Meaningful noise : auditory roughness and dissonance predict emotion recognition and cross-modal perception

Liew, Kong Meng

2018

Lim, K. M. (2018). Meaningful noise : auditory roughness and dissonance predict emotion recognition and cross-modal perception. Master's thesis, Nanyang Technological University, Singapore.

<http://hdl.handle.net/10356/73083>

<https://doi.org/10.32657/10356/73083>

**MEANINGFUL NOISE: AUDITORY ROUGHNESS AND
DISSONANCE PREDICT EMOTION RECOGNITION AND
CROSS-MODAL PERCEPTION**

LIEW KONG MENG

A thesis submitted to the School of Art, Design, and Media, Nanyang Technological University, in fulfilment of the requirements for the degree of Master of Arts (Research)

Supervisor:

Asst. Professor PerMagnus Lindborg, School of Art, Design, and Media

Co-Supervisor:

Nanyang Asst. Professor Suzy Styles, Division of Psychology, School of Social
Sciences

Singapore, 18 August 2017

Acknowledgements

I would like to thank all the people that have made this thesis possible, who have supported me one way or another over the past two years. Firstly, I'd like to thank my supervisor, Dr. PerMagnus Lindborg, for all his kindness and knowledge, the tireless encouragement and the support, and for believing in me even before I embarked on this journey in research. I am also grateful to my co-supervisor Dr. Suzy Styles for her dedication and wisdom, and for her unending patience in easing me into the intricacies of psychological research.

I'd like to thank the staff at the School of Art, Design and Media (ADM) and the Division of Psychology, in particular Hong Bee Kuen and Lela Ahmad for putting up with all my unusual requests, as well as the capable professors of the many graduate modules I took for their refinement of my thinking styles and research strategies: Andrea Nanetti, Michael John Kirk Walsh, Heitor Capuzzo, Xu Hong and Andy Ho. I am also grateful for all the friendships I've made while studying here, for all the times that we've discussed our research and helped each other out. I am thankful for the financial support from ADM and the opportunity to study as student of the school.

I'd also like to thank my family for all their love and trust, and for allowing me to take the leap of faith in pursuing this programme, while wholeheartedly supporting me in any way they could.

To my beautiful wife Alethea, thank you for walking this journey with me, be it proofreading my papers at 3am in the morning, or taking care of Miso when I was too immersed in writing. Thank you for loving me and I too look forward to spending the rest of our lives together.

Lastly, I'd like to thank God, for the blessings of wisdom and favour. Without him, none of this would have been possible.

Thesis Abstract	7
Foreword	8
1: Introduction to Noise Perception	9
1.1: Auditory Roughness	11
1.2: Dissonance	14
1.3: Perceiving Emotion Through Music	16
1.4: Cross-Modal Perception	19
1.5: Hypothesis	20
1.6: Summary of Methodological Approach	21
2: Roughness Predicts Emotion Recognition in Single Chords (Study 1)	22
2.1: Background	23
2.2: Procedure	23
2.3: Results	24
2.4: Discussion	28
3: The Effects of Dissonance in Single-Tone Instrumental Stimuli (Study 2)	29
3.1: Background	30
3.2: Corpus of Instrumental Sounds	30
3.3: Methods	31
3.4: Results	34
3.5: Discussion	42

4: Auditory Roughness and Dissonance in Sounds Generated Through Frequency Modulation Synthesis (Study 3)	43
4.1: Background	44
4.2: Methods	44
4.3: Results	46
4.4: Discussion	51
Chapter 5: <i>Rama</i> : An Interactive Musical Interface and Multimedia Performance Environment	57
5.1: Background to <i>Rama</i>	58
5.2: Graphical Interface for Audience Participation	61
5.3: Performance Premiere	63
Chapter 6: Cross-Modal Shape Perception in a Contemporary Music Concert Setting (Study 4)	64
6.1: Background	65
6.2: Methods	65
6.3: Results	69
6.4: Discussion	71
Chapter 7: Conclusion and Future Directions	73
7.1: An Evaluation of <i>Rama</i> : Environment versus Laboratory	76
7.2: Future Research Directions	77
Bibliography	81

Thesis Abstract

Auditory roughness perception is an important, low-level component in dissonance perception. Four experiments on single-note stimuli were conducted to explore how roughness and dissonance from instrumental timbres and synthesised sound affect emotion recognition and cross-modal perception in listeners. Results suggest that auditory roughness and dissonance are important in conveying emotional properties like arousal and valence, as well as information on visual shape and texture.

Keywords: auditory roughness; dissonance; emotion; cross-modal perception

Foreword

As a composer working with expanded musical definitions in the creation of art, I often have to work with the concept of *noise* as *music*. These are two seemingly opposed notions that the typical listener does not associate with each other. The conventional definition of noise already precludes music by stating that it refers to sound that is unmusical or unpleasant (noise, n.d.). As such, the intent behind this thesis is to explore on the lowest possible level, the implicit, arguably unconscious mechanisms at play behind the roughest sound that continually influences the listener. To accomplish this, this thesis will adopt the methodology of scientific rigour, applying empirical and quantitative analyses to examine the perceptual and cognitive mechanisms in noise processing.

By doing so, I hope this research will play a role in contributing to humanity's ongoing progress in understanding the science of music. Supplementary materials are available online at <https://osf.io/58udy>.

1. Introduction to Noise Perception

At the turn of the 20th Century, the Italian futurist composer Luigi Russolo came to a realisation that with the advent of noisy machinery, audiences in contemporary European societies had begun to lose their sensitivity to music. Composers had to expand their musical boundaries, by breaking free of centuries of tradition in music; of the excessive attention to Pythagorean ratios in musical harmony. With his 1913 manifesto *The Art of Noise* he declared that:

“Nowadays musical art aims at the shrillest, strangest and most dissonant amalgams of sound... we are approaching noise-sound” (Russolo, 1913/1967).

Today, over a hundred years later, noise has entrenched its place in the world of contemporary art and music, working its way into a wide range of music from experimental composers (e.g. Karlheinz Stockhausen and Iannis Xenakis), to pop musicians (e.g. The Beatles) and audiovisual artists (e.g. Ryoji Ikeda). While musicologists and philosophers have long debated the distinction between music and noise, such as Theodore Adorno’s (1970/1984) stand that music may share the same origins as noise but is differentiated on its rationality and organisation, this thesis will not address this discourse. Instead, this thesis will adopt a dictionary understanding of noise, defined as sound that is lacking “agreeable musical quality” or is “noticeably unpleasant” (noise, n.d.). In doing so, I hope to direct attention to the perceptual effects of noise, rather than a discussion on the aesthetic distinction of music from noise.

Despite the use of noise elements in music being widespread today, there is a lack of psychological research on the low-level perceptual and affective mechanisms involved when listening to noise music. This thesis seeks to address this gap by focusing on the base element of timbre: specifically examining the consequences of

dissonance caused by formant structure¹ from the spectrum² of a single sound devoid of its musical contexts, on emotion recognition and cross-modal perception.

Auditory Roughness

In reference to the earlier definition of noise as sounds that are unpleasant, auditory roughness is a property of sound that has been associated with annoyance (McDermott, 2012). It was first coined by the German psychologist Hermann von Helmholtz in 1885, in reference to the raspy, harsh buzz caused by sounds in close harmonic intervals (Vassilakis, 2005).

Auditory roughness is caused by constant, rapid amplitude fluctuations in the sound spectrum (resembling an auditory sensation of buzzing) in the range of 20 - 200 Hz (Parncutt, 1989), or 15 - 300 Hz (Fastl & Zwicker, 2007; Zwicker, 1961). In a harmonic relationship, it is perceived due to the inability of the human auditory system (basilar membrane) to perceive frequency differences within the critical band - the narrow range of frequencies where physical limitations of the auditory system impair the accurate perception of each frequency in the harmonic relationship as distinct (Campbell & Greated, 1987; Vassilakis, 2005).

One way of creating these sensations of auditory roughness is through the direct amplitude modulation (AM) of carrier frequencies, which is a form of external interference resulting in rapidly changing loudness levels (Fastl & Zwicker, 2007), found in, for example, a combustion engine of an old car. Secondly, auditory roughness

¹ Formant structure refers to selected frequencies in a sound spectrum of a complex sound which peak in intensity (Fant, 1960).

² The sound spectrum refers to the different frequencies that are present in a sound (Wolfe, 1997).

can also be created by phase interference from narrow harmonic intervals, and occurs in speech, musical instruments and other harmonic sounds. To understand the mechanisms behind phase interference, this explanation will start with a basic illustration of phase interference in two pure sinusoidal waves. A pure sinusoidal wave of equation $y = \sin(x)$ creates a graph as seen in Fig. 1.1. Note that the equation is for an ideal sinusoidal wave that has existed since the dawn of time and extends into eternity. In the diagram, only one period of the graph is shown. A pure sinusoidal wave with the inverse phase relationship creates a wave as seen in Fig 1.2. Should both sinusoidal waves be heard simultaneously, the inverse phase relationship would nullify the resultant waveform, reducing the amplitude to constant zero, and no sound will be heard.

Fig. 1.1. $y = \sin(x)$

Fig. 1.2. $y = -\sin(x)$

In the case of roughness, slight mistuning in the second frequency causes the relationship of the two waves to be slightly out of phase. As a result, the sound waves interfere with each other in a similar manner, but instead of completely cancelling each other out as in the perfectly inverse phase relationship of Fig. 1.1 and Fig. 1.2, only slight amplitude fluctuations will be observed, as seen in Fig. 1.3. In this example, the top and middle rows are representative of a sinusoidal wave at 10 Hz and 12 Hz respectively. When both tones are played simultaneously, the resultant wave is seen in

the bottom row, and has an amplitude fluctuation resulting in beatings of 2 Hz. The amplitude of the wave corresponds to intensity in sound, and the distinct amplitude increases and decreases in the wave translate to increases and decreases in volume. This creates amplitude fluctuations, which are similar to the direct amplitude modulation of a carrier frequency.

Figure 1.3: The top and middle rows are sinusoidal waves of 10 Hz and 12 Hz respectively, and the bottom row is the resultant wave. Image from Davis, D. (2003).

Beats (Graph illustration). Retrieved from

<http://www.ux1.eiu.edu/~cfadd/3050/Ch12Sound/beats.html>

Auditory roughness can thus be modelled by calculating the difference in amplitudes of these two frequencies against the critical band (Zwicker, 1961), and in complex sounds, the prominent formants in the spectrum of specific sounds are isolated into pairs and similarly analysed. (MacCallum & Einbond, 2008; Bernardes, Davies, Guedes, & Pennycook, 2014). The sum of these additional amplitude fluctuations in the

sound spectrum adds to the roughness and thus arguably affects the overall dissonance of a sound (Kameoka & Kuriyagawa, 1969; Sethares, 1993).

Dissonance

“The two concepts (consonance and dissonance) have never been completely explained, and for a thousand years the definitions have varied” (Hindemith, 1942).

The concept of dissonance correlating most with the auditory roughness of a sound is referred to as “sensory dissonance” (Sethares, 1993). However, a holistic definition of dissonance has eluded musicians and musicologists for generations. The early 20th Century German composer Hindemith, for example, laments this shifting of its definition throughout history, by citing the example of how an interval (such as a *major third*) can be considered both dissonance and consonant depending on the historical point of reference (Tenney, 1988). Dissonance was first conceptualised by the ancient Greeks more than two millennia ago. To them, consonance was based on certain intervallic ratios corresponding to *fourths*, *fifths* and *octaves* on a musical scale, when they noticed the how neighbouring strings vibrated sympathetically when a central string was vibrated according to these ratios (Tenney, 1988). Dissonance was then assumed to be all other intervals, and this ambiguity became a point of contention for the definition of dissonance in the rest of Western music history. Nevertheless, there appeared to be an innate understanding of dissonance, which corresponded to the properties of auditory roughness. As early as the Late Renaissance, the music theorist Johann Joseph Fux was aware of the tension and resolution implied by cadential points, which corresponded to a modern knowledge of auditory roughness (Parncutt & Hair, 2011). However, it was only with Hermann von Helmholtz’s discovery of the phenomenon of auditory roughness in the 19th Century that a link between roughness and dissonance could be identified. He discovered that auditory roughness was notably

lower in musical intervals that are traditionally known as consonant, such as the *octave* and the *fifth* (Ball, 2008). More recently, Johnson-Laird, Kang and Leong (2012) have also found that auditory roughness plays a role in dissonance perception in musical chords, particularly in tonal contexts, with major chords being less rough and perceptibly less dissonant than minor chords. Schoon, Regnault, Ystad & Besson, (2005) found that rough musical intervals presented simultaneously (harmonic) were rated as more dissonant than intervals presented sequentially (melodic) adding support to auditory roughness being a major perceptual feature in vertical harmony.

