Master’s Thesis

Mastering Object-Based Music with an Emphasis on Philosophy and Proper Techniques for Streaming Platforms

Michael Matsakis

Submitted in partial fulfillment of the requirements for the

Master of Music Technology

In the Department of Music and Performing Arts Professions

The Steinhardt School

New York University

Advisor: Paul Geluso

Reader: Dr. Schuyler Quackenbush

Spring 2021, April 20th

Abstract

The immersive medium has been slowly growing in popularity as the technology for consuming immersive media becomes more affordable and mainstream. Streaming services have become the most popular way for people to listen to music. Companies such as Tidal and Amazon music are already starting to showcase object-based immersive music content via Dolby Atmos or Sony 360. New and exciting software tools for creating and mixing object-based audio are emerging for artists and engineers alike. However, modern engineers are at a loss for any technology made for *mastering* audio objects–the traditional final step in the music creation process. Mastering object-based audio is possible with the current technology; but, it requires a deep understanding of the software's technical limitations, as well as a concrete philosophy for thinking about immersive audio itself.

Four professional mastering engineers who were currently working with immersive audio content were interviewed. The interviews were centered around each engineer's techniques and philosophies for working within the medium. They also noted the current aspects of mastering object-based music for streaming services, as well as their biggest issues with the softwares. Three different workflows were derived from the interviews, and a fourth was created by the testing engineer from a synthesis of the others.

In order to test the techniques proposed by the interviewees, each was practiced and utilized to master an immersive piece of music. The testing engineer took detailed notes on the strengths and weaknesses of each technique. Once the four masters were completed, they were shown to an expert panel as a means of further unbiased review. It was found that the each of the techniques were viable ways to master audio-objects. Although, each had their own unique dynamic quality and timbre, as is apparent in any artistic discipline. What was compiled is an evaluation of the cutting-edge object-based mastering techniques, as well as a philosophy for mastering audio-objects in any scenario.

Acknowledgements

I would like to thank my mother and father Meg Hanshaw & Manny Matsakis for continuously supporting me through my academic music journey as well as my big sister Meredith Matsakis for always checking in me and making sure I'm doing okay. Also, my best friends Owen Stone, Jake Lenderman, and Ethan Baechtold for being the *greatest* musical inspirations I could have every asked for in this life.

Thank you Paul Geluso as well for being incredibly patient and insightful during the entire thesis process.

Lastly, thank you to all my new friends and colleagues I've made while in grad school *and* during my undergrad at Duquesne. They have continuously demonstrated feats to me that I never thought were possible.

Michael Matsakis
Thesis Proposal and Method, Fall 2020
Advisor: Paul Geluso

Table of Contents

I. Introduction

The modern recording process is generally broken up into three respective sections, tracking, mixing, and mastering (Katz 2015). During the first two sections, artists focus more on crafting each individual song as a whole and less on the flow of the entire album. Mastering is the final step in every professional recording process that entails taking all the mixes from one project and creating something cohesive out of them in terms of frequency balance and loudness. This process is also subtly influenced by certain technical criteria set by whichever platform or medium it is intended to be distributed on. When mastering for broadcast/streaming services such as Spotify or Tidal, a professional mastering engineer should be aware of specific values regarding the integrated loudness (LUFS) of the master (Byers et al. 2015).

There are countless existing mastering methodologies in place for the *current* medium of audio consumption—channel-based audio. "Channel-based" implies that the master was created with the intention of being played back on discrete channels to specific speakers. In this situation each playback system varying in channel count would require a different mix and master. The growing popularity in 3D audio points to listeners potentially moving towards this new *object-based* medium for future listening. Object-based audio is described as a "format agnostic" medium of playback. Audio "objects" are spatialized by a renderer that will allegedly translate to whichever system they are played back on, from giant cinema systems, to stereo headphones, to a mono speaker. Although, this freedom from channel dependence creates room for many unexplored issues and ultimately alters the goal of the mastering engineer.

Mastering audio objects is vastly unexplored compared channel-based mastering. This study will focus on identifying and creating techniques for mastering audio objects for streaming platforms while exploring and noting the effects of experiencing the same object-based master across several playback systems. The result is not only a document for future engineers, noting the best current practices and options for working within this new object-based medium. But, also an analysis of the current technological state of mastering audio objects, and what tools may need to be created to help the next generation of object-based mastering engineers.

II. Literature Review

1. Loudness, Spatial Immersion, Dynamics, Audio Quality, & Format Preference

Measuring the subjective quality of object-based masters requires the use of specific parameters unique to the medium's traits. To distinguish the unique aspects of object-based playback, a study by Oramus and Neubar was examined. In their study, object-based masters made in Dolby Atmos were played through a 5.1 setup as well as a 5.1 channel-based master of the same material. They defined five measurable aspects of the sound; Spatial Immersion, Localization, Dynamics, Audio Quality, & Format Preference.

In the channel-based medium, it is rarely the mastering engineers job to spatialize the sources, this is more within the realm of mixing. But this changes when dealing with mastering audio objects. In the current state of media, music is generally not being produced or recorded for a 3D environment. So mastering a song that was made in a channel-based medium may involve the mastering engineer having to re-spatialize the mix as objects. This step falls under the new set of skills an object-based mastering engineer would need to practice.

Oramus' study is observed as a means of judging subjective quality for object-based mastering in this paper. Since loudness is determined by a mathematical equation (ITU 2012), this element will be judged objectively through appropriate metering.

1.1 Integrated Loudness (LUFS) & True Peak Level (TP)

A simple definition for Integrated Loudness is, "a measurement of the total amount of audio energy between two points in time divided by the duration of the measurement" (Katz 2015). It is an average of energy over time rather than a single value; in comparison to the True Peak measurement defined in section 2.1.2. The algorithm used to equate loudness was developed by the International Telecommunications Union and consists of four parts; Frequency weighting, mean square calculation for each channel, and channel weighted summation (this compensates for the fact that "surround channels have larger weights, and the LFE channel is excluded (ITU 2012).") The fourth component of the algorithm involves gating the audio into

400ms, overlapping buffers. These buffers create the period of time one can choose to average the signal by.

Broadcasting and streaming services normalize their streams "Full Scale" (Byers et al. 2015). This method of measuring loudness produces Loudness Units relative to Full Scale (LUFS/LKFS). Full Scale being in reference to the length of the broadcasted material. In this study, this refers to a single song and its entire dynamic range. It is important to understand that LUFS are different than a True Peak measurement, which is only a representation of how many bits are being used to represent the dynamic range of the audio material before clipping at its digital ceiling. True Peak measurement and Integrated Loudness have little to do with one another except for the idea of a Peak to Loudness Ratio (PLR). A higher PLR usually signifies a less fatiguing and more dynamically rich master (Byers et al. 2015).

1.2 Normalization

Normalization is a method of limiting and boosting the level of audio as a means of making a more consistent loudness experience for the listener. It is heavily used in broadcasting situations due to a long-term historical event known today as the "Loudness Wars." Dating as far back as Phil Spectre's Wall Of Sound approach to producing. Over the years, loudness was being pushed within music and advertisements so far that audio quality was beginning to suffer in an attention seeking contest (Vickers 2010). Certain content was perceived extremely louder than others due to higher amounts of dynamic compression and distortion. In the realm of music broadcasting, louder masters stood out on radio channels, theoretically resulting in more listeners stopping their tuners on "louder" stations. The goal of implementing normalization was to end the "loudness wars" by creating a Target Integrated Loudness Threshold (TILT) for broadcasts. By implementing a TILT, broadcasters could define a maximum loudness threshold and limit everything to make the broadcast adhere to that value (Byers at al. 2015). Within music streaming services, music is generally normalized to somewhere in-between -22 and -10 LUFS for channel-based content. This range compensates for the dynamic range necessary for varying genres of music. For example, a symphony with three movements will generally feature a much greater dynamic range than a pop single.

1.3 Spatial Immersion, Dynamics, Audio Quality, & Format Preference

Spatial Immersion and localization are judgements pertaining to the sonic clarity and depth of the material. Spatial Immersion relates to "the feeling of being immersed by the sound" (Oramus 2019). More specifically, does the master feel balanced in all appropriate directions? Immersion is not completely separate from the idea of realism, barring in mind that certain genres promote a variety of surrealism. Does the listener feel like the timbres and placements of objects within the material are believable? Localization relates to how easily the listener can tell the position of each sound source in the material. Immersion and localization are usually in the hands of the mixing engineer, but a bad master can ruin a great mix (Katz 2015).

Dynamics relates to the preservation of rhythm as well as an acceptable dynamic range for the genre. Does the applied compression throw off the rhythm of the track? Is one method more fatiguing to the ear than the other? Dynamic range is one of the most powerful and potentially transparent aspects of music that the mastering engineer has control over. There is more in-depth material on the tools available for modifying dynamic range in section 2.1.

Audio Quality and Format Preference are viewed as overarching judgements of the sound as a whole (Oramus 2019). The Quality is referring to any phase, sampling, or bit rate artifacts audible to a trained ear. Format preference is a general question regarding which listening medium the subject prefers overall.

2. Multichannel-Based Mastering

When mastering for discrete surround channels, the engineer should only be concerned with the master sounding great on systems able to play back the material with the correct setup (Katz 2015). Although, there are some exceptions; for example, a lot of media today is consumed via mono bluetooth speaker; Sonos, Beats Pill, & other portable speakers (Byers et al. 2015). This makes some mastering engineers want to check for mono compatibility. Interestingly, many new versions of bluetooth speakers designed for home use are now being replaced with 3D speakers; Amazon Alexa, Google Home, & Apple HomePod. Even still, most people who listen to music today are listening on headphones though personal music players (Byers et al. 2015).

Channel-based mastering engineers are most likely working in stereo, 5.1, or 7.1. The main distinguisher present is that, in channel-based mastering, the engineer is only concerned with one format at a time.

2.1 Multichannel-Based Mastering Tools and Techniques

Since the focus of this study will be on creating masters for *streaming* platforms specifically. This section will look into digital signal processing practices for tools that affect the loudness of a master made for digital consumption—dynamics processing, limiting, & EQ. Rather than other elements of mastering such as sequencing tracks, authoring metadata, or specific practices for analog mediums (how to master for vinyl, for example).

A dynamics processor is a tool mixing and mastering engineers use to compress or expand dynamic range. There are two types of dynamics processors and two varieties of each type. Illustrated in figure 2.1, the upwards/downwards *compressor* and the upwards/downwards *expander* are the four processors used by engineers, with the most popular being the downwards compressor.

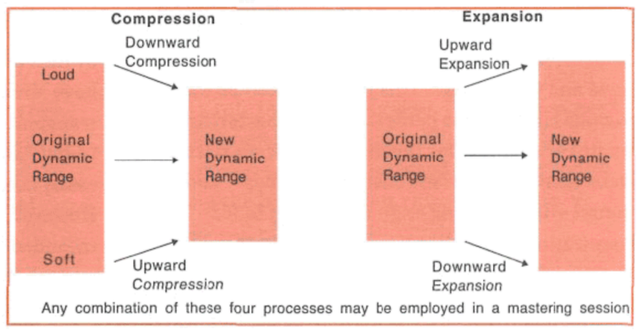

figure 2.1 (Katz 2015)
Visual examples of upward and downward compression and expansion

There is one other form of dynamics processing known as limiting, which is a form of downwards compression that uses an extremely high ratio.

