DEGREE PROJECT

The effects of distortion

Investigating how different types of distortion affect timbral attributes and subjective preference

André Waldton Lézin

Audio Technology, bachelor's level
2020

Luleå University of Technology
Department of Arts, Communication and Education

Abstract

The effects of distortion has been investigated prior to this study, however most of these studies focus on the objective physicalities of a certain type of distortion or they might apply distortion in static amounts to examine effects of loudspeaker distortion. Objectively the varying types of distortion may be different, however there are little explanations on how these types subjectively might sound different. This study aimed to investigate how subjective preference and perception of the timbral attributes warmth and roughness may vary between types of distortion, and if there was a pattern between these using three different types of distortion (zero-crossing, solid state and tube), applied at two different levels (high and low) and to two different instruments (guitar and vocals). The outcome indicated that subjects most prefer tube distortion and that this distortion was considered to provide the most amounts of warmth while also the least amounts of roughness. There were also interaction effects indicating guitar being less sensitive about the level of distortion while being more sensitive about the type of distortion for the measures of preference and amounts of roughness, when compared to vocals.

Table of Contents

Acknowledgement

I would like to thank my classmates for providing an inspiring environment the past three years. I would also like to thank my family for enduring my excruciatingly long back-and-forths about an area somewhat unknown to them. Thank you to Cleopatra Pike for responding to these back-and-forths as my supervisor. Also a big thank you to Nyssim Lefford and Jan Berg for supporting, while also questioning, any and all ideas, issues and dilemmas. Finally, a gentle pet to the cats of Ankarskatavägen for being cats.

1. Introduction

The perception and preference of the sound of distortion – the warping of a transmitted signal – might vary a lot depending on who is responding. A musician might say it is the sound of rock’n’roll guitar, a recording engineer might say it is the effects of a signal with too much gain going into a preamplifier, a mixing engineer might say it is the grit of the track and a mastering engineer might say it is the artifacts brought with downsampling or reducing bit depth. However, for researchers and manufacturers it might refer to undesired deviations and system faults in system design, such as loudspeakers, and for consumers of both music and audio products it might not mean anything at all. Altough differences in terms of objective characteristics may be of interest, this study looked into subjective characteristics, particularly the preference for different types of distortion and the way changes in timbre were linked to these preferences.

1.1. Background

Distortion is often thought of as an undesirable biproduct created from subpar analog or digital circuits, whether it is the microphone, preamplifiers, loudspeakers or other tools of the trade. To be specific (since there are multiple ways of regarding distortion), what is discussed is distortion which adds extra frequency components, sometimes called nonlinear distortion (Case, 2007). Nonlinear distortion occurs even in our auditory systems (Voishvillo, 2007) but we do not think of this the same way we think of the distorted sound emitting from a rock guitar or metal vocals, which might be even more distorted than our auditory system. This is most commonly measured in THD (Total Harmonic Distortion), a method which essentially measures and compares the input signal and output signal of a DUT (Device Under Test) to establish the ratio of extra frequency components relating to the fundamental frequency. However, there is more to nonlinear distortion than this. While an analog circuit might introduce the same distortion effect to each device, the timbral change that is perceived will depend on the device (e.g. the frequency response of the DUT), making the effect sound different on different devices.

Voishvillo (2007) explains that different types of nonlinear distortions such as hard-clipped distortion and zero-crossing distortion (crossover distortion) may differ significantly in terms of positive/negative listener response. McMullin, Brunet and Wang (2019) list attributes given by listeners in response to nonlinear distortion applied to musical pieces of varying genres. Some of the attributes listeners used were roughness, shrillness and buzzing. Herzog (2009) investigates the impact of nonlinear distortion on the psychoacoustical attributes loudness, sharpness, roughness, tonality and sensory pleasantness (see 1.3.2. Impact on loudness, and 1.5. Audibility and assessment of distortion).

How then can we properly evaluate the effects of distortion? THD does not account for what extra frequency components are added, only in what amount regardless of the perceived impact on sound quality (see 1.4. THD as a measure). Some forms of musical distortions we might find pleasing, and want to create on purpose. It is not necessarily the amount of THD that determines whether distortion will be deemed as pleasing or not, rather the type of distortion may be more important, as there are several different types. Further, the amount of distortion and how it is implemented, and on what instrument, might dictate the preferability of distortion. The aim and purpose of this bachelor thesis was to explore subjective evaluations of timbral attributes and preference for different types of distortion to further develop understanding of distortion.

1.2. Desirable distortion

Most sound engineers (music, films, games, sound design) would agree that distortion may not be all bad, but is dependent on how much, when and what kind of distortion is introduced. For example, many modern preamplifiers, microphones and compressors are reintroducing or mimicing the effect of having vacuum tubes as a part of the analog circuitry, a technology of old. This is somewhat similar to the desired distortion introduced by the tubes, which Poss (1998) describes as the nostalgic timbral richness and density of rock and soul vocals and the surest way to the speficic kind of distortion which is inherently pleasant to the human auditory system. He further argues that distortion might be considered as ”simply another dimension of timbre”.

Michael Beinhorn (via; Owsinski, 2005, p. 261) explains the similarities between a symphonic orchestra and multiple layered distorted guitars, claiming that there from a harmonic standpoint is no real difference between the two and that following that logic, close miking a string section to pick up the grit which has some characteristics of distortion we like would produce the same psychological effect as that of a distorted electric guitar.

1.3. Distortion and psychoacoustics

1.3.1. Types of distortion

When talking about distortion there is a common misconception that all distortion is harmonic in some way. Harmonic distortion may be one type of distortion, but the harmonics may be odd or even multiples of the fundamental frequency or both in varying degrees (Case, 2007). There are also intermodulated distortion, which is the extra frequency components resulting of two fundamental frequencies modulating eachother. As there is nonlinear distortion there is linear distortion, with the difference that nonlinear distortion adds additional frequency components while linear distortion does not (such as distortion of phase).

There is also zero-crossing distortion (also known as crossover distortion), which is distortion occuring at the intersection of positive and negative value of a frequency, and quantization distortion occuring in the conversion of analog-digital or digital-analog audio (i.e. approximation of real numbers to a finite set of values). This study focused on three different types of distortion; zero-crossing distortion, solid-state germanium distortion (nonlinear with emphasis on odd harmonics) and tube distortion (nonlinear with emphasis on even harmonics).

1.3.2. Impact on loudness

Loudness can be described as the attribute of human auditory system which determines how loud or quiet a sound is, not only dependent on the level and frequency but also the bandwidth of the sound. This means that nonlinear distortion increases loudness as the added frequency components will increase the bandwidth of the sound (Herzog, 2009). There is also a connection between increased loudness for types of nonlinear distortion adding mostly higher frequencies and various loudness models indicating an increased sensitivity of the ear toward the mid-high to high frequencies (Fastl & Zwicker, 2007). When assessing different types of distortion the impact of loudness has to be accounted for, as differences in loudness might affect the outcome of both preference ratings and perception of timbral attributes. An increase of preference due to application of a particular distortion should not be confounded with an increase of preference because the sound got louder after the distortion was applied.

To further understand the impact of nonlinear distortion on loudness and other psychoacoustic qualities we must first understand the human auditory system to some degree. There are several scales used to explain the human auditory system and the most common is the Bark-scale (Herzog, 2009). The scale ranges from 50 Hz to 16 kHz, divided into 24 critical bands. These bands are described as a spreading in the human auditory systems basilar membrane dependent on the spectrality and temporality of an incoming audio signal. The spectrality of the signal excites areas of the membrane correlating with the maximum excitation, and the bandwidth of that area may be considered an auditory filter, or critical band. Distortion which adds extra frequency components increases the spectral range of the signal which, as the bandwidth expands outside a critical band, affects the timbre of the sound. If we are to analyze any audio signal we must consider if the signal is within the spectra of one of these bands or multiple as the perceived loudness of the signal remains the same until it expands outside of that band into additional bands (Voishvillo, 2007).

1.3.3. Auditory masking of distortion

There are several circumstances to how nonlinear distortion integrates with a signal, making the signal be perceived as more or less ”natural”, or making the distortion more or less audible. One of these circumstances is the spectrality and temporality of the signal and in what way this affects masking of the distortion. Masking is the supression of signals with lower amplitude than a previous or following signal within the same frequency band. These masking bands gets wider the lower the frequency and become progressively asymmetrical at higher levels of masking (Voishvillo, 2007). Voishvillo (2007) explains that the masking acts on both spectral and temporal information, meaning a signal may be masked both pre and post the masker (backwards and forwards masking). This suggests that nonlinear distortion producing mostly high frequencies might be more audible than a similar distortion producing mostly lower frequencies due to the masking bands getting progressively narrower at high frequencies.

Voishvillo (2006) demonstrates how a hard clipping of a signal producing over 20 % THD had higher audio quality than a zero-crossing distortion producing about 3 % THD, which has its explanation in the peaks affected by the hard clipping being rare and the distortion products masked. This suggests that not only is the frequency range of the nonlinear distortion important to consider, but the order and type of nonlinearity also has a big impact on the perception and masking of distortion.

1.4. THD as a measure

THD is measured by repeatedly inputting a single sinusoid at different frequencies and levels and comparing the input with the output of the DUT. Any additional frequency components are explained to be a result of nonlinear distortion, which are compared to the fundamental frequency. The ratio of the summed power of the extra frequency components to the power of the fundamental frequency is the resulting THD. However as Voishvillo (2007) argues, the effects of critical bands and masking along with the nonlinear process of the basilar membranes excitation of hair cells, the limiting characteristics of the stapedius reflex, the mechanical nonlinearity of the ossicles and the inner ears compressive boost of low-level signals makes the human auditory system much more complex than any traditional analyzer of distortion such as THD. This points to the questionable relevancy of using a single static sinusoidal tone when assessing sound quality and distortion as a static audio signal and a time-varying complex audio signal differ significantly from each other, and as a single sinusoidal tone will not properly account for the critical bands or masking of music and speech (Voishvillo, 2007). Thus we must question the measure of THD when making subjective assessments.

Herzog (2009) further suggests that with the development of audio-codecs like MP3 the need for a more subjective measure than the technical THD arose. Examples of these are the GedLee Metric, DS and R_{nonlin} which quantify regular distortion occuring in analog systems with differentiable curves, which all provides a single number as a result. The differential curves makes a big part of the difference of these measures compared to THD. However, Herzog (2009) argues that these measures does not account for the irregular distortions of quantization in digital systems. Therefore Herzog proposes a measurement method to determine the audible effects of nonlinear distortions by means of the psychoacoustical qualities loudness, sharpness, roughness and tonality. This measurement method computes psychoacoustical measures by separating linear and nonlinear processing and utilising psychoacoustical models. The results of his proposed measurement method further reinforces the idea that THD may not correlate with subjective perception of nonlinear distortion.

1.5. Audibility and assessment of distortion

Herzog (2009) concludes that the thresholds of audible nonlinear distortion is strongly influenced by the source of the sound, an instrument with a lot of harmonics will have a higher threshold than the sound of an instrument with few dominant spectral components. The impact of the type of distortion is solidified, for example quantization distortion might produce artifacts over the whole audible frequency range which in turn would have a lower threshold of audibility than regular nonlinear distortion. Herzog (2009) did however not include subjective testings, but rather proposed a measurement method which would correlate better with psychacoustical measures.

Herzogs (2009) study showed that nonlinear distortions of 10 % THD or quantization with 8 - 10 bits can be considered as significantly influencing said psychoacoustic measures. Sharpness of a signal with predominant low-frequency content will increase. In the case of regular distortion the influence of roughness is the greatest at signals with high amplitude as the distortion products are in correlation with the original signal. This is not the case for quantization distortion as it produces distortion with near constant power which would have higher influence at signals with low amplitude. Tonality is reduced with the introduction of additional frequency components produced by nonlinear distortion which is most noticeable in signals with high original tonality. Finally, Herzog also points out that these measures do not describe the quality of music as a musical piece of tonal character may not sound better than one of more noise like character.

McMullin et al. (2019) evaluated listeners perception of how and when nonlinear distortion occured. This was done by modelling a loudspeaker known to produce audible nonlinear distortion at peak excursion and inputing a signal generating 8 % noncoherent distortion (NCD) as this level produces audible but not excessive nonlinear distortion. Musical pieces of varying genres was played back through the system and the subjects was asked to elicit attributes for the distorted signal compared to a reference audio file. The results of the experiment suggests that the audibility of nonlinear distortion is very track dependent while being less audible for speech and TV/film tracks, and it was the most common for the subjects to perceive distortions in vocals (music) and bass.

McMullin et al. (2019) also found that the descriptors ”roughness”, ”clipping” and ”buzzing” was commonly associated with nonlinear distortion of bass elements while ”shrillness”, ”buzzing” and ”modulation” was associated with nonlinear distortion of vocal elements. Note here that a distinction between modulation and roughness is made, while Herzog (2009) defines roughness as modulation. The subjects of McMullin et al. (2019) study tended to notice modulation in spectrally dense parts of a track, such as when instruments was added. Roughness was the most commonly perceived descriptor, both in sustained and intermittent parts of a track.

1.6. Aims and purpose

As described above, some research has been done on the timbral aspects associated with distortion, but research covering preference between different types of distortion and how timbral changes are linked to this has not been found. It has been described how previous research has examined the varying perceptions of distortion and some considerations when dealing with these. While types of distortion has been described to sound and behave in different ways, the subjective preferences and timbre of these are yet to be made clear. This study aimed to examine how different types of distortion relate to timbral perception and preference, with a view to determining whether differences in timbre were linked to differences in preference. Although distortions can affect the spatial and temporal qualities of sounds, this study focused more on how different distortions affect timbre, and how timbral differences were linked to preference. Ideally this study would have investigated the perception of distortion in an ecologically valid manner (e.g. by running a listening test in a live music enviroment), however for the purpose of scientific clarity measurements in a controlled enviroment on a limited amount of signals were conducted.

The independent variables used was instrument, type of distortion and level of distortion. The author hypothesized that there would be an audible difference between the types of distortion and that the preference ratings would reflect that. The zero-crossing (ZC) distortion was hypothesized to be the least preferred type of distortion.

