Evaluation of Binaural Mixes of Popular Music for Binaural Virtual 9.1 Auro-3D Reproduction

Christal Jerez
April 17, 2017
New York University
Advisor: Dr. Agnieszka Roginska

Submitted in partial fulfillment of the requirements for the Master of Music in Music Technology in the Department of Music and Performing Arts Professions in the Steinhardt School of Culture, Education, and Human Development, New York University.

Abstract

In January of 2017, Nielsen SoundScan reported that "music consumption was at an all-time high" (Washenko, 2017). This can partially be attributed to the increase of on-demand streaming services like Spotify, Apple Music, TIDAL, and more. In the same report, Nielsen reported that 80% of music listeners used one of the above named services (Washenko, 2017). An increase in music consumption can also be associated with an increase in music listening via headphones. In 2016, New York City reported that 42% of younger adult listeners (aged between 18-44 years old) used headphones to listen to a personal music player (Miller, C., & Martinez, J, 2016). With this noticeable increase in both music consumption and music listening via headphones, the question is posed if mixing engineers should mix binaurally for music reproduced over headphones.

In the attempt to apply the same immersive audio reproduction methods used for cinema to the world of popular music reproduction, this thesis explored binaural mixes of the increasingly popular Auro-3D 9.1 reproduction method. Using ProTools and the G'Audio Plug-In, three mixing engineers completed a stereo mix using loudspeakers, a binaurally rendered Auro-3D 9.1 mix completed over loudspeakers, and a binaurally monitored virtual 9.1 Auro-3D mix using headphones. This process was completed for two different songs resulting in three mixes per song and six mixes per engineer. Clips from each mix were compared against each other in an online listening survey. The goal was to find a listener preference between the different mixing methods from both the general public and music technology professionals. Results showed that listeners most often preferred stereo mixes, however, mixes completed by engineers with more experience and/or prior experience in 3D audio, produced binaural mixes that were better evaluated and more preferred than stereo mixes. This thesis concludes that binaurally monitored mixes were better evaluated than mixes that were binaurally rendered after the mixing process.

Acknowledgments

There are many people that helped guide me through this process that I would like to acknowledge. Thank you to my advisor, Dr. Roginska, for helping me formulate my ideas into cohesive thoughts. I'd also like to thank the Tonmeister 2016 cohort and Professor Paul Geluso for being available to share my ideas and experiments.

I'd like to thank my studio families including the Dolan Engineers and the Platinum Sound Engineers. Special thanks to Mooch, Cris G., and Hai Li for your time! You all helped make this thesis happen.

Thank you to my family for their continuous support including my Hermanas, Liddy, my siblings and their minis, and my parents. And of course, thank you to my number one: Kris Stith!

Table of Contents

List of Images

List of Figures

List of Tables

List of Appendix Figures

List of Appendix Tables

1

Introduction

1.1 Binaural Music Listening

Binaural audio, an immersive audio reproduction method that allows for sounds to be reproduced over headphones the way sounds would be heard under normal hearing conditions, has become the focus of music technology developments (Begault, 2000). However, binaural audio has been limited to music technologists and hobbyists of the field – ultimately dooming its potential for commercial markets; however, music has now left the confines of the home listening system. In days of past, people took pride in their elaborate home stereo systems. Now music is mobile. People listen to music on the go and take pride in their headphones. What if it were possible to take advantage of headphone monitoring while pushing binaural audio towards the consumer market?

The advance of virtual reality has slowly begun to bring binaural audio into consumer products. Similarly, the cinematic use of immersive audio reproduction systems like Dolby Atmos and Auro-3D has introduced 3D audio as the forefront of music technology; however, music reproduction and 3D audio have not been fully joined. Whether it be through binaural or multichannel reproduction, 3D music reproduction has still not found a place in consumer markets.

Previous research has concluded that binaural music listening was unimportant for an immersive listening experience (Fontana et al., 2007) and that binaural synthesis was insignificant for music listening (Payri et al., 2016). However, prior research either used binaural recordings or binaurally produced content that was not monitored binaurally during production. The distinction between this thesis and previous research is that the mixing monitoring method was changed in order to match the reproduction method. Binaurally monitored mixes created for binaural reproduction has not been the previously tested.

In this exploration of the potential for binaural 3D music listening, this thesis compared three different mixing approaches. Three professional mixing engineers created: (1) stereo mixes mixed over loudspeakers. (2) Auro-3D 9.1 mixes mixed over loudspeakers. (3) Virtual 9.1 Auro-3D mixes mixed binaurally over headphones. The mixing engineers mixed the same two songs that resulted in a total of 18 different mixes that were evaluated through listening surveys. The listening surveys evaluated listener

preferences and perceptions of the timbral and spatial qualities of each mix. The listening surveys were divided into two pools of participants: the general public and music technology professionals. The conclusions drawn from the listening surveys show the future potential of exclusive processing similar to *Mastered for iTunes*, but rather *Binaurally Mixed for Headphones.*

1.2 Motivation and Objectives

The motivation of this thesis was to evaluate and explore the different binaural mixing methods for popular music reproduction. The goal was to assess how binaurally monitored virtual Auro-3D 9.1 mixes would compare against stereo and virtual Auro-3D 9.1 binaurally rendered mixes. The objectives of this thesis therefore were:

(1) to conduct and evaluate a series of binaural mixing sessions
(2) to conduct listener surveys to evaluate the binaural mixes

1.3 Thesis Outline

This thesis is arranged into five separate chapters. Chapter 2 provides a background on music mixing, immersive audio, and perceptual audio evaluations. Chapter 3 discusses the different mixing procedures and how those procedures were evaluated by the mixing engineers. Chapter 4 reviews both the listening survey and analyzes those results through inferential statistics. This thesis concludes with Chapter 5 with a discussion on the significant findings of this thesis and how those findings contribute to the fields of music mixing and binaural audio.

2

Literature Review

2.1 Overview

The literature review ranges from topics that relate to music mixing, binaural audio, virtual surround sound, and perceptual testing. These topics were covered because they directly affected how the author approached the different mixing procedures and listening surveys that were implemented.

2.2 Common Music Mixing Techniques

Mixing music is often thought of as an art or a skill. Each twist of a knob or boost of a frequency is comparable to the way a painter uses color or stages a composition. There is no one book or guide to painting like Pablo Picasso; there is also no one set method to achieve "the perfect mix". There are many philosophies on how to approach a mix; what techniques to use, and which steps to take when working on a mix. Books on mixing generally recommend implementing a procedure to "simplify the process in order to spend more hours of quality time in your creative headspace…" (Izhaki, 2012, p. 153). Therefore, the session workflow was carefully designed for the mixing sessions that were carried out through this thesis.

The mixing process although not the same, has similarities across the books. Most often, it is recommended to organize your mix, balance levels, pan, equalize, compress, then add effects (Izhaki, 2012, p.154). While the order varies slightly across different resources, there is a general consensus of the steps that need to be taken to achieve a strong mix.

Vocal clarity for pop productions is a specific skill. While there are general mixing guidelines, more detailed instructions exist for vocal intelligibility. A comprehensive overview of some of the most common vocal mixing techniques were included in research by Ronen et al. including:

(1) Boosting the lead vocals by 2dB
(2) Compressing the lead vocals using a 4:1 ratio with a medium attack and release, while maintaining a gain reduction of 3-6 dB with a makeup gain
(Ronen et al., 2015, p. 4).

Because the songs used in this thesis were pop songs by a female vocalist, a combination of general mixing methods as well as those specific to vocal mixing were made available to each mixing engineer for their mixing sessions. The mixing session's design and workflow are discussed in detail in Chapter 3.

2.3 Headphone vs. Loudspeaker Monitoring

A major part of the mixing process is the way mixing engineers monitor their mixes. Each mixing engineer has their preference of monitors, but almost all would agree to mix on speakers and check on headphones. "Headphone monitoring while mixing music is considered advantageous by many mixers, as it eliminates the mixing room's acoustic properties and possible faults, including background noise, strong reflections, and heavy resonances" (King, 2011, p. 1). While acoustic isolation from headphones is beneficial, it also has drawbacks. Rarely is it encouraged to mix exclusively on headphones. "The isolation of headphone listening provides a further acoustical detachment from a listener's physical surroundings" (Toole, 1984, p. 9). A mixing engineer's physical surroundings greatly contributes to the decisions made when working on a mix.

Room resonances are common occurrences in listening spaces. Even the acoustically treated control rooms mixing engineers work in still effect how sound is perceived. Ideally, a mixing engineer wouldn't be sitting in a low-frequency room node, but not all mixing engineers work in acoustically treated rooms. If a mixing engineer was sitting in a low-frequency room node, then they might be over boosting the low-end to make up for the null they are experiencing. Similarly, room reflections, which also contribute to a room's reverberation time, can have a strong impact on mixing decisions. A room with many reflections can blur a sonic image preventing the mixing engineer from adding the appropriate amount of reverb. In other instances, a dry room might cause the mixing engineer to add too much reverb in the attempt to produce more reflections in a naturally non-reflective space. Even the monitors that are used for mixing greatly impact the mixing process. Mixing on near-field monitors that might have a brighter tone could cause a mixing engineer to not properly treat the low-end. Similarly, mixing without a subwoofer could cause a mixing engineer to over-emphasize the low-end that their monitors simply cannot reproduce.

Monitoring over loudspeakers provides acoustic interactions with the listening environment that is not possible with headphone monitoring. Although headphones are not the preferred monitoring method by mixing engineers, headphones, both in-ear and over-the-ear have become the standard monitoring method

for consumers. Currently, most mixes are made for loudspeaker reproduction, but are often reproduced exclusively on headphones. While mixing on headphones might not be the preferred mixing monitoring method for stereo mixes, it is proposed as a more appropriate mixing process for binaural mixes. Before discussing binaural audio, the concept of immersive audio needs to be introduced.

2.4 Immersive Audio

Immersive audio is a distant concept to those who are unfamiliar with the field. In the context of this thesis, immersive audio is relevant through "virtual acoustics, binaural audio, and spatialized sound" (Begault, 2000, p. X). Virtual acoustics can be used to represent the spatial attributes of a virtualized environment. This can be achieved through impulse responses that characterize a space. Azimuth and elevation represent the three dimensional sound field within spatialized and binaural audio. An example of spatialized sound in relation to this thesis is the Auro-3D listening system. A main layer of surround sound loudspeakers is paired with a height layer of loudspeakers representing surround sound in conjunction with elevation. This thesis deals directly with the Auro-3D 9.1 configuration and its binaural downmix.

2.4.1 Auro-3D

Auro-3D was introduced soon after the rise of 3D movies in 2005 (Auro Technologies N.V., 2015). Auro-3D's design focuses on the addition of the height, or the Z plane, in its reproduction system. While there are a few different Auro-3D configurations such as 9.1 and 11.1, this thesis implemented the Auro 3-D 9.1 reproduction configuration also shown in *Image 2.1*. The 9.1 Auro-3D configuration consists of:

- Main Layer: Left, Right, Center, Left Surround, and Right Surround
- Height Layer: Left Height, Right Height, Left Surround Height, and Right Surround Height
- LFE channel (Auro Technologies N.V., 2016).

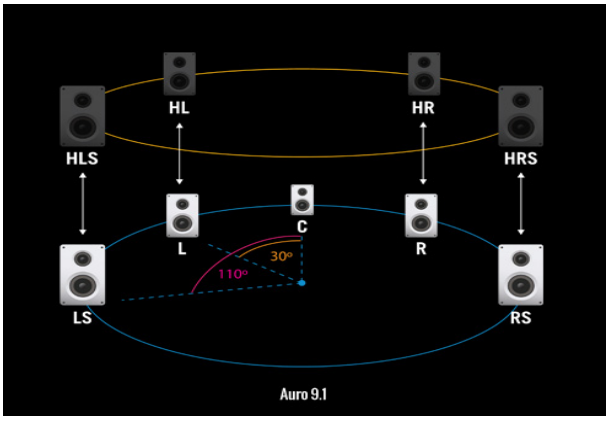


Image 2.1 **The Auro-3D 9.1 Setup (Auro Technologies, 2017).**
This image shows the main five channel layer along with the quad height layer, making up the 9.1 Auro-3D setup.

The largest benefit of the Auro-3D format is it backwards compatibility with the surround sound format. Because the main layer consists of the 5.1 setup, it is easy for individuals, studios, and production companies to transition to the Auro-3D format by simply adding at minimum, four height layer loudspeakers.

Auro-3D has also made their technology more accessible through their Digital Audio Workstation (DAW) compatible plug-ins. The Auro-3D Engine allows mixing engineers to incorporate the different Auro-3D formats into their workflow as long as they have the appropriate speaker reproduction systems. The different tools available within the Auro-3D Creative Suite include: Auro-Codec, Auro-Matic Pro, and Auro-Panner. The Auro-3D creative suite allows for upmixing, 3D panning, and 3D encoding (Auro Technologies N.V., 2016). In addition to being able to create an efficient Auro-3D workflow within Pro Tools, the Auro-3D engine also allows for downmixing of multichannel mixes.

Auro-3D created the Auro-3D Headphones plug-in which allowed for binaural monitoring of multichannel mixes; however, Auro-3D has since stopped distribution of the plug-in order to better integrate their headphone technology with mobile devices. Auro-3D Headphones is still offered through the Virtual Reality developer, AudioKinetic, who incorporate binaural audio into their virtual reality gaming designs. Auro-3D has made their technology an accessible and viable means for studios to transition to immersive audio making it an integral tool to incorporate as a part of this thesis.

2.4.2 Stereo and 3D Upmixing

The audio content used in this thesis used stems from already released popular music. These stems were not intended for any sort of surround or 3D reproduction. By using those stems and trying to maximize them for 3D audio playback, the content was upmixed from its original stereo format. Although this method doesn't meet the technical definition of upmixing, upmixing must be discussed to understand other approaches of creating 3D content from mono or stereo signals.

Upmixing takes a single signal and creates a different version of the signal for reproduction beyond mono. Manfred Schroeder first discussed upmixing in his 1958 paper, "An Artificial Stereophonic Effect Obtained from a Single Audio Signal". Schroeder defined upmixing as "…a situation in which a spatial subjective impression is created by applying to the two ears two different versions of a single audio signal" (Schroeder, 1958, p. 1).

One common upmixing process was established during the transition between mono to stereo reproduction. "This typically involved taking two copies of the signal, and delaying one copy by 30-40 ms, and then high-pass filtering one copy and low-pass filtering the other copy" (Fitzgerald, 2011, p. 1). Other

methods for mono-to-stereo upmixing which also can be applied for stereo-to-3D upmixing include signal decomposition and sound source separation.

An example of upmixing based on signal decomposition for mono-to-stereo upmixing was completed by Uhle et al. where sources were decorrelated into foreground and background signals using frequency domain processing (2016). Signals are decorrelated in such a way that vocals, drums, and guitar, for example, are "sufficiently separated to allow directionality to be imposed on the sources" (Fitzgerald, 2011, p. 2). The same concept applies for stereo-to-3D upmixing.

While these methods are advanced upmixing procedures, this thesis minimized processing and had the engineers choose where they wanted to output the stems within the Auro-3D 9.1 configuration. This method also helped investigate if it would be possible to create immersive content without upmixing. Would it be possible for engineers to use the same stems from the stereo mixing sessions and reimagine them for immersive binaural reproduction?

2.4.3 Binaural Audio

Binaural audio refers to "techniques where the outer ears (the pinnae) are either directly implemented or modeled as digital filters" (Begault, 2000, p. X). The techniques stimulate psychoacoustic cues that enhance inside-the-head sound source localization. One such technique is the application of Head Related Transfer Functions (HRTFs) which superimpose spectral characteristics on audio. HRTF measurements represent the Interaural Time Differences (ITDs) and Interaural Intensity Differences (IIDs) between the left and right ears, making each HRTF unique to an individual (Begault, 2000, p. 31). HRTFs specifically "capture the location-dependent characteristics of sound with respect to a person's anthropometric properties" (Andreopoulou et al., 2013, p. 1). Although research has shown that individualized HRTFs are better for localization, generalized HRTFs acquired from good localizers are more practical and will be used for the binaural synthesis portion of this research (Begault, 2000, p. 99).

Mixing binaurally for binaural reproduction is mostly unexplored territory. Rather, past studies have focused on mixing binaural content through post-binaural processing. A study in 2007 by Fontana et. al. combined binaural recordings from a Neumann KU-100 dummy head with traditional studio miking techniques to create binaural mixes of pop songs through post-HRTF processing. The mixes were heard by participants through listening tests which evaluated the spatial and timbral characteristics of the mixes. The timbral quality test results showed that stereo mixes were evaluated significantly better when compared to the binaurally synthesized mixes. The authors proposed that the HRTFs added unwanted coloration to the music as a result of the less refined state of binaural technologies (Fontana et al., 2007). The study outcomes were inconclusive, pointing that binaural mixes did not impact the listening experience.

The state of HRTFs research and technologies has progressed since the above-mentioned study. While it is known that non-individualized HRTFs introduce spectral coloration, research has gone into improving such coloration. One study by Takanen et al., found that "a significant amount of the coloration is introduced at high frequencies in the HRTF filter designs" (2012, p. 8). Using that information when binaural mixing can prevent mixing engineers from over-emphasizing the high-frequency spectrum in their mixes. Schönstein et. al conducted a study that focused on the reliability of listener judgements when evaluating HRTFs (2012). Their research results showed that more experienced listeners had less variance in their responses as opposed to less experienced listeners (Schönstein, 2012). Through the focused research on HRTF applications and evaluations, more reliable binaural listening tests were conducted for this thesis. The research above guided this thesis towards conducting two separate listening surveys towards music technologists and the general public.

More recent research done by Payri et al. (2016) directly compared listener perceptions of stereophonic and binaural mixes. In their research, binaural mixes of a string quartet, a jazz ensemble, and a chamber orchestra were binaurally processed through *Binaural Panner*, a tool featured in the popular music production DAW, Logic (Payri et al., 2016). The mixes were evaluated for perceptual differences between the stereo and binaural mixes. Listener test results showed that binaural synthesis in regards to the addition of elevation and spatialization were significantly irrelevant (Payri et al., 2016). While the listening test results concluded that binaural synthesis for music reproduction was not impactful, the binaural mixes were completed through a combination of binaural monitoring and post-binaural processing. Additionally, the authors focused their investigation on the distinct parameters of binaural mixes such as spatial and timbral attributes of binaural mixes.

Unlike the above-stated research, the binaural mixes were monitored binaurally during the entire mixing process. Although Fontana et al., concluded: "The results of perceptual tests on sound-caring listeners show that nowadays they do not seem sensitive to the benefits of binaural technologies…" (2007, p. 5), the procedures carried out through this thesis showed otherwise.

2.4.4 Virtual Speaker Reproduction

For this thesis, a virtual 9.1 Auro-3D mix was created through binaural rendering. Virtual speaker reproduction attempts to create the illusion of additional sound sources through binaural processing. An explanation of its use in creating Virtual 5.1 Surround Sound, for example, is provided by Pike et al.: "Binaural processing can be used to create an alternative signal for headphone listening from the 5.1 signal, attempting to give an improved spatial sound experience" (2013, p.1). Binaural synthesis creates virtual sound sources by superimposing HRTFs that correlate to speaker positions within the surround sound

listening environment. For example, a Left Surround Speaker can be virtually represented by processing it with an HRTF at 110º azimuth, 0º elevation. This process can be extended towards creating a virtual immersive audio environment. Binaural processing of a 9.1 mix would therefore include HRTF synthesis with both azimuth and elevation measurements. The elevation information helps create the illusion of the height speakers in the 9.1 Auro-3D environment.

As previously mentioned, binaural audio often introduces unwanted spectral characteristics, therefore, researchers have investigated other methods towards creating virtual immersive environments. One other method is through Vector Base Amplitude Panning (VBAP) which creates phantom images between speakers using amplitude, or intensity, panning. VBAP, a method proposed by Pulkki, simulates virtual sources based on the vectors produced between loudspeakers (Pulkki, 1997, p 458). This method can also be extended to three-dimensional sound fields to create elevated phantom images between main and height loudspeaker layers. While this method avoids the unwanted coloration of HRTF synthesis, this thesis focuses on binaural audio and will therefore incorporate binaural synthesis for virtual 9.1 Auro-3D reproduction.

2.4.4 2D vs. 3D Mixing

Research that helped focus the evaluation of virtual Auro-3D content is that which focused on mixing for immersive audio systems like Auro-3D and Dolby Atmos. Research by Cengarle et al., investigated the treatment of sound sources in 2D versus 3D playback (2013). An experiment which asked participants to identify hidden tones in stereo, 5.1, and 3D reproduction found that hidden tones were better located in 3D playback; therefore, “a wider panorama helps making individual sounds more audible and distinguishable” (Cengarle et al., 2013, p. 9). The researchers also gathered data during a second experiment which concluded that higher amounts of reverberation could be used in 3D playback without a loss of intelligibility (Cengarle et al., 2013).

Martin et al. (2015) investigated pop music mixing techniques for reproduction on 3D Audio systems like Dolby Atmos. What was unique about their study was their focus on stereo mixing tools such as discrete reverb and delay times in order to produce “believable three-dimensional immersion” (Martin et al., 2015, p. 2). The final 3D mixes were evaluated for: “(1) Immersion, (2) 3- Dimensionality, and (3) to determine the expansion of the sweet spot to the left and right of the center mix position” (Martin et al., 2015, p. 5). Although Martin et al. found that using conventional tools like discrete reverb and delay times to be inefficient, this thesis will incorporate a hybrid method that will include the use of common mixing methods along with a “simple, non-rendering 3D panning tool” (Martin et al., 2015, p. 7). The tool that was

used for the purposes of this was the G'Audio Works ProTools plug-in which is discussed in detail in Chapter 3.

2.5 Perceptual Audio Evaluations

The mixes created as a part of this thesis were evaluated through online listening surveys and were designed with guidance from existing research. As Rumsey recommended in his paper, the spatial attributes should be presented with clarity in order to not confuse participants and in order to attain reliable subjective data (Rumsey, 2002). In order to investigate parameters like timbre, naturalness, and spatialness - the definitions of those attributes were defined within the survey (Rumsey, 2002). The focus of the listener survey was to find a preference in the mixing method and not to judge mixing styles. Including a word bank with definitions of certain spatial attributes helped keep the participants focused on the sonic characteristics of each mix. The listening survey and survey design is discussed in Chapter 4.

3

Mixing Procedures and Results

3.1 Overview

The following sections describe the mixing engineers that participated in the mixing sessions, the software used for mixing and binaural processing, and the physical setups of the mixing environments. This section concludes with feedback from the mixing engineers on their experiences from each mixing session. All mixing sessions were held at NYU's Music and Audio Research Laboratory. Each mixing engineer completed three mixes for each song- meaning at six different mixes per each engineer, for a final sum of 18 mixes total.

3.2 Mixing Procedures

This thesis required three different setups for the mixing sessions that were used to create the test materials. The three mixing environments that were implemented were: a stereo loudspeaker setup, an Auro-3D 9.1 loudspeaker setup, and a Binaural Virtual Auro 3-D 9.1 setup over headphones. Each mixing session had time limitations. The stereo mixing sessions were allotted an hour and half as these were seen as the least demanding and most familiar mixes to complete. The Auro-3D 9.1 loudspeaker monitored mixes were allotted two hours to complete and the binaural Auro-3D 9.1 headphone monitored mixes were allotted two and a half hours to complete. The binaural Auro-3D 9.1 included a brief training session during the allotted time.

3.2.1 Mixing Engineers

Three professional mixing engineers were chosen to participate in the mixing phase of this thesis. A professional was defined as having at least three to five years of experience as a mixing engineer. The

engineers that participated in this thesis had a notable discography both as assistant and lead mixing engineers for many pop, R&B, and hip-hop artists. Based off of their professional mixing background, these engineers were ideal for this thesis.

A preliminary questionnaire was filled out prior to the mixing sessions in order to obtain information on the engineers' professional backgrounds. Engineers will anonymously be identified in this thesis as either Engineer X, Engineer Y, and Engineer Z. Engineer X was between 18-24 years old and Engineers Y and Z were between 25-34 years old. Engineer Y worked as a mixing engineer for 5-10 years, while Engineer X and Z said they worked as mixing engineers for 3-5 years. Only Engineer Z had prior experience with 3D audio recording and production, while the other two engineers did not. *Table A.2* in Appendix A can be referenced for more information on each mixing engineer.

3.2.2 Mixing Content

Consistency was important to maintain in order to minimize differences across all of the mixing sessions and across the three different mixing engineers. The songs chosen for the mixing sessions, both by pop artist Ariana Grande, were *Billboard Hot 100* hits "Into You" off of her album *Dangerous Woman* and "One Last Time" off of her album *My Everything* (Billboard Music, 2017). Both songs were of a similar tempo, instrumentation, and form. The similarities between the two songs helped carry similar parameters across all of the mixing sessions. By choosing a song that was already familiar to the general population, listeners would be less focused on whether they liked the song or artist and could therefore concentrate more on the overall listening experience.

The stems acquired from these two songs already included processing printed onto the tracks such as reverbs and delays. Premixed stems were used in order to maintain the focus of the listening surveys on differences between mixing methods and not mixing styles. If it were the case that the mixing sessions used raw tracks without any sort of processing, listeners would be more inclined to judge the better mix rather than the overall listening experience.

While mixing engineers were provided with preprocessed stems, one can challenge the purpose of having them further work on the songs. However, each engineer found different approaches, challenges, and workflows towards working with the stems. It was a completely new experience and mixing environment when working within the Auro 9.1 environment both over loudspeakers and binaurally over headphones. It turned out to be advantageous to have the mixing engineers focus on spatialness and the quality of sound rather than on first mixing steps like equalization and compression. Additionally, less of their personal time was required when using stems that were already preprocessed. Even though the engineers mixed the same content, under the same environments, each mix came out different enough for

listeners to be able to make decisions on preference, timbre, and spatial qualities of each mix.

3.2.3 Digital Audio Workstation Selection

ProTools was selected as the Digital Audio Workstation (DAW) because ProTools is the industry standard. Any valuable outcomes would therefore translate smoothly to the professional mixing world. Additionally, many binaural rendering tools already existed for ProTools, making this DAW ideal for the purposes of this thesis. Engineers were not given the option to use any other DAW in order to keep consistency across all of the mixing sessions, however, this was not an issue as each mixing engineer that participated in this thesis identified ProTools as their primary DAW and considered themselves proficient in using the software.

Each mixing engineer was allowed to use any three digital plug-ins from the standard ProTools HD plug-in bundle. Although mixing engineers were limited to using only three plug-ins, they could use the same three plug-ins any number of times within one session. While limiting plug-ins seemed like stripping away a carpenter's tools, the plug-in limitation helped control the goal of the listener evaluations. The focus of the listening surveys was to evaluate mixes based on the mixing procedures that were implemented and not on mixing techniques such as the applications of equalizers, dynamics, and other effects. Because the stems were already processed, there wasn't a necessity for a large application of plug-ins. At the end, the mixing tools that were available to the mixing engineers through ProTools included level balancing, panning, automation, access to 3 digital plugins varying from equalization, dynamic processing, reverberation, and more.

3.2.4 Stereo Mixing Session Setup

The stereo mixing sessions were the simplest and most familiar mixing environment to recreate. The stereo setup was based off of ITU standards with the left loudspeaker at -30° azimuth and the right loudspeaker at 30° azimuth from the listening center. The listening center was measured at 1.98 m, equidistant from the left and right loudspeakers. The mixing desk was placed to position the mixing engineer as close to the listening center as possible while leaving space to not impede entry or exit into the research lab. The desk was measured at 1.37 m equidistant from the left and right loudspeakers. The mixing engineers monitored through the Left and Right outputs as well as through a subwoofer positioned in the front left corner of the research lab.

Mixing engineers were also provided with Sennheiser HD 650 headphones to reference their stereo loudspeaker mix over headphones. They were only allowed to reference their mixes on headphones once

as the focus of this mixing session was on how well stereo loudspeaker mixes translated to headphones compared to binaurally monitored binaural mixes and binaurally rendered binaural mixes. Because the mixing engineers were most familiar with stereo mixing, they were allotted only one hour and a half to complete the mix. This was reasonable not only because stereo mixing was the easiest for them to complete, but also because the stems were pre-processed as mentioned earlier. In essence, very little work was required to complete the stereo mix except for level balancing and panning. As expected, each engineer spent the least amount of time working on their stereo mixes. *Tables A.3* and *A.4* can be referenced for a time table of how long each stereo mixing session lasted for each mixing engineer. The mixing engineers had very little issue with the stereo mixing sessions. Their feedback from the stereo mixing sessions is discussed in Section 3.3.3.

3.2.5 Auro-3D 9.1 Setup

The Auro-3D 9.1 loudspeaker mixing sessions for this thesis were setup according to the specifications listed in the Auro-3D technical document: Auro-3D's 9.1 Home Theater Setup (Auro Technologies N.V., 2015, Oct 8.) The setup was executed at the NYU Music and Research Laboratory using 9 Genelec 8403 Speakers and 1 Genelec 7070A subwoofer.

The Auro 3D 9.1 setup is configured so that the lower layer is 5.1 compatible. A 5.1 setup has the Left and Right speakers angled at 30 degrees from the listening center and the Left Surround and Right Surround speakers angled at 110 degrees from the listening center (International Technical Union, 2010). In addition to the 5.1 speaker setup, the height layer includes the addition of the Left Height, Right Height, Left Surround Height, and Right Surround Height speakers positioned at the same angles as their corresponding lower layer speakers. The height layer speakers are "elevated to an angle of 30º and tilted in such a way that the acoustical center is aimed at the listener's head when standing" (International Technical Union, 2010). These measurements were checked before each mixing session with respect to the listening center.

An image of the Auro-3D 9.1 setup that was used for this thesis is shown in *Image 3.1* at the top of the next page. This same image was provided to the mixing engineers to better illustrate the Auro 9.1 mixing environment. There was no training session for the loudspeaker monitored Auro-3D 9.1 mixing sessions, however, each engineer was allotted two full hours to complete their Auro-3D 9.1 mixes.

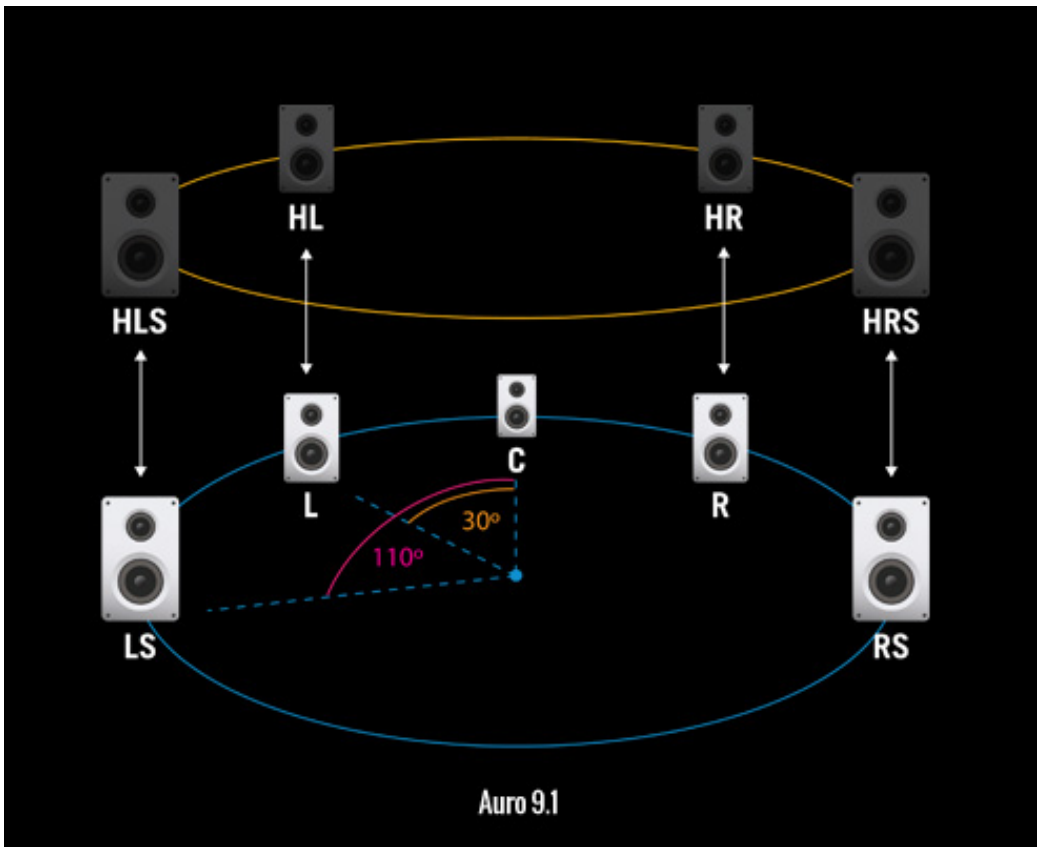


Image 3.1 **The Auro-3D 9.1 Setup (Auro Technologies, 2017).**
This image shows the main five channel layer along with the quad height layer, making up the 9.1 Auro-3D setup.

