A Multi-Dimensional Analytical Model for Musical Harmony Perception

Ning Cong

PhD

University of York

Electronics

November 2016

Abstract

Rules and conventions observed in western music harmony involve a number of psychophysical relationships between musical entities and perception responses in terms of consonance and dissonance concepts (*CDC*). These well-established relationships have informed a common psychophysical mechanism that can be studied and numerically modelled. In the literature, a number of physiological and psychological based theories have been proposed, but no one single theory is able to fully account for the phenomenon of music harmony perception.

This research deems musical consonance and dissonance to be a multi-dimensional concept that is underpinned by several psychoacoustic principles; it is hypothesized that perceived impression of a musical entity/structure is composed of a number of uncorrelated experiences that can be measured on a multi-dimensional space. The psychoacoustic model proposed here contains four types of actively defined dissonance concepts: namely *sensory*, *ambiguity*, *gloom* and *tension*. *Sensory* dissonance refers to the (primary and secondary) *beats effect* due to the physiological functions of auditory pathway organs; the *ambiguity* dissonance is developed from *harmonic-template based theories* in which sonorities with ambiguous tonal centres are considered dissonant; the *gloom* and *tension* dissonances are two fundamental dimensions of musical emotions that are related to raised and lowered pitch contours described by the *melodic expectation theory*. The correlation / independence between each type of dissonance concept is statistically analysed based on experimental results from newly conducted listening tests.

In the application of this analysis to musical triads, this work shows that: the secondary chord structures (simultaneous chords containing one or two semitone intervals) have the highest level of *sensory* dissonance; the suspended 4th chords have a higher level of *ambiguity* dissonance; diminished chords have the highest level of *gloom* dissonance (followed by minor structures); the augmented chord structures contain the highest level of *tension*; and, of all chord structures, all the dissonance types are at their lowest level in the major chords .

List of Contents

List of Tables

List of Figures

Acknowledgement

I would like to thank to my supervisors, David Howard and Helena Daffern, for their guidance and valuable discussions. I would also like to thank my thesis advisor, Tim Dozer for his efforts in polishing the thesis. Special thanks go to Andy Hunt for his research passion and inspiration.

Lastly, I would like to thank my family members for their love and support.

Declarations

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Chapter 1 Introduction

The theories surrounding western music harmony have depicted a number of subtle relationships between the use of certain musical chords and various types of psychological emotions. For instance, the major chord structures usually bring the sense of 'happy' and 'stable', whilst the minor chords sound more 'sad' and 'unpleasant'. Within a tonal system, musical chords or notes on different scale degrees[1] also evoke different feelings. For example, the first and major fifth scale degree usually sound more stable and grand than the other scale degree notes; the minor third and seventh scale degree are usually perceived to be 'sad'. Interestingly, many of these observations do not seem to vary between individuals and across different cultures, which makes music a 'universal language'.

The motivation of this research is simple but ambitious – to build a numerical model to account for music harmony perception. The main challenges of developing such an analytical model include:

1) Understanding the psychological experiences pertaining to music harmony perceptions. From a numerical model point of view, this defines the output of a psychoacoustic system. However, giving a proper definition to the psychological responses can be difficult: on one hand, the richness and complexity of human emotions makes language descriptors quite crude; and on the other hand, it is even more difficult to measure the emotions directly in a numerical way. In western music conventions, the terms 'consonance' and 'dissonance' have been used in a simple way to denote the perception of music harmony. However, these terms are far from fully communicating the perceptual experiences of music harmony. As a

[1] In a tonal music system, a scale refers to a set of ascending/descending musical notes (or root of chords) in reference to a tonal centre note. The intervallic relationships between each scale note and tonal centre is denoted by the scale degrees.

result, music educators are unable to fully interpret the psychological functions of music composition techniques; and a number of confusions and conflicts have been raised between psycho-musical scholars. The questions are hereby: ‘could we develop a new measuring system whereby the perception of music harmony can be better described than solely using the term consonance and dissonance’. Such a system is not only crucial for a numerical model for music harmony perception, but also provides a parametric way to describe and communicate the music harmony related emotions, akin to how the RGB system describes a specific colour.

2) Understanding the musical entities associated to western music harmony. From analytical point of view, this defines the input of a psychoacoustic system. In western music literature, music harmony perceptions do not correspond to a specific type of music entity; but a number of musical entities have been involved, for instances: the perception of a single isolated chord, a tonal system, melodic contour; or the relationship between a single note and a chord, between two successive musical chords/notes, or between a note/chord and a tonal system; etc. This poses the question, ‘do all these musical entities shares a ‘common law’ in term of harmony perceptions?’ If the answer to this question is negative, then it implies the perception of music harmony cannot be explained with a single, unified theory. Instead, the perception responses have to be analysed with respect to each type of musical entity.

3) Understanding the behaviour of our auditory perception system. Our perception responses can be viewed as a combination of both the lower-level physiological sensations and higher-level psychological principles. In the literature, a number of theories have been proposed in accounting for music harmony perception, and some of the theories even appear to contradict each other. The concern of this thesis is, ‘what are the inter-relationships between these theories?’ and ‘Is it possible that more than one school of theories are correct?’ For example, theories proposed based on the physiology of the inner ear could be relatively independent from the cognitive based theories, as they belong to different stages of auditory perception. They can both be true if they essentially correspond to two different types of musical experiences within the consonance and dissonance concepts. Furthermore, it then may be possible to convert these theories and combinations

of them into numerical models in order to predict the perception of music harmony.

4) Lastly, if there are multiple types of (i) psychological experiences (ii) musical entities and (iii) perception theories associated with music harmony, understanding how they are related to each other is fundamental to develop a more comprehensive theory to account for the phenomenon of harmony perception. For instance, does a particular musical entity correspond directly to a particular psychological experience, and does this psycho-musical relationship correspond directly with a particular explanatory theory? Once we can combine all these inter-relationships into a single picture, we may be able to develop a unified psycho-acoustic model accounting for the perception of music harmony. Such a model can then be used not only as a tool for music composition and analysis, but more importantly, it can push techniques of machine learning from artificial intelligence to the age of 'artificial emotions'. Imagine in the near future, we may able to use computer technology to record, transmit and reproduce our emotions, seize the sincere feelings in our lives, simply via the psycho-acoustic 'code' of music harmony.

1.1 Object and Scope

Like any perception phenomenon, music harmony perception has both innate and cultured aspects (Schönberg and Stein, 1985). The innate aspect is underpinned by the physiological and psychological functions, and its perception traits are commonly observed across different demographic groups; and the cultured aspect is closely related to personal memories/experiences, cultural backgrounds and music aesthetics. In perception and cognition process, the 'natured' aspect can be viewed as the 'lower' level response reflecting common physiological and psychological principles; and the 'nurtured' aspect can be viewed as the 'higher' level process which includes brain functions that interpret concurrent sensations with past memories. The focus of this research is mainly the natured aspect of harmony perception, for the reason that the perception responses of the natured aspect are: (1) **primarily deterministic**: therefore it is more favourable to theoretical tests, analysis and psychophysical models; (2) **relatively simple and consistent**: there have been a number of music theories and empirical tests suggesting common perception principles and results; (3) **more fundamental**: it reveals some lower-level perception mechanisms that serve as the theoretical ground before more complex music emotions can be analysed.

The consonance and dissonance concepts (*CDC*) in this research refer to a number of emotional-domain language descriptors that are categorized under the dichotomy of musical consonance and dissonance. This research aims to study the perceptions of *CDC* that are commonly observed among people regardless of their personal and cultural backgrounds. This includes:

(1) Musical theories pertaining to what type of musical entities have been involved in the discussion of consonance and dissonance. A more detailed review of the music harmony concept is presented in Chapter 2.
(2) Psychological theories including the discussion of the fundamental dimensions that describes the perception of musical consonance and dissonance as well as how the perception of *CDC* can be interpreted by psychological/cognitive based theories.
(3) Physiological theories that attempt to account for the perception of musical consonance and dissonance based on the function of the auditory pathway organs. Both psychological and physiological theories pertaining to *CDC* are reviewed in Chapter 3.

1.2 Methodology

A complete explanation of music harmony also involves the study of functional responses of each organ along the auditory pathway, and most importantly, the functioning of the brain. However, at this stage, neuroscience and associated technologies do not allow perceptual / emotional features to be precisely monitored, analysed, and interpreted. As an alternative, psychophysical theories and models are used to bypass the biological analysis, establishing direct relationships between acoustical domain features and perception impressions.

Typical psychoacoustic study pertaining to music harmony generally includes the following main tasks:

1. To understand musical entities involved under the music harmony concept; define and represent these musical entities with acoustic features.

2. To understand what psychological measurements should be used for music harmony. Psychological experiences are usually described with semantic descriptors; the challenge of a psychoacoustic approach is therefore to quantify psychological responses with numerical measurements.

3. To construct a psycho-acoustical model that takes in the acoustical features and produces a numerical prediction of the psychological measurement. A numerical model should reflect the central theories based on physiological or psychological principles (See the literature review of Chapter 3). In this research, the model design and simulations are implemented within Matlab.

4. To conduct listener experiments to evaluate the free parameters used in the numerical model. Electronic signal processing techniques are typically used, as this enables us to synthesise and modify the sounds that precisely meet our experimental needs.

1.3 Hypothesis

1.3.1 Statement of Hypothesis

Current analytical models for musical consonance and dissonance can be improved by implementing a pitch-based multi-dimensional harmonic analysis.

1.3.2 Analysis of Hypothesis

Under conventional numerical analysis, the perception responses of music harmony are typically measured on a one-dimensional scale using consonance dissonance concepts (*CDC*) – (or equivalent descriptors such as 'pleasant' – 'unpleasant', 'stable' – 'unstable') as its two extremes. This implies the literal meaning of music harmony is associated with a single psychoacoustic criterion. However, in music practice, there are a few types of dissonance experiences observed:

The first type of dissonance is known as *sensory beats*. Sensory beats are perceived as a 'rough, unsettling' sound that is usually observed between pure-tone interactions. Musical chords comprised of 1–3 semitone intervals (depending on the frequency centre) usually contain a higher level of beats. In this research, the term *sensory dissonance* is used to denote this type of dissonance.

The second type of dissonance arises when listeners have difficulties identifying the tonal centre of a specific group of simultaneously sounding sounds. For instance, an added fourth note in major triad structure is able to throw the root perception into confusion: the root of a major triad will be strongly contrasted to the fourth scale degree note, resulting in a particular type of dissonance. In this research, the term *ambiguity dissonance* is used to denote this type of dissonance.

The third type of dissonance is usually observed in augmented intervals or chords, such as the augmented fifth interval or augmented sixth chords. Such type of dissonance conveys musical emotions of 'tension', 'excitement', and 'anxiety' that many music composers referred to as the *sharp dissonance* (Cook, 2006). In this research, the term *tension dissonance* is used to denote this type of dissonance.

The fourth type of dissonance is associated to gloom and sad emotions, conveyed by music structures such as diminished intervals, minor mode, or the seventh chords. Some music composers refer to it as the *mild dissonance* (Schuijer, 2008). Such type of dissonance is relatively weak, leading many research scholars to believe that such effect has been accustomed in the western music literature. In this research, the term *gloom dissonance* is used to denote this type of dissonance.

If, as the theories discussed above imply, the meaning of musical dissonance has in fact multiple criteria, then each of the dissonance concepts need to be psychophysically measured and distinguished from each other in order to attribute the overall dissonance perception of a sound stimulus. To that end, a theoretical approach is proposed in this research to analyse the multi-dimensional nature of musical dissonance concepts.

The term ‘pitch-based’ in the hypothesis statement refers to an analytical method based on the pitch interactions. This differs from conventional psychoacoustic models where the dissonance effects are calculated based on pure acoustic analysis (time/frequency domain features). The strength of an acoustic-based approach is to model the functional responses for each organ along the auditory pathway in a ‘bottom-up’ based analysis. The research focus of this study is however not solely on the physiological aspect of sound perception, but also pertaining to the psychological principles for the perception of tonal music structures. Therefore, using pitch-based features can be more useful than solely acoustic features. Under a pitch-based approach, the perceived psychological features such as *sensory, tension, gloom* and *ambiguity* dissonances are estimated with respect to each audible pitch, instead of acoustic frequency partials. The overall consonance/dissonance perception of a given chord structure can be predicted by integrating the dissonance properties over all audible pitch components.

1.4 Main Contributions

This research work can be viewed as a comprehensive review of western music harmony. It includes discussion of musical theories, psychological descriptions, physiological and psychological interpretations as well as music mathematics and computational models pertaining to the concept of musical consonance and dissonance, and attempts to organize and integrate them in to a whole picture. It suggests that there are multiple meanings associated to the musical consonance/dissonance concepts, and each aspect of musical consonance and dissonance corresponds to a specific type of: (1) musical theory, (2) psychological experience and (3) psychophysical interpretation. (A detailed summary of *CDC* can be viewed in Table 29). By considering musical *CDC* as a multi-dimensional concept, we are able resolve a number of conflicts between previous theories, and build a comprehensive prediction model for the perception of music harmony.

This research also looks into some specific theoretical aspects and attempts to enhance the model performances in relation to previous theories, together with proposing new theoretical approaches to account for the perception of music harmony. This includes:

1. Enhancing the sensory dissonance based analytical model by addressing one of its theoretical problems: the secondary beats effect. The sensation of beats is in this research considered the first type of musical dissonance. Conventional sensory beats based computational models consider only the primary beat effect which occurs when two pure-tone partials are slightly separated in frequency. However, the discovery of the secondary beat effect means the accuracy of beats-based estimation can be further improved. In Chapter 5, a multi-peaked pure-tone dissonance curve is developed from a listening test, and modelled to estimate the overall sensory beats. It is observed that a better prediction result is achieved for musical dyads; however, like previous sensory-beats based models, incorporating the secondary beats effect is unable to predict the empirical rankings for tertiary triads.

2. Constructing an analytical model for *ambiguity dissonance* (Chapter 6). The concept of *ambiguity dissonance* is modelled based on how strongly a particular

pitch member competes with the tonal centre. This essentially categorizes the component pitches into three classes: a pitch supports the perception of a predetermined tonal centre; a pitch does not support the perception of a predetermined tonal centre but nor does it compete with the tonal centre; and lastly a pitch does not support the perception of predetermined tonal centre and acts as a competing tonal centre. In the computational model, only the last case is considered as the *ambiguity dissonance*. In this research, the *ambiguity dissonance* is viewed as second dimension of musical dissonance.

3. Proposing an analytical method for musical *gloom* and *tension*. Incorporating gloom and tension dissonances into the estimation of music harmony is a new theoretical attempt, they corresponds to the third and fourth dimension of the musical dissonance concepts. It is suggested that the perception of chord structures should be analysed in a context related to the tonal structure and progressions rather than solely in isolation (such as that for sensory dissonance and ambiguity dissonance). The model associates raised and lowered pitch contours to the *tension* and *gloom dissonances*. A detailed discussion and the rationales behind this approach are presented in Chapter 7.

On top of classifying musical dissonance into four dimensions, this research also compares the significance between four types of dissonance concepts in order to estimate perceived level of overall dissonance. A multi-dimensional dissonance concept means the presence of one type of dissonance may be more significant than the others or under certain condition, even mask their dissonance effect. To determine the significance that each type of dissonance concept contributing to the overall dissonance perception, a multivariable causal system is proposed. The significant coefficients of each type of dissonance are obtained through training data from the empirical rankings of musical triads. The results have shown that the overall dissonance effect is determined by the presence of (in the order of importance): sensory dissonance > tension dissonance > gloom dissonance > ambiguity dissonance. A more detailed discussion between *sensory, tension, gloom* and *ambiguity* dissonances is presented in Chapter 8.

1.5 Thesis Structure

Chapter 1 provides an overview of this Ph.D thesis, including the research background, motivations, object and scope, methodology as well as main hypothesis and contributions.

Chapter 2 provides a review of the music harmony concept, including the musical entities and psychological experiences involved, as well as the corresponding relationships between them, which are known as the music harmonic 'conventions' and 'laws'.

Chapter 3 reviews previous explanatory theories pertaining to the perception phenomenon of music harmony. The chapter begins with an introduction of the 'hardware' of music harmony perception – the auditory system, and followed by the acoustical, physiological and psychological based theories. The main focus is given to the introduction of two of the most prominent theories: *theory of beats* and *theory of harmonic template*.

Chapter 4 reviews previous psychoacoustic approaches for music harmony analysis. It firstly introduces the input and output requirements of the psychoacoustic systems; then it reviews the past numerical models that implement the *theory of beats* and *theory of harmonic template*.

Chapter 5 presents the modelling of *sensory dissonance*. The model presented in this chapter is closely related to the *theory of beats*, but with an extra consideration of the secondary beats effect. The chapter first introduces the definition of secondary beats effect and then obtains a quantitative measure of secondary beats effect from a listening test. Lastly, the prediction result is compared with previous *sensory dissonance* models.

Chapter 6 presents the modelling of *ambiguity dissonance*. This numerical model is constructed based on *theory of harmonic-template* but further emphasizes the role of the tonal centre (root of chord). The chapter first clarifies the role of tonal centre in relation to the *ambiguity dissonance* concept, and then the computational method is designed based on Terhardt's *virtual pitch determination algorithm*.

The model prediction results are compared with other harmonic-template-based analytical models by the end of this chapter.

Chapter 7 presents the modelling of *gloom* and *tension dissonance.* The chapter introduces the musical features corresponding to the gloom and tension emotions, and makes further use of such features to estimate the gloom and tension dissonances in musical chords. A listening perception study is also included to validate the theoretical thought presented in this chapter.

Chapter 8 presents the analytical model for pitch-based multi-dimensional dissonance. This model integrates the *sensory, tension, gloom* and *ambiguity* dissonances and produces a 'distribution of harmonic functions' for the input chord structure. For the 12-tet equal-temperament musical scales, the main harmonic function of each scale degree is summarized.

Chapter 9 concludes the thesis. The limitations and directions of further research work are also included.

1.6 Model Structure

One of the main research contributions is to construct an analytical model for the analysis of music harmony. The entire model is decomposed and introduced in four chapters (chapter 5–8). A general data flow diagram is presented in Fig.1.

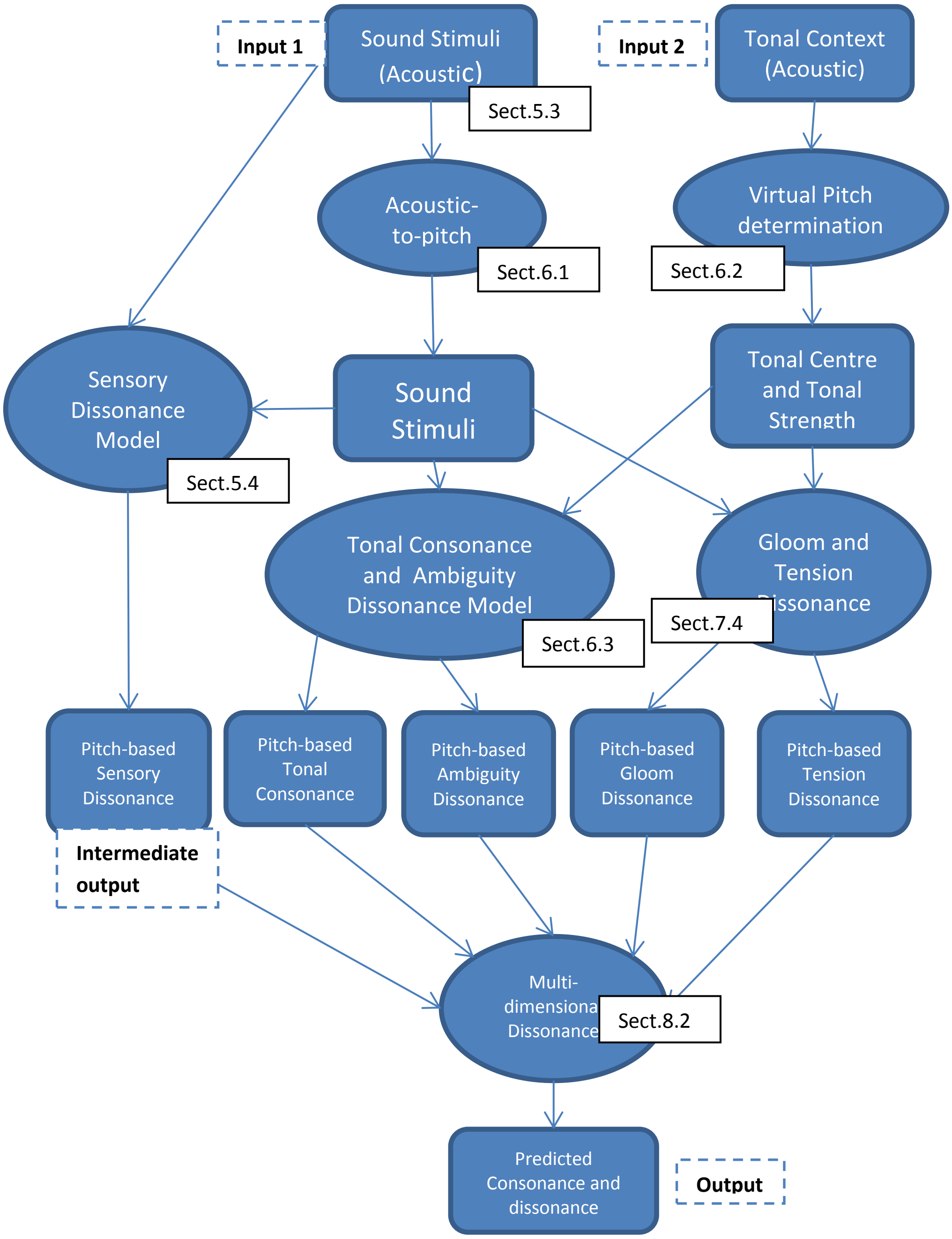


Figure 1 Model system diagram

Section 5.3 introduces the acoustic configuration of the object sound stimuli. Musical structures are converted into spectral features for further analysis. This corresponds to the first input of the entire model.