When dealing with dynamic range, Katz splits up the functions of dynamics processing into two sections, micro and macro. Microdynamics refers to millisecond changes to the amplitude only possibly through dynamics processing or hyper-automation. Macrodynamics refers to the loudness of specific sections of the song—chorus being too loud compared to verse etc. Macrodynamics changes are usually automated via fader level because the overall level of the section might as well be brought up. Doing this with just an automated level change can be much more transparent and less intrusive than compression. So it should be viewed as a strong starting point if the material seems to be too dynamic section-to-section.

Dynamics processors made for mastering generally have six basic controls on them: threshold, ratio, attack, release, knee, & makeup gain. During mastering, one can find use in extremely low thresholds, much lower than the average peak height, paired with low ratios of 1:1.1 - 1:1.8 (Katz 2015). This style of compression is extremely transparent and can bring a little more balance to a mix that is too jumpy/punchy. Katz refers to this style of compression as *invisible compression.* Attack and release are used in junction with threshold and ratio to govern how exactly the transients will be shaped. The quicker the attack, the duller the transients will become, with a slower attack (starting at around 30ms) allows transients to pass through resulting in a punchier sound. Knee is a subtle but powerful control that works with attack, release and threshold to shape the onset and offset of the compression. This onset and offset is technically known as a transfer curve. A sharp knee is more akin to a limiter, where *as soon as* the amplitude crosses the threshold the attack or release is triggered. A soft knee curves this transfer and 'spreads' the threshold out for a lighter attack/release.

Since compression alters the shape of transients, it is also considered a form of distortion. This makes the use of too much compression detrimental to the clarity and timbre of the master. An R&D study on the design of dynamic limiters done by the BBC found that humans can not hear distortions that are shorter than 6-10 milliseconds (Mayo 1940). Based on this information, Katz gives a recommendation, "short duration (A few milliseconds) transients of unprocessed digital sources can be reduced by 4-6db with little effect on the sound; however, this cannot be done with analog tape sources, which have already lost the short duration transients (Katz 2015)." If the entire mix needs to be raised in apparent loudness without severely effecting the sound, consider limiting instead of compression. In mastering, compression is generally a tool for smoothing out or creating stronger transients and effects the rhythmic perception of a mix.

The multi-band compressor/expander is described by Katz as "the most powerful and potentially deadly audio process that's ever been invented." Invented by TC Electronics within their M5000 unit, these processors are able to split the audio up into two to six frequency bands and compress each band individually. It is recommended to only use them in severe cases when the material sounds very unbalanced (Katz 2015). Due to its ability to modify specific frequency spectrum, the multi-band compressor has the power to drastically change elements regarding the balance of the mix. Because of its split band nature, the multi-band compressor can take on the form of an equalizer.

When faced with an unbalanced mix, how should one decide to choose between multi band compression and just plain equalization? Katz suggests, after honing in the macrodynamics of a song, one should use equalization to adjust *consistent* frequency imbalances. The equalizer is static in nature—unless it is automated. If certain frequencies jump out only during certain sections one can either decide to automate a band in their eq, or if the "jumps" are not constant, use a multi-band compressor or a dynamic equalizer to soften the unwanted frequency. One other common situation is balancing out low end energy. Since mix engineers tend to mix on limited range systems, sometimes without a subwoofer, the low end energy can be too dynamic between the kick and bass. A multi-band compressor band that is focused on the low end can catch the peaks of the kicks and attenuate them before bringing up the overall volume of the low end.

In mastering, the equalizer should be viewed in an extremely technical and surgical light, as well as a tool that effects the emotional content of a song. The mastering engineer's goal is to be as transparent as possible with their actions. Most DSP equalizers are not linear phase due to the amount of CPU power required to accomplish a real time linear phase filter. When filters are not linear phase, frequencies in the band affected by the equalizer will inherit a slight latency causing minor distortions to the now processed audio. Some refer to these distortions as the desirable "character" of the EQ. But, it is most likely that the mastering engineer will need their EQ adjustments to be completely transparent. Linear phase equalizers delay all other frequencies the same amount as the affected band so that they all exit the processor completely "in phase" with each other. This creates a much more transparent and uncharacteristic EQ.

When utilizing an equalizer in a mastering context, it is affecting all frequencies across all elements within the material. For example, if the guitar is too bright in a particular range in the mix, turning that frequency range down will turn down all frequencies in the entire master, as well as boost non-attenuated frequencies due to compression and limiting happening post EQ (Katz 2015). Issues involving frequency buildups in specific instruments should be taken care of through communication with the mix engineer, if absolutely necessary and always as a *last resort*. When using an equalizer, the mastering engineer is concerned first-and-foremost with the tonal balance of the material. Katz compares the idea of tonal balance to the balance of a symphony orchestra. A spectrogram of the average amplitudes of specific frequencies in an orchestra will show a gradual high end roll off. This same roll off is present in good pop music masters (Katz 2015).

The other tools a mastering engineer typically uses in channel-based masters are for more specialized situations, due to their destructive natures. For example, an engineer can use spectral imagers to spread out or narrow frequency ranges in a mix that come across as too narrow or wide; but, this could ruin a master’s stereo image if pushed too far. Subtle amounts of reverb can be used to give dry mixes more perceivable depth. The addition of reverb will also raise the LUFS meter reading due to the elongation of transients. This raises the average wave height, therefore raising overall loudness. Mastering engineers generally have an eye on their meters that tell them the loudness relative to full scale, true peak, and frequency spectrograms of the material they are mastering. Although, the most trustworthy piece of equipment is a trained ear, these meters should be used to assist the ear with knowledge of the strengths and weakness of the room acoustics in which the material is being mastered.

3. Object-Based Mastering

Object-Based mixes do not rely on panning to create discrete signals in speakers but instead, use metadata encoding to place a “sound object” in a virtual 3D scene that is sizable to any array of speakers hooked up to a proper decoder (Geluso, Roginska 2018). The audio in the object is processed as samples per usual, while the positioning of the sound objects, as well as their motion within the scene, is processed as frames per second. When working with Dolby Atmos files, matching sample rates with appropriate frame rates is a crucial part of mastering. If either of these aspects are accidentally changed during the mastering process, it could result in distortion of the final product (Dolby Laboratories 2019). The mastering engineer must be aware of the format of the material and what frame rates are respectively supported. Since object-based masters are independent of a discrete number of output channels, an engineer must develop techniques for compensating for artifacts caused by different playback systems. Due to this format agnostic medium, a different mastering approach must be adopted by future engineers.

3.1 Object-Based Media

When object-based audio content is streamed, the sound objects are stored in independent channels and positioned via metadata on each channel, leaving the "sonic result depend[ing] on the efforts of the renderer" (Rumsey 2018). A study on listening preference between object-based playback systems that found that it is rarely possible to create an identical listening experience on two different systems (Woodcock et al. 2018). This makes the goal of a mastering engineer different when working with object-based mixes, due to their need to compensate for loss-of-envelopment alongside user preference when the master is played back on lower channel systems. The benefit of the renderer being the bottleneck is that as rendering technology increases in quality, the content being rendered will increase in quality. A mentality for dealing with this is explained by the interviewed experts in section III, subheading 2.3.

3.2 Prospective Object-Based Mastering Tools

When dealing with mastering objects, due to their independence from output channels, every audio object needs to be adjusted individually. But adjusting each object individually was too reminiscent of channel-based mixing for researchers, HesterMann, Seideneck, & Sladeczek. They proposed the potential idea of introducing "Mastering Objects" as a tool to overcome this channel independence. The Mastering Object functions as a way to apply a certain processing to multiple audio objects at once. It can either be directly attached to certain objects (known as a "direct connection") so that the processing follows them if they move, or the mastering object can have a certain range (known as an "area of interest") to it that would affect an audio object when is enters that area. The area of interest contains a fade-in and fade-out area defined by the mastering engineer to create a smooth effect regardless of the playback system. Nothing is attached to specific channels still, only specific areas of playback.

They propose two other similar ways of utilizing Mastering Objects. One is called using the "Object-Interspace" which is the same concept of the Mastering Object having a radius in which it effects other objects, only now the radius does not fade in or out. The last method is called the "Angle of Interest" which takes all audio objects at a certain angle from the listener and applies the Mastering Object to them.

They created a prototype using *SpatialSound Wave* (SSW) technology developed by Fraunhofer IDMT. In their paper, the method of mastering must take place to all the audio objects in-between mixing and playback. So, the SSW acts as a mastering unit alongside the DAW. The SSW's task is to render the mixed audio objects as well as the new mastering objects and their respective processing. The communication between the two applications is done via Open Sound Control (OSC). It is unclear in the paper whether or not the mastering processing must be done in real time to the objects; but, it seemed like that was the case.

Currently, Dolby offers limited plug-ins relating to mastering in Atmos (Dolby Professional n.d.); mainly, several types of panners that can be utilized for both VR and Speaker arrangements. The panners connect to a renderer that is responsible for reading the metadata of the objects. The Atmos Conversion tool is a mastering tool made for changing the format of object-based masters as well as stitching multiple masters together. As of now, nothing has officially been released in terms of effects signal processing, or tools that assist in an object-based mastering workflow.

4. Listening & File Formats

When dealing with Object-Based Mastering, there is the issue of how to prepare the master for a certain format when no format is specified. How many speakers should the mastering engineer use for an accurate representation of the "best format?" Research currently states listeners actually prefer 7.2 and 5.1 over 22.2 listening formats. Due to this, mastering engineers need not be concerned with more than that amount of speakers unless their master is specifically going to be played in a bigger context, i.e movie theaters (Francombe et al. 2017). Currently, Dolby Atmos playback for music is supported on Tidal, Amazon Music, Netflix, iTunes, as well as in some other, lesser known softwares (Dolby Professional n.d.). The way these platforms are measuring their loudness is still not established in a global context. Recommendation ITU-R BS.1770 states that loudness in broadcasts should be measured by the LKFS scale, but does not go into detail on how to accomplish this for object-based media (ITU

2012). As of now, there are no concrete standards for measuring the loudness of object-based content (Norcross 2017). Although, when object-based audio is delivered, it comes with all objects and metadata as well as a 5.1 mix created as a side chain routing for backwards compatibility. It is recommended by Norcross that the broadcasters use this 5.1 render as the means for measuring the loudness of the object-based master. This is due to the change in loudness across multiple speaker formats being negligible (Norcross et al. 2017).