However, recent research in dissonance has suggested that the physical properties of sound that define dissonance, such as auditory roughness, may not account entirely for dissonance perception. Perceived dissonance, which in this paper refers to a perceptual understating of a sound as dissonant, has been shown to be more complex. For example, the integration of contextual information could be an important component of dissonance perception. Single cell recordings of neural activation in the auditory cortex of rats show enhanced representation on sounds when expectation is heightened (Jaramillo & Zador, 2011). This supports the growing evidence for familiarisation via short-term musical memory as an important factor in determining perceived dissonance (Jensen & Hjortljaer, 2012; McLachlan, Marco, Light, & Wilson, 2013). While this thesis will not be focusing on contextual information in perceived dissonance, it is important to note the difficulty of evaluating dissonance as being based solely on physical properties, and as such, the exact mechanism between the physical and the perceptual remains complex and undefined.

As such, in this thesis, perceived dissonance refers to the individual judgement of a sonic object as dissonant, based on both psychoacoustic factors and musical (contextual) factors (Johnson-Laird, Kang and Leong, 2012). Sensory dissonance refers

to the auditory sensation that arises solely from auditory roughness in the spectral domain of sounds, and may not necessarily be judged as dissonant by listeners.

An additional note of caution must be mentioned about the impact of roughness on consonance. Sethares (1993) has argued that the concept of consonance, as a polar opposite of dissonance, is simply defined by the absence of roughness. However recent research has suggested that roughness, although related to dissonance, is independent from consonance, in that decreasing auditory roughness does not imply an increase in consonance perception (McDermott, Lehr, & Oxenham, 2010; Bowling & Purves, 2015). As such, the findings from this thesis cannot be generalized to imply that consonance is associated with emotion recognition or cross-modal perception, even in the absence of dissonance.

Perceiving Emotion Through Music

Perceived emotion in music, or recognised emotion, is defined as the listener's ability to understand the expressed emotion in music, without necessarily feeling the emotion itself (Gabrielsson, 2001-2002; Evans & Schubert, 2008). This is different from the concept of felt or induced emotions, which refers to the actual emotion experienced by the listener at the point of listening to music.

The roots of emotion perception in music stem from early research directions by pioneers in music psychology such as Kate Hevner, Melvin Rigg and Carl Seashore. In the 1930s, researchers focused mainly on the perception of music through the use of complete musical stimuli, as many of the early psychologists were not able to successfully experimentally isolate features of music in their analyses (Hevner, 1936; Juslin & Sloboda, 2010). However, given the rise of the behaviourism movement in psychology in the 1950s and the cognitivism in the 1980s, research in music and emotion was largely side-lined, with the exception of a few psychologists like Leonard

Meyer (Juslin & Sloboda, 2010). His theoretical landmark *Emotion and Meaning in Music* pioneered the linking of mainstream psychology and music theory, by proposing several theories on how experiences of emotion relied on cognitive appraisals of the musical event (Meyer, 1956) and was a major influence on future music and emotion researchers. It was not until the 1990s with the increased focus on emotion psychology that research in music and emotion began to become more mainstream (Juslin & Sloboda, 2010).

A further distinction needs to be made with regard to the concept of emotion as opposed to mood states. While both come under the broader definition of affect, emotion differs from mood in that it is target-directed, high in intensity, short in duration, and resultant effects are distinct and specific to each emotion. Mood, on the other hand, is less intense, only positively or negatively valenced³, longer in duration, and usually lacks context (Frijda, 1993; Batson, Shaw, & Oleson, 1992). Music has thus been used in several studies to manipulate mood states to great effect (Avramova & Stapel, 2008), but its role in emotion is much less clear (Juslin, Friberg, & Bresin, 2002; Scherer, 2004; Juslin, 2013).

This thesis will approach music emotion perception from both discreet and dimensional perspectives; the former implies that emotions are fundamentally distinct from each other (Roseman, 1991; Eerola & Vuoskoski, 2011), and the latter argues that all emotions are on a two-dimensional circumplex plane with arousal and valence as its axes (Russell, 1980). For music and emotion research, arousal can be further discriminated into the dimensions of tension and energy (Vieillard, et al., 2007; Eerola & Vuoskoski, 2011). While research has suggested that the dimensional perspective is

³ Valence refers to the intrinsic positivity (attractiveness) or negativity (averseness) of a target event or stimulus (Frijda, 1986).

no longer effective in comprehensively understanding emotions (Lerner & Keltner, 2000), the dimensions identified here – valence and arousal – nevertheless form parts of the appraisal process of emotions (Tong, 2015). These may then constitute the appraisal dimensions, which an individual may use to construe a discrete emotion. For example, appraisals of high positive valence and high arousal may contribute partly to an individual's experience of happiness, alongside other appraisal dimensions like clarity and difficulty.

While much of the emotion in music literature has focused on musical emotions by examining excerpts from commercially available recordings of music (Eerola & Vuoskoski, 2011), musical emotions have been shown to be recognised through low-level musical and acoustical features such as timbre, tempo, melodic complexity, volume, and articulation (Balkwill, Thompson, & Matsunaga, 2004; Baraldi, De Poli, & Roda, 2006; Lahdelma & Eerola, 2016; Gabrielsson & Lindström, 2010; Hailstone, et al., 2009; Bowman & Yamauchi, 2015), utilising simple musical stimuli (e.g. single chords, simple melodies). Auditory roughness has also been shown to be one of these low-level features in emotion recognition, correlating strongly with the emotion domains of tension-arousal and energy-arousal in a study by Lahdelma and Eerola (2016b) with musical triads. Plomp and Levelt (1965) also discovered that roughness was related to participant ratings of pleasantness and unpleasantness, a concept linked to positive and negative valence in emotion (Lang, Bradley, & Cuthbert, 1997). Given that the classification of noise is also dependant on unpleasant appraisals, auditory roughness and dissonance might play an important role in the perception of musical emotion from noise in music.

There is additional biological evidence to suggest that auditory roughness affects emotion recognition through an unconscious process. The amount of unpleasantness

that depended on dissonance was found to be dependent on the relative position of the dissonant intervals to the critical band (Plomp & Levelt, 1965), mentioned earlier as a physical limitation of the human auditory system. This aversion towards auditory roughness is also based on neurobiology, and can be traced to brain activity in N1 - P2 components in musically trained participants (Schoon, Regnault, Ystad, & Besson, 2005). This is an Electroencephalogram (EEG) Event-Related Potential (ERP) component that is usually associated with musical experience and memory (Kuriki, Kanda, & Hirata, 2006; Tremblay, Ross, Kayo, McClannahan, & Collet, 2014), and the pre-conscious, low-level processing nature implied by these components (Bidelman & Krishnan, 2009; Lamy, Salti, & Bar-Haim, 2008) adds further support to the unconscious processing of auditory roughness. In a separate EEG experiment, the frequency-following responses (FFRs) to pitch salience in musical intervals suggested a neural preference in encoding consonant musical intervals over dissonant musical intervals for both musicians and non-musicians alike (Bidelman & Krishnan, 2009). Such preferences for consonant over dissonant sounds also appear to be present in humans from a very young age, suggesting that this preference could be biological (Zentner & Kagan, 1998).

However, even in such studies, the stimuli presented have been synthesised musical dyads, triads or chords, and this still implies tonality. It remains unclear if a single sound is able to affect emotion recognition through the roughness in its harmonic spectra, and comparisons across multiple instruments on the same single note might be effective in examining this effect. As such, the present research will focus specifically on a comparison of instrumental timbres and dissonance perception in single note stimuli, as well as evaluating its effects on emotion recognition.

Cross-Modal Sensory Mechanisms

Musical features are also associated with features in other sensory modalities. For example, faster music in the major key has been associated with higher saturation and brighter levels of yellow, and slower music in minor keys are associated with darker and more desaturated blue (Palmer, Schloss, Xu, & Prado-León, 2013). Furthermore, research has shown that music relates to taste (in the context of food), with higher pitches associated with sweet and sour tastes, and lower pitches with bitterness (Crisinel & Spence, 2009; Crisinel & Spence, 2010).

Studies in cross-modal transference with music have also suggested that emotion is a mediator in such relationships, especially between music and colour (Palmer, Schloss, Xu, & Prado-León, 2013; Lindborg & Friberg, 2015), and music and taste (Wang, Wang, & Spence, 2016; Kantono, et al., 2015). Other hypotheses include synaesthesia (Ramachandran & Hubbard, 2001), and cross-modal correspondences, where the matching of features across sensory modalities might be due to an innate learning of statistical regularities between modalities in our environment (Spence & Deroy, 2013).

Hypothesis

The present paper will thus focus on examining the effects of dissonance perception - a higher level cognitive function - and auditory roughness perception - a lower level perceptual feature of an acoustic phenomenon related to dissonance perception, on emotion recognition and cross-modal perception. Specifically, I hypothesise that an increase in arousal (energy and tension) ratings and a decrease in valence ratings should correspond to higher levels of dissonance and auditory roughness. Consequently, certain discrete emotions that are strongly associated with these appraisals (e.g. happiness)

should also be similarly affected by dissonance. Additionally, dissonance should also have a significant effect on cross-modal perception, though the direction of relationship needs to be first determined via an exploratory study (Study 2). While the present research does not aim to quantify a relationship between emotion recognition and cross-modal processing, possible links and implications will be discussed.

Summary of Methodological Approach

In order to investigate the effects of dissonance perception and auditory roughness on listeners, a total of four quantitative studies were conducted. The first study, an exploratory study on roughness and emotion, comprised a re-analysis of Lahdelma and Eerola's (2016a) study according to the additional factor of computed auditory roughness in their musical stimuli. This study confirmed the theory that emotion in music could be recognised by means of auditory roughness, but as the stimuli were musical chords, we had a confound of musical harmony. As such, the second study focused on a laboratory study on dissonance perception and roughness in single-tone stimuli over different instruments, with emotion variables and conceptual cross-modal variables as dependants. While dissonance perception could predict several dependant variables, the variation in roughness was too small for a robust effect of roughness could not be found, and a distinction to be made between dissonance perception and auditory roughness. Thus, the third study involved similar methods to the second study, but systematically varied auditory roughness through computer-generated tones. In relating the laboratory findings back to the context of noise in contemporary music, the fourth study explored the ecological validity of earlier findings in cross-modal perception through the guise of a contemporary classical music concert. This enabled a situation where the auditory roughness stimuli could be embedded in other musical sounds and systematically controlled in an environmental setting.

2. Roughness predicts emotion recognition in single chords

(Study 1)

Parts of this chapter were presented as a poster and an oral presentation:

Liew, K., Lindborg, P., & Styles, S. J. (2016, February). *Auditory roughness: A new dimension in cross-modal perception and mood?* Poster session presented at the International Symposium on Cognition and Neuroscience, Singapore.

Liew, K., Lindborg, P., & Styles, S. J. (2016, April). *An evaluation of sensory dissonance in the music of Singapore*. Paper presented at the Intercultural Transfers in Music Conference, Singapore.

Background

In a key study by Lahdelma and Eerola (2016a) single chords in a vertical relationship have been found to convey emotion to listeners. Participants in an online study rated a series of chords (major, minor, augmented, diminished and 7th chords) on several emotional qualities (such as *tension*, *energy*, *valence*, *melancholy*, *interest*, and *happiness*), and results suggested a consistency in emotion ratings on these chords across instruments. However, the mechanisms behind the listener recognition of emotions in these chords are unclear, and it remains a possibility that dissonance or roughness could explain part of this effect. As such, the present study explored the physical quality of auditory roughness in musical triads in existing data.

