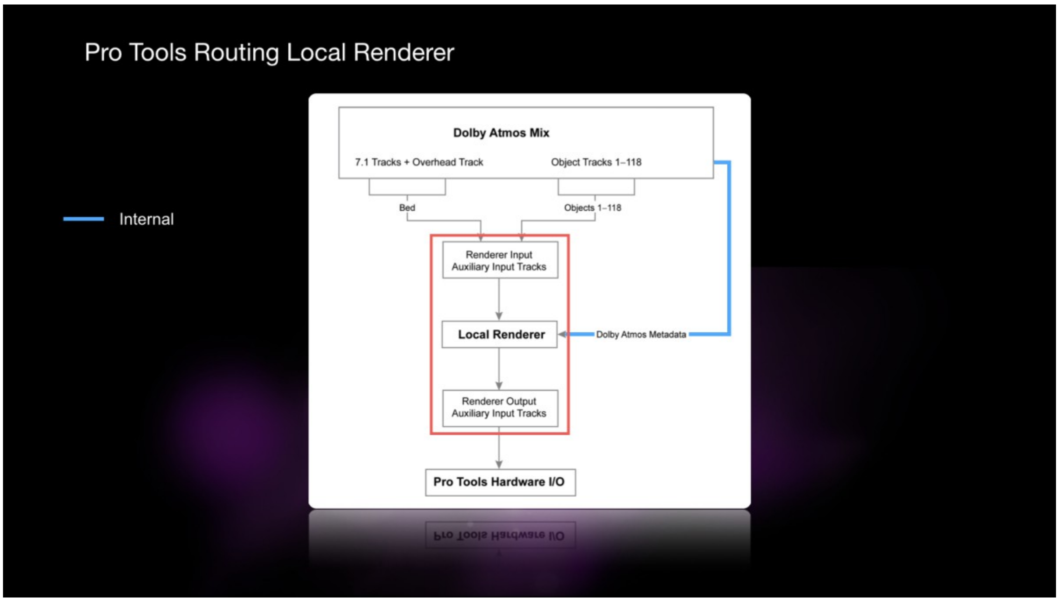
The input to the Local Renderer is via the Renderer Input plugin, which is inserted on an Aux as shown below. This plugin has different modes, which can be set in the source type within the plugin, for beds and objects.

Local Renderer Input



Local Renderer Output

The Return from the Local Renderer is taken into an Aux that can be sent to a speaker as in the number shown above. So, 1 will be Left, 2 will be right, 3 will be Center etc. up to a maximum of 16. The advantage with this method of routings is that since the sends to the Local Renderer are from a track via a Bus, this session can be opened in a mix stage and if needed sent back to the edit stage for correction ***without*** change in output routings. If you look at the Local Renderer Input Figure, you would see that the Object 1 is routed to MADI and also there is a send on Object 1 Bus to an Aux. Since this MADI isn't available in the Edit room, it would be greyed out and the Aux would be functional in sending the output to the Local Renderer. When in Mix stage, the AUX would not be functional as the MADI output is present. (It takes a few reads to understand that!)



Local Renderer Signal Flow. The Red box is indicative of internal Processing.

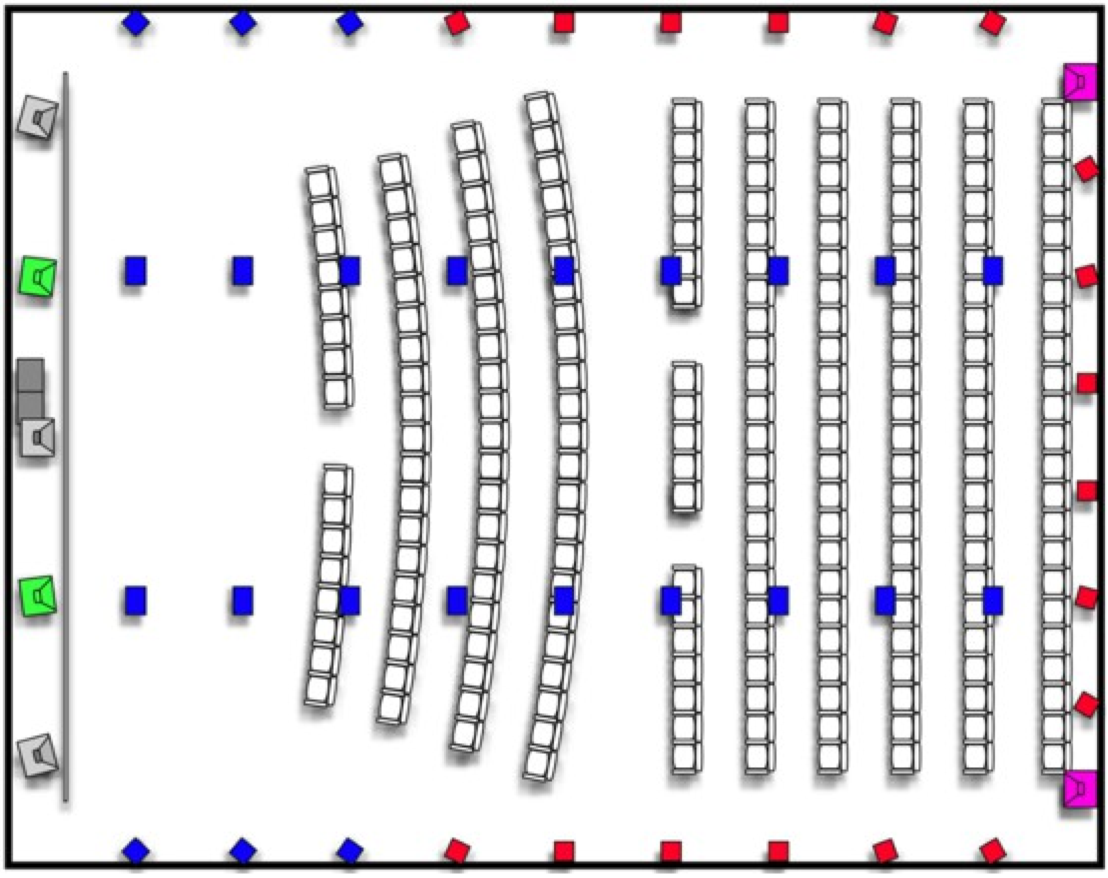
Routing Techniques

The Thought

It is important to know the thought behind preparing a session for mix. Usually we get the sessions from the Sound Designer and the Music Composer and then we re-route it to our comfortable bussing structure that we as mixers are used to. For example, there will be a dialogue Buss, Dia reverb busses, Foley, FX, Ambience, Score, song and crowd. Of late the FX and foley are also split into more wider stems like FX, Design 1, Design 1, Footsteps, Incidentals (taps, doors, cloth etc) and so on. This time too, Justin and I followed the same principle, but we had also come up with some ideas that are unique to this. I think we are the first to figure this out, but then if there is someone reading this who has done this before, then by all means this is a coincidental thought and not plagiarism!

The Mix Room

What is very unique about Futureworks Parel is that it is a Premiere Atmos Mix stage. The Only one in India and the first in the world. Now, what this means is that it has 5 screen speakers instead of 3.



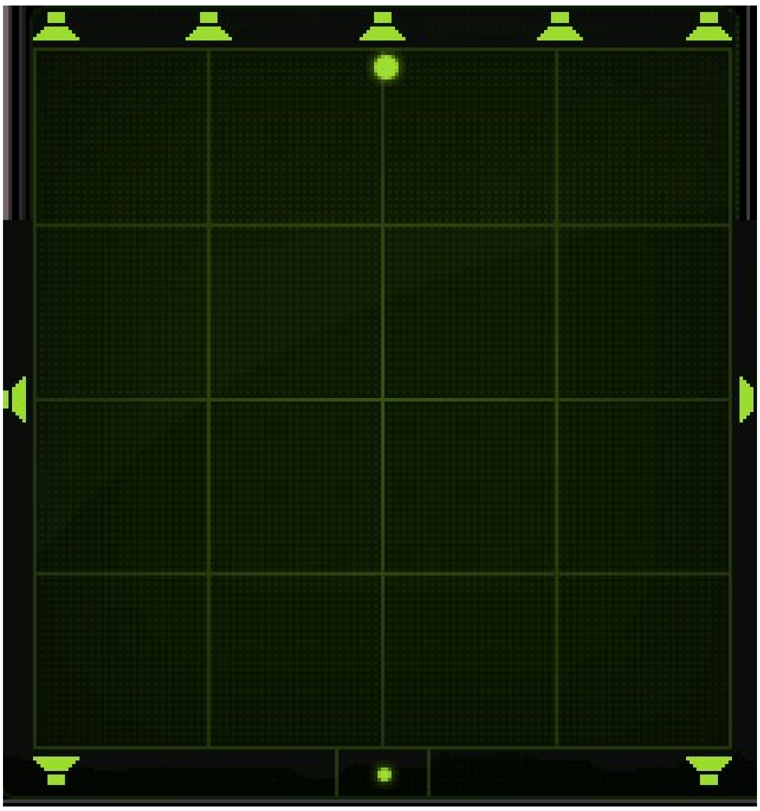
Just behind the screen we have Left, Center Right in gray color and two more in green color, which I will call Left Center and Right Center. Does this screen layout look familiar? If yes, it is because this is similar the **Screen** layout of the SDDS format. It is a very accurate representation of the screen and gives more resolution to the pan. But there are two issues with this.

1. The green speakers (Left Center and Right Center) are objects and not part of the bed.
2. The SDDS format has only 2 surrounds while Dolby Atmos has 4 surrounds (Left Side, Right Side, Left Rear and Right Rear.)

So, my thought was how could we combine both the SDDS for screen and 7.1 HD for Surrounds into one single Pan motion in Pro Tools? That's when a unique thing in Pro Tools struck me.

11.1 Pan in Pro Tools

Don't confuse this with the 11.1 in Auro 3D. That is 2 sets of 5.1s placed one above the other. (Actually it is a 5.1 and a 5.0 above it with the Voice OF God or VOG channel in the middle.) The 11.1 in Our Mix session for Bombay Velvet is different and not standard. It is laid out as Left, Left Center, Center, Right Center, Right, Left Side Surround, RIght Side Surround, Left Back Surround, Right Back Surround, LFE that makes 9.1 and then 2 overheads that makes a total of 11.1. So, we have a unique Set of resolution. How did we create this?



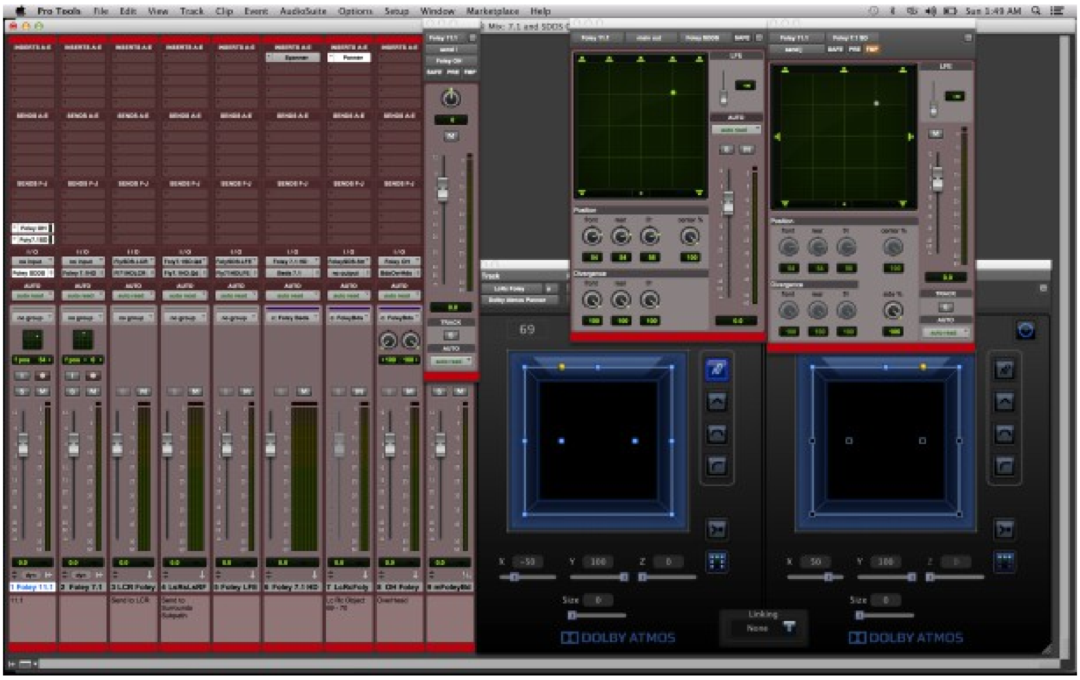
Simulated Image of the Panner

The first idea I had was to have the SDDS and 7.1 formats and then cut tracks into whichever is needed. But Soon I realized that's a lot of additional work with no benefit. Then I saw the option of **FMP** in Pro Tools.