Several file formats have been created that can handle the task of object-based audio: MPEG-H, Dolby Atmos, & Auro 3D. This study aims to look deeply into the Atmos format. To be considered an object-based format, it must be able to handle three components unique to the medium: expandability to up to the maximum amount of channels the decoder can handle, recognize the presence of audio objects, and contain the appropriate metadata to position the objects (Poers 2015). Dolby Atmos has several formats within itself as a means of preparing for specific kinds of Atmos playback systems. Each format contains a set of files which hold both the audio and the metadata positioning required for 3D playback. Each format is also only compatible with a specific list of frame rates, shown in Table 1. An ".atmos" file set is acceptable for home theatre as well as listening over headphones (Dolby Laboratories 2019). There is an older version of this format known as ".damf" compatible with legacy systems. If the material is being prepared for The Dolby Renderer for Cinema, the mastering engineer must use either the ".mxf" or ".rpl" file formats. The main difference between these two theater formats is that ".mxf" is more suitable for broadcasting purposes (Dolby Laboratories 2019). The ".dbmd" file format is an option for exporting only the metadata of the material. This can be combined with any channel-based audio and exported as any of the complete file sets listed previously.

Dolby Atmos master file format	**Supported target frame rates**
.atmos (or *.damf*)	23.976, 24, 25, 29.97, 29.97 drop frame (DF), or 30 fps
DCP *.mxf*	24, 25, or 30 fps
.mxf (IMF IAB)	23.976, 24, 25, or 30 fps
.rpl	23.976, 24, 25, 29.97, 29.97 drop frame (DF), or 30 fps
.wav (ADM BWF)	(as specified in the ADM BWF metadata): *.dbmd* chunk: 23.976, 24, 25, 29.97, 29.97 DF, or 30 fps
.xml (pmstitch)	23.976, 24, 25, 29.97, 29.97 DF, or 30 fps

Table 1 (Dolby Laboratories 2019)

5. Conclusion

As music platforms and creators begin to utilize object-based mixing as a norm, mastering engineers must figure out strong and effective methods for dealing with these new issues. Spatial immersion, dynamics, audio quality, and format preference are defined as important measurable aspects of an object-based master. For the sake of fidelity, these aspects must remain at a high level of importance to the mastering engineer, who's job is to prepare the media for the intended mediums of playback. It is also important that, as engineers move into a more object-based approach to streaming, loudness normalization doesn't get left behind or overlooked. This could–in the worst case scenario–result in a second 'object-based loudness war' until proper methods of limiting and measuring object-based audio are created and adhered to. Through the trial and analyzation of the latest techniques for mastering object-based material, this study aims to distinguish the strengths and weakness of the system as a whole. This gives object-based mastering engineers a philosophy and insight to proper techniques, so that they can continue to create high quality, dynamically rich music throughout the next era of object-based audio engineering.

III. Phase One: Information Collection

This study was a mixed approach of gathering the best practices regarding object-based mastering from interviews with engineers currently working in the field, followed by an applied practice of each learned technique, and closed with a final subjective review of each technique by the testing engineer and an expert panel. The overarching goal was to determine some of the best practices and clear up any discrepancies regarding current object-based mastering techniques. The techniques and practices attempted were derived from interviews conducted with expert mastering engineers who work with object-based musical content on the professional level. The end goal for the testing engineer was to create four different masters of the same song, each utilizing a different technique derived from the interviews. The song chosen was, "Mitad Del Mundo" by the artist, *Helado Negro* and was mastered by the engineer in a 9.1.4 Dolby Atmos environment. The object-based masters were then reviewed by an expert panel.

As a means of reflection and comparison, qualitative measurements were noted by the engineer during the mastering process. These qualitative measurements were defined in-depth in the literature review and consist of spatial immersion, dynamics, frequency roll off, and overall audio quality. This study is less concerned with format preference, and more with how the varying techniques affected the song as a whole. Because of this, the expert panel each chose their own preferred format and listened to all versions of the master in the same medium. A quantitative measurement of loudness using a LUFS scale was done to both masters as well. All LUFS measurements were done with the Dolby Atmos loudness meters in the renderer. Loudness measurements were taken from the full Atmos render as well as the binaural fold down of each master.

It should also be stated that, due to the current primitive state of object-based mastering, the engineer noted any technological issues that they faced while working. These short-comings were documented as a means of pushing the medium of object-based mastering to its fullest potential.

1. Interviewing process

The four engineers chosen to be interviewed were Ronald Prent of Valhalla Studios, Michael Romanowski of Coast Mastering, Reuben Cohen of Lurssen Mastering, and Ceri Thomas who works with Dolby on developing music production tools for Atmos. Prent, Romanowski, and Cohen are all mastering engineers that are currently working on mastering immersive projects. Romanowski and Cohen's experiences are mostly in music and soundtracks while Prent masters music as well as movies with dialogue. On the other hand, Thomas gave a deeper insight into how the Atmos renderer functions as well as important strengths and weaknesses to consider while working with it.

The interviews consisted of a supplementary document that the engineer created with basic talking points, making sure enough information was gathered to take what each interviewee said into practice. This included questions centered around their speaker setups, object-based mastering workflow, signal flow, how they preferred mixes to be delivered, as well as what they believed to be the current weakness of object-based mastering and what they would like to see in the future. After the interviews were completed, each engineers' thoughts and processes were compiled into the most significant subjects. Their specific mastering techniques were also noted and translated into processing block diagrams, giving the testing engineer several unique approaches to object-based mastering.

2. Interview Results

The next sections detail the most significant findings on mastering object-based content from the interviews, as well as the specific mastering signal flows from each interviewed engineer.

2.1 Understanding the Immersive Medium and System Design

The experts interviewed each individually made it a point to note that modern engineers who are used to working in channel-based environments must *first* relearn how they think about the positioning of sound sources in the space of a mix. The first mental block to overcome is somewhat obvious and is the fact that there are no discrete speaker channels anymore. Instead, when an engineer places objects within a 3D space, the renderer is cross-referencing that positional metadata with the speaker positions it knows about. It will then either put the sound where a speaker is or do it's best to create a phantom image where there is no speaker. In this scenario, the more speakers one has (to an extent), the better the spatial fidelity is experienced. So, it was explained that it is better to envision the system as if it were a wavefront. This ideology comes into practice constantly when working in the immersive medium. One example explains the issues regarding engineers creating object-based media on non-immersive systems with less speakers, such as headphones or a stereo speaker setup. It was explained by Thomas that these systems would not be ideal for a mixing or even a production environment because of the amount of phantom spatializing the renderer is having to do to the objects placed outside the present stereo field. Stereo may be adequate for a listening to immersive content, but from a production, mixing, mastering standpoint, it leaves too many elements uncertain to the engineer. It is because of this reality that every engineer interviewed recommended at least a 5.1.4 speaker setup for a production/mixing system. For mastering, it was recommended by all engineers to use at least 9.1.4 system depending on the size of the room the speakers are in. The added wide left and right speakers help fill in the sound-field in front of the engineer, where a human's spatial localization is much more accurate. The height channels seemed to be necessary for any professional setup. And, depending on the length of the room the engineer is working in, six or more may be required.

It is also important that object-based mastering engineers understand certain important features available in the Atmos renderer. First and foremost being Spatial Coding Emulation. This is a feature that can be toggled in the renderer preferences and, when switched on, dynamically re-renders all objects down into 16 static objects, equally placed around the listener. Turning this on will give the engineer a better example of how the content might be experienced on the streaming listeners end. The idea behind it is to break down all the objects in the entire 3D environment into 16 static "objects" that each represent a portion of the sound field. These 16 "objects" dynamically capture any audio that is present in them across the mix and plays it back through itself, reducing the number of objects the renderer has to deal with as well as the file size of the master. A mastering engineer should turn this off when working *but* be sure to check it after technical moves have been made. Spatial Coding Emulation is mostly unnoticeable but has been known to create unwanted masking as well as image distortion (Thomas).

Another notedly important renderer feature spoken about by Thomas was Atmos' object binaural render modes. In the Atmos renderer, one can set the perceived distance for an object in the binaural render modes screen. Setting each object's mode to Near, Mid, or Far applies a slightly different HRTF to the object, giving it more or less of a virtual room tone. This is useful when mixing, but when mastering (in certain situations, as will be explained), one should make sure this is off for all objects as to not re-convolve them with room tone over and over again.

Romanowski and Cohen spoke about the practice of cross referencing the material being worked in on several fold downs to test for compatibility. They both seemed to share a mentality along the lines of, their job is to make sure the material sounds great on the best system possible, and any system that is less than theirs will inevitably not sound as good. And, it's not productive for the mastering engineer to try and achieve something based on this problem. It was stressed by Cohen that channel-based fold downs aren't perfect yet, and trying to master content for them is counterproductive. Romanowski, Thomas, and Cohen each individually stated that the technology for personalized HRTFs *at the current moment* isn't great (and mostly doesn't exist). Because of this, binaural down mixes of object-based content are considered to be inaccurate (or not accurate enough for dense, spatialized music mixes) and may give a master a completely different sound to two different people. The way we hear is *much* to subjective to use the same pair of HRTFs for every listener, it simply does not work. Checking fold downs of the master is useful only in a sense that the engineer should make sure their master isn't getting "destroyed" by the re-rendering algorithm. The engineer should be mostly finished with the larger processing choices they've made before starting to listen to the fold downs. After a reference, one should make purely technical moves to remedy any dissatisfaction to a point of compromise, weighing heavier on preserving the quality of the full range master. The long term benefit of this mentality may be that, as fold downs increase in quality, the experienced object-based master itself will increase in quality over time. That is, if it already sounds great on a great sounding system, the technology will eventually catch up.

2.2 Object-Based Mix Delivery

There are currently two viable ways for an object-based mix to be delivered to a mastering engineer. The benefits of each will be explored in depth during the “In Practice” phase. The first way a mix could be delivered is as all the objects and metadata in the format of an .adm file. This gives the engineer access to all the individual objects as stems with any effects processing printed onto them. The second method is to deliver an export of the speaker channels as objects. Romanowski spoke about the benefits of this method saying that it helps him work on creating cohesion across speakers when processing is applied, so that timbres and levels are consistent across all speakers. Although, it was discovered by the testing engineer that one must be very careful when doing this as to not have a drop in overall quality when the objects are re-spatialized. Dolby Atmos may not be currently set up to work like this.

2.3 Immersive Mastering Goals/Mentalities

Romanowski stated in his interview that, “channel-based mastering is a different mindset and a different set of goals than channel-based mixing, Immersive mastering is even more so.” Because the added step of rendering is involved, this creates the opportunity for problems to arise. These are generally problems centered around phasing, level, and overall presentation. Romanowski and Cohen both noted that one of the greatest setbacks for object-based mastering at the moment is that nobody knows exactly *how* object-based music content will be experienced by the listener. *Plus*, the possibilities for listening are now much more vast than when dealing with channel-based stereo masters. The format has essentially dissolved and become fluid, meaning that a master that may have been created for a 9.1.4 setup could be experienced in *any* kind of playback system such as a car, headphones, sound bar, 3D speaker(s), or in a home entertainment system. The point being that one shouldn’t go crazy trying to get their master to sound perfect on all of these playback devices. As each device gets better at decoding immersive content, the fidelity of the content created will follow.