2. Method

In order to collect data for the study two sets of experiments in the form of listening tests was conducted. A prestudy was used to verify that there was a difference in preference between stimuli which suggested that the types of distortion were audible different. The first of the experiments asked participants to rate preference between samples and evaluate what influenced their preference ratings. Preference ratings was the main outcome variable, however qualitative data was also gathered in the form of reports on what influenced the subjects preference ratings. This qualitative data from the subjective evaluations was analyzed and categorized into attributes used for the second experiment where the same participants were asked to rate the value of the given timbral attributes. This in turn made linking of subjective preference and subjective perception of timbral attributes for each sample possible.

Both experiments were conducted running listening tests using a laptop and headphones and with students attending the musician or audio engineering program at Luleå University of Technology. The first experiment examined the dependent variable preference using three independent factors; instrument, type of distortion and level of distortion. The instrument had two factor levels, the types of distortion had three factor levels and the level of distortion had two factor levels. The second experiment measured the effect of the same three independent variables on two measures of timbre: warmth and roughness. Conclusions examined whether different types of distortions resulted in different preferences and different perceptions of warmth and roughness, and whether there was a link between preference and warmth or roughness. This was examined using a full factorial design meaning there was a crossing of every subcombination.

There were several evaluative tools which could have been considered for running a listening test, such as ABX, BS.1116 or interviews. However this study was based on a modified MUSHRA interface (see 2.5.1 MUSHRA interface) which used sliders on a scale of 0 – 100. This allowed subjects to accurately rate stimuli according to their will, leaving less room for arbitrary conclusions while providing average ratings represented by a number. Further description of the modified MUSHRA is given in 2.5.1. MUSHRA interface.

Table 2a – Combinations of independent variables

	ZC	***E***	***T***	***ZC***	***E***	***T***
Gtr	Gtr ZC Lo	Gtr E Lo	Gtr T Lo	Gtr ZC Hi	Gtr E Hi	Gtr T Hi
Vox	Vox ZC Lo	Vox E Lo	Vox T Lo	Vox ZC Hi	Vox E Hi	Vox T Hi
	Lo	***Lo***	***Lo***	***Hi***	***Hi***	***Hi***

2.1. Equipment and environment

Both experiments was conducted using a PC laptop and the software STEP, developed by Audio Research Labs. This software did not support 32 bit float meaning the stimuli had to be reduced to 24 bit. This bitreduction was done using dither which might have affected the outcome. The subjects monitored using Sennheiser HD650 headphones which are commonly accepted as within industrial standards. When choosing the monitoring, loudspeakers were considered but turned down due to the impact room acoustics has on sound. Since the choice of environmental space for the experiments was not without limit and repeatability was to be accertained, headphones was chosen as monitoring. The experiments took place in Piteå School of Music, Luleå University of Technology (LTU).

2.2. Target population

The target population was people within professions close to, or relating to, music and audio engineering. The reasoning for this was that the subjects had to name and rate timbral attributes of differently distorted signals which required some previous experience with the sounds of different distortions. Including people with no experience might have increased the range of understanding of how distortion is perceived by everyday listeners but would also had increased the variation of value ratings of timbral attributes. By excluding everyday listeners with no experience the study aimed to specify how pracitioners perceive and prefer different types of distortion. This might also have increased the coherency of preference ratings and agreement of timbral attributes amongst subjects. Some demographic data was collected.

The first experiment had 26 participants whereof 21 (81 %) were audio engineers and 5 (19 %) were musicians. The second experiment would have had the same participants, however two subjects did not participate resulting in 19 (79 %) audio engineers and 5 (21 %) musicians. The subjects ages ranged from 19 to 43 with a mean of 23.85, with musical preferences such as metal, CCM (contemporary christian music), electronica, indie pop, soul, jazz and rock amongst others.

2.3. Prestudy

Prior to the main study a prestudy was conducted to establish which of the types of distortion are likely to result in audible differences, at what levels the timbral effects of these types was barely audible and at what levels these differences was clearly audible, setting the foundation for the high and low levels of distortion used in the main study. This was done by preparing several combinations of each type of distortion at varying levels and to ask four subjects from the population pool who would not participate in the main study to rate preference between the stimuli and to, as part of the qualitative data collection, explain what motivated their ratings. If most of the subject could not differentiate between the types of distortion the level of distortion had to be higher, while if most of the subjects could barely identify some timbral attributes of the distortion the level of distortion was just right for the lower threshold. To determine the higher threshold of amount of distortion the subjects were asked to choose a stimuli with a level of distortion where the stimuli sounded clearly distorted but not overwhelmingly so.

The results of this prestudy showed that zero-crossing distortion was clearly different from the other types of distortion regardless of level. The E style (solid state germanium distortion emulated from Chandler TG Channel) and T style (tube distortion emulated from Thermionic Culture Culture Vultures triode setting) of Soundtoys (2019) differed the most, which was also true to the understanding that solid state distortion and tube distortion sound different. These types of distortion as well as the low and high levels for these provided the foundation for the rest of the study.

2.4. Creating stimuli

2.4.1. Source samples

When deciding on the source samples to which distortion is being applied the EBU (2008) *Sound Quality Assessment Material* (SQAM), recordings of music commonly used for research assessing audio quality, was considered but turned down in favor of recording original material. This allowed any errors of the recording procedures to be considered when analyzing the outcome as well as providing a more controlled experiment environment where the genre, chord progression, melody, choice of microphones, pre amplifiers and tonality could be adjusted as desired.

The source samples were to be performed in a genre with neutral expectations on the use of distortion, captured through commonly used microphones and standardized pre amplifiers. With no expectations on the use of distortion the subjects would be encouraged to motivate their preferences with less bias towards genre. However, this meant that any results would not show how genre and expectations on the use of distortion within that genre affect both subjective preferences and the perception of timbral attributes of the distortion. For example, an electric amplified guitar with distortion performing rock music might be considered more within the subjects expectations, and thus distortion might be more preferred, compared to if the same guitarists with the same guitar and the same settings had played R'n'B music. The genre chosen for the stimuli was pop/rock ballad.

By recording the source samples through commonly used microphones and pre amplifiers the tracks were closer to how the instruments are commonly heard in music, diminishing the chance that subjects would prefer a sample more, or less, based on lack of recognition of commonly heard combinations of instrument, microphone and preamplifier. Recording using equipment with transparent characteristics would on one hand put emphasis on the differences of timbre brought by the types of distortion, while on the other hand put the sounds in a more controlled and less ecologically valid environment. Therefore microphones and preamplifiers with neutral frequency response somewhat decreases ecological validity. For this thesis the main focus of creating the stimuli was to ensure neutral expectations of distortions as well as providing ecological validity.

The guitar and vocals each performed an original piece, different for each instrument, within the same genre. In order to have the different programs influence the results as little as possible the subjects were asked to disregard quality of performance and choice of melody/chords (see 6.2. (Appendix 2) Instructions and demographics for test one). The first of the source samples was to be an instrument which is commonly associated with distortion where the distortion might not be percieved as odd or disturbing, therefore electric guitar played through an amplifier was selected. The guitar, a Fender Telecaster, was played through a Fender Twin Reverb amplifier and recorded using a Shure SM57 microphone positioned close to the amplifier cone and slightly off-axis. The signal was amplified using Avid Pro Tools PRE and recorded in Avid (2019) Pro Tools using Avid Pro Tools HD I/O.

The second of the source samples intended to make the effects of the types of distortion more noticable than for the electric amplified guitar. McMullin et al. (2019) concluded that distortion was most noticable on bass and vocals. Of the two, vocals was selected for this study as it in many ways differed more from guitar than bass did and the difference of the types of distortions interactions with the source samples was sought to be evaluated. The recording was of a female vocalist singing an original piece of music in a similar genre to the piece performed by the guitarist, captured through a Neumann U89 microphone set to cardioid pick-up pattern. The signal was recorded using the same preamplifier, interface and DAW as the recording of the guitar. Both the guitar and vocals was recorded with a sample rate of 48 kHz with a bit depth of 32 bit float.

2.4.2. Applying distortion

There were several factors to consider when applying the distortion to the source samples. One was how distortion impacts the signal when applied post recording as opposed to earlier in the signal chain. For vocals, distortion is not commonly applied pre preamplifier, however it is for guitar. Applying the distortion post recording might have made it less integrated with the sound of the guitar than if the distortion was applied in-between the guitar and the amplifier. This might have affected the outcome of the study, however it was considered that in order to provide a more controlled environment the distortion would be applied post recording.

The stimulus were created using Soundtoys (2019) which had five types of distortion modelled from the famous preamps and distortion devices Ampex 350, Chandler TG Channel, Neve 1057 and Thermionic Culture Culture Vulture (triode setting and pentode setting), and Splice (2019) which had a function for introducing zero-crossing distortion. Out of the five types of distortion provided by Soundtoys (2019) the E (Chandler TG Channel) and T (Thermionic Culture

Culture Vulture triode setting) types were chosen as well as the ZC (zero-crossing, called zero-square in the software) provided by Splice (2019). These were chosen because they were deemed to differ the most from eachother based on the results from the prestudy as well as also being commonly used creatively in a musical context.

These distortions were applied to the recorded (i.e. clean) stimuli at two different levels, Lo (low) and Hi (high), using individual tracks in Pro Tools (2019) for each distortion type receiving the dry input signal (see 6.1. (Appendix 1) Signal flow when creating stimuli). These tracks were then mixed with the dry signal to provide a manually controllable dry/wet (clean/distorted) ratio while keeping the software plug-in settings drive and type identical for each sample. The drive, meaning the software plug-in internal input gain, resulting in additional distortion products to the signal, was set to 7/10 for Soundtoys (2019) and 10 % for Splice (2019). These relatively high settings were chosen so that subjects could identify the differences between types of distortion. This was confirmed by the prestudy.

This way of controlling the dry/wet ratio using multiple tracks allowed more precise observations of the difference in ratios between stimuli as the exact dB value of the distortion tracks could be identified. This also allowed for recreation of the study as well as preventing altering of the plug-in settings when creating the Lo and Hi levels of the stimulus. When creating the stimuli considerations to the aim of the study had to be made, to investigate the difference in preference and timbre between the types of distortion. In order to do this the prominence of certain types had to be compensated for, resulting in some of the tracks having significantly higher level than other. These level compensations were done in order to make the timbre of the distortions noteworthy rather than solely providing an increase of distortion. The exact values was a result of estimation by ear as well as the prestudy. For a visual representation of the signal flow, please look at 6.1. (Appendix 1) Signal flow when creating stimuli. The dB of the tracks was according to table 2.4.2a where the dry signal was at 0 dB using the PPM Nordic scale of +12 to -140 (∞) dB.

Table 2.4.2a – dB values of distortion tracks

	ZC Lo	E Lo	T Lo	ZC Hi	E Hi	T Hi
Gtr	- 22.0 dB	- 7.0 dB	- 2.0 dB	- 8.0 dB	+ 6.0 dB	+ 11.0 dB
Vox	- 25.0 dB	- 13.0 dB	- 8.0 dB	- 13.0 dB	+ 2.0 dB	+ 11.0 dB

In addition to the dB values of each distortion track (i.e. the dry/wet ratio), THD measurements of each stimuli were made. The relevance of THD has been made clear (see 1.4. THD as a measure) but this further increased repeatablity of the study and might provide some valuable insight in using THD as a measure of distortion. These measures were conducted by replacing the source tracks in the signal flow with a sinusoidal tone at 1 kHz while keeping the rest of the signal flow the same, including distortion plug-in settings and level compensation. The resulting distorted sinusoidal tracks were analyzed using Neutrik TT402. The results can be observed in table 2.4.2b.

Table 2.4.2b – THD+N measurements using a sinusoidal tone at 1 kHz

	ZC Lo	E Lo	T Lo	ZC Hi	E Hi	T Hi
(Gtr)	4 %	10 %	6 %	10 %	16 %	8 %
(Vox)	4 %	8 %	5 %	7 %	14 %	8 %

2.4.3. Loudness matching

Due to the variations in distortion track levels (dry/wet ratio) in order to provide equally audible timbral morphs introduced by the types of distortions rather than only a difference in the amounts of distortion (see 2.4.1. Applying distortion) the resulting stimuli were of varying levels which might affect both preferences and perceived timbre (see 1.3.2. Impact on loudness). Making the stimuli have equal loudness solved these issues. This was done by applying Waves (2019) WLM Plus Loudness Meter to the master bus and adjusting the level of each track to the designated value of -14 LUFS long term. The results of the adjustments can be observed in table 2.4.3a. For a visual representation of the signal flow, please look at 6.1. (Appendix 1) Signal flow when creating stimuli.

Table 2.4.3a – Loudness matching of the stimuli

	ZC Lo	E Lo	T Lo	ZC Hi	E Hi	T Hi
Gtr	+ 7.6 dB	+ 2.7 dB	+ 1.0 dB	+ 9.7 dB	- 1.5 dB	- 4.9 dB
Vox	+ 6.5 dB	+ 3.5 dB	+ 2.0 dB	+ 0.5 dB	- 4.1 dB	- 9.7 dB

2.5. Experiment design

2.5.1. MUSHRA interface

The traditional MUSHRA interface is a computer based interface that can accept and present sounds to participants in a chosen, or randomized, order and allow the participants to rate these sounds. It consists of several buttons connected to sliders on a scale of 0 – 100 as well as a reference button, hidden reference and anchors. The buttons plays the different stimulus which are to be examined and rated on the corresponding slider. The reference button is not connected to a slider but would play exactly the same sound as the hidden reference which is to be rated at 100. The hidden reference in turn is the same stimulus that is assigned one of the buttons meaning the subject would have to figure out which of the stimuli is exactly the same as the sound playing from the press of the reference button. The subject is expected to give that stimuli a rating of 100. This way you could figure out which of the subjects could hear an actual difference in the sounds and any who could not identify the hidden reference would be ruled out of the results. The anchors would act in a similar way by being stimuli estimated to be rated the lowest.