Pink noise was used to calibrate each speaker to 70 dBC. Although 85 dBC is the standard for which surround movies are mixed at according to the Society of Motion Picture & Television Engineers, 85 dBC has been regarded as "excessively high by consumer listeners" (Rumsey, 2001). 70 dBC was therefore set as the reference level for the loudspeakers. 70 dBC was not excessively loud for the small environment in the research lab. By setting a reference calibration, it was ensured that there was a unity reference level at the start of all the mixing sessions and that no one speaker was louder than another. Additionally, as stated by the Dolby Surround Mixing Manual, "When mixing music, set the 0 dB reference levels at the same SPL in each channel. Some engineers like to mix more loudly than others, so as long as all channels are calibrated at the same level, the overall volume setting is not crucial" (Dolby Laboratories Inc., 1998, 2-13). Since the focus of this thesis was music mixing, a reference level was set so that all channels were at the same level, however, the mixing engineers still had the option to increase or decrease their monitoring levels within the DAW.

At the end of each Auro-3D 9.1 mixing session, a print of the all the channels was completed so that there was a copy of each channel's mix. These printed tracks were later binaurally rendered in ProTools

using the G'Audio binaural processing detailed in the following section. The binaurally processed Auro-3D 9.1 mixes were used in the listening surveys and were compared against the stereo mixes and the binaurally monitored mixes.

3.2.6 Binaural Processing

One purpose of this thesis was to test if binaurally rendered mixes were preferred more when monitored binaurally rather than when binaurally rendered after the mixing process. In addition to evaluating binaural listening preferences, it was important to include easily adaptable immersive audio content. Therefore, the Auro-3D 9.1 configuration was integrated in order to have a well-defined guideline for the immersive audio environment that would be recreated over headphones. Because the objectives of this thesis were particular, the plug-in that was used to achieve binaural processing had to meet specific requirements. The plug in requirements at minimum included:

- ✓ Can be used in Pro-Tools as a digital plug-in
- ✓ Can achieve real-time binaural processing
- ✓ Can define azimuth and elevation levels
- ✓ Can have more than nine virtual sources
- ✓ Can export binaurally rendered content
- ✓ Can automate virtual sound sources

Many efforts were made to use the same tools that the leading immersive audio companies use to achieve immersive audio mixes. Unfortunately, many of these plug-ins had different limitations that didn't meet the needs of this thesis. Some of the immersive audio plug-ins investigated included Waves NX, TwoBigEars, and the Auro-3D Engine.

Waves NX, a recently developed binaural processing plug-in, achieves real-time binaural rendering and headtracking. This tool best emulates mixing environments such as surround sound, however, it is limited up to 7.1 speaker reproduction. Additionally, the Waves NX plug-in does not feature the option to include elevation in its binaural rendering. TwoBigEars, a free plug-in available through Facebook360, allows for binaural and ambisonic processing. While the plug-in does allow for real-time binaural processing, it has a limit of up to 8 instances of the plug-in, making it just short of meeting the processing demands of this thesis. The Auro 3D-Engine, a plug-in made available by Auro-3D, allows for 3D upmixing and panning of its audio sources. While this plug-in provides many great features for mixing 3D content, it doesn't allow for binaural processing. Auro-3D did have an Auro-3D Headphones plug-in, but it was no longer available when the thesis work was being completed.

While the above mentioned plug-ins each showed promise and interest in bringing binaural and 3D mixing tools to audio professionals, they were not used for this thesis. However, future work for this thesis would include adapting it to other binaural renders to see how results differ across different renderers. The plug-in that met the requirements for the mixing sessions specific to this thesis was G'Audio Works by G'Audio Lab.

3.2.7 G'Audio Works

G'Audio Works was recommended by colleagues who were also working with binaural content. G'Audio is a spatial audio company that released a free DAW binaural rendering plug-in in January of 2017. **"**G'Audio Works is the Digital Audio Workstation (DAW) plug-in that is available at no cost to anyone seeking to create realistic audio for cinematic VR and 360° video" (G'Audio LAB, Inc., 2017). Their plug-in, although new, met the many requirements of the binaural processing demands that this thesis required.

G'Audio Works was compatible with Mac OS X and Pro Tools 12 HD (G'Audio LAB, Inc., 2017). As advertised on their website, G'Audio Works allows for: "Simple placement of 3D sound in a virtual space: Works is easily added to Pro Tools as an AAX plug-in. There is no need to buy or learn a new system. Simply add Works at the end of the track's processing chain, and accurate spatial audio can be achieved" (G'Audio Lab, Inc., 2017b). Their easy workflow was proven to be true when setting up the binaural mixing sessions in Pro Tools. All that was required was to add the G'Audio Works plug-in on the desired track.

3.2.8 G'Audio Works Workflow

The G'Audio Works 1.0.0g User Manual released on January 23, 2017 was an essential resource that helped outline the Pro Tools workflow for the binaural mixing sessions. G'Audio Works establishes real-time binaural rendering by using two plug-ins that are part of the G'Audio Works suite: the G'Audio Slave plug-in and the G'Audio Master plug-in. In any Pro Tools session, there could only be one instance of a G'Audio Master plug-in. At the same, there could be multiple instances of G'Audio Slave plug-ins. Both plug-ins are inserted at the end of the processing chain on any Pro Tools track (G'Audio Lab, Inc., 2017a).

The G'Audio Slave plug-ins were set up as recommended by the manual. The G'Audio Slave plug-ins were inserted on the tracks that were to be binaurally monitored/rendered. In the case of these mixing sessions, the G'Audio Slaves were inserted on the 9.1 Aux tracks that simulated the Auro-3D 9.1 loudspeakers. Rather than setting the output of the audio tracks directly to an output, audio was sent through

buses to the aux tracks. These aux tracks, which had the G'Audio Slave Inserts, were then bused to a Binaural Stereo Master where the G'Audio Master plug-in was inserted.

The G'Audio Master plug-in is where the virtual sources were set with specific azimuth and elevation settings. There were ten virtual sources that made up the binaural Auro-3D 9.1 listening and mixing environment. Each virtual speaker was set to replicate the Auro-3D 9.1 environment. For example, the Left virtual speaker was set at -30° azimuth, 0° elevation and the Left Surround Height Speaker was set at -110° azimuth, 30° elevation. The virtual speakers were locked so that the engineers wouldn't change those settings.

The G'Audio Master plug-in also had the option of setting a virtual distance for each sound source. It was important to use a setting that sounded best by reducing internalization and enhance sound source externalization. During a trial period where time was spent studying the plug-in, it was felt that a setting of one meter internalized sounds while a setting of ten meters over attenuated sound source levels. After speaking with representatives at G'Audio Lab, it was stated that version 1.0.0 of G'Audio Works (which was used for these mixing sessions) did not yet incorporate a direct-to-reverb ratio; however, they stated that the distance parameter did do more than attenuate sound source levels. The method, however, was not disclosed. With this knowledge, it was decided to use five meters as the distance parameter.

3.2.9 Binaural Virtual Auro-3D 9.1 Mixing Sessions

Because the binaural mixing environment was the most unfamiliar setting for the mixing engineers, a training session was provided for the binaural mixing sessions. The binaural mixing sessions also had the longest allotted time, giving the mixing engineers two and half hours to complete their binaural mixes. The two and half hours included the training session.

The training sessions were provided to give the engineers background on the G'Audio plug-in as well as how to use it within ProTools. The training session included a self-made five-minute tutorial video that was made for the purposes of this thesis. The tutorial video helped keep the background information provided to all of the mixing engineers consistent. The topics that the tutorial video covered included how to send audio to the virtual speakers, how to add G'Audio Slave tracks, how to set the parameters of the virtual sound sources, and how to automate the parameters within the G'Audio Master plug-in. After the training session, engineers were given the option to ask any further questions on how to use the G'Audio plug-in. Questions lasted between five and ten minutes. On average, training sessions lasted between fifteen and twenty minutes, leaving the engineers about 2 hours and 15 minutes to work on their binaural mixes. All binaural mixes were monitored using Sennheiser HD 650 headphones. At the end

of each binaural mixing session, a print of the mix was completed which provided a stereo audio file of the binaural mix.

3.3 Mixing Session Results

Unfamiliar environments paired with foreign techniques resulted in valuable feedback from the perspective of the mixing engineer when trying to incorporate binaural and Auro-3D mixing processes to their everyday studio workflow. This feedback was obtained both through conversation and through a post-mixing session survey that was completed at the end of every mix.

3.3.1 Analysis of Mixing Approaches

Although the post-mixing session survey did not ask about workflow approaches, the mixing engineers provided comments through conversation on their approach to each mix. The stereo mix was standard for each engineer. Nothing was different than their usual workflow with the exception of the plug-in limitation and the environment. Only one engineer wanted to reference their headphones more than allowed. Aside from those limitations, the stereo mix was what the engineers were most familiar with.

The Auro-3D 9.1 loudspeaker sessions were approached differently. Engineer X and Z treated the 9.1 reproduction method as a large space that had to be filled. Engineer X found each channel in the 9.1 system as a variable that had to be treated independently in order to bring together the full 9.1 image. Engineer Z thought of the space as a puzzle that had to be worked through. Engineer Y treated the 9.1 environment as an extension of stereo pairs, i.e., (Left + Right, Left Surround + Right Surround, Left Height + Right Height, etc.). The reasoning behind this approach was because it was easy for them to translate their usual stereo mixing environment into mini-sets of stereo pairs that made up most of the 9.1 reproduction system.

The binaural mixing sessions were also approached differently amongst all of the engineers. Engineer X tried to simply recreate their Auro-3D 9.1 loudspeaker mix over headphones. Engineer Y engineer tried to treat it as a “full stereo mix” while Engineer Z tried to make their binaural mix a “more interesting stereo mix”. The differences amongst the binaural mixes is discussed in Chapter 4.

3.3.2 Post-Mixing Session Survey

The goal of the post-mixing session survey was to gather information from professional mixing engineers in regards to their opinions on the mixing session environment, workflow, and sound quality. The survey

was completed at the end of each mixing session in order to acquire the information. The format of the survey was seven "Yes" or "No" questions. If the questions were answered with a "Yes", then they were asked to explain why in a few short sentences. The questions from the survey, also shown in Appendix A, were:

1. Was this a difficult mixing session?
2. Was your workflow any different?
3. Was timbre affected by your mixing procedure
4. Was naturalness affected by your mixing procedure?
5. Did you have any difficulty with externalization? *
6. Was the quality of sound affected by the mixing procedure?
7. Was the diffuseness of sources affected by the mixing procedure? *

**Denotes questions that applied only to the binaural mixing sessions.*

Key words were defined at the top of the mixing survey in case the mixing engineer was unfamiliar with any of the terms from the questions. The words were defined based on Francis Rumsey's 2002 article "Spatial Quality Evaluation for Reproduced Sound: Terminology, Meaning, and a Scene Based Paradigm." The defined terms were:

Spatial Attributes - related to the impression of width, depth, and elevation
Externalization - related to the localization of sound outside of the head
Timbral Qualities - related to the spectral characteristics of a sound source

The mixing engineers were familiar with the terms and showed no difficulty understanding the questions in regards to their mixing session experience. Comments from the survey are discussed in the next section.

3.3.3 Stereo Mixing Sessions

Across the board, all three mixing engineers found the stereo mixing sessions as the easiest to complete. This was predictable as stereo mixing sessions were the most familiar mixing environment. Engineer Y commented that the stereo mix was "the easiest mix of them all." While each engineer had an hour and half to complete the mixing sessions, one engineer spent the full allotted time to complete the stereo mixes while the others finished their mix in half of the allotted time. All three engineers felt that the mixes were almost finished because the stems that were used were already processed, therefore, they felt that there was little work to be done to complete the mix. Engineer Z commented that "The mix is so good; I don't know what to do to make it better." While it was questionable to use pre-processed stems for the mixing sessions, it

proved valuable in the long-run because the listening test participants were able to better judge the listening experience rather than the processing on the mixes.

3.3.4 Auro-3D 9.1 Mixing Sessions

Two out of the three mixing engineers, Engineer X and Y, had no experience with mixing in 9.1. Engineer Z had experience with 9.1 mixing, but still found the task challenging. In general, all three engineers had difficulty with what seemed like an endless amount of possibilities of mixing with nine main channels. A common difficulty was achieving a spatial balance with all nine channels. For example, Engineer Y commented: “The balance and space of the mix is harder to judge then a normal stereo mix. My ears would need more time to adjust.” Even Engineer Z, who was familiar with 9.1 mixing commented that “I feel like 9.1 is like a puzzle. Sometimes I don't know which channel I want to use.” Because of the unfamiliarity with the 9.1 mixing environment, all engineers commented that their workflow was slower than usual because they spent a lot of time figuring out the space. That being said, there was interest from all of them to become more experienced with the Auro-3D environment.

3.3.5 Binaural Auro-3D 9.1 Mixing Sessions

All three engineers found the first round of binaural mixing as the most difficult mixing session for a number of reasons. First, the concept of binaurally monitored mixing was new to all of the engineers. This was expected as mixing binaurally was one of the tasks being investigated in this thesis. The added challenge of simulating the Auro-3D 9.1environment was the hardest concept for the mixing engineers to achieve. Engineers X and Y completed the Auro-3D 9.1 loudspeaker mixing session before the binaural mixing session, therefore, switching from Auro-3D 9.1 over loudspeakers to a binaural simulation of that experience was not an easy task. Engineer X “felt as if there was less space in the mix with headphone[s]” when switching from the loudspeaker to the binaural environment. The goal in simulating an Auro-3D 9.1 mixing environment was to see if mixes done in the increasingly common 9.1 mixing environment could be easily translated to a binaural environment through simple binaurally rendering without the necessity of binaural monitoring. While that was being tested through listener evaluations, the mixing engineers also offered valuable feedback on the task.

One common and important piece of feedback was that the spectral content of the mix was affected through binaural monitoring. This was expected as one of the most common setbacks of binaural rendering is spectral coloration. Binaural renderers, like the G’Audio plug-in, apply generic HRTFs to achieve the binaural synthesis. These generic HRTFs are heard differently by each listener because they are not

individualized HRTFs. While it is a goal for many of the existing binaural renderers to not affect the timbre of the content being processed, it is a difficult objective to achieve and it requires a lot of research and testing that is currently still being perfected by many audio technology companies. Future work past this thesis could test different binaural renderers to learn listener preferences among a few different binaural renderers. For this thesis, it became known that spectral content was affected as heard by the mixing engineers, however, feedback from the general public was still sought out through the online listening surveys that were conducted.

Another piece of feedback from the binaural mixing sessions was in regards to the headphone monitoring. There were split reactions about headphone monitoring from the mixing engineers. Engineer Z felt that headphone monitoring was always noticeable because of the fact that they were wearing headphones: "I feel the distance between each sound source but everything is still in my head." However, Engineer Y felt that the binaural mixing session was "far easier than the [Auro-3D 9.1] mix done over monitors". What was interesting, however, was that Engineer X and Y felt that their second binaural mix was the best mix they completed out of all of their mixing session. This was the complete opposite reaction after their first binaural mixing session where the same engineers commented that it was their worst mix of the day. With this knowledge, it could be said that with practice, binaural mixing could be something that could be mastered similar to how stereo mixing becomes a skill after a few years of practice. The effects of practice were noticed in the listening survey results discussed in Chapter 4.

3. 4 Mixing Sessions Discussion

Workflow seemed to be the most important factor to the mixing engineers. For future immersive audio projects, it is important that mixing engineers aren't slowed down because of processing, session set-up or content. While this point might seem obvious, this feedback might encourage developers to gain more input from mixing engineers during the development process. With that said, the mixing sessions that were carried out through this thesis showed that mixing engineers are willing and able to adapt to new procedures as long as it can be easily integrated within their existing workflow.

4

Listening Survey and Results

4.1 Overview

An online listening survey was designed in order to obtain opinions and listener preferences on the three different mixing methods that were carried out during the mixing phase. The goal was to gather feedback from both the general public and music technologists while also obtaining information on what variables affected listener preferences. This information was sought because it could help understand the prospect of binaural music reproduction in the consumer market.

4.2 Listening Survey Design

The listening survey was created with Google Forms. Google Forms was a free and easily accessible service that allowed individuals to complete the survey while also retaining their anonymity. Playback of each mix was completed through embedded YouTube videos that were included within the Google Forms webpage.

This survey was designed so that each mix by each engineer was played three times. The survey was separated into three rounds so that each engineer's mix for each song was played once per a round. Participants could opt out of continuing onto the next round. The three round survey structure allowed for the minimum number of responses to be collected within one round, while also obtaining repetitions in order to solidify listener responses. Each round was split up into two sections so that each section focused on one of the two songs. Each section had three sets of YouTube videos with three audio clips per video. Every video always featured one stereo mix, one binaural mix, and one Auro-To-Binaural mix. The arrangement of three videos per section allowed for all 18 mixes to be played at least once per a round. A sample list of what was covered in both sections of one round is shown at the top of the next page.

Round One Section One

- Engineer X's Stereo Mix of "Into You"
- Engineer Y's Stereo Mix of "Into You"
- Engineer Z's Stereo Mix of "Into You"
- Engineer X's Auro-to-Binaural Mix of "Into You"
- Engineer Y's Auro-to-Binaural Mix of "Into You"
- Engineer Z's Auro-to-Binaural Mix of "Into You"
- Engineer X's Binaural Mix of "Into You"
- Engineer Y's Binaural Mix of "Into You"
- Engineer Z's Binaural Mix of "Into You"

Round One Section Two

- Engineer X's Stereo Mix of "One Last Time"
- Engineer Y's Stereo Mix of "One Last Time"
- Engineer Z's Stereo Mix of "One Last Time"
- Engineer X's Auro-to-Binaural Mix of "One Last Time"
- Engineer Y's Auro-to-Binaural Mix of "One Last Time"
- Engineer Z's Auro-to-Binaural Mix of "One Last Time"
- Engineer X's Binaural Mix of "One Last Time"
- Engineer Y's Binaural Mix of "One Last Time"
- Engineer Z's Binaural Mix of "One Last Time

4.2.1 Video Preparation

Each video per a section featured the same eight to ten second audio clips. The only difference was which engineer's mix was played per a video. For example, Video One of Round One-Section One could have featured Engineer X's binaural mix, Engineer Y's stereo mix, and Engineer Z's Auro-To-Binaural mix while Video Two of Round One-Section One would have featured different mixes of the same audio clip. The goal was to ensure that all of the 18 mixes were heard at least once per a round.

The audio clips were embedded within a YouTube video so that the label of "A", "B", or "C" would appear while the audio clip played. The order in which each mix was played was randomized across the YouTube videos in order to prevent adaptation. The only information the listeners had was the randomly assigned labels of "A", "B", or "C" that appeared when a new mix would play.

The brief audio clips sampled from each song were selected so that they covered spectrally different content. For example, the clips used in both sections of Round One were from the introductions of both songs because they featured a low kick along with simple female pop vocals. For contrast, the clips featured in Round Three were selected from the choruses of both songs because they were spectrally busy with synthesizers, drums, vocals, and vocal effects.

4.2.2 Audio Preparation

The raw final mixes from the three mixing engineers varied in loudness levels and had to be level matched in order to have listeners properly judge the content. This was achieved in two stages. All 18 mixes were first imported into one ProTools session and were roughly matched against each other. Afterwards, the audio clips that were used in each YouTube video were level matched against each other. The level matching was done based on perceived loudness.

It was a difficult process to match the loudness levels for some of the audio clips because some mixing engineers balanced different aspects of the stems louder than others. For example, if one mixing engineer mixed the vocals higher in the mix than another engineer, it could be perceived that one mix was louder just because the vocals were louder in the mix. Therefore, the goal in loudness matching was to ensure that there was a smooth loudness transition when mixes "A", "B" and "C" were played sequentially within one YouTube video.

Additionally, because the stereo mixes were the only audio clips that did not feature any virtual LFE information, this was later added so that the three different mixing methods could be fairly judged against each other. It was thought that a pop mix without any low-end could be less preferred by listeners when compared against the Auro-To-Binaural mix that had a virtual LFE channel. Therefore, the LFE information from the Auro-To-Binaural mixes by each engineer was minimally added to their respective stereo mixes in order to match the loudness levels of the low-end content.

4.2.3 Listening Survey Design Summary

The survey was separated into three rounds with two sections per a round. Each section correlated to one of the two songs. The audio clips were featured within YouTube videos so that they could be easily embedded within the survey. There were a total of three audio clips per a video with three videos featured per a section of the survey. The survey structure allowed for one round of the survey to sample all 18 mixes at least once. By the end of the survey, a listener would have heard a sample of each engineer's mix three times.

4.3. Listening Survey Questions

The listening survey was separated into two sections: a background information section and the listening survey. The preliminary background section helped ensure that those participating in the survey were legally allowed to. It also helped screen out anybody who identified as having bad hearing since those participants were not ideal for the listening survey. Additionally, the background section was an opportunity to categorize the survey participants in order to later analyze if there were any correlations between survey participants and the survey results.

4.3.1 Preliminary Listening Survey Questions

The survey began with ensuring that participants consented to be a part of the listening survey through a simple agreement to participate question. Furthermore, it was essential to only allow adults to take the survey. Anyone who identified as "Under 18 years of age" was not allowed to continue with the survey. While the survey was open to the general public, if anyone identified as hard of hearing, then they were also not allowed access to the survey since they were not ideal listening subjects.

The remaining background questions were related to the participant's relationship to music and their music consumption habits. Participants were asked if they were either musicians, music technologists, or anyone who has studied or worked in the music business professionally. Then, participants were asked to identify their music consumption habits. Questions on a participant's music background would later be used to identify any correlations between the results and the participant's background. Lastly, participants were asked to state what type of headphones or earbuds they were using to take the listening survey. This information would later be used to see if the headphone reproduction method affected the listening survey results. Information on the participants from the listening surveys can be referenced in Appendix B. The listening portion of the listening survey commenced after obtaining background information on the survey participants.

4.3.2 Listening Survey Questions

The goal of the questionnaire was to obtain information on listener preferences and the contributing factors to those listener preferences. Therefore, the questions asked were based on preference, spectral qualities, and spatial qualities of the audio clips. Each set of questions featured the same seven multiple choice questions and one short answer question. The questions were placed after every YouTube video so each set of questions related directly to one series of audio clips that included one stereo mix, one binaural mix, and one Auro-To-Binaural mix.

The first two questions were in reference to listener preferences. Those questions were placed at the beginning of each questionnaire subsection because it was anticipated that listeners would be able to immediately judge their most preferred and least preferred clips. Therefore, Question 1 asked listeners to choose their most preferred audio clip, selecting between clips "A", "B", or "C" and Question 2 asked the listener to choose their least preferred audio clip by selecting between clips "A, "B", or "C". The exact language used in the survey is shown below:

1. Which audio clip did you most prefer?
2. Which audio clip did you least prefer?

Questions three through seven were multiple choice questions that focused on the qualities of sound of each audio clip. The answer options to questions three through seven were always either "A", "B", or "C". Key terms and definitions were provided at the start of every section in order to clarify the technical terms used within the questions. The key terms and definitions were based on references from two sources: Francis Rumsey's article, "Spatial Quality Evaluation for Reproduced Sound: Terminology, Meaning, and a Scene Based Paradigm" and Fontana et al.'s article "Binaural for Popular Music: A Case of Study".

Rumsey's article was useful in providing the type of language that should be associated with each key term. Rumsey specified in his article that the spatial attributes "should enable meaningful and sensitive distinctions… and [that] they should enable repeatable judgments" (Rumsey, 2002). Fontana et al. was useful when creating the word bank because their research carried out similar listening questionnaires related to binaural music. For example, Fontana et al. defined the parameter of spaciousness on a scale of 1 to 5 with 1 defined as "isolated and thin" and 5 as "enveloping and large" (Fontana et al., 2007). Based upon those references, the author put together the following word bank and provided it at the beginning of every section:

Key Terms:
Natural – without spectral/timbral coloration
Colored – the addition of unwanted spectral artifacts
Spatialness- related to the impression of width, depth, and elevation
Externalization – sense of a surrounding environment outside of the head
Immersive – sense of envelopment by external sound source

The goal in providing key terms and definitions was to ensure that those who had no experience in music, audio, or music technology would be able to correctly answer the questions with just enough background information. The definitions were kept clear of extremely technical terms and focused simply on what the listener should associate with the key terms. In addition to the word bank, some of the definitions were included within each question. This was also implemented in order to make the participant's experience easier by minimizing searching and scrolling within the survey. The exact questions from the listening survey are shown at the top of the next page.

3. Which audio clip sounded the most natural (without spectral/timbral coloration)?
4. Which audio clip was the most colored (the addition of unwanted spectral artifacts)?
5. Which audio clip had the best spatial qualities (i.e. width, depth, and elevation)?
6. Which audio clip had the best externalization (sense of a surrounding environment)?
7. Which audio clip was the least immersive?

The purpose of this set of questions was to ask participants to critically listen to each audio clip. By asking which audio clip sounded the most natural, there could be information on what type of mixing method most often produced the most natural sounding audio clips. Similarly, if there were audio clips that listeners identified as the most natural, then there could also be information on which mixing methods produced the colored audio clips. Therefore, the question was asked, which audio clip was the most colored. This was an important question to ask because binaural renders are often known to color audio content. While coloration could easily be identified by audio specialists, the question was also asked to see if the general public could distinguish any spectral coloration. The remaining questions focused on the spatial qualities of the audio clips. Did listeners more often find the stereo audio clips as more spatial than the Auro-to-Binaural audio clips? Or did they in fact find the stereo clips as the least immersive audio clips when compared to the binaural and Auro-to-Binaural audio clips. The focus of questions three through seven was to gather information on the sound qualities and characteristics of the audio clip.

The last question was a short-answer question which asked participants to provide any additional comments on the material they heard. Question eight was phrased “Provide comments, if any, on what you heard”. This gave the participants the opportunity to provide any additional feedback that might not have been directly asked within the survey. Questions one through eight were repeated after every video, totaling at 54 listening related questions for the entire survey.

4.4 Listening Survey Participants

The listening survey was directed at two pools of participants: music technologists and the general public. Carrying out an online survey allowed for the general public to easily access the survey. This thesis references the “Uncontrolled Listening Survey” as the one that was carried out without any oversight by the author. This survey was shared through social media, through family and friends of the author, and through emails to professional networks that the author was associated with.

In a more controlled setting, music technologists were directly sought out through the NYU Music Technology community. Those who agreed to participate and were members of the NYU Music Technology community took the listening survey in a controlled environment. The survey was the same as the online survey that was shared with the general public, but this listening survey was categorized as the “Controlled Listening Survey”.

The variables that were controlled included environment, the equipment, and the participant pool. The controlled setting took place at the NYU Music and Audio Research Laboratory, the same room where the mixing sessions were carried out. In addition, the participants who took the listening survey at NYU all

used Sennheiser HD 650 headphones to listen to the audio content. These were the same headphones that the mixing engineers used to complete the binaural mixing sessions. Having a separate pool of music technologists participate in the listening survey in a controlled setting allowed for the possibility to investigate the survey results from different perspectives such as controlled vs. uncontrolled and music technologists vs. the general public.

4.5 Listening Survey Results

The results from the listening survey were first analyzed on a large scale without taking into account the different variables such as survey type, song, engineer, and participant background. Following the large scale analysis is another analysis that separated the uncontrolled and controlled listening surveys. The results were divided between the two songs, “Into You” and “One Last Time” and then they were separated between the different rounds of the survey. This provided an investigation based on the results pool from rounds one through three of “Into You” and a results pool from rounds one through three of “One Last Time”. A closer level of analysis was also looked at that investigated any possible correlations between participants, the headphone reproduction method and the overall results of the survey. The results section concludes with a discussion on what can be inferred from the different levels of analysis that were carried out.

4.5.1 Combined Uncontrolled and Controlled Survey Results from Both Songs

The listening survey was designed so that participants could opt out of continuing onto the following rounds (three total). While the number of participants varied per each round for both the uncontrolled and controlled listening surveys, the repetition of the same questions helped gather a large response pool. Questions were repeated three times for each audio clip and nine times per song.

For the uncontrolled listening survey, there were a total of 20 participants for round one, 16 participants for round two, and 12 participants for round three resulting in a total of 288 responses for each question when considering both songs. For the controlled listening survey, there were a total of 7 participants in round one, 6 for round two, and 5 for round three totaling in 108 responses for each question when considering both songs.

The first level of analysis that was completed were a series of histograms that counted the frequency in which survey participants selected each mixing method in response to one of the seven multiple choice questions that were asked in the listening survey. The answer options were always either “A” “B” or “C”. Each option was blindly associated with either “Stereo”, “Binaural”, or “Auro-To-Binaural”. Because this

large scale analysis considered both songs and combined the results from both the uncontrolled and controlled survey results, N = 396.

The first two survey questions asked listeners to select their most preferred and least preferred audio clips. As shown in *Figures 4.1* and *4.2*, results showed that nearly 60% of the responses were that the stereo mixes were the *most* preferred audio clips and about 50% of the responses were that the Auro-To-Binaural mixes were the *least* preferred audio clips. 30% of the responses were that the binaural mixes were the *most* preferred audio clips while 35% of the responses were that the binaural mixes were the *least* preferred audio clips. The uncertainty between the binaural mixes was investigated in more detail within Section 4.5.4.

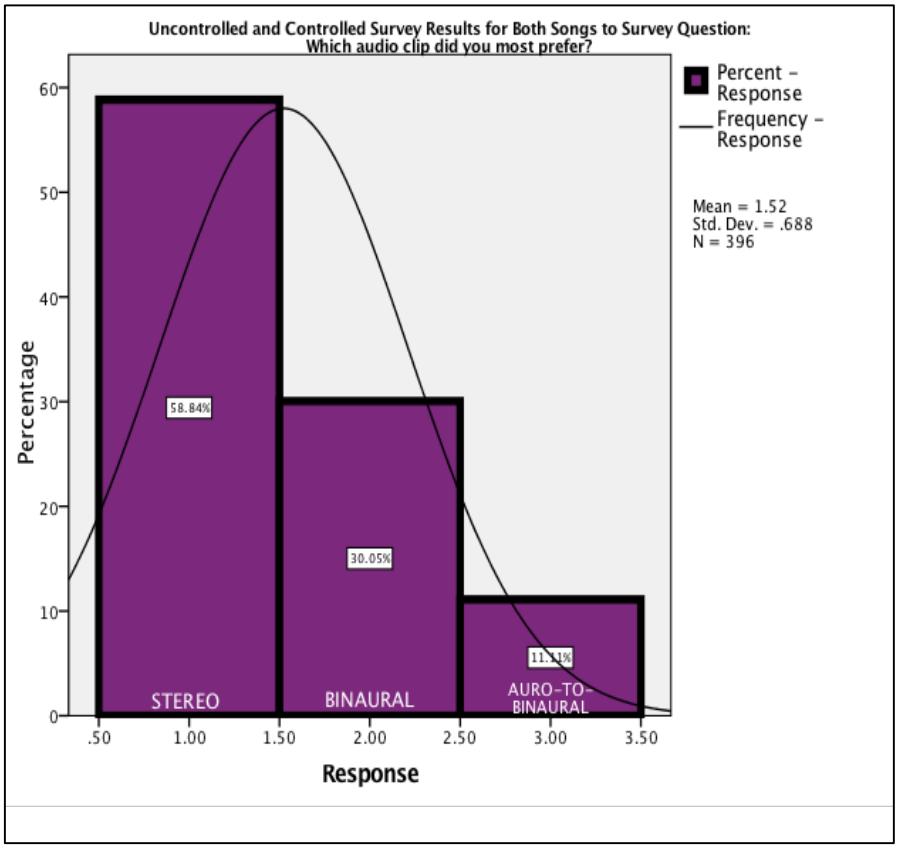


Figure 4.1 ***Histogram for Uncontrolled and Controlled Survey Results for Both Songs to Survey Question: Which audio clip did you most prefer?***
58.4% of the responses were that the stereo mixes were the most preferred audio clips.

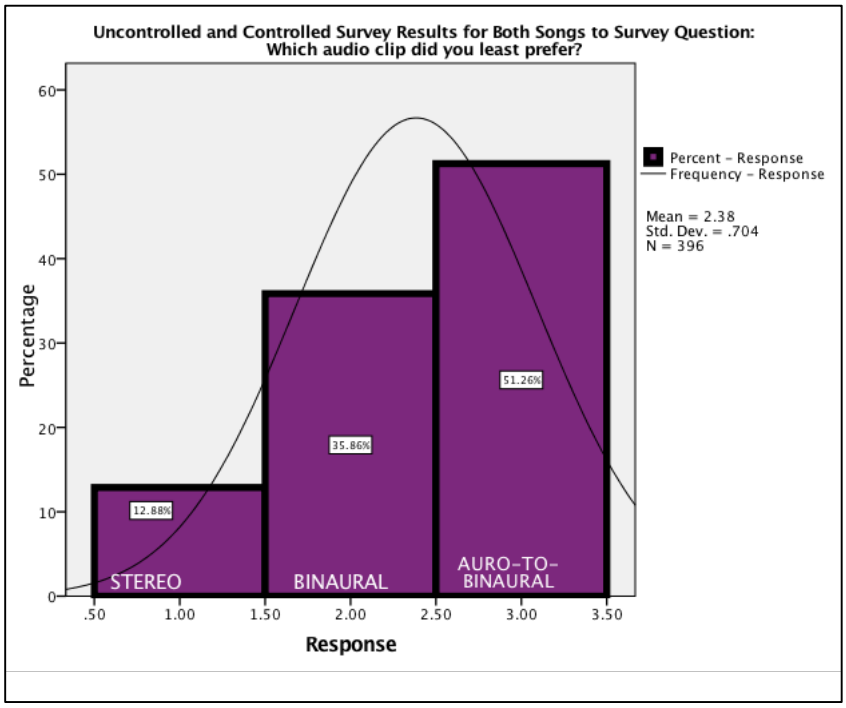


Figure 4.2 ***Histogram for Uncontrolled and Controlled Survey Results for Both Songs to Survey Question: Which audio clip did you least prefer?***
51.6% responses were that the Auro-To-Binaural mixes were the least preferred audio clips.