“Comparison is the seed of discontent” - George Massenburg

Each engineer had a somewhat different opinion on the topic of the adequate loudness for an object-based master. On one end, Romanowski spoke about how level is not so important to flex and mentioned the George Massenburg quote above. He felt good with finished masters metering around -18 LUFS. To him, the goal lies much less in the "making it loud" aspect of mastering and more in creating an immersive environment for the listener. Making something overtly loud can hurt the perceived reality of a project and, therefore make the spatial immersion suffer. Cohen's mentalities were somewhat parallel to this with the added idea that loudness is *also* somewhat dependent on what the song is asking for. In other words, heavier EDM or rock material may need to be pushed as far as -6 or -5 LUFS whereas more dynamic material such as orchestral recordings can lay closer to -20 LUFS. For Cohen, it was less about hitting a certain number more so than it is about doing what is necessary for the genre to compete in the current musical climate. Another noted benefit of *not* mastering to a loudness normalization value was that if, in the future services decide to change this value, your masters won't suffer. On the other side, Prent (who seemed to work mostly with film and TV) was focused on the practice of consistency. His goal when making object-based content was to consistently create deliverables that meter between -14 and -10 LUFS.

2.4 Immersive Mastering Techniques

Because the tools for mastering object-based audio are limited and constrained by the format agnostic medium it is currently existing beside in DAWs, mastering remains to be a subtle art. This is to say that larger processing moves such as multi-band compression become much less useful (if not useless) in the 3D environment. Thomas stated that the biggest benefit to mastering object-based content is that the mix will be heard and approved by another set of trained ears, in a great room on a great system. Even so, processing is usually still desired to give mixes that extra bit of polish. Each mastering engineer presented their own unique way of applying processing to an object-based master.

When Prent masters, he is mostly working with all objects as stems delivered as an .adm file. He also mentioned that, most of the time he is delivered a 7.1.2 channel-based bed as well as the objects. He has built an analog system for mastering object-based content that includes 24 mono limiters that can be linked (but mainly used independently), as well as 12 analog API compressors. He said he listens to the mix and applies EQ in the DAW if necessary, then runs the tracks through the decidedly necessary compressors and limiters. He then records the outputs into another completely different session that captures the processing as printed tracks. Those tracks are then routed into the renderer and spatialized. His main goal while performing these tasks is making sure the content is consistent in metering around -14 and -10 LUFS. The signal flow chart detailing this method is presented in figure 3.1.

Ronald Prent's Mastering Signal Flow

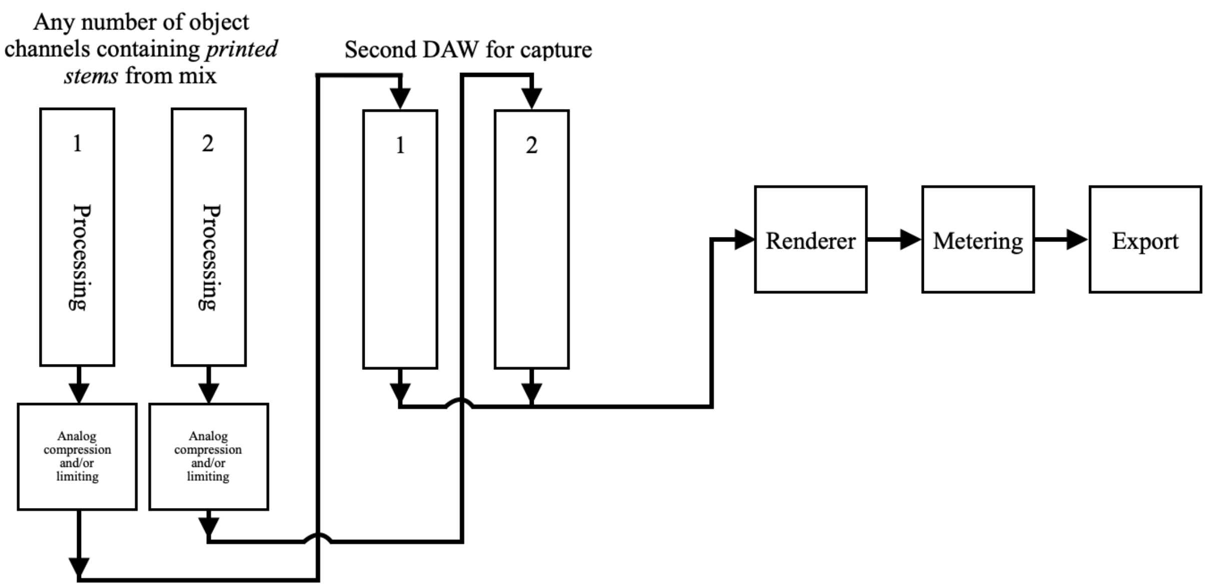

fig. 3.1

Romanowski presented a different approach for applying processing. His method starts with all the objects, exported as speaker channels and re-spatialized as "speaker objects." This way, the processing is applied to entire zones of the master instead of each object individually. He made clear that his main goal when mastering was to create a *realistic and immersive* space. Because of this, he believed that his listening environment was the greatest tool he had. When he applies processing, he has the goal of keeping a cohesive timbre across the entire immersive field in mind. If certain channels are generally louder or brighter, then moves would be made to adjust accordingly. To him, realism and immersion are the most important qualities to get right. It is also important to note, that when utilizing this method, since the "speaker objects" are created *post*-renderer, all the objects would have already had Binaural Render Modes applied to them. So, one should make sure to turn this *off* for all "speaker objects." This is to ensure that objects are not convolved with HRTFs twice. The signal flow chart detailing this method is labeled figure 3.2.

Michael Romanowski's Mastering Signal Flow

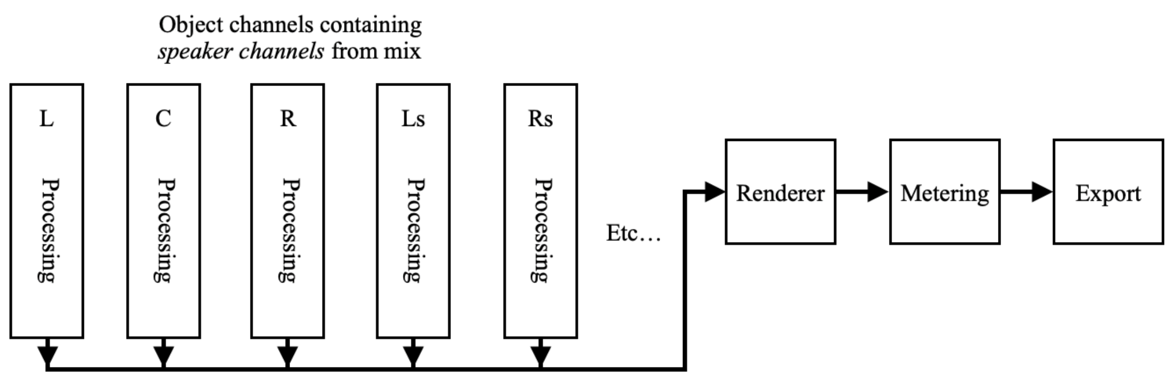

fig. 3.2

Cohen's method lied near Prent's in the practice of mastering all objects individually, but added another *special* step. Cohen wanted to figure out a way to mimic the master bus "glue" compression that so many engineers love to use in channel-based formats. He developed a method to accomplish this that first involves creating a mono summation from from all the objects. After the content is summed, an EQ will most likely need to be applied to compensate for the low end buildup. That summation's output is then sent to a "phantom bus" channel. That signal is used to trigger the thresholds of compressors placed on every object through their side-chain or key inputs. All the compressors are also set to have matching attack and release parameters. This technique applies compression while keeping phase correlation across objects and also creates an opportunity for the engineer to be slightly more aggressive with their compression. It is important to state that the delay compensation in DAWs is *not* set up for this complicated task. Meaning that it can *not* be done live. Because of this, one can not just send all objects to a mono bus and key compressors to that bus. The engineer will have to print out a mono summation of the mix so all the channels keep the correct phase relationship in the summation. The signal flow chart detailing this method is labeled figure 3.3.

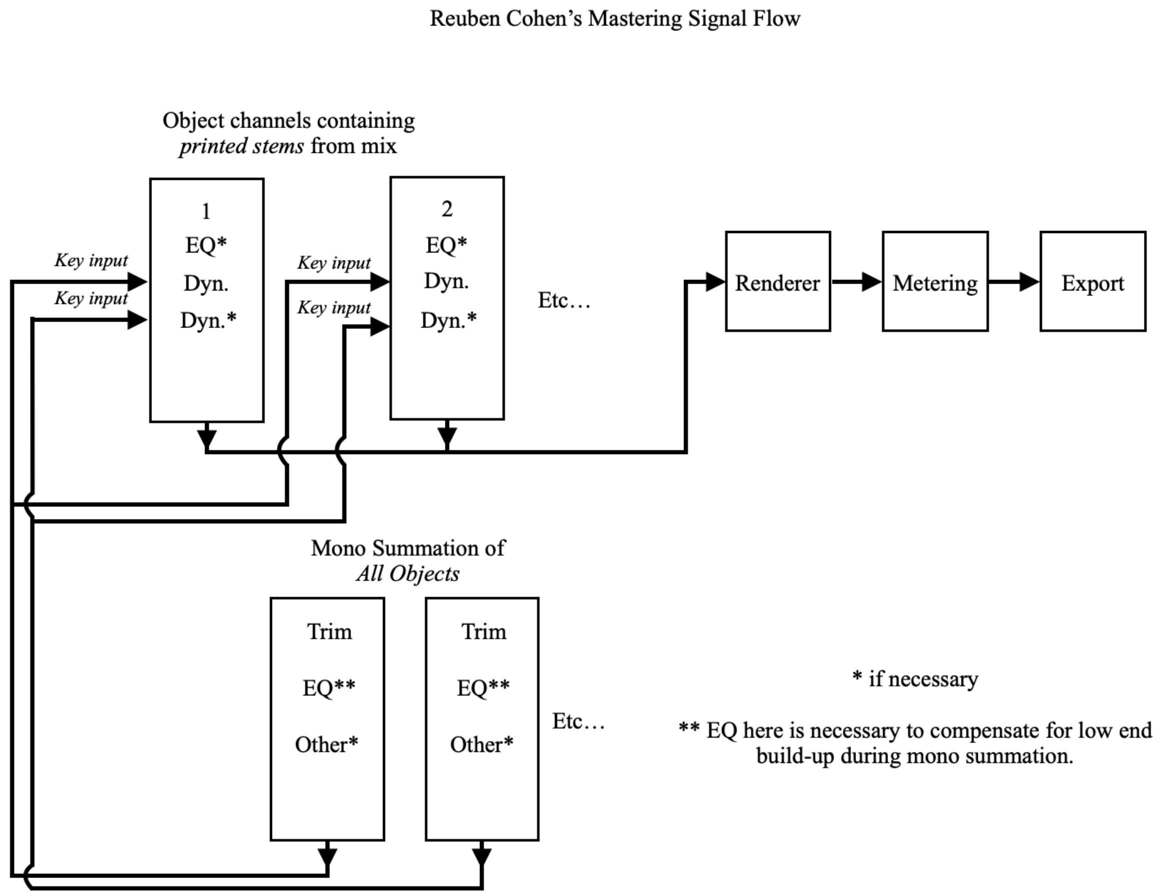

fig. 3.3

2.5 Preparing Object-Based Content for Streaming

Thomas gave incredible insight on the current state of how object-based content is being streamed and experienced. At the moment, only Tidal HiFi and Amazon music are streaming what they deem as "object-based content." In reality, this content is not object-based but a facsimile of one. This is due to the listening experience *still* being format dependent in both scenarios. The content is actually distributed to the services as an AC4 Immersive Stereo file, which is a fixed binaural print of an object-based master. This is how Tidal listeners experience the music over headphones or stereo speakers, only in stereo. On Amazon music, the material can also *only* be realized on an Amazon Echo Studio device which uses five speakers in a pod to emulate a 5.1.4 speaker setup. It uses microphones to judge the room and attempts to create phantom images of speakers in the space. This creates more of a novel experience due to it not being available on full range, high-fidelity speakers. Thomas also explained that streaming services are normalizing object-based content to -18 LUFS but allow a range of +-13 LUs above or below that value. The renderer normalizes the content for the listener to help combat any large jumps in volume while listening to a variety of music. It does this by analyzing the LUFS of the song currently being streamed and then uses that "dial norm" value to set the range of acceptable loudness for the next track.