What FMP stands for is Follow Main Pan. This is part of a Send and what it does is if you have a pan in an audio track, it will make the aux follow the same pan. Then it struck me that if I create all tracks with the main output to SDDS and also have a send to 7.1 HD with FMP on that send, this means when I do a pan in the main pan, it will be simultaneously sent to the bus.

Combining the different 7.1

In the above Picture, the Left Pan window is the main track output and the Right Pan window is the send with FMP on the top enabled in Orange. This means as you can see, any pan position in the main output is reflected in the send too, So, my next step was to make a set of Auxes that Route the L C and R of the main output to the L C R of the Bed. I would then discard the surrounds of that and instead use the surrounds of the send to the bed surrounds. (Note that the send is Post fader and also at unity so my levels are not changing.) My next step is to create a subpath for the Lc and Rc and route that to an object that is panned to that position using the Atmos panner. This is how a template would then look.



In this, if you see, Track no 1 is the main audio track routed to an SDDS output format. Track 2 is a routed to 7.1 HD in case we need that. Tracks 3, 4, 5 and 7 are what split the outputs. 3 and 4 combine to give the bed. Track 7 is the Lc Rc Subpath of the SDDS which is panned with the Atmos panner to those positions. Although it looks complex it really isnt once you have these auxes in place. Then anything will be routed to this and will be far easier to pan and achieve the changes, which previously wasn't possible at all. The biggest strength that this method has is that when you pan beyond the centerline, the audio no longer is removed from screen, which you will remember is a challenge from 7.1 HD.

Dolby Atmos Mix Stage Signal Flow

This project is unique in the fact that we are running the sessions on 96kHz, which is a world first for a Dolby Atmos movie, and all credit goes to Kunal Sharma for initiating that. All the Dialogues, Music and Effects were recorded at this Sample rate. It would have been practically impossible to run so many tracks at this scale unless it was an HDX system and we have 2 HDX2 systems.

The Systems

We are running the Dialogue and Music on a Pro Tools HDX2 System with 2 HD MADI IO and Version 11.3.1. The Effects, Foley and Ambience is running from another Pro Tools HDX2 system and 1 HD MADI IO with Version 12. There is a third system, which is an HD Native system with Pro Tools 10.3.9 that is used for recording the Mixdowns. All of these systems receive clock from a Rosendahl Nanosync Clock generator. They are also all on Satellite Link, which is included with Pro Tools 11 HD and above. We have kept the Dialogue and Music Machine as a Satellite Link Administrator. We also have a D-Command that is used for the Fader rides on Dialogue and Music and an Avid S3 on the FX.

The 96kHz Challenge

From the Outset, we knew that there was going to be a challenge to run so many tracks from one system. So we had 2 of them. The other main challenge was that the RMU would only accept signal at 48kHz. So, how do we supply that and why go through all this trouble?

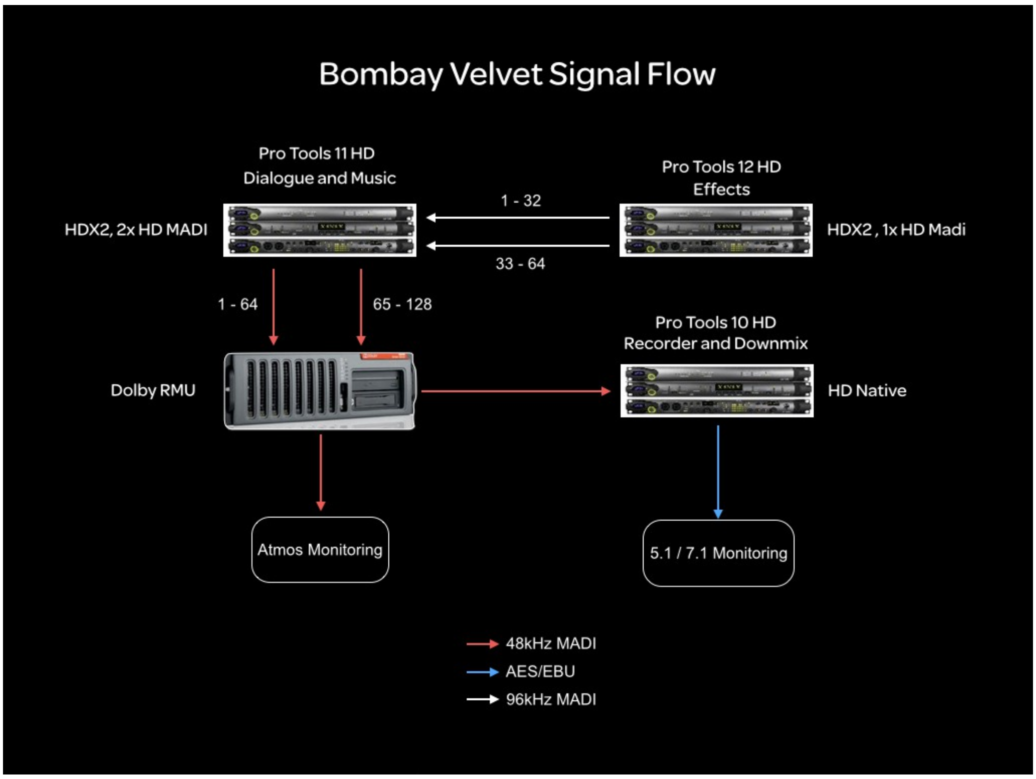
We solved the issue of the DSP and power by adding another HDX2 system with an HD MADI card. Now, at 96kHz, one issue is that the track count will be halved and so will the output. So instead of getting 64 IO it would be 32 IO. That wouldn't be enough for a film of this scale. But, the HD MADI is capable of delivering split Outputs from its 2 ports as in the table below.

Port	Channel	Sample Rate
1	**1–32**	**88.2/96 kHz**
2	**33–64**	**88.2/96 kHz**

All that we had to then do was to enable SRC (Sample Rate Conversion) on its output to the RMU to convert from 96kHz to 48kHz. But there was another issue. How do we send the beds from the FX Machine to the RMU? For this we sent the outputs of the FX Machine to the Input of the Dialog and Music Machine that was feeding the RMU. This way, the Dialog system got an additional 10 Objects and we were able to use the whole of 128 Outputs which otherwise would not have been possible. You can see this in the signal flow diagram below.

Why then take the pain of running everything at 96kHz if we are finally going to monitor in 48kHz? This is a difficult question to answer until you manage to hear the difference. We did tests with the session running at 96kHz and session converted to 48kHz. The Reverbs and EQs just sounded far cleaner and fuller on a 96kHz session even if the output was converted. This convinced us to stick to this. Moreover, the advantage of running everything at 96kHz also means there is more information in the higher frequency spectrum that has been recorded. Plus, the sounds that are being pitched for design sound more cleanly and with no artifacts. Plus with all the amount of analog emulation from Avid Heat, UAD and many more, the audio just sounded full.

The Signal Flow



Mixing Stage Signal Flow

So, if you look at the above diagram, you will see how we managed to achieve the flow in Dolby Atmos to maintain the sample rate through the session. Once we had this, we had budgeted our Objects and the split between the Dialog, music and Effects system. The split is:

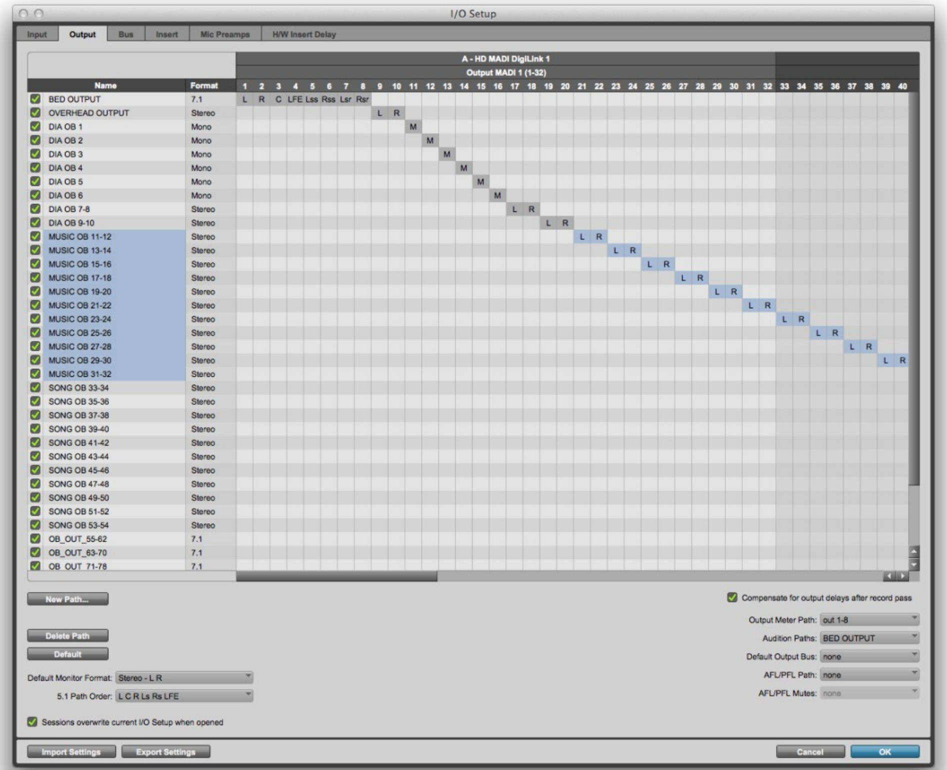
Stem	Object Start Number	Object End Number
Dialogue	1	10
BGM	11	30
Song	31	50
Additionals	51	64
Crowd	65	74
FX	75	99
Foley	100	106
Ambience	107	118

One thing that is really good with Pro Tools Version 12 is the ability to have unlimited Busses and also the ability to open a session with our IO setup or even import an IO setup from a session rather than saving the IO setup externally. That is really cool and does save us a lot of time. In addition, there is also the ability to automatically map the monitoring to what the monitoring output in the stage is without rerouting anything. This is a huge benefit and we decided to use and explore this on the FX machine. We did spend some time figuring this and testing the workflow before we started the whole mix. This was crucial as once we are in the mix, it becomes very difficult to troubleshoot and change things around.

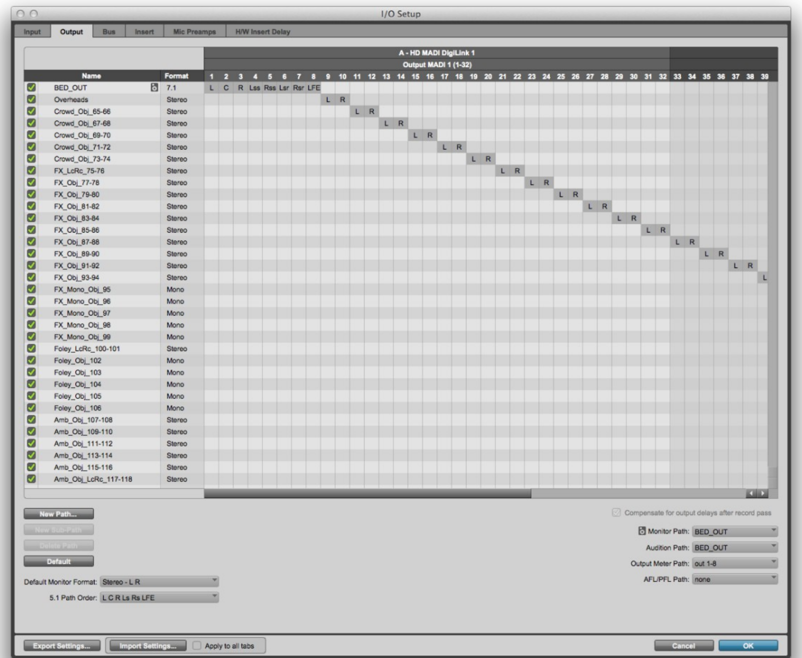
Splitting the Task

On the outset, we decided that Justin would look at the songs and the Background score, while I would take care of the Dialogues and Effects. We would then switch to see if we have ideas on each other's work or if something better needs to be done. This is an effective way of working so that the ideas are new and fresh and so is the tonality. But during the mix, since this is a very complex film when it comes to sound and the subtleties, we split the workload with me on the effects and him on dialogues and music and switch to review each other's ideas. Timing on achieving this is crucial as not all reels are equal in work and we need to complete for the other in time. So, Justin works day and I get in the evenings and cover up and on the Final Mix, we sit together with Kunal.

Setting up the Session



IO setup Dialogue and Music (Pro Tools HD 11.3.1)



IO Setup FX, Foley and Ambience (Pro Tools HD 12)

Dialogues

These comprised of the Booms, Lapels and multiple MS recordings. There were some techniques we employed for the MS, which I will cover, in a later post on the premix. I prefer to have the booms on the top followed by the Lapels and then the MS. There is a separator aux that divides this and the ADR. One thing that is also required is the fill track or a mono ambience that is preferably from the location. The Izotope RX4 does an incredible job of generating this from the location tracks. In addition, it is always nice to see the lapels split into characters and the Production Effects (effects that are on the dialogue tracks) split in to another track. This is where Zahir's talent comes into play. He was the dialogue editor on Gangs of Wasseypur and also on this. He takes care of the Phase match between the different mics and also matches the ADR very closely which I then only have to polish off in the mix.