The next two questions on the listening survey asked listeners to evaluate the timbral qualities of the audio clips. Question three asked listeners to select the audio clip that sounded the most natural and question four asked listeners to select the audio clip that sounded the most colored. The terms "natural" and "colored" were defined within the survey in order to clarify the language for listeners who had no background in music or music technology (see Appendix B for survey example).

A large percentage of the responses, 65.66%, showed that stereo mixes were selected as the audio clip that sounded the most natural. Only 11.11% of the responses were that the Auro-To-Binaural mixes were the most natural audio clips and only a few more, 23.23% of responses were that the binaural mixes were the most natural audio clips. This is shown in a histogram in *Figure 4.3* on page 32.

The Auro-To-Binaural mixes were selected as the most colored audio clips by 58.33% of the responses. A comparative study of *Figures 4.3* and *4.4* revealed that the most unnatural audio clips were most often selected to also be the most colored audio clips and vice versa. For example, 11.11% of the responses in *Figure 4.3* show that the Auro-To-Binaural mixes were less often selected as the most natural sounding audio clips while 58.33% of the responses in *Figure 4.4* show that the Auro-To-Binaural mixes were most often selected as the audio clips that sounded the most colored. The strong opposition between the responses for questions three and four hints that listeners were able to understand what was meant by "what sounded the most natural" and "what sounded the most colored".

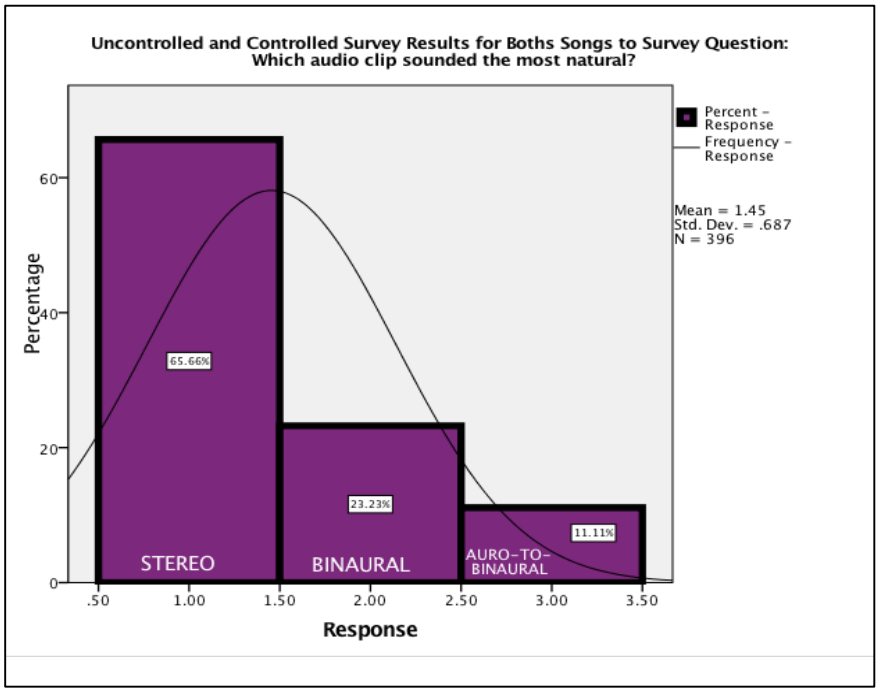


Figure 4.3 ***Histogram for Uncontrolled and Controlled Survey Results for Both Songs to Survey Question: Which audio clip sounded the most natural?***
Over half of the responses showed that stereo mixes were the most natural sounding audio clips.

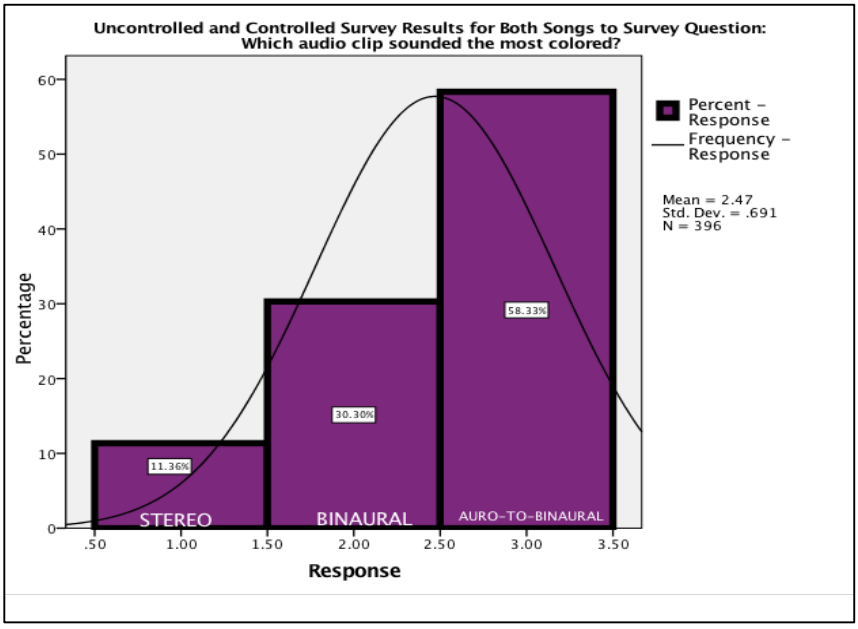


Figure 4.4 ***Histogram for Uncontrolled and Controlled Survey Results for Both Songs to Survey Question: Which audio clip sounded the most colored.***
Almost 60% of the responses were that the Auro-To-Binaural mixes were the most colored.

Questions five through seven of the listening survey asked listeners to evaluate the spatial qualities of the audio clips. The results of these questions were not as distinct as questions one through four. Rather, the responses were more distributed between the three mixing methods. Because the responses were more evenly distributed between the three mixing methods, a closer level of analysis in Section 4.5.5 looked at other contributing factors such as participant background and headphones used for the survey as possible connections to listener responses.

While about half of the listener responses in *Figure 4.*5 show that the stereo mixes had the best spatial qualities, there was less of a distinct majority. This was also true for *Figure 4.6* where Auro-To-Binaural mixes were selected almost 20% of the time in response to audio clips with the best externalization. *Figure 4.7* also shows only a 10% difference between binaural and Auro-To-Binaural mixes in reference to question: "Which audio clip was the least immersive?". Binaural mixes floated between 30 and 35% for questions five through seven, again, showing ambiguity in how binaural mixes fared amongst listener preferences. Did binaural mixes have the best spatial qualities or were they the least immersive audio clips?

The combined song and survey analysis, provided very broad information on listener preferences. While stereo mixes often fared as the most preferred, most natural, and most spatial mixes, the Auro-To-Binaural mixes were often on the other end of the spectrum as the least preferred, most colored, and least immersive audio clips. The binaural mixes wavered in between the two extremes, often around 30%, not providing very conclusive information. Therefore, the next level of analysis separated both the songs and the surveys into different categories.

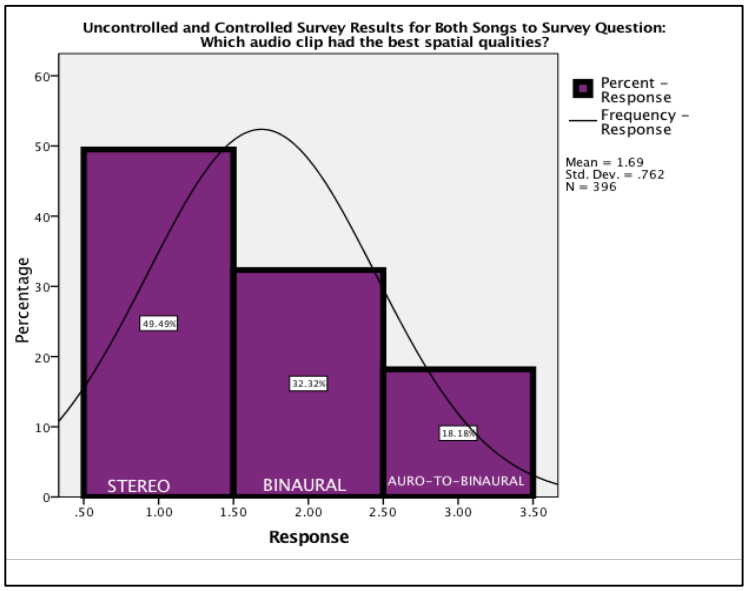


Figure 4.5 ***Histogram for Uncontrolled and Controlled Survey Results for Both Songs to Survey Question: Which audio clip had the best spatial qualities?***
Almost 50% of the responses were that the stereo mixes had the best spatial qualities.

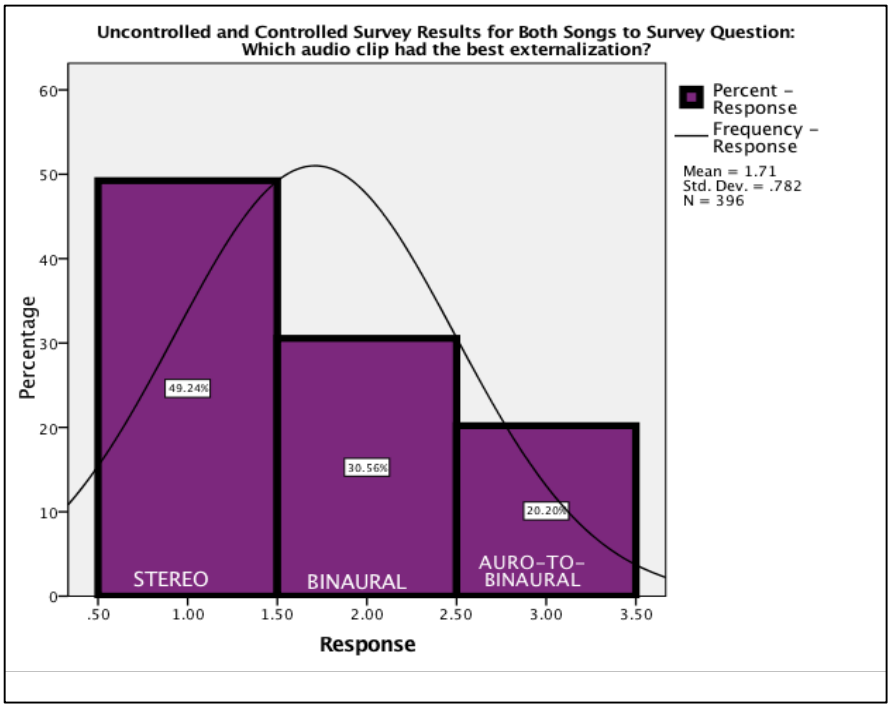


Figure 4.6 ***Histogram for Uncontrolled and Controlled Survey Results for Both Songs to Survey Question: Which audio clip had the best externalization.***
Almost 50% of the responses were that the stereo mixes had the best externalization.

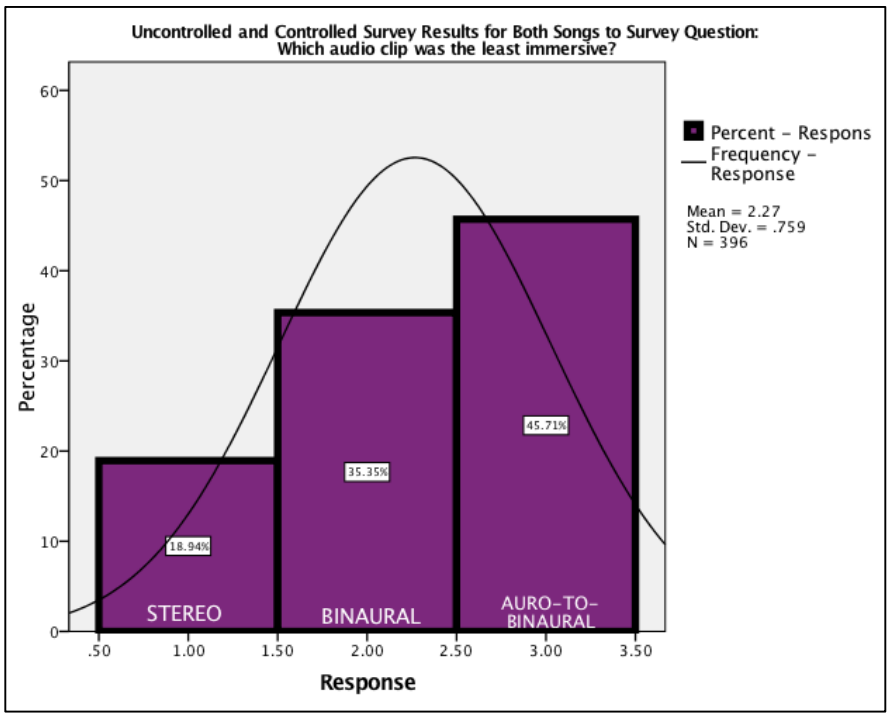


Figure 4.7 ***Histogram for Uncontrolled and Controlled Survey Results for Both Songs to Survey Question: Which audio clip was the least immersive?***
Almost 50% of the responses were that the Auro-To-Binaural mixes were the least immersive.

4.5.2 Uncontrolled vs. Controlled Survey Results for Each Song

The uncontrolled survey was distributed through email and social media platforms by providing an anonymous link to the online survey. This distribution format allowed for participants to take the online survey from anywhere and also opened the participant pool to the general public. Because the author had no control over the environment or equipment being used to take the listening survey, a controlled survey was also conducted where there was control over the equipment, the environment, and the participant pool. The survey was the same exact survey that was shared with the general public except participants identified themselves as professionals in the field of music technology. All seven participants who took the controlled survey used Sennheiser HD 650 headphones for the survey and took the survey at the NYU Music and Audio Research Laboratory.

The second level of analysis compared the uncontrolled and controlled survey responses while also separating the two songs that were used as stimuli: "Into You" and "One Last Time". Generally, the controlled survey responses mirrored the uncontrolled survey responses, but on a smaller scale. This is shown in *Figures 4.8 and 4.9* where the responses followed the same trend. For both the controlled and uncontrolled surveys, the stereo mixes were most often selected as the most preferred audio clips while the Auro-To-Binaural mixes were least often selected as the most preferred audio clips.

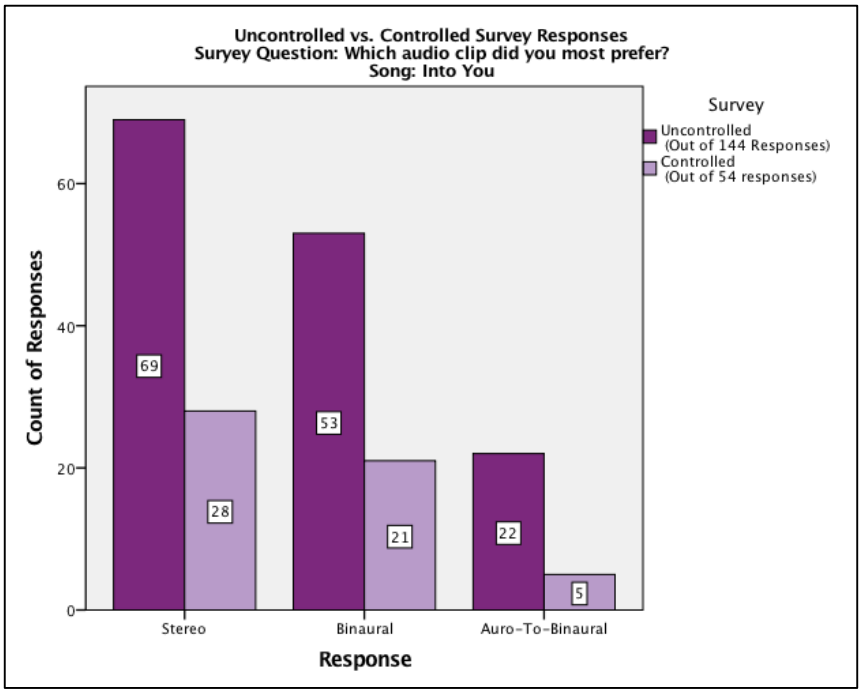


Figure 4.8 ***Bar Chart for Uncontrolled and Controlled Survey Results for "Into You" to Survey Question: Which audio clip did you most prefer?***
This bar chart shows that both uncontrolled and controlled survey participants most preferred stereo mixes.

While the responses between the uncontrolled and controlled surveys generally mirrored each other (reference Appendix D for more examples), the responses between the two songs showed dramatic differences. Fewer survey participants selected binaural and Auro-To-Binaural mixes as their most preferred audio clips for the song "One Last Time". For example, almost half as many of the uncontrolled survey participants selected the binaural mixes as their most preferred audio clip. Binaural mixes of "Into You" were selected as the most preferred audio clip 53 times (see *Figure 4.8*) as opposed to just 27 times for the song "One Last Time" (*Figure 4.9*)

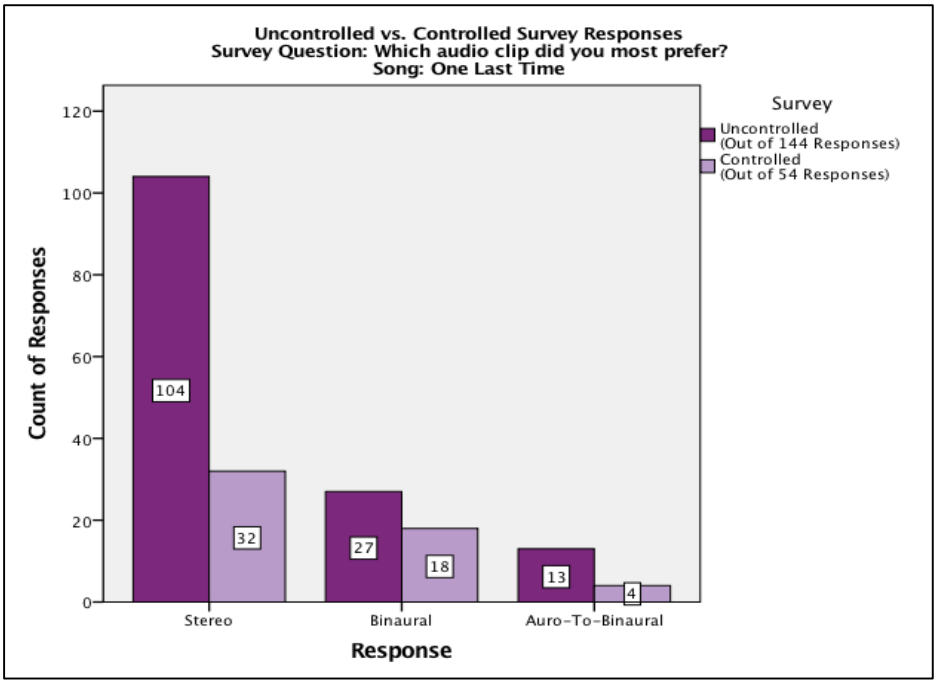
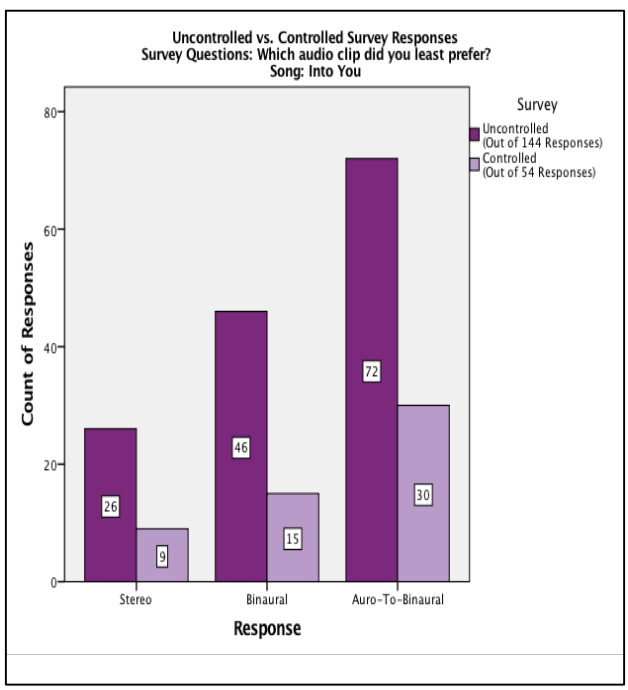


Figure 4.9 ***Bar Chart for Uncontrolled and Controlled Survey Results for "One Last Time" to Survey Question: Which audio clip did you most prefer?***
About 70% of participants selected the stereo audio clips and much fewer selected binaural and Auro-To-Binaural for both the uncontrolled and controlled surveys.

The difference between the two songs appeared most significantly for the questions that asked about listener preferences and timbral qualities of the audio clips. *Figures 4.10* and *4.11, s*hown on page 37, highlight the differences between the least preferred audio clips between each song. *Figures 4.12* and *4.13,* shown on the bottom of page 37, focus on the differences between the responses regarding the timbral qualities of each song. The binaural mixes of "One Last Time" were the least preferred audio clip by 43% of the responses as opposed to 31% for "Into You". Similarly, the stereo mixes of "One Last Time" were selected as the most natural sounding audio clip by 75% of the uncontrolled survey responses as opposed to 56% for "Into You". The difference in responses that ranged between 10 and 25% based on which song was being examined was further investigated in later sections of this analysis.



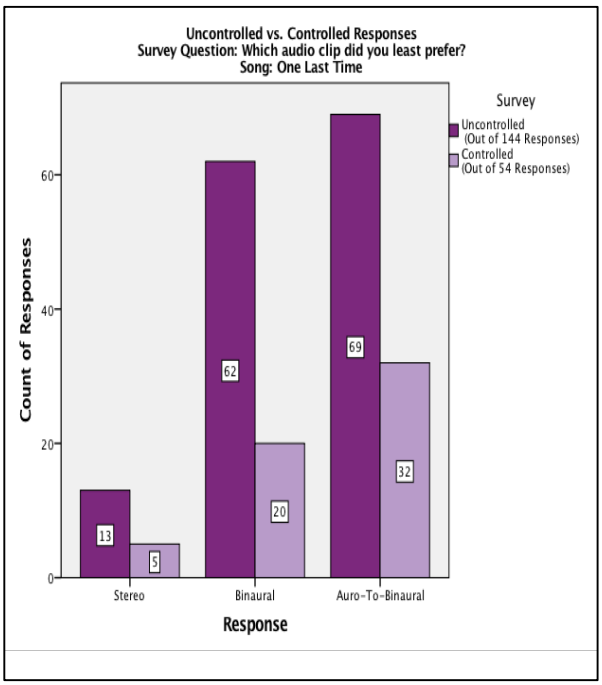
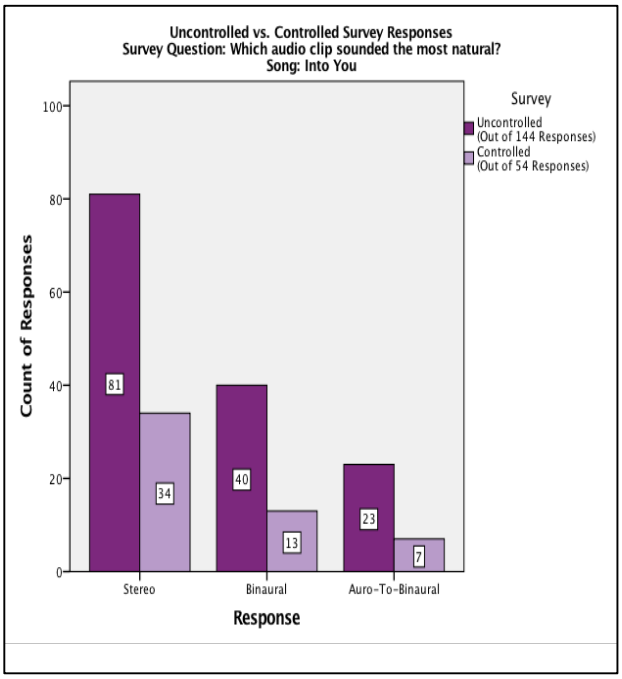


Figure 4.10 (left) and 4.11 (right) ***Bar Charts for Uncontrolled and Controlled Survey Results for "Into You" (left) and "One Last Time" to Survey Question: Which audio clip did you least prefer?***
About 20 more responses selected binaural mixes as the least preferred audio clip for the uncontrolled survey and 5 more responses for the controlled survey when asked about "One Last Time".



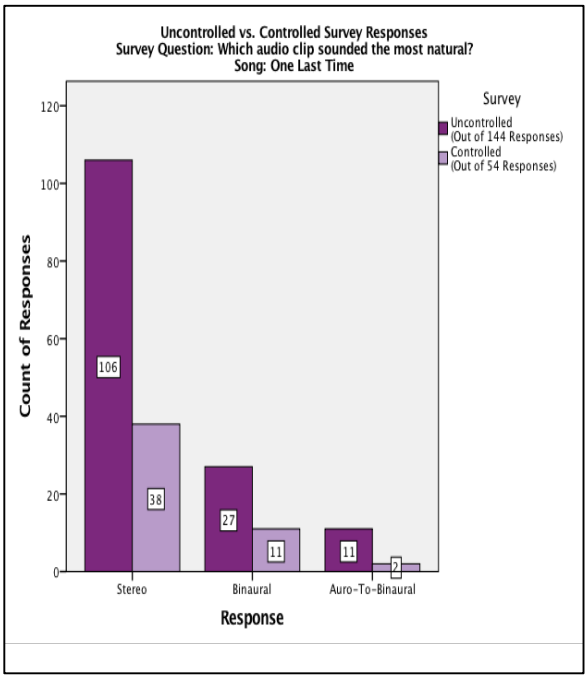


Figure 4.12 (left) and 4.13 (right) ***Bar Charts for Uncontrolled and Controlled Survey Results for "Into You" (left) and "One Last Time" to Survey Question: Which audio clip sounded the most natural?***
Fewer responses selected the binaural mixes as the most natural sounding audio clips for song "One Last Time" for both the binaural and Auro-To-Binaural mixes.

4.5.3 Uncontrolled vs. Controlled Survey Results for Each Song Compared by Rounds

Since the survey results from Section 4.5.2 showed differences between the two songs "Into You" and "One Last Time", a more detailed analysis was carried out which not only separated the songs, but also compared the survey results from rounds one through three for each survey type (uncontrolled or controlled). Because the controlled survey results generally mirrored the uncontrolled survey results but on a smaller scale, the bar charts for the controlled survey results are shown in Appendix E while some of the uncontrolled survey results are directly referenced within this section.

Figure 4.14 shows that listeners most preferred the binaural mixes of "Into You" in round one of the uncontrolled survey. Round one of "Into You" shows that 45% of the participants selected binaural mixes as their most preferred audio clip; however, the opposite was true for rounds two and three. The results are not similar when looking at the participant preferences for "One Last Time". As shown in *Figure 4.15* on page 39, 61% of the participants in round one of the uncontrolled survey selected the stereo mixes of "One Last Time" as the most preferred audio clip. The stereo mixes of "One Last Time" continued to be the most preferred audio clips for rounds two and three. The binaural and Auro-To-Binaural mixes were largely less preferred for rounds two and three, especially when compared to the uncontrolled survey results for "Into You". Nearly 30% of the participants preferred the binaural mixes in rounds two and three when asked about "Into You" while only about 8% of the participants selected binaural mixes as their most preferred audio clips in rounds two and three for "One Last Time".

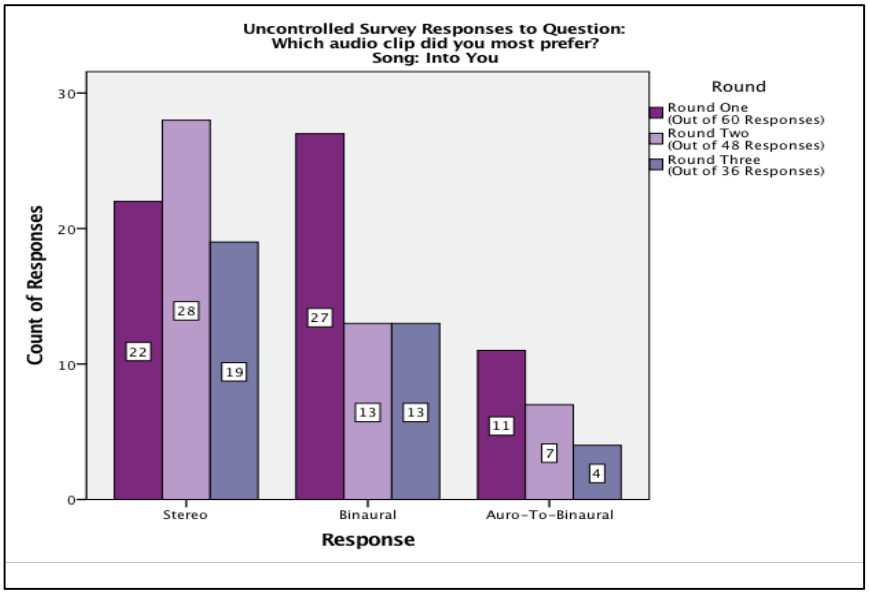


Figure 4.14 ***Bar Charts for Uncontrolled Survey Results from Rounds One –Three for "Into You" to Survey Question: Which audio clip did you most prefer?***
Binaural mixes were more preferred in round one while they were less preferred in rounds two and three.

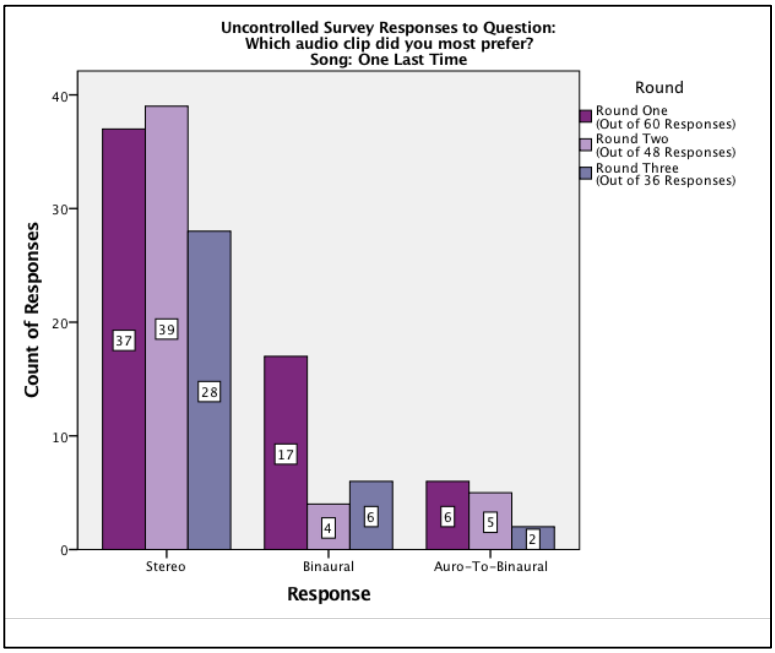


Figure 4.15 ***Bar Charts for Uncontrolled Survey Results from Rounds One –Three of "One Last Time" to Survey Question: Which audio clip did you most prefer?***
This bar chart reemphasizes the differences between responses for "One Last Time" and "Into You".

In order to prevent listener familiarity with structure, the arrangement of the rounds of "Into You" did not mirror that of "One Last Time"; however, the analysis showed that the non-stereo mixes of "One Last Time" were consistently less preferred. Similar differences between "Into You" and "One Last Time" were also apparent when examining timbral quality and immersion. For "Into You", about 20% of the listeners found the binaural mixes as the most colored audio clips for rounds two and three (see *Figure 4.16 on page 40)*. This number nearly doubles to 40% when asked the same question in regards to "One Last Time" as shown in *Figure 4.17* on page 40. The Auro-To-Binaural mixes were selected as the most colored audio clips for both songs and for both the uncontrolled and controlled surveys. An interesting note is that although the binaural mix of "Into You" was the most preferred audio clip by 45% of the participants in round one, it was also selected as the most colored audio clip by 36% of the participants in round one. This begs the question if coloration affects listener experiences.

Listener responses in terms of immersion were more scattered and showed less of a trend. For example, in round one of "One Last Time", 30% of listeners selected stereo mixes as the least immersive audio clip, but this dropped to 3% for rounds two and three as shown in *Figure 4.19* on the bottom of page 40. Another interesting point was that listeners more often selected stereo mixes as the least immersive audio clips for "Into You" showing that listeners considered the non-stereo mixes to have immersive qualities that competed with the spatial familiarities of stereo mixes.