Procedure

Participant ratings and stimuli were obtained from Lahdelma and Eerola's (2016a) study with the approval of the authors. The stimuli consisted of 14 MIDI piano and string chords in an mp3 format (sample rate: 48 kHz, bit rate: 224kbps, stereo), generated on ProTools HD10, using the virtual instruments Bosendorfer Imperial 10 (Ivory plug-in) for piano, and the Chamber Strings (Vienna Symphonic Library plug-in) for the strings. While a total of 14 chords were used in Lahdelma and Eerola (2016)'s study, the present study focused on the 6 Major and Minor chords (root and inversions) (Figure 2.1), as they formed a controlled subset which differed systematically from each other.

Figure 2.1: Musical chords used in this study

The discrete emotions that were rated in Lahdelma and Eerola's (2016a) study are: *nostalgia*, *melancholy*, *interest*, *happiness* and *tenderness*, along with ratings for the emotional dimensions of *valence*, *tension* and *energy*, as well as preference (*liking*). Each variable was evaluated on a scale of 1 to 5, with 1 being the lowest and 5 being the highest. Emotion ratings of each chord in the study were averaged.

To compute the roughness for each of the 6 chords, MacCallum's (2006) Roughness patch for Max was utilised. The original mp3 stimuli were converted to the WAV format (sample rate: 48 kHz, bit rate: 1536kbps, sample size: 16bit, stereo) with iTunes 12.3, for compatibility with the MacCallum (2006) Roughness patch. Custom sub-patches were added to automate the start and end of testing for each sample, as well as a processing check to ensure consistency and to check for optimisation and mechanical errors during processing. Another major addition was a custom sub-patch that converted the real time, change-based output of roughness into sequential samples taken every 20ms. Due to the stable nature of the stimuli, a mean roughness value was calculated for each stimulus. An independent panel of twelve expert raters then assessed these stimuli for their dissonance values to ensure that roughness and dissonance were closely related in this set of stimuli.

Results

A linear regression on stimuli ($N = 12$) conducted through SPSS on the mean roughness and dissonance ratings from the expert raters revealed that computed roughness was a significant predictor for both roughness (adjusted $R^2 = .74$, $\beta = .87$, $t(10) = 5.6$, $p < .001$), and dissonance (adjusted $R^2 = .55$, $\beta = .77$, $t(10) = 3.8$, $p < .003$).

Figure 2.2: Graph of computed roughness and expert ratings of dissonance for each stimulus

Figure 2.3: Graph of computed roughness and expert ratings of roughness for each stimulus

A series of ANCOVAs were conducted in SPSS to compare the effects of instrument (piano or strings) and chord quality (major or minor) on each emotion, with roughness as a covariate. Roughness had a positive effect on the emotional dimensions of energy, $F(1, 8) = 15.16, p = .005, \eta_p^2 = .66$, valence $F(1, 8) = 15.17, p = .005, \eta_p^2 = .66$, and tension $F(1, 8) = 13.29, p = .007, \eta_p^2 = .62$. Roughness also had a positive effect on the discrete emotions of happiness $F(1, 8) = 10.90, p = .011, \eta_p^2 = .58$, interest $F(1, 8) = 16.30, p = .004, \eta_p^2 = .67$, and a negative effect on melancholy $F(1, 8) = 7.76, p = .024, \eta_p^2 = .49$. Roughness had no significant effect on nostalgia, tenderness and liking, all $ps > .090$.

Figure 2.4. Roughness and its corresponding effects on mean emotion ratings for each of the chord stimuli. Legend: * = $p < .05$, ** = $p < .01$, *** = $p < .001$.

Discussion

The results corroborate recent findings by Lahdelma and Eerola (2016b) in that roughness is shown to affect ratings in the dimensional models of emotion – *valence*, *tension* and *energy*. Furthermore, the results of this study suggest that roughness has an effect on specific, discrete emotions: *happiness*, *melancholy* and *interest*, but not *nostalgia*, and *tenderness*. *Happiness*, *interest* and *melancholy* can be classified as positive and negative emotions (Solomon & Stone, 2002; Storm & Storm, 1987; Garrido & Schubert, 2011; Fredrickson, 2009), which are in turn, affected by valence (Russell, 1980). *Tenderness* and *nostalgia*, however, are much less clear in terms of valence, and as such are less affected by roughness than clearly positively or negatively valenced emotions. Furthermore, interest can also be classified as an epistemological emotion: a knowledge-related emotion that motivates learning towards the new and unfamiliar (Silvia, 2008). The effect of roughness on interest in this study appears to be larger than the effect of instrument (piano or strings) or chord quality (major or minor) in the original study (Lahdelma & Eerola, 2016), where it was reported that neither of these variables predicted interest. Hence, the epistemological aspect of interest could thus be strongly affected by roughness, in that sounds high in roughness can perhaps be appraised as complex or novel, both key cognitive appraisals in the emotion of interest (Silvia, 2008).

It is also noteworthy that unlike previous studies in “sensory” dissonance (Plomp & Levelt, 1965), roughness had no noticeable effect on liking. This could be due to the presence of confounding tonal factors, such as major and minor chords, which affect dissonance perception as well (Johnson-Laird, Kang, & Leong, 2012). As such a subsequent study is needed where musical chords are avoided altogether.

3. The effects of dissonance in single-tone instrumental stimuli

(Study 2)

Parts of this research were presented at two oral presentations, one of which with proceedings paper.

Liew, K., Lindborg, P., & Styles, S. J. (2016, October). *Audiences' perception of dissonance aids emotion recognition in abstract music*. Paper presented at the 2016 Performers Present Symposium, Singapore.

Liew, K., Styles, S. J., & Lindborg, P. (2017). Dissonance and roughness in cross-modal perception. *Proceedings of the 6th Conference of the Asia Pacific Society for the Cognitive Sciences of Music*, Kyoto, Japan.

Background

Following the findings of Study 1, Study 2 was designed to investigate roughness and dissonance on emotion, but without the additional complexity of tonal chords.

Furthermore, instead of computer-generated stimuli on virtual instruments, which could have been digitally processed and augmented by commercial developers, the stimuli for study 2 consisted of recorded sounds instead.

Corpus of Instrumental Sounds

As several recordings of live instruments were required to be the stimuli for Study 2, an archive of instrumental sounds was created. This archive consists of recordings of several stringed and wind instruments playing single notes, beyond the four selected as stimuli for this study. For each instrument, single note recordings from the major and minor triads on G, D, A, E, C were taken, with options for vibrato, non-vibrato, loud, soft, and various extended playing techniques (e.g. *sul. ponticello* for stringed instruments and ‘airy’ blowing techniques for flutes). The instruments used in this archive include: violin, viola, cello, flute, bass flute, *erhu* 二胡 (Chinese fiddle), *bangdi* 梆笛 (Chinese soprano flute), *qudi* 曲笛 (Chinese alto flute), *xiao* 箫 (Chinese end-blown flute), *hulusi* 葫芦丝 (Chinese cucurbit flute), *shakuhachi* 尺八 (Japanese end-blown flute), and *bansuri* (Indian flute). All recordings were made by professional musicians, and recorded with an AKG C414 XLS condenser microphone in a sound-attenuating recording studio at ADM, and recorded as a WAV file (sample rate: 48 kHz, sample size: 16bit, stereo).

The development of this archive should result in its eventual publication as a corpus of instrumental sounds with computed roughness values for each sound on open-access platforms such as the Open Science Framework. This may be of help to other

researchers of music who prefer live-recorded stimuli instead of computer generated ones and would not otherwise have had an access to such resources.

Methods

Participants

Participants ($n = 20$, mean age = 26.3, $SD = 5.6$) were recruited from the School of Art, Design, and Media (ADM), Nanyang Technological University (NTU) on a volunteer basis. Out of these participants, there were twelve females and eight males. Most of the participants had been in Singapore for a long time, with 18 having spent at least majority of the past five years in Singapore. The two exceptions were one who had spent the large part of the past five years in Germany, and another in China. Participants also had to complete the ten-item Ollen Musical Sophistication Index Questionnaire, with the final score being a probability that the participant was a musical expert (Ollen, 2006). In general, participants' mean = 0.16, $SD = 0.09$, suggesting that the majority of the participants had little expertise in music.

Stimuli

The stimuli, in the form of single note recordings, were obtained from four Western and Chinese instruments from two families (stringed: violin, *erhu*, and flutes: flute, *qudi*), from the above-mentioned archive of instrumental sounds. Each instrument provided two notes that were used as stimuli. Each recording was normalised to -3dB in Logic Pro X and exported as an mp3-format sound file (sample rate: 48 kHz, bit rate: 320kbps, stereo) for compatibility with the survey platform. Roughness was computed for each stimulus as in Study 1.

Procedure

Participants clicked an onscreen icon to play each stimulus, and were asked to rate the *dissonance, tension, energy, valence, interest, happiness, sadness* and *liking* of each sound on virtual sliders, before progressing to the next sound. The range of the sliders allowed for responses on a continuous scale from 0 (low) to 100 (high). The 8 stimuli were presented with the questionnaire in an internet-based survey platform (Qualtrics), and the order of the presentation of sounds was randomised for each participant.

Participants were then asked to rate each sound on a bidirectional slider (a continuous semantic scale from -50 to 50) on conceptually cross-modal dimensions of *texture* (rough to smooth), *shape* (spiky to curvy) and *hardness* (hard to soft). In addition, participants were also asked to identify the instruments used in each stimulus, and provide demographic information, such as age, gender, cultural background and musical sophistication. All experiments were conducted in sound-attenuating recording studios, and each studio consisted of one computer terminal and two studio monitors. This study received approval from the Institutional Review Board, Nanyang Technological University. All participants gave informed consent prior to participating in the study.

Data Handling and Analysis

Because of the complication of having both stimulus level effects (computed roughness), and participant level effects (perceived dissonance) in the same model, analysis was conducted using linear mixed-effect modelling, using the “lme4” package (Bates, Maechler, Bolker, & Walker, 2015) in R (R Core Team, 2016). This method allows for independent, crossed random effects to be incorporated in a single, comprehensive model (Baayen, Davidson & Bates, 2008), and to better account for variability through computing a random slope and random intercept for each participant and stimulus. Separate analyses were conducted for each of the rated variables (*tension,*

energy, valence, happiness, sadness, interest, liking, shape, texture and hardness), and dissonance, instrument, ethnicity (Western/Chinese), pitch (high/low) and a log-normal transformation of computed roughness were entered in as fixed-effect factors, and participants and stimuli as random-effect factors.

In selecting the best-fitting model, a maximum likelihood method was used to determine the contributions of the different fixed and random effects variables. This involved the systematic addition of variables to previous models and comparing the difference in fit between the two models. For example, if a model with dissonance is significantly better fitting than a model without dissonance in predicting a specific outcome variable, all other predictors staying constant, it suggests the importance of dissonance as a predictor in that outcome variable. This process was repeated multiple times for each outcome variable until the best-fitting model was selected.

Resultant p values for factors were calculated by the Monte Carlo Markov Chain (MCMC) technique, which is better suited to estimation of p values in more general and complex statistical models, such as linear mixed-effects modelling. Conventional hypothesis testing methods via p value calculation based on t -distributions and f -distributions do not factor in the difficulties in degree of freedom calculations across multiple random effects (participant and item) or the non-independence of random effects in their relation to fixed effect factors (Baayen, Davidson & Bates, 2008). MCMC simulations bypass such problems, by simulating means and variances for each of the factors over a set number of iterations in a stochastic manner; a selective sampling of variables fit into the statistical model in order to calculate posterior distributions according to Bayesian inference. Resultant fixed effect p -values calculated through this method are probabilities that the estimate is greater than zero, and function similar to traditional p -values in a frequentist context. (Andrieu, Freitas, Doucet &

Jordan, 2003; Baayen, Davidson & Bates, 2008; Hadfield, 2010; van Ravenzwaaij, Cassey & Brown, 2016). For this series of analyses, MCMC simulation was also done in the R programming environment, with the “MCMCglmm” package (Hadfield, 2010). Separate analyses were conducted for each outcome variable.

Results

Dissonance was found to be a strong predictor for the constituents of arousal in emotion: *tension* and *energy*, as well as the cross-modal concepts of *shape* (curvy to spiky) and *texture* (smooth to rough). No significant effects of computed roughness were found.

Tension

The best-fitting model for predicting *tension* included dissonance, pitch (high/low) and the interaction between instrument type (flute/strings) and ethnicity (Western/Chinese) as fixed effects, and participant and stimulus as random effects. The final results from the MCMC simulation are presented in Table 3.1 and Figure 3.1.