2.6 Issues with the Immersive Medium

As stated previously, mastering object-based content is still in its developing stages so there are still issues that need to be solved before the practice can become completely solidified. All the interviewees seemed concerned with creating a standard as to how the immersive content they create might be streamed and experienced by the listener. Cohen's issues were centered around the binaural fold downs for object-based music not sounding that great when compared to a similar channel-based stereo master. Also DAWs (mainly ProTools) will need to be somewhat redesigned to accommodate object-based projects. At the moment, ProTools does not have a way to create a 9.1.4 bus on one track and forces the user into a specific workflow that is difficult for the engineer to modify once things are set. Also, the issue with delay compensation holding back engineers from creating a live phantom bus for workflows such as Cohen's.

Romanowski spoke about wanting more channel scalable tools such as EQ's and Compressors. In the current climate, it is possible to perform mastering to object-based projects, but the systems being used are still primarily channel-based systems. Because of the core design of modern DAWs is inherently geared towards creating channel-based content, the object-based side is cluttered with redundancy and non-intuitive workflow practices. Either DAWs will have to create a "3D mode" for them to run in that prioritized an immersive workflow or a new DAW will take over to become the standard for object-based content.

IV. Phase Two: In Practice

1. Mastering Setup and Goals

Mastering took place in the NYU research lab with a 9.1.2 Genelec setup. This setup was chosen through suggestions made by Michael Romanowski and Ceri Thomas in their interviews. All of the mastering took place with ProTools Ultimate Version 2020.5.0 and the Dolby Atmos Renderer Version 3.4.0. For metering, the LKFS meter in the Atmos renderer was used to measure both the full Atmos render as well as the binaural render.

The engineer chose a single track to master several times, each time using a different method derived from the interviews. After completing each technique, the engineer wrote reflections on what they believed to be its strengths and weaknesses. With this information, they created a synthesized technique that took insight from each technique to create a different workflow. The final goal for this project was for the engineer to judge the best current practices for mastering object-based audio for streaming services. From this point on, the mastering techniques will now be referred to as followed: Prent's Technique will be A, Romanowski's Technique will be B, Cohen's Technique will be C, and the Synthesized Technique will be D.

1.1 Immersive Mastering Goals

The ultimate goal for mastering the same song several times is to test the strengths and weakness of different methods for applying processing. Each engineer presented a slightly different technique for applying dynamics and frequency based processing. Prent's technique applied it to all objects individually, Romanowski's applied it to speaker channels as objects, and Cohen's also worked with all objects but applied compression differently with his phantom master bus compressor technique. "Mitad Del Mundo" is an electronic pop track and seemed to sound the most natural in its genre sitting around -12 to -14 LUFS. The engineer tried to be somewhat systematic in terms of how they were applying gain reduction. Somewhere around -2 to -3 db of gain reduction was applied during each technique. The biggest difference being *when* the processing was applied.

Another note on loudness: A mastering engineers working on tracks intended to be listened to on a streaming platform such as Spotify, Tidal, or Apple Music should note that these streaming platforms normalize to values around -12 -18 LUFS; *but,* they also allow a range between -5 and -31 LUFS depending on the genre of music (Thomas 2020). As Cohen suggested, loudness should be taken on a case by case basis. In the real musical world, some tracks may require a less dynamic master to do the genre justice without sacrificing musicality or sound quality. The normalization standards are only considered and tested as reference points to make sure that they will *still* sound decent at the normalized level.

2. Object-Based Mastering Process

The notes from the interviews were used to create three methods for mastering object-based audio, described in detail in section 3.2.4. The engineer did several passes at mastering "Mitad Del Mundo," each time trying a different method. They also created a fourth method (D) from a synthesis of methods B and C. Their notes for each master entail detailed block diagrams of signal flow and processing used, as well as a written reflection about the mastering process as a whole. The engineer also made sure to include their thoughts about each masters' quality by the subjective terms defined in the literature review: spatial immersion, dynamics, frequency roll-off, and audio quality. The loudness of these masters was captured with the meters within Atmos Renderer. After all the masters were perceivably finished, the engineer chose what they believed to be their final version of each master to show to the expert panel interviewed for feedback.

It is important to note that the song being mastered in this thesis was not knowingly mixed *or* produced for an object-based listening experience. Because of this, the engineer needed to perform minimal adjustments to the printed stems in order to help them better translate to the immersive medium. This preliminary spatializing aspect of the process should also be considered mastering, due to it being a process that prepared the audio for its intended medium of distribution. The new reality now being that modern object-based mastering engineers may have more roles to fill in the area of spatializing channel-based mixes. Details of this process are described in the following section.

2.1 Preparing Mitad Del Mundo for Immersive Mastering Notes

The first step in preparing this song for the immersive medium was to spatialize the mix as objects so that it would fill up the space and sound as big as it did in stereo, in the immersive environment. To do this, objects were created for each mono track in the mix, stereo tracks receiving two mono objects. The position of each sound within the 360º format was loosely derived from the stereo master. All the objects stay static throughout the song with the exception of the vibraphone track, which was automated to spin around the listener when it enters in the bridge at 1:03. Because this track was stereo, the left vibraphone object was initially placed in the front left speaker while the right object was placed 180º across from it in the back right speaker. From there, they slowly rotated counter clockwise around the listener throughout the musical line they played.

The vocals seemed to be covered up by the rest of the mix so an "immersivising" technique was used. This technique involved first creating what is known as FX objects; static objects that represent each speaker channel are to be fed audio data through auxiliary channels. Each channel has a discrete bus as an input. This setup enables any channel-based effect to be used in channel format and then sent to the appropriate object positions via bus sends, so not *every* effect requires its own objects. This practice is also fully scalable to any output setup. The "immersivisor" was created by making a mono bus send the same signal to three tiers of quick, slap-back delays. The first tier was set to a time of 33ms and sent to FX object in front of the listener. The second was doubled in time at 66ms and sent to the next row of FX objects; on the sides of the listener. The third and final tier was doubled in value again at 132ms and sent to the FX objects in the back of the room. The auxiliary channels used as well as their FX bus sends are shown in figure 4.1. Doing this subtly to the vocals made them stand out and fill up the immersive space while still keeping the integrity of the mix in tact.

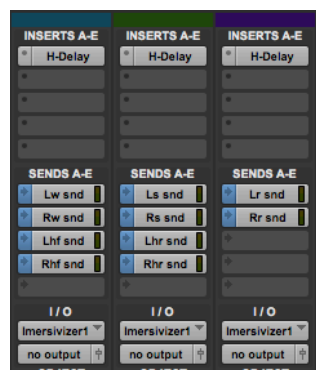

fig. 4.1

Now that the vocals took up a wider space in the mix, the reverb that was printed on them before didn't seem adequate compared to the increased size of the vocals. So, a new reverb was added and bussed to the two rear FX objects. The reverb chosen was a convolution with an impulse response from a plate. It was tuned to sound as close as possible to the reverb printed on the vocal stems.

Lastly, when compared to the stereo mix, the punch of the kick was nowhere near as present in the lower register once it was converted to audio objects. This was a small issue that the engineer decided to leave in the mix to see how effective each mastering technique was at remedying it.

Two loudness measurements were taken from the object-based mix, the first was right after the objects were generally positioned while the second was done after all the moves were done to help translate the mix detailed above. These moves include the “Immersivisor,” FX Objects, and vocal reverb. These moves ended up adding exactly .9 LUFS to the immersive mix. The loudness readings from all the masters completed by the engineer in this next phase are shown below in Table 2. The atmos mix that includes the added processing will be the starting point for each immersive mastering technique in the following sections.

Table 2

Technique	LUFS
Original Mix (Not Spatialized)	-15.6 LUFS
Original Mix (Spatialized)	-14.7 LUFS
Technique A	-13.42 LUFS
Technique B	-12.92 LUFS
Technique C	-13.69 LUFS
Technique D	- 13.99 LUFS

LUFS of Masters

2.2 Mitad Del Mundo Technique A Notes

After preparing Mitad Del Mundo in the first stage, the engineer exported an .adm file of the spatialized mix to master. Technique A was centered around processing each individual object with its own discrete processing. In total, there were 33 mono objects; 10 stereo pairs of instruments and 13 mono FX objects.

The engineered applied dynamics while taking into account what the experts had said about perceiving the system as a wavefront. This idea helped with the issues that may surface when listening to fold downs of the master. When a 9.1.4 master gets folded down into 2 channel stereo, there will be a boost in overall level and volume as the renderer is moving energy that previously existed 360º around the listener to the 60º stereo angle. It was found that, when dealing with objects that already existed within the stereo angle, more compression/limiting could be applied with less impact on the bass/level jump when switching to a fold down with less speakers. This mentality persisted to be valid throughout each mastering technique.

For this technique, limiting was applied to all objects across the board with objects in front receiving slightly more gain reduction. Compressors with matching attack and release times were applied to all instrument tracks. Threshold and make-up gain were adjusted so that each object was receiving around -2db of gain reduction for objects in the front angle and -1.5db for surrounding objects. The drums were treated with an EQ to boost the low end that was diminished upon spatialization. The keys as well as all of the surrounding FX objects were treated with similar multi-band exciters to bring more energy and clarity to the master. This master was finished and metered at -13.42 LUFS. The signal flow chart detailing this method is labeled figure 4.2.