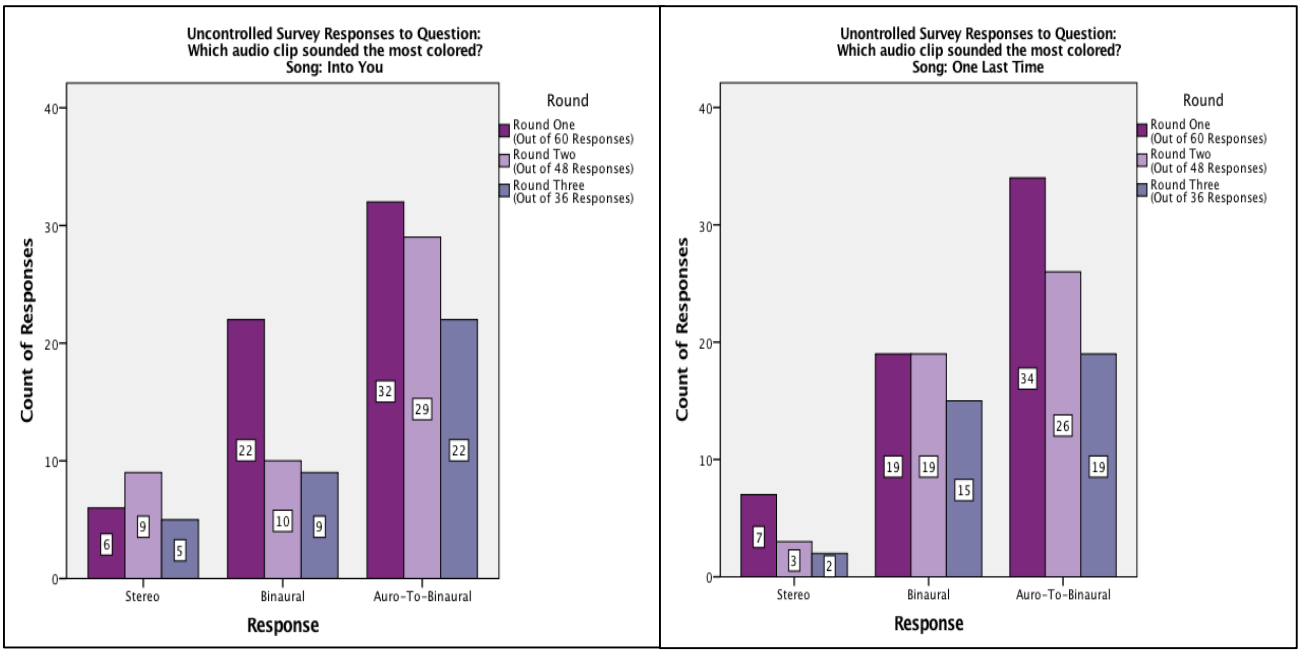


Figure 4.16 (left) ***Bar Charts for Uncontrolled Survey Results from Rounds One –Three for "Into You" to Survey Question: Which audio clip sounded the most colored?***
Auro-To-Binaural mixes were consistently selected as the most colored audio clips for rounds one through three.
Figure 4.17 (right) ***Bar Charts for Uncontrolled Survey Results from Rounds One –Three for "One Last Time" to Survey Question: Which audio clip sounded the most colored?***
Binaural mixes were most often selected as the most colored audio clips for round one through three when compared to the responses for "Into You".

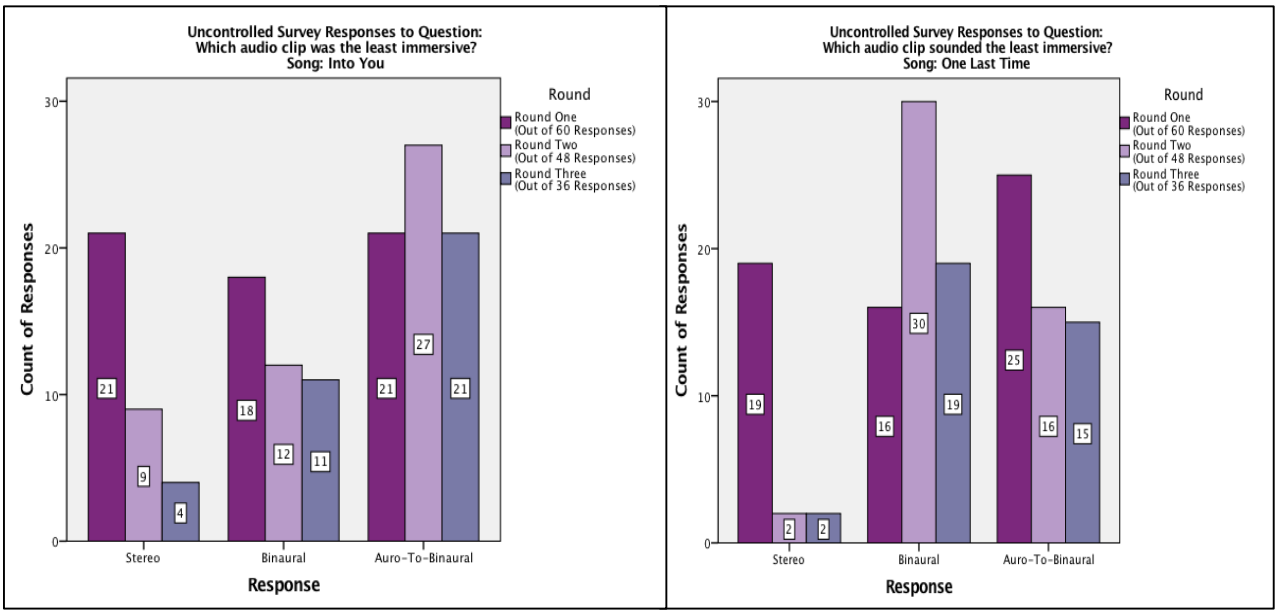


Figure 4.18 (left) ***Bar Charts for Uncontrolled Survey Results from Rounds One –Three for "Into You" to Survey Question: Which audio clip was the least immersive?***
Results were more distributed, but Auro-To-Binaural mixes held the majority for the least immersive audio clips.
Figure 4.19 (right) ***Bar Charts for Uncontrolled Survey Results from Rounds One –Three for "One Last Time" to Survey Question: Which audio clip was the least immersive?***
Results were more distributed and less clear on what listeners found to be the least immersive audio clips.

This analysis further verified that the Auro-To-Binaural mixes were generally not preferred, were the most colored, and had the least favorable spatial qualities when compared to the other mixing procedures. However, a variance between stereo and binaural mixes was still noticed across the survey results for both the uncontrolled and controlled surveys. The anomalies between each round and between each song in regards to the binaural mixes led to another level of analysis that investigated each engineer's binaural mixes.

4.5.4 Evaluation of Each Engineer's Binaural Mixes

The engineers selected to participate in the mixing process of this thesis all had different backgrounds and experiences. For example, Engineer Y was the most experienced engineer who had 5-10 years of mixing experience as opposed to the other engineers who both had 3-5 years of mixing experience; however, only Engineer Z identified as having prior experience with 3D audio recording and production. The other engineers stated that they had no prior experience with 3D audio. Lastly, Engineer X was the youngest mixing engineer of the three while Engineer Z was the oldest. A table of each engineer's history can be referenced in Appendix A.

Other differences aside from professional history were also the differences in which the mixing sessions were completed. Engineer Z first worked on "Into You" as opposed to Engineers X and Y who both worked on "Into You" on their second day of mixing. Engineer X and Y both completed the binaural mixes of "One Last Time" second while it was the fourth mix completed by Engineer Z.

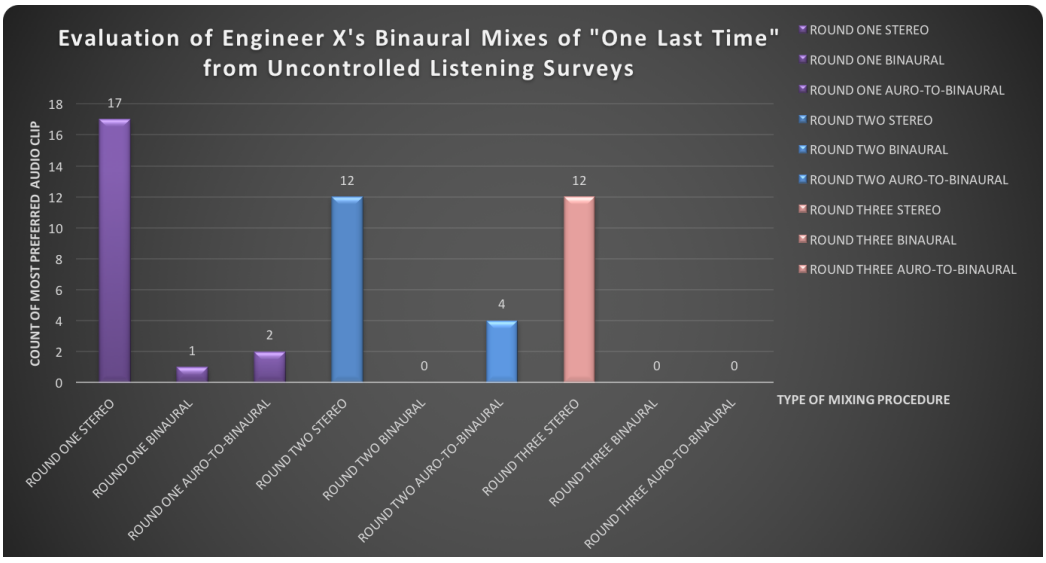


Figure 4.20 ***Bar Chart for Uncontrolled Survey Results from Rounds One –Three for Engineer X's Binaural Mixes of "One Last Time" to Survey Question: Which audio clip did you most prefer?***
Engineer X's binaural mixes were almost never preferred.

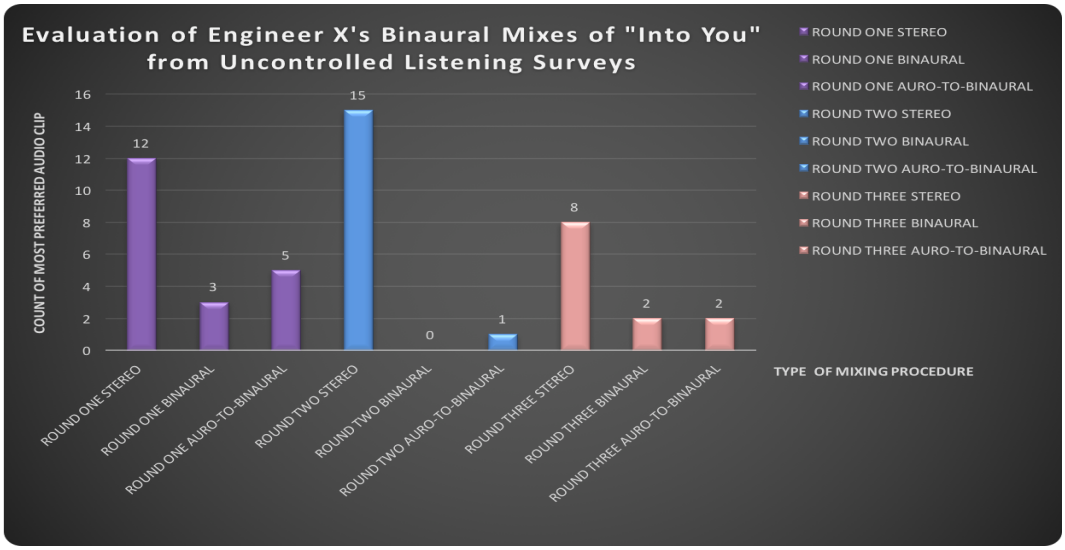


Figure 4.21 ***Bar Chart for Uncontrolled Survey Results from Rounds One –Three for Engineer X's Binaural Mixes of "One Last Time" to Survey Question: Which audio clip did you most prefer?***
Engineer X's binaural mixes were only slightly more preferred for "Into You".

Figures 4.20 and *4.21* show the uncontrolled survey evaluations of the binaural mixes completed by Engineer X. Out of the three rounds, only once was Engineer X's binaural mix of "One Last Time" selected as the most preferred audio clip. This number only increased to five for "Into You" which was completed on the second day of mixing for Engineer X. Engineer X was the youngest engineer who identified as having no prior experience with 3D audio recording and production.

Engineer Y's binaural mixes had better evaluations compared to that of Engineer X's, especially when comparing the two separate songs. For example, Engineer Y's binaural mix of "One Last Time" was selected as the most preferred audio clip only four times throughout all three rounds (see *Figure 4.22* on page 43). For "Into You", the results were quite different as shown in *Figure 4.23* on page 43. In total, Engineer Y's binaural mix of "Into You" was selected as the most preferred audio clip 17 times across the three rounds. Additionally, Engineer Y's binaural mix of "Into You" was the majority preference in round one and fared closely behind the stereo mix for round three. Round two was the first time throughout the whole survey where the Auro-To-Binaural mixes were evaluated the same as the stereo mixes. An explanation for this is that in that section of round two, the stereo mix, binaural mix, and Auro-To-Binaural mix were all completed by Engineer Y. Therefore, in that instance, listeners were not comparing mixes from different engineers, but were comparing one engineer's different mixing methods. This was even noticed by one participant of the listening survey who commented: "All three of these mixes were fairly similar. A was the plainest but C [Auro-To-Binaural] and B [binaural] were both very close sonically".

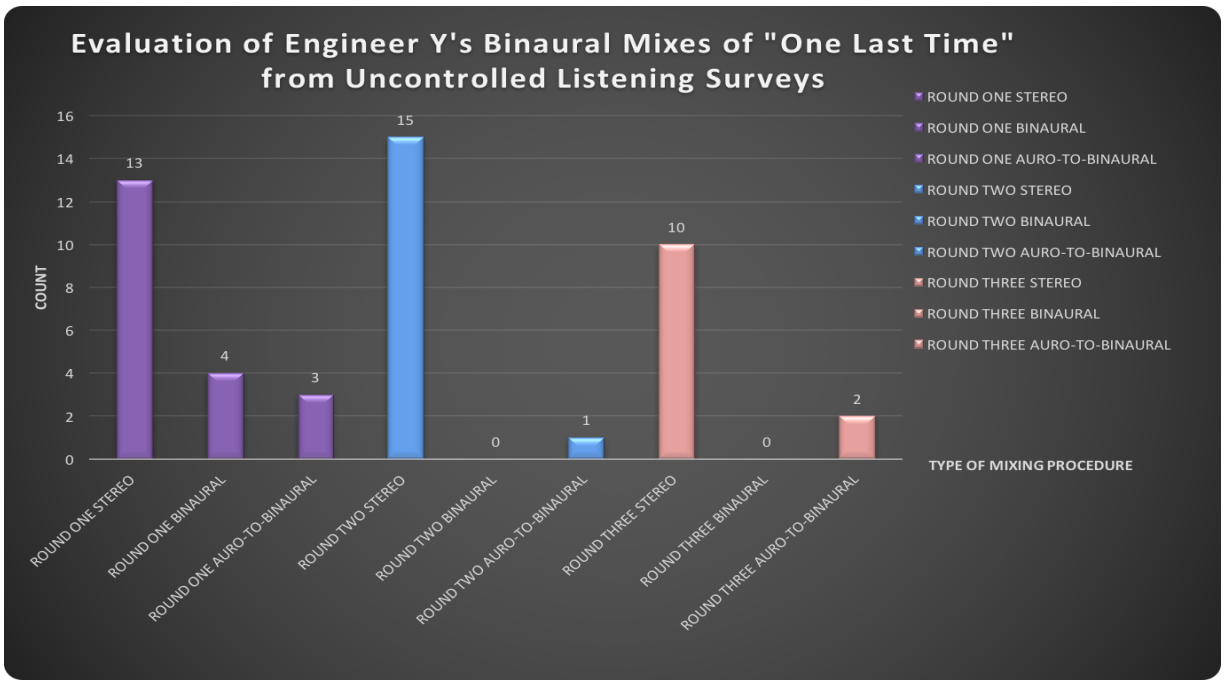


Figure 4.22 ***Bar Chart for Uncontrolled Survey Results from Rounds One –Three for Engineer Y's Binaural Mixes of "One Last Time" Survey Question: Which audio clip did you most prefer?***
Engineer Y's binaural mixes were only slightly more preferred for "Into You".

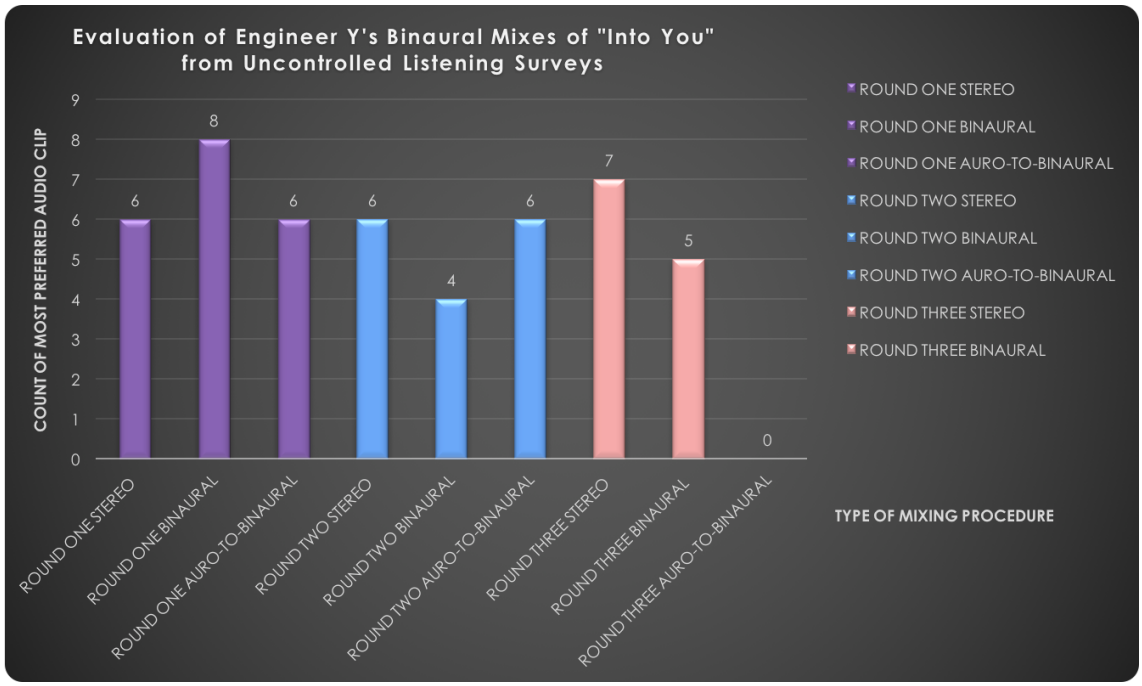


Figure 4.23 ***Bar Chart for Uncontrolled Survey Results from Rounds One –Three for Engineer Y's Binaural Mixes of "Into You" to Survey Question: Which audio clip did you most prefer?***
Engineer Y's binaural mixes had better evaluations for "Into You" as opposed to Engineer Y's binaural mixes of "One Last Time".

After evaluating *Figures 4.20 – 4.23* one could argue that the mixes for "One Last Time" were not as preferred as the mixes for "Into You" for Engineers X and Y since those mixes were completed during the first mixing sessions. Similarly, because Engineer Y was the most experienced engineer, it could be said that this engineer was better able to quickly learn and adapt to the new mixing methods presented during the binaural and Auro-To-Binaural mixing sessions.

Previous experience with 3D audio was also reflected in the engineer's overall evaluations, especially when comparing Engineer Z's binaural evaluations to Engineer X's and Y's evaluations. For example, in *Figure 4.24*, Engineer Z's binaural mixes were consistently more preferred than the other mixing procedures. Round one shows the most significant difference where 80% of the 20 participants selected Engineer Z's binaural mixes as the most preferred audio clip. This was also true in Round One of "One Last Time" where 60% of the participants selected Engineer Z's binaural mixes as the most preferred audio clip as reflected in *Figure 4.24*. Although round two had more participants select the stereo mix over Engineer Z's binaural mix, round three shows that 50% of the participants selected Engineer Z's binaural mix as the most preferred audio clip.

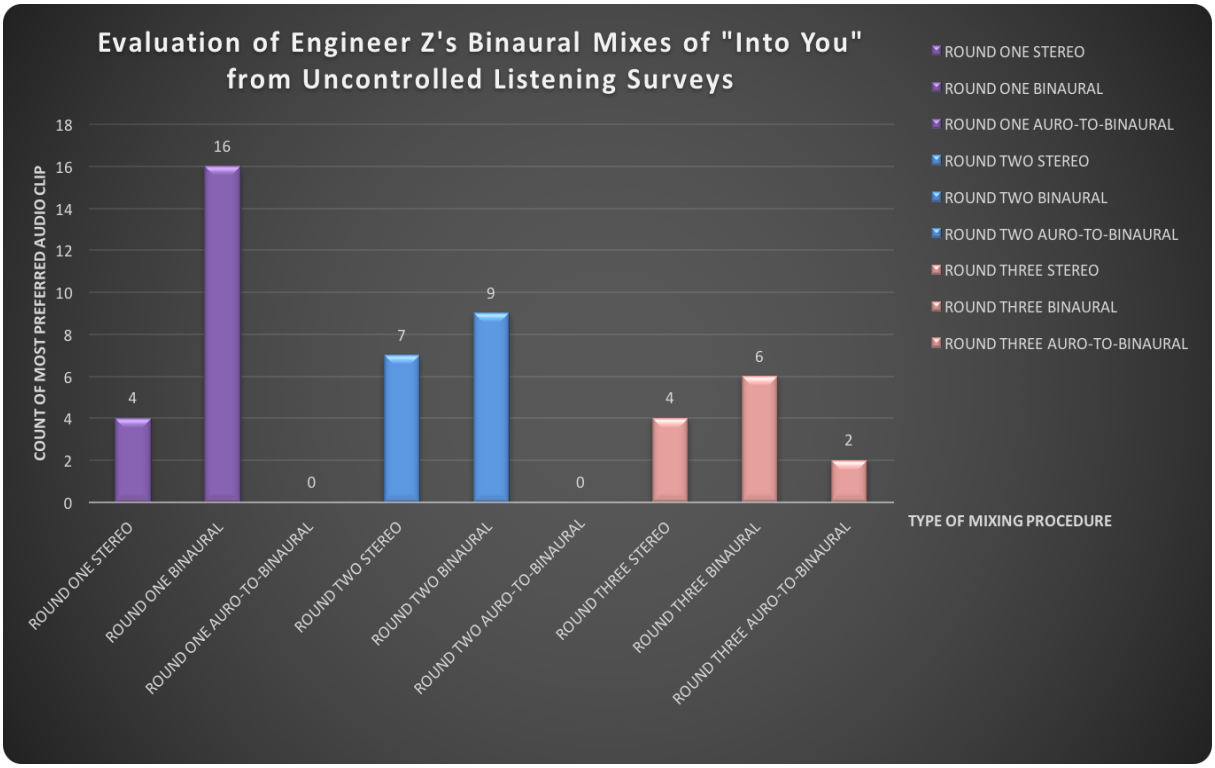


Figure 4.24 ***Bar Chart for Uncontrolled Survey Results from Rounds One –Three for Engineer Z's Binaural Mixes of "Into You" to Survey Question: Which audio clip did you most prefer?***
Engineer Z's binaural mixes of "Into You" were always more preferred than the other mixing procedures for all three rounds.

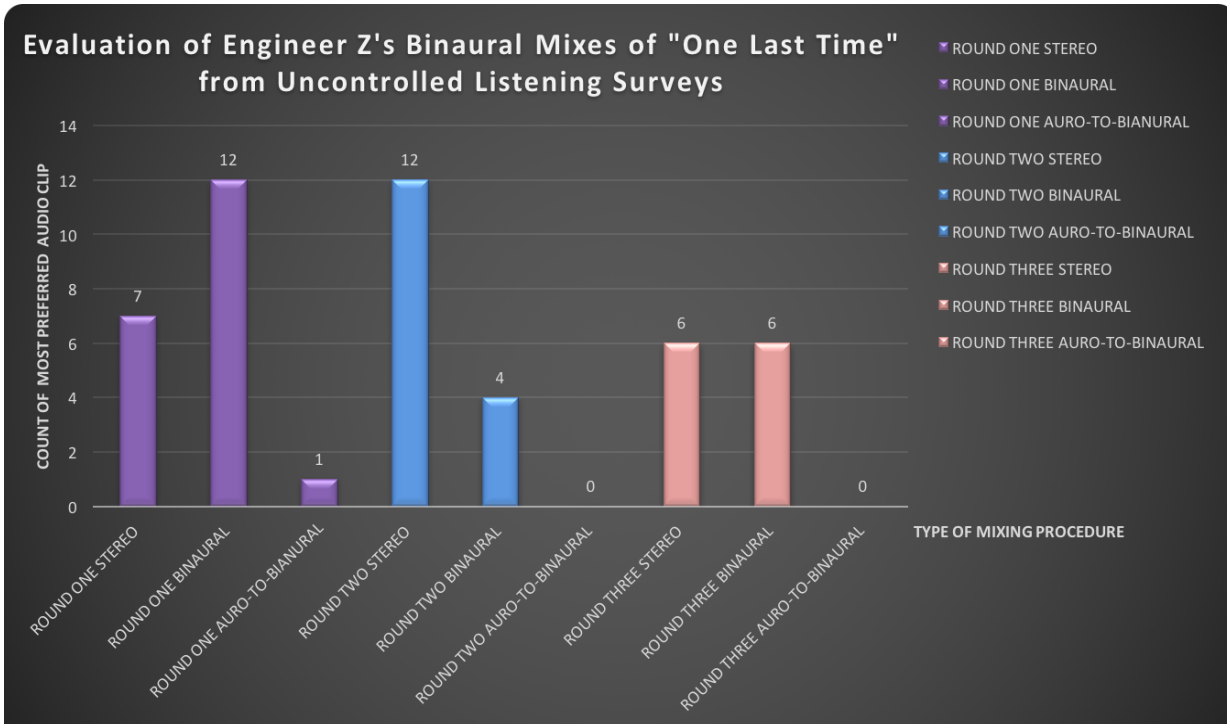


Figure 4.25 ***Bar Chart for Uncontrolled Survey Results from Rounds One –Three for Engineer Z's Binaural Mixes of "One Last Time" to Survey Question: Which audio clip did you most prefer?***
Engineer Z's binaural mixes of "One Last Time" was the majority preference for round two and was generally preferred throughout the remaining rounds.

Engineer Z's binaural mixes did not follow the same trend as Engineer X's and Y's binaural evaluations. Rather, the mixes completed on Engineer Z's first day of mixing were evaluated better than the mixes completed on the second day of mixing. It could be argued that "One Last Time" was a song that did not translate well to binaural reproduction.

It could also be argued that because Engineer Z had prior experience with 3D recording and production, the binaural mixes required less of a learning period and allowed the engineer to spend more time focused on the actual mixing. The prior experience with 3D audio is reflected in the fact that Engineer Z's binaural mixes garnered the most preferred evaluations for both songs when compared to the evaluations of the other engineers. This was also the case for the controlled listening surveys as shown in *Table E.1.5* and *E.1.6* where Engineer Z's binaural mixes were the most preferred audio clips across all rounds for "Into You" and held a high preference for rounds one and three for "One Last Time". Even some of the feedback in reference to Engineer Z's binaural mixes included, "Mix A [Engineer Z's Binaural Mix] is very good. There is minimal coloration and the spatial effects are incredibly immersive and engaging. Mix B [Engineer X's Auro-To-Binaural mix] features the most coloration and least amount of immersiveness."

The analysis completed on each engineer's binaural mixes helped clarify some of the ambiguity on how binaural mixes fared between the stereo mixes (the better evaluated audio clips) and the Auro-To-Binaural mixes (the lesser evaluated audio clips). Although on the larger scale, binaural mixes were not more preferred, the evaluations of the binaural mixes improved with the addition of experience.

Type of Mixing Procedure	**Rounds One - Three**		
	Round One	*Round Two*	*Round Three*
Stereo	20.00%	43.75%	33.33%
Binaural	80.00%	56.25%	50.00%
Auro-To-Binaural	0.00%	0.00%	16.67%

Table 4.1. ***Evaluation of Binaural Mixes of Engineer Z's "Into You" from Uncontrolled Listening Surveys***
Listeners most preferred Engineer Z's binaural mixes. Results were from the survey question: Which audio clip did you most prefer?

Type of Mixing Procedure	**Rounds One - Three**		
	Round One	*Round Two*	*Round Three*
Stereo	35.00%	75.00%	50.00%
Binaural	60.00%	25.00%	50.00%
Auro-To-Binaural	5.00%	0.00%	0.00%

Table 4.2 ***Evaluation of Binaural Mixes of Engineer Z's "One Last Time" from Uncontrolled Listening Surveys***
Listeners were divided between stereo and Engineer Z's binaural mixes as their most preferred audio clips. Results were from the survey question: Which audio clip did you most prefer?

4.5.5 Other Contributing Factors to Listener Evaluations

Two of the most significant differences between the controlled and uncontrolled listening surveys were the participant pool and the equipment. Participants who were music technologists were specifically enlisted to be participants of the controlled listening survey. Additionally, the participants of the controlled listening survey all used the same pair of Sennheiser HD 650 headphones while the uncontrolled listening survey participants were asked to use any pair of headphones or earphones that were available to them. These differences were analyzed to see if they contributed to the responses of the listening survey.

The listening survey asked participants to identify their music history background. The three largest categories of groups that resulted were (1) I am a music technologist (2) I am a musician and (3) None of the above. Therefore, an analysis based on those three categories was completed to see if they contributed to listener preferences, however, this was not the case. More often than not, each category responded with the same trend. Some examples of how responses remained consistent no matter the participant's music background are shown in *Figures 4.25* and *4.26* on page 47 and more can be referenced in Appendix F.

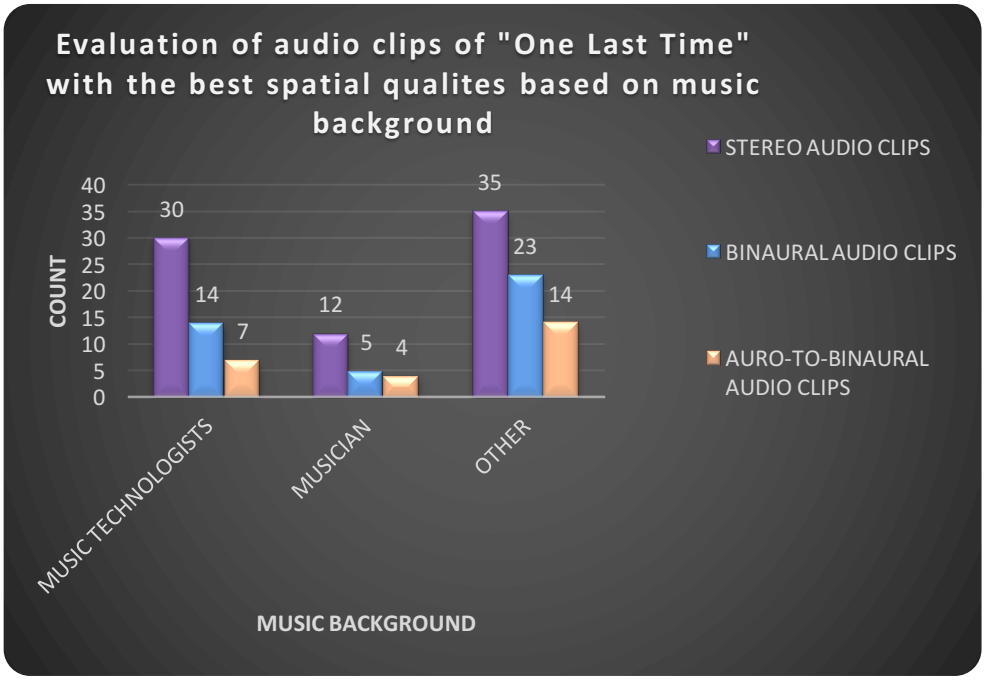


Figure 4.26 ***Bar Chart of the audio clips with the best spatial qualities of "One Last Time" based on music background of the survey participants***
Bar chart shows that all categories of music backgrounds answered the question more or less in the same fashion.

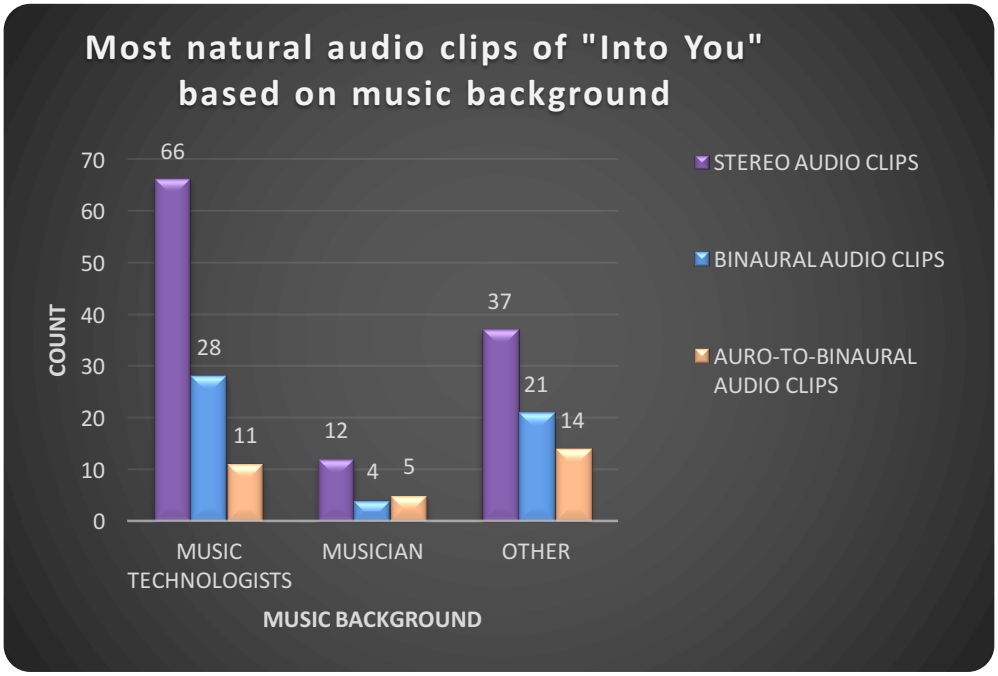


Figure 4.27 ***Bar Chart of the most natural audio clips of "Into You" based on music background of the survey participants***
Categorized results based on music background that participants answered similarly without any strange anomalies.