Section 5.4 presents the modelling of the sensory dissonance model.

Section 6.1 converts acoustic features into a set of audible pitch components for pitch-based tonal functional analysis.

Section 6.2 introduces the modelling of the second input – the tonal context, and a method to estimate its tonal centre and tonal strength.

Section 6.3 considers the tonal context and estimates the tonal consonance and ambiguity dissonance functions with respect to each audible pitch component.

Section 7.4 considers the tonal context and estimates the gloom and tension dissonance functions with respect to each audible pitch component.

Section 8.2 presents a method to integrate pitch-based harmonic functions into an overall consonance and dissonance perception for input sound stimuli.

Chapter 2 Music Harmony

The concept of music harmony generally refers to the use of simultaneous pitches (tones, notes) in achieving certain musical functions or effects. However, along with the development of western music, the meaning of harmony has strayed from its original form (what the Ancient Greeks referred to as *harmon ú* before the 4th century B.C.) (Tenney, 1988) and left us with many definitions. Some of the older understandings have never been completely replaced by the later ones, causing much confusion in music harmony related research. This chapter aims to resolve those confusions by summarising major music harmony related concepts.

One of the main features of music harmony is that its perception responses are usually described by the dichotomy of consonance and dissonance concepts *CDC* (Tenney, 1988). Consonance is somewhat associated with positive emotions such as ‘pleasant’, ‘relaxed’, and ‘agreeable’ or ‘concordant’ whereas dissonance generally has negative associations. The Consonance and dissonance concepts (*CDC*) however need to be distinguished from the concepts of ‘beauty’ and ‘ugliness’ in music aesthetics. The CDC in this research are limited to the ‘common and universal’ aspects of sound perception where the perception responses can be determined by the acoustical features. By contrast, aesthetics are largely influenced by an individual’s emotional state, cultural background and personal experience. From another perspective, *CDC* can be viewed as a lower-level informational ‘tool’ which music composers may use to achieve higher-level artistic emotions. Great musical works often contain a large proportion of consonance, but there are those who have achieved a successful balance between consonance and dissonance. The focus of this research is however on the *CDC* only.

Although most consonance and dissonance related concepts were formalized and developed under western music literature, many of the harmonic theories have also been globally recognized (Cook, 2006; Tenney, 1988) looked into the history of western music and outlined five stages of *CDC* (*CDC*1–*CDC*5). Each stage distinguishes itself from the others by: (1) the historical background, (2) the musical entity involved, and (3) the popular descriptors used for consonance and dissonance. Tenney's primary purpose is to provide a historical view of how *CDC* were derived from music practice. The review in this chapter is however result-oriented, focussing on the conclusions that have been made in the literature. This review also probes the harmonic reasons behind the music notations, such as tuning systems, chord systems, diatonic scales and major and minor tonalities. From a musical entity point of view, five categories of musical entities pertaining to *CDC* are visited in this research:

- single musical tone (section 2.1),
- between two musical tones (section 2.2),
- musical chord (section 2.3),
- between two musical chords (section 2.4), and
- musical modes (section 2.5).

By the end of this chapter, a conclusion (section 2.6) is presented where cross-comparisons between all the musical entities are made, and the main observations for the object of this study will be summarized.

2.1 Musical Tone

For the perception of an isolated single tone, it can be generally concluded that: A ‘musical tone’ usually has a salient pitch perception at comfortable loudness range (Lucker, Grzybmacher and Ventry, 1978). Compared to a random sound spectrum, a musical tone is usually perceived as consonant.

Acoustically, a ‘musical tone’ is associated with periodic waveforms and harmonic frequency spectra (Fig.2). In contrast, the pitch perception for noise can be ambiguous or lost. The boundaries between musical tones and noise are however not clear-cut: a noise tone may have harmonic sound components, and a musical tone may also contain a noise-like component, such as that in pitched percussion instruments. There is a scale by which the clearness of the pitch can be measured: more salient pitch perception gives rise to the clearer perception of a musical tone.

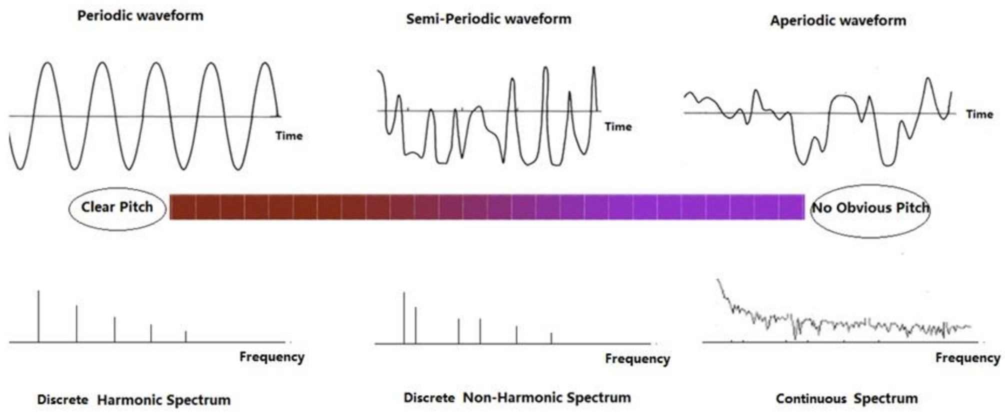


Figure 2 Musical tone vs. noise on a one-dimensional scale

Pitch sensation is an important feature of a musical tone. Almost within the same time period, eastern (Chinese Spring and Autumn Period, ~700–400 B.C) and western music theorists (Ancient Greek Era, ~800–400 B.C) discovered that perceived pitch is inversely-related to the length of a vibrating cavity/ string (Rossing, 2010). The length of vibration later led to the concept of frequency that has been commonly used for pitch measurements, where a higher vibration rate produces a higher pitch perception. Although modern theories pertaining to pitch perception have been more complicated, frequency can still be viewed as the primary determinant of pitch perception.

Besides pitch perception, the loudness of a musical tone should also be kept within a comfortable range. Empirical studies (Punch, et al, 2004) have concluded that the most ccomfortable loudness (MCL) level for musical sound averages between 40–60 dB whereas uncomfortable loudness levels (ULL) are generally above 90dB. However, the comfortable and uncomfortable loudness levels are also subject to the types of music, the environment, and the age group of subjects. The comfortable loudness level is also frequency dependent. Tones with frequency components below 40Hz generally sound ‘stressful’ and ‘muddy’; and those with frequency components above 2.5kHz generally appear to be ‘harsh’ and ‘sharp’; therefore, a comfortable frequency range is deemed to be between 40Hz and 2.5kHz. The general loudness range (in term of sound intensity) for musical tones can be viewed in Fig.3 below. Tones with uncomfortable loudness perceptions will be unconditionally recognized as noise therefore produce dissonant sensations.

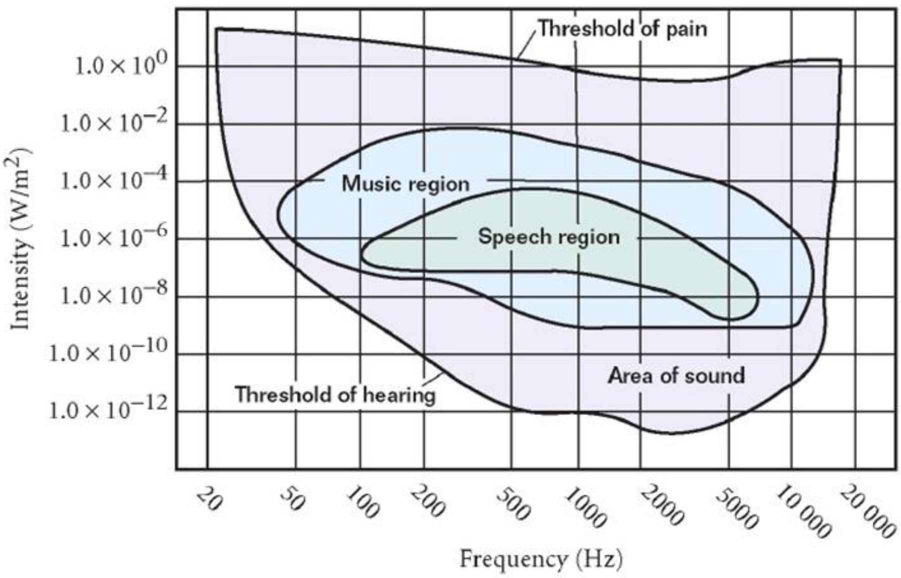


Figure 3 Sound intensity and frequency regions for musical tones

(Figure taken from Peterson, 1974)

While comparing two musical tones, some people may also prefer certain timbres over others. The timbre property, also known as ‘tone quality’ or ‘tone colour’, describes the innate characteristics of the sound source. Compared to pitch and loudness, timbre is the sound property that cannot be measured from ‘high’ to ‘low’ on a one-dimensional scale. In the literature, there was a debate of whether the timbre of a tone could be quantitatively measured. More recent research has indicated that timbre is associated to 2–4 semantic dimensions such as promoted in (Rasch and Plomp, 1999). For instance, the research studies from (Pratt and Doak, 1976) have identified three principal axis of timbre, namely:

- 'dull' – 'brilliant',
- 'cold' – 'warm'
- 'pure' – 'rich'.

The preference of timbre of a tone is typically influenced by personal emotional states and cultural factors, and less perception traits have been observed between the timbre of a tone and *CDC* (Phillip, 2014). This thesis is concerned with *CDC* perception between (a number of) musical tones, and it does not investigate the timbre perception of individual tones.

Lastly, we should also notice that not all musical forms make use of the musical tones. Contemporary noise music has developed a concept of noise aesthetics (Demers, 2010). Unconventional sound instruments, non-harmonic tones, silences, extreme volume and distortion have all been used by many noise and atonal musicians; some of them are well-received by the audience. Moreover, using electronic techniques to generate noise components has also been a very common way to create novel timbres or to produce custom sound effects. This thesis focusses on *CDC* perception within musical tones, although the models presented could also be relevant to these genres of music.

2.2 Between Musical Tones

The perceived harmony (degree of consonance/dissonance) *between* musical tones is primarily characterized by the size of the pitch interval[2] separating them. The historical definitions of consonance and dissonance intervals are however not static, and have been influenced by the musical forms that prevailed at each stage of western music history.

Early music (Ancient Greek 4th – 6th B.C.) was essentially melodic and monophonic[3]. The *CDC* at this stage (*CDC*1) concerns how the music tones progress from one tone to another. It was discovered that when successive pitches have simple integer frequency ratios, it will generally sound concordant and *symphonos*[4] (the first consonance concept); whereas complex and irresolvable ratios correspond to dissonant perceptions (Tenney, 1988).

The *Pythagoreans*[5] also developed an arithmetic approach where all musical intervals could be classified into six broad categories, namely (from consonance to dissonance): *equal, multiple, epimore, epimere, multiple epimore* and *multiple epimere* (see Table 1). Among these, *equal, multiple,* and *epimore* were viewed as the consonance intervals and the rest are dissonances.

---

[2] A pitch interval measures the pitch distance between two musical tones.

[3] Monophony is the simplest of textures, consisting of a single melodic line, without accompanying harmony.

[4] *Symphonos* is a Greek term used by Aristoxenus to describe the consonance intervals.

[5] Pythagoreans refers to a set of teachings and belief held by the Greek philosopher Pythagoras (580-500 BC) and his followers

Categories	***f.r***	**For 1 ≤ *f.r* ≤2**	**Diatonic Name**
equal	*fr* =1	1	Unison
multiple	*fr* =*n*	2	Octave
epimore	*fr*=(*n*+1)/*n*	3/2	Perfect Fifth
		4/3	Perfect Fourth
		5/4	Major Third
		6/5	Minor Third
epimere	*fr*=(*n*+*p*)/*n*	5/3	Major Sixth
		8/5	Minor Sixth
multiple epimore	$fr=((n+1)/n)^m$	25/16	≈Minor Seventh
multiple epimere	$fr=((n+p)/n)^m$	49/25	≈Major Seventh

Table 1 *Pythagorean* classifications of musical intervals

where *n*, *p*, *m* are positive integers; *fr* is the abbreviation for frequency ratio; and the corresponding names under western music are listed in the last column (Tenney, 1988)

Later in the Medieval Era[6], the advent of polyphony and the emerging compositional techniques led music theorists to consider music harmony for *simultaneous* pitch pairs (musical dyads) (*CDC*2). The 'consonance' intervals were classified into three subcategories: *perfect, intermediate* and *imperfect* consonances. In later years, dissonance intervals were also classified similarly into three categories, namely *perfect, intermediate* and *imperfect* dissonances (Tenney, 1988).

[6] Periods and Eras of Western music literature

Antiquity	Ancient Greece	b.c. 800-300
	Ancient Roma	b.c. 300-c.500
Early	Medieval	c. 500-1400
	Renaissance	c.1400-1600
Common practice	Baroque	c.1600-1760
	Classical	c.1730-1820
	Romantic	c.1780-1910
	Impressionist	c.1875-1925
Modern and contemporary	Modern	c.1890-1975
	Postmodern	c.1975-present

Moving forward in history to the Renaissance period, (*CDC*3: 14th – 17th century), with the prevalence of counterpoint[7], the dissonance classes began to merge into one, the intermediate consonance class was withdrawn; and the overall number of consonance and dissonance classes began to decrease. The consonance and dissonance classes were simplified into three main categories (Tenney, 1988):

1. *Perfect Consonance*: Unison, Octaves, Perfect Fifth
2. *Imperfect Consonance*: Major and Minor Third, Major and Minor Sixth
3. *Dissonance*: Perfect Fourth, Major Second, Major Seventh, Augmented Fourth and Minor Second.

The entire evolution of consonance and dissonance intervals is illustrated in Figure 4.

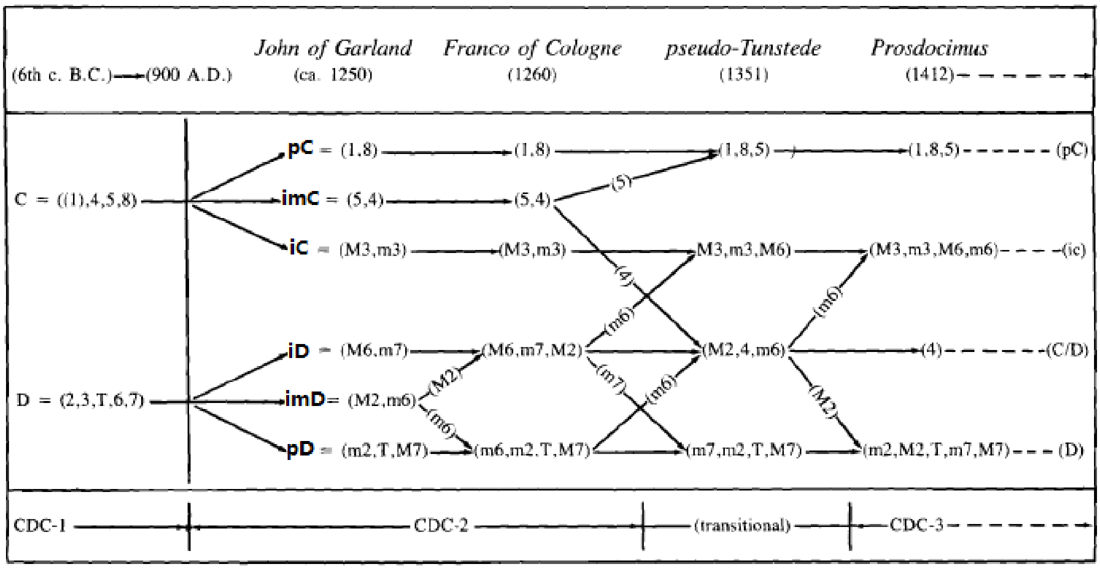


Figure 4 Evolution of consonance and dissonance intervals

Legend:

M = major, m = minor, T = tritone, p = pure, im = intermediate, i = imperfect, C = consonant, D = dissonant, Numbers = diatonic intervals. Figure copied from (Tenney, 1988)

The music harmony view of musical intervals is generally summarised by the *Proportion theory* (Randel, 1990), which is a line of thinking originating from the *Pythagoreans* and supported by music theorists such as Euler and Lipps at the beginning of the 20th century. The central hypothesis of the *Proportion theory* is that

[7] In music, *counterpoint* describes the harmonic relationships between two or more voices contours

the degree of dissonance is proportional to the complexity of frequency ratios. Euler (Euler, 1766) provided a numerical method to analyse the complexity of ratios. In his approach, musical intervals (in terms of frequency ratios) can be mathematically represented with the use of prime numbers. Juan (Juan, 2006) extended his thought and developed the concept of *harmonic field*. Below is an example of how *harmonic field* works:

With a prime number base of {2,3,5}, diatonic intervals can be generated under the specification of three integer indexes. When larger integer indexes appear, the interval tends to be more dissonant:

$<0, 0, 0> = 2^0 \cdot 3^0 \cdot 5^0 = 1 = 1/1$ unison;

$<-1, 1, 0> = 2^{-1} \cdot 3^1 \cdot 5^0 = 3/2 = 3/2$ fifth;

$<1, 1, -1> = 2^1 \cdot 3^1 \cdot 5^{-1} = (2x3)/5 = 6/5$ minor third;

$<-5, 2, 1> = 2^{-5} \cdot 3^2 \cdot 5^1 = (3^2x5)/25 = 45/32$ tritone;

However, there are some serious shortcomings of the *Proportion theory*.

The first is that most intervals may be assigned to more than one ratio, depending on the prime numbers used. For example, a *major second* can be both 9/8 ($3^2/2^3$) or 10/9 ($2x5/3^2$). Therefore, based on the prime number used, the consonance and dissonance properties can be completely changed.

Secondly, the use of an accurate number to represent an interval means that a slight mistune may cause a drastic change of the complexity of the frequency ratio, therefore the perception of consonance and dissonance should be significantly altered. For example, a perfect octave has a ratio 2/1, but a slightly mistuned octave might have the ratio 51/25. The latter consists of big integer numbers which should (according to the theory) cause a big dissonance, but in reality this mistuned effect is almost imperceptible to our ears. To resolve this, a degree of tolerance which reflects the just-noticeable-difference of intervals (1% according to Kollmeier et al., 2008) has been allowed.

The third and also the most critical problem with the *Proportion theory* is that it cannot be used to interpret the perception of the fourth interval. Under *Proportion theory*, the perfect fourth is without doubt a consonant interval; however, in *CDC*3, it can be observed that the fourth interval was moved from the consonant class to the dissonance class. Such peculiar observations had also been noticed by Tenney, which forced him to conclude that "*the perfect fourth interval cannot be explained with a previous way, a second factor is invoked... the emergence of a new criterion... involving another aspect of the sonorous character of simultaneous dyads.*"

The *Proportion theory* works well for isolated tone pairs, whereas the dissonance of the perfect fourth can be viewed as a result of considering its musical context (*harmonic relation theory*) that will be discussed in the next section.

The *CDC* observations of musical intervals are also one of the major reasons behind musical tuning systems. A music tuning system defines a series of pitch scales used for music composition and performance. The primary purpose of a tuning system is to 'contain' those consonance intervals (such as those with frequency ratios of 2:1 3:2, 4:3, 5:4 and so on) on its scales (*requirement 1*). On the other hand, a tuning system should also make sure that same set of frequency ratios can be found with respect to each individual scale position (*requirement 2*) which infers the use of geometric sequence. A natural mathematical conflict occurs when a tuning system tries to satisfy both requirements (1&2) at the same time. To demonstrate such conflict, let's assume a reference note (the first scale position) whose frequency is x Hz: suppose we want the second scale position to be defined by the 5/4 interval, and the third scale position by the 4/3 interval; when looking at the second scale position, we notice that its next note (the third scale position) does not give the desired 5/4 ratio (but 16/15 instead). In the literature the *just-tuned* scales generally manage to satisfy the first requirement, whereas the *equal-temperament* scales focus on the second requirement.