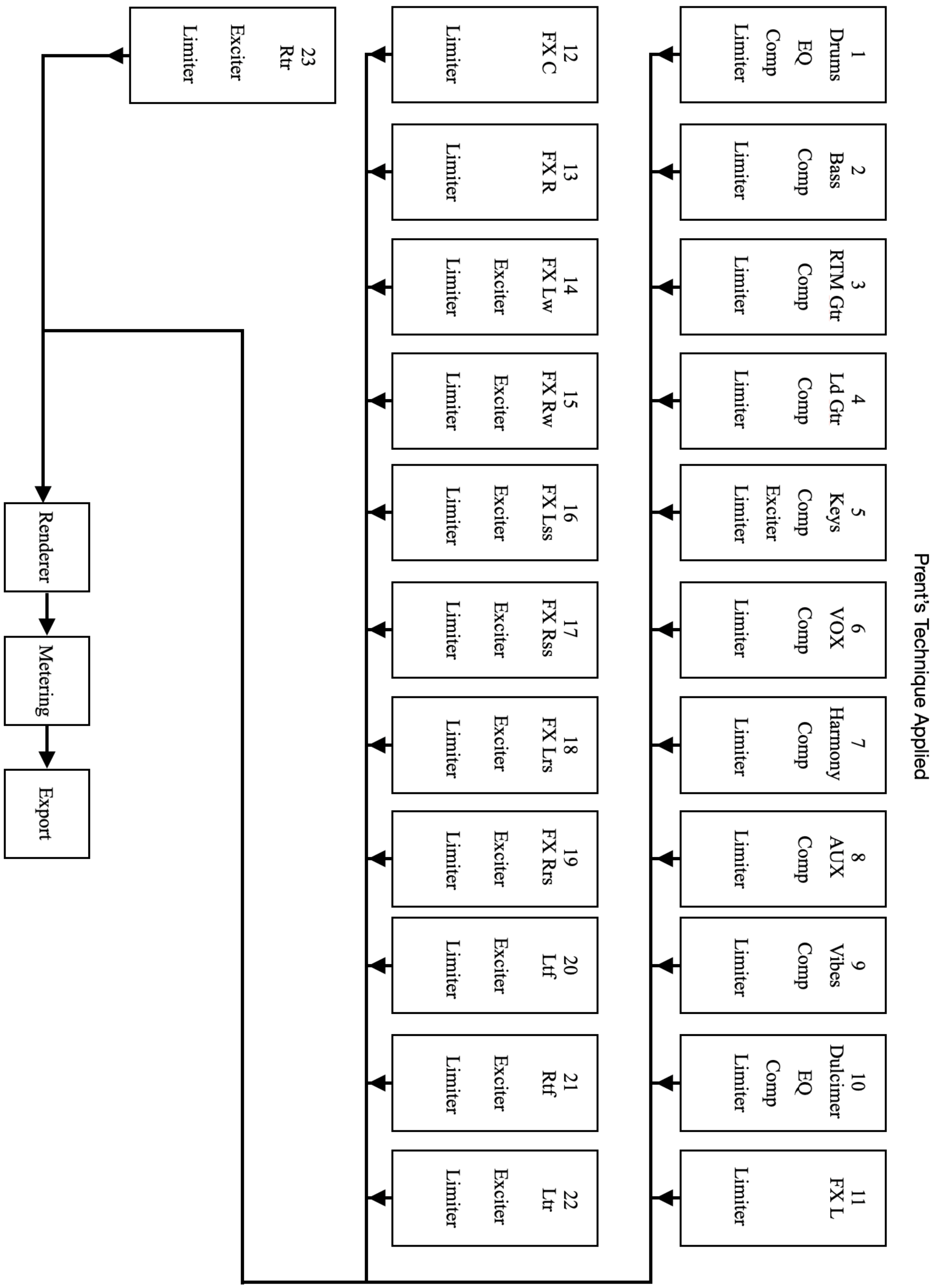
Prent's Technique Applied
1 Drums EQ Comp Limiter
2 Bass Comp Limiter
3 RTM Gtr Comp Limiter
4 Ld Gtr Comp Limiter
5 Keys Comp Exciter Limiter
6 VOX Comp Limiter
7 Harmony Comp Limiter
8 AUX Comp Limiter
9 Vibes Comp Limiter
10 Dulcimer EQ Comp Limiter
11 FX L Limiter
12 FX C Limiter
13 FX R Limiter
14 FX Lw Exciter Limiter
15 FX Rw Exciter Limiter
16 FX Lss Exciter Limiter
17 FX Rss Exciter Limiter
18 FX Lrs Exciter Limiter
19 FX Rrs Exciter Limiter
20 Ltf Exciter Limiter
21 Rtf Exciter Limiter
22 Ltr Exciter Limiter
23 Rtr Exciter Limiter
Renderer
Metering
Export

fig. 4.2

2.3 Mitad Del Mundo Technique B Notes

In this method, the mastering engineer hypothetically receives only the mix as objects representing speaker channels. It is important to note that all the objects that have been rendered into the speaker channel objects most likely had binaural rendering done to them in the Atmos renderer. So when setting up the renderer, make sure all the speaker channel objects do not have any binaural rendering applied.

All objects were treated with the same level of limiting. The L,C & R channel objects also had compression, exciting, and EQ applied to them for presence. The rear and height channel objects had similar exciting and EQ applied to them as well. This master was finished and metered at -12.92 LUFS. The signal flow chart detailing this method is labeled figure 4.3.

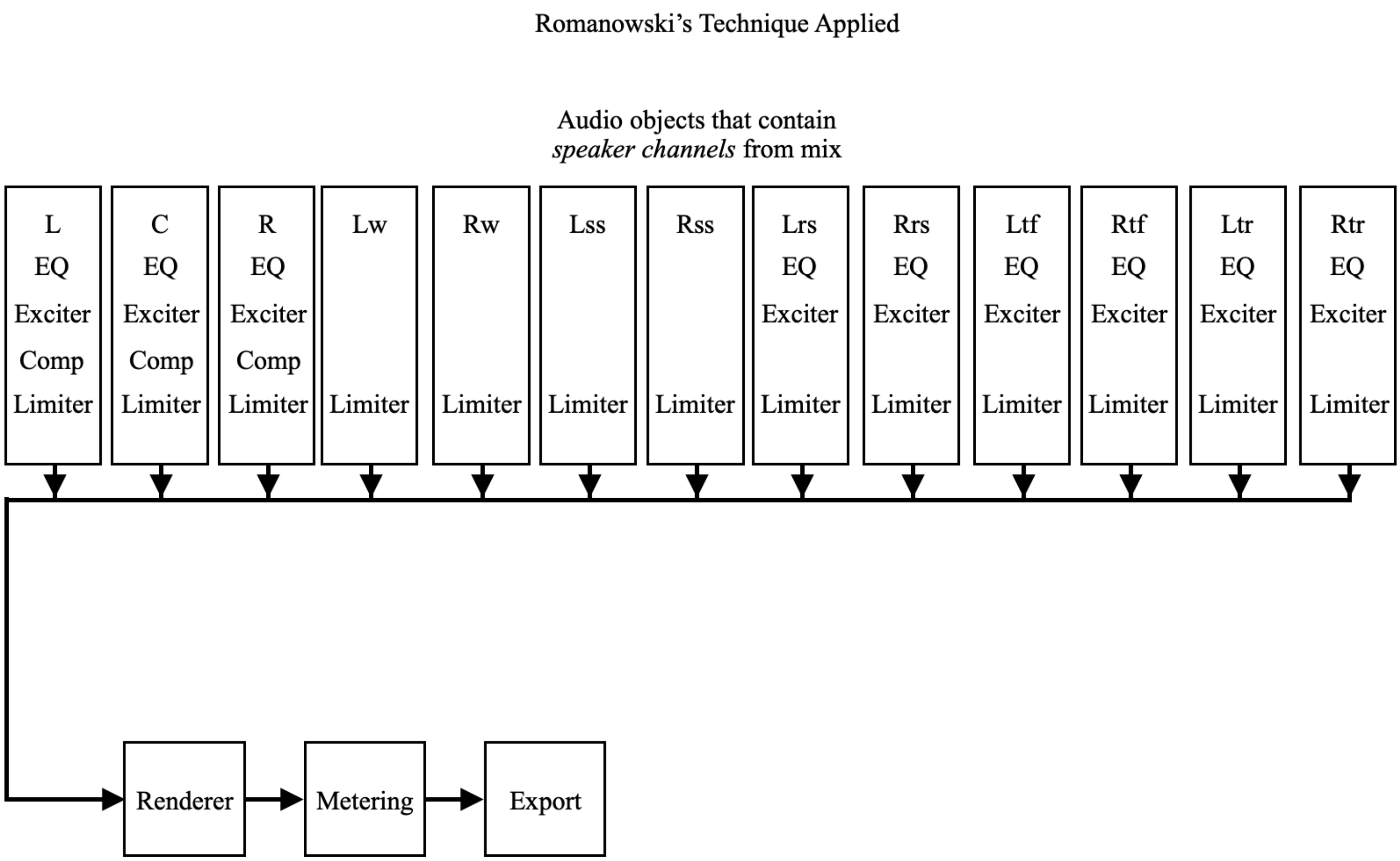

fig. 4.3

2.4 Mitad Del Mundo Technique C Notes

Technique C was similar to A in the way that it was dealing with an .adm file containing all objects in the mix. The major difference was in the way the compression was applied. First, the engineer created a mono audio track and bussed all tracks to it. When recorded, this created a mono summation of all objects. After it's finished recording, the engineer listened to the mono sum and applied EQ to subdue the bass buildup that occurs. After the mono track sounded balanced, its output was set to a discrete bus named "Trigger." This bus was routed to the key input of all compressors across all objects. This technique was supposed to emulate a master bus compressor such as in stereo mixing and mastering where the loudest element in the mix pushes down the other sounds as it increases in volume.

For this master, every track had a compressor that was being triggered by the mono summation bus as well as an independent limiter. After the dynamics were applied, some excitement and EQ was needed on a few objects to maintain clarity in the master. This master was finished and metered at -13.69 LUFS. The signal flow chart detailing this method is labeled figure 4.4.