The most significant differences based on music background were in the comments that were provided by the different categories of test participants. For example, those who identified as having no professional relationship with music rarely provided any comments. The comments that were provided were, for example, “A's vocals sounds underneath the beat, B's vocals sound above and C's is the best of both”. That sort of feedback was more related to the mixing style rather than the sonic qualities of the mix. Those who identified as music technologists and were a part of the controlled listening surveys were much more critical in their feedback. For example, one participant stated:

> *Although I enjoyed C [Engineer Y’s binaural mix] the most for its spatial qualities, everything was more colored than when compared to A [Engineer Y’s stereo mix], thus creating a less natural experience.*

Another example of written feedback from the controlled listening survey was:

> *Mix b [Engineer Z’s binaural Mix] & c [Engineer X’s stereo mix] both had a great sense of space while Mix A [Engineer Y’s Auro-To-Binaural mix] seemed pretty boxed in and uncomfortable. Though Mix B had more low-end impact, Mix C had a lesser amount of coloration.*

While it is hard to say what affected the decision making for the participants of the uncontrolled survey, the participants of the controlled survey seemed to base their answers of preference based on the other questions of the survey. Participants of the controlled survey seemed to consider coloration and spatialness of each audio clip when basing their decisions their most preferred and least preferred mixes.

There was also no apparent trend when evaluating the headphones used for the survey. One might have expected that those participants that used in-ear headphones such as Apple AirPods would not have been able to fairly or consistently judge timbral or spatial qualities of the audio clips. Similarly, listeners who used the “controlled” survey headphones (Sennheiser HD 650 headphones) were expected to be more consistent in their responses than those listeners that used in-ear headphones, however, this was not the case. For example, *Figure 4.28* on page 48, shows that participants answered similarly no matter the headphone reproduction method used for the listening survey. Participants in general found that the stereo audio clips sounded the most natural, followed by the binaural mixes and the Auro-To-Binaural mixes in that order.

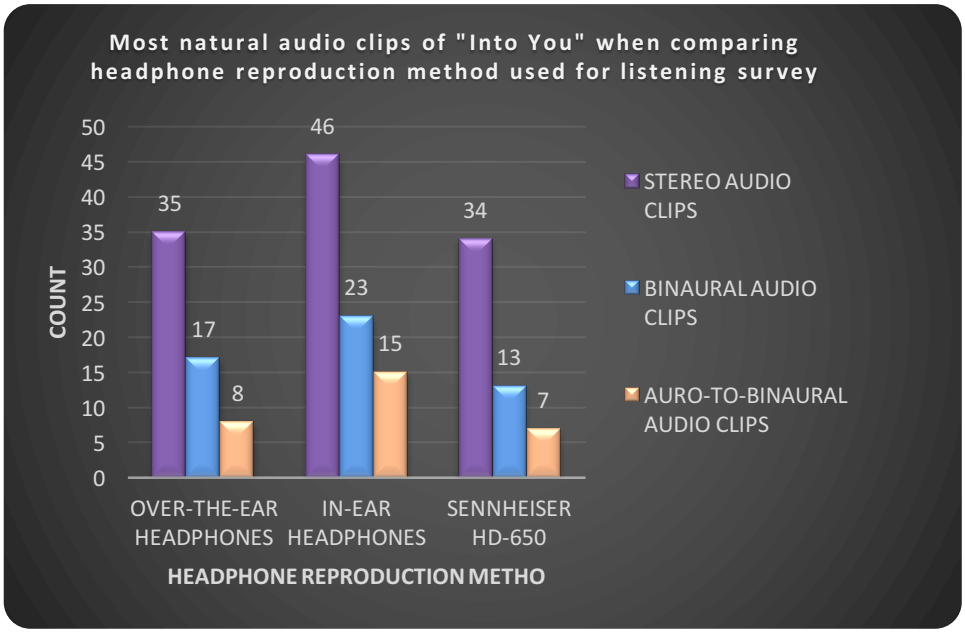


Figure 4.28 ***Bar Chart of the most natural audio clips of "Into You" based on headphone reproduction method.***
Categorized results show that participants answered in a similar fashion no matter the headphone reproduction method.

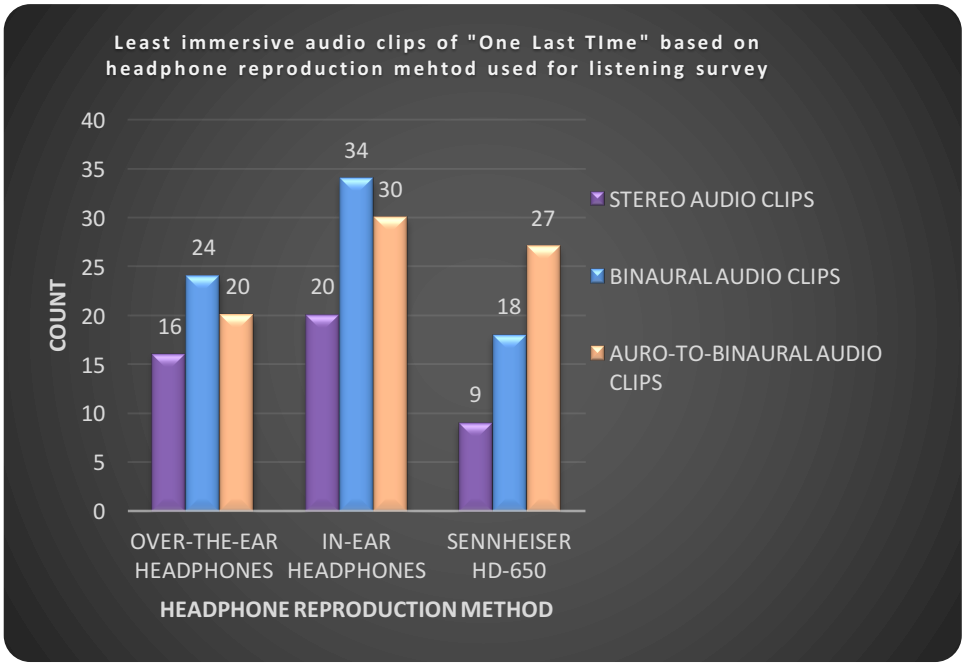


Figure 4.29 ***Bar Chart of the least immersive audio clips of "One Last Time" based on headphone reproduction method.***
Categorized results show that there was no clear distinction as to which audio clip was the least immersive.

Figure 4.29 also shows that the headphone reproduction method did not significantly impact the survey responses. While stereo mixes were less often selected as the least immersive audio clips of "One Last Time", whether the Auro-To-Binaural mixes or the binaural mixes were the least immersive remained ambiguous. The uncontrolled survey participants that used either any pair of over-the-ear or in-ear headphones showed very slight differences between the binaural and Auro-To-Binaural mixes as the least immersive audio clips. Only the controlled survey participants that used the Sennheiser HD 650 headphones had showed a majority reflecting that the Auro-To-Binaural mixes were the least immersive. While controlled survey participants were the only ones to show a majority, there were less of those participants making it difficult to judge if those headphones truly made an impact on listener preferences.

4.6 Listener Evaluations Discussion

There were many different layers of analysis that were completed in order to distinguish which variables made a significant impact to listener responses. The large scale conclusions made in Section 4.5.1 showed that the stereo mixes were the most preferred audio clips while the Auro-To-Binaural mixes were the least preferred. In terms of timbral and spatial qualities, stereo mixes were most often selected as the most natural sounding clips with the best spatial qualities and best externalization. The Auro-To-Binaural mixes were most often selected as the most colored sounding audio clips and the least immersive audio clips. However, this large level analysis provided little detail on how the binaural mixes were evaluated.

Section 4.5.2 focused more on how each song could have contributed to listener preferences. In this analysis, it was discovered that the evaluations for "Into You" were significantly different the evaluations for "One Last Time". More specifically, evaluations of the non-stereo mixes of "One Last Time" were less selected in terms of preference and timbral qualities, especially when compared to the results of "Into You". The significant difference between the two songs led to further investigation that not only separated the results for each song, but also compared rounds one through three for each song.

While rounds one through three for "Into You" and "One Last Time" did not mirror each other in terms of sequence, this analysis further emphasized the different reactions towards the mixes of "Into You" and "One Last Time". As shown in Section 4.5.3, listeners were less likely to select the non-stereo mixes of "One Last Time" as their most preferred audio clips. Additionally, the binaural and Auro-To-Binaural mixes of "One Last Time" were more often selected as the most colored audio clips and the least immersive audio clips. However, this analysis still did not provide any more details on what contributed to the very low evaluations of the binaural mixes for "One Last Time."

The discrepancy between how the binaural and Auro-To-Binaural mixes were ranked led to an investigation on how each engineer's mixes were evaluated between each round in Section 4.5.4. This

analysis provided the most information on what variables contributed to listener preferences. Bar charts in Section 4.5.4 reflected that the more experienced engineers were able to produce the better evaluated binaural mixes. Furthermore, the engineer with prior experience in 3D audio almost consistently produced binaural mixes that were better evaluated than the stereo mixes. This makes one wonder if Grammy nominated mixing engineers with 10 or even 20+ years of experience would be able to release binaural mixes that would be evaluated better than their own stereo mixes.

A final analysis was completed in Section 4.5.5 that investigated any further factors that could have contributed to listener preferences. While the results showed that music background or headphones did not significantly impact listener preferences, the information showed that listeners even with no music background were able to distinguish the same sonic qualities as those individuals that were either trained musicians or music technologists. The most significant differences between music backgrounds showed that music technologists were more critical in selecting their most preferred and least preferred audio clips. It seemed that music technologists took into account coloration and immersion towards making their decisions on their most preferred and least preferred audio clips.

5

Discussion and Conclusion

5.1 Significant Findings

This thesis began with the motivation to discover the best approaches towards mixing binaural music; therefore, there were two phases of this thesis that were evaluated: (1) binaural mixing procedures and (2) listener evaluations of those mixing methods. The mixing procedures that were tested in this thesis were:

(1) stereo mixing using loudspeakers
(2) virtual Auro-3D 9.1 binaural mixing monitored binaurally using headphones
(3) Auro-3D 9. 1 mixing monitored over loudspeakers and post-binaurally rendered

During the mixing phase, the mixing engineers were able to quickly adapt towards the non-stereo mixing methods; however, the process was still strange and unfamiliar. For example, requiring the engineers to mix using only headphones during the binaural mixing phase was difficult for some of the engineers. While the engineers were able to incorporate the G'Audio binaural plug-in within their workflow, it still changed their habits making it a slower and uncomfortable process. Each engineers first experience with the binaural mixing sessions was noted as their most difficult while their second experience with binaural mixing was marked as either their favorite or easiest mix. All three engineers found the Auro-3D 9.1 mixing sessions as the most difficult because there were so many channels to work with. However, all three engineers showed genuine interest in continuing to learn and produce binaural and Auro-3D 9.1 mixes.

The mixes that were produced were evaluated through two online listening surveys. One was uncontrolled and open to the general public. The other survey was controlled and was limited to music technologists where the environment and equipment were controlled variables set in place by the author. The results from the survey show that listeners no matter their music background consistently found the stereo mixes as the most natural sounding mixes, which could be a result of familiarity with the format. Additionally, the large scale analysis showed that the stereo mixes were the least colored and had the best spatial qualities in terms of externalization and immersion. On the other hand, the Auro-To-Binaural mixes

were the least preferred, most colored, and least immersive while the binaural mixes sat somewhere in between.

An evaluation based on each engineer's binaural mixes showed more interesting findings. The engineer with prior experience in 3D audio and production consistently produced binaural mixes that were more preferred than the stereo and Auro-To-Binaural mixes. Additionally, the engineer with the most experience produced binaural mixes that were more preferred than the stereo mixes for the song that was worked on second, showing a high rate of learning and adaption.

Other degrees of analysis revealed that the controlled survey participants used the timbral and spatial qualities to affect their judgement on preference while the uncontrolled survey participants did not. For example, there were instances when the most preferred audio clip was also the most colored and least immersive. This was not the case for the controlled survey participants who even in their comments mentioned that coloration affected their preference.

The listener evaluations revealed that binaural mixes that were monitored binaurally during the mixing process were more preferred than just the average stereo mix. However, the mixes that were post-binaurally rendered were consistently the worse evaluated mixes. While coloration and spatialness were noticed within binaural mixes, these factors did not greatly contribute to listener preferences from the uncontrolled survey participants who represented a small sample size of the general public.

5.2 Contributions

Previous research on the subject of binaural music reproduction concluded that binaural music was less preferred and insignificant to the average music listener; however, prior research approached binaural mixing using either binaural recordings or binaural post-processing. This thesis specifically employed binaurally monitoring throughout the mixing process to test if those mixes were better evaluated than mixes that were binaurally rendered after the mixing process. While stereo mixes were better evaluated than binaural mixes, a closer analysis revealed that binaural mixes were consistently more preferred when completed by more experienced engineers.

5.3 Future Work

This thesis specifically used the G'Audio binaural renderer, therefore making much of the data reliant on that factor. It would be interesting to see if the results would still hold true when using other binaural renderers like Facebook 360, for example. Additionally, a second survey would like to be carried out where the participants were the mixing engineers that participated in this thesis in order to see how each engineer evaluates their own mixes and mixing methods.

5.4 Conclusions

With the increased usage of headphones and headphone monitoring for music reproduction, it was asked if music should be monitored binaurally during the mixing process. Although stereo mixes were evaluated as the most preferred, less colored, and most spatial, the binaural mixes that were monitored binaurally were more preferred when completed by engineers who had more experience and/or familiarity with 3D audio. This knowledge can help push binaural mixes of popular music towards the average music listener. If binaural mixes were completed by the same engineers who mix the songs that populate today's Top 40 Radio, maybe there could be a separate market for binaural music listening? The potential of having a similar model of processing like *Mastered for iTunes* except *Binaurally Mixed for Headphones* could exist through the research and conclusions that were made from this thesis.

References

Andreopoulou, A., Roginska, A., Bello, J. (2013, October 17). Reduced Representations of HRTF datasets: A Discriminant Analysis Approach. *Presented at the 135th Convention of the Audio Engineering Society*, (8949).

Auro Technologies N.V. (2015, October 8). Auro-3d® Home Theater Setup.

Auro Technologies N.V. (2015, November 4). Auro-Max Next Generation Immersive Sound System.

Auro Technologies N.V. (2016). Auro-Matic Pro - User Guide.

Auro Technologies N.V. (2017). The Auro-3D 9.1 Setup. Retrieved March 5, 2017 from

Begault, D. (2000, April). 3-D Sound for Virtual Reality and Multimedia. *NASA Center for Aerospace, Ames Research Center.*

Billboard Music (2017). Ariana Grande Chart History. Retrieved February 27, 2017, from http://www.billboard.com/artist/1484343/ariana-grande/chart

Cengarle, G. & Perada, A. (2013, May 4). The Perception of Masked Sounds and Reverberation in 3D vs. 2D Playback Systems. *Presented at the 134th Convention of the Audio Engineering Society*, (8822).

Dolby Laboratories Inc. (1998). Dolby Surround Mixing Manual.

Fitzgerald, D. (2011). Upmixing from Mono; a Source Separation Approach. *17th International Conference on Digital Signal Processing,*

Fontana, S., Farina, A., & Grenier, Y. (2007, June 26). Binaural for Popular Music: A Case Of Study. *Proceedings of the 13th International Conference on Auditory Display.*

G'Audio Lab, Inc. (2017a). Works 1.0.0g User Manual.

G'Audio Lab, Inc. (2017b) *Cinematic VR*. Retrieved on March 6 from http://www.gaudiolab.com/cinematicvr#works

Izhaki, R. (2012). *Mixing audio: Concepts, practices and tools*. Amsterdam: Focal Press.

King, R., Leonard, B., & Sikora, G. (2011, October 20). The Effects of Monitoring Systems on Balance Preference: A comparative study of mixing on headphones versus loudspeakers. *Presented at the 131st Convention of the Audio Engineering Society*, (8566).

Martin, B. & King, R. (2015, October 29). Three Dimensional Spatial Techniques in 22.2 Multi-channel Surround Sound for Popular Music Mixing. *Presented at the 139th Convention of the Audio Engineering Society*, (9432).

Miller, C., & Martinez, J. (2016, October 6). Health Department Releases Data on Exposure to Loud Sounds and Headphone Use Among Teens and Younger Adults in New York City In Recognition Of National Protect Your Hearing Month. Retrieved April 13, 2017, from https://www1.nyc.gov/site/doh/about/press/pr2016/pr081-16.page

Owsinski, B. (2006). *The mixing engineer's handbook*. Boston: Thomson Course Technology.

Payri, B., & Sanchis-Rico, J. (2016, June 4). Perceptually significant parameters in stereo and binaural mixing with Logic Pro Binaural Panner. *Presented at the 140th convention of the Audio Engineering Society.*

Pike, C. & Melchour, F. (2013, May 4). An assessment of virtual surround sound systems for headphone listening of 5.1 multichannel audio. *Presented at the 134th Convention of the Audio Engineering Society*, 8819).

Pulkki, V. (1997, June). Virtual Sound Source Positioning Using Vector Base Amplitude Panning. *Journal of the Audio Engineering Society* (45:6).

Ronen, Y. (2015, October 29). Vocal Clarity in the Mix: Techniques to Improve the Intelligibility of

Vocals. *Presented at the 139th Convention of the Audio Engineering Society*, (9445).
Rumsey, F. (2001). Spatial Audio. Chapter 4.
Rumsey, F. (2002). Spatial Quality Evaluation for Reproduced Sound: Terminology, Meaning, and a Scene Based Paradigm. *Journal of the Audio Engineering Society*, 50(9), 651 666.
Schroeder, M. (1958, April). An Artificial Stereophonic Effect Obtained from a Single Audio Signal. *Journal of the Audio Engineering Society* (6: 2).
Schönstein, D., Katz, B. (2010, May 22). Variability in perceptual evaluation of HRTFs. *Presented at the 128th Convention of the Audio Engineering Society*, (8153).
Takanen, M., Hiipakka, M., Pulkki, V. (2012, March 1). Audibility of coloration artifacts in HRTF filter designs. *Presented at the AES, 45th International Conference*.
Toole, F. (1984, May 11). The Acoustics and Psychoacoustics of Headphones. *Presented at the 2nd AES International Conference*.
Uhle, C. & Gampp, P. (2016, June 4). Mono-to-Stereo Upmixing. *Presented at the 140th Convention of the Audio Engineering Society*.
Washenko, A. (2017). Nielsen 2016 report: "Music consumption is at an all-time high". Retrieved April 13, 2017, from http://rainnews.com/nielsen-2016-report-music-consumption-is-at-an-all-time-high/

A

Mixing Session Survey and Information

A.1 Mixing Engineer Survey

1. What is your age?
   - 18-24
   - 25-34
   - 35-44
   - 45-54
2. How many years have you been working as a mixing engineer?
   - 1-3
   - 3-5
   - 5-10
   - Over 10 years
3. In a typical day, about how many hours a day do you spend mixing?
   - 0-2
   - 2-4
   - 4-6
   - Over 6 hours
4. How do you most often listen to music? (check all that apply)
   - Over-the-ear headphones
   - In-ear headphones
   - Computer speakers
   - Loudspeakers or monitors
   - Car speakers
5. Please check all the DAWs you are proficient in.
   - ProTools
   - Logic
   - Abelton Live
   - Studio One
   - Garageband
   - Others: _____
6. Do you have any prior experience with 3D Audio Recording and Production? If yes, please describe.
   - Yes
   - No

7. Which mixing session did you just complete?
   - o Stereo
   - o Auro-3D 9.1 over loudspeakers
   - o Binaural Virtual Auro-3D 9.1 over headphones

Key Terms:
Spatial Attributes - related to the impression of width, depth, and elevation
Externalization - related to the localization of sound outside of the head
Timbral Qualities - related to the spectral characteristics of a sound source
Naturalness - related to the most familiar listening environment
Quality- fidelity was not obstructed by spectral coloration

1. Was this a difficult mixing session. If yes, please describe.
   - o No
   - o Yes

     _____
     _____No
   - o Yes

     _____
     _____
2. Was timbre affected by your mixing procedure. If yes, please describe.
   - o No
   - o Yes

     _____
     _____
3. Was naturalness affected by your mixing procedure. If yes, please describe.
   - o No
   - o Yes

     _____
     _____
4. Did you have any difficulty with externalization? If yes, please describe.
   - o No
   - o Yes

     _____
     _____
5. Was the quality of sound affected by the mixing procedure? If yes, please describe.
   - o No
   - o Yes

     _____
     _____
6. Was the diffuseness of sources affected by the mixing procedure? If yes, please describe.
   - o No
   - o Yes

     _____
     _____
7. If you have any additional comments or feedback on your mixing experience, please provide them below:

A.2 Mixing Engineer History

Question	Engineer X	Engineer Y	Engineer Z
Age	18-24 years old	25-34 years old	25-34 years old
Years Mixing	3-5 years	5-10 years	3-5 years
Hours a day mixing	6+ hours	6+ hours	0-2 hours
Prior Experience with 3D Audio	No	No	Yes

Table A.2 ***Background information and mixing history of the mixing engineers.***

A.3 Mixing Session Completion Time for "Into You"

Session	Engineer X	Engineer Y	Engineer Z
Stereo	50 minutes	1 hour 15 minutes	1 hour
Binaural	1 hour 50 minutes	1 hour 50 minutes	2 hours 30 minutes
Auro-3D 9.1	2 hours	1 hour 40 minutes	2 hours

Table A.3 ***Mixing Session Completion Time for "Into You" for each engineer***

A.4 Mixing Session Completion Time for "One Last Time"

Session	Engineer X	Engineer Y	Engineer Z
Stereo	30 minutes	1 hour 20 minutes	40 minutes
Binaural	1 hour 30 minutes	2 hours 30 minutes	1 hour 30 minutes
Auro-3D 9.1	1 hour 45 minutes	2 hours	2 hours

Table A.4 ***Mixing Session Completion Time for "One Last Time" for each engineer***

B

Listening Survey and Participant Demographics

B.1 Listening Survey

Key Terms:

Natural – without spectral/timbral coloration

Colored – the addition of unwanted spectral artifacts

Spatialness- related to the impression of width, depth, and elevation

Externalization – sense of a surrounding environment outside of the head

Immersive – sense of envelopment by external sound sources

1. Which audio clip did you most prefer?
   - A
   - B
   - C
2. Which audio clip did you least prefer?
   - A
   - B
   - C
3. Which audio clip sounded the most natural (without spectral/timbral coloration)?
   - A
   - B
   - C
4. Which audio clip was the most colored (the addition of unwanted spectral artifacts)?
   - A
   - B
   - C

5. Which audio clip had the best spatial qualities (i.e. width, depth, and elevation)?
    - A
    - B
    - C
6. Which audio clip had the best externalization (sense of a surrounding environment)?
    - A
    - B
    - C
7. Which audio clip was the least immersive?
    - A
    - B
    - C
8. Provide comments, if any, on what you heard. (7 responses)

B.2 Uncontrolled Listening Survey Participant Age Groups

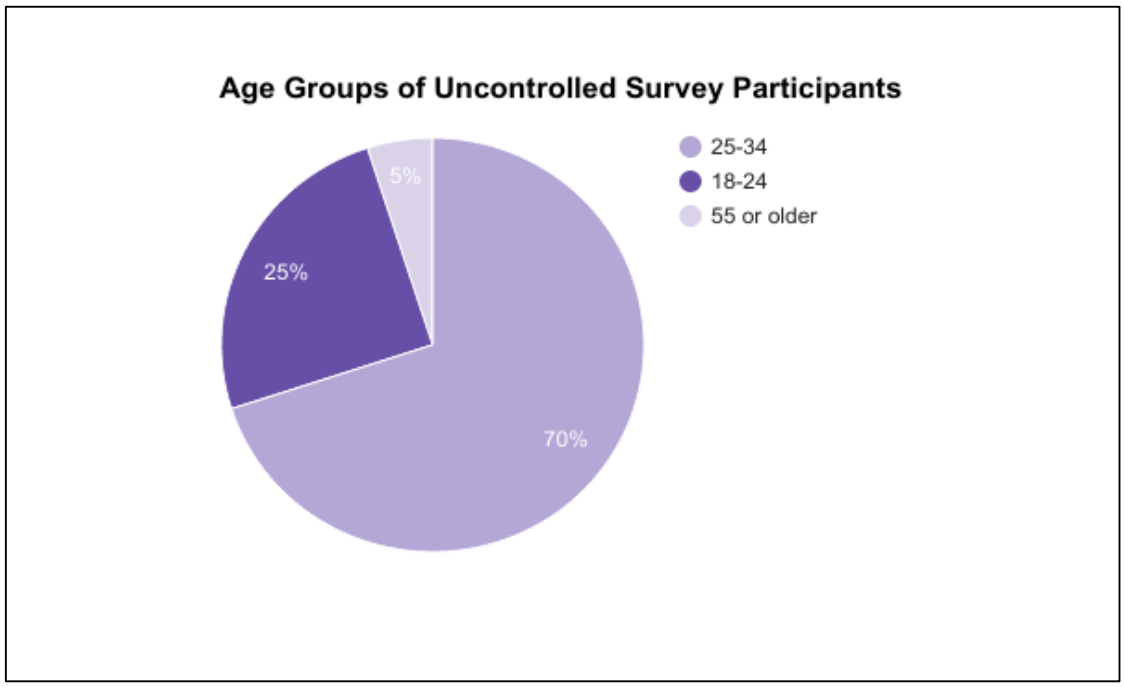


Figure B.2 ***Age Groups of Uncontrolled Survey Participants***

B.3 Controlled Listening Survey Participant Age Groups

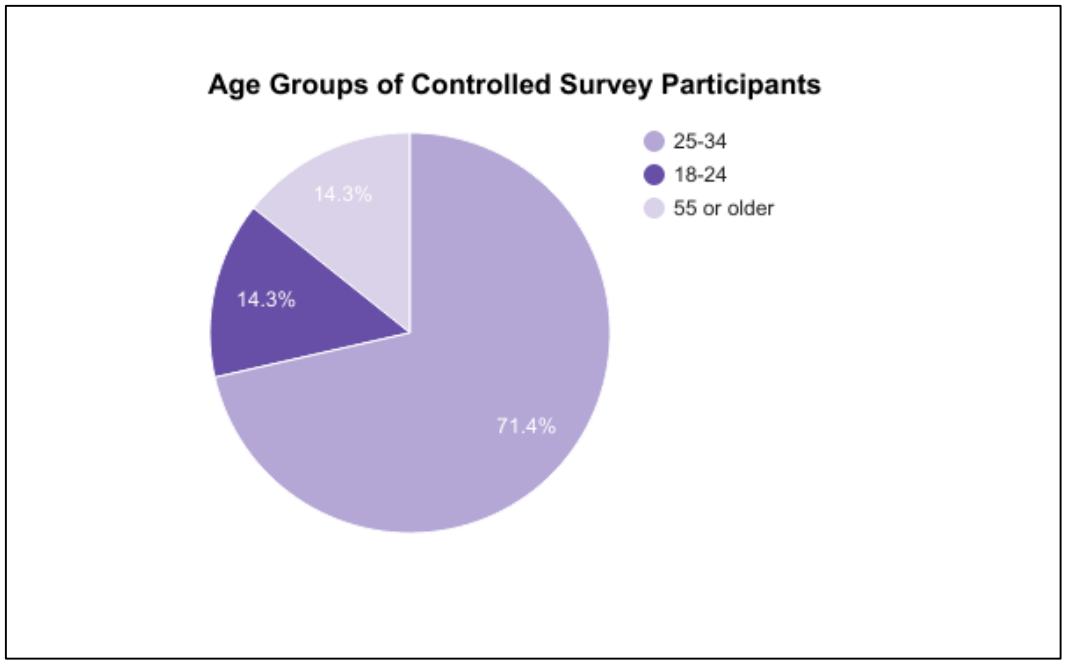


Figure B.3 ***Age Groups of Controlled Survey Participants***

B.4 Uncontrolled Survey Participants Music Consumption Information

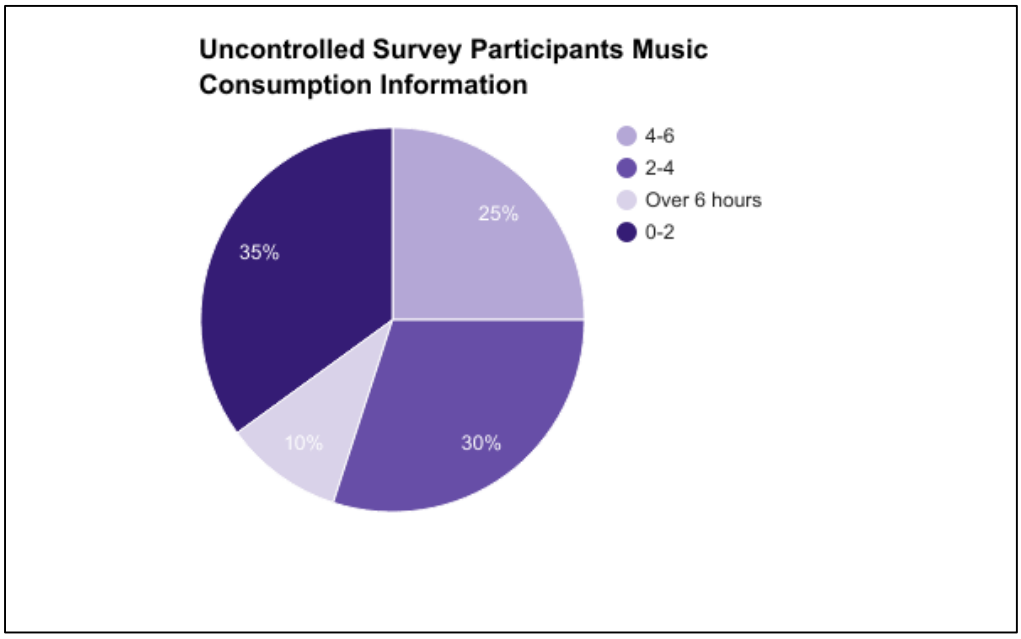


Figure B.4 ***Uncontrolled Survey Participants Music Consumption Information***

B.5 Uncontrolled Survey Participants Music Consumption Information

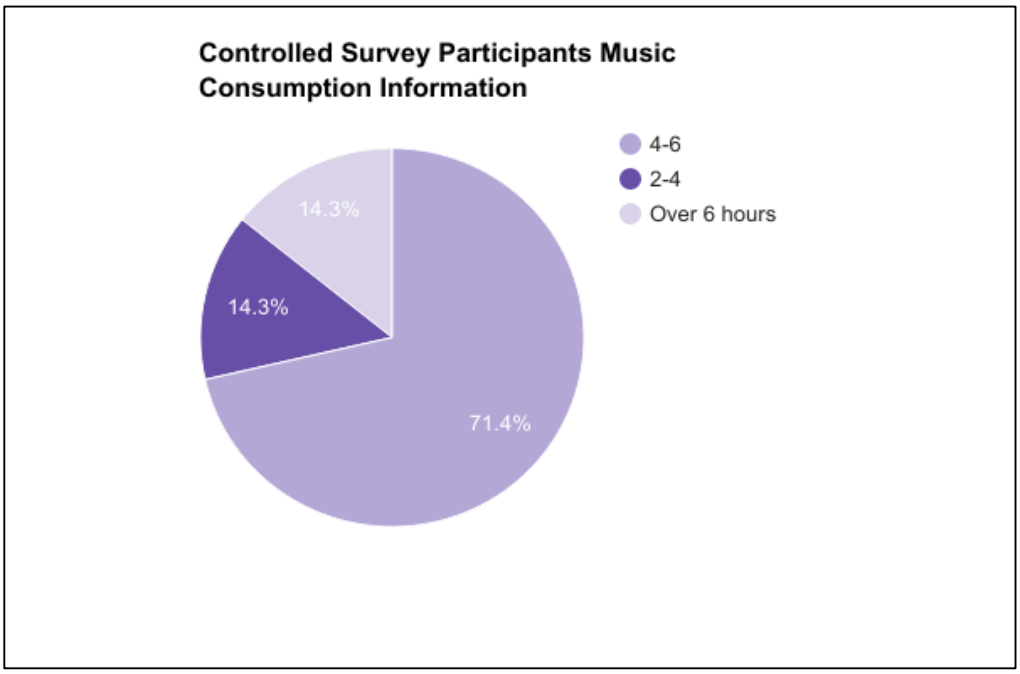


Figure B.5 ***Controlled Survey Participants Music Consumption Information***

C

Figures of Overall Listening Survey Results

C.1 Histograms for Uncontrolled and Controlled Survey Results for "Into You"