There are 10 object tracks created for this so we have a resolution in pans when it is done across the surrounds or screen. (Oh yes, you can pan dialogues into the surrounds in Dolby Atmos. No exit sign effect!) The reverbs are set up as a 7.1 and a stereo Overhead. I am using Phoenixverb Surround from Exponential Audio for this. These are brilliantly designed reverbs with a very smooth tone for dialogues and score. One clever thing this has is a 3D link that can link the beds and overheads to create a very realistic 3D space. I then have a VCA master for the Beds and Objects. All of this is taken care of by Sarath Mohan who is brilliant in setting up these things.

Effects, Foley and Ambience

The setup for these is pretty similar to the dialogues, but on a different machine. One change here is the use of the SDDS and HD 7.1 combination I mentioned here. This is employed in the Effects and Ambience routings. Again the reverbs and the Masters are set in the same way as the above. I prefer to have the foley on top, followed by its auxiliary, the Effects and then the Ambience. I leave my VCA masters on the top of the session marked by a vFX, vAmb and vFoley. The reason is that we are using a Pro Tools S3 on the Effects machine and it is a bit easier for me to do the rides when things get crucial. I also have Spanner on all the beds, as it has become an indispensable tool for the mix.

Music and Songs

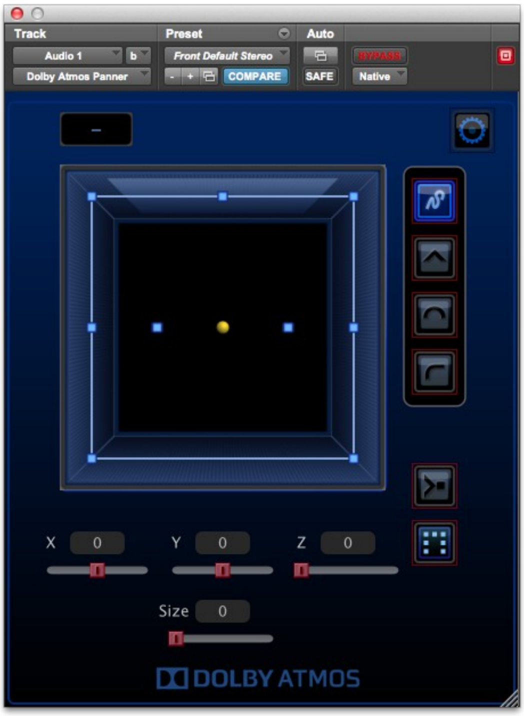
One of the challenges we have is fitting everything into a master session. So, the Music is split into different 7.1 stems and the corresponding objects. The beds are then recorded with split reverbs and imported into the master session with all the raw tracks deactivated and hidden. We then have a window config that allows us to go into the BGM or Dialogues if we need to fix something in the recording. The good thing is that we don't leave the session to do this. Spanner is again a tool here to fit the music in the space. Another reason we did this is because we are using a lot of analog emulation for the Dialogues and effects. Not something that would be standing out, but for the warmth and to match the look and feel of the movie. We went through a lot of plugins to fix in on some very cool ones. I also took some help from Farhad Dadyburjor who in my opinion is an encyclopedia on this!

The Master Session

The Master session for the Effects machine doesn't change but the one on the Dialogue and Music has a difference where they are combined and an additional set of 64 auxes are created to send the signal from the Effects machine to the RMU as in the earlier chapter I mentioned about the signal flow.

Defaulting the Dolby Atmos Plugin

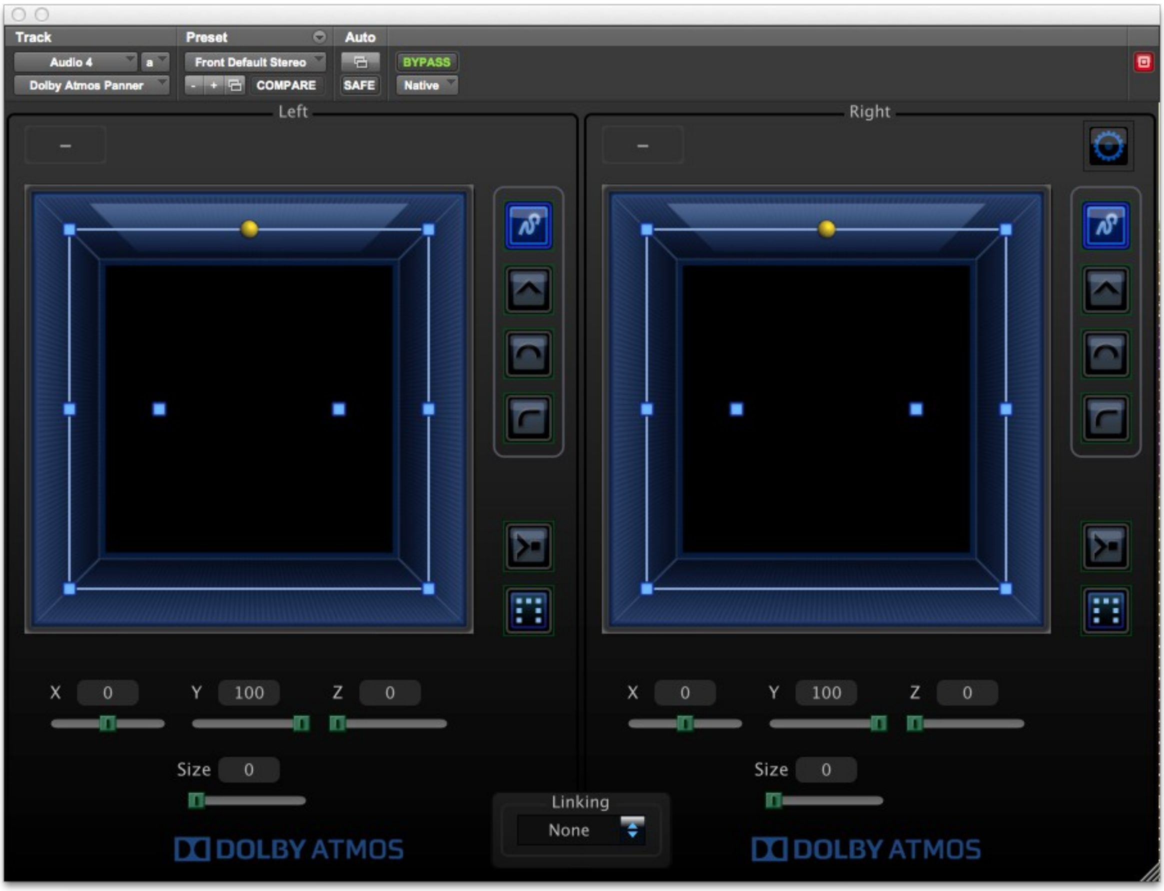
One issue I have with the Dolby Atmos panner plugin is that when we instantiate it, the panner comes up like this.



Dolby Atmos Default

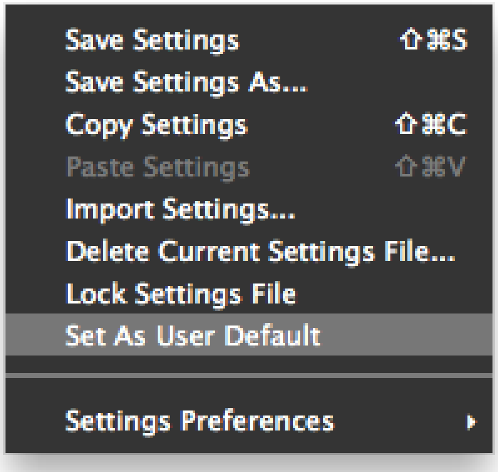
The issue here is that the pan position is in the center of the room which isn't a very good place to start. If we are used to the Pro Tools Panner, then it becomes a challenge. But like all good things, Pro Tools allows us to use our own setting as a default while loading. This is how it is done.

Load a stereo instance of the plugin on to a track and set the Y to 100. This makes the plugin look like this.

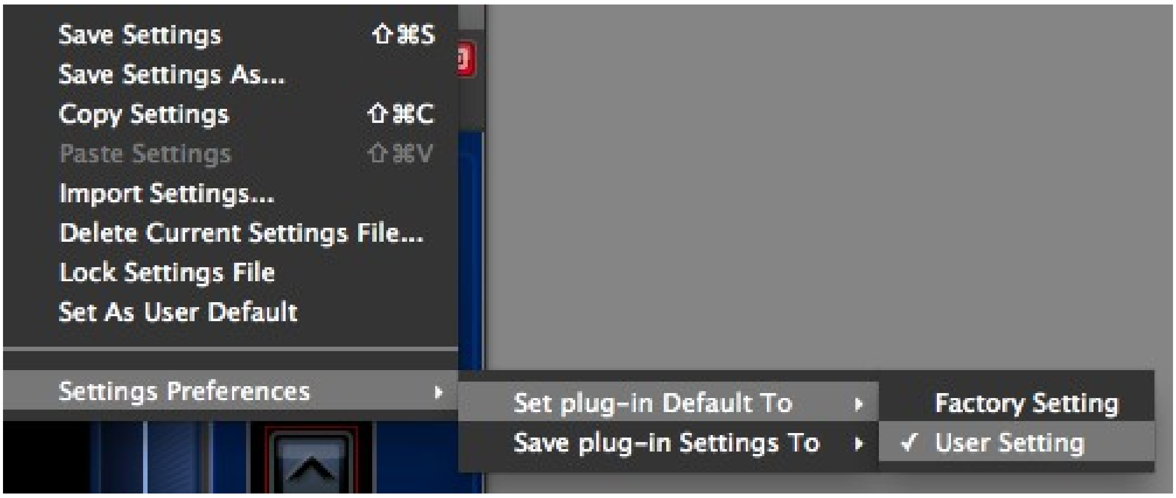


Atmos Stereo Default

To make this load as a default, all we have to do is save this as something. I saved this preset as Front Default Stereo as you can see above. Then there are 2 steps to do.



Set as User Default



Set load User Default

Once you do both of these, what it does is essentially make this as a default setting while loading. The reason you have to do this on a stereo file is if you do it on a mono, the right channel on the stereo version of the plugin will default to the first picture. Once you do this, then things are smooth to go. You can also use the above technique for any plugin not just the panner.

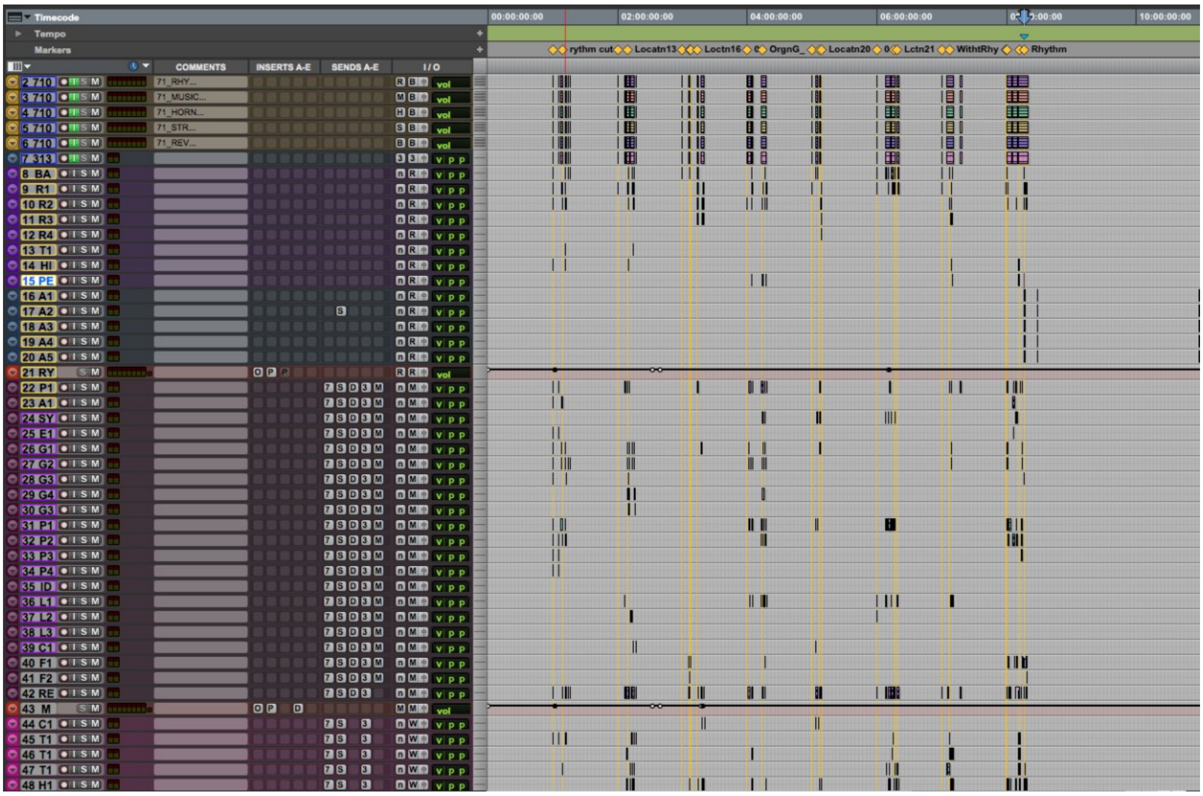
The Music Mix

The Score was done by Amit Trivedi and is mixed by Justin. So, what I am writing in this chapter is all Justin's work and techniques that he was very happy to share. All credit in this post is his.