Effect	Posterior Mean	Confidence Interval		<i>p</i>
		L – 95%	U – 95%	
(Intercept)	26.3	14.8	37.3	<.001***
Dissonance	0.4	0.2	0.6	<.001***
Instrument	-11.2	-20.8	0.1	.035*
Ethnicity	-16.7	-26.2	-6.6	.002**
Pitch	9.7	2.5	16.7	.011*
Instrument:Ethnicity	34.7	20.9	49.9	<.001***

Significance codes: <.001 ‘***’; <.01 ‘**’; <.05 ‘*’

Table 3.1. Full results for fixed effects factors on *tension*.

Figure 3.1. The Pirate Plots⁴ illustrate a clear interaction between instrument and ethnicity. Pitch also predicted tension, with the lower pitch rated overall as higher in tension. Among instruments, participants rated the *dizi* as consistently higher in *tension*.

The scatterplot also highlights a relationship between dissonance and *tension*.

⁴ For an introduction to Pirate Plots, see Phillips (2016).

As shown in Table 3.1, participants' ratings of dissonance were consistent with their ratings of tension, suggesting that stimuli rated as more dissonant were also more likely to be rated as higher in *tension*. Other significant effects showed that stimuli of lower pitch were consistently rated as higher in dissonance, and stimuli played by the Chinese instruments (*erhu* and *dizi*) were similarly rated as higher in dissonance. In all, the model identified significant main effects of dissonance, instrument, ethnicity and pitch, as well as a significant interaction of instrument and ethnicity on *tension*.

Energy

The best-fitting model for predicting *energy* included dissonance and the interaction between instrument type (flute/strings) and ethnicity (Western/Chinese) as fixed effects, and participant and stimulus as random effects. The final results from the MCMC simulation are presented in Table 3.2 and Figure 3.2.

Effect	Estimate	Confidence Interval		<i>p</i>
		L – 95%	U – 95%	
(Intercept)	20.5	11.6	29.3	<.001***
Dissonance	0.4	0.2	0.5	<.001***
Instrument	1.6	-8.7	11.3	.77
Ethnicity	-7.4	-16.9	3.0	.13
Instrument:Ethnicity	14.9	0.92	29.0	.046*

Significance codes: <.001 '***'; <.01 '**'; <.05 '*'

Table 3.2. Full results for fixed effects factors on *energy*.

Figure 3.2. The Pirate Plots illustrate a clear interaction between instrument and ethnicity. Pitch also predicted *energy*, with the lower pitch rated overall as higher in tension. Among instruments, participants rated the *dizi* as consistently higher in *energy*. The scatterplot also highlights a relationship between dissonance and *energy*.

Table 3.2. highlights the consistency between participants' ratings of dissonance and their ratings of *energy*, suggesting that stimuli rated as more dissonant were also more likely to be rated as higher in *energy*. In addition to dissonance, the instrument itself (interaction between the instrument type and ethnicity predictors) had an effect on the perceived *energy* of the stimuli. In all, the model identified a significant main effect of dissonance and a significant interaction of instrument and ethnicity on *energy*.

Shape

The best-fitting model for predicting *shape* included dissonance, pitch (high/low) and the interaction between instrument type (flute/strings) and ethnicity (Western/Chinese) as fixed effects, and participant and stimulus as random effects. The final results from the MCMC simulation are presented in Table 3.3 and Figure 3.3.

Effect	Estimate	Confidence Interval		<i>p</i>
		L – 95%	U – 95%	
(Intercept)	-23.4	-34.8	-13.1	<.001***
Dissonance	0.5	0.2	0.6	<.001***
Dissonance:Instrument	-0.8	-20.8	0.1	.035*
Dissonance:Ethnicity	-11.2	-20.8	0.1	.035*
Dissonance:Instrument:Ethnicity	-11.2	-20.8	0.1	.035*
Instrument	-16.7	-26.2	-6.6	.002**
Ethnicity	-16.7	-26.2	-6.6	.002**
Instrument:Ethnicity	34.7	20.9	49.9	<.001***
Pitch	9.7	2.5	16.7	.011*

Significance codes: <.001 '***'; <.01 '**'; <.05 '*'

Table 3.3. Full results for fixed effects factors on *shape*.

Figure 3.3. The Pirate Plots illustrate a clear interaction between instrument and ethnicity. Pitch also predicted *shape*, with the lower pitch rated overall as lower in spikiness. The scatterplot also highlights a relationship between dissonance and *shape*.

Table 3.3 highlights the consistency between participants' ratings of dissonance and their ratings of *shape*, suggesting that stimuli rated as more dissonant were also more likely to be rated as spikier. In addition to dissonance, the instrument itself (interaction between the instrument type and ethnicity predictors) had an effect on the conceptualised *shape* of the stimuli. In all, the model identified a significant main effect of dissonance, instrument, ethnicity and pitch, as well as a significant interaction of instrument and ethnicity on tension.

Texture

The best-fitting model for predicting *texture* included dissonance only. The final results from the MCMC simulation are presented in Table 3.4 and Figure 3.4.

Effect	Estimate	Confidence Interval		<i>p</i>
		L – 95%	U – 95%	
(Intercept)	15.4	0.9	29.7	.049*
Dissonance	-0.4	-0.7	-0.1	<.001***

Significance codes: <.001 '***'; <.01 '**'; <.05 '*'

Table 3.4. Full results for fixed effects factors on *texture*.

Figure 3.4. The scatterplot highlights a relationship between dissonance and *texture*.

Table 3.4 highlights the consistency between participants' ratings of dissonance and their ratings of *texture*. Sounds that were rated as high in dissonance were also rated as less smooth (more rough).

Discussion

Compared to Study 1 (Chapter 2), *valence* and its associated emotions variables (*happiness* and *sadness*) were not predicted by dissonance or roughness. Also different from study 1 is the relevance of dissonance in predicting liking. However, the role of dissonance in predicting the emotion dimensions of *tension* and *energy*, which together form the 'arousal' component in circumplex models of emotion when applied to music (Eerola & Vuoskoski, 2011), was notably strong. This suggests a consistent relationship between sounds that are perceived as dissonant and arousal evaluations, and it is a relationship that is likely based purely on the spectral properties of instrumental timbre, and not context, short-term memory or tonal effects. However, it is unclear if these effects are due to auditory roughness. It could be that the differences in auditory roughness was too small to be accurately picked up by MacCallum's Roughness patch, and may require other forms of analysis. Alternatively, roughness alone may not be enough to discern roughness in single-tone sounds, and other forms of psychoacoustic analyses, such as loudness or sharpness, may also be important.

The results also suggested a cross-modal link between auditory and visual modalities facilitated by the perception of dissonance, particularly on the shape and texture dimensions. However, because of the semantic nature of the questionnaire, this effect cannot be interpreted as a necessarily cross-modal sensorial link, but instead a cognitive link between the concepts of shape and texture, and dissonance in music.

4. Auditory Roughness and Dissonance in Sounds Generated Through Frequency Modulation (FM) Synthesis

The content of Chapter 4 was part of an oral presentation with proceedings paper:

Liew, K., Styles, S. J., & Lindborg, P. (2017). Dissonance and roughness in cross-modal perception.
Proceedings of the 6th Conference of the Asia Pacific Society for the Cognitive Sciences of Music, Kyoto,
Japan.

Background

Study 2 highlighted the role of dissonance perception in conceptual cross-modal links and emotion recognition. However, auditory roughness did not vary enough to be notably linked to either emotion or cross-modal perception. In Study 3, auditory stimuli were electronically synthesized to systematically vary according to predetermined roughness levels, in order to determine a link between auditory roughness and dissonance with cross-modal perception and emotion recognition. Concurrently, unlike the abstract metaphors used in Study 2, cross-modal perception in this study was tested visually through 3D rendered graphics that varied on dimensions of shape and texture.

Methods

Participants

Participants ($n = 41$, mean age = 23.2, $SD = 4.6$) were recruited from the psychology participant pool at the School of Social Sciences, Nanyang Technological University (NTU) on a volunteer basis. Out of these participants, there were 24 females and 17 males. 30 of them reported growing up in Singapore, and the remaining 11 were various international students. 35 participants had also spent the majority of the last 5 years in Singapore. Participants could not be considered experts in music, scoring low on the Ollen Musical Sophistication Index (OMSI) with mean = 0.21, $SD = 0.26$.

Stimuli

Audio stimuli were generated using frequency modulation (FM) synthesis on Max. A FM synthesis patch was designed and embedded within MacCallum's (2006) real-time roughness analysis patch. Carrier frequencies (base frequencies) were set at 2 predetermined pitches: A1 (110 Hz) and A2 (220Hz). At each specified pitch, 4 levels of roughness were set according to MacCallum's (2006) roughness estimations: high

roughness (1.0), mid roughness (0.5), low roughness (0.01) and no roughness (0). The consistent variation in roughness levels was achieved through the adjustment of the modulation index for each level, which varied the intensity of harmonic bands across the frequency spectrum (Hass, 2001). The harmonicity of the FM synthesis was kept consistent at 6.7, which meant that the harmonic series of the synthesised sound did not follow typical frequency ratios. Amplitude was kept constant for all synthesised sounds. Each stimulus was then exported through the `sfrecord~` object in Max and converted to an mp3-format sound file (bit rate: 320kbps) through iTunes for compatibility with the survey platform.

Study 3 also made use of visual stimuli. Eight pairs of computer-generated graphics were randomly selected from a pre-validated library by Styles (2016). Each pair consisted of two separate images, one depicting a spiky and rough circular object, and the other a smooth and curvy circular object. Each pair was randomly yoked to a specific auditory stimulus.

Procedure

Participants clicked an onscreen icon to play each stimulus, and were asked to rate the *tension*, *energy*, *valence*, *interest*, *happiness*, *sadness*, and *roughness* of each sound on virtual sliders, before progressing to the next sound. The range of the sliders allowed for responses on a continuous scale from 0 (low) to 100 (high). The 8 stimuli were presented with the questionnaire in an internet-based survey platform (Qualtrics), and the order of the presentation of sounds was randomized for each participant. Participants were then asked to rate each sound on a bidirectional slider (a continuous scale from -50 to 50) on *liking* (how much do you like this sound) and *valence* (how positive or negative do you think this sound is). They then undertook a two-alternative forced choice (2AFC) task, matching an image from the pair of graphics to the auditory

stimulus. Participants were also asked to provide demographic information, such as age, gender, cultural background and musical sophistication. All experiments were conducted in the university's psychology laboratories with a computer and Sennheiser HD280 headphones from the School of Art, Design, and Media. An additional rating of dissonance was obtained for each stimulus by having independent expert raters ($N = 10$ after one exclusion due to poor quality of listening device, median OMSI = 0.83, $SD = 0.41$) rate stimuli for perceived dissonance levels. This study was approved by the Institutional Review Board, Nanyang Technological University. All participants gave informed consent prior to participating in the study.

Results

Similar to Study 2 (Chapter 3), the experiment design incorporated both stimulus level effects (computed roughness), and participant level effects (dissonance) in the same model. As such linear mixed-effect modelling (LME), using the “lme4” package (Bates, Maechler, Bolker, & Walker, 2015) in R (R Core Team, 2016) with its maximum likelihood comparisons was used to determine the best-fitting model. In addition to median dissonance ratings, pitch (high/low) and computed roughness level (high, medium, low and no roughness) were entered as fixed effect factors. Resultant p values for fixed factors were estimated by Monte Carlo Markov Chain (MCMC) simulations with 50000 iterations via the “MCMCglmm” package (Hadfield, 2010), with participant and stimuli additionally entered as random effects factors. Separate analyses were run for each outcome variable. Note that for all outcome variables, pitch and its associated

interactions did not significantly influence model fit during likelihood tests. The significance of results was evaluated at the pre-set level of $\alpha = 0.05$.⁵

Tension

The LME selected a model with perceived dissonance and roughness as the best fitting model. The MCMC simulation revealed a significant effect of dissonance ($p = .03$), and computed roughness ($p = .05$).

Effect	Estimate	Confidence Interval		<i>p</i>
		L – 95%	U – 95%	
(Intercept)	11.7	1.5	22.7	0.04*
Dissonance	0.5	0.1	0.9	0.03*
Computed Roughness	8.9	-0.4	17.1	0.05*

Table 4.1. List of predicting variables for *tension*

Energy

The LME selected a model with perceived dissonance and roughness as the best fitting model. The MCMC simulation revealed a significant effect of dissonance ($p < .001$), and computed roughness was not significant ($p = .07$).