An equal-tempered system (constant frequency ratios between adjacent scale positions, a geometric series) is mathematically unable to contain all simple frequency ratios with respect to a specific scale note. Finding a proper equal-temperament system can be viewed as an optimization problem where those consonant intervals should be contained as accurately as possible. Besides unison (frequency ratio of 1:1), the octave interval (frequency ratio of 2:1) has been considered as the most consonant interval, therefore the frequency ratio of 2:1 has the first priority to be accurately presented on the scale. The geometric frequency series of $2^{m/n}$ thus can be used for the equal-temperament scale. Under such a system, n divisions are made between the 1:1

and 2:1 frequency ratios where m represents the scale degrees running from 1 to *n*. When *m*=*n*, the frequency ratio of 2:1 can be perfectly contained. The remaining task is to find a proper value of n such that the errors of frequency ratios of other consonance intervals, especially the diatonic fifth and fourth intervals can be minimized. A mathematical simulation (Sethare, 1999) has shown that a 12 division (*n*=12) within an octave is one of the optimized solutions to cover the consonance intervals at minimum cost (see Fig.5) comparing to other equal-temperament systems. Thus the 12-tet equal temperament can be viewed as the mathematical result of consonance intervals (Sethare, 1999).

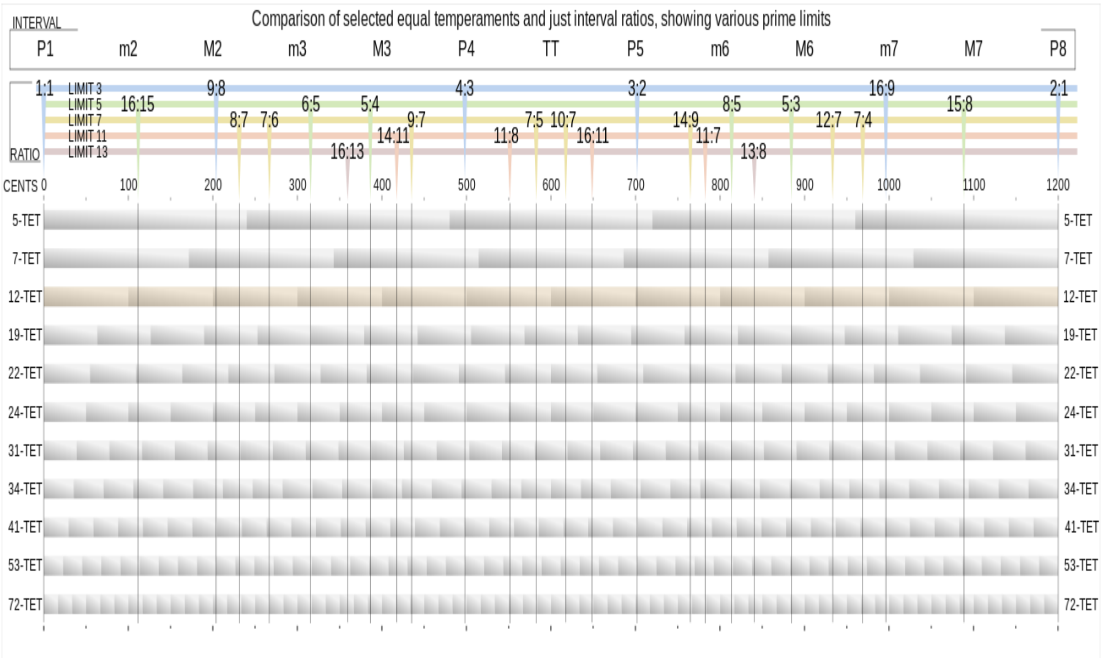


Figure 5 Comparison between N-tet division methods

The figure is taken from (Sethare, 1999 Fig.4.6 p. 58)

Table 2 presents the errors introduced by 12-tet equal temperament between its 12 scale positions and the simple integer frequency ratios. It can be observed that the errors for the frequency ratios of 3/2 (diatonic fifth) and 4/3 (diatonic fourth) have been kept below 2 cents[8]. Besides unison and octave intervals, the diatonic perfect fifth and fourth intervals have smaller error than other intervals, therefore under 12-tet

[8] The unit cent corresponds to a frequency ratio of $2^{1/1200}$.

equal temperament, the consonance of simple frequency ratios such as 1:1 (unison), 2:1 (octave), 3:2 (perfect fifth) and 4:3 (perfect fourth) can be perceived.

Interval	No. of Semitones	Frequency Ratios	Equal Temperament Ratios	Error in Cents
Unison	0	1/1 = 1	$2^{0/12} = 1$	0
Minor Second	1	16/15 = 1.06666	$2^{1/12} \approx 1.0595$	-11.73
Major Second	2	9/8 = 1.1250	$2^{2/12} \approx 1.12246$	-3.91
Minor Third	3	6/5 = 1.2000	$2^{3/12} \approx 1.18921$	-15.64
Major Third	4	5/4 = 1.2500	$2^{4/12} \approx 1.25992$	+13.69
Perfect Fourth	5	4/3 = 1.3333	$2^{5/12} \approx 1.33483$	+1.96
Tritone	6	7/5 = 1.4000	$2^{6/12} \approx 1.41421$	+17.49
Perfect Fifth	7	3/2 = 1.5000	$2^{7/12} \approx 1.49831$	-1.96
Minor Sixth	8	8/5 = 1.6000	$2^{8/12} \approx 1.5874$	-13.69
Major Sixth	9	5/3 = 1.6667	$2^{9/12} \approx 1.68179$	+15.64
Minor Seventh	10	7/4 = 1.7500	$2^{10/12} \approx 1.7818$	+31.17
Major Seventh	11	15/8 = 1.8750	$2^{11/12} \approx 1.88775$	+11.73
Octave	12	2/1 = 2	$2^{12/12} = 2$	0

Table 2 Intervals under diatonic scales

2.3 Musical Chord

A musical chord contains at least three musical tones, and it can be played simultaneously (simultaneous chord) or successively (broken chords or *arpeggios*). In the western music literature, diatonic chords typically have following features[9] (Schönberg and Stein, 1985):

(1) A musical chord should consist of musical notes from different pitch classes. According to (Károlyi, 1995), each individual note in a chord should be identifiable and distinguishable from other notes; therefore the basic form of chord should avoid using notes from the same pitch class[10].

(2) The minimum spacing of two adjacent musical notes in a chord should be at least three semitones. Simultaneous tones at one or two semitones generally produce a significant dissonance effect (later known as sensory dissonance, see section 3.2). To avoid this, basic musical chords are typically built by a stack of major or minor third intervals (three or four semitones), and such chords are also known as tertian chords.

(3) One of the methods to denote a chord is according to its *root*. During the practice of polyphony in the medieval era, it was discovered that certain combinations of musical tones tend to merge as one, therefore these tones could be combined and treated as an independent musical element in the progressions of music harmony; such combinations of tones typically revolve around a pitch centre, known as the root of chord. *Root* is to a chord as pitch is to a tone: *root* somewhat represents the holistic pitch perception of the chord.

The *CDC* of a chord are closely related to the salience of its root: musical chords with strong root perceptions are likely to be classified as a consonant chord; those with weak or no root perception are typically recognized as a dissonant chord structure.

[9] Such requirements were typically seen in Baroque period (c.1600-c1760), where contemporary harmony does not necessarily obey.

[10] Octave related musical tones (frequency ratio of 2:1) are highly concordant; when played simultaneously, they will be perceived as highly merged. In music, such tones are grouped under a same class, called the pitch class.

Based on the chord structure, the salience of a root can be determined by the *harmonic relation theory* (Rameau and Wundt, 1721).

Harmonic relation theory was postulated by the French music theorist J.P. Rameau in the 18th century. Rameau believed that the intervals that appear in the natural harmonic series (what he proposed as the *Corps Sonore*) have a strong function to indicate a root at its fundamental frequency (see Fig.6), a mechanism akin to the overtones (partials) suggesting the pitch perception at fundamental frequency. Different from the frequency ratios, Corps Sonore can be used to estimate the consonance and dissonance perception of a single isolated sound entity instead of measuring the harmonic relationships between two musical tones (or frequencies).

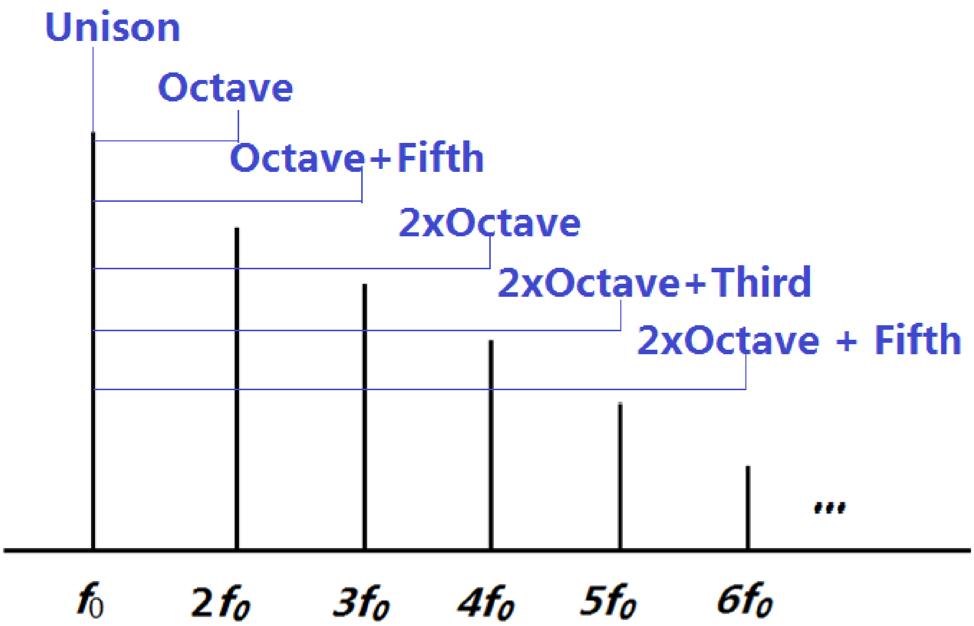


Figure 6 Intervals of the first six natural harmonic overtones (Illustration of *Corps Sonore* concept)

The intervals formed by f_0 and its first five overtones are:

- octave (frequency ratio: 2:1),
- fifth above one octave (3:1),
- two octaves (4:1),
- (major) third above two octaves (5:1), and
- fifth above two octaves (6:1).

Higher harmonics generally contribute less to perception at f_0. The role that the next harmonic ($7f_0$) contributes to the root perception is also significantly weakened; therefore, it is not shown in Fig.6.

Under *harmonic relation theory*, the fifth interval (frequency ratio of 3/2) and major third interval (frequency ratio of 5/4) have special functions indicating that its lower tone is the root. For this reason, musical chords containing fifth or major third intervals tend to have a clearer root. For instance, the major triad (root position) contains both major third and fifth intervals and therefore is deemed the most stable of triads; the fifth interval in the minor triad (root position) also makes it relatively consonant than other triadic structures. Conversely, the fundamental frequency of the upper note in a perfect fourth (frequency ratio of 4/3) is not present as one of the natural harmonic frequencies of the lower (tonic) note. Therefore, during the Renaissance period where musical chords had been frequently used in music compositions (*CDC*3), the fourth interval was moved from a consonance class to dissonance (Fig.4).

Besides the harmonic relations, the position of the root also has an impact on root perception (Schönberg and Stein, 1985). People tend to treat the lowest note as the root as compared to the higher notes in a chord. When a chord is arranged at its root position, the lowest note is perceived as the root; perception-wise, it is deemed to be more consonant than its first inversion where the highest note is now the root. Furthermore, when neither major third or fifth interval exists in a chord (the case where no obvious root is perceived according to the *harmonic relation theory*), the lowest note serves as a 'default root' to the chord.

Within a chord, a note that is neither a fifth nor a major third above the root can be viewed as a dissonant component. However, since the 19th century, more dissonant intervals were allowed to increase the complexity of harmony. In particular, (Roberts, 1986) observed that listeners judge *just* intervals as being less pleasant than slightly mistuned intervals. Vos (1986) had also noted that '*some people may rate pure intervals to be "insipid" and therefore prefer intervals slightly tempered as expressing greater "warmth."*' In another listening test (Huron, 1993), listeners were presented with A: a complex tone with 10 harmonics at 100Hz fundamental frequency and B: a dyad of pure tones at 200Hz and 300Hz. Most listeners considered B as being more ‘pleasant’, ‘euphonious’ or ‘consonant’ than A. All the above-mentioned evidence seems to overthrow the hypothesis of *harmonic relation theory*; however, such observations may reflect listeners’ preferences, but not consonance concepts. The personal and cultural aspects should be ideally removed before *CDC* are considered.

Depending on the number of tones involved, musical chords can be further categorized into triads, tetrads and extended chords. Among them the musical triad is the most important musical component for chord harmony. With the requirements (1), (2) and (3) stated at the beginning of this section, the possible combinations of intervals give rise to four basic types of musical triads (and their inversions): major, minor, diminished, and augmented (see Table 3).

Intervallic structure	The position of Root	Triad types
3–3	Lowest note	Diminished (root position)
3–4	Lowest note	Minor (root position)
3–5	Middle note	Major (1st inversion)
3–6	Middle note	Diminished (1st inversion)
4–3	Lowest note	Major (root position)
4–4	Lowest note	Augmented
4–5	Middle note	Minor(1st inversion)
5–3	Highest note	Minor(2nd inversion)
5–4	Highest note	Major (2nd inversion)
6–3	Highest note	Diminished (2nd inversion)

Table 3 the triadic structures

The ‘intervallic structure’ of triads is represented with the two intervals (Lowest to middle note – Middle note to highest note in term of semi-tones)

The empirical studies from (Robert, 1986) had demonstrated that the ‘stability’ (what he used as an equivalent consonance concept) of the triad decreases in the order of: Major > Minor > Diminished > Augmented, regardless of the subjects’ cultural and musical backgrounds. The harmonic relation theory mentioned previously can be used to explain Robert’s empirical result:

- a major triad contains both major third and fifth intervals therefore it is most consonant;

- a minor triad contains the fifth interval only, therefore it is next consonant;
- diminished and augmented triads contain no fifth intervals and are therefore treated as dissonant chord structures.

However, the *harmonic relation theory* fails to explain why the augmented triad is less consonant than the minor, since the augmented triad has a major third interval whereas the diminished triad does not. Moreover, Roberts also showed that for each triadic structure, the root positions are more ‘stable’ than the first inversions, which are in turn more consonant than the second inversions.

Musical chords tend to have their own featured ‘quality’ or ‘colour’, as James Linderman (2014) described:

‘*The quality of the chord has everything to do with the mood it helps create. Major chords tend to sound happy, while minor chords evoke a feeling of sadness. Diminished chords can help create a feeling of anticipation or a discontented mood depending on their application, while augmented chords tend to sound anxious or sometimes remind me of what a hangover would sound like, if a hangover made a particular sound (though sometimes, of course, they do!).*’

Similar to the timbre preferences of a musical tone, the preferred ‘colour’ of chords may also be influenced by personal and cultural factors.

One way to construct a musical tetrad is by adding an extra note on top of a musical triad. Under condition (2) stated at the beginning of this section, adding a tone that is a second, fourth, or sixth interval higher than the root will introduce sensory dissonance. Therefore, a tone that is a seventh higher than the root is usually added to form a musical tetrad.

In early times (roughly before the 17^{th} century), the seventh interval in tetrad was considered to be dissonant as it ‘destabilised’ the musical triads. But later it became frequently used in contemporary music, especially in Jazz. It is believed that the harmonic perception of the seventh interval has been ‘accustomed’ so that it does not sound as dissonant as it used to (Benward and Saker, 2003).

By adding a seventh interval into four basic types of triads (root position), we can obtain the following combinations of seventh chords that are commonly used (Table 4):

Chord name	Symbol Notations	Intervals above the root		
		2nd note	3rd note	4th note
Major seventh	maj^{7} M^{7} Δ	major third	perfect fifth	major seventh
Major minor seventh	^{7}	major third	perfect fifth	minor seventh
Minor major seventh	$\mathrm{m}^{\mathrm{maj7}}$ m^{M7} $\mathrm{m}^{\Delta 7}$ - j 7	minor third	perfect fifth	major seventh
Minor seventh	min^{7} m^{7} -7	minor third	perfect fifth	minor seventh
Diminished major seventh	$\mathrm{m}^{\mathrm{M7}\flat 5}$ $-7\ \Delta^{\flat}5$	minor third	diminished fifth	major seventh
Half-diminished seventh	m^{7} $-7\ (^{\flat}5)$	minor third	diminished fifth	minor seventh
Augmented major seventh	$\mathrm{maj}^{7\ (\sharp 5)}$ $+^{\mathrm{M7}}$ $+^{\Delta 7}$	major third	augmented fifth	major seventh
Augmented minor seventh	aug^{7} $+^{7}$	major third	augmented fifth	minor seventh

Table 4 Common seventh chords

The name for a seventh chord has a general form of: ‘name of the triad it contains’ and ‘the type of the seventh interval’. For instance, minor–major seventh means it contains a ‘minor’ triad and a ‘major seventh’ interval

The consonance and dissonance rankings for the types of seventh chords generally follow those of their contained triads (order of consonance: major > minor > diminished > augmented). However, the seventh interval is not the only interval which ‘extends’ the harmony of the triads. Higher intervals such as the ninth, eleventh and thirteenth extend music harmony in a similar way. Chords contain such intervals are generally referred as extended chords.

2.4 Between Music Chords

In music literature, there are also composition 'rules' on how musical chords should progress from one to another. Sequential chords generally progress 'well' when they are harmonically related. The harmonic relationships between two musical chords are slightly more complicated than between two musical notes, as two relationships have to be considered: the first one is the harmonic relationship between the roots of chords; and the second is the relationships between the note members of the two chords. According to Parncutt (1987), sequential chords generally progress well (consonance) when:

(1) The roots of two chords are harmonically related.

As reviewed in section 2.2, unison, octave, fifth, and fourth are the consonant intervals. When the roots of two sequential chords have such intervals, they are deemed to be harmonically related. For instance, the progressions from a C-major triad to F-major or G major-triad are quite concordant as the interval between C & F is a fourth, and between C & G is a fifth.

(2) The note components of two chords are harmonically related.

One special case of this is when two chords share many common notes (unison intervals). Taking the example of C-major triad again, it has three note components: C, E, and G. Searching for other major or minor triads that share two musical notes with C-major triads, we get E minor triad (shares E and G with C-major), A-minor triad (shares C and E) and C-minor triad (shares C and G). Therefore, the E-minor, A-minor and C-minor triads are also harmonically related to the C-major triad.

Many uses of such harmonic relationships have been developed, such as the harmonic table (Vancouver, 2008) (see Fig.7).

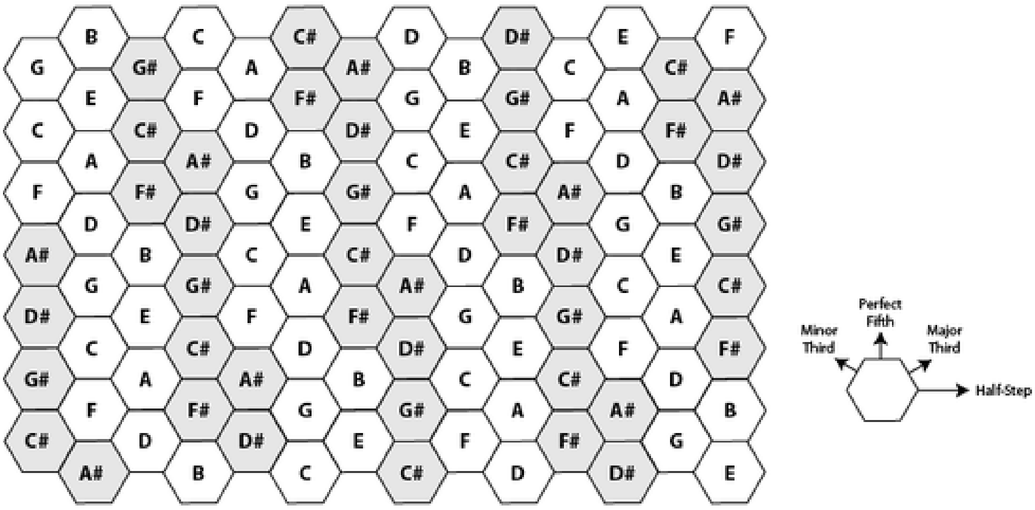


Figure 7 Chromatic note layout for the harmonic table (Vancouver, 2008)

The above-mentioned criterion of consonance solely considers the isolated progressions between two musical chords. However, under the context of a specific musical mode, the consonant and dissonant progressions can be different. For instance, a G-major chord within the C major mode is usually perceived to be stable and consonant, whilst it is less consonant under F major mode as the B note in G-major chord does not feature in F-major diatonic scale.

2.5 Musical Mode

In western music literature, the concept of a *mode* (seen in *Gregorian chant theory* or in *Renaissance polyphonic theory*, or *functional harmony*[11]) consists of a set of frequently used scales. Seven pitch classes were typically involved to form a musical scale.

Under each musical mode (in tonal music applications), musical tones are categorized into a series of scales. In tonal music, a particular pitch class provides a subjective sense of 'home base' where the music wants to be resolved to, such pitch class is referred as the *key*. A key does not necessarily remains the same throughout the whole piece; but for a specific segment of tonal music, there is usually a single, identifiable tonal center.