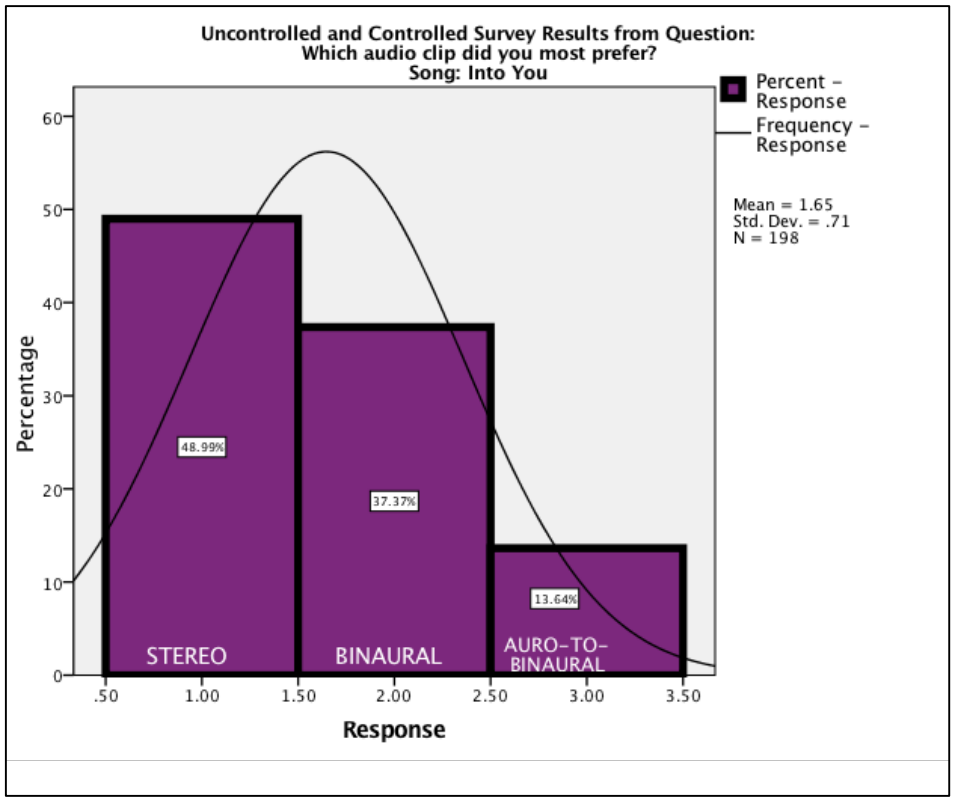
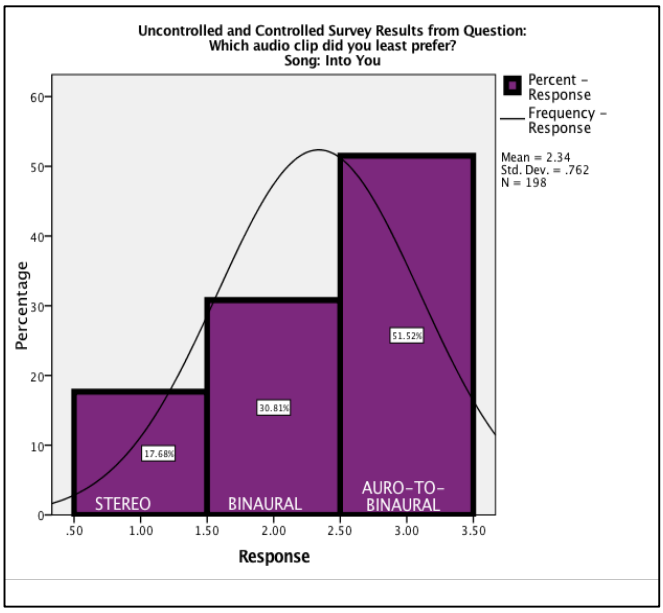


Figure C.1.1 ***Uncontrolled and Controlled Survey Results for "Into You" to Question: Which audio clip did you most prefer?***



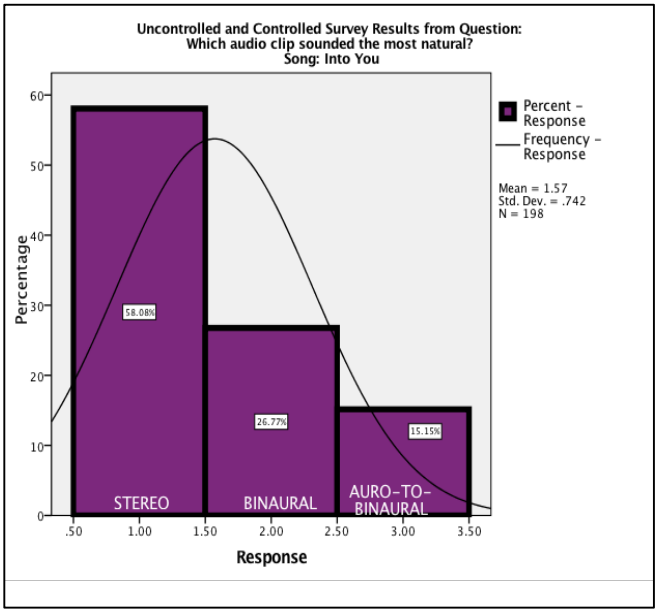
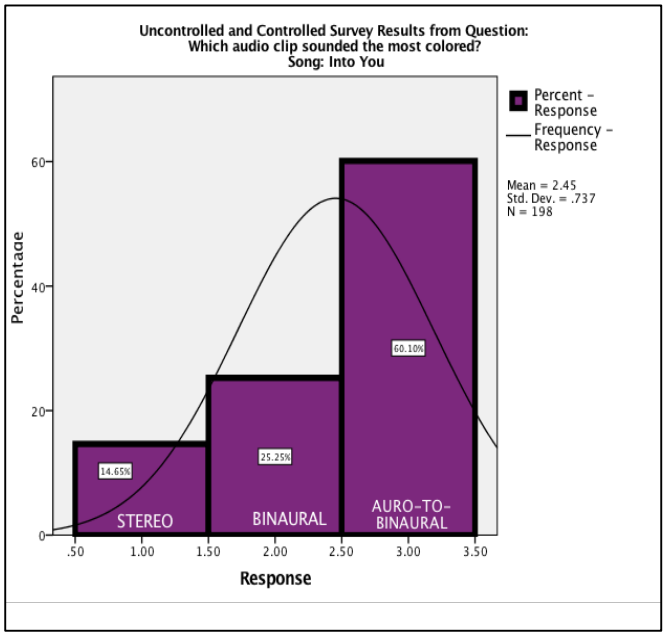


Figure C.1.2(left) ***Uncontrolled and Controlled Survey Results for "Into You" to Question: Which audio clip did you least prefer?***
Figure C.1.3 (right) ***Uncontrolled and Controlled Survey Results for "Into You" to Question: Which audio clip sounded the most natural?***



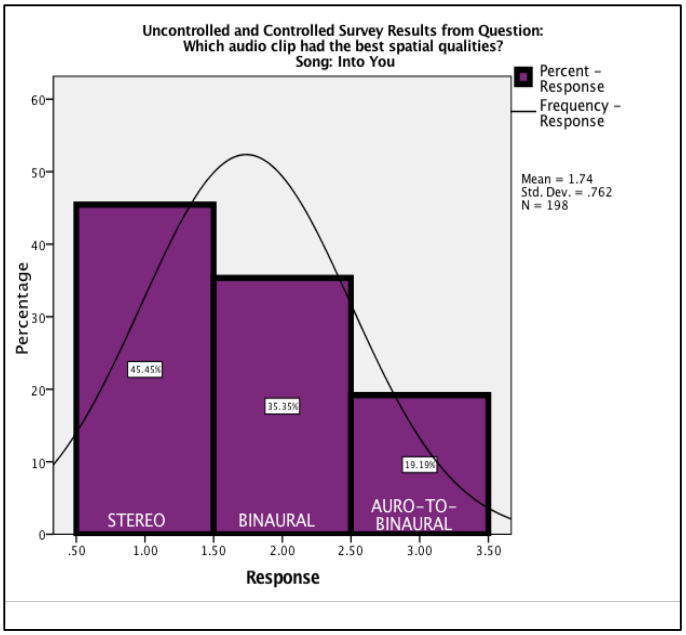
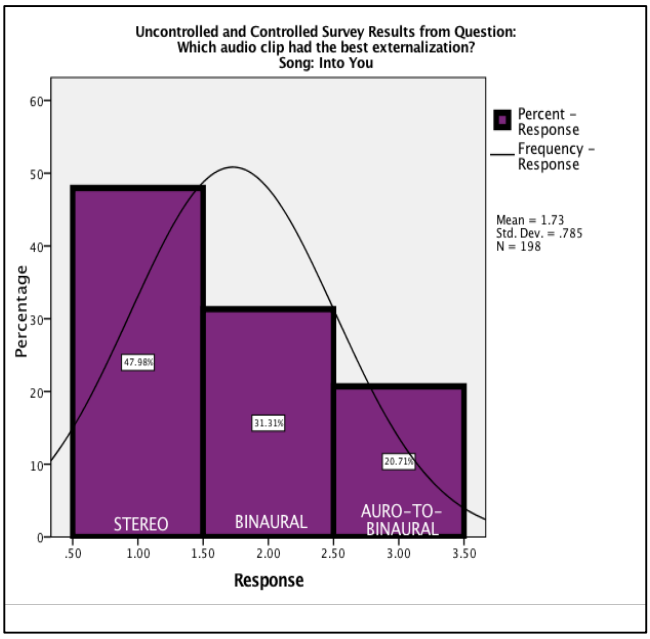


Figure C.1.4 (left) ***Uncontrolled and Controlled Survey Results for "Into You" to Question: Which audio clip sounded the most colored?***
Figure C.1.5 (right) ***Uncontrolled and Controlled Survey Results for "Into You" to Question: Which audio clip had the best spatial qualities?***



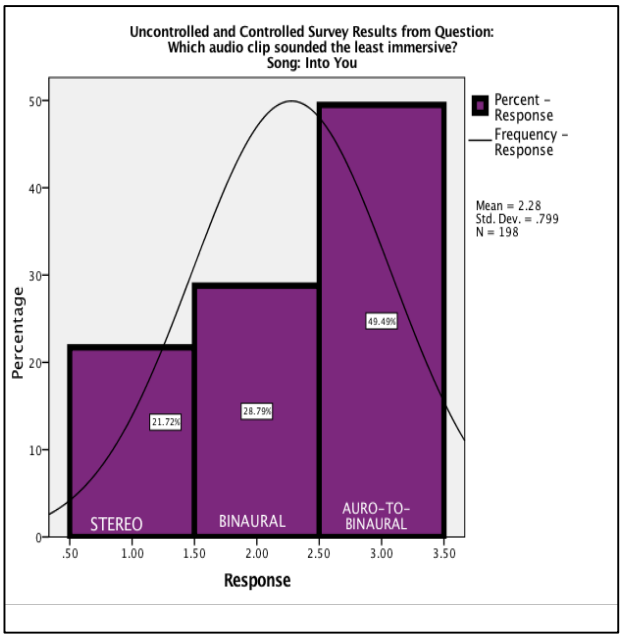
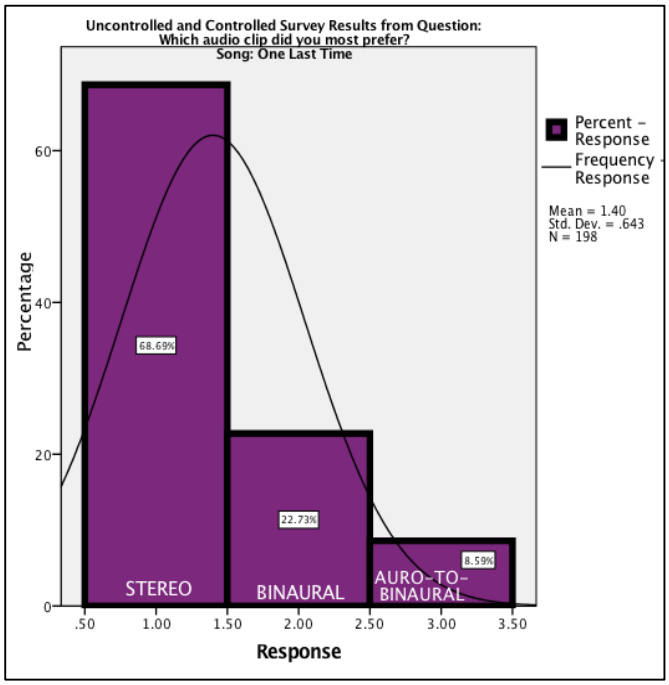


Figure C.1.6 (left) ***Uncontrolled and Controlled Survey Results for "Into You" to Question: Which audio clip had the best externalization?***
Figure C.1.7 (right) ***Uncontrolled and Controlled Survey Results for "Into You" to Question: Which audio clip sounded the least immersive?***

C.2 Histograms for Uncontrolled and Controlled Survey Results for "One Last Time"



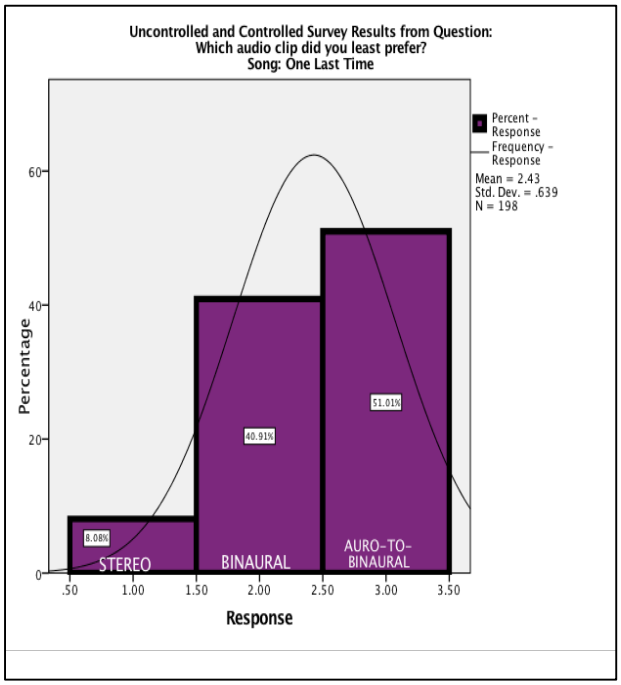
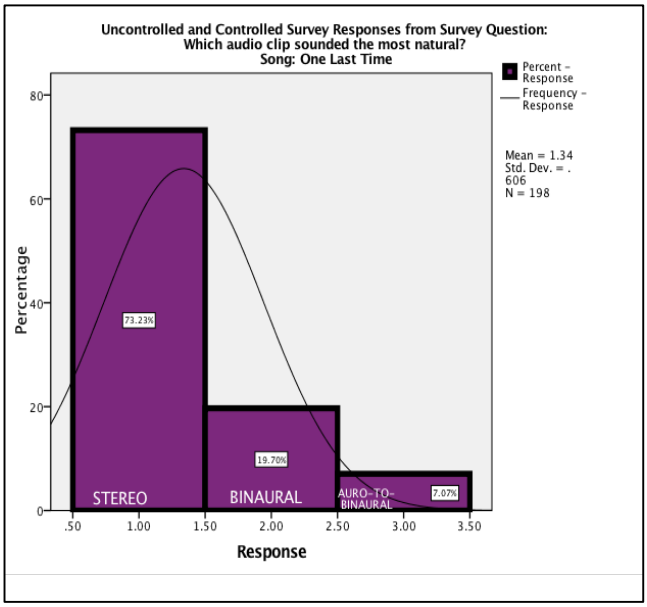


Figure C.2.1 (left) ***Uncontrolled and Controlled Survey Results for "One Last Time" to Question: Which audio clip did you most prefer?***
Figure C.2.2 (right) ***Uncontrolled and Controlled Survey Results for "One Last Time" to Question: Which audio clip did you least prefer?***



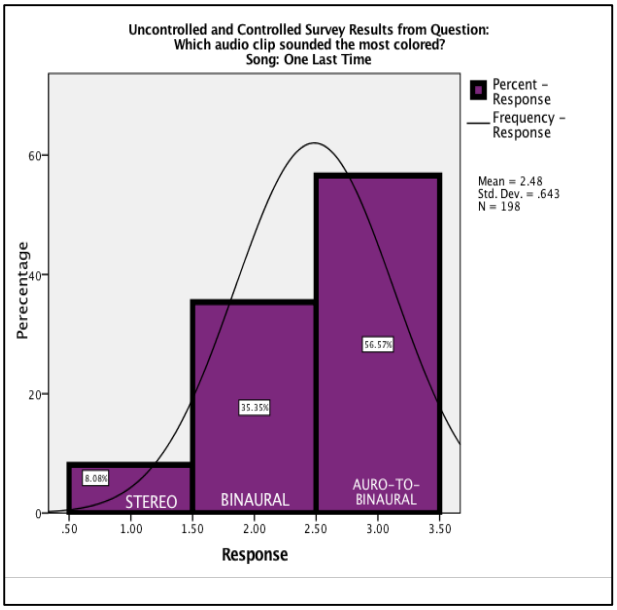
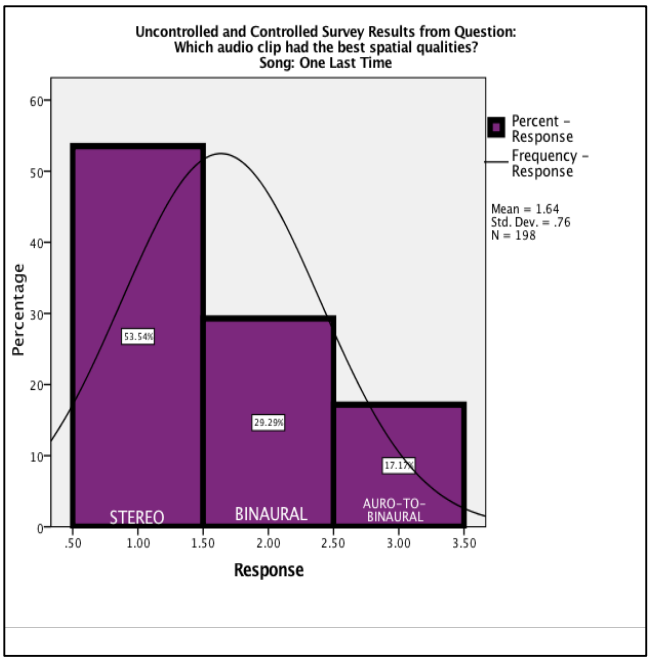


Figure C.2.3 (left) ***Uncontrolled and Controlled Survey Results for "One Last Time" to Question: Which audio clip sounded the most natural?***
Figure C.2.4 (right)***Uncontrolled and Controlled Survey Results for "One Last Time" to Question: Which audio clip sounded the most colored?***



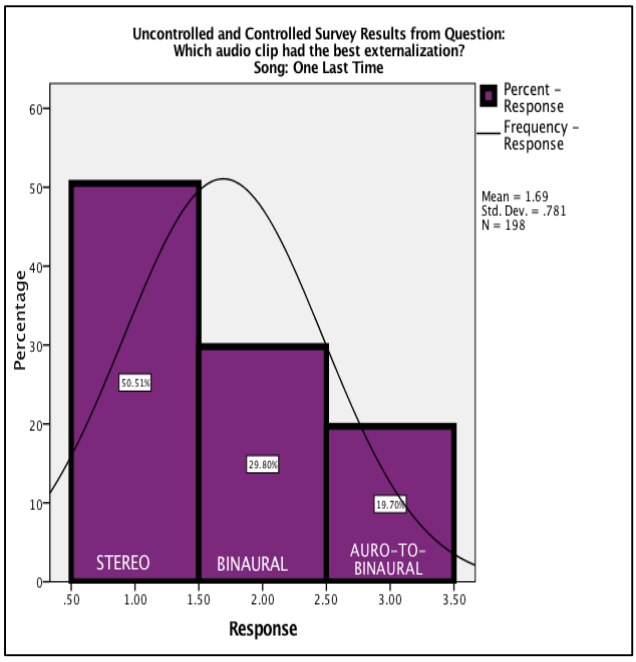


Figure C.2.5 (left) ***Uncontrolled and Controlled Survey Results for "One Last Time" to Question: Which audio clip had the best spatial qualities?***
Figure C.2.6 (right) ***Uncontrolled and Controlled Survey Results for "One Last Time" to Question: Which audio clip had the best externalization?***

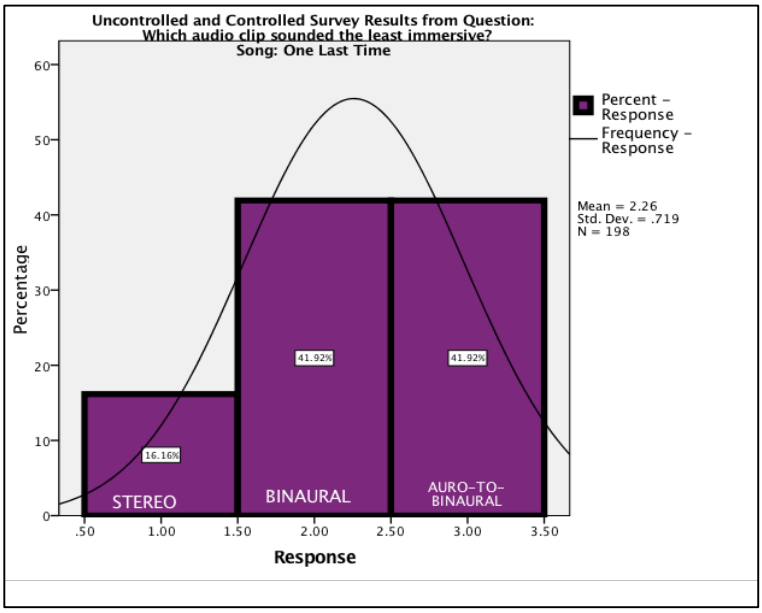
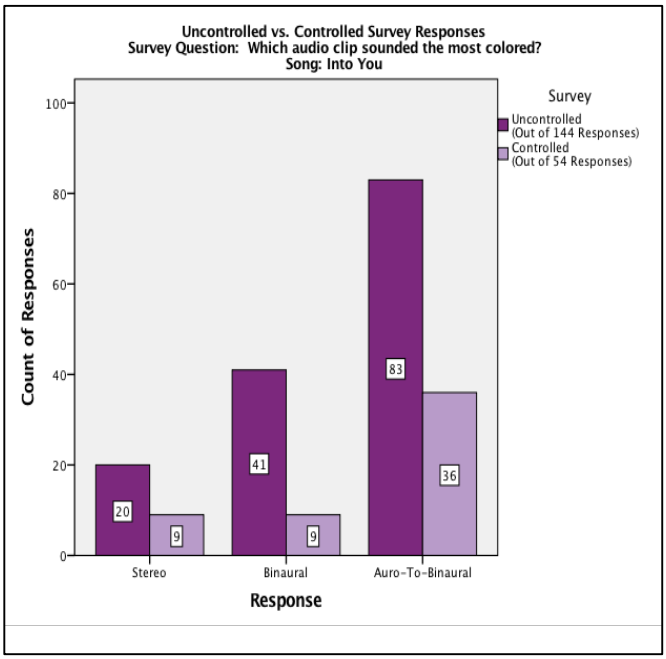


Figure C.2.7 ***Uncontrolled and Controlled Survey Results for "One Last Time" to Question: Which audio clip sounded the least immersive?***

C.3 Bar Charts of Uncontrolled and Controlled Survey Results for "Into You"



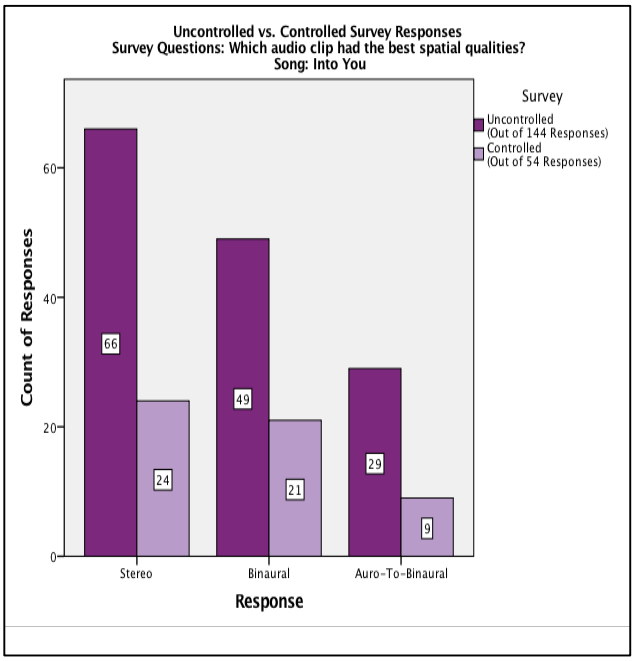
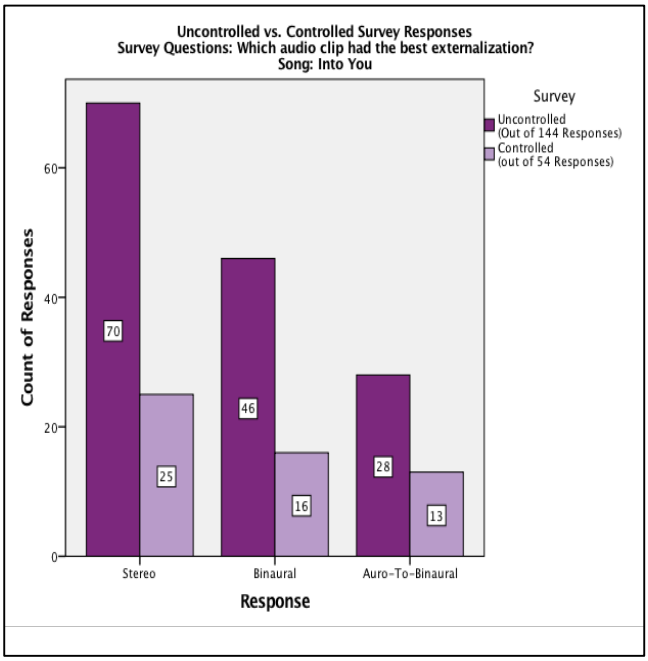


Figure C.3.1 (left) ***Uncontrolled and Controlled Survey Results for "Into You" to Question: Which audio clip sounded the most colored?***
Figure C.3.2 (right) ***Uncontrolled and Controlled Survey Results for "Into You" to Question: Which audio clip had the best spatial qualities?***



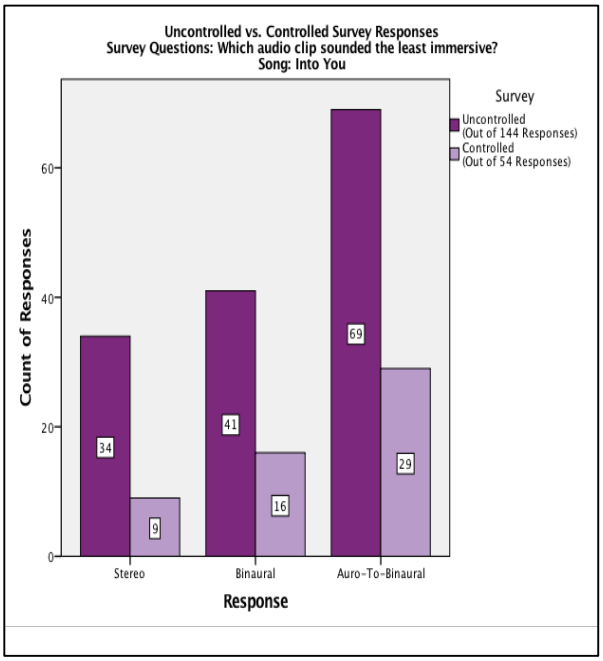
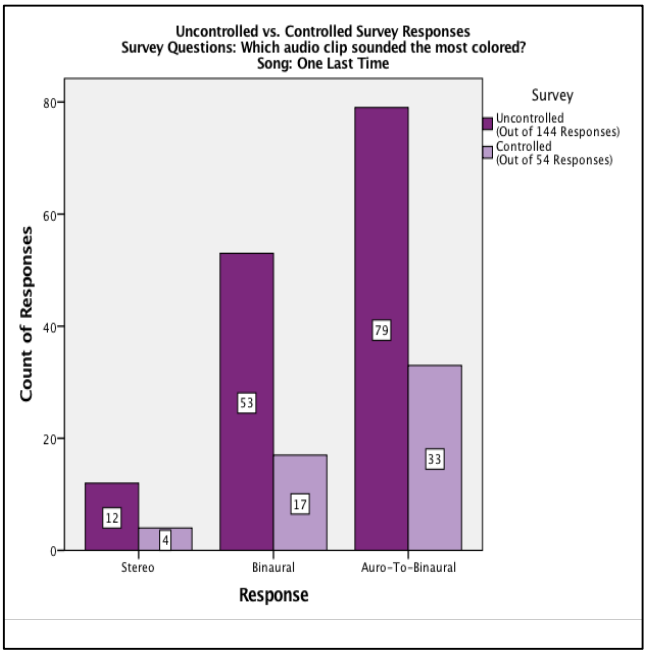


Figure C.3.3 (left) ***Uncontrolled and Controlled Survey Results for "Into You" to Question: Which audio clip had the best externalization?***
Figure C.3.4 (right) ***Uncontrolled and Controlled Survey Results for "Into You" to Question: Which audio clip sounded the least immersive?***

C.4 Bar Charts of Uncontrolled and Controlled Survey Results for "One Last Time"



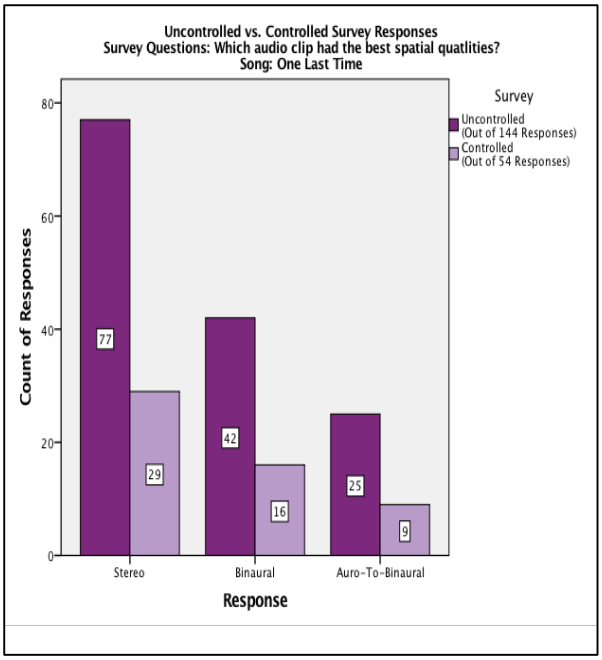
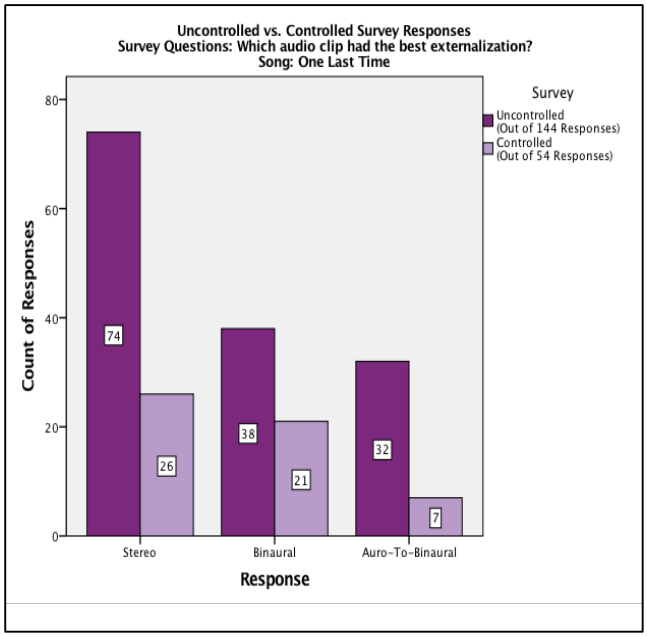


Figure C.4.1 ***Uncontrolled and Controlled Survey Results for "One Last Time" to Question: Which audio clip sounded the most colored?***
Figure C.4.2 ***Uncontrolled and Controlled Survey Results for "One Last Time" to Question: Which audio clip had the best spatial qualities?***