Since the original MUSHRA interface was designed to rate audio quality of codecs and not subjective evaluations of preference the interface was tweaked for this study. The buttons and sliders remained the same as well as having a hidden reference, although not in the same way. Instead of having a reference button and have the subjects rate the hidden reference at 100 the hidden reference was a copy of another stimuli existing amongst the buttons and sliders. This means there was no reference button playing a sound which the subjects could expect to be rated the highest, while still providing a way to see what subjects could hear an actual difference as the hidden reference was expected to be rated about the same as its copy. The subjects did not know about the hidden reference which was removed when analyzing the main results but used for checking consistency. The reason for this tweak of the MUSHRA interface was the difference in purpose of the study. When rating preference there was no sense in providing a sound which was to be rated the highest since this means any other sounds would be a deterioration of this sound, removing the subjective qualities of the study. The anchors were removed for the same reason.

The tweaked MUSHRA interface consisted of two pages presented in a random order where one page had all of the guitar stimuli and the other page had all the vocal stimuli. The stimuli of each page was also randomized. This quasi-randomized way of presenting the stimuli (since within a page there was no randomization between guitar and vocal stimuli) was chosen due to the possibility that intermixed stimuli of guitar and vocals might distract subjects preference ratings between types of distortion while highlighting differences of instrument. While this might provide interesting results regarding the interactions of distortion and different instruments it withdraws from the focal point of differences between the types of distortion.

The final tweaked MUSHRA therefore consisted of two pages with seven stimuli on each page (six to be examined and a hidden reference, see table 2.a) divided into guitar stimuli on one page and vocal stimuli on the other page along with the seven buttons and sliders. There was also a ”play/pause” button, however pressing the button of another stimuli would seamlessly play that stimuli at the same point in time as the previous stimuli was stopped at. In addition to these buttons there was a ”Loop” button, allowing continuous repeated playback of sections of the stimuli selected with additional timeline-based sliders. A button labled ”Next” prompted the subjects to move on to the next page once all the sliders had been placed at a rating corresponding to the subjects evaluations. The subjects could not go back to a previous page.

2.5.2. Experiment one

The first of the experiments had the subjects rate preference between the stimuli using the tweaked MUSHRA interface (see 2.5.1 MUSHRA interface) with an additional copy of each page. The reason for having a copy of each page was to provide an estimated average rating for each stimulus for each subject, which provided greater precision in the preference estimates. It also helped to further investigate the accuracy of which the subjects had identified the hidden reference. The instructions as well as demographics were printed and handed to the subjects at the start of the experiment (see 6.2. (Appendix 2) Instructions and demographics for experiment one). For this experiment the subjects were allowed to freely adjust the audio playback level. By providing this option the subjects could listen to the stimuli as they would in any other situation – professionally or casually. This might have increased the subjectivity of the experiments.

After submitting the pages of the MUSHRA the subjects were asked to describe what influenced their preference ratings. This paper was handed to the subjects after submitting the MUSHRA (see 6.2. (Appendix 2) Instructions and demographics for experiment one). Figure 2.5.2a shows the frequency of words influencing preference ratings with no categorization. This quantitative analyzis of qualitative data shows that subjects listened mostly for clarity and noise when deciding on preference.

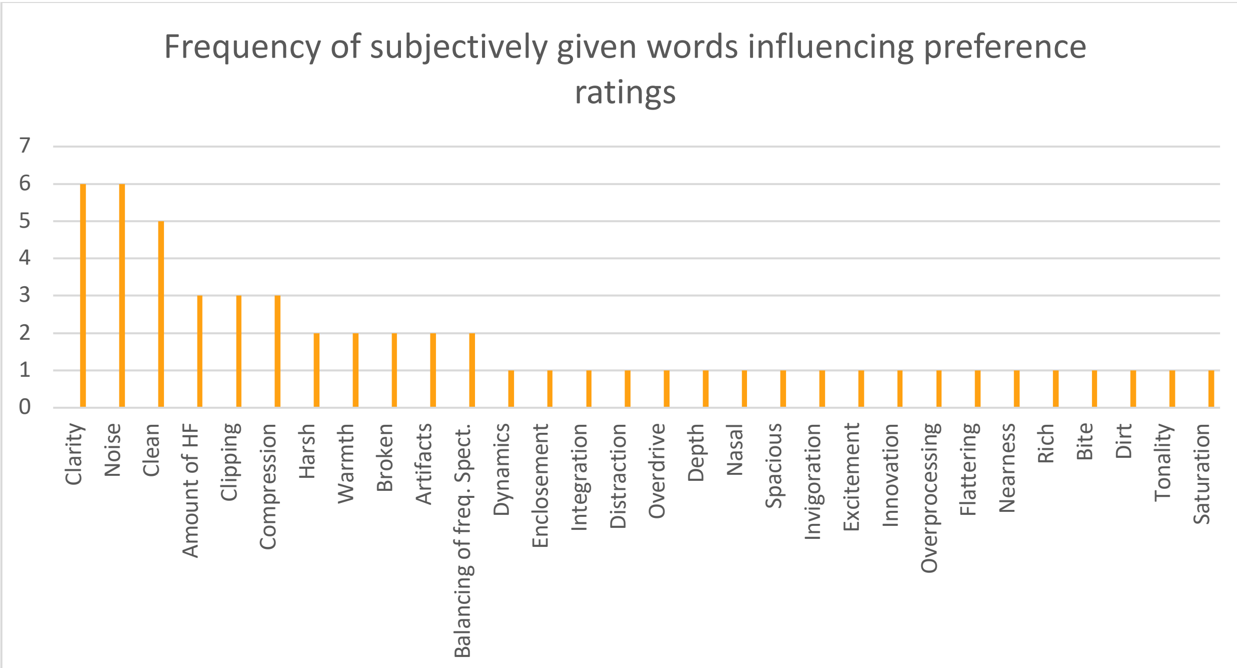


Figure 2.5.2a – Frequency of subjectively given words influencing preference ratings

Out of the subjectively given words some proved irrelevant or too vague for this study. Examples of these were ”noise”, ”clean” and ”artifacts”. The reasoning behind this was that while noise and clean were close to antonyms they described the amount of existing distortion rather than the timbral attributes of these distortions. Clarity, which was closely related to clean, might be considered a timbral attribute, however given the context the author of this study deemed the term too vague as it might imply the use of distortions itself. After sorting, the remaining words which could be considered to describe timbre were ”Harsh”, ”Warmth”, ”Amount of HF (high frequencies)”, ”Nasal”, ”Rich”, ”Bite”, ”Dirt” and ”Saturation”. Words which had similar meanings were categorized into timbral attributes describing the common timbral effect of these words. Harsh, bite and dirt was categorized to ”Roughness”, warmth, rich and saturation was categorized to ”Warmth”, amount of HF was categorized to ”Sharpness” and nasal was categorized into ”Nasality”. The frequency of occurance of the categorized timbral attributes can be observed in figure 2.5.2b. The results provided the two most frequent timbral attributes to be used in experiment two; roughness and warmth. The words given by the subjects were both in Swedish and in English meaning a translation had to be made where possible errors might have occured. The attributes were not further described to the subjects in experiment two.

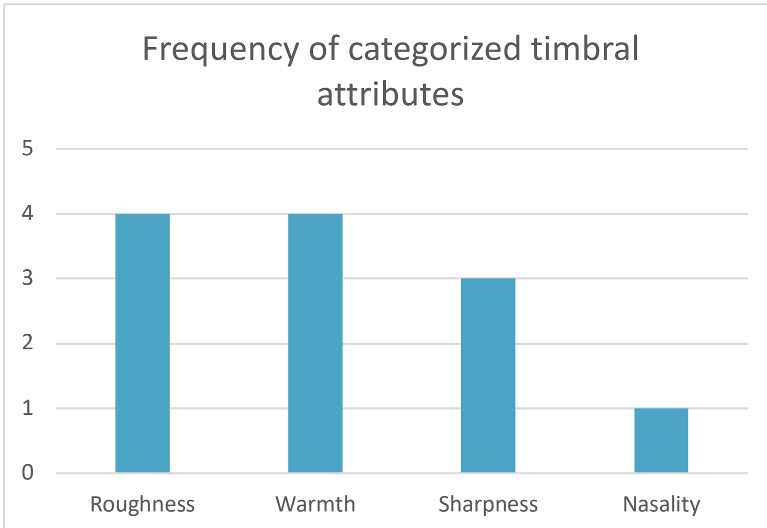


Figure 2.5.2b – Frequency of categorized timbral attributes

2.5.3. Experiment two

The second experiment, conducted two weeks after the first, consisted of two tests named test A and test B. These tests, in order to keep the length of the experiment roughly the same as for the first experiment, did not include copies of the MUSHRA pages (i.e. all the same stimuli were presented but the participants did not make repeated measurements of each stimuli). In all other regards the MUSHRA was identical to the first experiment but with different instructions handed to the subjects. Test A asked to subjects to rate the amount of ”warmth” between the stimuli and test B asked them rate the amount of ”roughness” (see 6.3. (Appendix 3) Instructions for experiment two test A, and 6.4. (Appendix 4) Instructions for experiment two test B). Warmth and roughness were chosen as timbral attributes that might vary between the distortions (and therefore explain preference), because these were the most prominent attributes listened out for when rating preference in the first experiment.

The instructions also specified that the subjects could not adjust the audio playback level for experiment two contrary to experiment one. The reason for this is that the audibility of some timbral attributes might vary depending on playback level. This was described by the equal-loudness curves (ELC) explained by Fastl & Zwicker (2007), at low levels the amount of audible low frequencies diminsh while the most sensitive area of frequencies exists around 3 kHz. Having each subject listen at the same playback level removed the possibility of some subjects adjusting the level to properly hear and evaluate timbral differences where others might not have. The playback level was therefore set to a level decided on by a small committee as the subjects would monitor through headphones and dB SPL (sound pressure level) could not properly be measured.

3. Results and analysis

The results was achieved by utilizing Microsoft Excel and IBM SPSS. A full factorial repeated measures ANOVA was conducted using Bonferroni corrected p values as well as Greenhouse-Geisser corrected degrees of freedom, and therefore Greenhouse-Geisser corrected p values, where necessary.

3.1. Experiment one – preference test

This section presents the results for experiment one which examined the effect of distortion type, distortion level and instrument type on preference. The results for experiment two, which investigated the effect of these factor on timbre, will be presented in section 3.2. Experiment two – test A and 3.3. Experiment two – test B. Of the 26 participants of experiment one, 21 were deemed to have identified the hidden reference and as such 21 was the number of participants which the data was based on.

3.1.1. Main effects analysis

Appendix 6.5.1. shows the results of Mauchly's Test of Sphericity for main effects as well as the violation of assumption of sphericity for the main effects of type. This meant the degrees of freedom (*df*) had to be corrected using the Greenhouse-Geisser correction. The within-subjects effects table was examined next.

The main effects of instrument yielded an F ratio of $F(1, 20) = 0.011$, $p = 0.916$ providing a non-significant result. This meant that there was a greater than 5% chance that any observed difference in preference score between vocals and guitar was due to chance factors alone, rather than a true difference.

The main effects of type violated the assumption of sphericity, therefore GG (Greenhouse-Geisser) correct degrees of freedom were used to obtain the p values. This effect yielded an F ratio of $F(1.344, 19861.516) = 42.174$, $p < 0.001$. This meant that any differences between distortion types were unlikely to be due to chance. However, since there were three types of distortion one could not know based on these results which of the types of distortion were statistically significant.

In order to find which of the types were statistically significant a pairwise comparison had to be made – see further below as well as appendix 6.5.3.

The main effects of level yielded an F ratio of $F(1, 20) = 34.898$, $p < 0.001$. This meant that the main effects of level was deemed statistically significant. This meant that the observed difference in preference between the two different levels was unlikely to have come about by chance. Therefore there was a confidence that the difference we see represents a true difference in preference.

Appendix 6.5.3. shows the results of pairwise comparisons of the main effects of type. The results indicated a stastical significance in preference between all the types of distortion as well as depicting the differences in mean between the types. The mean differences show that type 1 (ZC) was the least preferred (with a mean preference rating of 23) and type 3 (T) the most preferred (with a mean preference rating of 53).

Figure 3.1.1a shows the average mean of preference for all main effects including 95 % confidence intervals. The main effects of instrument was included even though there was no statistic significance as these might prove interesting to some.

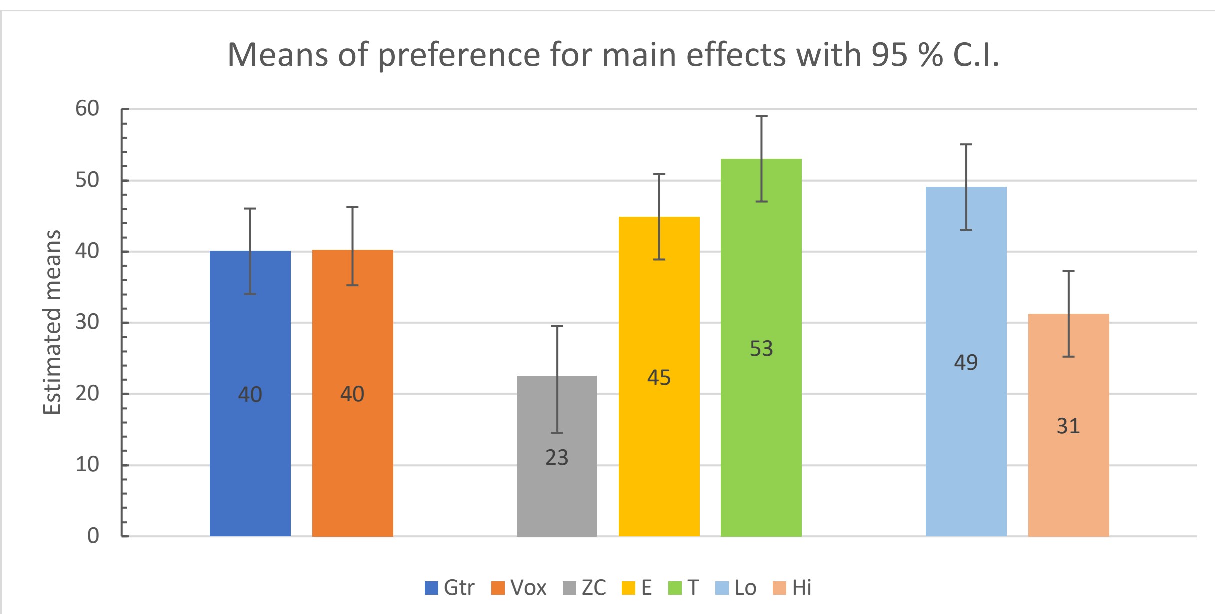


Figure 3.1.1a – Means of preference for main effects with 95 % confidence intervals

3.1.2. Interaction effects

Appendix 6.5.4. shows the results of Mauchly's Test of Sphericity, which tests for the assumption of sphericity, for all interaction effects ($p = 0.005$). All results were non significant so no correction of degrees of freedom had to be made.