Based on the intervallic distances from the key, the seven diatonic scale degrees can be described in Table 5.

Scale degree (number)	Distance to key	Scale degree (Name)	Figure bass	Solfege
1st	0 semitones	Tonic	i	Do
2nd	2 semitones	Supertonic	ii	Re
3rd	3 or 4 semitones	Mediant	iii	Mi
4th	5 semitones	Subdominant	iv	Fa
5th	7 semitones	Dominant	v	So
6th	8 or 9 semitones	Submediant	vi	La
7th	10 or 11 semitones	Leading tone	vii	Ti

Table 5 Diatonic major scale degrees

The first scale degree contains the music key, therefore contains a sense of 'finality' and 'arrival', known as the *tonic*. Compared to other scale degrees, the first scale degree is the most consonant.

[11] Functional harmony was pioneered by the German music theorist Hugo Riemann in the late 19^{th} century. Riemann's initial purpose was to study why and how certain chords led to other chords. Under functional harmony, there are primarily three types of functions in harmonic chord progressions: tonic, dominant and subdominant.

The fifth scale degree is second in importance in a diatonic scale. The fifth scale degree is a pitch class that is fifth interval higher than the tonic. The fifth scale degree has a harmonic function to establish the tonic through music progressions. In functional harmony, a fifth scale degree chord (dominant) usually resolves to the tonic, known as the perfect authentic cadence (Kaplan, 1996).

The fourth scale degree can be viewed as a fifth interval lower than the tonic, therefore it is known as the sub-dominant. The subdominant scale degree is one of the three core scale degrees under functional harmony (the other two are tonic and dominant); its primary role is to lead to the dominant. Furthermore, modulation to subdominant usually creates a sense of relaxation.

The third scale degree is known as the mediant. The mediant is deemed as an extension of the tonic scale degree (tonic parallel) by Schenkerian analysis (Salzer, 1952), and it is also viewed as having dominant functions (dominant parallel) under functional harmony. Different from the fourth and fifth scale degrees, the third scale degree can be either major or minor, resulting in two fundamental types of tonalities: the major and minor tonalities.

The sixth scale degree can be viewed as a third interval lower than the tonic, therefore it is also known as the sub-mediant scale degree. In chord progressions, the sub-mediant is usually preceded by mediant scale degree. The sixth scale degree also has major and minor types.

The major seventh scale degree (a semitone lower than the tonic) is known as having a leading function that is urged to be resolved to the tonic. Interestingly, for the minor second scale degree (which is a semitone higher than the tonic), the leading function is not observed.

There are fundamentally two types of diatonic scales, known as the diatonic major and minor scales, which can be viewed as essentially two musical modes derived from the major and minor triads. To illustrate this, we may start from a tonic major triad at root C (C–E–G). From section 2.4, we observed that the G major (G–B–D) and F major (F–A–C) triads are the only two triads where both roots and individual notes are harmonically related to the C major triads. The contained notes in the C, F, and G major triads form seven scale degrees (A–G). In order to share more notes with C, F, or G major triads, the triads at second, third, sixth scales degree are under the minor structure; and the seventh scale degree is a diminished triad.

Similarly, considering the C minor triad as the tonic triad, we can obtain its harmonic relatives of the F minor and G minor triads. The C, F, and G minor triads contain seven scale degrees of C, D, E^b, F, G, A^b, B^b. For harmonic reasons (sharing common notes), the triads at the third, sixth and seventh scale degrees are major and the second scale degree is diminished. However, the minor scale has a weakened leading tone function as the interval between the seventh to the tonic is two semitones instead of one. To resolve this, the seventh scale degree is raised by one semitone and thus forms a modified minor scale called the harmonic minor scale. Moreover, as the harmonic minor scale creates a 'gap' between the sixth scale degree and seventh scale degree (three semitones), the sixth scale degree is therefore also raised by one semitone for melodic motions, forming the so-called *melodic* minor scale.

The use of the first, third (major) and fifth scale degrees can also be seen in eastern music cultures. Early in the Chinese Spring and Autumn Period (770 B.C – 476 B.C), a pentatonic scale consists of '宫'(gong), '商'(shang), '角'(jue), '徵'(zhi), '羽'(yu) was used in court music. Under the Chinese pentatonic scale, '宫', '角', '徵' correspond to the first, third and fifth diatonic scales in western harmony. The common use of the first, third and fifth scale degrees can be a result of the major triad which has been universally recognized as consonant (Chen, 2002).

Besides the first, third (major) and fifth scale degrees, other scale degrees can be generally viewed as the 'added colour' or 'dissonance' to the music mode as they do not 'confirm' the tonality of tonic. However, the use of such dissonance scale degrees reflects the features of a specific music mode. For instance, in the Japanese pentatonic mode, the fifth scale degree is missing, whereas the fourth and seventh (major) scale degrees play an important role. In Jazz music, the use of the second and major/minor seventh scales add a degree of 'warmth' and 'graceful sadness' to the music emotions (Duncan and Barrett, 2007); in Baroque music, the use of the augmented sixth scale degree introduced a unique sense of 'tension' that needs to be resolved to the dominant (Tenney, 1988).

Today, a broader definition of a music mode has become: ‘a set of scale degrees that are frequently used within certain music styles’ (Benward and Saker, 2003). The choice of scale degrees is no longer limited to consonance and dissonance analysis or ‘harmonic functions’, but has become more focused on the ‘colours’ of chords introduced by the scale degrees. In other words, more auditory dissonance has been introduced into music composition, which diversifies the harmony beyond the consonance of the tonic major triads.

2.6 Chapter Summary

The musical entities reviewed in this chapter can be summarized with a three-layer system: namely *tone*, *chord* and *mode*.

Throughout this thesis the term *mode* refers to a series of scale degrees; a chord consists of a series of musical intervals; and a music tone consists of a set of frequency partials.

For each layer, there is a correspondence to the *tonal centre concept*: for a musical tone, this is simply its pitch; for a musical chord, it is the root of chord; and for a musical mode, it is the tonic (key).

Each layer also has its own perceived 'quality' (the 'timbre aspect' of harmony): for a tone, this means its timbre perception characterized by its spectral content; for a chord, it is the quality of the chord characterized by its intervallic structure; for a mode, it is perceived tonality that is determined by its scale degrees.

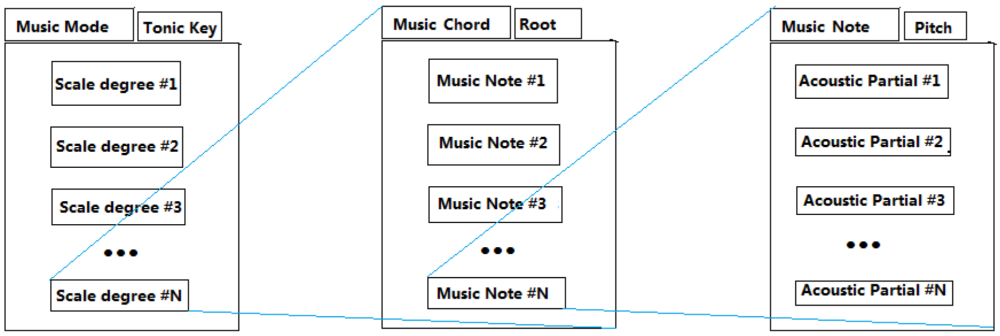


Figure 8 The three-layers of musical entities

The musical relationships pertaining to *CDC* can be categorized into three types, summarized as 'isolated', 'mutual' and 'contextual' type herein. The 'mutual' and 'contextual' relationships are illustrated in Fig.9.

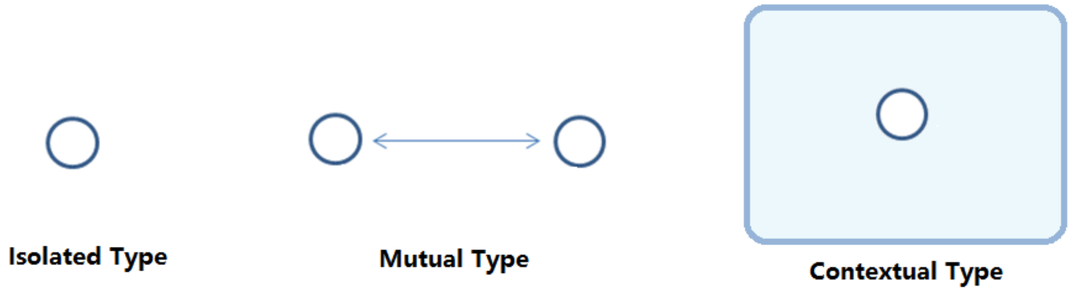


Figure 9 Three types of harmonic relationships

The illustration of isolated (left), mutual (middle) and contextual relationships (right), the white circles represent musical entities

The ‘isolated’ type refers to the perception of a single musical entity on its own. The perception conventions for an isolated tone (section 2.1) or chord (section 2.3) can be considered under this type. The consonance and dissonance perception of an isolated musical entity is determined by its internal structure. For instance, a musical tone / chord structure with clearer pitch/ root perception is considered to be more consonant.

The ‘mutual’ type of harmonic relationship is defined between two music entities, and these two entities are on the same layer (see the three-layer representation of music harmony, Fig.9.); for instance, the relationship between two musical notes (section 2.2), and the relationship between two chords (section 2.4). Such a type of music harmony is generally governed by *proportion theories*, where simple frequency ratios correspond to the consonance, and complex frequency ratios correspond to dissonance.

The ‘contextual’ type is the relationship between an individual music entity and its tonal context. Following relationships are categorized under the ‘contextual’ type of harmony: the relationship between a particular acoustic partial in a tone and the pitch of that tone (section 2.1); the relationship between a particular note in a chord and the root of that chord (section 2.3); and the relationship between a particular chord under a music mode and the tonic key of that mode (section 2.5). The ‘contextual type’ of harmony is governed by *harmonic relation theory*, where a frequency belonging to the overtone series of another reference frequency is generally considered as consonance (see section 2.3).

Music harmony is, however, not limited to the musical entities that have been reviewed in this chapter. The fundamental harmonic laws as summarized in this chapter are able to provide some basic guidelines for more complicated music composition rules to be analysed, such as counterpoint, functional progressions, musical composition period, and the structure of musical forms.

Lastly, *CDC* have also been associated with melodic motion. In a melodic progression, without considering the consonance and dissonance property of musical intervals, step-wise motions (consecutive notes with a difference of one scale degree) are usually considered as being more consonant than melodic leaps (consecutive notes with a difference of more than two scale degrees) (Delone, 1975). In music, the concepts of passing notes/chords, and auxiliary notes, are also examples of melodic progression, where the passing notes/chords usually happen at the bridging sections and the auxiliary notes are usually fast and transient that create a dynamic effect.

The main focus of this research is the tonal music structure. The harmonic laws involved are summarized in the following points:

(1) For musical dyads: empirical studies have shown that the unison, octave and perfect fifth intervals generally correspond to the consonance concepts; whereas the harmonic properties of the major third, major sixth and perfect fourth have been moved back and forth between *CDC*; other chromatic intervals are generally regarded as the dissonance intervals.

(2) For musical triads: It is universally clear that the perceived level of consonance follows the sequence: major > minor > diminished > augmented triads regardless of their inversions. It has also been observed that the existence of one or two semi-tones plays a dominant role for the dissonance perceptions (Sethare, 1999).

(3) When musical chords consist of four or more notes, the perceived harmony is primarily influenced by their lower triads. Additional higher notes generally enrich the colours of the chords (Eric, 2012). There is no general agreement in the literature concerning the ‘functions’ of each added note.

To conclude, this chapter reviewed and summarized the major harmonic laws and conventions observed in western music literature. The primary goal of this chapter is to review the musical-related terms that clarifies the object of this research study.

With a general understanding of 'what' music harmony is, and 'how' music harmony works, in the next chapter, we shall start to explore 'why' music harmony works in such a way by looking into the physiological and psychological aspects of sound perceptions.

Chapter 3 Physiological and Psychological Account

Explanations of music harmony can be dated back to Ancient Greek times (4th century BC). With very limited understanding of acoustics, auditory physiology and cognitive psychology, the mystery of musical consonance and dissonance took on a strong theological direction: The *Pythagoreans* believed that the frequency ratios of consonance (*symphonic intervals*) were those that could be found directly from the *τετρακτύς* (*tetractys*, see Fig.10).

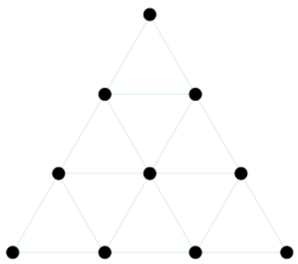


Figure 10 Pythagorean Tetraktys

A mystical symbol consisting of ten points arranged in a triangular form. The integer numbers that appeared in the *Tetraktys* (1 to 4) construct five frequency ratios: 1/1, 2/1, 3/1, 3/2, 4/1, and 4/3 which were considered to be the frequency ratios of the consonant intervals (Tenney, 1988)

Modern theories are developed based on the psychophysical insight of sound perception. These theories can be further categorized into two main camps – physiological-based theories and psychological-based theories. Physiological-based theories attribute harmony perceptions to the functions of auditory pathway organs, therefore the consonance and dissonance perception responses are preliminarily determined by the acoustic features of the input sound stimuli. Psychological-based theories on the other hand believe the musical perception of sound entities are not physiologically determined, but interpreted through brain/ psychological activities. In this chapter, relevant theories pertaining to physiological and psychological based theories are reviewed in section 3.1 and 3.2 respectively. In section 3.3, these two camps of theories are compared and justified to provide a more comprehensive view for music harmony perception.

3.1 Physiological-Based Theories

Physiological theories pertaining to music harmony perception are developed based on the functions of our auditory 'hardware' system – the auditory pathway. Human auditory pathway consists of three sub-systems: peripheral auditory system (ear), central auditory system, and brain.

Our ear is the sound sensing system; anatomically, it consists of outer ear, middle ear and inner ear (see Fig11.). The shape of the outer ear can help to gather acoustic energy and capture the main features of the sound source. The acoustic signal is further resonated and transmitted through the ear canal and reaches the eardrum. The eardrum is the most important organ of the middle ear; its main function is to amplify the acoustic signal for the inner ear functions. The core functioning organ of the inner ear is the cochlea; it essentially converts acoustical signals into electrochemical impulses which are further converted into neural signals.

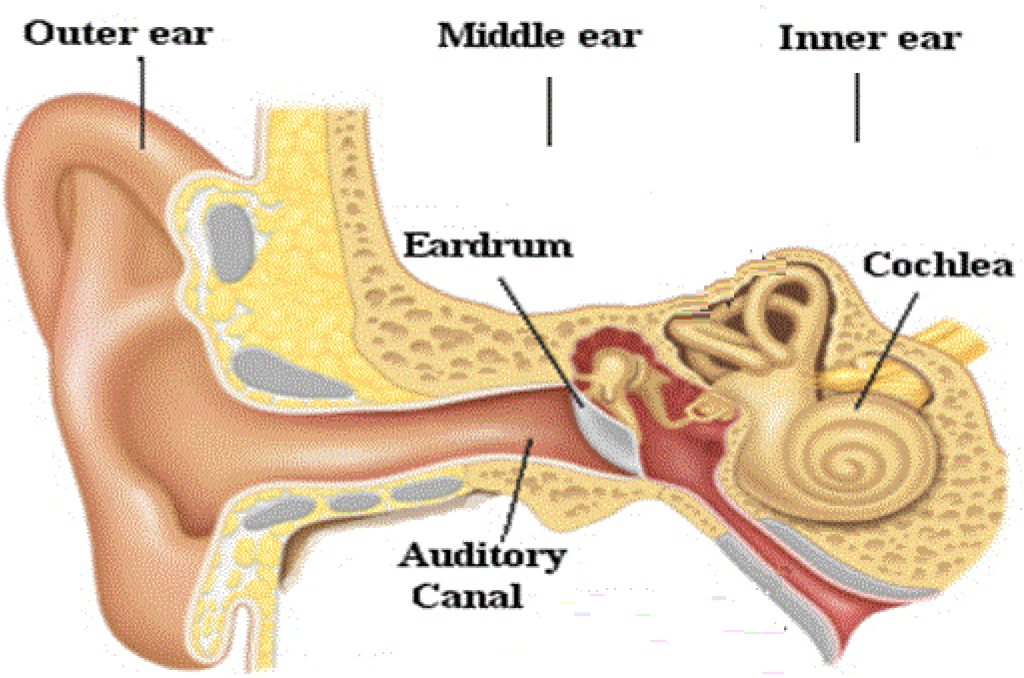


Figure 11 Anatomical structure of ear

This figure is produced from (Lawrence and Yantis, 1956)

The central auditory system transmits and processes the auditory bioelectrical signal and delivers it to the auditory cortex. The relay organs along the auditory pathway include the *Cochlear nucleus, Trapezoid body, Superior olivary complex, Lateral lemniscus, Inferior colliculi, and Medial geniculate nucleus* (Fig.12). The functional responses in term of neural signals at the central auditory system are complex, and many crossing and feedback loops are observed within and between intermediate organs. The idea of building a precise functional model of the central auditory system is fundamentally hindered by the complexity of neural structure.

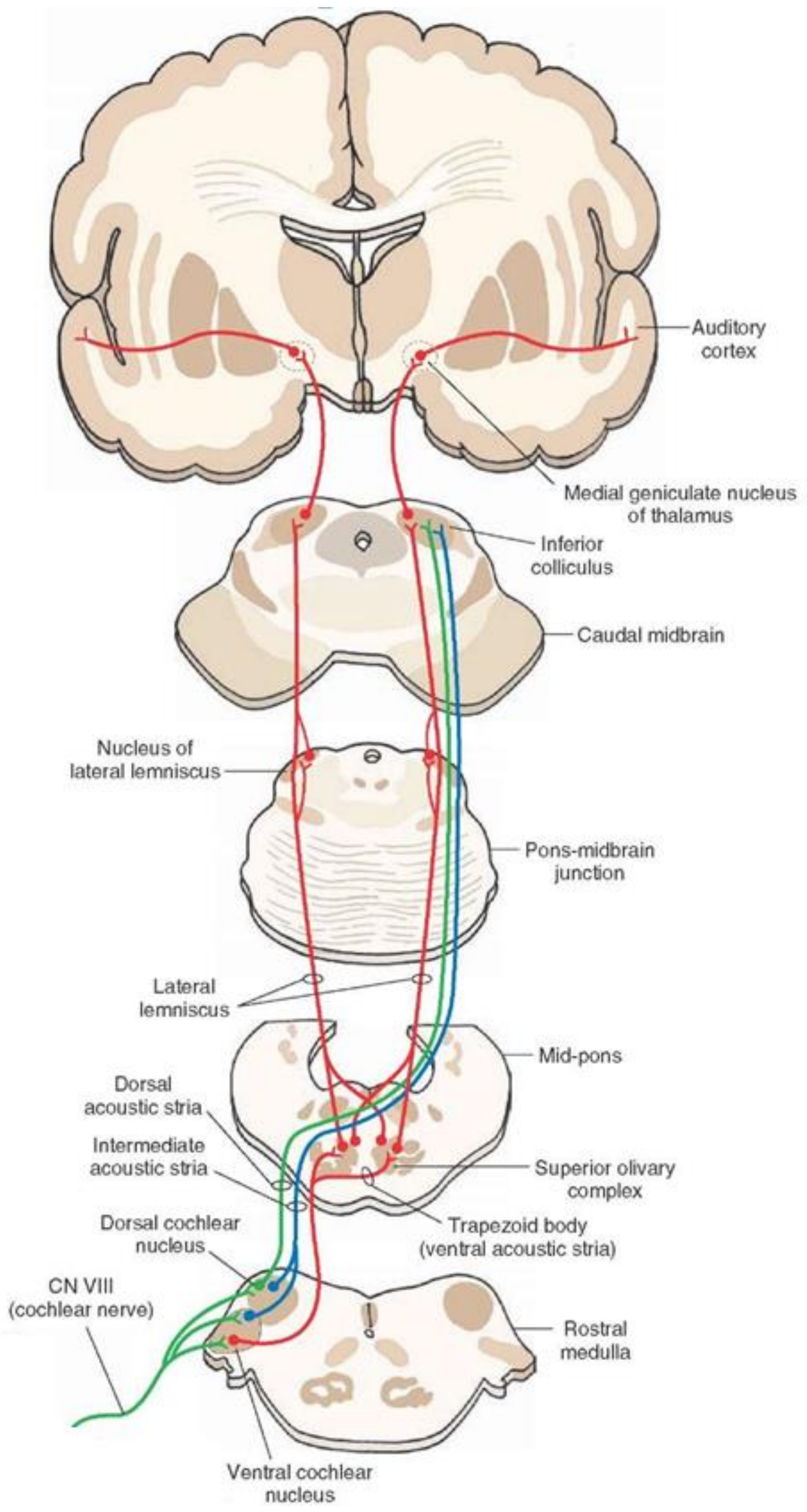


Figure 12 Central Auditory System

This figure is taken from (Patterson et al., 1992)

The brain system under physiological-based theories simply refers to the auditory cortex. The auditory cortex receives neural signals and interprets it as ‘auditory sensations’. For the perception of music harmony, auditory sensations are directly linked to the consonance or dissonance concepts. Brain activity is typically monitored with the techniques of functional magnetic resonance imaging (fMRI[12]), positron emission tomography (PET[13]), transcranial magnetic stimulation (TMS[14]), magneto-encephalography (MEG[15]), and electroencephalography[16] (EEG). However, data obtained this way is only parametric; it needs to be further explained with human descriptors.