⁵ “p = probability of obtaining a test statistic result at least as extreme as the one that was actually observed, under the assumption that the null hypothesis (no effect) is true”.

(Lindborg, 2016)

Effect	Estimate	Confidence Interval		<i>p</i>
		L – 95%	U – 95%	
(Intercept)	14.5	6.3	22.3	0.001**
Dissonance	0.4	0.2	0.6	<0.001***
Computed Roughness	3.8	-0.3	7.9	0.07

Table 4.2. List of predicting variables for *energy*

Valence

The LME selected a model with perceived dissonance and roughness as the best fitting model. While the MCMC simulation revealed a significant effect of computed roughness ($p = .003$), dissonance was not significant ($p = .2$).

Effect	Estimate	Confidence Interval		<i>p</i>
		L – 95%	U – 95%	
(Intercept)	-2.4	-7.5	2.8	0.4
Dissonance	-0.1	-0.3	0.07	0.2
Computed Roughness	-6.3	-9.7	-2.8	0.003**

Table 4.3. List of predicting variables for *valence*

Interest

While the LME model comparison selected a model with perceived dissonance only as the best fitting model, the MCMC simulation did not reveal any relationship between dissonance and *interest*.

Happiness

A likelihood comparison revealed that the stimuli as a random effect did not influence participant ratings of happiness. The LME selected a model with perceived dissonance

as the only factor in the best fitting model, as confirmed by the subsequent MCMC simulation ($p = .002$).

Effect	Estimate	Confidence Interval		p
		L – 95%	U – 95%	
(Intercept)	14.6	10.4	18.8	<0.001***
Dissonance	-0.1	-0.2	-0.04	0.002**

Table 4.4. List of predicting variables for *happiness*

Sadness

As with happiness, the likelihood comparison revealed that the stimuli as a random effect did not influence participant ratings of sadness. The LME selected a model with perceived dissonance as the only factor in the best fitting model, as confirmed by the subsequent MCMC simulation ($p < .001$).

Effect	Estimate	Confidence Interval		p
		L – 95%	U – 95%	
(Intercept)	41.4	34.3	47.8	<0.001***
Dissonance	-0.4	-0.6	-0.2	<0.001***

Table 4.5. List of predicting variables for *sadness*

Roughness

Participants were also asked to provide ratings corresponding to their concept of roughness for each auditory stimulus. As expected, the LME selected a model with perceived dissonance and computed roughness as the best fitting model. The MCMC simulation revealed a significant effect of both computed roughness ($p = .008$) and dissonance ($p = .004$).

Effect	Estimate	Confidence Interval		<i>p</i>
		L – 95%	U – 95%	
(Intercept)	4.5	-2.8	11.2	0.2
Dissonance	0.3	0.1	0.6	0.004**
Computed Roughness	6.7	2.0	11.5	0.007**

Table 4.6. List of predicting variables for *roughness*

Liking

The LME selected a model with perceived dissonance and roughness as the best fitting model. The MCMC simulation revealed a significant effect of dissonance ($p = .006$), and computed roughness ($p = .002$).

Effect	Estimate	Confidence Interval		<i>p</i>
		L – 95%	U – 95%	
(Intercept)	6.3	1.0	11.6	0.02*
Dissonance	-0.3	-0.4	-0.1	0.006**
Computed Roughness	-8.7	-12.5	-5.5	0.003**

Table 4.7. List of predicting variables for *liking*

Cross-modal Shape (Virus choice)

The Generalised (logistic) LME selected a model with perceived dissonance and roughness as the best fitting model. The MCMC simulation (family = categorical) revealed a significant effect of computed roughness ($p < .001$), and computed roughness was marginally significant ($p = .049$).

Effect	Estimate	Confidence Interval		<i>p</i>
		L – 95%	U – 95%	
(Intercept)	-60.9	-110.1	-3.8	<0.001***
Dissonance	0.5	-0.04	1.5	0.049*
Computed Roughness	25.6	1.4	50.3	<0.001***

Table 4.8. List of predicting variables for virus choice

Discussion

Auditory roughness, dissonance and liking

Given that Study 1 replicated earlier studies (Plomp & Levelt, 1965) by showing a strong correlation between computed roughness and perceived roughness and dissonance, it was unexpected that these results were not found in Study 2: computed roughness did not account for any variation in perceived roughness. This was attributed to the low amount of variance in computed roughness levels in the stimuli. In Study 3, with the change in computed roughness now organised into distinct levels of roughness, the effect of computed roughness and dissonance becomes clearer in perceived roughness. Furthermore, the effects of computed roughness and dissonance are seen even in liking, which was not significant in the previous two studies, but are consistent with past studies on dissonance and preference (Trainor & Heinmiller, 1998; Zentner & Kagan, 1998). These suggest that perhaps a large difference in auditory roughness is needed before affective and cognitive changes can be noticed, even in musically inexperienced populations. This could account for certain recent studies (McLachlan, Marco, Light, & Wilson, 2013) not finding any stable relationship between perception of dissonance and auditory roughness, commonly attributed to differences in musical experiences.

Emotion Dimensions, Happiness and Sadness

Dissonance, computed roughness, or a combination of both, predicted *valence*, *tension* and *energy*. Consistent with Study 2, ‘arousal’ (*tension* and *energy*) was robustly predicted by dissonance. Computed roughness, which predicted *tension*, *energy* and *valence* in Study 1, also predicted *valence* in this study. Both computed roughness and dissonance were predictors for *tension*. *Interest* was not significantly predicted by both computed roughness and dissonance. These findings suggest that while arousal and valence can be closely related to the concept of dissonance and auditory roughness, perceptual differences exist in the way dissonance and auditory roughness are processed, consequently affecting the mechanisms behind emotion recognition. Arousal appears to be linked to a combination of both computed roughness and dissonance. *Valence*, itself closely related to the concept of *liking* (Witvliet & Vrana, 2008), appears to be linked primarily to auditory roughness instead of dissonance perception. This suggests that if the variation is large, auditory roughness, as a low-level acoustic feature⁶, may unconsciously affect judgement of positivity and negativity in auditory stimuli. Conversely, perceptual dissonance, which is arguably a higher level construct due to its encompassment of top-down cognitive functions, did not affect *valence* in both Studies 2 and 3, and appears not to have a role in positive or negative assessment of auditory stimuli.

A possible source of explanation for this pattern of results can be found in research on approach – avoidance motivations, in that stimuli with positive valence generally motivate approach behaviour, and stimuli with negative valence generally motivate avoidant behaviour, in an often-unconscious process (Elliot & Covington,

⁶ For more information, see Chapter 1, p. 17.

2001; Cunningham, Raye, & Johnson, 2004). This is consistent with the results in that only auditory roughness had an effect on *valence* and *liking*, in that higher roughness corresponded to lower pleasantness and negative valence. Conversely, dissonance, a higher level cognitive function, predicted *happiness* and *sadness*: emotions that are related to *valence* but much more complicated in their construal. For example, *happiness* can also be dependent on appraisals in clarity, difficulty, and locus of control⁷, in addition to pleasantness (positive valence) (Tong, 2015). Furthermore, *happiness* may not always be construed as positive and *sadness* not always as negative, depending on culture (Uchida & Kitayama, 2009). As such, it makes sense that dissonance, a higher level function, is related to perceptions of *happiness* and *sadness*, which as emotions are complex, higher level functions, since both require subjective evaluations. Contrarily, auditory roughness, a low level percept, is related to perceptions of general positivity and negativity, which is similarly unconscious judgements tied to low-level approach – avoidance motivations. As a cause of annoyance (McDermott, 2012), it makes sense that sounds with high auditory roughness are best avoided, but more research is needed to fully understand the negative impact of roughness before a conclusive link with avoidance motivation can be drawn. Thus, *Interest*, which is a higher-order emotion based on epistemological and cognitive responses to stimuli and does not depend strongly on valence, is not easily influenced by the same low-level approach – avoidance motivations, and thus not robustly predicted by roughness or dissonance.

⁷ Locus of control refers to an individual's belief in whether events in one's life are within one's control (internal) or outside one's control (external) (Rotter, 1966).

Cross-Modal Perception

Auditory roughness, when accentuated, display strong effects on visual shape and texture preference, going beyond the conceptual understanding of shape and texture as examined in Study 2. However, the marginal significance of perceived dissonance as a factor raises additional issues.

Compared to auditory roughness - a bottom-up, low level process - dissonance, often involves the processing of contextual information and experience as a higher level cognitive process. Cross-modal perception here could thus be a result of higher level cognitive links, such as semantic and emotional associations. As such, it could have less to do with the unconscious correspondences between sensory modalities, as compared to auditory roughness. This adds weight to the argument that audio-visual correspondence in cross-modal processing for noise with visual shape and texture, is a similarly low-level, possibly unconscious process.

Furthermore, when considering visual shape perception, spikiness and curviness have long been associated with approach - avoidance motivations. Objects with curved contours are generally preferred over spiky, angular objects, and implicitly associated with pleasantness, safety, positive concepts, as well as approach tendencies. In contrast, spiky, angular objects are implicitly associated with danger, threat and other negative concepts (Bar & Neta, 2006; Palumbo, Ruta, & Bertamini, 2015; Gómez-Puerto, Munar, & Nadal, 2016; Bertamini, Palumbo, Gheroghes, & Galatsidas, 2016).

In the context of earlier discussions on auditory roughness, the unconscious approach – avoidance motivations might offer an additional explanation for the underlying roots of cross-modal perception, particularly in situations where cognitive processing is unconscious and automatic. This is conceptually opposed to higher level mechanisms in cross-modal perception, including, for example, emotion mediation of

cross-modal perception, or more abstract semantic linkages (e.g. this shape reminds me of a bomb, and bombs explode, therefore, this shape sounds like an explosion).

As approach – avoidance motivations have been argued to have survival and adaptive significance (Zajonc, 1998; Elliot & Covington, 2001; Rutherford & Lindell, 2011), related cross-modal mappings may perhaps be more automatic than others that do not directly involve the approach – avoidance motivation. If so, the high degree of transference between auditory roughness and spiky-curvy shapes, and the related lower level valence judgement, may all stem from this primal adaptive mechanism to efficiently decide if an object is safe to approach or better avoided.

Limitations

Firstly, while amplitude was controlled in sound synthesis, actual perceived loudness may have varied systematically between stimuli as a result of equal loudness curves. However, due to the narrow range of stimuli pitch in this experiment (one octave), the variation in perceived loudness is expected to be negligible, not affecting the direction of results in this experiment.

Additionally, the extreme difference in auditory roughness values in Study 3 should also be noted. Here, the stimuli used consisted partially of sustained high-roughness sounds which rarely occur in music, and are more characteristic of sounds that have unexpected harmonic formants, such as sustained cymbals and gongs. Commonly occurring sounds in music often have lower auditory roughness levels. Furthermore, there remains the possibility that sounds with high auditory roughness may be masked by other sounds when not attended to, thus nullifying the effect that auditory roughness has on shape and texture perception. As such, more research is

needed to determine if auditory roughness is still important in cross-modal perception in a setting with higher ecological validity.

**5: *Rama*: An Interactive Musical Interface and Multimedia Performance
Environment**

Background to Rama

In order to examine the robustness of the previous findings pertaining to cross-modal perception in a more ecologically valid situation, *Rama*, an interactive multimedia performance piece, was created with the intent of collecting real-time data from audience members. To this end, the different experimental paradigms were embedded within the time-based structure of the performance, and audience members provided real-time feedback through a web-based interface on their mobile phones. This mobile interface enabled access to data collection, which consequently affected the generated visuals of the performance in real time. *Rama* was an entirely artistic endeavour, with predetermined experimental paradigms subsumed within the structure of the work. In this chapter, I will detail the procedures and considerations in the programming of the piece *Rama*.

Rama

Three key elements of interest were taken into consideration when building *Rama*. Firstly, the differing levels of auditory roughness (from Study 3) had to be played and controlled. Secondly, the resultant audience experience should still be classifiable as an aesthetic experience, and not a laboratory experiment, such that the change in auditory roughness levels should not be a distinct feature of the piece. Lastly, data collection from cross-modal perception as the dependent variable of the experiment had to be possible.