The test of within-subjects effects for interaction effects (appendix 6.5.5.) indicated there were statistical significance for the interaction effects of instrument * type, and instrument * level. Again this meant that the differences observed in this experiment were unlikely to simply represent chance variation.

The interaction effects of instrument * type yielded an F ratio of $F(2, 40) = 15.117$, $p < 0.001$. The interaction effects of instrument * level yielded an F ratio of $F(1, 20) = 8.007$, $p = 0.010$. This meant both of these interaction effects were deemed stastistically significant.

The interaction effects of type * level yielded and F ratio of $F(2, 40) = 0.152$, $p = 0.859$, and the interaction effects of instrument * type * level yielded an F ratio of $F(2, 40) = 1.106$, $p = 0.341$. These interaction effects were therefore not statistically significant and was considered as caused by chance. These effects was therefore not further analyzed or discussed.

Appendix 6.5.6. shows the mean, standard error and 95 % confidence intervals for the interaction effects of instrument * type and instrument * level. Figure 3.1.2a and figure 3.1.2b provides a visual representation of the results stated in appendix 6.5.6. These were the estimated means of preference for the interaction effects of instrument * type and instrument * level. Overlapping confidence intervals is normally undesired, however these effects had already been proven to be significant.

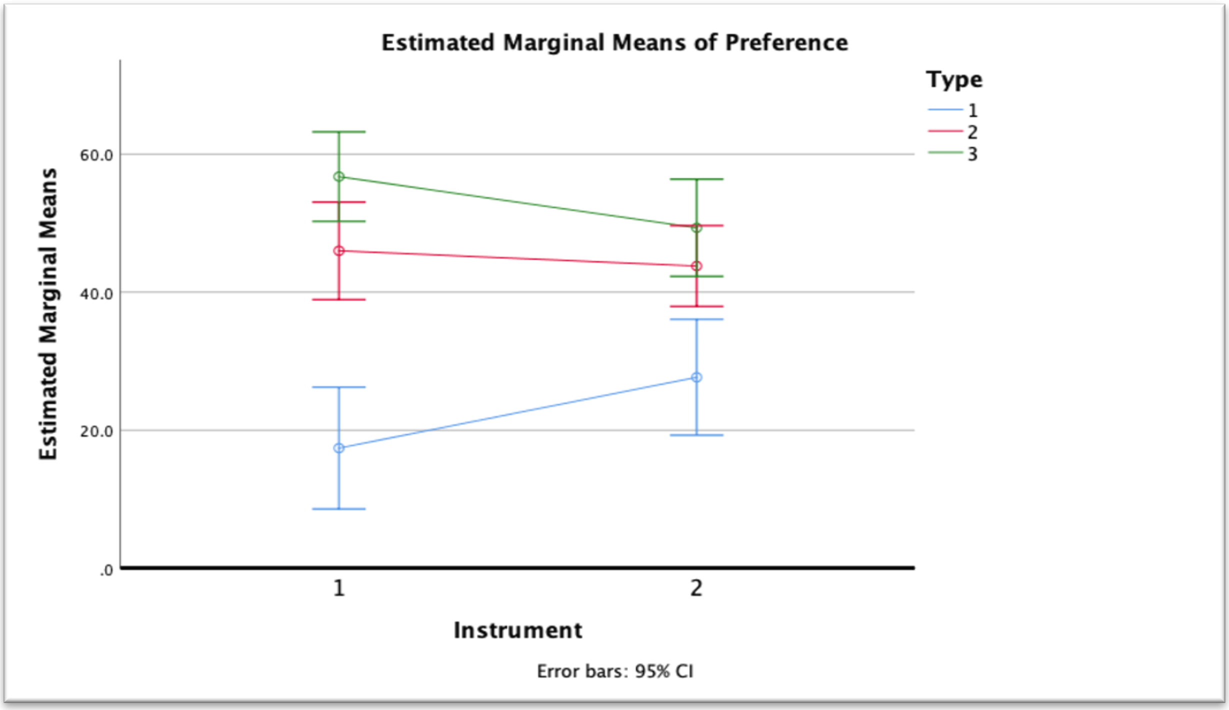


*Figure 3.1.2a – Line graph depicting interaction effects of instrument * type. Instrument 1 = Gtr, instrument 2 = Vox, type 1 = ZC, type 2 = E, type 3 = T*

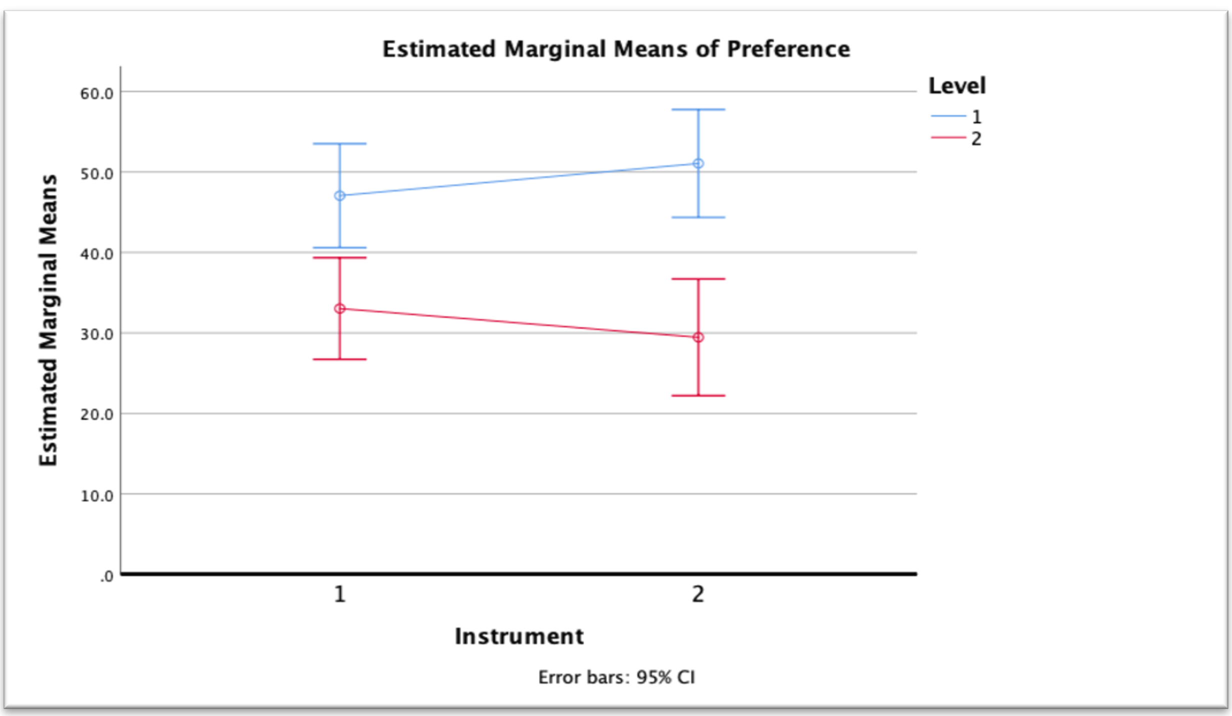


*Figure 3.1.2b – Line graph depicting interaction effects of instrument * level. Instrument 1 = Gtr, instrument 2 = Vox, level 1 = Lo, level 2 = Hi*

3.1.3. Analysis of experiment one

3.1.3.1 Main effects

The results of the main effects showed that the preference of the types of distortion vary, with ZC (zero-crossing distortion) being rated the lowest and T (tube distortion) being rated the highest. The gap in estimated mean was however closer between E and T (8 score) than for ZC and E (22 scores). This meant that while there was a difference in preference, the preference between E and T was more equal than the preference of ZC. Since there was a statistic significance in the difference of types of distortion the null hypothesis could be rejected with 95 % confidence. According to the hypothesis set out in section 1.6. Aims and purpose, these results were expected. Particularly it was expected that ZC would be the least preferred distortion as this had been observed in previous literature.

There was also a noteworthy difference in preference of level (18 scores). The lower amount of distortion was more preferred (49 scores) than the higher amount (31 scores), meaning the subjects might have disliked the distortions altogether and would have preferred a clean stimuli. This was however not part of this study as a clean signal provides no information on preferences or perception of timbre between different types distortion, as a clean signal in itself indicates a lack of distortion, and the independent variable of level was included mainly to observe any interaction effects. See section 2.5.2 Experiment one, for further information.

3.1.3.2 Interaction effects

The interaction effects of instrument * type (figure 3.1.2a) showed that while ZC was the least preferred type of distortion overall, it was even more so for the guitar stimuli. The differences in preferences of types of distortion was closer for vocals as the preferences of E and T both were lower than for the guitar, while the preference of ZC was higher than for the guitar. Therefore, these interaction effects was identified as the importance of type of distortion being less prominent on vocals than guitar.

The interaction effects of instrument * level (figure 3.1.2b) indicated that there was a bigger difference in preference of level of distortion when applied to vocals than guitar. The higher level of distortion was least preferred for both instruments but even more so for vocals. On the other hand, the preference of the lower amount of distortion was increased when applied to vocals rather than guitar, meaning the level of distortion had less impact on overall preference for guitar than for vocals.

What had been discovered was that for guitar the type of distortion seemed to be of more importance than the level of distortion when assessing subjective preferences, while the opposite could be said for vocals. The preference ratings on vocals seemed to be motivated more by the level of distortion rather than the type of distortion.

3.2. Experiment two – test A

This section presents the results for experiment two test A which examined the effect of distortion type, distortion level and instrument type on the perceived amount of warmth. Of the 24 participants of experiment two test A, 20 were deemed to have identified the hidden reference and as such 20 was considered the number of participants which the data was based on.

3.2.1. Main effects

Appendix 6.6.1. shows the results of Mauchly's Test of Sphericity for main effects as well as the violation of assumption of sphericity for the main effects of type. This meant the degrees of freedom (*df*) for the main effect of type had to be corrected using the Greenhouse-Geisser correction.

The results of the within-subjects effects test showed that there was no statistical significance for the main effects of instrument while there was for the main effects of type and level. The main effects of instrument yielded an F ratio of $F(1, 19) = 2.121$, $p = 0.162$. The main effects of type yielded an F ratio of $F(1.159, 22.027) = 13.109$, $p = 0.001$ and the main effects of level yielded an F ratio of $F(1, 19) = 49.523$, $p < 0.001$. These main effects could be considered statistically significant. The signifance of type did however need to be further looked into as there was a possibility only some of the types might have been statistically significant. This was done using pairwise comparisons.

Appendix 6.6.3. shows the results of pairwise comparisons of the main effects of type. Reported was a statistic signifance for all of the types as well as the mean differences between these types. It could be concluded that ZC was considered to have the least amount of warmth (with a mean rating of 35) while T was considered to have the most amount of warmth (with a mean rating of 57).

Figure 3.2.1a shows the estimated means of the amount of warmth for each of the main effects. The main effect of instrument was not statistically significant meaning there was a greater than 5 % probability any differences were chance products. The 95 % confidence intervals would ideally not overlap, however these main effects had already been proven to be stastically significant and as such these could be ignored.

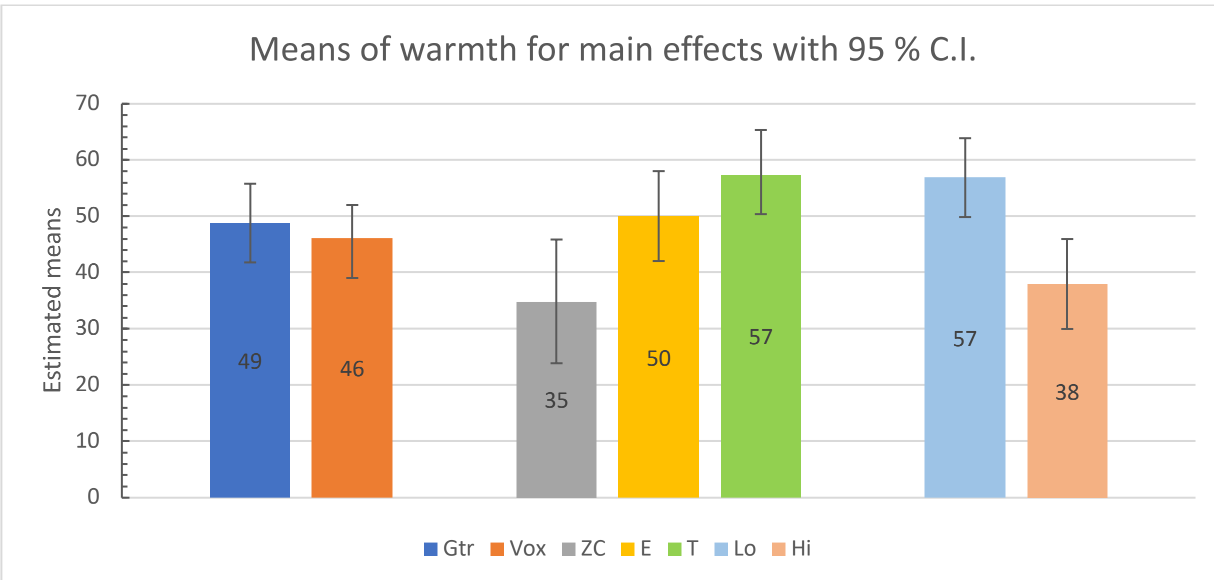


Figure 3.2.1a - Means of warmth for main effects with 95 % confidence intervals

3.2.2. Interaction effects

Appendix 6.6.4., showing Mauchly's Test of Sphericity, shows that sphericity could be assumed for all of the interaction effects. This meant no correction of degrees of freedom had to be made. Appendix 6.6.5. shows the results of the within-subjects tests for interaction effects. None of these interaction effects were statistically significant.