The Approach

The movie is well connected with songs and score. The whole music was recorded on 96kHz and our sessions are running at 32-bit float. The reason for this I will mention soon. But the decision that Justin took and we discussed at the start of the mix was that the Score wont have an LFE track. Yes, the whole Background score of Bombay Velvet is mixed without using the LFE. This was a difficult move, but Justin has cracked this. The concept is that the stage speakers are full range and if used right, we can get the score to sound really good, thereby leaving a lot of space for the FX tracks when they need the energy. A lot of emotion is reliant on the score which both of us felt that should be achievable even without the extended bass.

Background Score



The Score

The approach to the score was to keep it minimal when we need to ride it. But that being said, we made sure there was space for everything to fit. Amit has done a fabulous score and so we had to do full justice to it. The first thing was to split the score into the components so that we record the 7.1 beds separately. So, we ended up recording the Rhythm, Instruments, Horn, Strings and the Reverbs separately. And anything we had for objects were kept on a separate track. We had to do this because we were losing out on Voices in the master session with all the Dialogue, Music and the FX section Pass-through. Also, because we were on 32bit Float, we would never clip in an internal recording. Pro Tools has got a 64bit-summing engine, which means the headroom is petty good. And if in case we clip, that is a signal indication. The file being a 32bit float file will have the capability to restore that just by gain reduction. But with this score, we didn't clip and we didn't use a limiter. Both of us aren't fans of using that in the score. It is important to let the score breathe its natural movement.

One thing to realize is that it is very important to mix to the visual. What does the screen tell you? That will give you the reason for certain instruments balance that can be changed. For example, very close shots don't require a very lush mix because the shot defines the character there. Also, the pans are very important to be placed properly. But on this movie, we had a lot of ideas that were only possible because of Spanner. Sometimes the score is emotional, sometimes it is a filler, sometimes it is part of the location and very often it can morph between all of this very quickly in a single part. This is very tricky to achieve if you have already eq'd the track a lot. That is where the approach is important.

Many times, we all put in EQs because we want to shape the tone or the sound. This time, we didn't at all. What is important to realize is that the score has come to us after a music engineer has balanced this. This means that the overall balance works when it is in stereo. What is done different as a film mix engineer then? Two things.

1. The Pans and placement
2. The dialogues and effects

Both the above will affect the rides that happen on a score. For example, there may be scenes where the ambience is taken off as a ride based on the dialogues and the expression of the scene. In such cases, the score becomes quite prominent. How then can we make the score seem part of the movie? By making sure the score has tonality that matches the ambience. Like, no over compression, or High Frequency etc. But remember, like all things this too changes. Many times for example if there is an open hat played for a scene link and between two pieces, I have the tendency to drop the score to maintain the tiptoe feel we are aiming for. At that time, yes I may raise the extreme highs a bit for that shimmer to sustain, but then if there are night crickets, then it will be a different call. So, how can this be done before hand? Simple, it can't be. There WILL be changes and corrections and that is the fun of it. And the rides on the score too are tried to be as musical as possible by following the tempo. But sometimes that may not be possible due to an effect or a dialog on screen. In such cases, the ride will need to be covered up with something. That something can be a car pass, an ambience ride, a shot change etc. There is a lot of misdirection that can be done in a film. It is like Magic. The audience's attention is taken up by screen. This can be effectively used to cover the rides as long as the surrounds don't drop drastically. This is what will cause the move to be noticed. Those surrounds are covered by ambience.

This is exactly how Justin and I work. After he has done the mixes, I sit on the score to understand what is there and what are the elements. This is because I need to know how to blend my effects around the score and at which portions the score goes around the effects and dialogue. Once I understand this, I can anticipate his rides and move my faders based on that. It will not be perfect the first time, but both of us get a sense of the space and the area we need to cover and also very importantly what works for the movie. I am not hell bent on my effects being heard or Justin on the score. We have to compliment each other. The other big advantage is that both of us see the movie in a new perspective. Me with score and him with effects.

Placement of Instruments

The placement of the instruments within the score was planned based on an initial discussion on how we wanted the mix to grow. The scenes define how much they open up and how much is in the front and how much is in the surrounds. As we progress, the usage of objects in Atmos also changes and helps define the whole mix. This will be talked about more in a later note on the mix. But in general the instruments had different placements based on the genre of the score. If it went from Jazz to a synth, the element placements would move from a lush sound to a more modern tone. To create a warmer and comfortable score mix, we avoided placing elements with higher frequency in the surrounds. So Strings, Horns, Pads, Reverbs all had the leak and sometimes placement in the surrounds. We also avoided the Wall Mix (where the panner in Pro Tools is on the border) and brought a lot of the instruments into the space. This also was done keeping in mind that the film uses this space a lot. It is a very effective way of transforming the audience from being a spectator of events to being in the events. We didn't follow a pattern but this was the general idea.

Songs

Because I am not at a liberty to talk about the scenes or the songs in particular, I can tell you this. The Songs were mixed to the space it was in. This was an approach we hadn't taken before but came during the mix. And Justin's work on this is fab. The reason was that once we realized this, the positioning of the elements and the reverbs changed. We mixed songs that go through multiple locations and scenes and placements. For this, the balance was very carefully done to the stereo mix, then changed slightly in order to fit the dialogue and ambience texture and an additional reverb and movement was created for the space of the song and its context. Little things like this take a lot of time but that's the fun! Once we get a good result, then we have a responsibility of making the audience feel the same as we did! Kunal, Justin and I truly believe we managed to do this, but there are days ahead, more work ahead and we may change if change is needed. But all for the good of the film.

The Dialog Mix

This film is quite an experience for me and frankly is the most complex film I have attempted to mix. While like all films, the dialogue plays an important and central character in the movie, the dynamics of it is so much that we had to figure out many ways to make it sit in the mix without the listener knowing that something was raised or lowered. Of course, all seasoned Mix engineers have their techniques and something that works for them. I will be talking about what I do and what Justin later does.

The Recording

During Gangs of Wasseypur, Kunal had recorded some tracks using an MS technique (Mid-Side where the sound source can be captured as a true LCR with the Middle and the Side images. More info here.) At that time, after decoding the MS to a stereo, I used the Dolby LtRt Decoder to decode that stereo into LCR. This time, he went ahead and had captured multi MS with a separate one for Crowds. So, there was a forward or camera perspective and a rear MS for some cases. The recordings were on a Deva and at 96kHz.

He also had lapels and booms in the location. The idea is that we always have a camera perspective of the shot with the MS. If we want to get close, we have the lapel and if we want the room and overall tonality, we have the Boom. I always mix them together with the balance needed to maintain consistency. The recordings were clean and I can happily say that around 85% or more of the movie is location sound. The only ADR that was done was for creative decisions or in some really challenging sections.

Dialogue Editing

This was Zahir's forte. One of the challenges was that the Avid Media Composer was having challenges editing using the 96kHz audio. They were down converted to 48kHz. The Pro Tools Session was then reassembled from the EDL using Synchroarts Titan. The most challenging part begins then. What takes to keep, what to ADR, how to clean, creating the fill track, splitting the Production Effects into a different track for the MnE etc. And all of these keeping in mind the requirements for a mix stage of track arrangements.

He had supplied me with the tracks assembled as Boom, Lapels, MS, ADR, Production fx (PFX). One of the most important things to look and understand is that the Boom and the Lapel may differ in phase. It is not always 180 degrees. This is also important because the tonality will change and that is not something that can always be corrected with EQ.

The Mix

I have a confession here. The first reel I start off will never be the perfect reel. It is true because it takes some time to set the tone and the signal chain. By the time I am on the second reel, I usually have a good idea of this and then go back to the first one after a few reels. There is nothing wrong in not getting it right the first time. This used to be a stress for me initially but now I have gotten around it. The first thing I do is listen to the dialogue on the first reel. I have my EQs and Compressors ready at the factory default for EQ and 1:1 ratio for the compressors. I don't want them to act out just yet. I also have my reverbs and delays set up in the session and leave some space for any crazy thing I might do if I strike upon an idea in the middle of the night! There are something's I look for in this run. I have a detailed writeup here and here and not much has changed in the overall workflow so I will write about the additional things I do.

MS Decoding

The ambience and camera perspective of the shots were recorded in MS too along with the Boom and Lapel. What this gives is a very clean space of the location. Including some very beautiful Room reverbs, Exterior slaps, exterior reverbs etc. But the issue with MS is that it is decoded in stereo. The center element is made from a phantom center. I came up with a technique to decode this. If you remember the mix notes of Gangs of Wasseypur, I had briefly mentioned this technique. What I did here was to take the MS Decoded track and run it through a Pro Logic II Decoder from Neyrinck. What this does is essentially treat the stereo as an LtRT track and generate a 5.1 from it. This means the Camera perspective will be accurate in 5.1 and the Rooms you will hear in Bombay Velvet are the Real rooms of the location and not heightened by reverbs. Reverbs were used where needed and for matching.

But Kunal went one step further and recorded Dual MS in opposite directions. One facing the Character and the other the other way. What we did in the Mix was to decode both and flip the rear MS Pan so that the LCR is actually along the Back Surrounds and the Ls Rs would be along the front. When combining these two tracks, the space in the theater is amazingly real. Not just that, when there are crowds or characters falling off camera or talking and walking away or even having car passes, they are real and you get the sense of being in the location. To this, we added some subltle voices and crowds as objects to make the feel even more real by matching camera pans and panning these across the speakers.

To summarize, the Beds will pan automatically due to the recording, the ambience will be spannered and the real feel of the individual speakers will come from the few objects that create the realistic feel.

Decoding

Dialogue Matching and EQ

This is a very important aspect of the dialogue premix and what takes the most amount of time to achieve. What I used was this beautiful AAX plugin from Nugen called Seq-S. This is my go to tool and my not-so-secret tool. The Seq-S is a linear phase EQ. This means that there are minimal shifts that will happen when an eq match is done. This is most noticeable in the lower spectrum if done with a regular eq. Especially if you want to match something in a Room or an exterior shot. This Magic plugin takes the entire burden out of the job and goes one step ahead by letting you decide how deep or accurate you want it to match. Sometimes you don't need the accuracy because the shot changes will require you to have that slight difference.

But that is not the only place this shines. It is an amazing tool for invert EQ. If for example I want the dialogues to sit within the crowd without lowering the crowd or music, I can take the spectral sample of the dialogues, and apply the *INVERSE* on the Crowd or Background. The depth setting also allows me to decide how deep I want the dialogue to fit in the range. Because it is a very complex filter and linear phase, it lets me fit it with just the precision I need and not letting the listener feel that we have cut out frequencies for the mix to fit.

My Go to EQ for notching and tonal shaping is the [Channel strip](#) from Avid. This plugin is in fact an exact replica of the Channel strip of the System 5. This means we get the sound of the large format console right here on a Pro Tools System and it's free. It also doesn't take a lot of resources and can easily sit within the mix.

Channelstrip

The Oxford EQ was my clarity and body EQ. The reason I took this was this is the only EQ that I know (other than the MDW) that has frequencies up to 40kHz while on 96khz sample rate. (Remember Nyquist Theorem?) What this also gives is a very clean sheen and finesse to the track without the noise floor being pushed up.

Sonnox

Reverbs

I have two reverbs that I frequently use. The Revibe and [Phoenixverb](#) Surround from Exponential audio. The Phoenixverb has a setting called 3D. What this allows us is to link two instances of phoenixverb and assign one of them as the bottom bed and the other as a top. This is also quite good for Auro 3D format. Now, what I did is to have the 7.1 and a stereo setup and linked and the stereo would be routed to the overheads. This gives a very realistic overhead reflection and creates the space very well. The revibe is my instant reverb for slaps and halls. Although the Phoenixverb was used in this wherever the dubs were to be matched.

Another technique I used to match reverbs is with Altiverb. What I did was to open up the Dialogues to an empty section with room noise and export a small section of it to the desktop. Then I drag and drop that into the IR section of Altiverb. This creates a reverb that is very close to the real space. Then, I use this to process the foley or fill tracks to create the room space to match.

Multiband Compression

My new favorite multi band compressor is the Pro Multiband from Avid. This is a hybrid plugin in that the algorithm is processed on the DSP of the HDX card while the graphics on screen is taken care of by the CPU. One thing very interesting about this is that the crossover filters are Linkwitz-Riley 8th order filter. That's a very strange word to hear if you haven't understood what they do. Essentially a band is made by a set of 2 filters that have a level drop over a frequency range. So, 6dB per octave at 100 Hz means it will drop by 6dB at 200 Hz and 12 dB at 400 etc. Now these shelving filters will cause a phase shift depending on the frequency you are applying it at. It is important to know the filters and their dependency on the signal level.