In order to control visual and sonic information simultaneously, Max was used as the main programming language, and NexusUI, a JavaScript toolkit running on HTML5 was used to program the mobile interface for data collection.

Figure 5.1. A screenshot of the Max performance interface for *Rama*.

Auditory roughness was controlled through embedding the frequency modulation synthesis patch used in Study 3 within the main Rama framework. These controlled roughness sounds formed the experimentally levels which similarly differed on roughness (Modulation Index) and pitch (Carrier Frequency). MacCallum's (2006) Roughness patch was also incorporated to provide quick visual confirmation of roughness levels during playback.

Figure 5.2. Roughness controls within the main patch, and roughness selection subpatch.

The aesthetic experience was essential in the ecological validity of the experiments. As such, the roughness conditions were hidden in the backdrop of other aesthetic and musical devices. A four-channel input with live processing (multichannel delay, sample and hold) was used to enable live input from performing musicians on live instruments (not fixed) into the electronic soundscape of the piece, with the *Max urn* object allowing for randomization of the roughness conditions in order to reduce bias from deciding the order. A buffer was also created for each channel to allow spontaneous recordings and reduced or inverted playback to create several drone-like sounds. This also allowed for the control of the frequency range of the overall sonic output, allowing for all four levels of pitches in the controlled roughness sounds to be partially masked regardless of instrument selection. For example, if the selection of instruments were limited to high-pitch instruments like violins and flutes, this processing allows for low-pitch sounds to be generated from these instruments, thereby masking lower-pitched controlled roughness sounds and preventing them from standing out too much.

In order to examine audiences' visual associations to an auditory source, a 3D rendering of a circular object was generated in Max through the *jit.gl.gridshape* object, which would change its shape on a continuum between spiky or curvy depending on audience members' feedback through the mobile interface in real time. If they determined a sound went best with a spiky shape, they could increase the size of the spikes on the object (spikiness), and if they determined a sound to be 'curvy', they could similarly increase the curvature of the object (curviness).

Figure 5.3. (from left) default object, object at maximum curviness, object at maximum spikiness

This was achieved through the creation of several conical and spherical structures embedded within the base circular object, and mapping the variation of the size and stretch of each structure to a single function. This created the illusion of spikes or spheres gradually emerging from the circular object as audience members' gradually increased their rating of curvy or spiky. Through this, the resultant shape of the object can be traced to specific points in time in the presentation of the piece, allowing for a proper quantification of an averaged shape according to audience decisions with each roughness condition.

Graphical Interface for Audience Feedback

Audio-visual artistic works often involve a direct modulation of visual object properties to sonic properties, or vice versa, at a programming or conceptual level (e.g. Ryoji Ikeda, Authecre, Patrick Hartono). However, for *Rama*, this modulation of sonic to visual properties occurs through 'perceptualisation' (Grinstein & Smith, 1990) as a medium; the mapping of visual object to sonic properties is achieved through a real-time visualisation of audience responses to the musical output.

To incorporate the responses from audience/participants into the shape of the onscreen graphic, a simple interface was needed to link audience involvement with the

larger Rama Max patch. Ultimately, an interactive web interface accessible by smartphones was created on HTML5 with elements from Nexus UI (Taylor, Allison, Oh, & Piquemal, 2011-2014) to encode audience responses as OSC information, which could then be imported in real time from the *udpreceive* object in Max. Participants would need to be on the same wireless network as the host device (concurrently running the main Rama patch), in order for each participant's feedback to be downloaded conveniently and automatically.

A Quick-Response (QR) code with the link embedded enabled quick access to the site. Participants first had to consent to participating in the study before being redirected to the interface. The interface allowed for 2 main responses: “More Curvy” or “More Spiky”. Each click on the “More Curvy”/ “More Spiky” button would increase the size of the desired attribute (spikes or spheres) in the 3D object by 2% of the maximum spiky/sphere size (1 point out of a bidirectional 100 point scale). If the object were already spiky or curvy, each click of the button in the opposite direction would shrink the size of the existing spikes or spheres by a corresponding 2%. The differing choices made by each individual member should result in an eventual consensus in object shape, with sounds in the higher roughness conditions having a higher agreement in spikiness, and hence a spikier shape, allowing for a statistical analysis of shape and roughness level. More information on the methods and analysis will be given in the following chapter.

Figure 5.4. The buttons used in the audience interface

Performance Premiere

Through the support of the Singapore-based New Music Ensemble, the South-eastern Ensemble for Today's and Tomorrow's Sounds (SETTS), Rama was programmed in one of their regular concerts of new music at the Visual Arts Centre, Singapore in March, 2017. This concert featured electroacoustic music by Singapore-based composers, and Rama was scheduled to be both the opening and the closing item. For this performance, the ensemble comprised of four musicians: a pianist, bassoonist, flutist and violist.

6: Cross-Modal Shape Perception in a Contemporary Music Concert Setting
(Study 4)

Background

Study 3 found a link between cross-modal dimensions of visual shape/texture with auditory roughness and perceived dissonance. However, due to the stronger, lower level relationship between auditory roughness and visual shape, only auditory roughness was tested in Study 4. It consisted of two experiments. The first was an examination of these effects in the context of an electroacoustic music concert, thereby prioritising ecological validity over laboratory control. The second was a laboratory replication, where participants were played an audio recording of the concert. The purpose of Study 4 was to investigate whether the cross-modal sensory linkages discovered in Studies 2 and 3 would also appear in situations of higher ecological validity where control over auditory roughness was subtler. In a live concert setting, elements of noise were frequently woven into the musical experience. Furthermore, going beyond the traditional two-alternative forced choice approach, I developed a novel response method for cross-modal shape-to-sound mapping by having participants control the exact extent of the spikiness for each sound.

Methods

Participants

In the first experiment, audience members at a live concert event were asked to participate. The concert was held at the Visual Arts Centre in Singapore in March 2017. The number of participants was estimated to be 16 but for technical reasons it could not be exactly determined. No demographic information was collected. The second experiment was conducted in a laboratory setting. Participants (N = 26) were recruited from the undergraduate research participant pool at the Division of Psychology at Nanyang Technological University.

Stimuli

The interface for sonic expression, *Rama*, was designed in the Max programming environment, with additional interfaces designed in HTML 5 (see Chapter 5). This allowed for improvising musicians to build a spectrally full sound through live signal processing and complex delay algorithms, over a series of 16 sequentially presented 20-second electronic tones in a random order. The roughness in these stimuli tones was made to vary systematically (in a similar manner to that in Study 3). A two-dimensional graphical interface was also used, where participants could control the shape of the graphic (spiky or curvy) through a HTML 5 webpage hosted through a local area network. A detailed description of the programming behind this interface can be found in Chapter 5. For experiment two, participants listened to a recording of the performance in experiment one (recorded on a Zoom H6n device in stereo .WAV format, sample rate: 48kbps, bit depth: 24).

Procedure

In the first experiment, audience members were briefed the procedure for data collection through a talk and a demonstration at the start of the concert. Then four musicians (piano, flute, viola and bassoon) from the SETTS Ensemble, a Singaporean professional new music ensemble, improvised according to a list of set parameters. Audience members were invited to collectively reshape the computer graphic in a manner fitting to the music. Based on live feedback from each participant via a LAN webpage connected to the main *Rama* interface through Open Sound Control (OSC) data, an object with varying shape (curvy to spiky) was displayed on a monitor. The ultimate shape was determined by a subtraction of the number of curvy votes from the number of spiky votes, from a minimum of -50 (curvy) to a maximum of 50 (spiky).

The music itself was structured into 16 different segments, each 20 seconds long. The segments differed in the background electronic tone that was held in a drone-like manner throughout the length of the segment. The order of the presentation of segments was randomized.

Due to a technical error, the initial data collected by the experiment was unusable, and an immediate repeat was conducted in the exact same manner at the end of the concert. The data presented in this study is from this repeat. Due to problems with the real-time audio rendering, out of the original 16 stimuli generated, only 10 of these were played, with the initial 6 muted. Fortunately, the random order of presentation, coupled with the systematic variation in roughness, provided enough control for the experiment to still be meaningful.

In the laboratory experiment, participants dragged a bidirectional slider on a Max interface, allowing individual participants to shape the object (on a continuum from spiky to curvy), as they desired. Data was collected based on the position of the slider, on a continuous scale from 0 to 1. Note that in experiment one, the resultant shape was based on a collection of audience responses, but in experiment two, participants designed the shapes independently.

Figure 6.1: A screenshot of graphic shape corresponding to different slider points. The top row displays a spiky graphic (top right) corresponding to the slider position demarcated by a white line (top left), and the bottom row displays a curvy graphic (bottom right) corresponding to slider position (bottom left).

Data Handling and Analysis

Participant response data were transformed to a continuous scale from 0 to 1 for the analysis. Time and final shape information (outcome variable) were captured and stored every second, allowing for convenient mapping to auditory roughness condition in encoding at a rate of 20 data points for 20 seconds of audio for each auditory roughness condition.

The first 5 data points were discarded for each condition in both experiments, allowing time for participants to adjust to the onset of each sequence. The resultant shape information was then recorded as an average of participant feedback on a continuous scale of 0 to 1, with 1 being extremely spiky, and 0 being extremely curvy.

For data analysis, the 10 stimuli were divided into two groups, according to ‘lower’ and ‘higher’ roughness. The cut off was set at roughness = 0.5 according to McCallum’s (2006) model: The Lower Roughness group consisted of the no roughness and low roughness stimuli, while the Higher Roughness group consisted of the mid roughness and high roughness stimuli. This was to reduce the unequal weightage of each level brought on by the erroneous 6 segments worth of roughness levels⁸. As the analysis required a comparison of stimuli, the sample size ($N = 10$) of stimuli was too small for the application of parametric analyses. As such, a non-parametric test was used to compare the group medians. Two separate Mann-Whitney U tests were conducted for experiment one (concert) and experiment two (laboratory) respectively.

Results

For experiment one, the Mann-Whitney U test revealed that the medians for the Lower Roughness group ($N = 4$) and Higher Roughness group ($N = 6$) were different ($U = .67$, $p = .048$). Participants created a significantly spikier object for musical segments with auditory stimuli from the Higher Roughness group (mid and high auditory roughness levels).

⁸ See p.65.

Figure 6.2: The relationship between auditory roughness levels and participant determination of shape in experiment one.

For experiment two, the Mann-Whitney U test revealed that the medians for the Lower Roughness group ($N = 4$) and Higher Roughness group ($N = 6$) were equally different ($U = .67, p = .048$). Participants consistently created significantly spikier objects for musical segments with auditory stimuli from the Higher Roughness group (mid and high auditory roughness levels).

Figure 6.3: The relationship between auditory roughness levels and participant determination of shape in Study 4b.

Discussion

Study 4 demonstrated the robustness of cross-modal auditory roughness - shape mapping even in situations beyond laboratory testing. Despite the musical improvisation being the main focus of the performance, participants were nevertheless able to infer shape information for the roughness stimuli in the background. This suggests that even with auditory distractors present, the perception of auditory roughness can influence shape determination, strengthening the case for automaticity in sound-shape mapping for spiky/curvy domains and auditory roughness. As these stimuli were designed to meet the definition of noise as “unpleasant” and “unmusical” sound in this thesis, we can likely generalise the results of this study to the perception of noise in music in the

context of a musical concert. The fact that noise in music is still able to unconsciously influence the conceptualisation of shape and texture, opens up possibilities for future experimentation in other forms of unconscious influences that noise may convey in this context.

7: Conclusion and Future Directions

In this series of studies, a clear, consistent effect of dissonance and auditory roughness on the recognition of emotion arousal (tension and energy) is demonstrated. However, a greater distinction between auditory roughness and dissonance was demonstrated only in Study 3, where auditory roughness varied sufficiently for a detailed analysis to be carried out. Dissonance was also found to predict arousal, and certain higher level emotions such as happiness and sadness, while auditory roughness predicted lower level emotional judgements, such as arousal and valence. Even then, most of the effects displayed similar patterns of results as Study 1 and 2, affirming past speculations that auditory roughness and dissonance are closely related concepts even when devoid of other musical information (such as melody and harmony).