The interaction effects of instrument * type yielded an F ratio of $F(2, 38) = 1.985$, $p = 0.151$, instrument * level yielded an F ratio of $F(1, 19) = 0.090$, $p = 0.768$, type * level yielded an F ratio of $F(2, 38) = 1.997$, $p = 0.150$ and instrument * type * level yielded an F ratio of $F(2, 38) = 0.065$, $p = 0.937$.

3.2.3. Analysis of experiment two test A

3.2.3.1. Main effects

The main effects of type and level, proved to be stastically significant, indicated that the amount of warmth was perceived to be the greatest for distortion typ T and low levels of distortion. The least perceived amount warmth was deemed to be distortion type ZC and high levels of distortion. Since the interaction effects proved to be nonsignificant there was a limit to how many assumptions could be made regarding these results, however there was an interesting observation to be made. Distortion, commonly associated with warmth, was deemed to have less warmth at increased levels of distortion. This might have meant that the amount of distortion was too high and therefore was no longer perceived as bringing warmth, or it might mean that there was another type of distortion believed to bring warmth. Of the types of distortion used the T type, emulated from tube distortion, was perceived as the most warm and this ringed true with the common assumptions that tubes add warmth to the signal.

There was also a similarity in the ratings of preference and warmth. Both results showed the main effects of instrument as nonsignificant, ZC as the least preferred/least amount of warmth, T as the most preferred/most amount of warmth and guitar as the more preferred instrument in conjunction with type and with the most warmth. Therefore the assumption that subjects preferred distortion which adds warmth and mostly on guitar could be made. There was now a link between timbre, specifically warmth, and preference of distortion.

3.3. Experiment two – test B

This section presents the results for experiment two test B which examined the effect of distortion type, distortion level and instrument type on the perceived amount of roughness. Of the 24 participants of experiment two test B, 19 were deemed to have identified the hidden reference and as such 19 was the number of participants which the data was based on.

3.3.1. Main effects

Appendix 6.7.1. shows the results of Mauchly's Test of Sphericity for main effects as well as the violation of assumption of sphericity for the main effects of type. This meant the degrees of freedom (*df*) had to be corrected using the Greenhouse-Geisser estimates of sphericity.

The results of the within-subjects effects test (see appendix 6.7.2.) showed that there was no statistical significance for the main effects of instrument while there was for the main effects of type and level. The main effects of instrument yielded an F ratio of $F(1, 18) = 0.005$, $p = 0.942$. The main effects of type yielded an F ratio of $F(1.375, 24.749) = 21.465$, $p < 0.001$ and the main effects of level yielded an F ratio of $F(1, 18) = 139.046$, $p < 0.001$. The signifance of type did however need to be further looked into as there was a possibility that only some of the types might have been statistically significant. This was done using pairwise comparisons.

Appendix 6.7.3. shows the results of pairwise comparisons of the main effects of type. Reported was a statistic signifance for all of the types as well as the mean differences between these types. It could be concluded that ZC was considered to have the most amount of roughness (with a mean rating of 66) while T was considered to have the least amount of roughness (with a mean rating of 43). Figure 3.3.1a shows the estimated means of the amount of roughness for each of the main effects. The 95 % confidence intervals would ideally not overlap, however these main effects had already been proven to be stastically significant and as such these could be ignored.

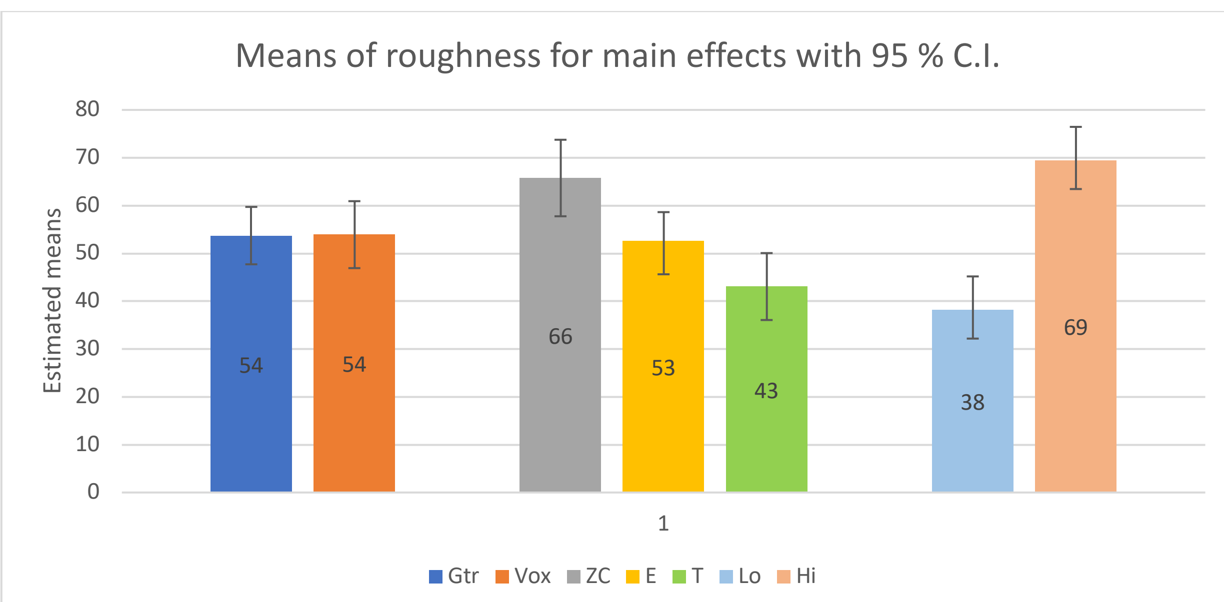


Figure 3.3.1a – Means of roughness for main effects with 95 % confidence intervals

3.3.2. Interaction effects

Appendix 6.7.4. shows Mauchly's Test of Sphericity for interaction effects which resulted in assumptions of sphericity for all interaction effects. This meant no corrections of degrees of freedom had be made.

As seen in appendix 6.7.5. there was a statistic significance for the interaction effects of instrument * type and instrument * level. There was however no significance for the interaction effects of type * level or instrument * type * level and the data for these interaction effects could be considered as caused by chance.

The significant interaction effects of instrument * type yielded an F ratio of $F(2, 36) = 6.371$, $p = 0.004$ and the interaction effects of instrument * level yielded an F ratio of $F(1, 18) = 5.161$. The nonsignificant interaction effects of type * level yielded an F ratio of $F(2, 36) = 2.238$, $p = 0.121$ and the interaction effects of instrument * type * level yielded an F ratio of $F(2, 36) = 0.216$, $p = 0.807$.

Appendix 6.7.6. shows the mean, standard error and 95 % confidence intervals for the interaction effects of instrument * type and instrument * level. Figure 3.3.2a and figure 3.3.2b provides a visual representation of the results stated in appendix 6.7.6. These were the estimated means of perceived roughness for the interaction effects of instrument * type and instrument * level. Overlapping confidence intervals is normally undesired, however these effects had already been proven to be significant and could therefore be ignored.

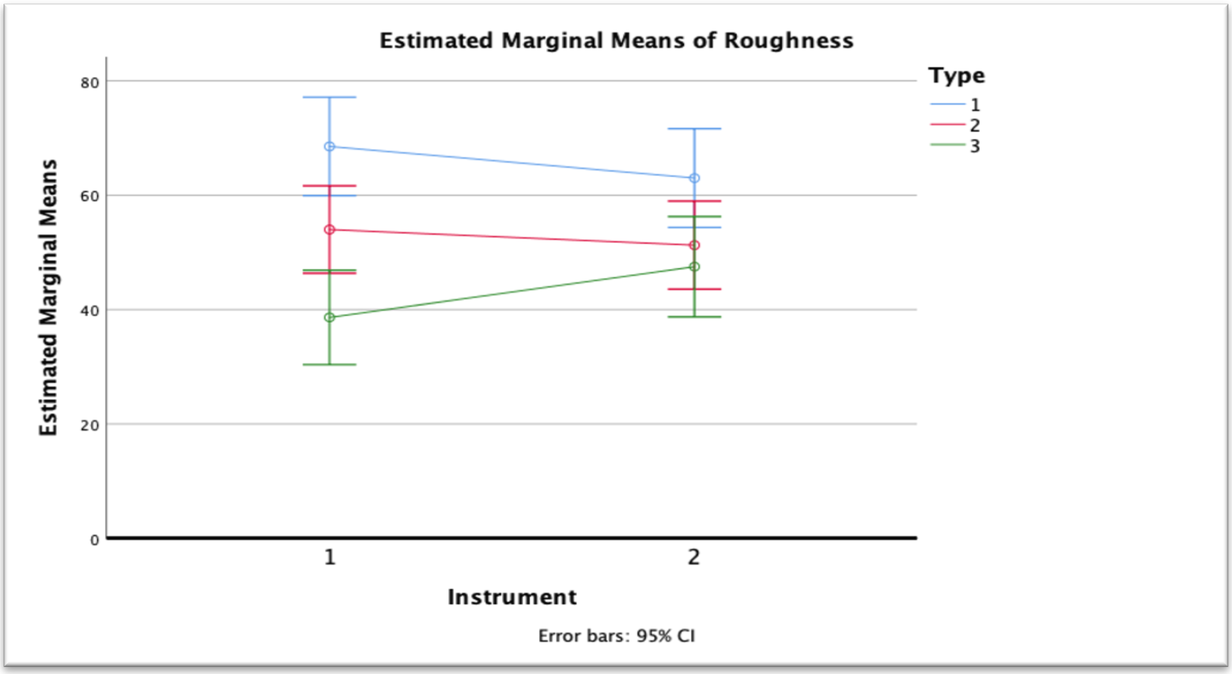


*Figure 3.3.2a – Line graph depicting interaction effects of instrument * type. Instrument 1 = Gtr, instrument 2 = Vox, type 1 = ZC, type 2 = E, type 3 = T*

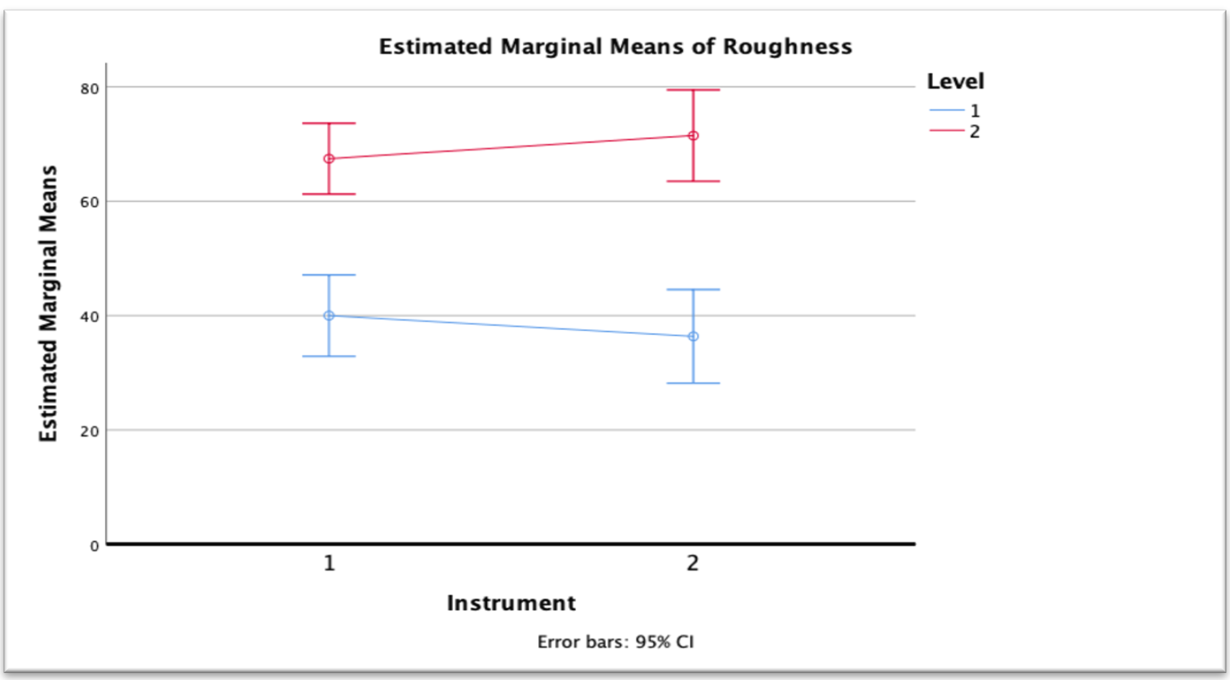


*Figure 3.3.2b – Line graph depicting interaction effects of instrument * level Instrument 1 = Gtr, instrument 2 = Vox, level 1 = Lo, level 2 = Hi*

3.3.3. Analysis of experiment two test B

3.3.3.1. Main effects

The main effects of type and level indicated high amounts of roughness for type ZC and high levels of distortion, while the lowest amounts of roughness came with type T and low levels of distortion. This was the polar opposite of the main effects for both preference and amounts of warmth. This might have indicated that the subjects generally preferred warmth but not roughness as the types and levels of distortion are rated as such. However, since there were no significant interaction effects for warmth, while there were for roughness, the pattern of warmth being polar opposite of roughness broke.

3.3.3.2. Interaction effects

The interaction effects instrument * type indicated that the amounts of roughness for the types of distortion varied depending on the instrument. While both ZC and E had a linear decline of roughness going from guitar to vocals, T had an increase of roughness placing it very close to the perceived amount of roughness for E. This meant that the amount of roughness depending on type of distortion relied on interaction, and possibly integration, with the instrument as the sensitivity of perception of this timbral attribute seemed higher for guitar than for vocals.