A comprehensive physiological-based theory requires the functional response of each auditory pathway organ to be clearly understood. However, even with state-of-the-art medicine, performing neuron-level measurements along the auditory nerve is still quite an intractable task (Patterson et al., 1992). Instead of understanding the physiological functions of the entire auditory pathway organs, current psychoacoustic theories typically focus on the functional response of a particular part along the auditory pathway and hypothesize it to be the major organ that causes the polarization of consonance and dissonance sensations. In literature, *theory of beats* and *theory of harmonic-template* are two of the most popular physiological-based theories. The former attribute dissonance sensations to inner ear functions, and the latter is underpinned by the central auditory system functions. A more detailed review of *theory of beats* and *theory of harmonic-template* is presented in next sections (section 3.1.1 and 3.1.2 respectively).

[12] Functional magnetic resonance imaging or functional MRI (fMRI) is a using MRI to measures brain activity by tracing the dynamic changes of the blood flows.

[13] Positron emission tomography (PET) is a technique that make use of the radioactive chemicals to trace the organ activities. .

[14] Transcranial magnetic stimulation (TMS) is a technique to use generated magnetic field to stimulate a small regions of brain and to observe relevant reactions.

[15] Magnetoencephalography (MEG) is able to investigate brain activity per milliseconds. It is achieved mainly by detecting the natural brain electrical current.

[16] An electroencephalogram (EEG) detects electrical activity using electrodes.

3.1.1 The Theory of Beats

According to the place theory of pitch perception, each audible frequency component excites a specific 'place' along the basilar membrane. Following such a thread, when two frequency components are present simultaneously, (ignoring the masking effect) we may expect to hear two isolated pitches. However, in reality, this is not how the human hearing system works. Helmholtz exemplified the perception responses for two sine-wave tones separated at different frequency intervals, from which he observed that: when the separations are close enough, a volume fluctuation effect is observed which is known as the beating effect (Cross and Goodwin, 1893). When the separations were increased further to 30–40Hz, a 'harsh, rattling' sound is heard, also known as the *roughness* sensation or the rapid beating effect; further separations will decrease the roughness until the tones can be heard as 'clearly two' (see Fig.13).

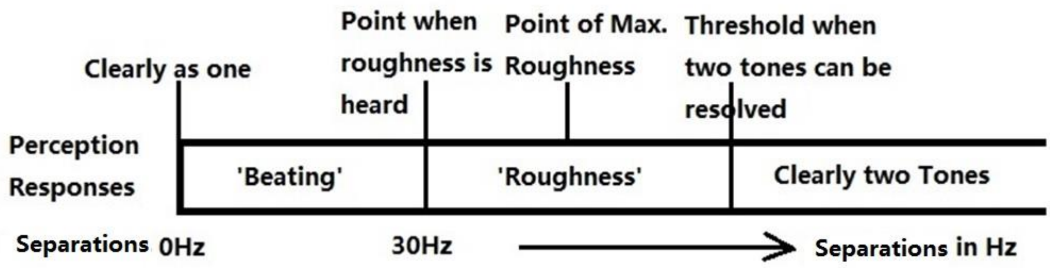


Figure 13 Illustration of Helmholtz's pure-tone interferences

Helmholtz interprets beating as a result of temporal interferences, with the beating frequencies and patterns identical to the envelope of the combined waveforms. More importantly, he also hypothesised that the roughness sensation is essentially the 'true cause' of musical dissonance, whereas the consonance concept is the relative lack of the roughness.

Helmholtz's hypothesis of dissonance was later supported by (Greenwood, 1961) who further linked it with the concept of *critical bandwidth* (CB). Critical bandwidth is a concept introduced by American Physicist H. Fletcher in 1940. It refers to the tonotopical limitation (see Fig.14) of the cochlea that within a certain frequency bandwidth (a *critical bandwidth)*, two pure-tone components interfere with each other producing a non-linear effect such as mutual masking effects. Greenwood believes the (rapid) beating sensation is another result of pure-tone interferences that happen within a *critical bandwidth.* When the frequency separation of two simultaneous pure tones exceeds one *critical bandwidth*, the interference effect is eliminated and two clear pitches are perceived.

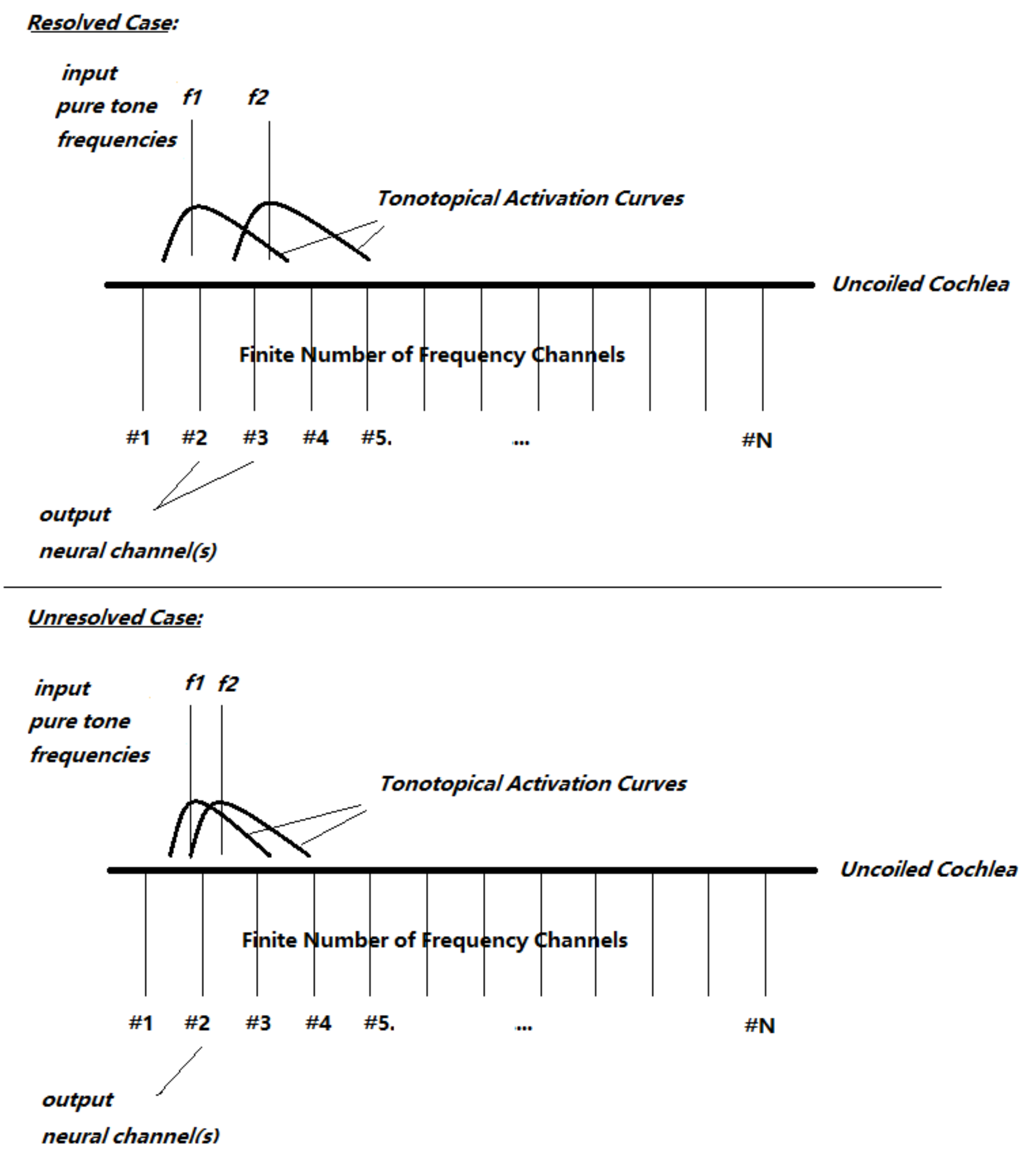


Figure 14 A diagram showing the frequency limitation of inner ear

The upper diagram demonstrates two pure tone partials f1 and f2 are resolved and coded into two neural channels; and the lower diagram demonstrates two pure tone partials f1 and f2 cannot be resolved and coded into two neural channels, this is when mutual masking effects and roughness sensations generally occur.

As most musical tones are complex tones, the overall dissonance of musical tone(s) can be estimated by the exhaustive sum of quantified *roughness* effect between each of the possible pure-tone pairs within that sound aggregate. Following this line of thinking, a major success was achieved in accounting for the consonance and dissonance perceptions of musical intervals in line with *proportion theories* (see section 2.3). In particular, the beats theory has also achieved a notable success in accounting for the unconditional dissonance introduced by the simultaneous dyads separated at one or two semitones. The dissonance sensations associated under the roughness concept is therefore also known as *sensory* dissonance (Terhadt, 1979).

The theory of beats prevailed in the 1970s but it has been questioned in more recent research.

One problem with the beats-based theory is the assumption that beats are a result of critical band. According to Greenwood, zero beats should be perceived when two pure tones are separated beyond the critical bandwidth. However, empirical studies (Hindemith, 1984 and Benade, 1976) have shown that beats also occur when the two pure-tone partials have a frequency ratio approaching (but not equal) to *m: n*, where *m* and *n* are positive integers (such effects are also known as the *secondary beats* effect). Such empirical observations are therefore contradicted to Greenwoods' hypothesis. In chapter 5, a theoretical attempt is made to incorporate *secondary beats* effect into the estimation of the total sensory dissonance effect.

Moreover, the critical band cannot be used to explain *binaural beats*. The concept of *binaural beats* refers to the phenomenon that when one pure-tone is played to the left ear, and the other to the right ear, beats can still be perceived. As the two pure tones clearly belong to two critical bands (one of the left inner ear and one of the right), theoretically zero sensory dissonance is expected. Helmholtz also noticed such a problem and he attempted to account for secondary beats with a non-linear inner ear transfer function (Plomp and Mimpen, 1968). Nevertheless, recent research (Wright, 1986) has demonstrated that the perception of beats must involve higher-level functions in the central auditory system (CAS) and brain, and roughness sensations cannot be interpreted by the inner ear function.

The second problem is the method of estimating the overall dissonance for complex tone(s). As musical tones consist of a number of pure-tone partials, in order to estimate the overall dissonance, we have to consider the total amount of beats within the sound aggregate. However, it is extremely difficult to quantify and estimate the beating effect. The amount of beats introduced by a particular pure-tone pair may be masked by the beats introduced by other pure-tone pairs. The amplitude, numbers, as well as the spacing of pure-tone partials will also have a complex role in the estimation of overall beats which cannot be modelled using a simple linear summation algorithm.

Lastly, although the computational model based on the theory of beats has been generally successful to account for musical dyads, none of the models is able to produce a consistent result for musical triads, which (Cook, 2006) concluded as “a complete failure thus far to account for the core phenomena of diatonic harmony on psychophysical principles.”

Beyond all doubt, the dissonance effect introduced by auditory beats (sensory beats) is evidential. However, the *theory of beats* does not consider any musical context (such as musical key, root of chord, preceding chords and so on), which has been quite important in music harmony. Thus more complex physiological and psychological process must be involved in music harmony perception. In the next section, another theoretical thread is introduced pertaining to the sensory response of music consonance and dissonance.

3.1.2 Theory of Harmonic Template

Harmonic-template theories were inspired by the *harmonic relation theory* (see section 2.5 proposed in Rameau’s seminal work early in 1721). Following Rameau’s hypothesis, Terhardt proposed an idea linking musical consonance and dissonance to our pitch perception mechanisms. His theory hypothesizing that musical consonance and dissonance are related to the effort our brain makes in order to recognise a *pitch centre* of the input sonority: more effort means more dissonance. The ‘least effort’ happens when the input sonority has harmonic spectral structure, therefore an acoustic harmonic-template is deemed as an acoustic template of consonance (Terhardt, 1979).

On the question of why our auditory system is inherently adept at resolving harmonic sound spectra but not others, Shamma and Klein introduced a physiological model that simulates the process of how harmonic templates can be formed by the exposure of random noise over a certain time period (Shamma and Klein, 2000) (see Fig.15).

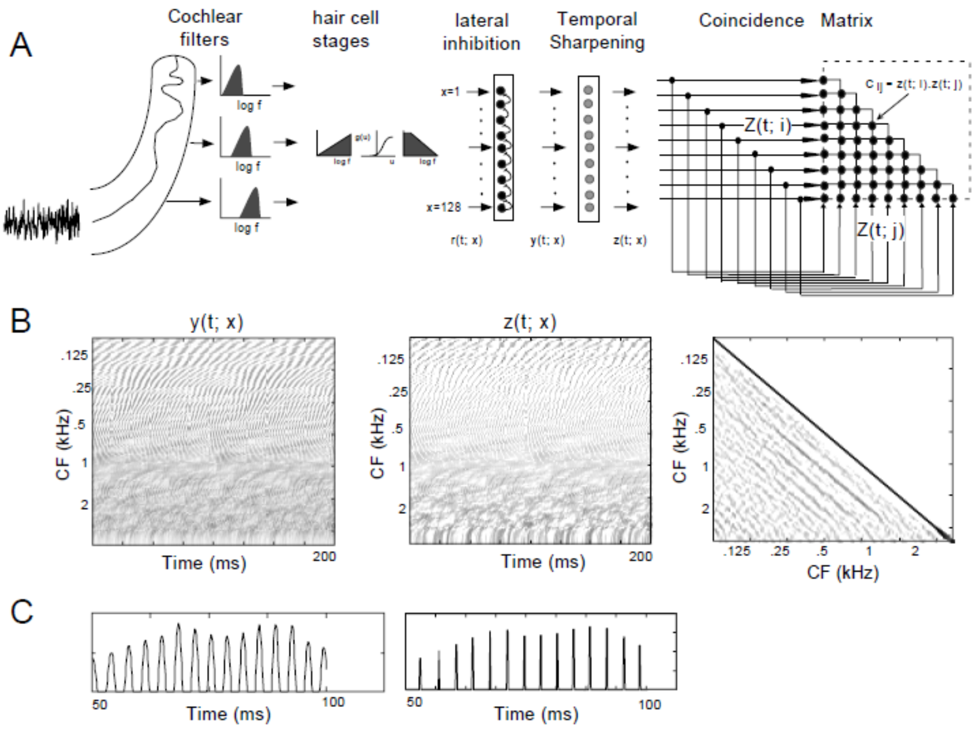


Figure 15 Diagram of noise training of harmonic-template (Shamma and Klein, 2000)

In above figure, part *A* illustrates random noise input is put through a bank of 128 cochlear filters between 100Hz–4kHz; the output waveform of each filter is then passed through a hair cell model with the two main functions of cochlear filtering: lateral inhibition and temporal sharpening. A coincidence between all cochlear filters is then calculated. In part *B*: the spatiotemporal responses of the channel array (from left to right) after lateral inhibition, temporal sharpening and coincide matrix. In part *C*: The integration waveform for a particular critical band (from left to right) after lateral inhibition, temporal sharpening. As we can see, a harmonic structure is observed after temporal sharpening.

Under harmonic-template theory, consonance and dissonance perception is essentially a pattern matching process: when the input sonority has a spectral template similar to harmonic-template spectra, a 'sensory consonance' concept will be perceived; and vice versa. In general, musical intervals with simple, integer frequency ratios can be viewed as the overtone partials of a specific (possibly virtual) fundamental partial, therefore the harmonic-template theory can be used to interpret the *proportion theories* summarised in section 2.6. Meanwhile, the harmonic-template theory is inherently consistent with the *harmonic relation theory*; therefore, two of the main observations (*proportion theories and harmonic relation theory*) can be generally interpreted by the harmonic-template theory.

Resnick took another approach: he measured the average time that a listener spent in trying to determine the pitch of a sound stimulus (Resnick, 1981). Resnick hypothesized a dissonance concept, the *pitch resolution dissonance,* to those sounds that required a longer time for pitch recognition. He also found that for harmonic tones the time taken is relatively fast — a result that is very much in line with harmonic-template theory.

In contrast to frequency-domain analysis, the harmonic-template in the time-domain involves the periodicity detection of a waveform. Therefore, instead of harmonic pattern matching, periodicity detection can also be used for the analysis of musical consonance and dissonance. For example, F. Stoltenberg proposed a concept of *relative periodicity* where the lowest tone of a triad is used as the reference and from which the perceived level of consonance is determined. The details of his numerical approach will be reviewed in section 4.4.3.

3.2 Psychology-Based Theories

The perception domain concepts (such as emotions, moods, affects, etc.) can be very obscure and confused, as many terms are used interchangeably in daily life. However, they need to be clearly defined in order to provide a common ground for scientific discussions. One way to organize the perception domain descriptors is according to the psychological stage of the perception mechanism. Like general perception models, psychological perception of music harmony can be broken into two layers: *lower-level psychological affects* and *higher–level psychological experiences* (Resnick, 1981).

The *psychological affect* is the subject description of auditory sensations, it refers to our innate ability to recognize and perceive sound stimuli. The perception behaviours that happen at this layer should be universally applicable to anyone. The term *psychological affect* is the subjective counterpart of the physiological *sensations*; and it is considered to be the non-conscious psychological experience. Language-wise, there are many adjectives describing the sense of *psychological affects*; nevertheless, they can be classified according to the three principal dimensions: *valence, arousal,* and *motivational intensity. Valance* refers to the positive-to-negative property of the affect; *Arousal* corresponds to the level of excitement; and *motivational intensity* concerns the urge to act (differing from the arousal concept which requires no action implication). Fig.16 shows a graph demonstrating some of the common affection descriptors under the first two dimensions (*valence* and *arousal*).

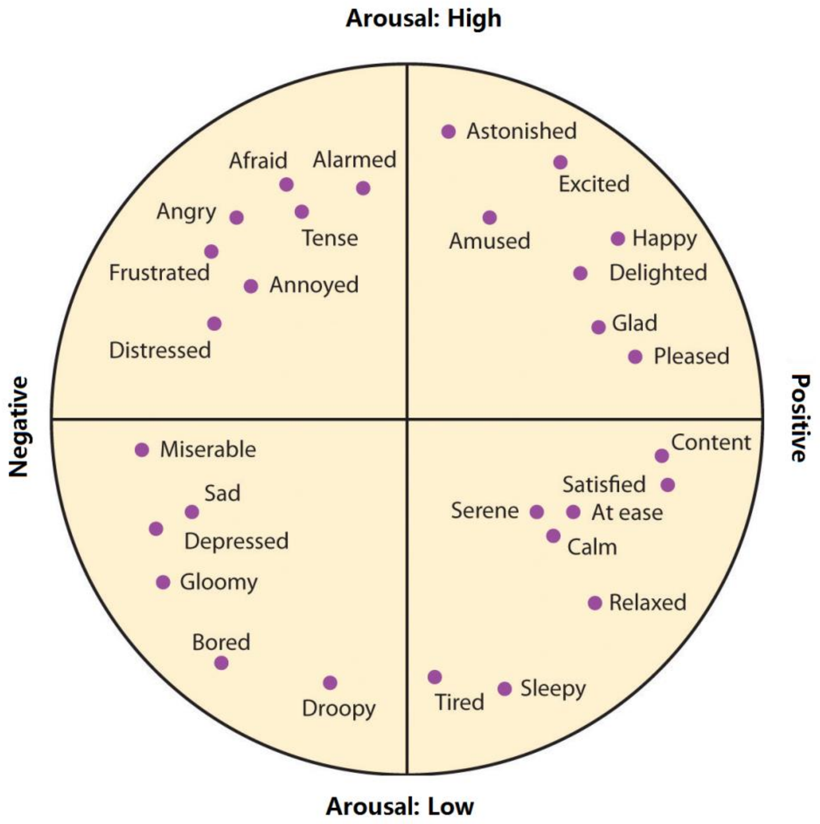


Figure 16 Distribution plot of the psychological affects (Gable and Harmon-Jones, 2013)

The *higher psychological layer* refers to a higher-level perception mechanism where more complicated psychological activities are involved. A key feature, and one that distinguishes the *higher psychological layer* from the *affect* layer, is the interaction between concurrent *psychological affects* with past experiences, memories or emotional states. There are two types of higher-level psychological experiences: the first one is based on *cognitive features*, corresponding to the rational thought of *psychological affects*; and the second one is associated to personal emotional state, corresponding to the *psychological feelings* evoked by *psychological affects*.