For cross-modal perception, dissonance and roughness consistently predicted texture and shape, both in situations where cross-modal transference was conceptual, and when participants had control over the precise extent of cross-modal transference for shape in visual objects. The automaticity of cross-modal processing implied in Study 4 further enhances the case for unconscious, fast cross-modal mapping between auditory roughness and shape perception. As such, with regard to the original hypothesis, this paper has demonstrated a clear connection between dissonance and auditory roughness with emotion recognition and cross-modal perception.

Study	Description	Stimuli	Predicting Variables	Significant Results	Main Limitations
1	Re-analysis of existing study (Lahdelma & Eerola, 2016a) with an additional variable of computed roughness.	Synthesised musical chords (Major/Minor, piano/strings).	Auditory roughness, chord quality, instrument.	Auditory roughness: <i>happiness, melancholy, interest, tension, energy, valence.</i>	Musical chords do not sufficiently isolate auditory roughness as a variable. Synthesised sounds for commercial use may be inherently low in roughness.
2	Laboratory study with single tone instrumental sounds.	Recorded instrumental tones: flute, violin, <i>erhu</i> and <i>qudi</i> .	Auditory roughness, perceived dissonance, instrument type, pitch, ethnicity.	Perceived dissonance: <i>tension, energy, shape, texture.</i>	Auditory roughness did not differ enough between stimuli for an adequate comparison.
3	Laboratory study with single tone synthesised sounds.	Sounds generated through FM synthesis, differed systematically on roughness.	Auditory roughness, perceived dissonance, pitch.	Auditory roughness: <i>tension, valence, liking, roughness, shape choice.</i> Perceived dissonance: <i>tension, energy, liking, happiness, sadness, roughness, shape choice.</i>	Low ecological validity, as high auditory roughness stimuli had computed roughness values that were rarely heard in everyday life.
4	Environmental study at a concert with single tone synthesised sounds in the background of a performance item, and laboratory replication.	Sounds generated through FM synthesis (same as Study 3), presented in <i>Rama</i> , an interactive audience interface and multimedia performance environment, at a music concert.	Auditory roughness.	Auditory roughness: <i>cross-modal shape design.</i>	High experimental noise, difficulty in real-time data collection.

Table 7.1. Summary table of results and findings.

An Evaluation of Rama: Environment versus Laboratory

Rama was introduced as an interface for a multimedia performance environment that allowed collection of audience response data. To have ecological validity, aesthetic experience was prioritised over laboratory control, and this carried several inevitable implications. Firstly, to ensure that the experimental elements did not ‘pop out’ from the general performance context, the combinations of noise and improvisation were not as discretely controlled as would be possible in a non-performance context - indeed, as the musicians were improvising, the nature of their performance may have differed in the context of different kinds of noise, as the performers incorporated the noise into their improvised performance. These interactions between experimentally controlled ‘noise’ and live improvised musical performance may have created self-reinforcing feedback between the instrumental and synthesised elements of the composition. Secondly, due to the uncontrolled nature of the live concert scenario, not all audience members elected to take part in the interactive element of the performance. This may have been due to the size of the display (a large digital monitor), which did not deliver a fully-immersive visual experience, or to individual differences in audience interactivity preferences (only a small proportion of audience members chose to participate). Finally, *Rama* was performed in a somewhat niche electro-acoustic concert, meaning the audience was somewhat unusual in their level of musical knowledge and engagement. This ‘expert’ audience may show a different pattern of audio-visual mapping preferences, when compared to a less-musically experienced audience. Nevertheless, the experimental results were robust, when compared to the subsequent laboratory replication, and successfully replicated the results of Studies 2 and 3. As such, it was able to quantify participants’ responses in a highly unusual experimental setting, generalising the cross-modal effects of auditory roughness (noise) back to the original research question on the

perceptual effects of noise in music. As a sizeable portion of electronic music today revolves around the aesthetic of noise, the findings demonstrate that there are subtle effects of noise that go beyond an auditory listening experience to influence processes in other sensory modalities. The fact that noise in music, through auditory roughness, is able to unconsciously influence the conceptualisation of shape and emotion, opens up possibilities for future experimentation in other forms of influence that noise may convey. In the context of art and music, such findings can help composers and artists deliver a stronger artistic message through pieces that are grounded in a unified perceptual experience.

An interface for audience response like *Rama* also has the potential to bring a new dimension into perception and cognition research in the arts. With the ability to consolidate participant responses outside of the laboratory and into an environment like a concert hall or an art gallery, researchers can develop testing methods suited to the sampling of situations with high ecological validity. Given that a platform like Max allows for easy integration of various plug-ins, future studies could explore other quantification methods, such as motion tracking, that are less intrusive to the aesthetic experience of audience members, and reducing the limitations of performance-based environmental research.

Future Research Directions

Future research can focus on proving a definitive link between auditory roughness and approach – avoidance motivation. This can provide a possible explanation linking all the above concepts (dissonance, auditory roughness, emotion recognition and cross-modal perception) from an evolutionary standpoint: in the need to efficiently evaluate information about the environment around us, we as humans have

adapted by unconsciously assimilating cross-modal information and forming positive or negative judgements in order to decide if something should be approached or avoided for survivability. This could explain why certain sound – shape mappings appear to be consistent between cultures and languages (such as the bouba-kiki effect [Ramachandran & Hubbard, 2001], where spikiness is strongly associated with fricatives that are high in auditory roughness), while others appear to be strongly mediated by emotion or statistical regularities in a specific environment instead. Future studies can thus explore the validity of this theory by expanding the scope of low-level musical and psychoacoustic features studied. For that purpose, it may be necessary to examine other aspects of perception and cognition, such as attention and memory, to properly understand if dissonance, auditory roughness, cross-modal perception and emotion recognition are related through the approach – avoidance motivation.

If proven, this may help answer certain questions commonly raised by arts practitioners and composers as to the accessibility of contemporary music by the general public. The avoidance and innate dislike of noise in music may be due to a primal human mechanism, one that had evolved over time to ensure the survival of our species, which intuitively categorises an object of noise as something to be avoided. However, higher level cognitive processing of emotions like happiness and sadness suggest that while this initial preference might be unconscious, conscious attitudes and a subjective emotional experience of music may override this initial dislike. As such, a person's environment and upbringing may have an even greater role in overcoming that person's innate dislike of noise in music. This could have future implications for policymaking in arts education, in order for today's flourishing and diverse experimental arts scene to be further appreciated by the general public.

These research findings can also be relevant to existing branches of research in psycholinguistics, especially with regard to cross-modal perception in the form of sound symbolism. Sound symbolism refers to the non-arbitrary consistencies between language structure and meaning, which are commonly understood across speakers of a language. In particular, the structures of certain word forms show unconscious bias towards corresponding physical characteristics, in areas such as sound-shape mapping, across different languages and cultures (Kohler, 1947; Ramachandran & Hubbard, 2001; Kovic, Plunkett, & Westermann, 2010; Maurer, Pathman, & Mondloch, 2006; Hung, Styles, & Hsieh, 2017). Research in sound-symbolism, particularly in linguistic sound-shape mapping, has revealed a relationship between formant structure and various visual properties in various verbal and nonverbal vocalisations, focusing in particular, on the relationship between f_0 and f_1 formants in vowel production (Lockwood & Dingemanse, 2015; Morton, 1977). For example, Morton (1977), argued that one aspect of such a relationship, specifically between physical size and voice fundamental frequency (f_0), can be ascribed to a lower voice implying larger vocal folds, and correspondingly, a larger organism. Ohala, Hinton and Nichols (1997) build on this theory by arguing that in addition to the f_0 , because of the relationship between the size of the vocal tract and the overall size of an individual, the higher vocal tract frequencies are also important in sound symbolism. While these studies provide a biological basis for understanding cross-modal perception, they do not go any deeper into the formant structure of these sounds. For this purpose, auditory roughness, inherently an estimate of formant structure based on distances between prominent pairs of formants for any give sound, may offer further insight into the internal mechanisms of sound symbolism, thus providing an alternative means of quantifying formant relationships in vocal tract frequencies, contributing towards a better understanding of

the innate mechanisms in cross-modal processing. While this thesis has focused on sounds in the context of music, further studies can thus replicate the findings and research methods in the context of verbal sounds, such as vowels and consonants.

Bibliography

- Adorno, T. W. (1970/1984). *Aesthetic Theory*. (G. Adorno, R. Tiedemann, Eds., & C. Lenhardt, Trans.) London: Routledge.
- Avramova, Y. R., & Stapel, D. A. (2008). Mood as spotlights: The influence of mood on accessibility effects. *Journal of Personality and Individual Differences*, 95, 542-554.
- Balkwill, L.-L., Thompson, W. F., & Matsunaga, R. (2004). Recognition of emotion in Japanese, Western, and Hindustani music by Japanese listeners . *Japanese Psychological Research*, 46(4), 337-349.
- Ball, P. (2008). Science & Music: Facing the Music. *Nature*, 453, 160-162.
- Bar, M., & Neta, M. (2006). The proactive brain: using rudimentary information to make predictive judgements. *Journal of Consumer Behavior*, 7, 319-330.
- Baraldi, F. B., De Poli, G., & Roda, A. (2006). Communicating Expressive Intentions with a Single Piano Note. *Journal of New Music Research* , 35(3), 197-210.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effect models using lme4. *Journal of Statistical Software*, 67(1), 1-48.
- Batson, C. D., Shaw, L. L., & Oleson, K. C. (1992). Differentiating affect, mood, and emotion: Toward functionally-based conceptual distinctions. In M. Clark (Ed.), *Review of Personality and Social Psychology* (Vol. 13, pp. 294-326). Newbury Park, CA: Sage.
- Bernardes, G., Davies, M., Guedes, C., & Pennycook, B. (2014). Considering roughness to describe and generate vertical musical structure in content-based algorithmic-assisted audio composition. In A. & Georgaki (Ed.), *Proceedings of the 40th International Computer Music Conference/Sound and Music Computing Conference*. Athens, Greece.

- Bertamini, M., Palumbo, L., Gheroghes, T. N., & Galatsidas, M. (2016). Do observers like curvature or do they dislike angularity? *British Journal of Psychology*, 107, 154-178.
- Bidelman, G. M., & Krishnan, A. (2009). Neural correlates of consonance, dissonance, and the hierarchy of musical pitch in the human brainstem. *Journal of Neuroscience*, 29(42), 13165-13171.
- Bowling, D., & Purves, D. (2015). A biological rationale for musical consonance. *Proceedings of the National Academy of Sciences*, 112(36), 2015.
- Bowman, C., & Yamauchi, T. (2015). Perceiving Categorical Emotion in Sound: The Role of Timbre . *Psychomusicology: Music, Mind and Brain, Advance Online Publication*.
- Campbell, M., & Greated, C. (1987). *The Musician's Guide to Acoustics*. NY: Schirmer.
- Crisinel, A.-S., & Spence, C. (2010). As bitter as a trombone: synesthetic correspondences in nonsynesthetes between tastes/flavors and musical notes. *Attention, Perception & Psychophysics*, 72(7), 1994-2002.
- Crisinet, A.-S., & Spence, C. (2009). Implicit association between basic tastes and pitch . *Neuroscience Letters*, 464(1), 39-42.
- Cunningham, W. A., Raye, C. L., & Johnson, M. K. (2004). Implicit and explicit evaluation: fMRI correlates of valence, emotional intensity and control in the processing of attitudes. *Journal of Cognitive Neuroscience*, 16(10), 1717-1729.
- Eerola, T., & Vuoskoski, J. K. (2011). A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music*, 39(1), 18-49.
- Elliot, A., & Covington, M. (2001). Approach and avoidance motivation. *Educational Psychology Review*, 13(2), 73-92.