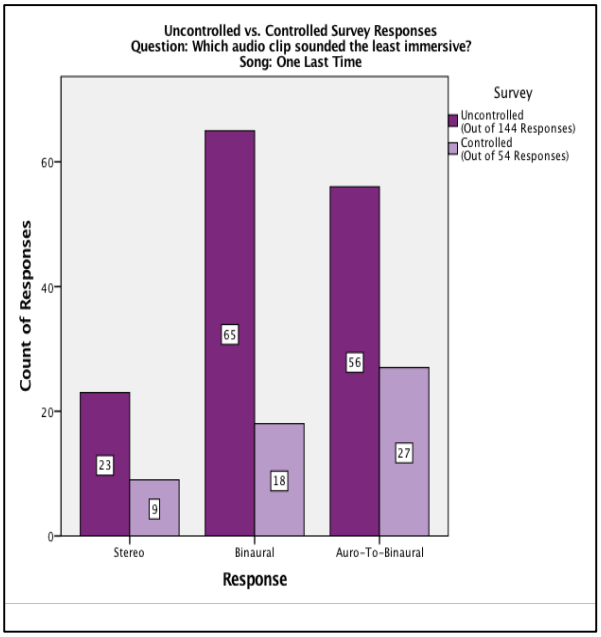
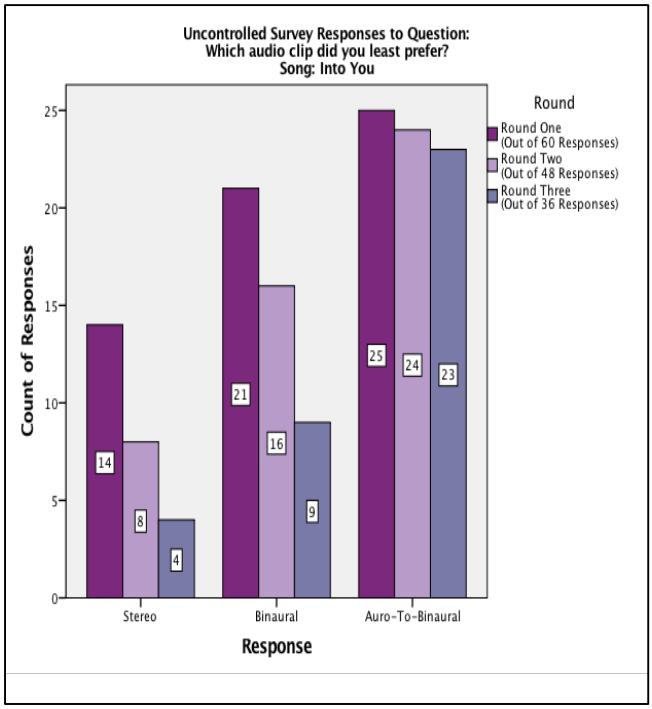


Figure C.4.3 (left) ***Uncontrolled and Controlled Survey Results for "One Last Time" to Question: Which audio clip had the best externalization?***
Figure C.4.4 (right) ***Uncontrolled and Controlled Survey Results for "One Last Time" to Question: Which audio clip sounded the least immersive?***

C.5 Bar Charts for Uncontrolled Survey Results from Rounds One –Three for "Into You"



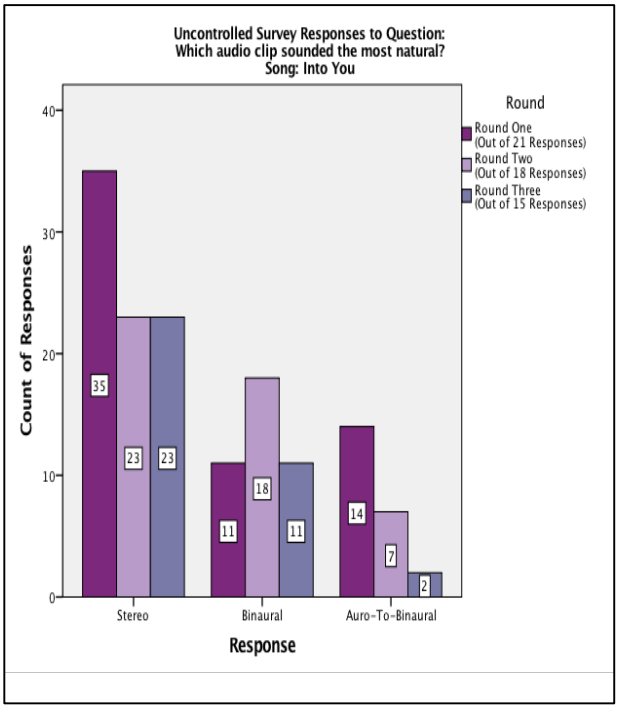
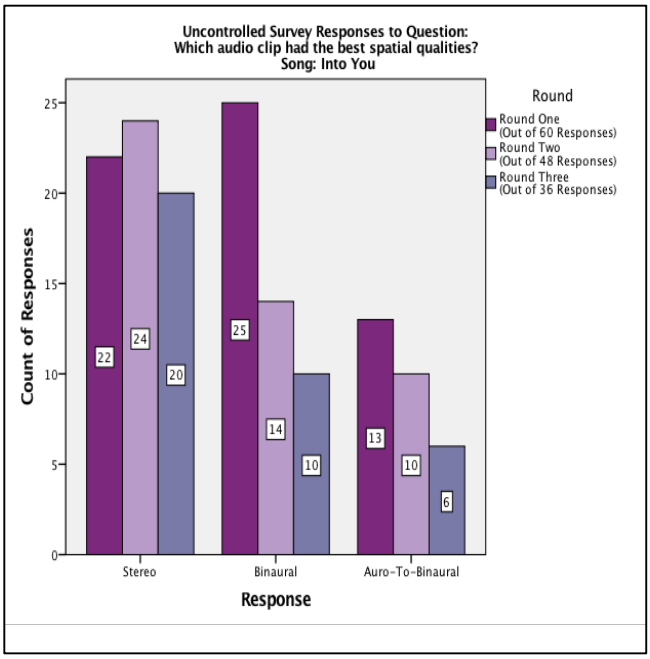


Figure C.5.1 (left) ***Uncontrolled Survey Results for "Into You" to Question: Which audio clip did you least prefer?***
Figure C.5.2 (right) ***Uncontrolled Survey Results for "Into You" to Question: Which audio clip sounded the most natural?***



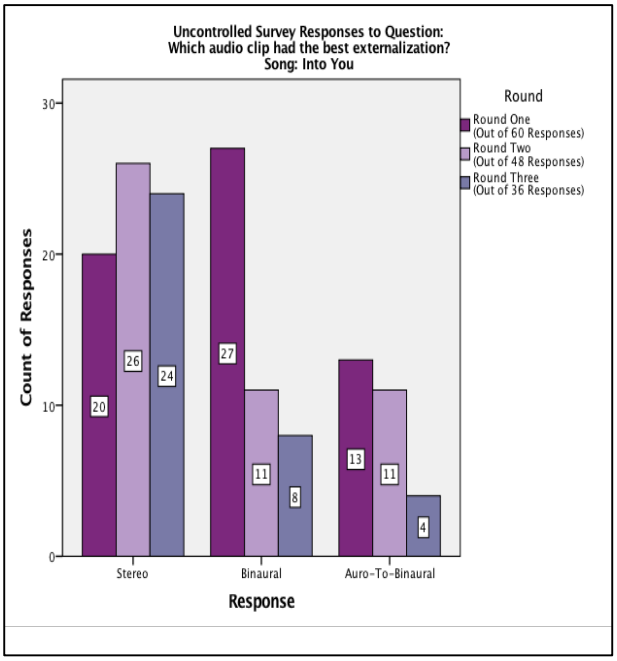
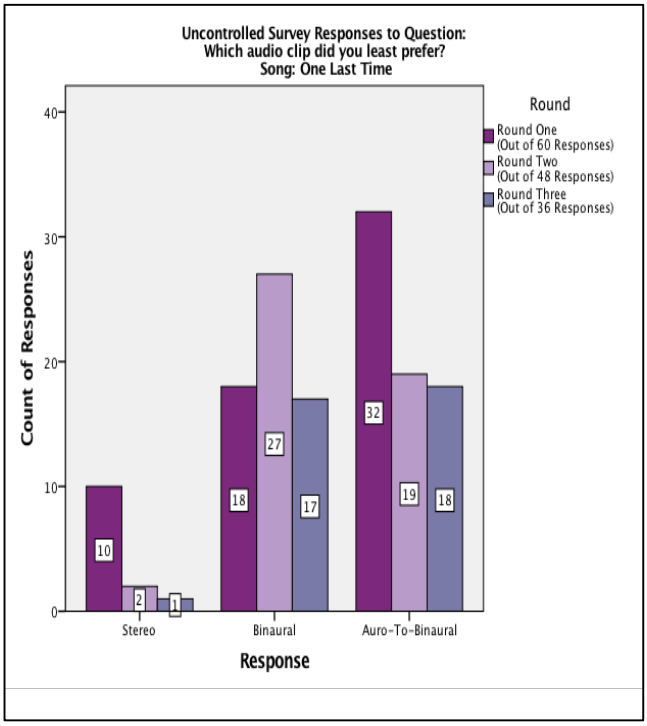


Figure C.5.3 (left) ***Uncontrolled Survey Results for "Into You" to Question: Which audio clip had the best spatial qualities?***
Figure C.5.4 (right) ***Uncontrolled Survey Results for "Into You" to Question: Which audio clip had the best externalization?***

C.6 Bar Charts for Uncontrolled Survey Results from Rounds One –Three for "One Last Time"



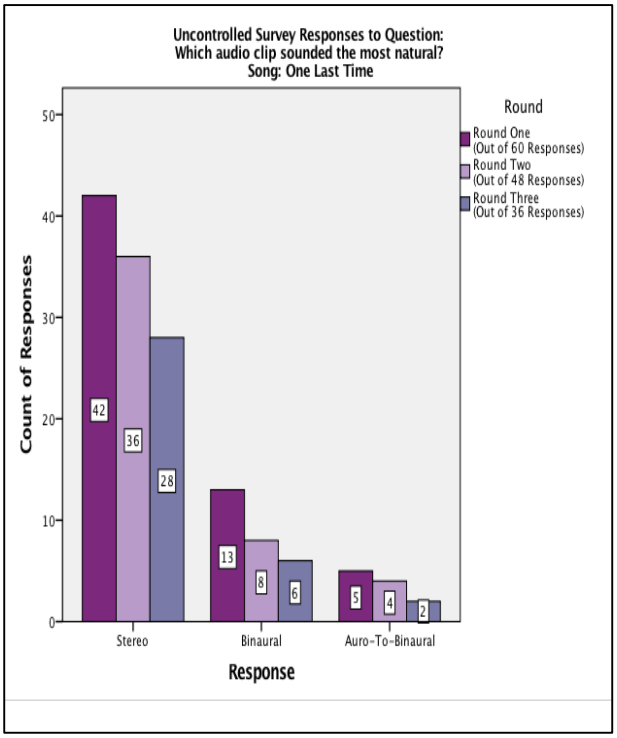
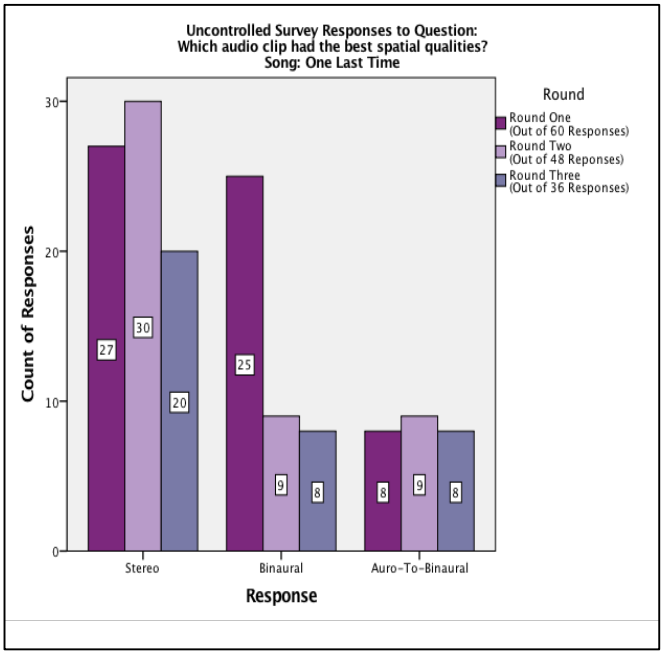


Figure C.6.1 (left) ***Uncontrolled Survey Results for "One Last Time" to Question: Which audio clip did you least prefer?***
Figure C.6.2 (right) ***Uncontrolled Survey Results for "One Last Time" to Question: Which audio clip sounded the most natural?***



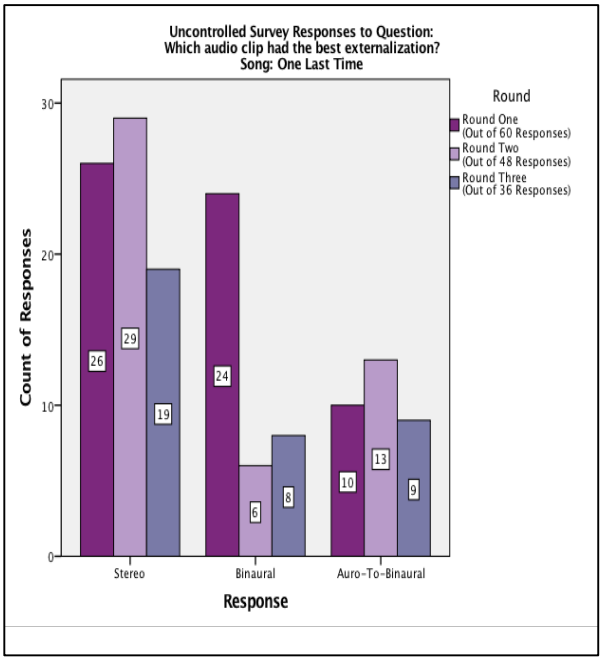
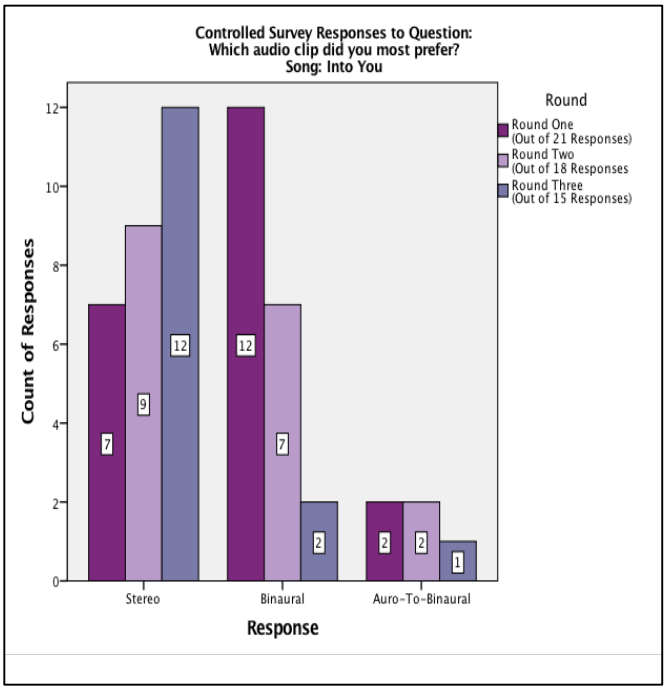


Figure C.6.3 (left) ***Uncontrolled Survey Results for "One Last Time" to Question: Which audio clip had the best spatial qualities?***
Figure C.6.4 (right) ***Uncontrolled Survey Results for "One Last Time" to Question: Which audio clip had the externalization?***

C.7 Bar Charts for Controlled Survey Results from Rounds One –Three for "Into You"



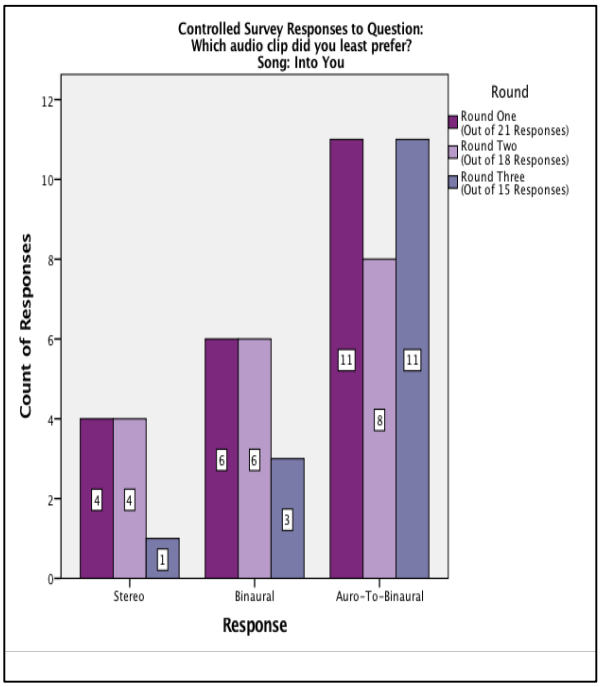
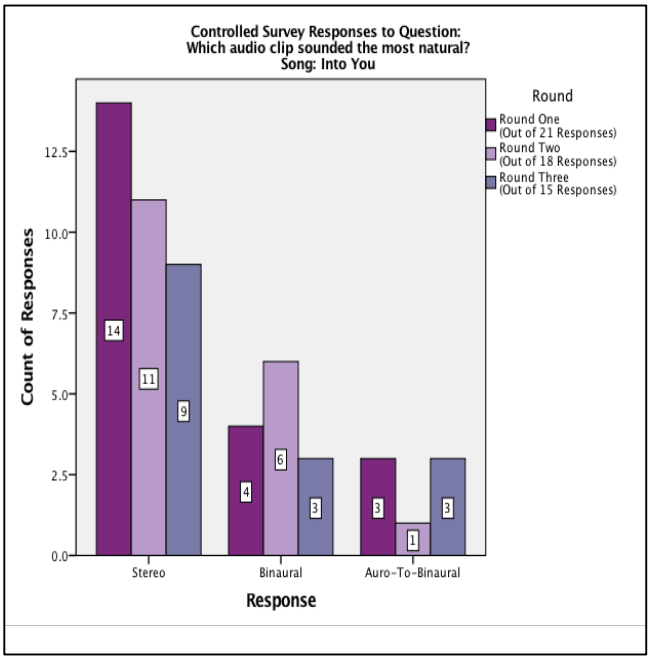


Figure C.7.1 (left) ***Controlled Survey Results for "Into You" to Question: Which audio clip did you most prefer?***
Figure C.7.2 (right) ***Controlled Survey Results for "Into You" to Question: Which audio clip did you least prefer?***



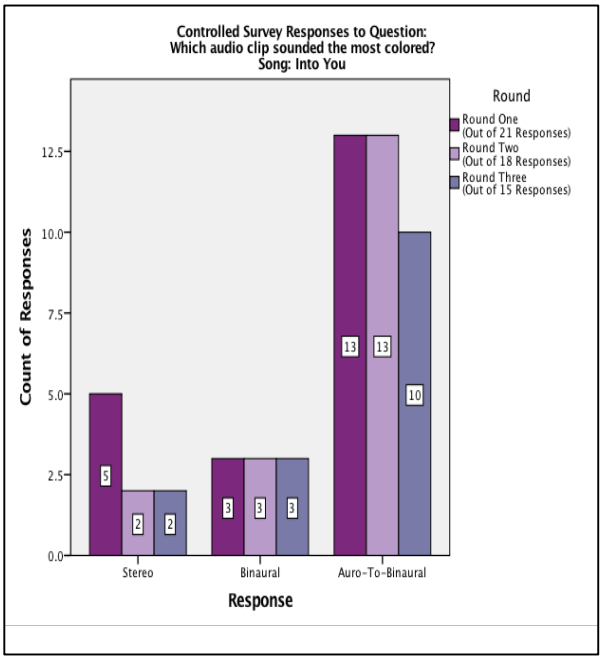
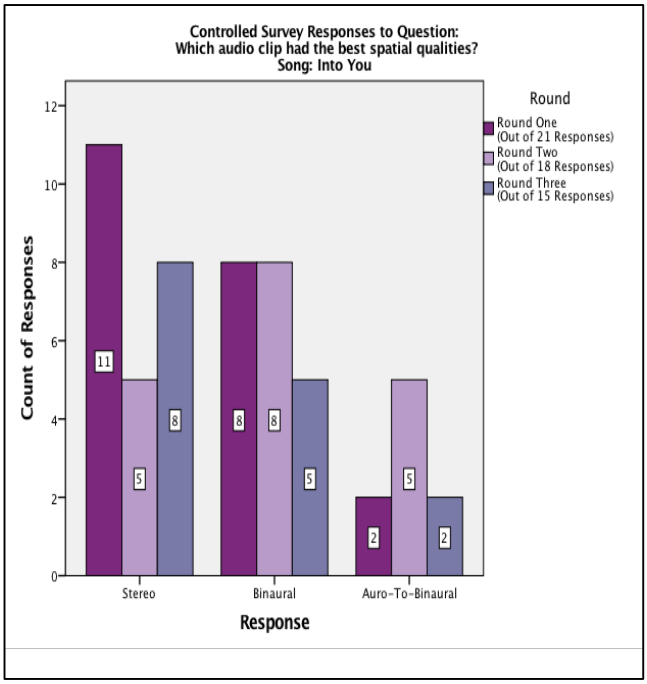


Figure C.7.3 (left) ***Controlled Survey Results for "Into You" to Question: Which audio clip sounded the most natural?***
Figure C.7.4 (right)***Controlled Survey Results for "Into You" to Question: Which audio clip sounded the most colored?***



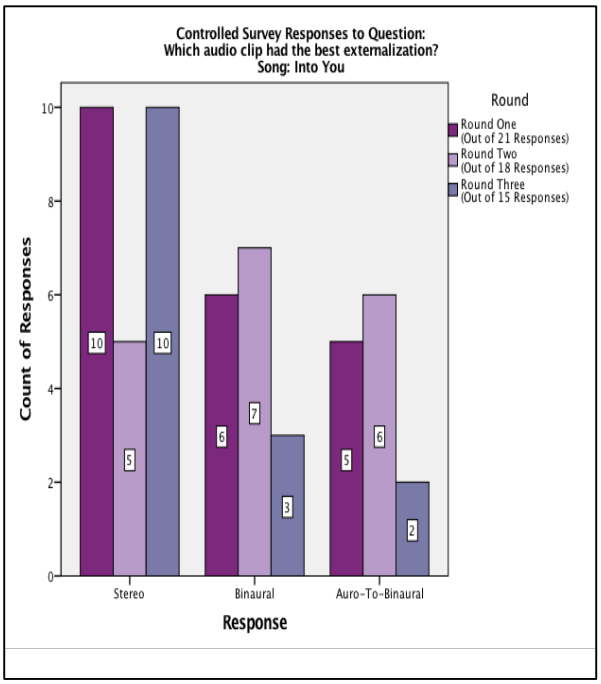


Figure C.7.5 (left) ***Controlled Survey Results for "Into You" to Question: Which audio clip had the best spatial qualities?***
Figure C.7.6 (right)***Controlled Survey Results for "Into You" to Question: Which audio clip had the best externalization?***

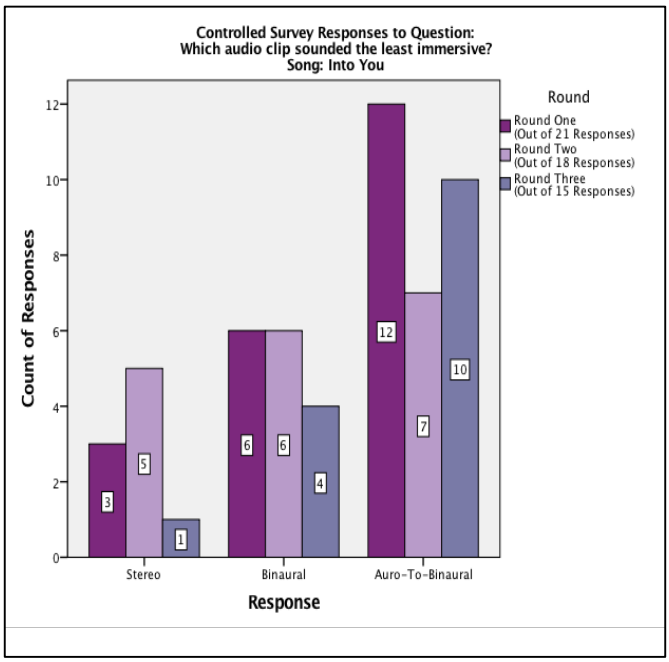


Figure C.7.7 ***Controlled Survey Results for "Into You" to Question: Which audio clip sounded the least immersive?***

C.8 Bar Charts for Controlled Survey Results from Rounds One –Three for "One Last Time"

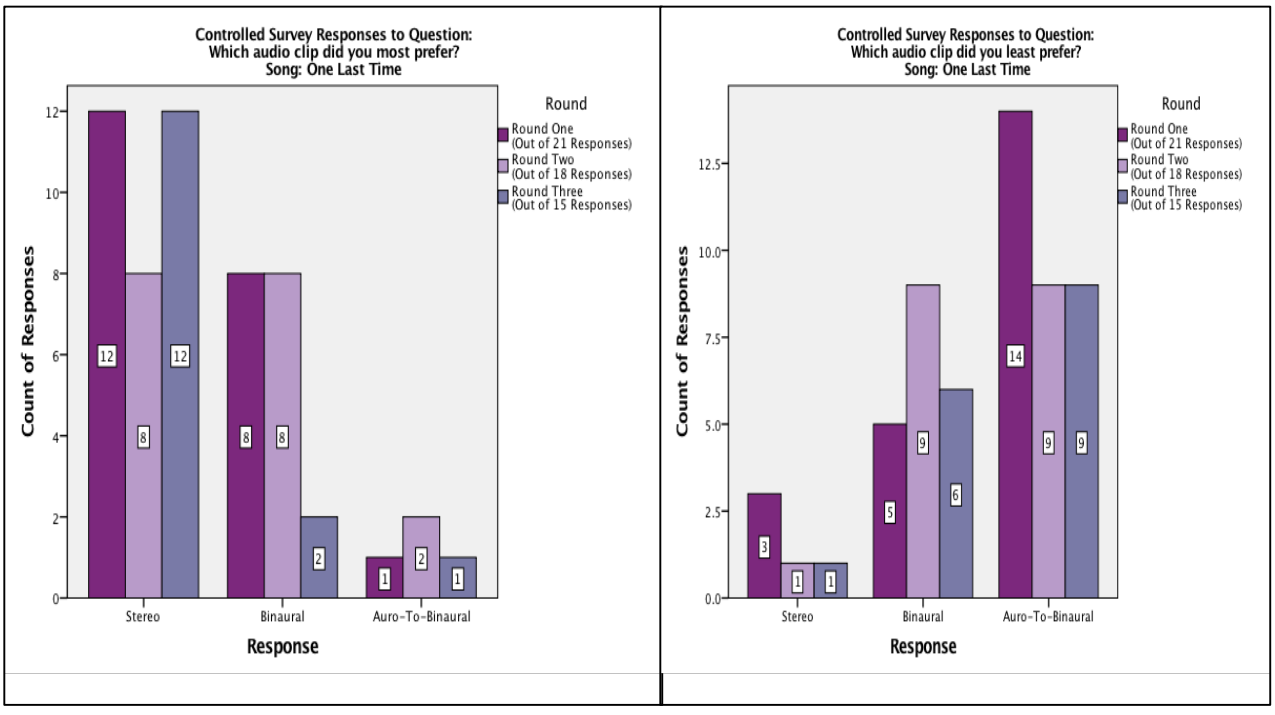
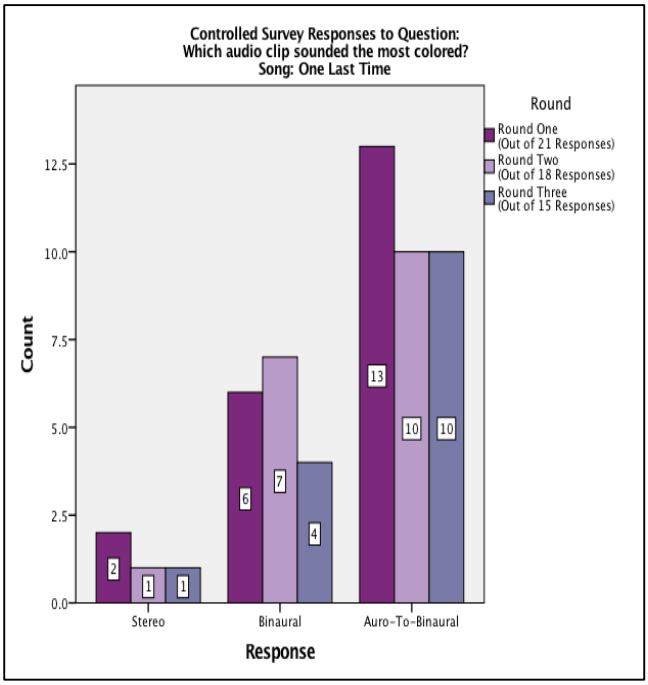


Figure C.8.1 (left) ***Controlled Survey Results for "One Last Time" to Question: Which audio clip did you most prefer?***
Figure C.8.2 (right) ***Controlled Survey Results for "One Last Time" to Question: Which audio clip did you least prefer?***



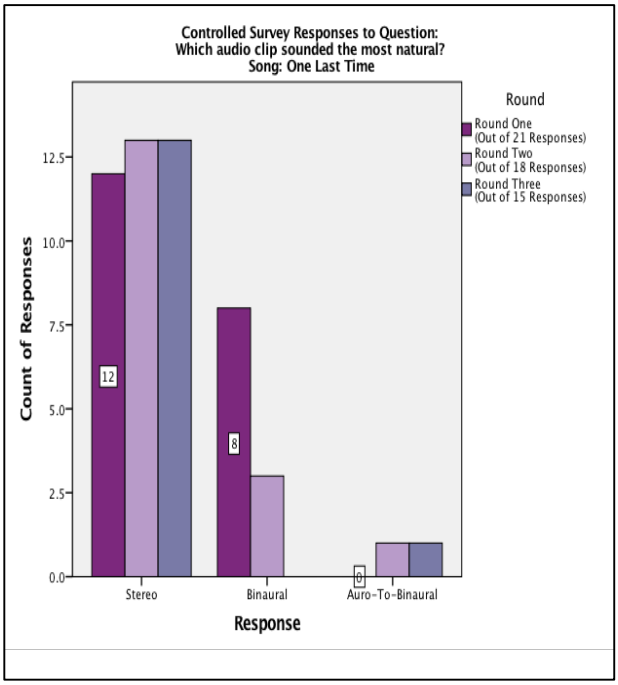
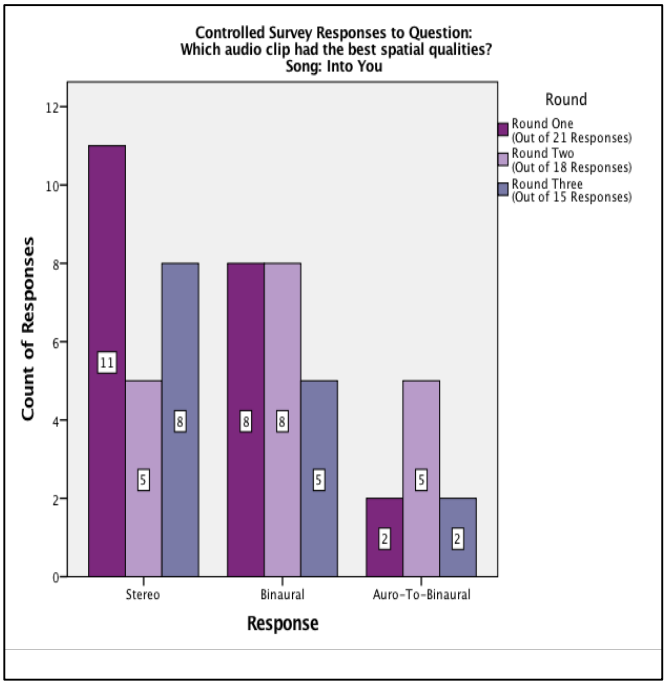


Figure C.8.3 (left) ***Controlled Survey Results for "One Last Time" to Question: Which audio clip sounded the most natural?***
Figure C.8.4 (right) ***Controlled Survey Results for "One Last Time" to Question: Which audio clip sounded the most colored?***



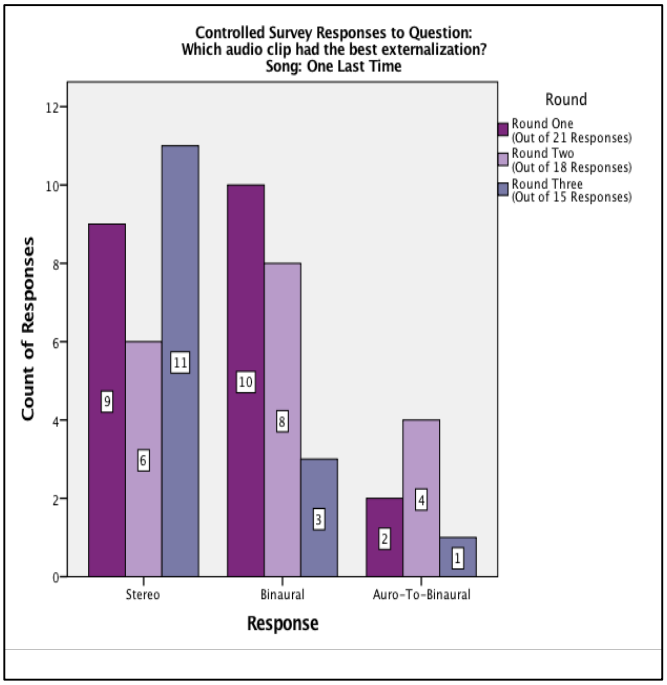


Figure C.8.5 (left) ***Controlled Survey Results for "One Last Time" to Question: Which audio clip had the best spatial qualities?***
Figure C.8.6 (right) ***Controlled Survey Results for "One Last Time" to Question: Which audio clip had the best externalization?***

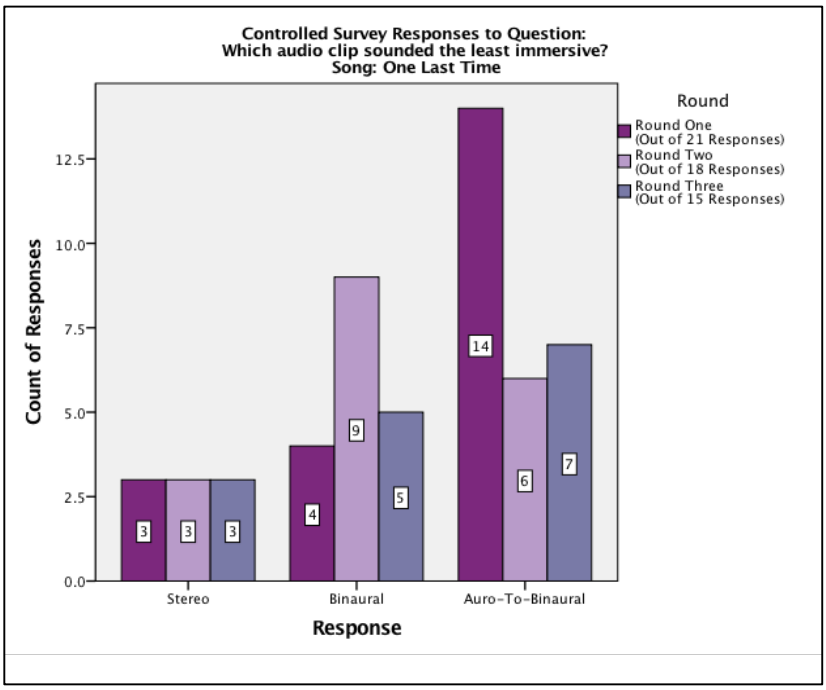


Figure C.8.7 ***Controlled Survey Results for "One Last Time" to Question: Which audio clip sounded the least immersive?***

D

Tables of Overall Listening Survey Results

D.1 Listener evaluation results of "Into You" from both the uncontrolled and controlled surveys

Type of Mixing Procedure	Most Preferred vs. Least Preferred	
	Most Preferred	*Least Preferred*
Stereo	52.02%	17.68%
Binaural	34.34%	30.8%
Auro-To-Binaural	13.63%	51.52%

Table D.1.1 ***Listener preferences of "Into You" from both the uncontrolled and controlled surveys.***
Results showed that from 198 responses, just over 50% of listeners most preferred stereo mixes and least preferred Auro-To-Binaural mixes. Results were from the survey questions (1) Which audio clip did you most prefer? (left column) and (2) Which audio clip did you least prefer? (right column).