The interaction effects instrument * level provided additional interesting results. While this study has concluded that the type of distortion seemed to matter less for vocals than for guitar when assessing both the amounts of roughness as well as preference, it seemed that the level of distortion matters *more* for vocals than for guitar for the same assessments. This might have indicated that what type of distortion was causing the increase of roughness was less prominent for vocals, while none the less different levels of distortion affected the amount of roughness more heavily for vocals than for guitar.

3.4. Patterns

This section highlights some patterns in the ratings between preference, warmth and roughness. The reader may want to look at figure 3.1.1a, 3.2.1a and 3.3.1a for a reminder of outcomes. Warmth and roughness seemed to be on opposite ends in terms of significant main effects, meaning the type and level of distortion. Preference seemed to be allied with high warmth and low roughness. There were similar patterns in interaction effects for preference and roughness as well, where the mean differences in preference and amount of roughness between types of distortion increased depending on instrument. The same could be said for the mean differences in preference and amount of roughness between levels of distortion, increasing depending on instrument.

4. Discussion

This study aimed to clarify how subjective preference for different types of distortion may vary and how these differences were perceived by applying these types of distortion to two different instrument and at two different levels. Two timbral attributes were choosen from the qualitative data taken in experiment one where participants were asked what they were basing their preference on. These attributes were evaluated in experiment two in order to explain the differences between the types of distortion as well as what was sought after in distortion.

4.1. Conclusions

The main effects of the independent variable instrument proved to be nonsignificant for all of the tests, which was suprising as McMullin et al. (2019) concluded that distortion was most audible on vocals (and bass), meaning any nonpreferred distortion might have had a bigger impact on the overall preference of instrument. The interaction effects of instrument, however, proved significant for experiment one (preference) and experiment two test B (roughness). What this meant was that while the differences in preference, warmth or roughness between types of instrument might have been product of chance, the combination of instrument and type of distortion or level of distortion and how it affected the ratings could be considered as not caused by chance. A nonsignificance did not mean that there was no difference of preference, warmth or roughness, but that the spread of data provided indicated a cause of chance. Additional instruments, playing varying genres, might have provided results showing differences of preference, warmth or roughness between types of instrument. Therefore it remains possible that future experiments would show a difference due to program item.

There was statistic significance between all types of distortion for all of the tests, meaning any mean differences were likely not due to chance. The outcomes showed that T type (tube distortion) was considered the most preferred, most warm and least rough, while ZC (zero-crossing) was the least preferred, least warm and most rough. E type (solid-state germanium distortion) was considered the middle ground for all tests. The results were somewhat expected as Poss (1998) claimed tube distortion was inherently pleasing, and Voishvillo (2006) demonstrated how zero-crossing distortion was perceived as of bad audio quality. The interaction effects of type and instrument proved statistically significant for the test of preference and the test of roughness. Therefore some conclusions could be made. The type of distortion did influence the subjective preference ratings as well as the amount of warmth and amount of roughness, but even more so for preference and roughness when combined with the type of instrument (i.e. guitar, more so than vocals, influencing preference and percieved amount of roughness between the types). The null hypothesis could therefore be rejected.

The results of type of distortion also correlated to the hypothesis that types of distortion would be rated differently as neither the type with most or the least amount of THD was the most preferred type (i.e. ZC had the least amounts of THD and was least preferred, E had the most amount of THD and was not more preferred than T). This was true to Voishvillos (2007) conclusion that the type of distortion was more relevant when assessing preference than the amount of THD as the THD for ZC was 4 % at low levels and 10 % at high levels for guitar and 4 % at low levels and 7 % at high levels for vocals, while the T type had a THD of 6 % at low levels and 8 % at high levels for guitar and 5 % at low levels and 8 % at high levels for vocals. The argument that an increase of THD might have been more preferred did not ring true as the E type had THD of 10 % at low levels and 16 % at high levels for guitar and 8 % at low levels and 14 % at high levels for vocals (see 2.4.2. Applying distortion). While the preferences for the types of distortion might have differed if they had all exhibited the same amount of THD, this concluded that THD did not solely dictate preference (see 1.4. THD as a measure).

There was a significance for the main effects of level in all tests. The level of distorion could be said to have influenced the subjects preference, the amount of warmth and the amount of roughness, with high levels showing less preference and warmth but more roughness. Only the preference and roughness tests showed statistically significant level by instrument interactions. There was an increased importance of level with the vocals compared to the guitar, specifically the effect of level became more apparent with the vocals both when measuring preference and roughness. There were no level by type interactions. This meant that while the sensitivity about the type of distortion was greater for guitar than for vocals, the amount of distortion was of more importance for vocals than for guitar.

The patterns showed that while subjects seemed to prefer low levels of distortion, these differences were smaller for guitar than for vocals where the preference of low levels of distortion was increased and the preference of high levels of distortion was decreased. Further, this might have indicated that not the type of distortion, but distortion in itself brought perceived roughness to vocals. What was interesting was that, yet again, the opposite could be said for guitar where the type of distortion influenced perception of roughness more than the level of distortion when compared to vocals. The cause of this might have been the expectations on distortion when applied to instruments playing a specific genre. The preference of guitar seemed less sensitive about the amount of distortion compared to vocals, which might have been expected as guitar is more commonly associated with distortion. Including different genres as well as instruments in the study might have proved that varying amounts of distortion influenced the instrumental preferences more or less depending on the performed genre.

The ratings of the dependent variable of warmth seemed to be influenced by the type of distortion as well as the level of distortion. The T type was considered to exhibit the most amount of warmth while ZC was considered to exhibit the least. A connection could be made between these results and the common conception of distortion adding warmth to the signal, meaning the distortion referred to in these contexts might be tube distortion. However, as previously stated, the type of distortion rather than the amount of THD seemed to have influenced the amount of warmth as T type had the higher average rating of warmth when compared to E, which had the higher amount of THD. The fact that T type was rated as the most preferred and the most warm of the types of distortion might have meant that subjects prefer distortion adding perceived warmth to the signal (exploring the physicalities of different types of distortion might provide some more insight). However, as subjects also seemed to perceive more warmth for low levels of distortion than for high levels of distortion, it might have beeb that a clean signal (i.e. signals with no added distortion) would be rated as having even more amounts of warmth. Since there were no interaction effects of type and level for the amount of warmth this could only be speculated on.

The ZC type of distortion was rated to have the most amount of roughness, followed be E and lastly T. This might have indicated that subjects preferred distortion not adding roughness to the signal as ZC was rated the least preferred and most rough. However the differences were smaller in combination with type of instrument. E and T were rated very closely at 51 and 47 for vocals while the same types were rated 54 and 38 for guitar. This might have meant that there was less sensitivity for the types of distortions effect on perceived amount of roughness depending on the type of instrument. The types seemed to matter less for vocals than for guitar, which yet again might have been the cause of expectations on the use of distortion for these specific instruments.

4.2. Validity and reliability

The validity, meaning if the answers provided relate to the posed question, of this study is affected by a number of things. What had been sought after was an answer to the differences in preference between types of distortion as well as how these differences are perceived by subjects. In order to examine this, stimuli had to be created. This stimuli could have been a single sinusoidal tone at 1 kHz, however as concluded in 1.4., this method lacked the complexity of sounds more commonly heard by subjects. Therefore, the use of isolated instruments performing a specific genre were chosen to increase ecological validity. Two different instruments were chosen as the effects of types of distortion might have proven to be perceived as different depending on the type of instrument and the expectations on the use of distortion on these. Therefore, while the answers provided by this study did answer the question of subjective preference between types of distortion and how these are perceived, the answers were also dependent on the types of instrument and the performed genre. To conclude the reasoning of validity it could be said that the results of the main effects of type of distortion proved that regardless of the instruments and levels of distortion there was a statistically significant difference between the types of distortion for all tests.

The reliability of this study could be accepted as the parameters and equipment used has been stated. The clean stimuli is available to anyone who wish to repeat the study (contact the author). The mindset and mood of the subjects could of course have affected the outcome of the tests, however the sessions for experiment two test A and experiment two test B was due two weeks after experiment one, decreasing the likelyhood that mindset and mood had an impact on the results. The relatively large subjectpool also increased the likelyhood that a repetition of this study would have provided equal results. This could however be affected by expectations on use of distortion, and depending on when such repetition would take place these expectations may be different from when this study took place. This would probably yield different results.

4.3. Critique of methodology

As by nature every choice made for the methodology had a counteroption, these options would have yielded different experiments and therefore factors which might have affected the results. This section covers some of these options as well as suggestions on improvements to the choices made.

4.3.1. Recordings of stimuli

While the clean recorded stimuli is available those asking for it, the SQAM (see 2.4.1. Source samples) which was commonly used in audio research would have provided not only the clean audible signal used immediately, but would also have made this study more comparable to other studies using the SQAM such as Herzog (2009). By not using the SQAM more freedom of stimuli was granted, resulting in a more controlled environment, however this had the mentioned downsides.

Having guitar and vocals perform the same program (i.e. the same piece of music) might have made the chord progression or melody make sense for one instrument but not the other. By having different programs the instruments were performed in ways more commonly heard within musical contexts. On the other hand, the difference of programs might have made subjects assess based on the differences between programs rather than differences between distortion. Then again, following this reasoning, by having the instruments perform the same piece of music the assessments could be based on how fitting the program was between instruments. The easy way out of this loop would have been to either exclude different types of instrument, or to have informed the subjects to disregard quality of performance, chord progression and melody, the latter of which was done.

There were of course different expectations on the use of distortion depending on the performed genre. The stimuli of this study was performed in a single genre (mostly due to the time limit and envelopment of the research) believed to have no specific expectations on the use of distortion. However, this was an area of research which could very well be delved deeper into as different genres might certainly have provided different outcomes. The same goes for the amount of instruments. Only two instruments were investigated in this study, selected on opposite ends of expected preference outcome based on the reviewed literature, but the effects of distortion when applied to the whole plethora of instruments used in music could certainly have been explored for interesting results.

The question of musical context (i.e. multiple instruments performing a program such as any commercial song) might hav been the most complex of the ones posed. Not only did it include the questions of genre, program and instruments involved the same way the above-mentioned did, but it also alters how the distortion would be applied and in what amount it had to be applied in order to make the differences between the types audible. For example, a song performed including mutiple instruments might have been considered a more broadband sound, which in turn might have masked the effects of distortion (see 1.3.3. Auditory masking of distortion). In that case, the distortion would have had to be applied at a higher level, possibly resulting in a non-ecologically valid stimuli none the less (assuming musical context would be used to increase ecological validity). Another possibility was subjects not identifying what change there was between stimuli which might have resulted in an increased spread of data. In that case different instructions may have had to be provided, and at that point the study had taken a whole different form, leaning more towards the research of McMullin et al. (2019) where distortion was applied to known songs, along with questions such as the audibility of distortion rather than how these might have differed.

4.3.2. Application of distortion

As mentioned in 2.4.2. Applying distortion, distortion on guitar is not commonly applied post recording but rather between the guitar and the amplifier. Applying distortion unorthodoxly post recording did decrease the ecological validity to some degrees as the effects of distortion applied this way was not the same as the effects of distortion applied the traditional way pre recording. This might have been a reason why there was no statistical significance in the main effects of instrument for any of the tests.

While references had been used to explain the differences in amount of THD, a different study might have wanted to include an additional experiment where the different stimuli all exhibited the same amounts of THD. The reasoning for this was that while some types of distortion may have detrimented from this equality of THD in terms of preference, some may have considered it as yet another aspect to the differences between types of distortion. However, as has been explained, the non-equality of THD in this study was due to the research examining not whether different types of distortion were considered as introducing more or less distortion (i.e.

the distortion effects being differently prominent between types) but rather in what terms these differences could be described and how they were differently preferred.

The results of the qualitative data gathered in experiment one proved that ”clean” and ”noise” were the most common descriptors influencing preference ratings. To understand why these descriptors were not chosen as the attributes for experiment two test A and experiment two test B one must understand the nature of these descriptors in this context. A clean signal referred to the amount of distortion of the stimuli (i.e. the dry/wet ratio), and subjects reporting this descriptor having influenced their preference ratings might not have paid attention to the timbral differences between the types but rather which stimuli was considered to have the most, or least, amount of distortion. Noise, in this context being the antonym of clean, followed the same logic. Including these descriptors in experiment two would not have been consistent with the aims of this study as it would then have been about how the prominence of different types of distortion affected preferences, rather than how these types of distortion were perceived differently.

4.3.3. Methodology of experiments

The prestudy could have been bigger and more in-depth. One of the reasons is that, as mentioned above in 4.3.2. Application of distortion, some subjects might have considered the prominence of certain types of distortion unequal. A more developed prestudy, covering more variations of stimuli (i.e. stimuli with more variations of levels of distortion for each type) and including more subjects, could have provided stimuli which more accurately corresponded to subjects ratings of preference, warmth and roughness. The stimuli provided, meaning the types and levels of distortion as well as resulting THD, would also with more certainty have been considered as appropriate for this study.

There was also a question about the free-text box in addition to MUSHRA. The second part of experiment one, where subjects rated attributes, could have been obsolete if text boxes were provided for each sample and preference rating. The subjects own descriptions could then have been analyzed to establish what attributes were influencing and being considered when rating preference. This might however have proven difficult to analyze and not very practical since there would have been many different reasonings for many pairs of preference ratings.

When asking subjects to describe what influenced their preference ratings, the subjects could have been guided towards describing timbral attributes rather than any descriptors. The reason this was not done was because such guidance occuring prior to, or during, the ratings of preference might have brought a bias towards certain types of distortion which consciously ”should” have been more preferred as specific timbral attributes were being asked for. Asking the subjects to describe what timbral attributes influenced their preference ratings prior to, or post, the preference ratings might have been the better option, however this was also not entirely true to subjectivity as timbral attributes (i.e. differences of timbre between types of distortion) might not have been a part of their preference ratings at all. As such, not guiding the subjects towards the timbral aspects of the distortion was considered to be the most subjective way of gathering data for experiment two.

4.4. Further research

Some variations which might have provided a different outcome has been stated in section 4.3. Critique of methodology. This section focus less on these different outcomes and more on how the outcomes provided by this study may be further explored.