Cognition is a process which turns the *psychological affect* into ‘something that we know’ (*information*): it compares the present *affect* with past experience; a pattern recognition mechanism is typically involved. For instance, when we are trying to figure out the name of the music interval that has been just played, our brain compares the sound stimuli with our past experience of musical intervals; if a good match is found, we are able to tell the name of that interval (Resnick, 1981). Most psychological-based theories are proposed based on cognitive features, on the hypothesis that certain cognitive features are commonly recognized as consonance (or dissonance) concepts.

Feelings can be viewed as 'personalized' psychological *affects.* Depending on the emotional state and personal experiences, the same auditory sensation/ *psychological affects* can result in different kinds of *feelings*. For instance, a particular song to someone may sound happy and sweet when he/she is with his/her loved ones; but becomes sad and heart-breaking if the lover is no longer there. As *psychological feeling*s are highly dependent on personal and cultural influences, therefore musical consonance and dissonance based on this concept should be excluded from the study of music harmony conventions.

Psychological *feelings* should not be confused with *emotions, mood* and *music aesthetics* (Duncan and Barret, 2007). The psychological term *emotion* is the outward manifestation of *feelings.* Psychological *emotion* can be either genuine or feigned: one may pretend to behave against their inner *feelings*. The psychological term *Mood* is similar to *emotions,* but tends to be more unfocused and diffused (Gabrielsson and Justin, 2003) and is generally not specific to any sound stimulus but associated with a period of emotional state. *Music aesthetics* is a more complicated concept, which involves both music *cognition* and *feelings.* For example, a certain piece of music may evoke a certain period of life in the past which we may miss, cherish, or regret. The whole picture of music perception mechanism can be illustrated in Fig.17:

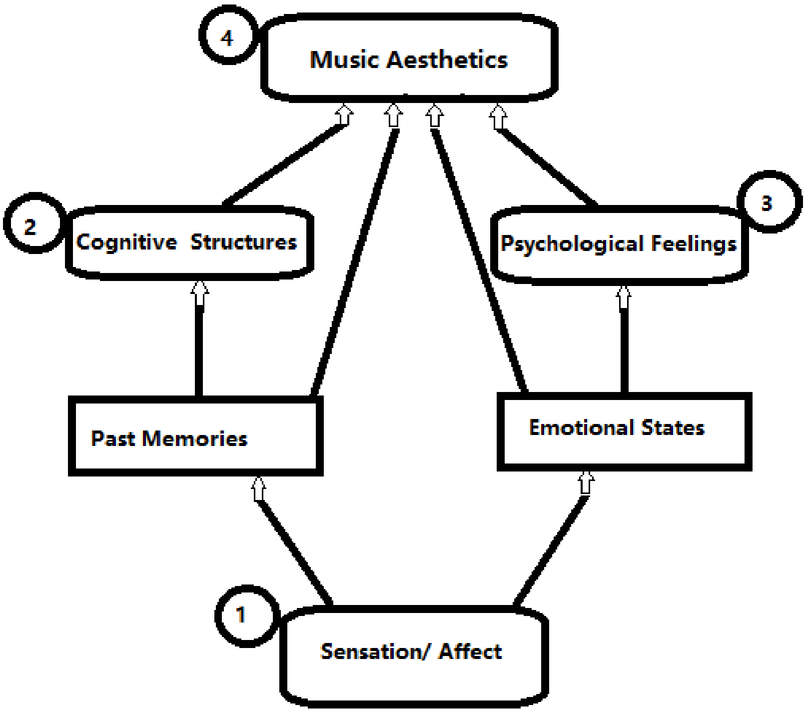


Figure 17 Psychological meanings of consonance and dissonance related concepts

In this diagram, all the objects (①②③④) may relate to description of musical consonance and dissonance, but only object ② is studied under psychological-based theories.

In the literature, many psychological-based theories have been proposed pertaining to music harmony perceptions, such as Krumhansl's *tonal harmonic hierarchy theory* (Bharucha and Krumhansl, 1983), Cook's *intervallic equivalence theory* (Cook, 2006), *tonal fusion/numerosity conjecture theory* (Schneider, 1898; Huron, 1991) and *melodic expectation theory* (Margulis, 2007). These theories are however referring to the perception of different musical objects: *tonal harmonic hierarchy* refers to the perception of a particular tone/ chord within a tonal context; *intervallic equivalence theory* and *tonal fusion/numerosity conjecture* theory refers to the perception of an isolated sound object; and *melodic expectation theory* discusses the perception of motions between musical entities. The following sections (sect.3.2.1–3.2.4) will review these theories in more details.

3.2.1 Tonal Harmonic Hierarchy

Under *Tonal harmonic hierarchy theory* (Bharucha, 1983), the consonance and dissonance perception of a musical note/ chord has to refer to its tonal context (tonic key). In reference to the tonic key, a musical note/ chord is noted by its intervallic relationships (scale degrees) rather than frequency ratios. *Tonal harmonic hierarchy theory* essentially categorizes musical scales into five hierarchical levels (see Fig.18).

On the top level is tonic pitch class; on the second level is the fifth scale degree; the 'triadic' level refers to the third scale degree appeared in major/minor triads; the 'diatonic' level refers to other diatonic scale degrees, these scale degrees are either melodically or harmonically related to the tonic key (see section 2.2); Remaining scale degrees and micro scale degrees are grouped at the 'chromatic' level, these correspond to the highest dissonance concept under *tonal harmonic hierarchy*.

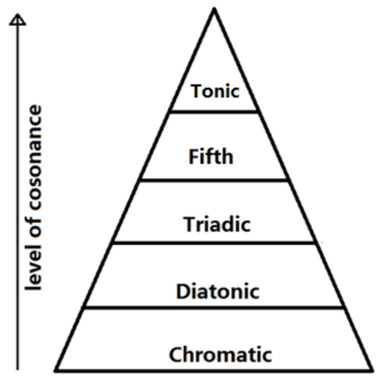


Figure 18 Hierarchical structure of tonal consonance (Bigand et al., 1988)

Tonal harmonic hierarchy has been studied and verified by a number of empirical tests, such as (Dowling & Harwood, 1986) and (Krumhansl, 1997). It is shown that even inexperienced subjects were able to perceive a five-level harmonic structure under the description of *tonal harmonic hierarchy*. Comparing to *Pythagorean* classification of musical intervals (see section 2.2), a few number of inconsistencies can be noticed (See Table 6 below):

Pythagorean	**Diatonic Name**	**Tonal Hierarchy**
equal	Unison	Tonic
multiple	Octave	
epimore	Perfect Fifth	Fifth
	Perfect Fourth	Diatonic
	Major Third	Triadic
	Minor Third	
epimere	Major Sixth	Diatonic
	Minor Sixth	
multiple epimore	≈Minor Seventh	
multiple epimere	≈Major Seventh	

Table 6 Pythagorean classifications of musical intervals and tonal harmonic hierarchy

First of all, tonal hierarchy has an extra category (chromatic level) dedicated to general dissonance. Secondly, the fourth scale degree (on the *triadic level)* is considered to be less consonant than the third scale degree (on the *diatonic level)* under *tonal harmonic hierarchy* whereas under *Pythagorean* classification method, fourth interval is more consonant than the third intervals. Such inconsistencies are also observed by Tenney in the literature review of consonance and dissonance concepts (*CDC*, see section 2.2). A possible explanation for this inconsistency is that *Pythagorean* classification method only considers the isolated mutual relationships between two musical tones without considering tonal context; therefore, the fourth interval is more consonant than the third interval. Under tonal hierarchy structure, third scale degree is closer related to tonic key than the fourth scale degree, therefore it is more consonant.

Tonal harmonic hierarchy theory is proposed base on the tonal functions of musical entities. The layers in *tonal harmonic hierarchy* are highly associated to the perception of tonal *CDC*. The proposed tonal structure is highly in line with the empirical observations in western tonal music.

3.2.2 Intervallic Equivalence Theory

The *intervallic equivalence theory* proposed by (Cook, 2006) focused on the consonance and dissonance perception for musical triadic structures. Under *intervallic equivalence* theory, a triad consisting of two equal intervals, such as diminished (2 × three semitone intervals) or augmented (2 × four semitone intervals) has a higher salience of tension/dissonance. Unequal intervallic structures (such as major and minor triads consisting of one interval of three semitones and one interval of four semitones) on the other hand correspond to a more consonant concept (see Fig.19). Cook’s model prediction is successful in accounting for the perceived consonance and dissonance levels of most musical triads. However, the *intervallic equivalence* theory does not explain why augmented triads sound more dissonant than the diminished triads as they both have intervallic equivalence; neither can it explain why major triads are more consonant than minor as they are all one semitone shift away from intervallic equivalence.

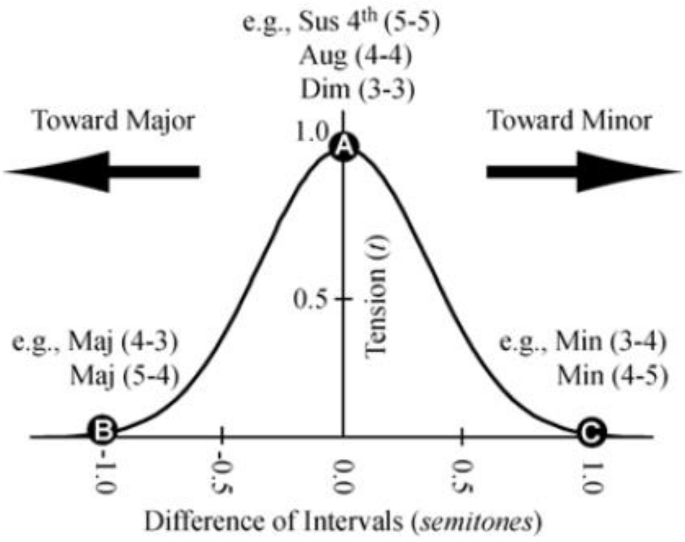


Figure 19 Intervallic equivalence model (taken from Cook, 2006)

In Figure 19, the difference of interval is calculated by subtracting lower-to-middle-tone interval from middle-to-higher-tone interval under triadic structure. For instance, for a minor triad, the 'lower-to-middle-tone interval' is 3 semitones, and the middle-to-higher-tone interval is 4 semitones, therefore the difference of interval is 1 semitone. According to Fig.18; this is a consonant structure. For a diminished triad, the difference of interval is zero (3 – 3 semitones), therefore result is in tension. The dissonance structures according to intervallic equivalence theory are therefore suspended 4th, augmented and diminished structures

Cook's psychophysical model is not built on the interaction of cognitive features such as musical note to compute the perceived level of 'tension'. Instead, unit 'tension' is given to three frequency partials with equal frequency intervals. The total 'instability' of the input sound stimuli is estimated with a linear summation function akin to the estimation of overall sensory roughness (see section 4.3.2). The theoretical challenge of using a frequency-based numerical method is whether the 'tension' concept under *intervallic equivalence theory* can be linearly summed in order to estimate the overall 'tension' effect.

3.2.3 Tonal Fusion / Numerosity Conjecture Theory

The *tonal fusion theory* was proposed by Carl Stumpf early in 1898. In his view, consonance and dissonance perceptions are closely related to how well a number of tones are 'fused' as one. This is somewhat similar to the *harmonic relation theory*, as pitch sets in harmonic relations would be merged to create a single complex tone sensation that defines the tonal fusion concept. For example, 200Hz, 300Hz, and 400Hz pure tones when played together would be perceived as a 100Hz musical (complex) tone, but the 200Hz 300Hz, 400Hz partials will not be noticed individually. Albert Bregman drew attention to the past confusion between sensory consonance ('smooth sounding') and tonal fusion ('sounding as one') (Bregman, 1994). A notable affirmation of this distinction is found in (Huron, 1991) who carried out a statistical analysis of a sample of music by J. S. Bach who attempted to avoid tonal fusion while pursuing tonal consonance.

In another perception test (Huron, 1993), listeners were presented with a complex tone consisting of 10 decaying harmonics with 100Hz fundamental frequency and another dyad of pure tones with frequencies of 200Hz and 300Hz. Most listeners tended to judge the second dyad as less fused than the first complex tone, yet most listeners consider the dyad as sounding more 'pleasant', 'euphonious' or 'consonant' than the complex tone. All this evidence indicated that consonance and dissonance are not directly related (but contradict) to the tonal fusion concept. Conversely, Huron proposed a *numerosity conjecture* theory (Huron, 1996) in which the perceived level of consonance is highly proportional to the perceived number of tones. Huron considers the diversified pitch classes within a chord or piece of music as the 'added colour or warmth' to avoid the dullness of music harmony. The concept of dullness, however, requires a past experience of tonal consonance. The *numerosity conjecture* theory thus can be viewed as a higher-level cognitive behaviour.

3.2.4 Melodic Expectation Theory

Under *melodic expectation theory*, *CDC* are used to describe the motions from one musical entity to another. The premise of this theory is that listeners are able to have predictive expectation on next-coming musical note based on current and previous melodic lines or musical notes. When the listener's expectation is realized, melodic consonance is perceived. In the literature, many theories have been proposed to model listeners' expectation behaviour, such as *Schenkerian analysis* (Salzer, 1952) and the 'five principles' under Narmour's implication-realization model (Narmour, 1977).

More recent research pertaining to the principles of melodic expectation is presented by (Margulis, 2007). The following conclusions can be made from Margulis' research experiments:

(1) Predictive features can be pitch, tempo or loudness;

(2) music variables exceeding such expectation (such as higher pitch, ascending conjunct motions [17], or higher loudness level) generally introduce psychological tension;

(3) when music variables are under melodic expectation (such as lower pitch, descending conjunct motions, or lower loudness level), emotions such as gloom or sadness can be perceived;

(4) the melodic progression from higher level sound features to the 'expected' level leads to an emotion of 'relaxation'; and

(5) the melodic progression from lower level sound features to the 'expected' level invokes a psychological experience of 'happiness' and 'pleasantness' (gaiety emotion).

The relationships between gaiety & relaxation and gloom & tension emotions are illustrated in Fig.20.

[17] Conjunct motions refer to step-wise melodic motions, typically with a step-size of one or two semitones. This term contrasts with disjunct melodic motions, also known as skip-wise motions, with an intervallic jump of more than two semitones.

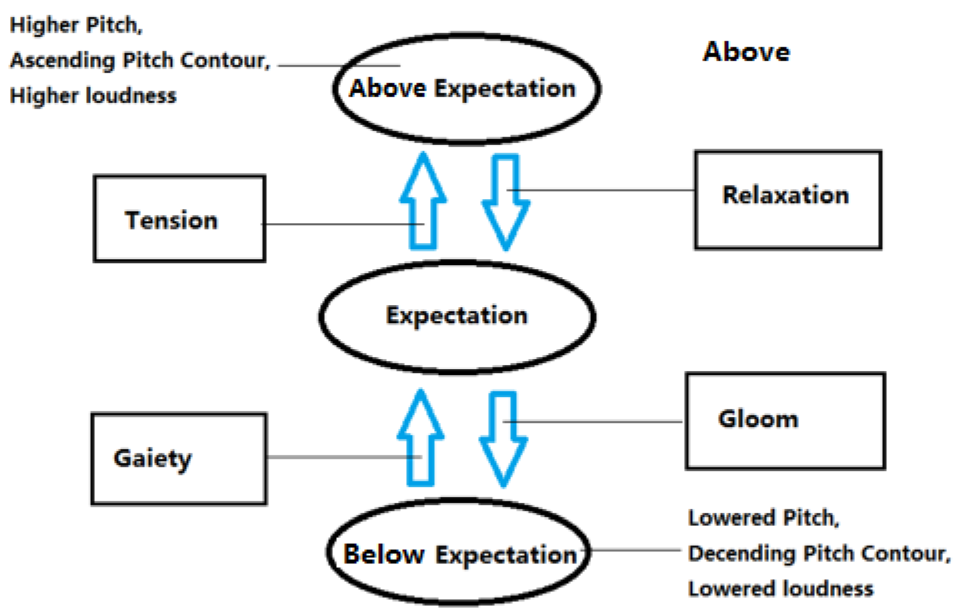


Figure 20 Music expectation and two-dimensional music emotions.

In Fig.20, both ‘gaiety’ and ‘relaxation’ correspond to a musical consonance concept, and ‘tension’ and ‘gloom’ correspond to musical dissonance concepts. According to the study of music emotions (Juslin and Sloboda, 2010), ‘gloom’ and ‘tension’ are however two uncorrelated dissonance concepts. In this research, these two concepts are modelled separately in Chapter 7.

3.3 Between Physiological and Psychological Theories

Comparing physiological and psychological based theories, the following conclusions can be made:

(1) Physiological based consonance and dissonance perceptions are underpinned by the biological functions of the auditory pathway; Psychological based consonance and dissonance theories are proposed based on common psychological principles (such as Gestalt psychology, tonal structures, and melodic expectations).

(2) Physiological based harmonic analysis is based on the principles (hypothesis) of pure-tone interactions; Psychological based harmonic analysis is based on the principles (hypothesis) of pitch interactions.

(3) Physiological based CDC are also known as the sensory response; Psychological based CDC are usually referred as the cognitive responses.

(4) Physiological theories typically study the perception of isolated sound objects; psychological based theories typically study the perception of a particular entity under musical context.

(5) Under physiological based theories, the perception of a sound object is absolute and fixed (determined by the acoustic property of the sound object); under psychological based theories, the meaning of each sound object is not fixed, but depends on the presence of other cognitive features.

The key differences between physiological and psychological based theories are highlighted in Table 7.

	Psychological Theories	Psychological Theories
'hardware'	auditory pathway	brain
input features	time/frequency features	pitch, musical notes
output response	sensory response	cognitive response
meaning of *CDC*	absolute and fixed	depend on other cognitive features

Table 7 Physiological and psychological based theories

Based on the perception mechanism, theories reviewed in this chapter can be classified into three main categories:

The first category is the *theory of beats*. The *theory of beats* analyses acoustic signals and predicts sensory *roughness* based on the physiological function of inner ear organs. *Roughness* sensation is a unique dissonance concept occurring at the lower physiological level, and *theory of beats* can thus be distinguished from any other theories.

The second category involves a discussion of the *tonal centre concept*. This including the *harmonic-template based theories, tonal fusion theory* and *tonal harmonic hierarchy theory*. A common feature among these theories is that they involve the discussion of *tonal centre concept*:

> Under *harmonic-template based theories*, acoustic partials contributing to a clearer pitch sensation (*tonal centre concept* [18]) are associated to the consonance concept;

[18] Recall under the three-layer system of musical entities (section 2.6), *tonal centre concept* can refer to the pitch of a tone, root of chord, and tonic of a musical mode.

Under *tonal fusion theory*, musical structures enhancing a clear root / pitch perception (*tonal centre concept* [17]) are considered to have consonance property;

Under *tonal harmonic hierarchy theory*, note/ chord components are close related to tonic key (*tonal centre concept*[17]) are located at higher level of tonal hierarchy.

Thus, it is generally observed that: tonal consonance concept is usually associated to the acoustical or musical entities those are highly concordant with the perception of *tonal centre concept*; and tonal dissonance concepts on the other hand, usually against with the perception of *tonal centre concept.*

The third type of theory *is melodic expectation theory*. *Melodic expectation theory* is associated to other cognitive features such as melodic contours. And it is considered as one of unique type of psychological mechanism.

Chapter 4 Psychoacoustic Models

The psychoacoustic model is one of the numerical approaches to implement psychoacoustic theories. The challenges of building a proper psychoacoustic model are the quantization method for the input (physical) and output (psychological) features as well as developing computational algorithms based on theoretical hypothesis. The model predictions are expected to account for the empirical observations.

In section 4.1, typical methods used to convert the input features – the musical entities into acoustic measurable are reviewed, and in section 4.2, the methods used to describe the output features – the perceived musical consonance and dissonance are introduced. Conventional psychoacoustic models have been focused on the *sensory* aspect of music harmony only, therefore section 4.3 and 4.4 review the computational algorithms involved in two of the most prominent physiological based theories: the theory of beats (see section 3.1.1) and harmonic templates theories (see section 3.1.2) respectively. In section 4.5, *the theory of beats* is compared with *theory of harmonic template*. And section 4.6 concludes this chapter.

4.1 Acoustic Features

As summarized in section 3.3, the musical note is one of the most fundamental concepts in tonal music. The use of notes constitutes most of musical harmony's related concepts, such as intervals, chords, and tonal systems. For pure tone partials, there is a one-to-one relationship between notes and acoustic frequencies, and this is reviewed in section 4.1.1. For musical tones (complex tones), frequency analyses are involved, and this is reviewed in section 4.1.2.

4.1.1 Musical Notes and Frequencies

The pitch perception of a pure tone (sinusoidal wave) is related to the fundamental vibration frequency of its waveform. Thus for pure tones, the pitch sensation of musical notes can be measured on a one-dimensional frequency scale from 'high' to 'low'.

In music tuning system, the A_4 note is standardised at 440Hz under **ISO 16** (International Standard Organization: https://www.iso.org/standard/3601.html). The frequencies of other musical notes can be calculated by:

$$440 \times 2^{n/m}$$

Where m=12 under 12-tet equal-temperament, and n refers to the difference (in semitones) from A_4. The musical note to frequency conversion can be viewed in Table 8.