- Evans, P., & Schubert, E. (2008). Relationships between expressed and felt emotions in music. *Musicae Scientiae*, 12, 75-99.
- Fant, G. (1960). *Acoustic Theory of Speech Production*. The Hague, Netherlands: Mouton & Co.
- Fastl, H., & Zwicker, E. (2007). *Psychoacoustics: Facts and models*. New York: Springer.
- Förster, J. (2003). The influence of approach and avoidance motor actions on food intake. *European Journal of Social Psychology*, 33, 339-350.
- Fredrickson, B. L. (2009). *Positivity*. New York: Three Rivers Press.
- Frijda, N. H. (1986). *The Emotions*. Cambridge: Cambridge University Press.
- Frijda, N. H. (1993). Moods, emotion episodes and emotions. In M. & Lewis (Ed.), *Handbook of emotions* (pp. 381-403). New York: Guildford Press.
- Gabrielsson, A. (2001-2002). Emotion perceived and emotion felt: Same or different? *Musica Scientiae*, 5(1), 123-147.
- Gabrielsson, A., & Lindström, E. (2010). The role of structure in the musical expression of emotions. In P. N. Juslin, & J. A. Sloboda (Eds.), *Handbook of Music and Emotion: Theory, Research and Applications*. Oxford: Oxford University Press.
- Garrido, S., & Schubert, E. (2011). Negative emotions in music. What is the attraction? A qualitative study. *Empirical Musicology Review*, 6(4), 214-230.
- George, D., & Mallery, M. (2010). *SPSS for Windows Step by Step: A Simple Guide and Reference, 17.0 update*. Boston, MA: Pearson.
- Gómez-Puerto, G., Munar, E., & Nadal, M. (2016). Preference for curvature: A historical and conceptual framework. *Frontiers in Human Neuroscience*, 9(712).
- Grinstein, G. G., & Smith, S. (1990). Perceptualisation of scientific data. *Proceedings of the Society of Photo-optical Instrumentation Engineers, Society for Imaging*

- Science and Technology, Extracting Meaning from Complex Data: Processing, Display, Interaction*, 1259, 190-199.
- Hailstone, J. C., Omar, R., Henley, S. M., Frost, C., Kenward, M. G., & Warren, J. D. (2009). It's not what you play, it's how you play it: Timbre affects perception of emotion in music . *Quarterly Journal of Experimental Psychology*, 62(11), 2141-2155.
- Hevner, K. (1936). Experimental studies of the elements of expression in music. *American Journal of Psychology*, 48(2), 246-248.
- Hindemith, P. (1942). *The craft of musical composition* (Vol. 1). (A. Mendel, Trans.) New York: Associated Music Publishers.
- Hung, S. M., Styles, S. J., & Hsieh, P. J. (2017). Can a word sound like a shape before you have seen it? Sound-shape mapping prior to conscious awareness. *Psychological Science*, 28(3), 263-275.
- Jaramillo, S., & Zador, A. M. (2011). The auditory cortex mediates the perceptual effects of acoustic temporal expectation. *Nature Neuroscience*, 14(2), 246-253.
- Jensen, K. K., & Hjortljaer, J. (2012). An improved dissonance measure based on auditory memory. *AES*, 60(5), 350-354.
- Johnson-Laird, P. N., Kang, O. E., & Leong, Y. (2012). On Musical Dissonance . *Music Perception: An Interdisciplinary Journal* , 30(1), 19-35.
- Juslin, P. N. (2013). From everyday emotions to aesthetic emotions: towards a unified theory of musical emotions. *Physics of Life Reviews*, 10, 235-266.
- Juslin, P. N., & Sloboda, J. A. (2010). The past, present and future of music and emotion research. In P. N. Juslin, & J. A. Sloboda (Eds.), *Handbook of Music and Emotion: Theory, Research, and Applications*. Oxford: Oxford University Press.

- Juslin, P. N., Friberg, A., & Bresin, R. (2002). Toward a computational model of expression in music performance: The GERM model. *Musicae Scientiae, Special Issue*, 63-122.
- Kameoka, A., & Kuriyagawa, M. (1969). Consonance theory, parts I & II. *Journal of the Acoustical Society of America*, 45(6).
- Kantono, K., Hamid, N., Shepard, D., Yoo, M., Carr, B., & Grazoli, G. (2015). The effect of background music on food pleasantness ratings. *Psychology of Music*, 44(5), 1111-1125.
- Kohler, W. (1947). *Gestalt Psychology: An Introduction to New Concepts in Modern Psychology*. New York: Liveright.
- Kovic, V., Plunkett, K., & Westermann, G. (2010). The shape of words in the brain. *Cognition*, 114(1), 19-28.
- Krpan, D., & Schnall, S. (2014). Too close or comfort: Stimulus valence moderates the influence of motivational orientation on distance perception. *Journal of Personality and Social Psychology*, 107(6), 978-993.
- Kuriki, S., Kanda, S., & Hirata, Y. (2006). Effects of musical experience on different components of MEG responses elicited by sequential piano-tones and chords. *The Journal of Neuroscience*, 26(15), 4046-4053.
- Lahdelma, I., & Eerola, T. (2016b). Mild Dissonance Preferred Over Consonance in Single Chord Perception. *i-Perception*, 1-21.
- Lahdelma, I., & Eerola, T. (2016a). Single chords convey distinct emotional qualities to both naive and expert listeners. *Psychology of Music*, 44, 37-54.
- Lamy, D., Salti, M., & Bar-Haim, Y. (2008). Neural correlates of subjective awareness and unconscious processing: An ERP study. *Journal of Cognitive Neuroscience*, 21(7), 1435-1446.

- Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). *International Affective Picture Systems (IAPS): Technical manual and affective ratings*. NIMH Center for the Study of Emotion and Attention. International Affective Picture Systems (IAPS).
- Lerner, J. S., & Keltner, D. (2000). Beyond valence: Toward a model of emotion-specific influences on judgement and choice. *Cognition and Emotion*, 14(4), 473-493.
- Lindborg P. M. (2016). A taxonomy of sound sources in restaurants. *Applied Acoustics*, 110C. Doi: 10.1016/j.apacoust.2016.03.032.
- Lindborg, P. M., & Friberg, A. K. (2015). Colour association with music is mediated by emotion: evidence from an experiment using a CIE Lab interface and interviews. *PLoS One*, 10(12), e0144013.
- Lockwood, G., & Dingemanse, M. (2015). Iconocity in the lab: A review of behavioral, developmental and neuroimaging research into sound-symbolism. *Frontiers in Psychology*, 6(1246).
- MacCallum, J., & Einbond, A. (2008). Real-time analysis of sensory dissonance. In R. Y. Kronland-Martinet (Ed.), *Computer Music Modeling and Retrieval. Sense of Sounds. Proceedings of the 4th International Symposium, CMMR 2007*. Copenhagen, Denmark.
- Maurer, D., Pathman, T., & Mondloch, C. J. (2006). The shape of boubas: sound-shape correspondences in toddlers and adults. *Developmental Science*, 9(3), 316-322.
- McDermott, J. H. (2012). Auditory Preferences and Aesthetics. In R. J. Dolan, & T. Sharot (Eds.), *Neuroscience of Preference and Choice: Cognitive and Neural Mechanisms* (pp. 227-256). London: Academic Press.

- McDermott, J., Lehr, A., & Oxenham, A. (2010). Individual differences reveal the basis of consonance. *Current Biology*, 20(11), 1035-1041.
- McLachlan, N., Marco, D., Light, M., & Wilson, S. (2013). Consonance and pitch. *Journal of Experimental Psychology*, 142(4), 1142-1158.
- Meyer, L. B. (1956). *Emotion and Meaning in Music*. Chicago: Chicago University Press.
- Morton, E. S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *American Naturalist*, 111, 855-869.
- noise. (n.d.). in *Merriam-Webster.com*.
- Ohala, J., Hinton, L., & Nichols, J. (1997). Sound symbolism. *4th Seoul International Conference on Linguistics (SICOL)*, (pp. 98-103).
- Palmer, S. E., Schloss, K. B., Xu, Z., & Prado-León, L. R. (2013). Music–color associations are mediated by emotion. *Proceedings of the National Academy of Sciences*, 110(22), 8836–8841.
- Palumbo, L., Ruta, N., & Bertamini, M. (2015). Comparing angular and curved shapes in terms of implicit associations and approach/avoidance responses. *PLoS One*, 10(10).
- Parncutt, R. (1989). *Harmony: A psychoacoustical approach*. New York: Springer.
- Parncutt, R., & Hair, G. (2011). Consonance and dissonance in music theory and psychology: Disentangling dissonant dichotomies. *Journal of Interdisciplinary Music Studies*, 5(2), 119-166.
- Phillips, N. D. (2016). *Yarrrr! The pirate's guide to R*. <http://thepiratesguidetor.com>.
- Plomp, R., & Levelt, W. J. (1965). Tonal consonance and critical bandwidth. *Journal of the Acoustical Society of America*, 38.

- R Core Team. (2016). R: A language and environment for statistical computing. Vienna, Austria.
- Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia — a window into perception, thought and language. *Journal of Consciousness Studies*, 8(12), 3-34.
- Roseman, I. J. (1991). Appraisal determinants of discrete emotions. *Cognition and Emotion*, 5(3), 161-200.
- Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General and Applied*, 80(1), 1-28.
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161-1178.
- Russolo, L. (1913/1967). *The Art of Noise (Futurist Manifesto)*. (R. Filliou, Trans.) NY: Something Else Press.
- Rutherford, H. J., & Lindell, A. K. (2011). Thriving and surviving: Approach and avoidance motivation and laterization. *Emotion Review*, 3(3), 333-343.
- Scherer, K. R. (2004). What emotions can be induced by music? What are the underlying mechanisms? And how can we measure them? *Journal of New Music Research*, 33(3).
- Schoon, D., Regnault, P., Ystad, S., & Besson, M. (2005). Sensory consonance: An ERP study. *Music Perception: An interdisciplinary Journal*, 32(2), 105-118.
- Sethares, W. A. (1993). Local consonance and the relationship between timbre and scale. *Journal of the Acoustic Society of America*, 94(3), 1218-1228.
- Silvia, P. J. (2008). Interest - the curious emotion. *Current Directions in Psychological Science*, 71(1), 57-60.

- Solomon, R. C., & Stone, L. D. (2002). On “Positive” and “Negative” Emotions. *Journal for the Theory of Social Behaviour*, 32(4), 417-435.
- Spence, C., & Deroy, O. (2013). On why music changes what (we think) we taste. *iPerception*, 4, 137-140.
- Storm, C., & Storm, T. (1987). A taxonomic study of the vocabulary of emotions. *Journal of Personality and Social Psychology*, 53(4), 805-816.
- Taylor, B., Allison, J., Oh, Y., & Piquemal, S. (2011-2014). Nexus UI (Software).
- Tenney, J. (1988). *A History of Consonance and Dissonance*. NY: Excelsior Music.
- Tong, E. (2015). Differentiation of 13 positive emotions by appraisals. *Cognition and Emotion*, 29(3), 484-503.
- Trainor, L. J., & Heinmiller, B. M. (1998). The development of evaluative responses to music: Infants prefer to listen to consonance over dissonance. *Infant Behavior and Development*, 21(1), 77-88.
- Tremblay, K. L., Ross, B., Kayo, I., McClannahan, K., & Collet, G. (2014). Is the auditory evoked P2 response a biomarker of learning? *Frontiers in Systems Neuroscience*, 8.
- Uchida, Y., & Kitayama, S. (2009). Happiness and unhappiness in East and West: Themes and variations. *Emotion*, 9(4), 441-4556.
- Vassilakis, P. N. (2005). Auditory roughness as a means of musical expression. In R. A. Kendall, & R. W. Savage (Eds.), *Selected Reports in Ethnomusicology: Perspectives in Systematic Musicology* (Vol. 12, pp. 119-144). Los Angeles: University of California Los Angeles.
- Vieillard, S., Peretz, I., Gosselin, N., Khalfa, S., Gagnon, L., & Bouchard, B. (2007). Happy, sad, scary and peaceful musical excerpts for research on emotions . *Cognition and Emotion*, 22(4), 720-752.

- Wang, Q., Wang, S., & Spence, C. (2016). "Turn up the taste": Assessing the role of taste intensity and emotion in mediating crossmodal correspondences between basic tastes and pitch. *Chemical Senses*, 41, 345-356.
- Wolfe, J. (1997). *What is a Sound Spectrum?* Retrieved from Music Science at UNSW: <https://newt.phys.unsw.edu.au/jw/sound.spectrum.html>
- Xu, R. (2003). Measuring explained variation in linear mixed effects models. *Statistics in Medicine*, 22, 3527-3541.
- Zajonc, R. (1998). Emotion. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *The Handbook of Social Psychology* (4 ed., pp. 591-632). New York: McGraw-Hill.
- Zentner, M. R., & Kagan, J. (1998). Infants' perception of consonance and dissonance in music. *Infant Behavior & Development*, 21(3), 483-492.
- Zwicker, E. (1961). Subdivision of the audible frequency into critical bands. *Journal of the acoustical society of america*, 33(2), 248-249.