Michael Matsakis
Thesis Proposal and Method, Fall 2020
Advisor: Paul Geluso

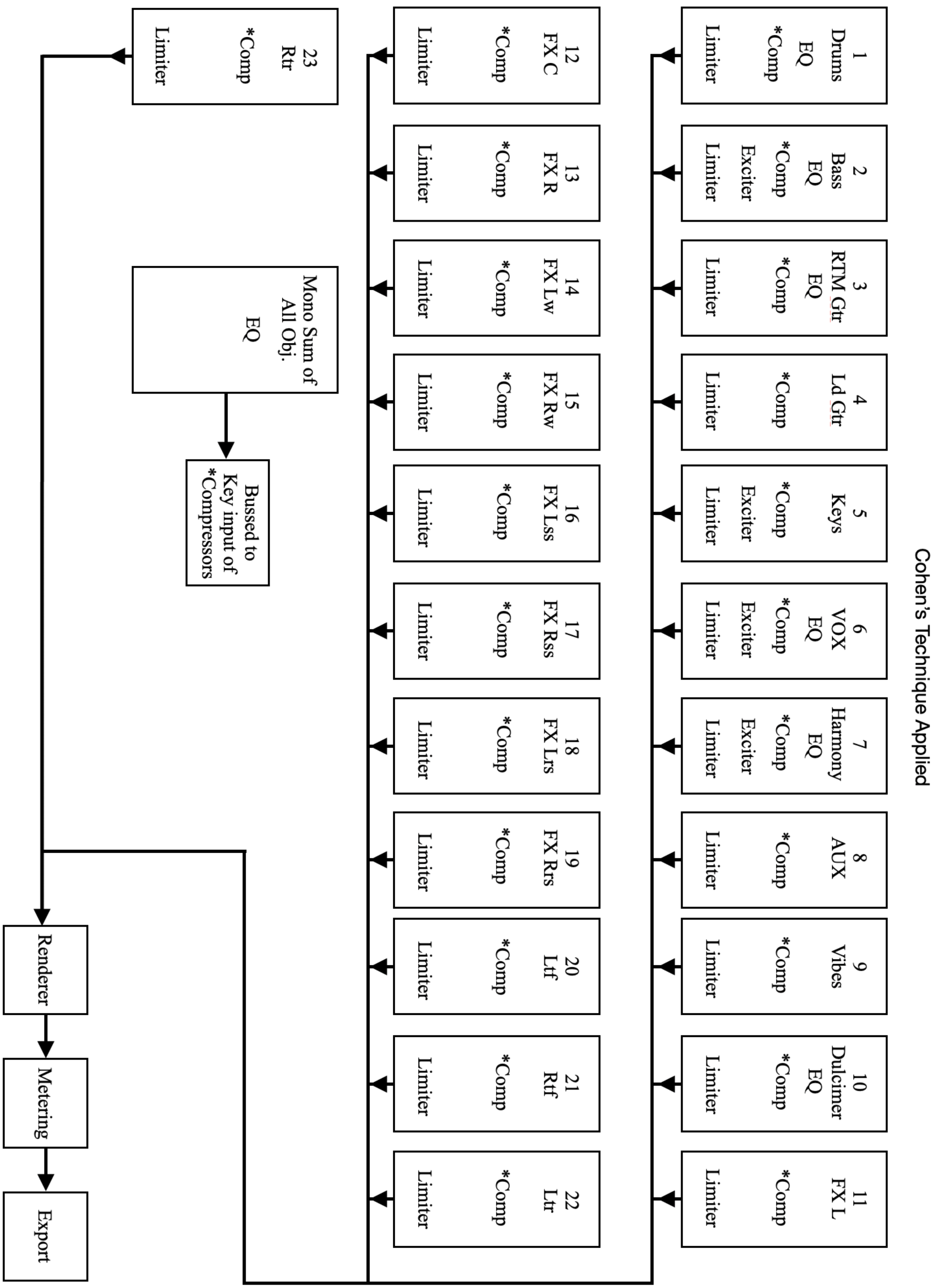

fig. 4.4

2.5 Mitad Del Mundo Technique D Notes

During the process of applying the learned mastering techniques, the engineer felt as if compression/limiting was easiest and most transparent when applied to speaker channels as objects. Although, more fine tuning to specific sounds can be done when working with all objects individually. So, it was thought that if frequency based processing such as EQ and excitement were applied in a setting like Prent's and Cohen's with all objects, while dynamics processing is applied afterwards in a "speaker channels as objects" setting, one might get the benefits of both.

For this technique, there were two phases. The first phase began with importing the same .adm file into ProTools and applying only frequency based processing. After completing this step, the phase 1 master was exported out of the renderer as speaker channels and imported into a new ProTools session. Phase 2 began with turning off the binaural rendering for the speaker channel objects in the new Atmos master file. Then, in ProTools, all tracks were summed into a mono audio track which was sent to a discrete bus to trigger compressors on each speaker channel object. Individual limiters were also applied on every track. The master was completed and metered at -13.99 LUFS. The signal flow charts detailing both steps for this method are labeled figures 4.5.1 and 4.5.2.

Combination Technique Phase 1

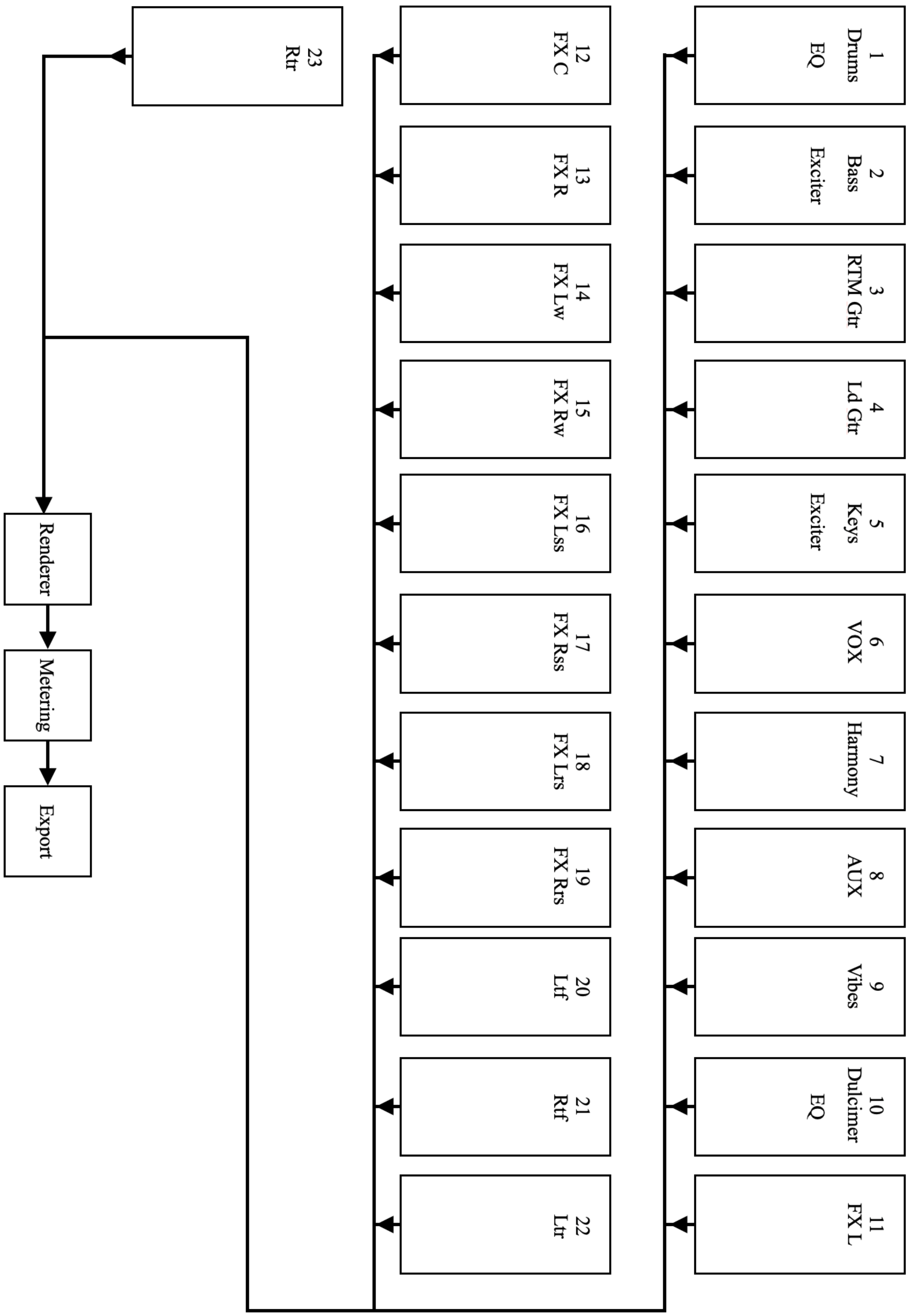

fig. 4.5.1

Combination Technique Phase 2

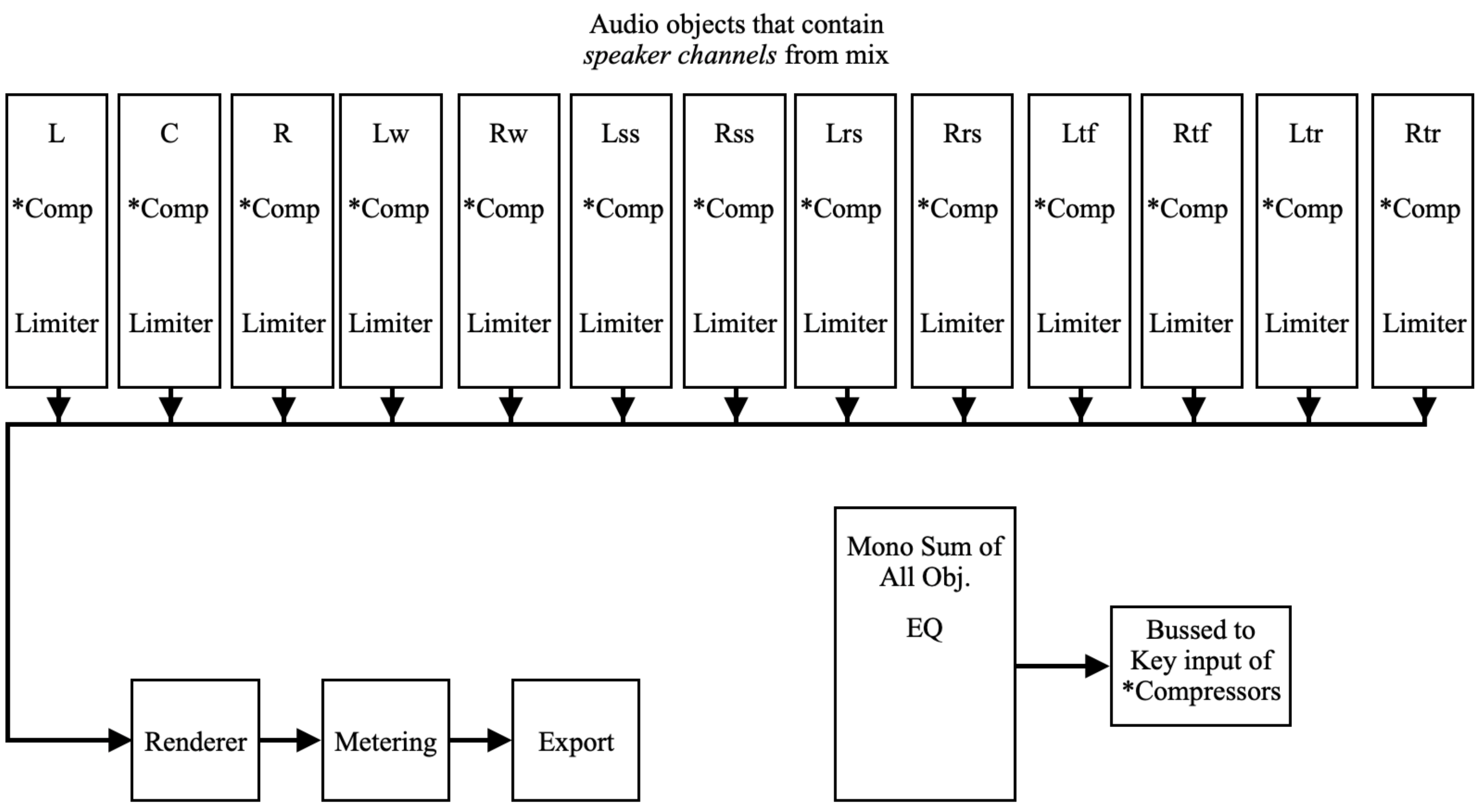

fig. 4.5.2

V. Expert Panel Evaluation

To subjectively judge the quality of the masters created by the engineer, three expert opinions were sourced. The experts each had varying specialties, each within the audio engineering field; Expert A was the channel-based audio expert, Expert B was the immersive audio expert, was Expert C had experience in both so their role was to bridge the gap between the other two experts.