Essentially, phase shift is delay. The reason it is in degrees is because the delay time is dependent on the frequency. So when we say like 180 degrees phase shift, it means
Phase = freq*360 / (time delay).

So in this though the phase shift is fixed for a type of crossover, the time will change for frequency. A Linkwitz-Riley is actually 2 butterworth filters in series. This is why LR filters are only even ordered. (2nd order, 4th Order, 6th Order, 8th order etc.). The phase shift is order number * 45 so for 8th order it is 360 degrees.

This should not be confused with 0 degrees because 360 degrees in the above formula will give you a different time of arrival for THAT frequency. This will be different for other frequencies. So, addition and subtraction will vary.

What is also interesting is that the slope although is 48 dB/Octave, the frequency is dependent on input signal level! That means if you push the signal, for example at 100 Hz, the level is calculated as 100 Hz will be -6 dB and 200 Hz will be -48 dB. But this is provided the signal is at 0 dB. If the input is increased by let's say 6 dB (easy for calculation) then at 100 hz, the level is 6-6 ie 0 dB! This means the center frequency is no longer 100 Hz but 110 or so. Because that is the frequency at -6 in this case. :) -6 is the value because for the Linkwitz Riley (LR), the center frequency is the one that drops by 6 dB. The advantage is that when you have a High pass and low pass LR the total drop is -6 at the crossover but since acoustical summing of 6 db happens it becomes flat. Acoustical summing of 6dB always happens at all crossovers. If we used a 4th order LR filter instead, the total drop will be 3 dB so at the crossover the signal will be 6-3=3dB more. And this is why splitting the signal with the multiband splitter plugin and recombining it doesn't produce any change. It's the 8th order LR advantage!

This is also why the middle bands are called floating bands and not crossover bands because for crossover band it needs this summing to happen while we can decide the mid bands and slopes in this case. Also floating bands are defined by the center frequency unlike the crossover frequency, which is defined by the dip. One thing in the C6 is that you can screw up the bypass state. If you bypass the plugin, adjust the crossover frequency and then unbypass and bypass again, the signal changes! This means it is not a true bypass. It still runs the signal through the plugin filters and recombines them; and just bypasses the individual bands and because of the filter property we just saw, it can introduce phase issues even when it is bypassed!

I know some of this may be a bit difficult to grasp in the first read, but this is also what is very important about these compressors. If you try parallel compression and then get phase issues, you know why. In addition, it is also useful to know why sibilance and some frequencies suddenly jump up when doing a level ride into a multiband compression. As you see above, it changes the band itself. All this being said, this is by far the most transparent Multiband I have used and is going to be my weapon of choice for a long time.

Avid Pro Multiband Dynamcs

Revoice Pro

An amazing piece of software made by [Syncroarts](). This is what saved us a ton of time in matching takes. Not only does it match intonations, it can also tune pitch and energy of the Dubs against the Location. A brilliant piece of work.

Analog emulation

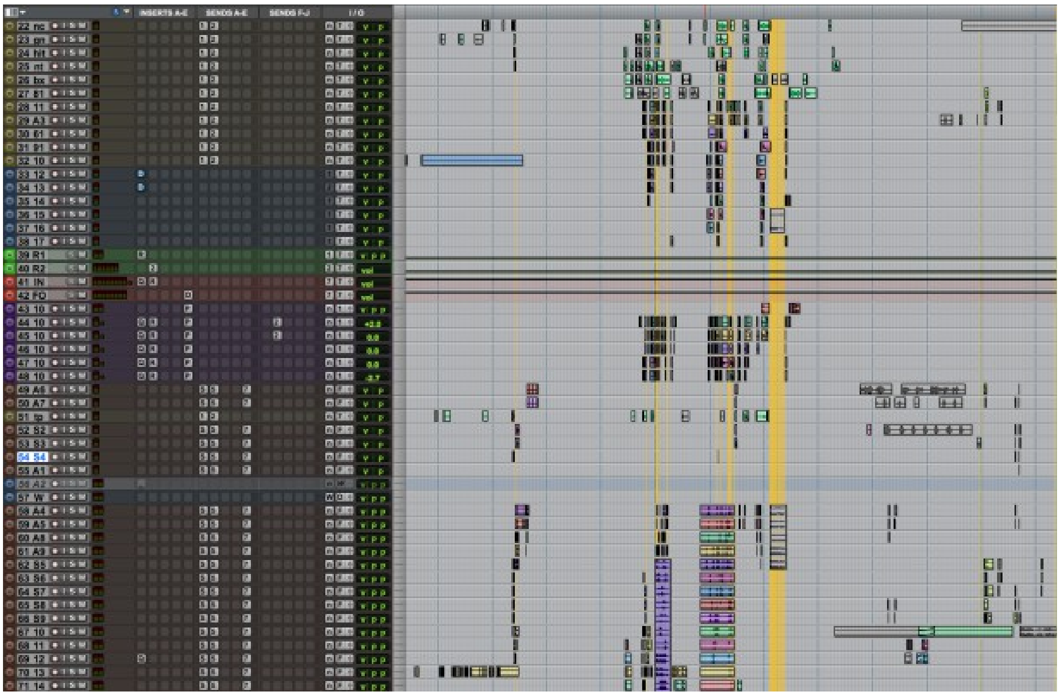
This was the most difficult and time consuming part. Our constant companion was Heat from Avid. This is a harmonic enhancer made by Dave Hill from the designer of Cranesong. Once we had this on our mix session, there was no turning back. It had to stay. Our next step was analog emulation in some parts for which we used the UAD Studer A800. This added that missing element to the body of the dialogues when we had cleaned up certain sections. It was also something that gave us so much warmth on very close shot and intense dialogues. That being said, adjusting the latency was a bit time consuming and yet we preferred to use this in portions of the movie. The other plugin on VO and other characters was the Reel Tape Saturation. This too was something that wouldn't work on all characters but when it did, it was amazing.

Once I am done with my premix, Justin then takes lead and tone shapes the dialogues to blend with the score. He uses a combination of EQ, Expanders and DeEssers in his chain so that the tone that needs to be attained sits well within the mix with no sibilance or annoying bumps that may creep in when the dialog levels will be changed on the demand of the scene.

Effects Foley and Ambience Mix

There is nothing fancy or new in this but there are some techniques that I used this time. Before I get to that, I would say that all the premixes that I did were by listening to the Score and Dialogues playing back from the first machine. This helps me to quickly workaround whatever clashes were there. Also what it helps is to find what scene works with the score and what would need to be based on effects. In such cases I would mix the effects with prominence and then Justin, Kunal and I would figure out the scene. But then that too would change based on what The Director would want. By then we already have an idea of how it would work either ways.

The Session



FX Session

The effects were running from a Laptop with 16Gb of Ram and an HDX2 on a thunderbolt chassis with Pro Tools HD V12. We had close to 200 tracks running at 96kHz. These were original source files recorded at 96k by Kunal and Gokul and not up converted from 48kHz. Preparing the session involved some effort, as we needed to know what to compile and how to group them. Earlier in my mixes, I would route separate auxes for elements if needed. This time, I used VCAs. For example, if a session had guns, ricochets, passbys, hits, breakings etc. on separate tracks, then rather than creating separate auxes and use Voices, I would send them to common outs and control them via VCA. If they need any common processing then I would use grouped plugins. My session would have masters in the order of FX, Ambience, Foley, Incidentals, FX Objects, Ambience Objects, FX Bed, Ambience Bed and Crowd. If there were guns etc, those would also be grouped. All of this was discussed with the sound editor Gokul. He gave me the sessions in the required split and this also helped in sorting the session to a manageable one. I was using the Avid S3 to do the fader rides and plugin control and this allowed me to create layouts in count of 16 Faders. They were also color-coded tracks.

I had LFE constantly sent through the Pro Subharmonic that gave me beautiful options for different frequency bands that were automated. My go to EQ here was the Fabfilter Pro Q2. This had frequency bands up to 30kHz, which was very useful in a 96kHz project. The R2 from Exponential Audio was my reverb. It is a very cool plugin that is absolutely remarkable for non-musical real elements like Foley and effects. On the foley I used the RA for a bit more body on the footsteps. Spanner was on the entire Master surround busses that I used.

The Routing involved the SDDS technique I have explained before. The object tracks were placed just below the last track of each section e.g. Foley will be followed by its Object, then FX and Ambience. One thing to note about the Dolby Atmos system as of now is that you cannot have a gap between object numbers. So if you have FX objects from 20 to 40 and Ambience from 45 to 55, the numbers from 41 to 44 cannot be left blank. There has to be a dummy object or instead of starting from 45, shift those to 41. In our case, we literally had no spare objects and ended up using all the 118 object tracks. This was also because the SDDS technique by itself uses objects for the pan.

Spanner to Object

The above figure shows how I used the Spanner to Object app in my session. Basically the beds would be rotated using spanner where needed. Once done, I would take the individual element like a wind or cutlery in a room (if the camera pans across) and put that on a track that is routed to an output that I called FX_Object_Pan. This bus has its separate elements to object outputs. So, the Sub path L would be in 95, C in 96, R in 97 etc. Each of these auxes will have the Dolby atmos panner with the pan position as L, C, R, Lss, Rss, Lsr, Rsr, respectively. Then by running the app I mentioned above, the automation would be copied into the Dolby plugin. Simple! A 2-minute job in less than 10 seconds!

EQ and Compression

I generally cut off the Highs and Lows in a Foley track but this time, the production recording was so good that I used most of that and ended up using Foley only where it was absolutely needed. The truth is we don't hear every footstep, tap or shuffle. So, there were a lot of rides on the footsteps to make them sound natural and also on the taps. In fact we took off most of them so that it makes the film more natural rather than forced. My major use of the objects was for panning within the field and not for the overheads. One thing we had decided in this movie was to make the mix grow as the film does. So, this is where the accuracy of Atmos helped. Very little compression was used in the effects, as they were extremely dynamic in nature going from barely heard levels to good ones. We made sure the mids were controlled so that we were able to create loudness without the pain of it. There were sequences that were really difficult to do, as it had to be big yet not hurt. I always managed to push it even further by controlling the upper frequencies. One subtle thing this does is it gives a very rounded tone with the attack where needed so that the mind makes up for the missing frequencies.

Panning

If you read my previous dialog post, you would have noticed that I like to have things in the Center Channel. Basically my thought is if there are speakers on the stage, use them to reproduce the sound and not rely on the phantom center. This is also why I like to have ambiences panned through the center and never leave it on Left and Right Channel. I am also not a big fan of using Divergence as it means we are simply sending the same signal across channels and not really using the space to be created. Instead it's like having a big Mono. So, what I generally do is rely on the Pan Law or use a separate part of the same signal (eg crickets or night ambience) in the surrounds. This gives a bigger separation to the space that is created. A lot of the effects that were MS recorded were also decoded as I had mentioned in the dialog blog post. Another thing that I used was to use the inner Left Center and Right center speaker for the low elements and the LCR for attack and brighter elements. This gave a lot of separation in the sounds and the Atmos Renderer took care of rendering it back to a 5.1 or 7.1 track.

Techniques

Invert EQ

This is one that I used with Nugen Seq-S on the Crowd tracks. What I did was to get a feed of the dialogue recorded, get the Tonal spectrum and use the invert on the crowd with automation on the depth of the EQ cut so that it is not felt that the crowds have a dip while the dialogues can clearly push through without boosting it or lowering the crowds.

Ghost Signals

I didn't know what to call this technique so I named it this way. Basically the idea is to open a signal triggered by another. Of course I could have used a gate, but I wanted the smoothness of a compressor's attack and release in this. The requirement was in a certain sequence where the character only would hear the crowd sounds when he was hit. It would vary based on the intensity of the hit and also the duration of it. I came up with this weird way of achieving it. I used the Pro Compressor with the settings as in the figure.

If you look above you can see there is a key signal called CrowdSC and what I did was to switch the speaker icon on. (The icon is just to the right of smart RMS AVG PEAK FAST setting).

How this works is like this. Whenever there is a sidechain the compressor would compress. What this speaker icon would do is play the compression or gain reduction signal rather the compressed signal. It is actually there to know what you are compressing. So this was put on the design crowd track and whenever there was a hit that came through the CrowdSC this would play the compression. Adjusting the attack and release gave me the option for a smooth open and close and the ratio gave me gain reduction based on the input signal. So a harder hit had more level of the crowd. This was then sent to a reverb so that it tailed out in the ear of the character. So the effect in words is as he is hit he hears the crowd open and die out for that one moment with reverb. It sounds very surreal but is actually something that would happen in a fight.