Octave	Frequency (Hz)											
	C	C#	D	D#	E	F	F#	G	G#	A	A#	B
0	16.35	17.32	18.35	19.45	20.6	21.83	23.12	24.5	25.96	27.5	29.14	30.87
1	32.7	34.65	36.71	38.89	41.2	43.65	46.25	49	51.91	55	58.27	61.74
2	65.41	69.3	73.42	77.78	82.41	87.31	92.5	98	103.8	110	116.5	123.5
3	130.81	138.6	146.8	155.56	164.8	174.6	185	196	207.7	220	233.1	246.9
4	261.63	277.2	293.7	311.13	329.6	349.2	370	392	415.3	440	466.2	493.9
5	523.25	554.4	587.3	622.25	659.3	698.5	740	784	830.6	880	932.3	987.8
6	1046.5	1109	1175	1244.5	1319	1397	1480	1568	1661	1760	1865	1976
7	2093	2217	2349	2489	2637	2794	2960	3136	3322	3520	3729	3951
8	4186	4435	4699	4978	5274	5588	5920	6272	6645	7040	7459	7902

Table 8 Musical note to frequency conversion table

Other than using alphabetic symbols (A–G), musical notes can also be represented with pure numerical notations. Under psychoacoustic modelling, using numerical representations as input can be beneficial to the model computations. MIDI note numbers are one of the standard numerical representations of musical notes (Table 9).

Octave	Midi Note Numbers											
	C	C#	D	D#	E	F	F#	G	G#	A	A#	B
0	0	1	2	3	4	5	6	7	8	9	10	11
1	12	13	14	15	16	17	18	19	20	21	22	23
2	24	25	26	27	28	29	30	31	32	33	34	35
3	36	37	38	39	40	41	42	43	44	45	46	47
4	48	49	50	51	52	53	54	55	56	57	58	59
5	60	61	62	63	64	65	66	67	68	69	70	71
6	72	73	74	75	76	77	78	79	80	81	82	83
7	84	85	86	87	88	89	90	91	92	93	94	95
8	96	97	98	99	100	101	102	103	104	105	106	107
9	108	109	110	111	112	113	114	115	116	117	118	119
10	120	121	122	123	124	125	126	127				

Table 9 MIDI note numbers

For more general models whose input musical notes are not limited under **ISO 16**, the frequency values can be used directly as the numerical measurements of input.

4.1.2 Frequency Analysis

There are a few acoustic features that characterise a musical tone. In the time domain a tone typically consists of an onset, steady-state (or main body) and offset (Fig.21).

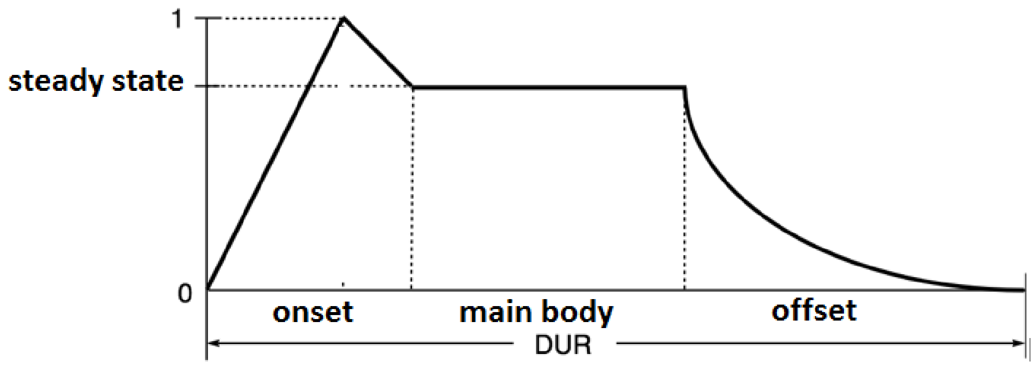


Figure 21 Time envelope of a musical tone

where DUR is the time duration of the tone.

In frequency-domain analysis, a musical tone in the sustain stage typically has a harmonically-patterned spectrum (Fig.22):

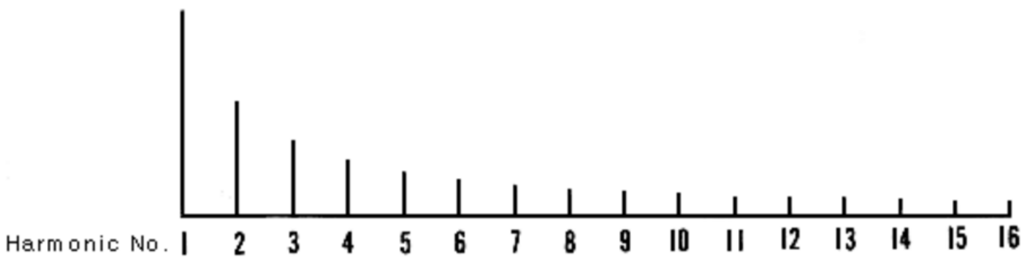


Figure 22 A harmonic spectral template

As observed from above figure, a harmonic spectrum is discrete and the frequency components are ideally equally spaced. Each frequency component is considered as a *harmonic*. The frequencies of all the harmonics are positive integer multiples of the lowest partial's frequency, called the *fundamental* frequency; and all the partials excluding fundamental frequency (if any) are also called *overtones*. Continuous or non-harmonic spectral structures on the other hand creates sound perceived as noise.

However, the spectral features for even the same sound source do not always remains stationary in time for the duration of the musical tone. The spectrogram plot allows us to observe dynamic features of the frequency partials (see Fig.23).

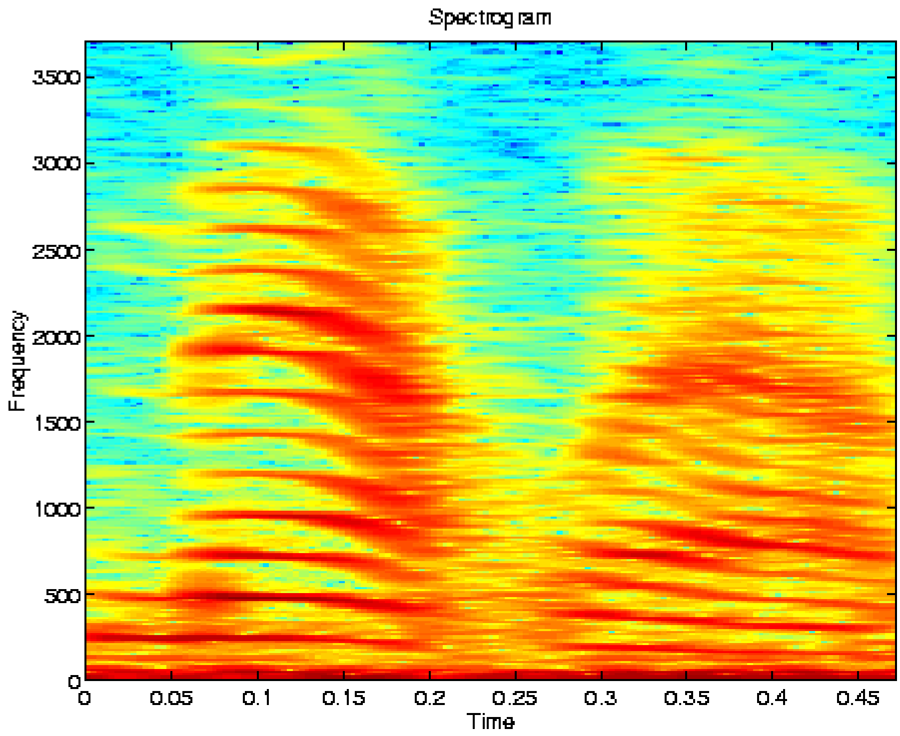


Figure 23 A sample spectrogram plot (Patterson et al., 2014)

where the dark lines represent how harmonics change over time

Although in reality the frequency spectrum of a particular note does not remain stationary, under psychoacoustic modelling the main body a musical tone is usually represented by the frequency features of its main body. Models using frequency features as input configurations indicate a frequency-based analytical method.

Under frequency-based approaches, a musical tone is typically modelled by a series of harmonic pure-tone partials with amplitude-decaying characteristics (see Fig.22). For instance, in Parncutt's model (Parncutt, 1989), the amplitude ratio of n^{th} over $(n+1)^{th}$ harmonic partials is set at 8 – 10 dB; whereas in Sethare's model (Sethare, 1999) the ratio is fixed at a constant of 0.88. The amplitude-decay reflects the dominant roles of lower partials in pitch perceptions. Higher partials (beyond 4kHz) are seldom important for the reason of *spectral dominance*[19].

4.2 Psychological Measurements for CDC

While trying to explore the universal sensitivity to music consonance and dissonance, psychoacoustic modelling requires consonance and dissonance concept (*CDC)* to be a quantifiable and measurable concept. However, at the present stage of research the *CDC* are understood via the subject's verbal descriptors rather than measurable biological variables. Thus the task of measuring the level of consonance (or the level of dissonance) can be quite complex. The semantic meaning of *CDC* is multifaceted, with synonyms such as *tense/relaxed*, *centric/acentric*, *diatonic/chromatic*, *primary/subordinate*, *stable/unstable*, *close/distant*, *similar/different*, *rough/smooth*, *fused/segregated*, *implied/realized*, and *tonal/atonal* being used in music harmony.

According to (Parncutt and Hair, 2011), *CDC* are primarily described by the *concord/discord* or *pleasant/unpleasant* dichotomies. *Tension*/*relaxation* have been referred to as the 'close relatives' of *CDC*; and while considering pitch hierarchies[20], descriptors such as *primary/subordinate, centric/acentric*, and *stable/unstable* are also involved. From the psychological foundation point of view, Parncutt and Hair asserted two physiological and psychological processes: *sensory dissonance* measured by *roughness* (see *theory of beats* in section 3.1) and *tonal consonance* measured by *harmonicity* (see *theory of harmonic-template* in section 3.2). Therefore, depending on the physiological and psychological theories involved, consonance and dissonance can also be measured by the level of either *harmonicity* or *roughness*.

Other numerical approaches generally presume that the *CDC* can be measured on a one-dimensional scale by assessing relative extremes of consonance and dissonance

---

[19] Spectral dominance refers to a set of lower harmonics having the dominant role in the determination of spectral pitch. Moore (1985) concluded the first six harmonics were the spectrally dominant partials.

[20] Under a pitch hierarchy system, certain pitches are considered to be more important and stable than others. In musical chords or scales, pitch hierarchy has been observed.

or equivalent semantic descriptors such as stable–unstable in (Kaestner, 1909), or pleasant–unpleasant in (Guthrie and Morrill, 1928) and (Plomp and Levelt, 1965).

4.3 Sensory Beats based Models

In the literature, the ‘beat’-informed dissonance concepts are generally modelled in a two-step approach:

(1) Observe the level of sensory beat effect based on pure-tone partial interactions. Such effects have been studied under the discussion of dissonance curves, and this is discussed in the next section (sect. 4.3.1).

(2) Extend the observation from pure-tones to complex tone(s) by a summation algorithm. The overall level of sensory beat effect is assumed to be the exhaustive sum of pure-tone pairs interactions. Detailed summation methods for the overall beat sensations for complex tone(s) are discussed in section 4.3.2.

Such a two-step approach reflects a physical research philosophy which tries to attack a complex phenomenon from simple, fundamental observations. However, the perception of music harmony can be different from pure physical interactions as many psychological processes are involved. Therefore, there is a dispute over whether it is possible for the observations from pure-tone partial interactions to be ‘summed’ to account for music harmony consisting of complex tone(s) (Sethare, 1999). Nonetheless, according to the original hypothesis (Plomp and Levelt, 1965) that sensory beats are a result of physiological inner-ear function, the analysis of the overall sensory dissonance can be precisely modelled based on the use of dissonance curves.

4.3.1 Dissonance Curves

A dissonance curve depicts the perceived level of sensory beat effect as a function of the frequency difference between two pure-tone partials. In the early 20th century, many empirical studies were carried out to obtain a quantitative plot of such a dissonance curve. Kaestner (1909) produced measurements of the ‘pleasantness’ against intervals within one octave (1:2) (Kaestner, 1909) (Fig.24). Another investigation by Guthrie and Morrill, measured the level of both the ‘consonance’ and ‘pleasantness’ for the intervals from Unison (1:1) to slightly beyond the Fifth (2:3) (Guthrie and Morrill, 1928) (Fig.25). A high correlation was found between ‘consonance’ and ‘pleasantness’ from Guthrie and Morrill’s plot. On the whole, both plots confirmed Helmholtz’s thoughts, with only minor data discrepancies.

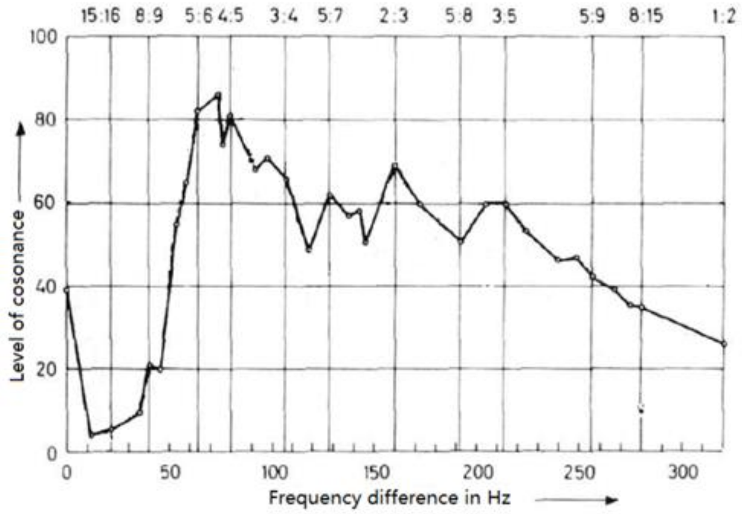


Figure 24 Kaestner’s ‘pleasantness’ curve

where 30 discrete intervals were rated from the subjects, with a lower frequency fixed at 320Hz. Figure taken from (Kaestner, 1909)

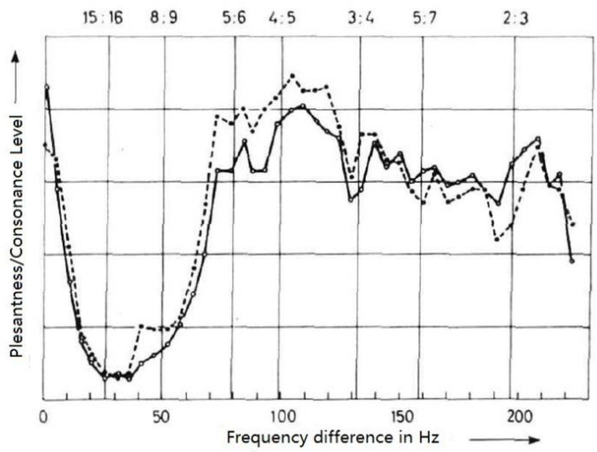


Figure 25 Guthrie and Morrill’s curve of consonance

where the consonance (solid curve) and pleasantness (dashed curve) of 44 discrete intervals were examined with lowest frequency at 395Hz. Figure taken from (Guthrie and Morrill, 1928)

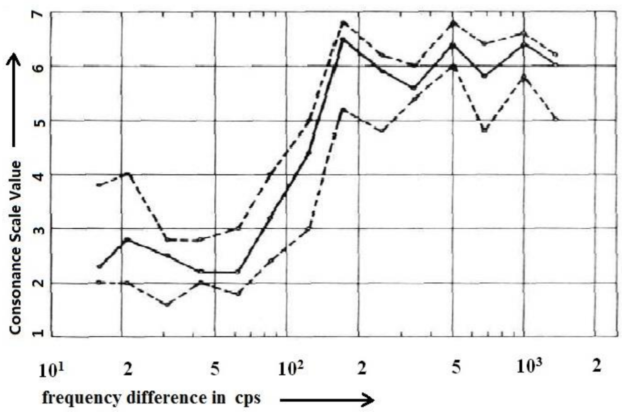


Figure 26 Plomp and Levelt’s consonance curve at mean frequency of 500Hz

Figure taken from (Plomp and Levelt, 1965)

Plomp and Levelt further complete the *tonotopical* theory with an exploration of the frequency dependency of the consonance perceptions with mean frequencies of 125, 250, 500, 1000, 2000Hz respectively (see Fig.26). Within each frequency centre, several selected frequency intervals up to one octave were studied. The result showed that the frequency intervals where ‘maximum consonance’ occurs increase with frequency; they roughly correspond to 25% of the critical bandwidth of the respective auditory filters. A smoothed dissonance curve (Fig.27) is theorized to represent the roughness sensations within a critical bandwidth.

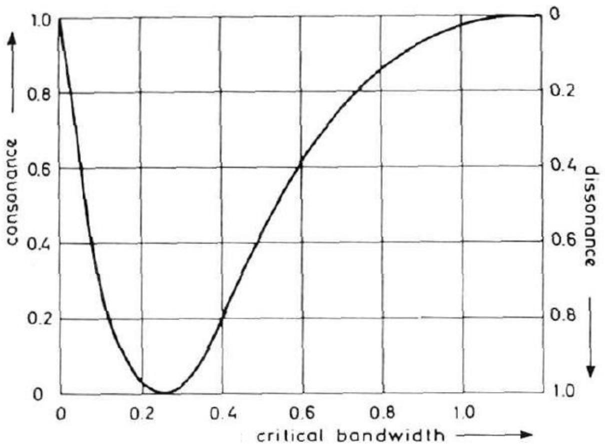


Figure 27 Plomp and Levelt's theoretical curve (taken from Plomp and Levelt, 1965)

Many other related experiments were done in the late 20th century such as (Kameoka and Kuriyagawa, 1969) (Fig.28), and (Hutchinson and Knopoff, 1978), They differ from Plomp and Levelt's model in either the experimental data or the critical bandwidth model used. But similar conclusions were made about the characteristics of the dissonance curve:

1. There is a single-peak dissonance observed within a critical bandwidth; and
2. A minimum amount of sensory dissonance is observed when the separation between two pure-tones exceeds one *CB* (critical bandwidth).

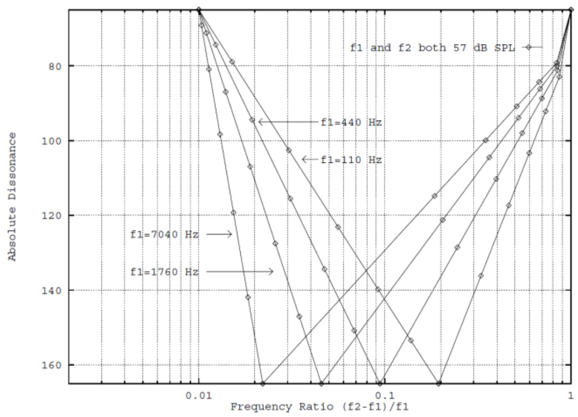


Figure 28 Kameoka & Kuriyagawa's 'V-shape' plot (taken from Kameoka and Kuriyagawa, 1969)

Based on the theoretical curve from (Plomp and Levelt, 1965; Sethare, 1999) developed a numerical model for dissonance curves which applies to a range of base frequencies:

$$d\,(f_1, f_2, v_1, \mathrm{v}_2) = \mathrm{v}_1 \bullet \mathrm{v}_2\,(\exp(a{\bullet}s(f_1-, f_2))\;\; - \exp\,(b{\bullet}s(f_1-, f_2)\,))$$

Where d is the dissonance curve as a function of two pure tone partials with frequencies of f_1, f_2 and amplitudes of v_1, v_2; a and b are the tuned constant taking the values of 3.5 and 5.75 respectively; s is the parameter which allows the curve to interpolate between different base frequencies (illustrated in Fig.29):

$$s = 0.24 \,/\, (0.021\, f_1 + 19)$$

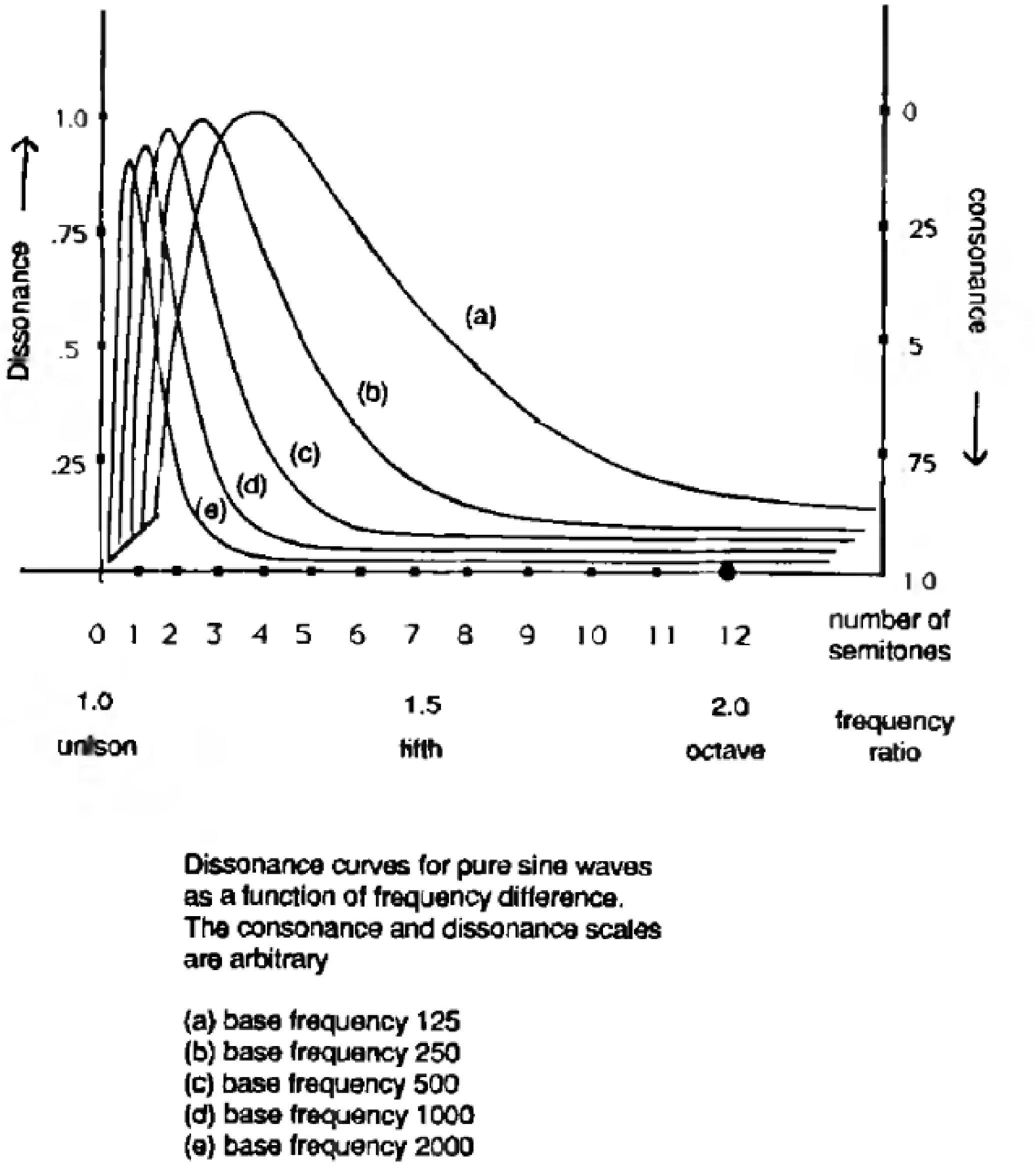


Figure 29 Dissonance curves as a function of base frequencies

Figure taken from (Sethare, 1999)

The dissonance curves generally have a single peak roughly corresponding to an interval of two to three semitones. Such an observation is however contradictory to the C&D observations of music intervals described in section 2.3. Further research (Hindemith, 1984; Benade, 1976) had shown that the beating effect can also be observed slightly below and above integer frequency ratios. For instance, when two pure tones are separated at an interval slightly above or below one octave (frequency ratio: 2/1), the beating effects can also be clearly observed; such a phenomenon has been referred as *secondary beats*. The *secondary beat* effect is one of the issues *that c*annot be explained by the *tonotopical* theory; so higher-level organs along the auditory pathway must be involved in order to account for *secondary beats*.