Since it had been proven that there was a difference between types of distortion in preference as well as perceived amount of warmth and roughness, a further investigation might explore the physicalities of these types of distortion in order to provide a more technical explanation for these outcomes. While this study has been focusing on subjective characteristics, such research would focus on the objective characteristics.

There might also be further research done covering more variations of demographics. This study included students attending the musician and audio engineering program at Luleå University of Technology, with most of the subjects being 20 – 27 years old. While their musical preferences of genre varied, this subject group may be considered as too narrow to make any general conclusions regarding casual listeners subjective preferences and perceptions of distortion. Further research may therefore explore how the outcomes of this study compares to subject pools of different ages, professions or culture.

5. References

Avid. (2019). Pro Tools (Version 2019.6) [Digital Audio Workstation]. Retrieved from https://www.avid.com/pro-tools

Case. A. (2007). *Sound FX* (1. ed.). Elsevier: Oxford.

European Broadcasting Union. (2008). Sound Quality Assessment Material Recordings For Subjective Tests. Retrieved 2020 February 8 from https://tech.ebu.ch/publications/sqamcd

Fastl, H., & Zwicker, E. (2007). *Psychoacoustics facts and models* (3. ed.). Berlin: Springer-Verlag.

Herzog, S. (2009). *Investigations of the Effects of Nonlinear Distortions on Psychoacoustical Measures* (AES Convention Paper 7751). Retrieved from the website of Audio Engineering Society: http://www.aes.org/e-lib/browse.cfm?elib=14947

McMullin, E., Brunet, P., Wang, Z. (2019). *Describing the Audible Effects of Nonlinear Loudspeaker Distortion* (AES Convention Paper 10258). Retrieved from the website of Audio Engineering Society: http://www.aes.org/e-lib/browse.cfm?elib=20631

Owsinski, B. (2005). *The Recording Engineer's Handbook* (1. ed.). Thomson Course Technology: Boston.

Poss, R. M. (1998). Distortion Is Truth. *Leonardo Music Journal,* (8), 45-48.

Soundtoys. (2019). Decapitator (Version 5.3.2) [Software plug-in]. Retrieved from https://www.soundtoys.com/product/decapitator/

Splice. (2019). Serum (Version 1.23b2) [Software plug-in]. Retrieved from https://splice.com/plugins/15493-serum-vst-au-by-xfer-records

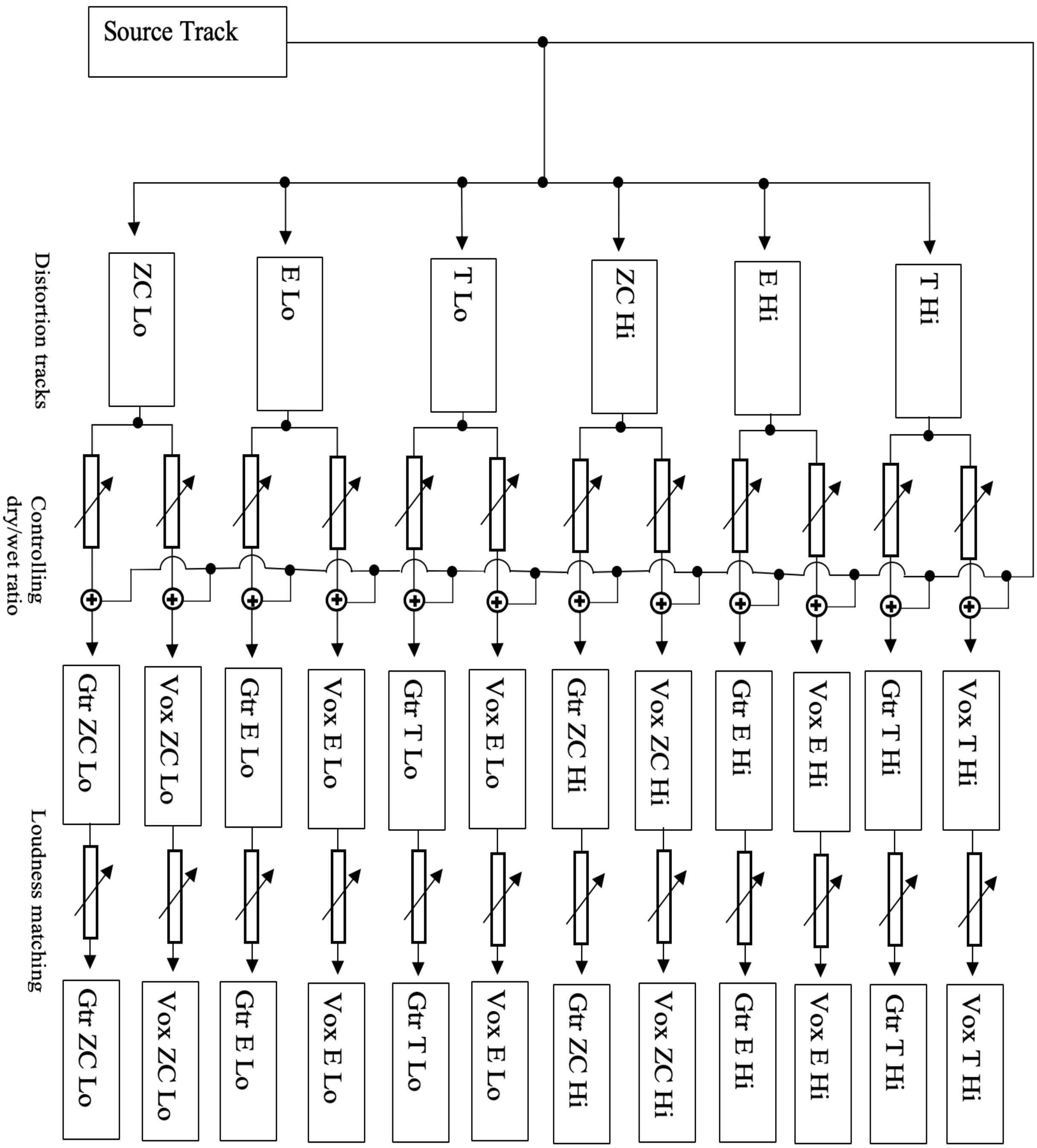
Voishvillo, A. (2006). *Assessment of Nonlinearity in Transducers and Sound Systems – from THD to Perceptual Models* (AES Convention Paper 6910). Retrieved from the website of Audio Engineering Society: http://www.aes.org/e-lib/browse.cfm?elib=13744

Voishvillo, A. (2007). *Measurements and Perception of Nonlinear Distortion – Comparing Numbers and Sound Quality* (AES Convention Paper 7174). Retrieved from the website of Audio Engineering Society: http://www.aes.org/e-lib/browse.cfm?elib=14232

Waves. (2019). WLM Plus Loudness Meter (Version 11.0.0) [Software plug-in]. Retrieved from https://www.waves.com/plugins/wlm-loudness-meter#how-to-set-loudness-levels-for-streaming-wlm

6. Appendix

6.1. (Appendix 1) Signal flow when creating stimuli



6.2. (Appendix 2) Instructions and demographics for experiment one

Please answer the following

How old are you?

What is your profession?

How long have you been in your profession?

What is your preferred genre of music?

Thank you for your participation

Instructions

Press the buttons labeled A, B, C, D, E, F and G to select what sound to play. Press the play (▶) button to play the select sound and pause with the pause (⏸) button. The sound will start and stop at the selected values on the horizontal sliders and the position slider shows current play-position. To loop the selected section, press the loop button. Feel free to adjust the audio level.

Please listen through all the sounds and place each vertical fader at your corresponding preference rating for each sound. Disregard quality of performance and choice of melody/chords when assessing your preferences. When you are satisfied with your ratings, press next to go to the next page. Bear in mind there will be more stimuli on the following pages that you might prefer more or less, and you can not go back to previous pages.

Please call out when you are done with all the pages or if help is needed.

What influenced your preference ratings? Please describe with as much detail as possible.

6.3. (Appendix 3) Instructions for experiment two test A

Instructions

Press the buttons labeled A, B, C, D, E, F and G to select what sound to play. Press the play (▶) button to play the select sound and pause with the pause (⏸) button. The sound will start and stop at the selected values on the horizontal sliders and the position slider shows current play-position. To loop the selected section, press the loop button. You may ***not*** adjust the audio level.

Please listen through all the sounds and place each vertical fader at a position corresponding to your subjective perception of the amount of ***warmth*** in each sound. Disregard quality of performance and choice of melody/chords when assessing. When you are satisfied with your ratings, press next to go to the next page. Bear in mind there will be more stimuli on the following pages and you can not go back to previous pages.

Please call out when you are done with all the pages or if help is needed.

6.4. (Appendix 4) Instructions for experiment two test B

Instructions

Press the buttons labeled A, B, C, D, E, F and G to select what sound to play. Press the play (▶) button to play the select sound and pause with the pause (⏸) button. The sound will start and stop at the selected values on the horizontal sliders and the position slider shows current play-position. To loop the selected section, press the loop button. You may ***not*** adjust the audio level.

Please listen through all the sounds and place each vertical fader at a position corresponding to your subjective perception of the amount of ***roughness*** in each sound. Disregard quality of performance and choice of melody/chords when assessing. When you are satisfied with your ratings, press next to go to the next page. Bear in mind there will be more stimuli on the following pages and you can not go back to previous pages.

Please call out when you are done with all the pages or if help is needed.

6.5. (Appendix 5) Results experiment one

6.5.1. Mauchly's Test of Sphericity (main effects)

Mauchly's Test of Sphericity[a]							
Measure:	Preference						
Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Instrument	1,000	0,000	-	-	1,000	1,000	1,000
Type	0,512	12,704	2	0,002	0,672	0,703	0,500
Level	1,000	0,000	-	-	1,000	1,000	1,000

6.5.2. Tests of Within-Subjects Effects (main effects)

Tests of Within-Subjects Effects						
Measure:	Preference					
Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Instrument	Sphericity Assumed	2,893	1	2,893	0,011	0,916
	Greenhouse-Geisser	2,893	1,000	2,893	0,011	0,916
	Huynh-Feldt	2,893	1,000	2,893	0,011	0,916
	Lower-bound	2,893	1,000	2,893	0,011	0,916
Error(Instrument)	Sphericity Assumed	5119,274	20	255,964	-	-
	Greenhouse-Geisser	5119,274	20,000	255,964	-	-
	Huynh-Feldt	5119,274	20,000	255,964	-	-
	Lower-bound	5119,274	20,000	255,964	-	-
Type	Sphericity Assumed	41881,526	2	20940,763	42,174	0,000
	Greenhouse-Geisser	41881,526	1,344	31151,555	42,174	0,000
	Huynh-Feldt	41881,526	1,406	29783,609	42,174	0,000
	Lower-bound	41881,526	1,000	41881,526	42,174	0,000
Error(Type)	Sphericity Assumed	19861,516	40	496,538	-	-
	Greenhouse-Geisser	19861,516	26,889	738,652	-	-
	Huynh-Feldt	19861,516	28,124	706,215	-	-
	Lower-bound	19861,516	20,000	993,076	-	-
Level	Sphericity Assumed	20000,099	1	20000,099	34,898	0,000
	Greenhouse-Geisser	20000,099	1,000	20000,099	34,898	0,000
	Huynh-Feldt	20000,099	1,000	20000,099	34,898	0,000
	Lower-bound	20000,099	1,000	20000,099	34,898	0,000
Error(Level)	Sphericity Assumed	11462,151	20	573,108	-	-
	Greenhouse-Geisser	11462,151	20,000	573,108	-	-
	Huynh-Feldt	11462,151	20,000	573,108	-	-
	Lower-bound	11462,151	20,000	573,108	-	-

6.5.3. Pairwise Comparisons (main effects)

Pairwise Comparisons

Measure:	Preference					
(I) Type	(J) Type	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
1	2	-22.351*	3,450	0,000	-31,366	-13,337
	3	-30.494*	4,351	0,000	-41,861	-19,127
2	1	22.351*	3,450	0,000	13,337	31,366
	3	-8.143*	2,152	0,003	-13,766	-2,520
3	1	30.494*	4,351	0,000	19,127	41,861
	2	8.143*	2,152	0,003	2,520	13,766

6.5.4. Mauchly's Test of Sphericity (interaction effects)

Mauchly's Test of Sphericity[a]

Measure:	Preference						
Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Instrument * Type	0,851	3,073	2	0,215	0,870	0,946	0,500
Instrument * Level	1,000	0,000	-	-	1,000	1,000	1,000
Type * Level	0,995	0,102	2	0,950	0,995	1,000	0,500
Instrument * Type * Level	0,762	5,155	2	0,076	0,808	0,869	0,500

6.5.5. Tests of Within-Subjects Effects (interaction effects)

Tests of Within-Subjects Effects						
Measure:	Preference					
Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Instrument * Type	Sphericity Assumed	3465,554	2	1732,777	15,117	0,000
	Greenhouse -Geisser	3465,554	1,740	1991,518	15,117	0,000
	Huynh-Feldt	3465,554	1,892	1831,917	15,117	0,000
	Lower-bound	3465,554	1,000	3465,554	15,117	0,001
Error(Instrument*Type)	Sphericity Assumed	4584,905	40	114,623	-	-
	Greenhouse -Geisser	4584,905	34,803	131,738	-	-
	Huynh-Feldt	4584,905	37,835	121,181	-	-
	Lower-bound	4584,905	20,000	229,245	-	-
Instrument * Level	Sphericity Assumed	906,683	1	906,683	8,007	0,010
	Greenhouse -Geisser	906,683	1,000	906,683	8,007	0,010
	Huynh-Feldt	906,683	1,000	906,683	8,007	0,010
	Lower-bound	906,683	1,000	906,683	8,007	0,010
Error(Instrument*Level)	Sphericity Assumed	2264,817	20	113,241	-	-
	Greenhouse -Geisser	2264,817	20,000	113,241	-	-
	Huynh-Feldt	2264,817	20,000	113,241	-	-
	Lower-bound	2264,817	20,000	113,241	-	-
Type * Level	Sphericity Assumed	40,300	2	20,150	0,152	0,859
	Greenhouse -Geisser	40,300	1,989	20,258	0,152	0,858
	Huynh-Feldt	40,300	2,000	20,150	0,152	0,859
	Lower-bound	40,300	1,000	40,300	0,152	0,700
Error(Type*Level)	Sphericity Assumed	5288,825	40	132,221	-	-