Apart from this, it didn't really change what I did in my earlier notes except in the mix where some more things were tried out. Some of the scenes like a moving car would have a mondomod process on the car sound to give the sense of vibration to the sound, as you would have in a car. A lot of the effects were done to match the music, which is why I used the score while premixing. I got a rough sense of the tempo and used that to set my pre delay on reverbs for musical scenes so that the effects and the space feel part of the score. It was also easy for me to fit the effect within the space with minimal rides in this case as it blends more easily. There were places where we treated the effects and blended them into the score so morphing the transition from score to effects to score. The LFE I mentioned was not present in the BGM while Justin mixed it and it was left to the Effects to compliment that.

Final Mix

The final mix is where Justin and I run the film together doing our rides at the same time. It wont be a polished run at all, and we get an idea of what the other persons moves are and how the scene is shaped. It is very important to realize that you have to trust your instinct. Many times, the first ride that I do is roughly close to what I want to achieve but I don't spend a lot of time to get there. I can achieve that later while polishing the mix.

The Mix

The Pass

The first pass is where all of our premixes are laid up and run. We don't stop at what is wrong or right and correct it there. We do our rides as we decide to approach the scenes. Then both of us get an idea of what each other has in mind and what needs to be worked out. After that, we go into it scene by scene fixing the music dialogues and effects. Justin takes care of the score and dialogue then and I would handle the FX, Ambience and Foley. The whole movie has a lot of detailing and we went to the extent of adding quite some Easter eggs for people who want to hear and find out.

The benefit of having two mixers is that there are very different approaches that we take as individuals to a scene or emotion. For example, Justin is a more soothing and smooth mixer whose fader rides are very subtle, while I on the other hand like to keep things real and raw even on the rides. I like to make it non linear where it demands. But the beauty of this is that this movie will get a benefit of having very radical mixes for scenes and it helps both of Kunal and us to get what he wants. He wants a particular emotion or approach and we treat it in our own way but with me on one side and

Justin on the other without clashing into each other. This is a beautiful way to work and it doesn't tire you. In fact when we mix, we constantly ask each other on the opinion of our moves as then that is a fresher perspective. So, I would comment on the music and dialogue rides while he would comment on the effects ride.

It is at this stage where we also get into a lot of design changes, as the tone of the mix will be clearer now. When I say tone, I don't mean EQ or dialogues. But what we are creating for the audience. For example, one thing we did is to drop music to create anticipation or drop effects for the same. The advantage is if we then want to take the audience into an uncomfortable or expected stage, we just do this. Psychologically the audience is ready for this then.

I have always maintained that a good mixer should have a good amount of editing skills. Now, this is just my personal opinion and I will say why. If I want to achieve something that I have in my head, I know what to do with the tracks I have or if I am not having the required elements within. As a mixer, it is very important to realize that you don't necessarily need to make the audience hear every sound that is there within the track. Kunal understands this well. So, if something is low or masked by some other sound, we just mute it rather than lowering it. It makes no sense to have that playing if it can't be heard.

Similarly, it isn't necessary to have the sound of everything that is on screen to be heard. It depends on what is the required emotion needed. If it is a busy street but we are staying with the character, then the business of the street can be present but it isn't necessary to hear everything that we see on screen. Remember even when we are walking or talking in a busy place, we don't listen to everything around us, yet we can hear the person we are talking to. This is a very important aspect to understand because the mic and our ear will pick up anything and everything. Our brain decides what to concentrate on. This is the pattern we followed on this movie. But that is not all. We have also worked a lot to set a pattern and then break it. This is a very important way by which one can create dynamic range.

This is also why this phase is very strainful and can have a lot of stress. It is being deeply involved with what's on screen, what's the sound, what's that you want to achieve, how do you want to achieve, what is right what is wrong, what can be better what is the edit, why was this edit etc. all at the same time. None of this is a conscious decision. It comes only by being associated with films and by starting off as an assistant and learning whatever is being thrown at you. The importance of having been there can never be understated.

How loud is loud?

This is a question that I have been asked a few times. In the music industry and the broadcast world we have the R-128 or the ITU-BS-1770 and the variants for measuring loudness. They are very much needed in today's loudness wars in that industry too. But the same also happens in movies. It is a common fear for Mix engineers that the projectionist will lower the mix and the same for the Projectionist that will clip the amplifiers. One thing that happens is that the mixes will get louder and the projection will get softer. What can be done in such a case? One thing I have seen is that controlling the mids will give a lot of power and also helps cut dialogues through cleanly. When something is supposed to be loud, it is for the power and not for the level. Because when you are looking at the screen there is no level meter, just what you feel and hear. What is termed or felt, as power is the lower mids and lows. So in fact taking off the highs will get the audience to fill that missing elements if the attack is left in. Another technique we have used in the mix is kind of like fill-in-the-blanks. We build the spectrum with more frequency elements of the one we want to remove. For example, if it is an explosion, the score will have that element of highs or brightness in it for an extended period before it. What happens is that the mind will automatically mask it off. It's like a blind spot. Once you reach that point, then the next sound you introduce need not have that frequency as the mind will cover and make up for that missing piece. This means you can have a cleaner and bigger sound without being harsh on the ears. Of course then managing this across the channels and speaker will take time and is something that one will have to do only by thinking over it and making sure that the right notes are hit. This is where I said that as a mixer, you need to know if you have all the elements or no. If it is less then you need to know what is needed or how you rearrange the tracks and elements to make it sound good. Because lets be real. At the end of the day, the audience isn't going to comment on the bird tweet you put in a scene but rather on the emotional value of the scene which would have been lost if something as simple as the bird tweet wasn't there.

Both of us use the metering of the DMU (Dolby Mastering unit) for the levels and peaks. While it is useful to have a PPM metering, we are used to VU on this and help us to set the dialogue levels on an average. What is an optimal level? If the dialogues are between +12 and +6 on this scale, then it is ok. Anything below this will only be useful in scenes and not if it is consistent. Once the dialogue levels are taken care of, everything else rides around it. During the Mastering process when the final is being made, we look at the LEQ values. This is an averaging loudness for the program. Trailers are fixed to be 85. A film usually cannot contain a whole LEQ level because reels can be extremely dynamic. Yet, it is a responsibility to have a proper listening level so that the audience is not blanketed with sound and they are not fatigued.

Reel	Beep				FF				LFOA				LEQ
1	00	59	44	01	00	59	46	01	01	17	22	09	82
2	01	59	58	00	02	00	00	00	02	16	58	17	80
3	02	59	58	00	03	00	00	00	03	20	59	07	79
4	03	59	58	00	04	00	00	00	04	18	30	02	79
5	04	59	58	00	05	00	00	00	05	09	00	00	78
6	05	59	58	00	06	00	00	00	06	20	53	09	79
7	06	59	58	00	07	00	00	00	07	19	15	02	84
8	07	59	58	00	08	00	00	00	08	21	00	00	84
9	08	59	58	00	09	00	00	00	09	05	15	00	84

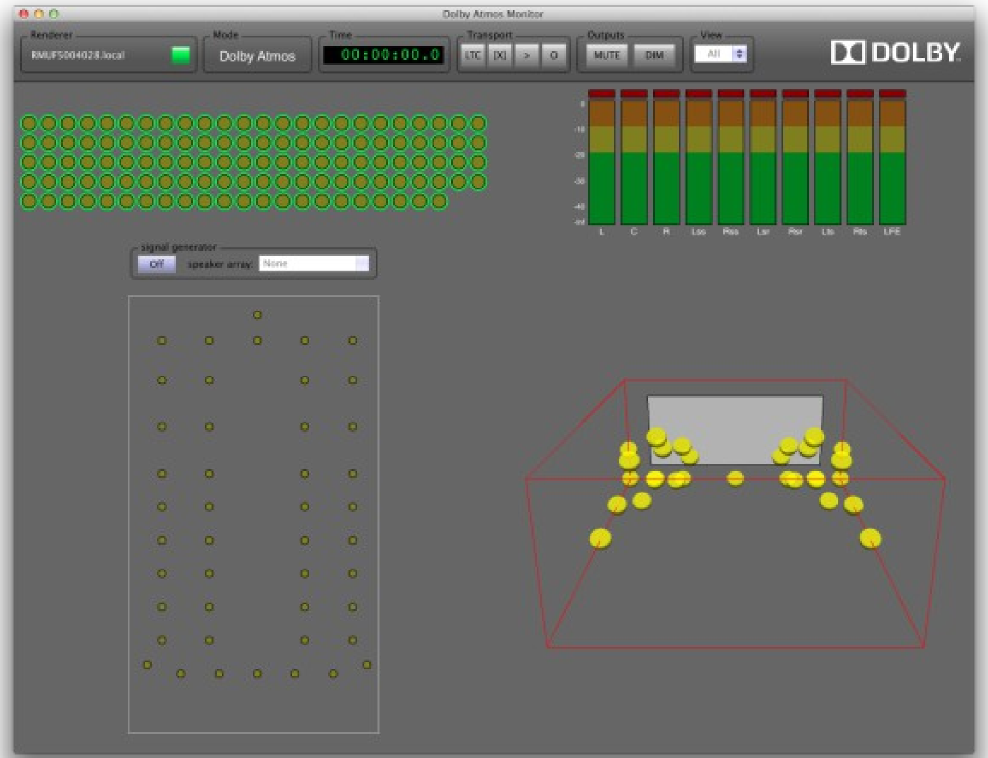
Our loudest reel is 84 LEQ.

Dolby Atmos Mixing

Let me tell you something that I have experienced on this project. An Atmos mix won't take considerably longer time than a regular mix. It will take more time, but not too much for the mix. It's the Decisions that hold you. The thing is that pans and resolution changes in Atmos compared to a traditional mix. What changes is for the mix engineer to set up his master faders and prepping the tracks so that the mix is achieved in the smoothest way. That's it. Once the pans and overheads are placed, then the mix is not different from a traditional mix in terms of the philosophy or techniques. This is also why I am now fully convinced that getting a 5.1 or 7.1 downmix from an Atmos mix will give you an unbelievable result and I can personally say that it is impossible to achieve that 5.1 mix on its own. Maybe you can but that would take considerable time to get to at which point one will lose all objectivity of the mix. So, even if the mix isn't releasing in Atmos, I would suggest doing it in Atmos and taking the down render. The way it handles a pan across the surrounds in 5.1 is pure magic for me and is really something I look forward to in the future too.

The Mastering Process

In this final installment, lets look at the Mastering process for Dolby Atmos during Bombay Velvet. Bharat Reddy and Dwarak Warrier gave the Mastering and workflow help from Dolby India. They are dear friends and the resident Dolby Jedi Masters! This step is very little documented online and I will try to explain it as detailed as I can.



Mastering

One thing that is different in a Film mastering compared to mastering songs for audio is the meaning of the term itself. In the Mastering process for a CD or iTunes etc., great care is taken to ensure the different songs are similar in levels, the dynamics are clean, and the song breaks are ok. There will be a master compressor, Limiter, EQ, etc, and many times, the song mix will sound different after mastering. None of this happens in a film mix master!

The reason I mention this is because I was asked quite a few times as to what mastering plugins do I use, what compression is used during the final master etc. The reason is that Film Sound is spread over a very long period and so the mix itself is done to sound the way it is intended to sound. There is no final overall glazing process that I use. I am not sure if that is the case worldwide but I would definitely think so. Any compressor or limiter would be in chain during mixing itself.

Preparing for Mastering

For the mastering, the sessions are prepared with all the start marks and first frame of action (Commonly called as FFOA, which stands for the first frame of the reel where the movie starts. It is 48 frames after the beep) and the Last frame of Action (LFOA). Once these are done, the timecodes are entered into the RMU's (Rendering and Mastering Unit) interface. The control is via a web interface on the RMU itself. Once done, the playback is started 15 seconds before the beep. The reason is to also have a buffer for the file while doing overlaps. This time, since we were running two systems in sync and didn't have an additional system to record the Atmos Printmaster, the final atmos mix was recorded only on the RMU. Simultaneously, the downmix is recorded onto a separate recorder in 7.1 from which we created the 5.1.

The Mastering Process

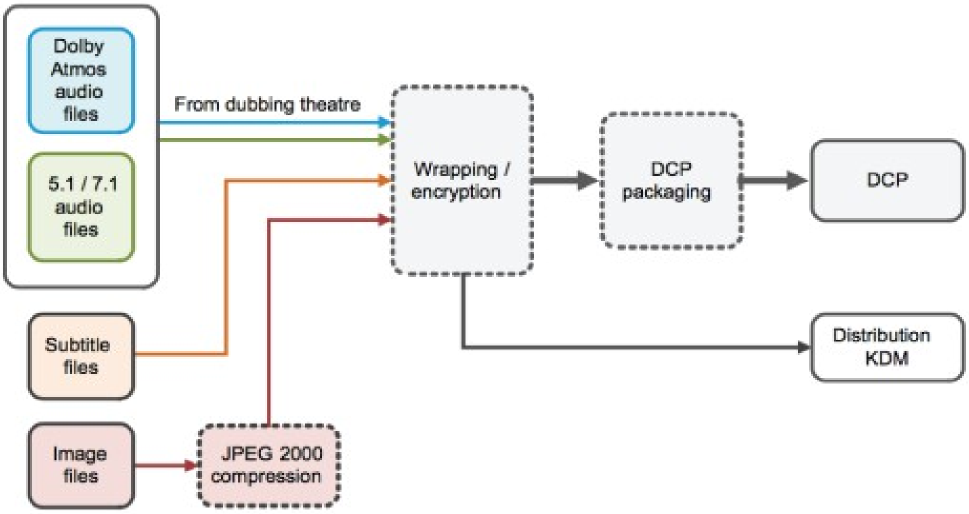
The Mastering Process involves the RMU recording the mix that we send to it via the 128 inputs on MADI. The process is done reel wise. Basically we run the mix and the RMU records.