Type of Mixing Procedure	Most Natural vs. Most Colored	
	Most Natural	*Most Colored*
Stereo	58.08%	14.65%
Binaural	26.77%	25.25%
Auro-To-Binaural	15.15%	60.10%

Table D.1.2 ***Listener choices of the Most Natural and Most Colored audio clips of "Into You" from the both uncontrolled and controlled surveys.***
Results showed that from 198 responses, just nearly 60% of listeners found stereo mixes as the most natural audio clips and about 60% of listeners found the Auro-To-Binaural mixes as the most colored. Results were from the survey questions (1) Which audio clip sounded the most natural (left column) and (2) Which audio clip sounded the most colored (right column).

Type of Mixing Procedure	Spatial Quality Evaluations		
	Best Spatial Qualities	*Best Externalization*	*Least Immersive*
Stereo	*45.45%*	*46.96%*	*21.72%*
Binaural	35.35%	34.34%	28.79%%
Auro-To-Binaural	19.19%	1970%	49.49%

Table D.1.3 ***Listener evaluation results of the spatial qualities of "Into You" from both the uncontrolled and controlled surveys.***
Results showed that from 198 responses about listeners found that the stereo and binaural mixes had the best spatial qualities and best externalization, respectively, while the Auro-To-Binaural mixes were the least immersive. Results were from the survey questions 1) Which audio clip had the best spatial qualities (left column) (2) Which audio clip had the best externalization (middle column) and (3) Which audio clip was the least immersive?

D.2 Listener evaluation results of "One Last Time" from both the uncontrolled and controlled surveys

Type of Mixing Procedure	Most Preferred vs. Least Preferred	
	Most Preferred	*Least Preferred*
Stereo	68.8%	21.72%
Binaural	27.73%	35.35%
Auro-To-Binaural	8.59%	42.39%

Table D.2.1 ***Listener preferences of "One Last Time" from both the uncontrolled and controlled surveys.***
Results showed that from 198 responses, almost 70% of listeners most preferred stereo mixes and generally least preferred the Auro-To-Binaural mixes. Results were from the survey questions (1) Which audio clip did you most prefer? (left column) and (2) Which audio clip did you least prefer? (right column).

Type of Mixing Procedure	Most Natural vs. Most Colored	
	Most Natural	*Most Colored*
Stereo	73.74%	8.08%
Binaural	19.70%	35.35%
Auro-To-Binaural	6.57%	56.57%

Table D.2.2 ***Listener choices of the Most Natural and Most Colored audio clips of "One Last Time" from the both uncontrolled and controlled surveys.***
Results showed that from 198 responses, over 70% of listeners found stereo mixes as the most natural audio clips and about 60% of listeners found the Auro-To-Binaural mixes as the most colored. Results were from the survey questions (1) Which audio clip sounded the most natural (left column) and (2) Which audio clip sounded the most colored (right column).

Type of Mixing Procedure	Evaluation of Spatial Qualities		
	Best Spatial Qualities	*Best Externalization*	*Least Immersive*
Stereo	53.54%%	50.51%	22.73 %
Binaural	29.29%	29.80%	38.38%
Auro-To-Binaural	17.17%	19.70%	38.89%

Table D.2.3 ***Listener evaluation results of the spatial qualities of "One Last Time" from both the uncontrolled and controlled surveys.***
Results showed that from 198 responses about listeners found that the stereo mixes had the best spatial qualities and best externalization, while the Auro-To-Binaural and binaural mixes were the least immersive. Results were from the survey questions (1) Which audio clip had the best spatial qualities (left column) (2) Which audio clip had the best externalization (middle column) and (3) Which audio clip was the least immersive?

D.3 Listener evaluation results of "Into You" from uncontrolled surveys

Type of Mixing Procedure	Most Preferred vs. Least Preferred	
	Most Preferred	*Least Preferred*
Stereo	52.08%	18.06%
Binaural	34.64%	31.94%
Auro-To-Binaural	15.28%	50.00%

Table D.3.1 ***Listener preferences of "Into You" from the uncontrolled surveys.*** Results showed that from 144 responses, about 50% of listeners most preferred stereo mixes and least preferred Auro-To-Binaural mixes. Results were from the survey questions (1) Which audio clip did you most prefer? (left column) and (2) Which audio clip did you least prefer? (right column).

Type of Mixing Procedure	Most Natural vs. Most Colored	
	Most Natural	*Most Colored*
Stereo	56.25%	13.89%
Binaural	27.78%	28.47%
Auro-To-Binaural	15.97%	57.64%

Table D.3.2 ***Listener choices of the Most Natural and Most Colored audio clips of "Into You" from the uncontrolled surveys.*** Results showed that from 144 responses, that nearly the same amount of listeners found stereo mixes as the most natural and the Auro-To-Binaural mixes as the most colored. Results were from the survey questions (1) Which audio clip sounded the most natural (left column) and (2) Which audio clip sounded the most colored (right column).

Type of Mixing Procedure	Evaluation of Spatial Qualities		
	Best Spatial Qualities	*Best Externalization*	*Least Immersive*
Stereo	45.83%	45.83%	23.61%
Binaural	34.03%	36.11%	28.74%
Auro-To-Binaural	20.14%	18.06%	47.92%

Table D.3.3 ***Listener evaluation results of the spatial qualities of "Into You" from the uncontrolled surveys.*** Results showed that from 144 responses, that nearly the same amount of listeners found that stereo mixes had the best spatial qualities, best externalization and that the Auro-To-Binaural mixes were the least immersive. Results were from the survey questions 1) Which audio clip had the best spatial qualities (left column) (2) Which audio clip had the best externalization (middle column) and (3) Which audio clip was the least immersive?

D.4 Listener evaluation results of "One Last Time" from uncontrolled surveys

Type of Mixing Procedure	Most Preferred vs. Least Preferred	
	Most Preferred	*Least Preferred*
Stereo	72.22%	27.78%
Binaural	18.75%	35.42%
Auro-To-Binaural	9.03%	36.81%

Table D.4.1 ***Listener preferences of "One Last Time" from the uncontrolled surveys.*** Results showed that from 144 responses, over 70% of listeners most preferred stereo mixes and least preferred both the binaural and Auro-To-Binaural mixes. Results were from the survey questions (1) Which audio clip did you most prefer? (left column) and (2) Which audio clip did you least prefer? (right column).

Type of Mixing Procedure	Most Natural vs. Most Colored	
	Most Natural	*Most Colored*
Stereo	73.61%	8.33%
Binaural	18.75%	36.81%
Auro-To-Binaural	7.64%	54.86%

Table D.4.2 Table D.3.2 ***Listener choices of the Most Natural and Most Colored audio clips of "One Last Time" from the uncontrolled surveys.*** Results showed that from 144 responses, over 70% of listeners found stereo mixes as the most natural and nearly over 50% of listeners found the Auro-To-Binaural mixes as the most colored. Results were from the survey questions (1) Which audio clip sounded the most natural (left column) and (2) Which audio clip sounded the most colored (right column).

Type of Mixing Procedure	Evaluation of Spatial Qualities		
	Best Spatial Qualities	*Best Externalization*	*Least Immersive*
Stereo	53.47%%	51.39%	25.00 %
Binaural	29.17%	26.39%	40.28%
Auro-To-Binaural	17.36%	22.22%	34.72%

Table D.4.3 ***Listener evaluation results of the spatial qualities of "One Last Time" from the uncontrolled surveys.***
Results showed that from 144 responses, that nearly the same amount of listeners found that stereo mixes had the best spatial qualities and best externalization and the binaural mixes as the least immersive. Results were from the survey questions 1) Which audio clip had the best spatial qualities (left column) (2) Which audio clip had the best externalization (middle column) and (3) Which audio clip was the least immersive?

D.5 Listener evaluation results of "Into You" from both the controlled surveys

Type of Mixing Procedure	Most Preferred vs. Least Preferred	
	Most Preferred	*Least Preferred*
Stereo	51.85%	16.67%
Binaural	38.89%	27.78%
Auro-To-Binaural	9.26%	55.56%

Table D.5.1 ***Listener preferences of "Into You" from the controlled surveys.***
Results showed that from 54 responses, about 50% listeners most preferred stereo mixes and least preferred the Auro-To-Binaural mixes. Results were from the survey questions (1) Which audio clip did you most prefer? (left column) and (2) Which audio clip did you least prefer? (right column).

Type of Mixing Procedure	Most Natural vs. Most Colored	
	Most Natural	*Most Colored*
Stereo	62.96%	16.67%
Binaural	24.07%	16.67%
Auro-To-Binaural	12.96%	66.67%

Table D.5.2 ***Listener preferences of "Into You" from the controlled surveys.***
Results showed that from 54 responses, about 60% listeners found stereo mixes as the most natural and Auro-To-Binaural mixes as the most colored. Results were from the survey questions (1) Which audio clip sounded the most natural (left column) and (2) Which audio clip sounded the most colored (right column).

Type of Mixing Procedure	Evaluation of Spatial Qualities		
	Best Spatial Qualities	*Best Externalization*	*Least Immersive*
Stereo	44.44%	46.03%	16.67%
Binaural	38.89%	29.63%	29.63%
Auro-To-Binaural	16.67%	24.07%	53.70%

Table D.5.3 ***Listener evaluation results of the spatial qualities of "Into You" from the controlled surveys.***
Results showed that from 54 responses, the spatial qualities were similar for stereo and binaural mixes while stereo mixes had the best externalization. Auro-To-Binaural mixes were the least immersive by 50%. Results were from the survey questions 1) Which audio clip had the best spatial qualities (left column) (2) Which audio clip had the best externalization (middle column) and (3) Which audio clip was the least immersive?

D.6 Listener evaluation results of "One Last Time" from both the controlled surveys

Type of Mixing Procedure	Most Preferred vs. Least Preferred	
	Most Preferred	*Least Preferred*
Stereo	59.26%	5.56%
Binaural	33.33%	35.19%
Auro-To-Binaural	7.41%	59.26%

Table D.6.1 ***Listener preferences of "One last Time" from the controlled surveys.***
Results showed that from 54 responses, about 59% listeners most preferred stereo mixes and least preferred the Auro-To-Binaural mixes. Results were from the survey questions (1) Which audio clip did you most prefer? (left column) and (2) Which audio clip did you least prefer? (right column).

Type of Mixing Procedure	Most Natural vs. Most Colored	
	Most Natural	*Most Colored*
Stereo	74.07%	7.41%
Binaural	22.22%	31.48%
Auro-To-Binaural	3.70%	61.11%

Table D.6.2 ***Listener preferences of "One Last Time" from the controlled surveys.***
Results showed that from 54 responses, that over 70% of listeners found stereo mixes as the most natural and 60% of listeners found the Auro-To-Binaural mixes as the most colored. Results were from the survey questions (1) Which audio clip sounded the most natural (left column) and (2) Which audio clip sounded the most colored (right column).

Type of Mixing Procedure	Evaluation of Spatial Qualities		
	Best Spatial Qualities	*Best Externalization*	*Least Immersive*
Stereo	53.70%	48.15%	16.67 %
Binaural	29.63%	38.89%	38.89%
Auro-To-Binaural	16.67%	12.96%	50.00%

Table D.6.3 ***Listener evaluation results of the spatial qualities of "One Last Time" from the controlled surveys.***
Results showed that from 54 responses, that about 50% of listeners found that stereo mixes had the best spatial qualities and best externalization and that the Auro-To-Binaural mixes were the least immersive. Results were from the survey questions 1) Which audio clip had the best spatial qualities (left column) (2) Which audio clip had the best externalization (middle column) and (3) Which audio clip was the least immersive?

E

Most Preferred Binaural Mixes Based on Engineers

E.1 Evaluation of Most Preferred Binaural Mixes from Controlled Listening Survey

Type of Mixing Procedure	**Rounds One - Three**		
	Round One	*Round Two*	*Round Three*
Stereo	51.14%	83.00%	100.00%
Binaural	28.57%	17.00%	0.00%
Auro-To-Binaural	14.29%	0.00%	0.00%

Table E.1.1 ***Evaluation of Binaural Mixes of Engineer X's "Into You" from Controlled Listening Surveys***
Listeners most preferred stereo mixes as opposed to Engineer X's binaural mixes. Results were from the survey question: Which audio clip did you most prefer?

Type of Mixing Procedure	**Rounds One - Three**		
	Round One	*Round Two*	*Round Three*
Stereo	85.71%	83.33%	100.00 %
Binaural	14.29%	16.67%	0.00%
Auro-To-Binaural	0.00%	0.00%	0.00%

Table E.1.2 ***Evaluation of Binaural Mixes of Engineer X's "One Last Time" from Controlled Listening Surveys***
Listeners most preferred stereo mixes as opposed to Engineer X's binaural mixes. Results were from the survey question: Which audio clip did you most prefer?

Type of Mixing Procedure	**Rounds One - Three**		
	Round One	*Round Two*	*Round Three*
Stereo	42.86%	0.00%	100.00%
Binaural	42.86%	66.67%	0.00%
Auro-To-Binaural	14.29%	33.33%	0.00%

Table E.1.3 ***Evaluation of Binaural Mixes of Engineer Y's "Into You" from Controlled Listening Surveys***
Listeners were most preferential towards Engineer Y's binaural mixes except in Round 3. Results were from the survey question: Which audio clip did you most prefer?

Type of Mixing Procedure	**Rounds One - Three**		
	Round One	*Round Two*	*Round Three*
Stereo	57.14%	66.67%	80.00 %
Binaural	28.57%	16.67%	0.00%
Auro-To-Binaural	14.29%	16.67%	20.00%

Table E.1.4 ***Evaluation of Binaural Mixes of Engineer Y's "One Last Time" from Controlled Listening Surveys***
Listeners were less preferential towards Engineer Y's binaural mixes of "One Last Time". Results were from the survey question: Which audio clip did you most prefer?

Type of Mixing Procedure	**Rounds One - Three**		
	Round One	*Round Two*	*Round Three*
Stereo	0.00%	67.00%	40.00%
Binaural	100.00%	33.00%	40.00%
Auro-To-Binaural	0.00%	0.00%	20.00%

Table E.1.5 ***Evaluation of Binaural Mixes of Engineer Z's "Into You" from Controlled Listening Surveys***
Listeners mostly preferred Engineer Z's binaural mixes except in round two. Results were from the survey question: Which audio clip did you most prefer?

Type of Mixing Procedure	**Rounds One - Three**		
	Round One	*Round Two*	*Round Three*
Stereo	29.00%	67.00%	60.00%
Binaural	71.00%	33.00%	40.00%
Auro-To-Binaural	0.00%	0.00%	0.00%

Table E.1.6 ***Evaluation of Binaural Mixes of Engineer Z's "One Last Time" from Controlled Listening Surveys***
Listeners showed a liking to Engineer Z's binaural mixes, but the preference was less than that of stereo mixes in rounds two and three. Results were from the survey question: Which audio clip did you most prefer?

F

Contributing Factors to Listening Survey Results

F.1 Evaluation of audio clips of "Into You" based on Headphone Reproduction Method for both controlled and uncontrolled listening surveys

Type of Mixing Procedure	Uncontrolled Group		Controlled Group
	Over-ears	*In-ears*	*Sennheiser HD 650 Headphones*
Stereo	22.22%	29.86%	51.85%
Binaural	15.28%	17.36%	38.89%
Auro-to-Binaural	4.17%	11.11%	9.26%

Table F.1.1 ***Most preferred audio clips of "Into You" when comparing headphone reproduction method used for listening Survey***
Listeners from both groups most preferred stereo mixes. Results were from the survey question: Which audio clip did you most prefer?

Type of Mixing Procedure	Uncontrolled Group		Controlled Group
	Over-ears	*In-ears*	*Sennheiser HD 650 Headphones*
Stereo	6.25%	11.81%	16.67%
Binaural	15.97%	15.97%	27.78%
Auro-to-Binaural	19.44%	30.56%	55.56%

Table F.1.2 ***Least preferred audio clips of "Into You" when comparing headphone reproduction method used for listening survey***
Listeners from both groups least preferred the Auro-To-Binaural mixes. Results were from the survey question: Which audio clip did you least prefer?

Type of Mixing Procedure	Uncontrolled Group		Controlled Group
	Over-ears	*In-ears*	*Sennheiser HD 650 Headphones*
Stereo	24.31%	31.94%	62.96%
Binaural	11.81%	15.97%	24.07%
Auro-to-Binaural	5.56%	10.42%	12.96%

Table F.1.3 ***Most natural audio clips of "Into You" when comparing headphone reproduction method used for listening survey***
Participants from both groups found the stereo audio clips as the most natural. Results were from the survey questions (1) Which audio clip sounded the most natural?

Type of Mixing Procedure	Uncontrolled Group		Controlled Group
	Over-ears	*In-ears*	*Sennheiser HD 650 Headphones*
Stereo	3.47%	10.42%	16.67%
Binaural	13.89%	14.58%	16.67%
Auro-to-Binaural	24.31%	33.33%	66.67%

Table F.1.4 ***Most colored audio clips of "Into You" when comparing headphone reproduction method used for listening survey***
Both groups found the Auro-To-Binaural audio clips as the most colored. Results were from the survey question: Which audio clip sounded the most colored?

Type of Mixing Procedure	Uncontrolled Group		Controlled Group
	Over-ears	*In-ears*	*Sennheiser HD 650 Headphones*
Stereo	21.53%	24.31%	44.44%
Binaural	10.42%	23.61%	38.89%
Auro-to-Binaural	9.72%	10.42%	16.67%

Table F.1.5 ***Evaluation of audio clips of "Into You" with the best spatial qualities when comparing headphone reproduction method used for listening survey***
Participants who used in-ears were divided between stereo and binaural mixes as those with the best spatial qualities. Results were from the survey question: Which audio clip had the best spatial qualities

Type of Mixing Procedure	Uncontrolled Group		Controlled Group
	Over-ears	*In-ears*	*Sennheiser HD 650 Headphones*
Stereo	18.75%	27.08%	46.30%
Binaural	12.50%	23.61%	29.63%
Auro-to-Binaural	10.42%	7.64%	24.07%

Table F.1.6 ***Evaluation of audio clips of "Into You" with the best externalization when comparing headphone reproduction method used for listening survey***
Controlled group users found that the stereo audio clips had the best externalization. Results were from the survey question: Which audio clip had the best externalization?

Type of Mixing Procedure	Uncontrolled Group		Controlled Group
	Over-ears	*In-ears*	*Sennheiser HD 650 Headphones*
Stereo	8.33%	15.28%	16.67%
Binaural	13.89%	14.58%	29.63%
Auro-to-Binaural	19.44%	28.47%	53.70%

Table F.1.7 ***Least immersive audio clips of "Into You" when comparing headphone reproduction method used for listening survey***
Auro-To-Binaural mixes were amongst the least immersive audio clips from both groups. Results were from the survey question: Which audio clip was the least immersive?

F.2 Evaluation of audio clips of "One Last Time" based on Headphone Reproduction Method for both controlled and uncontrolled listening surveys

Type of Mixing Procedure	Uncontrolled Group		Controlled Group
	Over-ears	*In-ears*	*Sennheiser HD 650 Headphones*
Stereo	29.17%	43.06%	59.26%
Binaural	6.94%	11.81%	33.33%
Auro-to-Binaural	5.56%	3.47%	7.41%

Table F.2.1 ***Most preferred audio clips of "One Last Time" when comparing headphone reproduction method used for listening survey***
Stereo mixes were the most preferred from both groups. Listeners from both groups most preferred stereo mixes. Results were from the survey question: Which audio clip did you most prefer?

Type of Mixing Procedure	Uncontrolled Group		Controlled Group
	Over-ears	*In-ears*	*Sennheiser HD 650 Headphones*
Stereo	14.58%	13.19%	5.56%
Binaural	13.89%	21.53%	35.19%
Auro-to-Binaural	13.19%	23.61%	59.26%

Table F.2.2 ***Least preferred audio clips of "One Last Time" when comparing headphone reproduction method used for listening survey***
Controlled group participants dramatically found the Auro-To-Binaural mixes as the worst. Listeners from both groups least preferred the Auro-To-Binaural mixes. Results were from the survey question: Which audio clip did you least prefer?

Type of Mixing Procedure	Uncontrolled Group		Controlled Group
	Over-ears	*In-ears*	*Sennheiser HD 650 Headphones*
Stereo	35.42%	38.19%	74.07%
Binaural	2.78%	15.97%	22.22%
Auro-to-Binaural	3.47%	4.17%	3.70%

Table F.2.3 ***Most natural audio clips of "One Last Time" when comparing headphone reproduction method used for listening survey***
Both groups found stereo mixes as the most natural audio clips. Results were from the survey question: Which audio clip sounded the most natural?

Type of Mixing Procedure	Uncontrolled Group		Controlled Group
	Over-ears	*In-ears*	*Sennheiser HD 650 Headphones*
Stereo	2.78%	5.56%	7.41%
Binaural	18.75%	18.06%	31.48%
Auro-to-Binaural	20.14%	34.72%	61.11%

Table F.2.4 ***Most colored audio clips of "One Last Time" when comparing headphone reproduction method used for listening survey***
Both groups found the Auro-To-Binaural mixes as the most colored audio clips. Results were from the survey question: Which audio clip sounded the most colored?

Type of Mixing Procedure	Uncontrolled Group		Controlled Group
	Over-ears	*In-ears*	*Sennheiser HD 650 Headphones*
Stereo	18.75%	34.72%	53.70%
Binaural	12.50%	16.67%	29.63%
Auro-to-Binaural	10.42%	6.94%	16.67%

Table F.2.5 ***Evaluation of audio clips of "One Last Time" with the best spatial qualities when comparing headphone reproduction method used for listening survey***
Controlled group participants found the stereo mixes as those with the best spatial qualities. Results were from the survey question: Which audio clip had the best spatial qualities?

Type of Mixing Procedure	Uncontrolled Group		Controlled Group
	Over-ears	*In-ears*	*Sennheiser HD 650 Headphones*
Stereo	14.58%	36.81%	48.15%
Binaural	12.50%	13.89%	38.89%
Auro-to-Binaural	14.58%	7.64%	12.96%

Table F.2.6 ***Evaluation of audio clips of "One Last Time" with the best externalization when comparing headphone reproduction method used for listening survey***
Stereo audio clips were found to have the best externalization amongst both groups. Results were from the survey question: Which audio clip had the best externalization?

Type of Mixing Procedure	Uncontrolled Group		Controlled Group
	Over-ears	*In-ears*	*Sennheiser HD 650 Headphones*
Stereo	11.11%	13.89%	16.67%
Binaural	16.67%	23.61%	33.33%
Auro-to-Binaural	13.89%	20.83%	50.00%

Table F.2.7 Least immersive audio clips of "One Last Time" when comparing headphone reproduction method used for listening survey
Controlled group participants found that the Auro-To-Binaural mixes as the least immersive while uncontrolled participants found binaural mixes as the least immersive. Results were from the survey question: Results were from the survey question: Which audio clip was the least immersive?

F.3 Evaluation of audio clips of "Into You" based on music background for both controlled and uncontrolled listening surveys

Type of Mixing Procedure	Participant Backgrounds		
	Music Technologists	*Musicians*	*Other*
Stereo	29.80%	5.05%	17.17%
Binaural	19.19%	3.54%	11.62%
Auro-to-Binaural	4.04%	2.02%	7.58%

Table F.3.1 ***Most preferred audio clips of "Into You" based on music background for both controlled and uncontrolled listening surveys***
All participant backgrounds most preferred stereo mixes. Results were from the survey question: Which audio clip did you most prefer?

Type of Mixing Procedure	Participant Backgrounds		
	Music Technologists	*Musicians*	*Other*
Stereo	6.06%	3.03%	8.59%
Binaural	17.17%	2.53%	11.11%
Auro-to-Binaural	29.80%	5.05%	16.67%

Table F.3.2 ***Least preferred audio clips of "Into You" based on music background for both controlled and uncontrolled listening surveys***
Participants least preferred the Auro-To-Binaural mixes. Results were from the survey question: Which audio clip did you least prefer?

Type of Mixing Procedure	Participant Backgrounds		
	Music Technologists	*Musicians*	*Other*
Stereo	33.33%	6.06%	18.69%
Binaural	14.14%	2.02%	10.61%
Auro-to-Binaural	5.56%	2.53%	7.07%

Table F.3.3 ***Most natural audio clips of "Into You" based on music background for both controlled and uncontrolled listening surveys***
Music technologists most preferred stereo mixes more than other participants. Results were from the survey question: Which audio clip sounded the most natural?

Type of Mixing Procedure	Participant Backgrounds		
	Music Technologists	*Musicians*	*Other*
Stereo	*5.56%*	3.54%	5.56%
Binaural	*14.14%*	2.53%	8.59%
Auro-to-Binaural	*33.33%*	4.55%	22.22%

Table F.3.4 ***Most colored audio clips of "Into You" based on music background for both controlled and uncontrolled listening surveys***
Auro-To-Binaural mixes were identified as the most colored audio clips amongst all backgrounds. Results were from the survey question: Which audio clip sounded the most colored?

Type of Mixing Procedure	Participant Backgrounds		
	Music Technologists	*Musicians*	*Other*
Stereo	28.28%	2.53%	14.65%
Binaural	15.66%	5.05%	14.65%
Auro-to-Binaural	9.09%	3.03%	7.07%

Table F.3.5 ***Evaluation of audio clips of "Into You" with the best spatial qualities based on music background for both controlled and uncontrolled listening surveys***
Other participants were divided between binaural and stereo as the audio clips with the best spatial qualities. Controlled group participants found the stereo mixes as those with the best spatial qualities. Results were from the survey question: Which audio clip had the best spatial qualities?

Type of Mixing Procedure	Participant Backgrounds		
	Music Technologists	*Musicians*	*Other*
Stereo	26.26%	3.54%	16.16%
Binaural	14.65%	5.05%	14.65%
Auro-to-Binaural	12.12%	2.02%	5.56%

Table F.3.6 ***Evaluation of audio clips of "Into You" with the best externalization based on music background for both controlled and uncontrolled listening surveys***
Stereo mixes were identified to have the best externalization however, binaural mixes fared closely behind. Results were from the survey question: Which audio clip had the best externalization?

Type of Mixing Procedure	Participant Backgrounds		
	Music Technologists	*Musicians*	*Other*
Stereo	7.58%	4.55%	9.60%
Binaural	17.17%	2.538%	9.09%
Auro-to-Binaural	28.28%	3.54%	17.68%

Table F.3.7 ***Least immersive audio clips of "Into You" based on music background for both controlled and uncontrolled listening surveys***
Auro-To-Binaural mixes were identified as the clips with the least externalization. Results were from the survey question: Which audio clip was the least immersive?

F.4 Evaluation of audio clips of "One Last Time" based on music background for both controlled and uncontrolled listening surveys

Type of Mixing Procedure	Participant Backgrounds		
	Music Technologists	*Musicians*	*Other*
Stereo	29.17%	13.19%	29.86%
Binaural	4.86%	1.39%	12.50%
Auro-to-Binaural	1.39%	0.00%	7.64%

Table F.4.1 ***Most preferred audio clips of "One Last Time" based on music background for both controlled and uncontrolled listening surveys***
Music technologists and others most preferred stereo mixes while musicians were divided between binaural and stereo mixes. Results were from the survey question: Which audio clip did you most prefer?

Type of Mixing Procedure	Participant Backgrounds		
	Music Technologists	*Musicians*	*Other*
Stereo	10.42%	3.47%	13.89%
Binaural	11.81%	5.56%	18.06%
Auro-to-Binaural	13.19%	5.56%	18.06%

Table F.4.2 ***Least preferred audio clips of "One Last Time" based on music background for both controlled and uncontrolled listening surveys***
There is no significant majority over the least preferred audio clips.

Type of Mixing Procedure	Participant Backgrounds		
	Music Technologists	*Musicians*	*Other*
Stereo	33.33%	11.11%	29.17%
Binaural	1.39%	1.39%	15.97%
Auro-to-Binaural	0.69%	2.08%	4.86%

Table F.4.3 ***Most natural audio clips of "One Last Time" based on music background for both controlled and uncontrolled listening surveys***
All participants found that the stereo mixes sounded the most natural. Results were from the survey question: Which audio clip sounded the most natural?

Type of Mixing Procedure	Participant Backgrounds		
	Music Technologists	*Musicians*	*Other*
Stereo	0.00%	0.00%	8.33%
Binaural	17.36%	4.86%	14.58%
Auro-to-Binaural	18.06%	9.72%	27.08%

Table F.4.4 ***Most colored audio clips of "One Last Time" based on music background for both controlled and uncontrolled listening surveys***
Participants were divided between binaural and Auro-To-Binaural mixes as the most colored mixes. Results were from the survey question: Which audio clip sounded the most colored?

Type of Mixing Procedure	Participant Backgrounds		
	Music Technologists	*Musicians*	*Other*
Stereo	20.83%	8.33%	24.31%
Binaural	9.72%	3.47%	15.97%
Auro-to-Binaural	4.86%	2.78%	9.72%

Table F.4.5 ***Evaluation of audio clips of "One Last Time" with the best spatial qualities based on music background for both controlled and uncontrolled listening surveys***
Stereo mixes were identified as those with the best spatial qualities. Results were from the survey question: Which audio clip had the best spatial qualities?

Type of Mixing Procedure	Participant Backgrounds		
	Music Technologists	*Musicians*	*Other*
Stereo	17.36%	9.03%	25.00%
Binaural	9.72%	2.78%	13.89%
Auro-to-Binaural	8.33%	2.78%	11.11%

Table F.4.6 ***Evaluation of audio clips of "One Last Time" with the best externalization based on music background for both controlled and uncontrolled listening surveys***
Stereo audio clips were identified as those with the externalization. Results were from the survey question: Which audio clip had the best externalization?

Type of Mixing Procedure	Participant Backgrounds		
	Music Technologists	*Musicians*	*Other*
Stereo	5.56%	4.17%	15.28%
Binaural	15.28%	6.25%	18.75%
Auro-to-Binaural	14.58%	4.17%	15.97%

Table F.4.7 ***Least immersive audio clips of "One Last Time" based on music background for both controlled and uncontrolled listening surveys***
Binaural mixes were identified as the clips with the least externalization. Results were from the survey question: Which audio clip was the least immersive?