	Greenhouse-Geisser	5288,825	39,787	132,930	-	-
	Huynh-Feldt	5288,825	40,000	132,221	-	-
	Lower-bound	5288,825	20,000	264,441	-	-
Instrument * Type * Level	Sphericity Assumed	300,264	2	150,132	1,106	0,341
	Greenhouse-Geisser	300,264	1,616	185,806	1,106	0,332
	Huynh-Feldt	300,264	1,737	172,846	1,106	0,335
	Lower-bound	300,264	1,000	300,264	1,106	0,305
Error(Instrument*Type*Level)	Sphericity Assumed	5428,611	40	135,715	-	-
	Greenhouse-Geisser	5428,611	32,320	167,963	-	-
	Huynh-Feldt	5428,611	34,744	156,248	-	-
	Lower-bound	5428,611	20,000	271,431	-	-

6.5.6. Mean, standard error and 95 % confidence intervals (interaction effects)

Instrument * Type					
Measure:	Preference				
Instrument	Type	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
1	1	17,405	4,231	8,578	26,231
	2	45,988	3,388	38,922	53,054
	3	56,738	3,108	50,255	63,221
2	1	27,667	4,024	19,273	36,061
	2	43,786	2,805	37,934	49,637
	3	49,321	3,378	42,275	56,368

Instrument * Level					
Measure:	Preference				
Instrument	Level	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
1	1	47,056	3,095	40,599	53,512
	2	33,032	3,027	26,717	39,347
2	1	51,063	3,212	44,363	57,764
	2	29,452	3,479	22,196	36,709

6.6. (Appendix 6) Results experiment two - test A

6.6.1. Mauchly's Test of Sphericity (main effects)

Mauchly's Test of Sphericity[a]							
Measure:	Warmth						
Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Instrument	1,000	0,000	-	-	1,000	1,000	1,000
Type	0,275	23,250	2	0,000	0,580	0,594	0,500
Level	1,000	0,000	-	-	1,000	1,000	1,000

6.6.2. Tests of Within-Subjects Effects (main effects)

Tests of Within-Subjects Effects						
Measure:	Warmth					
Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Instrument	Sphericity Assumed	456,504	1	456,504	2,121	0,162
	Greenhouse-Geisser	456,504	1,000	456,504	2,121	0,162
	Huynh-Feldt	456,504	1,000	456,504	2,121	0,162
	Lower-bound	456,504	1,000	456,504	2,121	0,162
Error(Instrument)	Sphericity Assumed	4089,079	19	215,215	-	-
	Greenhouse-Geisser	4089,079	19,000	215,215	-	-
	Huynh-Feldt	4089,079	19,000	215,215	-	-
	Lower-bound	4089,079	19,000	215,215	-	-
Type	Sphericity Assumed	21041,308	2	10520,654	13,109	0,000
	Greenhouse-Geisser	21041,308	1,159	18150,079	13,109	0,001
	Huynh-Feldt	21041,308	1,188	17718,932	13,109	0,001
	Lower-bound	21041,308	1,000	21041,308	13,109	0,002
Error(Type)	Sphericity Assumed	30497,858	38	802,575	-	-
	Greenhouse-Geisser	30497,858	22,027	1384,591	-	-
	Huynh-Feldt	30497,858	22,563	1351,701	-	-
	Lower-bound	30497,858	19,000	1605,150	-	-
Level	Sphericity Assumed	21451,504	1	21451,504	49,523	0,000
	Greenhouse-Geisser	21451,504	1,000	21451,504	49,523	0,000
	Huynh-Feldt	21451,504	1,000	21451,504	49,523	0,000
	Lower-bound	21451,504	1,000	21451,504	49,523	0,000
Error(Level)	Sphericity Assumed	8230,079	19	433,162	-	-
	Greenhouse-Geisser	8230,079	19,000	433,162	-	-
	Huynh-Feldt	8230,079	19,000	433,162	-	-
	Lower-bound	8230,079	19,000	433,162	-	-

6.6.3. Pairwise Comparisons (main effects)

Pairwise Comparisons						
Measure:	Warmth					
(I) Type	(J) Type	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
1	2	-15.150*	5,345	0,032	-29,181	-1,119
	3	-22.488*	5,352	0,001	-36,538	-8,437
2	1	15.150*	5,345	0,032	1,119	29,181
	3	-7.338*	1,726	0,001	-11,868	-2,807
3	1	22.488*	5,352	0,001	8,437	36,538
	2	7.338*	1,726	0,001	2,807	11,868

6.6.4. Mauchly's Test of Sphericity (interaction effects)

Mauchly's Test of Sphericity[a]							
Measure:	Warmth						
Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Instrument * Type	0,867	2,568	2	0,277	0,883	0,966	0,500
Instrument * Level	1,000	0,000	-	-	1,000	1,000	1,000
Type * Level	0,978	0,408	2	0,816	0,978	1,000	0,500
Instrument * Type * Level	0,959	0,754	2	0,686	0,961	1,000	0,500

6.6.5. Tests of Within-Subjects Effects (interaction effects)

Tests of Within-Subjects Effects						
Measure:	Warmth					
Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Instrument * Type	Sphericity Assumed	825,408	2	412,704	1,985	0,151
	Greenhouse-Geisser	825,408	1,765	467,572	1,985	0,158
	Huynh-Feldt	825,408	1,933	427,118	1,985	0,153
	Lower-bound	825,408	1,000	825,408	1,985	0,175
Error(Instrument*Type)	Sphericity Assumed	7901,758	38	207,941	-	-
	Greenhouse-Geisser	7901,758	33,541	235,586	-	-
	Huynh-Feldt	7901,758	36,718	215,203	-	-
	Lower-bound	7901,758	19,000	415,882	-	-
Instrument * Level	Sphericity Assumed	19,837	1	19,837	0,090	0,768
	Greenhouse-Geisser	19,837	1,000	19,837	0,090	0,768
	Huynh-Feldt	19,837	1,000	19,837	0,090	0,768
	Lower-bound	19,837	1,000	19,837	0,090	0,768
Error(Instrument*Level)	Sphericity Assumed	4208,079	19	221,478	-	-
	Greenhouse-Geisser	4208,079	19,000	221,478	-	-
	Huynh-Feldt	4208,079	19,000	221,478	-	-
	Lower-bound	4208,079	19,000	221,478	-	-
Type * Level	Sphericity Assumed	641,058	2	320,529	1,997	0,150

	Greenhouse-Geisser	641,058	1,956	327,707	1,997	0,151
	Huynh-Feldt	641,058	2,000	320,529	1,997	0,150
	Lower-bound	641,058	1,000	641,058	1,997	0,174
Error(Type*Level)	Sphericity Assumed	6098,108	38	160,477	-	-
	Greenhouse-Geisser	6098,108	37,168	164,070	-	-
	Huynh-Feldt	6098,108	38,000	160,477	-	-
	Lower-bound	6098,108	19,000	320,953	-	-
Instrument * Type * Level	Sphericity Assumed	18,025	2	9,012	0,065	0,937
	Greenhouse-Geisser	18,025	1,921	9,382	0,065	0,931
	Huynh-Feldt	18,025	2,000	9,012	0,065	0,937
	Lower-bound	18,025	1,000	18,025	0,065	0,801
Error(Instrument*Type*Level)	Sphericity Assumed	5233,808	38	137,732	-	-
	Greenhouse-Geisser	5233,808	36,503	143,382	-	-
	Huynh-Feldt	5233,808	38,000	137,732	-	-
	Lower-bound	5233,808	19,000	275,464	-	-

6.7. (Appendix 7) Results experiment two – test B

6.7.1. Mauchly's Test of Sphericity (main effects)

Mauchly's Test of Sphericity[a]							
Measure:	Roughness						
Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Instrument	1,000	0,000	-	-	1,000	1,000	1,000
Type	0,545	10,305	2	0,006	0,687	0,726	0,500

6.7.2. Tests of Within-Subjects Effects (main effects)

Tests of Within-Subjects Effects						
Measure:	Roughness					
Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Instrument	Sphericity Assumed	2,320	1	2,320	0,005	0,942
	Greenhouse-Geisser	2,320	1,000	2,320	0,005	0,942
	Huynh-Feldt	2,320	1,000	2,320	0,005	0,942
	Lower-bound	2,320	1,000	2,320	0,005	0,942
Error(Instrument)	Sphericity Assumed	7776,596	18	432,033	-	-
	Greenhouse-Geisser	7776,596	18,000	432,033	-	-
	Huynh-Feldt	7776,596	18,000	432,033	-	-
	Lower-bound	7776,596	18,000	432,033	-	-
Type	Sphericity Assumed	19737,535	2	9868,768	21,465	0,000
	Greenhouse-Geisser	19737,535	1,375	14354,861	21,465	0,000

	Huynh-Feldt	19737,535	1,451	13601,834	21,465	0,000
	Lower-bound	19737,535	1,000	19737,535	21,465	0,000
Error(Type)	Sphericity Assumed	16551,298	36	459,758	-	-
	Greenhouse-Geisser	16551,298	24,749	668,753	-	-
	Huynh-Feldt	16551,298	26,120	633,671	-	-
	Lower-bound	16551,298	18,000	919,517	-	-
Level	Sphericity Assumed	55742,215	1	55742,215	139,046	0,000
	Greenhouse-Geisser	55742,215	1,000	55742,215	139,046	0,000
	Huynh-Feldt	55742,215	1,000	55742,215	139,046	0,000
	Lower-bound	55742,215	1,000	55742,215	139,046	0,000
Error(Level)	Sphericity Assumed	7216,035	18	400,891	-	-
	Greenhouse-Geisser	7216,035	18,000	400,891	-	-
	Huynh-Feldt	7216,035	18,000	400,891	-	-
	Lower-bound	7216,035	18,000	400,891	-	-

6.7.3. Pairwise Comparisons (main effects)

Pairwise Comparisons						
Measure:	Roughness					
(I) Type	(J) Type	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
1	2	13.132*	3,477	0,004	3,955	22,308
	3	22.697*	4,378	0,000	11,143	34,252
2	1	-13.132*	3,477	0,004	-22,308	-3,955
	3	9.566*	2,245	0,001	3,642	15,490
3	1	-22.697*	4,378	0,000	-34,252	-11,143
	2	-9.566*	2,245	0,001	-15,490	-3,642

6.7.4. Mauchly's Test of Sphericity (interaction effects)

Mauchly's Test of Sphericity[a]							
Measure:	Roughness						
Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Instrument * Type	0,911	1,577	2	0,455	0,919	1,000	0,500
Instrument * Level	1,000	0,000	-	-	1,000	1,000	1,000
Type * Level	0,908	1,639	2	0,441	0,916	1,000	0,500
Instrument * Type * Level	0,880	2,177	2	0,337	0,893	0,984	0,500

6.7.5. Tests of Within-Subjects Effects (interaction effects)

Tests of Within-Subjects Effects						
Measure:	Roughness					
Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Instrument * Type	Sphericity Assumed	2214,588	2	1107,294	6,371	0,004
	Greenhouse-Geisser	2214,588	1,837	1205,382	6,371	0,006
	Huynh-Feldt	2214,588	2,000	1107,294	6,371	0,004
	Lower-bound	2214,588	1,000	2214,588	6,371	0,021
Error(Instrument*Type)	Sphericity Assumed	6257,246	36	173,812	-	-
	Greenhouse-Geisser	6257,246	33,071	189,209	-	-

	Huynh-Feldt	6257,246	36,000	173,812	-	-
	Lower-bound	6257,246	18,000	347,625	-	-
Instrument * Level	Sphericity Assumed	837,583	1	837,583	5,161	0,036
	Greenhouse-Geisser	837,583	1,000	837,583	5,161	0,036
	Huynh-Feldt	837,583	1,000	837,583	5,161	0,036
	Lower-bound	837,583	1,000	837,583	5,161	0,036
Error(Instrument*Level)	Sphericity Assumed	2921,000	18	162,278	-	-
	Greenhouse-Geisser	2921,000	18,000	162,278	-	-
	Huynh-Feldt	2921,000	18,000	162,278	-	-
	Lower-bound	2921,000	18,000	162,278	-	-
Type * Level	Sphericity Assumed	494,851	2	247,425	2,238	0,121
	Greenhouse-Geisser	494,851	1,832	270,162	2,238	0,126
	Huynh-Feldt	494,851	2,000	247,425	2,238	0,121
	Lower-bound	494,851	1,000	494,851	2,238	0,152
Error(Type*Level)	Sphericity Assumed	3979,649	36	110,546	-	-
	Greenhouse-Geisser	3979,649	32,970	120,704	-	-
	Huynh-Feldt	3979,649	36,000	110,546	-	-
	Lower-bound	3979,649	18,000	221,092	-	-
Instrument * Type * Level	Sphericity Assumed	66,009	2	33,004	0,216	0,807
	Greenhouse-Geisser	66,009	1,785	36,971	0,216	0,782
	Huynh-Feldt	66,009	1,969	33,528	0,216	0,803
	Lower-bound	66,009	1,000	66,009	0,216	0,648
Error(Instrument*Type*Level)	Sphericity Assumed	5500,158	36	152,782	-	-
	Greenhouse-Geisser	5500,158	32,137	171,146	-	-
	Huynh-Feldt	5500,158	35,438	155,206	-	-
	Lower-bound	5500,158	18,000	305,564	-	-

6.7.6. Mean, standard error and 95 % confidence intervals (interaction effects)

Instrument * Type

Measure:	Roughness				
Instrument	Type	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
1	1	68,526	4,098	59,918	77,135
	2	54,000	3,636	46,360	61,640
	3	38,632	3,931	30,373	46,890
2	1	63,000	4,113	54,360	71,640
	2	51,263	3,663	43,568	58,958
	3	47,500	4,175	38,729	56,271

Instrument * Level

Measure:	Roughness				
Instrument	Level	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
1	1	40,000	3,387	32,884	47,116
	2	67,439	2,944	61,253	73,624
2	1	36,368	3,896	28,184	44,553
	2	71,474	3,805	63,479	79,469