The Transport section that you see is only active during mastering. The RMU requires LTC (Linear Time Code) for sync and mastering. Without that, the mastering wouldn't trigger. The Web interface on the Mastering unit has the option to show where the files are being recorded to. It creates 48kHz and 24bit .wav files. The first 10 are labelled as the beds and the remaining are the objects. So, what is created after a recording pass is:

- Ten mono .wav files, which make up the 9.1 bed
- One .prm and one .wav file per object
- One dub_out.rpl file

The dub_out.rpl file is basically an xml format file that has the details of all the associated wav files that are recorded. The .prm file contains the panner automation that we make on the plugin. This is also recorded and each object will have its associated prm file.

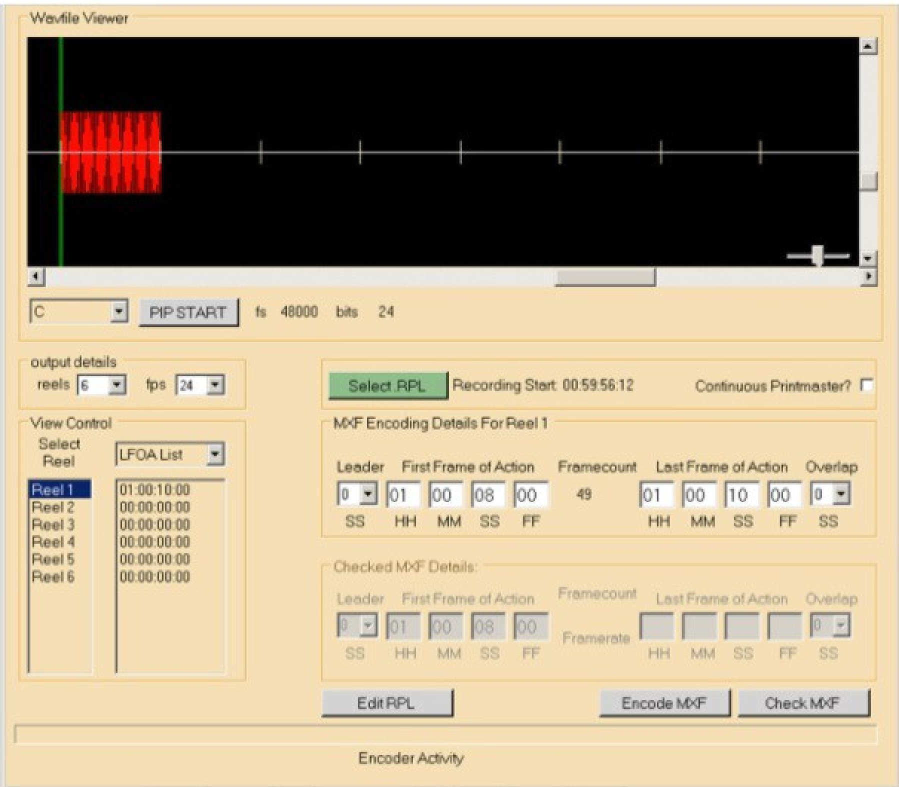
The Encoding



Atmos DCP Process

Once the Mastering is done, it has to be encoded into the [DCP](). The DCP that is made for atmos has some requirements. The Original file comes from the DCP package that contains the 5.1 mix. Once that is made, the file is sent to the Dolby consultant with a KDM. KDM stands for Key Delivery Message. It is a license file that specifies the limitations of an encrypted file like when should it play, which theater it should play etc. The KDM is provided for a particular reason. When the consultant gets the DCP with the 5.1 embedded, they have to unwrap it, and add the Atmos tracks into it. At this stage, there is one step that is done. Once the mastering is done, it has to be encoded into an MXF format. It is this MXF that is added into the DCP. But, the DCP is always made, as a first half and second half, each of which is an individual clip. How is it that the atmos that has been mastered reelwise converted into this?

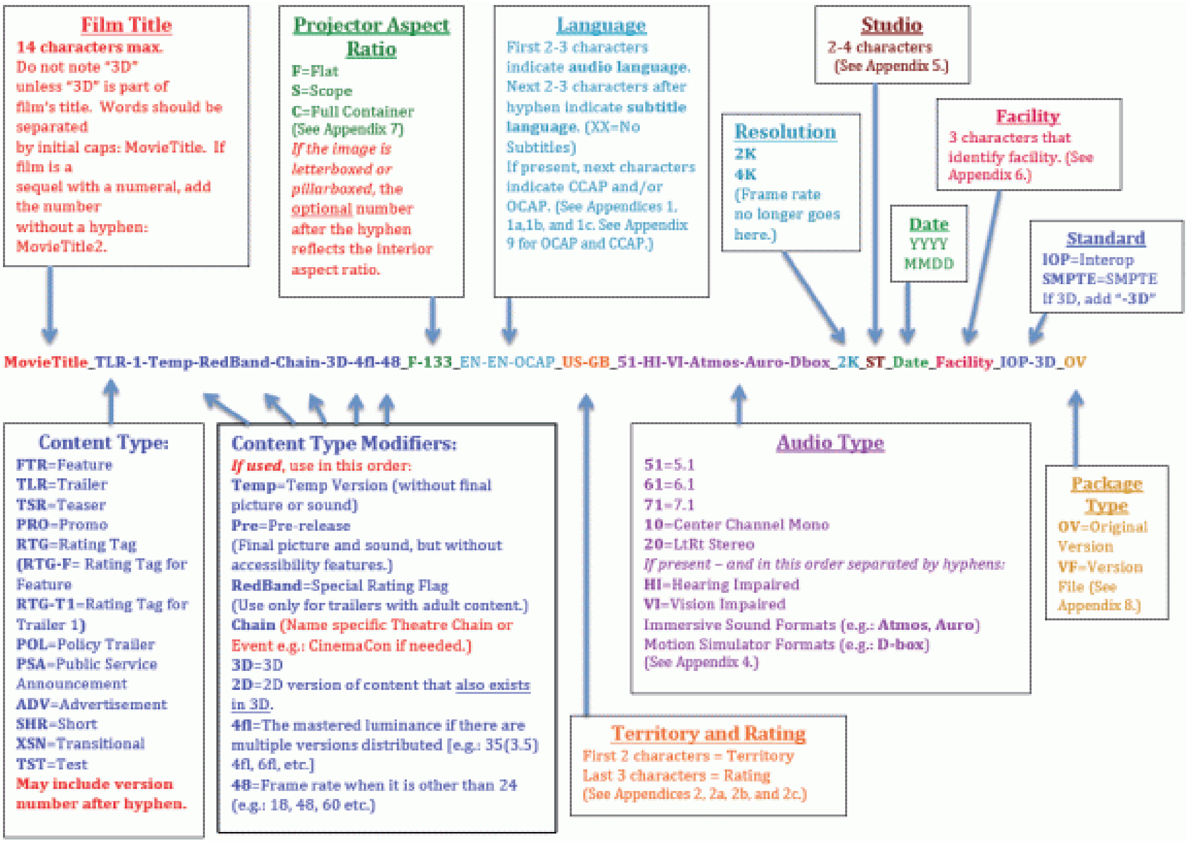
There is a tool that is used which can match the Beep and also “stitch” the multiple reels into a single first half and second half audio. One of he biggest issue is usually sync. The Atmos encoding method allows for the tracks to be slipped as they have buffers to do so.



Atmos Tool (Image taken from the RMU)

At this stage, there is one important thing that needs to be considered. It is called overlaps. An overlap is basically any music or sound that extends beyond a reel and needs to be taken into the next reels beginning. This is usually music or the tails. Now, the issue with the digital format is it is sample based. If there is a cut that is made on a sound that contains low frequency then you will hear a click at the reel change. To prevent this usually the score is ended before the reel ends or has a highpass that can be run at the very end on the transition.

So, once the reels are stitched, the atmos tracks are added into the DCP. The DCP has a naming convention that is followed.



DCP Naming Convention

The DCP created by the Atmos encoder is a SMPTE standard. The usual standard followed by Qube or Scrabble is an interop standard although there is talk that the SMPTE will be the standard for DCPs in the future. You can read more about it [here](). Once all of this is done, we have an Atmos Package that can then be played back in theaters.

This concludes the entire workflow that was used during the mix in Bombay Velvet. I hope you had a great time reading it as I had mixing it and documenting it. I hope it was useful for all of you. Here's wishing to great soundtracks and techniques and more importantly learning and sharing.

Fun in the Mix

There were somethings we did in the mix which were there just as a reminder to the older era and some as an easter egg. Now that the movie has released and the story is out, I will mention some of them.

1. The title sequence is mono. It is run through a Studer plugin from UAD at 15 ips and also had an EQ that was very close to the RIAA curve but tweaked for come clarity.
2. As the title opens to the song, the EQ opens to full spectrum and there will be a perceived fullness and clarity because we band limited the audio till now and also opened to 5.1 giving a feel of more openness than what is actually there.
3. The first sound is actually in the very back and is the sound of a projector. The room of the projector was closely matched to the reverb in the actual Projector Room in Rajkamal. It is positioned there as a reference and ode to the projectionists. This technique was an afterthought that came at the very end on the last day of the mix!
4. The film story starts in Mono and Ends in Mono. The mix graduates to 5.1 and then 7.1 and then overhead speakers come in towards the 3rd reel nearing the interval. Post interval it is mostly object based. This was also to represent the change in time of the different format across different periods as the story progressed. Although no one other than audio guys will get this!
5. There are several songs from the time period embedded along with the score that one would hear only if they noticed. It was like you hear it or feel it.
6. The predelays in the fights during a score or song is timed to the tempo of the song or score to make it sit better and feel musical.
7. Pans of sounds that were non digetic (off screen) would be done opposite to the character's look on screen for 2 reasons. One was to create an additional depth, the second was to show conflict of the characters on screen.
8. All the songs were mixed and then reverbed to the location it was playing in. The reverbs also graduated from plate to room modelling as the movie progressed. So, dingy clubs and the finesse clubs had different reverbs. (more like the poor man's reverb to the rich posh one!)
9. Vocal perspectives were done based on the focus of the camera. So, if it was blurred, highs would be rolled of a bit and bought back as the character came in focus. This was subtle and done to match the focus on screen to what the audience has to notice at the time.
10. All of the pans were achieved with Spanner. The dialogues were panned on the Decoded MS tracks to get a real space movement.
11. We didn't duck effects, score or crowd for the dialogues. We let the dialogue cut through it.
12. Dialogues were pitched to the score in some places so that it fit well without riding too much on the music.
13. We took audio or visual cues to drop or raise the score drastically. Like a clap or a scene cut or look.
14. All major action are followed by silence to create a breathing space. To heighten tension, sometimes we just kept the inhale of the character and cut off the exhale so that it is more like holding one's breath.
15. Effects were shifted in sync by a few frames especially on guns to match the score. We treated the FX as part of the percussion and there were instances we took it off to let the percussion drive it and come back to fit the space. This created an additional dynamics where it wasnt level or pan but different elements coming and going that made things more fluid in the scenes.
16. The Mixing style changed from using reverbs and delays for design in the beginning of the movie to complete Modulation and morphs and Filters as the movie progressed serving as a reminder of the way design changed according to time. The mix also started with Dialogue prominence to Dialogue Music Prominence finally going to Dialogue Music FX blends.

Not all of these are meant to be noticed or to be heard, but this was what we thought and discussed in the beginning amongst us and this was how we achieved some of that.

Additional Thoughts

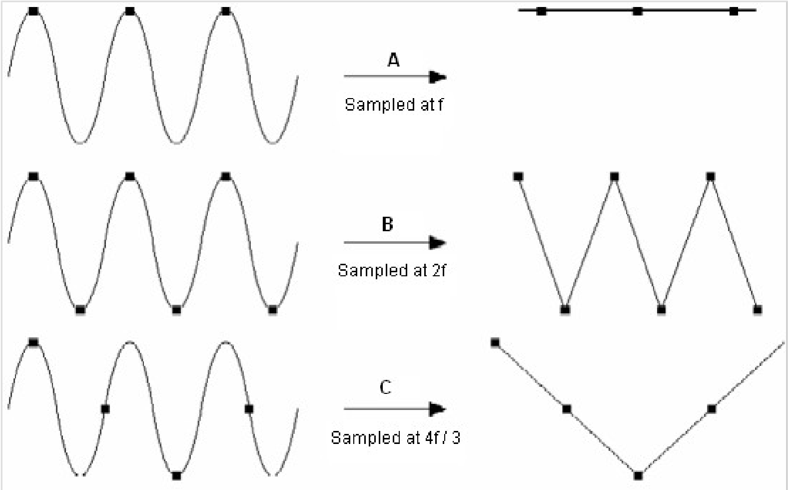
I have decided to write some additional notes on what I thought would be useful to help in the understanding of some of the techniques above.