The engineer prepared the four gain-matched masters created with different methodologies for each of the experts. They were told to choose their most preferred listening medium for monitoring the masters and the engineer delivered a re-render of the atmos mix in that format. This scenario seemed to be the most realistic regarding how immersive masters might realistically be experienced by most listeners today. All of the experts chose to listen in two channel stereo. During the panel, they listened to each master and gave feedback regarding the dynamics, frequency roll-off, immersion/localization, and overall audio quality. Overall the judges preferred technique C to all others, followed by A. While techniques B and D were the least preferred.

1. Techniques A and C Review

Being the overall preferred two masters, technique C was slightly better received by the experts due to the quality of its compression. In technique A, Expert C noted that they could hear too much of the compression pumping and it was throwing off the rhythm of the track. While C's compression felt better in terms of keeping the punch and rhythm from the track in tact. Expert C also thought the frequency response in A was a little too bright compared to the other masters. Expert B thought that both A and C were clearest in terms of their ability to localize sounds. Expert A described these two masters as superior in overall audio quality to B and D. They felt as if both of these masters drew their interest, kept their attention, and were the most enjoyable in terms of their inherit musical quality.

2. Techniques B and D Review

These techniques were overall rated lower than A and C. The most apparent critique the experts made regarding these masters is that they both seemed to sound a bit "lossy," as in .mp3 compressed. Through the discussion, it was deduced that this was most likely a result of re-spatializing the object-based mix *again* as speaker channels. Something that the engineer did in the process of creating the object-based mix as speaker channels made the overall audio quality drop. Other than that larger error, Expert B also thought D sounded slightly distorted and "boomy" in the low end. Expert A thought D sounded the least processed but not as exciting as the other masters. Expert C agreed with Expert A and added that master D sounded like a less impactful version of C.

VI. Discussion

1. Personal Review

Each technique deemed to have its own strengths and weaknesses. A and C were similar in the way that all objects were being dealt with individually. To the engineer, these techniques felt more like mixing, *but* could be all the master needs on a case by case basis. Thomas spoke about sometimes the most useful thing one can receive from a master is another set of trained ears listening to the mix on a trustworthy system. In this sense, just an approval could be considered a complete master. The biggest benefit from mastering in this “all objects” scenario is that frequency based processing such as EQ and saturation can be applied to specific objects. This is very powerful but at the same time, one must be careful to not re-mix the song in this environment, only to enhance and clarify the good that already exists.

In the engineers’ case, most of the objects already had compression on them, so by applying more to each track (with technique A) seemed to increase the probability of messing up the overall rhythm of the track. In Technique C, the compression was more uniform and flexed as if it were a master bus compressor. But this pulsing tended to diminish the high end of the master when compared to Technique A. Technique C’s master did not meter quite as loud as A’s while at the same time, felt more tightly compressed. Also, the sound quality was overall slightly better in A’s compared to C’s.

Technique B felt more like traditional channel-based mastering because the processing was now being applied in a broader sense. It was noted that using EQs in this method seemed to be much more intrusive than when dealing with all objects individually–akin to channel-based mastering. Although, dynamic processing or saturation seemed to work quite well. The overall rhythm of the track was more pronounced with the applied compression in Technique B compared to A’s or C’s. The biggest setbacks to this method were mainly centered in not being able to adjust a specific object or perform any modifications to the spatialization. The engineer noted that the roles of the object-based mastering engineer would change in this new era of mastering. So, Technique B works unless the client mixed the song on a two channel system (which will probably be likely in the future) and sent the mix to the mastering engineer to check spatialization as well as loudness and frequency rolloff. In that scenario, the object-based mastering engineer would need to receive the mix as all objects. The engineer believes that this is where the Technique D could come in useful.

Taking the strengths from each method, one could be very effective with their mastering with little processing used if the processing is applied at the correct moment. The dynamics were most transparent and useful when applied to speaker channels as objects while the EQs were more so in the "all objects" environment. The engineer also wanted to utilize Cohen's idea of the master bus compressor to try and push the master a but further than the others. This synthesized approach worked well but led to some over-compression during the first attempts at mastering. As far as the current track "Mitad Del Mundo" goes, the dynamic range was already somewhat heavily limited and compressed already. The compression applied with the Cohen's master bus technique was pretty intense and maybe not quite necessary in this context. The first couple of masters done with Technique D turned out to sound over-compressed compared to the others. When an object-based master is over compressed, the spatialization/immersion is greatly diminished. The phantom imaging gets thrown off resulting in the sound "shrinking" into each speaker instead of blooming and creating a sound field around the listener.

2. Expert Panel Review

The experts each first noted that the four masters were more similar to each other than they were different, and that they all could pass individually as a professional masters. But, due to some errors that the engineer was not hearing while working, techniques B and C came across worse than anticipated. The biggest takeaway is the note about the drop in overall quality when the track was put through the renderer more than once. Even with the binaural rendering modes turned off on the second rendering step, there was a notable drop in quality for the experts while listening to techniques B and D. To the engineers knowledge, the speaker channels were created correctly out of the Atmos renderer. If this was the case, then there seems to be some kind of loss of information when the Atmos renderer is creating the speaker channels. And just re-spatializing those channels as objects does not work as seamlessly as thought initially.

If the re-spatialization of the speaker channels as objects was able to be done in a seamless and lossless way, then techniques B and C may actually have the chance to be much more useful. Maybe a different renderer is able to do it better than Atmos at the moment. It seems like the idea was solid, but the technology was not designed to be used exactly in that way.

VII. Conclusions and Future Work

Out of the four distinct methodologies recorded, most of the weakness could be remedied through more practice and/or by using different rendering/DSP software than the testing engineer. After attempting the techniques, the engineer felt that object-based mastering was possible to accomplish with the current technology. The expert panel unanimously agreed that each master presented to them could pass as professional work. They felt as though techniques A and C were the best overall for *this* master. Technique D seemed as if it has the potential to be the framework for most mastering workflows due to its flexibility. But, it was found that putting an Atmos re-render back into the renderer degraded the sound quality, making it impossible to use techniques B or D with the Atmos renderer.

The reality of working with audio content is that there will never be *one* way that will work for *every* situation. The techniques and methodologies learned in this paper should be taken broadly as starting points for different kinds of object-based masters. Engineers should be most cautious about the current technological issues discovered with the software used: ProTools is not a DAW that is optimized for object-based music workflows, the Dolby Atmos renderer does not do well re-spatializing speaker channels from a re-render, & creating a phantom master bus compressor can not be done in real time.

This study was not entirely comprehensive of the current state of object-based mastering, but more of an evaluation of popular techniques that are currently being used and developed in the professional world. If this research were to be taken further, more DAWs and renderers should be used on a wide variety of musical genres. Particularly for techniques B and D, which were put at a disadvantage due to the loss of quality when put through the Atmos renderer twice. It seems as if the streaming world is wanting to support object-based music content more as time goes on; but, for now the mediums has a hard time stacking up to the stability of channel-based audio. Although, as media consumption itself becomes more immersive for the user, music is bound to follow accordingly.

References

Byers, R., Johnston, J., Kean, J., Lund, T., Orban, R., & Wisbey, A. (2015). *Recommendation for Loudness of Audio Streaming and Network File Playback.* In Audio Engineering Society Conference: 2015. AES E-Library

Cohen, R. (2021, March 14). Object-Based Mastering Interview [Online interview].

Dolby Laboratories. (2019). *Dolby Atmos Conversion Tool v1.9 Software Guide.* Retrieved from https://professional.dolby.com/

Dolby Professional. (n.d.). *Dolby Atmos Music for Professionals*. Retrieved from https://professional.dolby.com/

Francombe, J., et al., (2017). *Evaluation of Spatial Audio Reproduction Methods (Part 2): Analysis of Listener Preferences.* In Audio Engineering Society Conference: 2017. AES E-Library

Hestermann, S., Seideneck, M., & Sladeczek, C. (2018, July). *An Approach for Mastering Audio Objects*. In Audio Engineering Society Conference: 2018 AES International Conference on Spatial Reproduction-Aesthetics and Science. Audio Engineering Society.

ITU. (2012). Algorithms to Measure Audio Program Loudness and True Peak Audio Level. Retrieved from http://www.itu.int/publ/R-REC/en. International Telecommunication Union.

Katz, R. A. (2015). *Mastering audio: the art and the science.* Burlington, MA: Focal Press, Taylor & Francis Group.

Mayo, C. G., (1940). *Limiter Design*. British Broadcasting Corporation R&D. Retrieved from http://downloads.bbc.co.uk/rd/pubs/reports/bbc-rd-report-list.pdf

Norcross, S., Pribadi, M., Nanda, S. (2017). *Practical Loudness Measurement and Management for Immersive Audio*. In Audio Engineering Society Conference: 2017. AES E-Library

Oramus, T., & Neubauer, P. (2019). *Comparison Study of Listeners Perception of 5.1 and Dolby Atmos*. In Audio Engineering Society Conference: 2019. AES E-Library

Poers, P., (2015) *Monitoring and Authoring of 3D Immersive Next Generation Audio Formats*. In Audio Engineering Society Conference: 2015. AES E-Library

Prent, R. (2021, February 1). Object-Based Mastering Interview [Online interview].

Roginska, A., & Geluso, P. (2018). *Immersive sound: the art and science of binaural and multi-channel audio*. New York, NY: Routledge.

Romanowski, M. (2021, February 21). Object-Based Mastering Interview [Online interview].

Rumsey, F. (2018). *Spatial Audio, Channels, Objects, Or Ambisonics.* In Audio Engineering Society Conference: 2018. AES E-Library

Thomas, C. (2021, March 5). Object-Based Mastering Interview [Online interview].

Vickers, E. (2010). *The Loudness War: Background, Speculation, and Recommendations*. In Audio Engineering Society Conference: 2010. AES E-Library

Woodcock, J., et al. (2018). A Framework for Intelligent Metadata Adaption in Object-Based Audio. In Audio Engineering Society Conference: 2018. AES E-Library