The adventures of 96 kHz

The reason we decided to go for 96 kHz is definitely quality. But if you think about it, there is no speaker that can reproduce anything above 24 kHz. In fact there are some amplifiers that will trip if the input frequency crosses 21kHz. What reason did we have then to go through all this trouble? Because if you think of it, the file size is almost double, track counts are less, CPU load is more, not all plugins work, you record unwanted frequencies, etc. But I wanted to know more about the pros and cons and the benefit of this. Surely it cant just be a number and so people want to look at it.

Sampling

Sampling frequency is a very important factor in this. I have heard comments sometimes saying that high frequency tones above 30kHz etc were used in scenes in a film that was mixed at 48kHz. Whether this really had an impact or not will purely be a placebo effect, but yes it will be reproduced, at a different frequency. The reason is Nyquist Theorem. The theorem states that: ***A sampled waveforms contains ALL the information without any distortions, when the sampling rate exceeds twice the highest frequency contained by the sampled waveform.*** This means that if your session runs at 48 kHz, it can only reproduce frequency upto 24kHz of audio. There is no way to reproduce a 30kHz. In fact, there are some interesting things that will happen at higher frequencies. Lets say there is a tone at 30kHz. Because it is higher than 24kHz, it ends up being reproduced as roughly 16kHz (look at the image). As the frequency goes nearer to the sample rate, the reproduction approaches 0. So, if you have a frequency of 48KHz in a sample rate of 48 khz, as far as the sampler goes, it will be 0 Hz. Its like film. Film is 24 frames per second. So, if you have a light that flashes 24 times a second, when you shoot that, it will either be on or off because that is the state captured on film. (If you look at a TV scene in an old movie, you can see the image shaking and a line travelling across the TV. This is the reason)



Sampling at different rates. C is equivalent to higher frequency than what the Nuquist theory says.

This reproduction of a different frequency at a higher than Nyquist Limit is called Aliasing. (From the word Alias which means false or assumed identity. This is exactly the case where 30kHz is represented as 16 kHz!). Many people compare high sample rate to a high resolution picture. The analogy being that just like higher Pixel count (resolution) means a clear picture, means higher sample rate means a clear sound. No. If the source does not produce anything beyond 24 kHz, then recording the source at 96 kHz will not yield any information because 96 kHz can record frequency upto 48 kHz and there is nothing above 24 kHz in the source! That being said, aliasing is a very noticable issue. This is what differentiates many Analog to Digital converters. To prevent recording of anything above half of the sample rate, there will be very steep filters in the conversion. (Lets take 48 kHz as our base sample rate). So, to prevent frequencies above 24 kHz, a filter will be introduced. Now, with steep filters, there is phase issues, and aliasing issues at that frequency. So realistically, the converter will have to filter from around 22 kHz to be on the safe side. This will bring tonal changes in the audible region.

Why 96 kHz?

If we use the above analogy, we can see that the converter filter in this case would have to be at 48 kHz. That means that it is far above the audible range even if aliasing occurs. This means that we get much clearer highs. Another benefit is that when we have pitch changes, we are shifting lets say by an octave. In the earlier scenario in 48 kHz, the highest frequency we have is 24 kHz that would end up at 12 kHz after the shift. There would be nothing above that. This is also why pitch shifting in sound design sounds dull. But think of 96 kHz. We have information upto 48 khz. Pitching would still give us upto 24 kHz and that means cleaner highs! But we cant hear anything above 20 kHz, so why so much effort? Well, we cant hear it but can certainly feel it. How many amps and speakers can reproduce is another question. But still we get so much detail without aliasing that eqing, compression, pitching and processing become really clean. Not only that, it can also distinguish between arriving times. The human brain can discern a difference in a sound’s arrival time between the two ears (Interaural Time Difference) of better than 15 microseconds, which is around the time between samples at 96 kHz sampling. (96000 samples is 1 second. So, 1 sample is roughly 10 micro seconds long, or distance between samples is 10 micro seconds. This doubles in 48 kHz samplerate.)

Is 48 khz Bad?

Absolutely not! Today, many eqs, limiters and processing plugins can oversample, thereby giving you a cleaner sound. The converters today are more better designed. That being said, I would still think that for designing and pitching etc, if the source is recorded in high sample rate, it finally gives a clean sound. Personally i have felt one thing, and that may be my imagination as I couldnt test it scientifically. I dont need to push levels in 96 kHz to get detailing. The sheen and the mix can be much cleaner and more present. But, the final delivery still remains at 48 kHz for films.

So what did we do?

We did the whole work at 96 kHz, using the entire benefit of this workflow and then did a Sample rate conversion at the very end thereby removing any frequencies above 24 kHz. This means we get a very clean signal without any artifacts of processing which even when played at lower levels would still sound very clean as the masking effects would not be present at all.

The thought of the Phantom Center

I have mentioned before that I am not a big fan of phantom center. I will get into the reason in a moment but lets first understand what phantom center is.

Phantom center is essentially the illusion of sound that seems to come from the acoustical center of two speakers when they reproduce the exact sound. This means that if we are mixing for stereo and we place a mono kick panned center, it would be reproduced at -3 dB from Left and Right. (-3dB is also something that varies from system to system. Pro Tools has the option to change that pan law from -2.5 to -6 dB). It all sounds good right? Yes but not in a surround environment.

Surround sound has a dedicated speaker for the center channel. So if you think that since an equal signal produces sound in the center, well, it does but in the acoustical center. That means it also depends on where you sit in the theater. So, if you begin placing your music elements only on the Left and Right, then anyone sitting a bit to the off of center would get a left or right heavy sound. The whole beauty of the balance will be thrown off. My thought and principle is that if we want to achieve a sound from the center, might as well use the speaker dedicated to it. This is a very important fact to understand especially when you have to realize that you are mixing with the L-C-R in view so might as well do full justice to it.

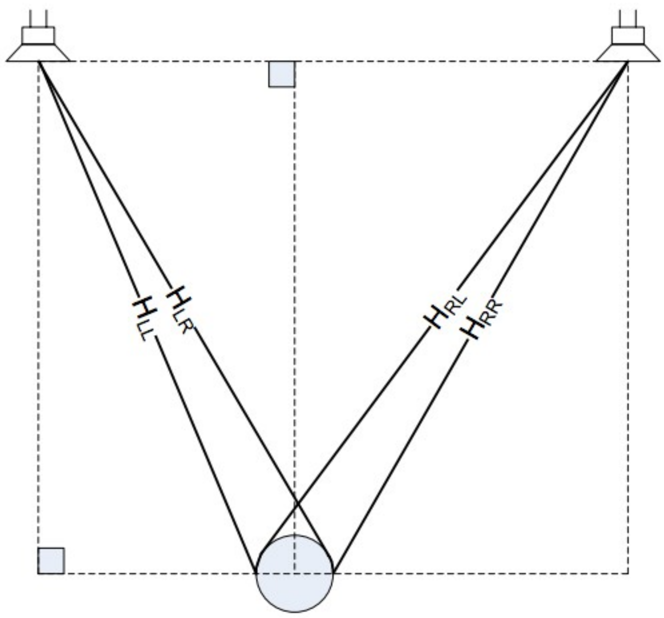
There is another reason why I dont prefer this. And this is the fact that there will be loss of frequencies depending on where we sit. But before that there are other issues we need to look at.

The Haas Effect

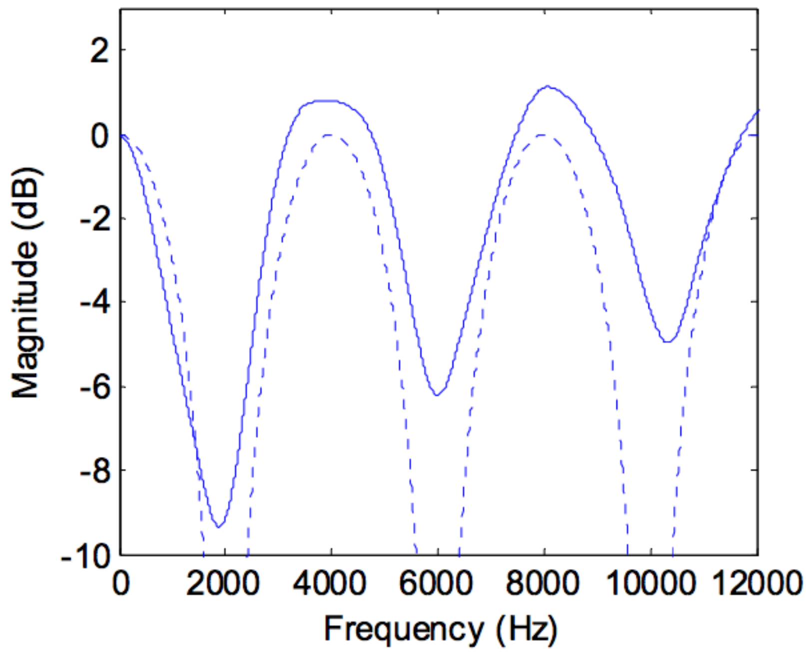
This is also called as the law of the first wavefront or the Precedence effect. The real statement is that *when a sound is followed by another sound separated by a sufficiently short time delay (below the listener's echo threshold), listeners perceive a single fused auditory image; its perceived spatial location is dominated by the location of the first-arriving sound (the first wave front). The lagging sound also affects the perceived location. However, its effect is suppressed by the first-arriving sound*. It essentially states that if there are identical sounds from two speakers, then it will appear to come from the closest source. This means that if you are sitting off center towards left in a theater with the kick coming from the Left and right, the kick will seem to be left heavy.

Comb Filtering

When the same signal from two speakers reach the ear at different times, it creates what is called an acoustic crosstalk. This is perceived as comb filtering. What is significant about this is that the filtering frequency falls in the speech range.



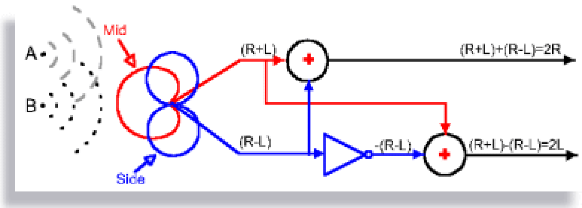
Cross talk created between the ears. Note that the difference of time between HLR and HLL as well as the difference in HRL and HRR across different frequencies will create a comb filter as below.



This means that the Phantom center will be less clear than a real center. This was always my main concern with using this especially when you have a center channel. One common method of mixing music is to have just the voice in the center. One justification I heard is so that it is to avoid clashes with the vocal. Personally I don't agree with that. Any mix will always have a way to carve the space for the vocal based on the frequency distribution. It should not rely on the speaker placement to do the job. This is also the reason why I had decoded the Stereo that I obtained from the MS track into 5.0. This gave me a clean center channel rather than trying to rely on the created center which would incidentally also make the resulting sound too wide. Because if you look at the distribution, all the main information or speech would be present in the Phantom center. And that is what will face this filtering while the noise or unwanted side signals would reach the listener directly thereby resulting in a low signal to noise ratio.

Converting MS to 5.0

This was a debated technique for sometime when I posted the dialogue mix notes. I wanted to clarify the process and thought behind this a bit more and hopefully it would be clearer. MS or Mid -Side recording essentially is a technique where the signal is captured with two microphones, one of which is a cardioid for mid and the other is a figure of 8 for sides.



So, if you see the above figure, the mid microphone wont have much difference when the source is moved from A to B. The Side Microphone will however have one side on phase and the other off phase while rejecting anything in between the lobes. So, this signal is captured as M on Left channel and S on the right Channel. If you look at the logic above, you will see that to decode this into a stereo track, it would be

Left = M+S

Right = M+ (-S)

where -S stands for the S signal phase inverted. Lets call this decoded stereo signal as StereoDecode. Now, if you were to hear this decoded signal, you will get the actual spacialisation of the location with source on A on the Right, when it moves to B on the left and that will be via the center. This center is essentially a Phantom center because it is reproduced as a common signal from two speakers. Now, if you read the above note on my thoughts about Phantom speaker, you will see why I decoded this StereoDecode into 5.0. This is so that I get the signal in the center channel actually derived from the center in the stereo signal. The logic used here is the Dolby Pro Logic II in a plugin from Neyrinck. How this works is a sophisticated decoding where signal unique on Left will be in Left Channel, Signal equal on both Left and Right on the Center Channel, Signal unique on the right will be on the Right Channel. Then it has its own algorithm to generate the surrounds which is essentially phase inverted signal on both the channels. It intelligently creates a stereo surround. The beauty of this method is that the decoded signal is fully mono compatible.

This was my thought behind using the decoded MS for achieving a surround ambience of the location.