4.3.2 Estimating Roughness Sensations for Complex Tones

The precise modelling of the beat effect in a complex tone scenario can be very problematic for the following reasons:

The influence of amplitudes of partial components has not been thoroughly studied. Present numerical approaches such as Sethare's assume that the perceived level of roughness to be proportional to the amplitude of the partials, implying the perceived overall dissonance is proportional to the sound volume of musical entities. But such an implication has not been observed in music literature.

When the amplitude effect is involved in model calculations, we have to consider the auditory mutual masking effect. However, the amount of masking for a pure-tone under multiple maskers at this stage cannot be explicit modelled. Instead, many simplified methods have been proposed. In Terhardt's approach, a level of 15–25dB per critical band was applied to the masked pure-tones for a single pure-tone masker (Zwicker, 1970).

According to (Taylor, et al. 1974), the beats effect introduced by a particular pair of partials can either mask or be masked by the beats effect of the combinations of other pure-tone pairs, this is also known as "beats masking beats". But no empirical data had been provided in order to model such a masking effect.

In order to simplify the modelling for complex tones, a linear summation algorithm is typically used to calculate the overall dissonance for complex tones:

$$D = \sum d\ (f_i, f_j, v_i, v_j)$$

Where d is the level of dissonance introduced between two pure-tone partials (i, j) whose frequencies are f_i, f_j and amplitudes are v_i, v_j respectively. The dissonance effect is summed over all possible combinations of pure-tone pairs within the input sonority.

Such a model leads to an internal dissonance for the musical tones, as the interactions of pure-tones within a musical tone will also produce beats for higher harmonics (typically above seventh harmonics[21]).

Instead of a linear summation method, (Zwicker, Flottorp, and Stevens, 1957) proposed a logarithmic summation algorithm in relation to the power law of psychological significance[22]:

$$D = \sum \{d\ (f_i, f_j, v_i, v_j)\}^{\beta} \quad (f_i \neq f_j)$$

where $\beta = 0.75$ according to (Zwicker, 1970).

Huron (2000) pointed out that the problem with such kinds of dissonance model is that the total dissonance will always tend to increase with the number of tones, which is contradictory to common perception. The effect of magnitude and the number of partials can be normalised by the overall energy of the spectrum, denoted by $\sum (v)$:

$$D_{norm} = D / \sum (v)$$

---

[21] According to Helmholtz, higher harmonics can fall into a single critical bandwidth, therefore producing a sensation of "rough" and "cutting" (Gray, 1977).

[22] Under the power law of psychological significance, the perceived magnitude of stimuli is proportional to the acoustic features to the power of certain values.

Huron also hypothesised that such a dissonance effect would be neutralized when sonority consisted of multiple tones/notes. This effect is taken into account by dividing the total normalised dissonance by the noticed multiplicity (M) of pitches:

$$D = D_{norm} / M$$

Even with such simplified model designs, plausible predictions have been made to account for simultaneous musical dyads, which are discussed in the next section.

4.3.3 The Predicted Results of Beats-Based Models

Plomp and Levelt's (1965) research is notably the first beats-model that is able to account for intervallic harmony (see section 2.2). A dissonance curve of a complex tone pair spaced at various intervals was produced (see P&L curve, Fig.30).

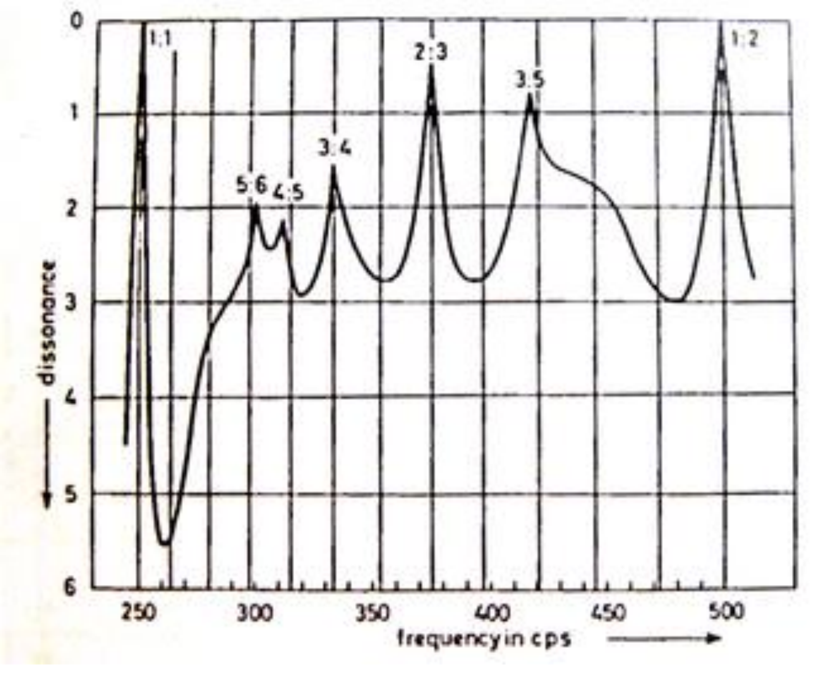


Figure 30 P&L complex tone dissonance curve (taken from Plomp and Levelt, 1965)

The P&L curve has a few 'peaks' of consonances generally corresponding to the diatonic intervals of the unison (1:1), minor third (5:6), major third (4:5), fourth (3:4), fifth (2:3), and major sixth (3:5). A harmonic complex tone (with the first six harmonics) was used to produce the P&L complex tone dissonance curve.

However, the same summation method fails completely to account for the common perception of musical triads (Table 10)

Triad Class	Inversions	Empirical Ranking	Beats based models			
			Helmholtz	P&L	K&K	Sethare
Major	Root Position	**1**	3	3	1	3
	1st Inversion	**2**	6	8	8	5
	2nd Inversion	**3**	1	1	3	1
Minor	Root Position	**4**	3	3	1	3
	1st Inversion	**5**	1	1	3	1
	2nd Inversion	**6**	6	8	8	5
Diminished	Root Position	**7**	10	10	3	9
	1st Inversion	**8**	8	5	6	7
	2nd Inversion	**9**	8	5	6	7
Augmented	Root Position	**10**	5	7	10	9

Table 10 Beats-based model predictions for musical triads

The intervallic structure shows the intervals of the lower note to middle note and middle to higher note in semitones; the empirical ranking is obtained from Roberts (1986), 1–10 with 1 being the most consonant; the model predictions correspond to Plomp & Levelt (1965), Kameoka & Kuriyagawa (1969), Sethares (1999); data for P&L, K&K, Sethare and Parncutt cited from (Cook 2006). It can be observed that none of the models is fully able to simulate the empirical rankings

In the next section, a mathematical discussion of why the beats-based computational method succeeds in accounting for musical intervals but fails for the triads will be presented.

4.3.4 Mathematical Discussion of Beats-Based Computational Method

The computational method that appears in the beats-based model is mathematically equivalent to the determination of the level of ‘shared frequency components’ between two sound spectra. The more that two input spectra share frequency components, the more likely they are to be considered as mutually consonant. For example, consider a harmonic tone played at C4, and compare it to when it is played at G4: as the two harmonic spectra share a good number of frequency components, they are quite consonant in reference to each other. In contrast, for the case of C4 and D$^{\#}$4, these two harmonic spectra share almost no frequency components, and therefore they are considered as dissonant in reference to each other.

Based on the fundamental principle that consonance and dissonance are determined by the degree of shared frequency components, we may derive that two harmonic complex tones with simple integer ratios will inevitably have a higher degree of partial-overlapping as compared to random intervals. Such an idea can be illustrated by the following example:

Harmonic tone A: fundamental frequency f_0; therefore, it consists of frequency partials:

$$\{f_0, 2f_0, 3f_0. nf_0\}$$

Harmonic tone B: fundamental frequency of pf_0/q; therefore, consists of frequency partials:

$$\{pf_0/q, 2pf_0/q, 3pf_0/q. pnf_0/q\}$$

When both p and q are simple integers, the likelihood of $pnf_0/q = mf_0$ will be increased. (Where m and n are the number index of harmonics, which are also simple integers).

Therefore, the dissonance curve is not necessarily the only condition to produce the P&L curve (Plomp and Levelt, 1965). However, the use of harmonic tones is a necessary condition to produce a prediction result that has good alignment with the music literature (such as the P&L curve).

The consonance and dissonance perception for a music triad is not simply based on the complexity of frequency ratios (see section 2.6); it therefore fails under beats-based models. For instance, the middle note of a minor triad (in root position) barely overlap with either the bass or the highest note, thus the consonance of the minor triad cannot be explained in this way.

When the same analogy is applied to non-harmonic tones, the consonance and dissonance can also be estimated. For instance, when two non-harmonic church bells are played together, the algorithm seeks the level of their shared frequency components; if a high degree of overlapping is found, it is considered as 'well-tuned' (or consonant), and vice versa. The degree of frequency overlapping is also affected by the fundamental frequency of both non-harmonic tones. Figure 31 below shows two non-harmonic tones played at the same pitch (left) and a different pitch (right). It can be observed that in the left case, more partials are overlapped, and thus will be more consonant.

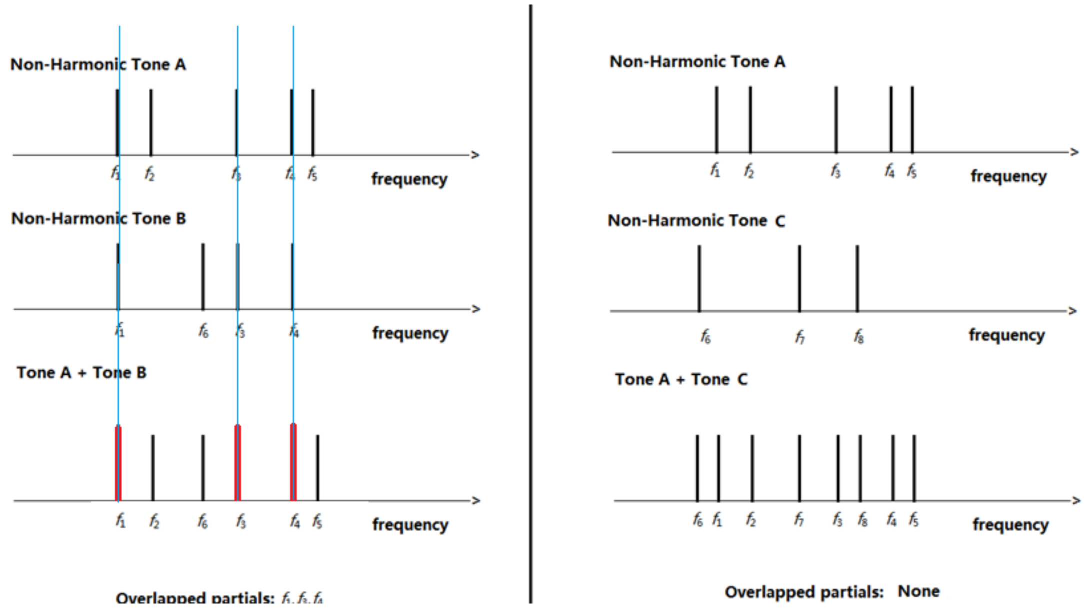


Figure 31 Illustration of Beats based computation mechanism

Sethare demonstrated a mathematical approach where – for given any spectrum – a series of solutions of the consonance and dissonance intervals can be generated. For instance, a stretched/compressed harmonic spectrum is defined as:

$$f_i = f_1 A^{\log_2 i}$$

Where f_i is the frequency value for the i^{th} partial, and f_1 is the frequency of the first and lowest partial; A>1;

For such kind of stretched/compressed harmonic spectrum, the interval of consonances will also be stretched or compressed (see Fig.32).

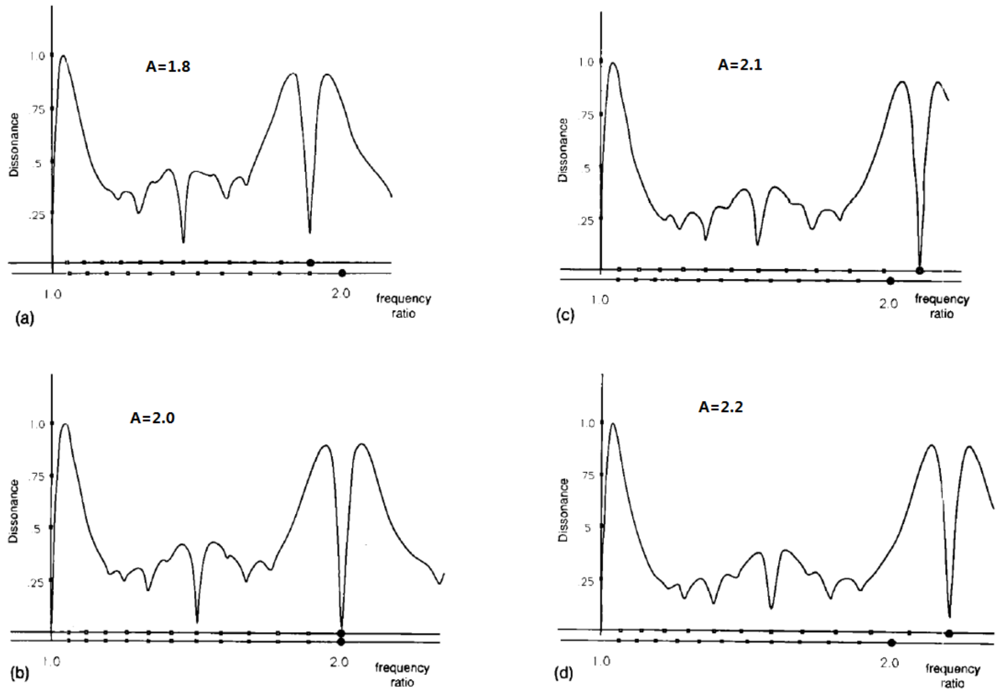


Figure 32 Sethare's plot of the compressed and stretched 'octaves' (taken from Sethare, 1999) Where the effect of using different timbre can be viewed from the location of 'consonance dips' on the frequency scale

Sathare's research work (1999) provided a thread of using non-harmonic timbres for music compositions. He also established a mathematical framework between timbre structures and the tuning system. However, as non-harmonic timbre usually does not have strong pitch perceptions; the scales for non-pitched tones can be strange: as the intervals are defined with pitch perceptions only. In Sathare's demonstrations (*A Bell, A Rock*, *A Crystal*), the lowest partial is always assumed to be the pitch. Such assumptions have not so far been verified in past empirical studies.

4.3.5 A Conclusion for Beat-Based Models

The beats-based computational method can be viewed as a 'bottom- up' physiological modelling rationale that requires the precise functional responses of each organ along the auditory pathway. However, the functional responses of the entire auditory pathway have not been fully understood from psychological or neural measurements at the present stage. To overcome such uncertainties, model simplifications and hypothetical conditions have been used in beats-based modelling, leading to a number of theoretical approaches and inconsistent prediction results. The dispute has been mainly focused on the role of partial amplitude and the summation method of the beat effect.

However, the phenomenon of sensory beats is evidential, especially observed in simultaneous chords that contain the intervals of one or two semitones. The auditory beats are primarily due to the physiological functions of the auditory critical bands.

4.4 Harmonic-template Based Models

Harmonic-template based models are developed in relation to the harmonic-template harmony perception theories. The central thread of harmonic-template based models is to estimate how well the sound stimuli match to a harmonic acoustic spectrum, known as the *Corps Sonore* (Rameau and Wundt, 1721). A higher degree of matching is prone to the consonance concepts. As discussed in section 2.1, harmonic spectrum usually corresponds to a music tone, such type of consonance is also known as tonal consonance.

On the other hand, the distortions from this harmonic spectrum are considered as dissonance (or tonal dissonance). The extreme of dissonance is referred to complete random sound spectrum with no identifiable harmonic spectral patterns. Therefore, the computational algorithm of harmonic-template based models can be viewed as a feature matching process or pattern recognition.

In the literature, the number of numerical analytical models proposed based on harmonic-template theories are significantly less than that of beats-based theories, and their proposals are comparatively new. In this section, three of the harmonic-template based models will be reviewed:

Section 4.4.1 Parncutt's *tonalness* model (Parncutt, 1989)

Section 4.4.2 Hofmann-Engl model (Hofmann-Engl, 2006)

Section 4.4.3 Stolzenburg's *relative periodicity* model (Stolzenburg, 2012).

In section 4.4.4, the prediction result of harmonic-template-based models and the theoretical limitations are summarised and discussed.

4.4.1 Parncutt's Tonalness Concept

The *Tonalness* concept is one of the central parameters in Parncutt's *CDC* model (Parncutt, 1989). It roughly represents how 'tone-like' a sound input is. A higher value of *Tonalness* implies the input sound object has a better match with harmonic tones, which according to Terhardt's harmonic-template hypothesis (Terhardt, 1979), corresponds to a more consonant sensation. The entire psycho-acoustic model contains two main sequential stages: estimation of pure-tone audibility and the estimation of complex-tone audibility.

In the first stage, the windowed sound input is transformed into the frequency domain, represented by the amplitude and frequency values. The model considers both the threshold of hearing and mutual masking effect, where the amplitude of most frequency components will be attenuated or zeroed. The remaining amplitudes of a finite number (n) of frequency partials are used to denote the probability of noticing the pure-tone frequencies, also known as the pure-tone audibility (A_p).

$$A_p(f) = [A_p(f_1), A_p(f_2), A_p(f_3), \dots A_p(f_n)];$$

In the second stage, the audibility of pure-tone components $A_p(f)$ are compared to a predefined harmonic spectral template in order to determine the complex-tone audibility. The computation of *Tonalness* is further associated with the complex tone with the highest audibility. This process is illustrated as follows:

The harmonic-template used in his approach is defined as:

$$A_T(f_i) = A_0/i$$

Where $A_T(f_i)$ is the audibility of the i^{th} harmonic of the harmonic template, and is equal to the audibility of the fundamental partial A_0 divided by the harmonic index i. This is fundamentally an impulse train with amplitude-decaying characteristics (see Fig.33).

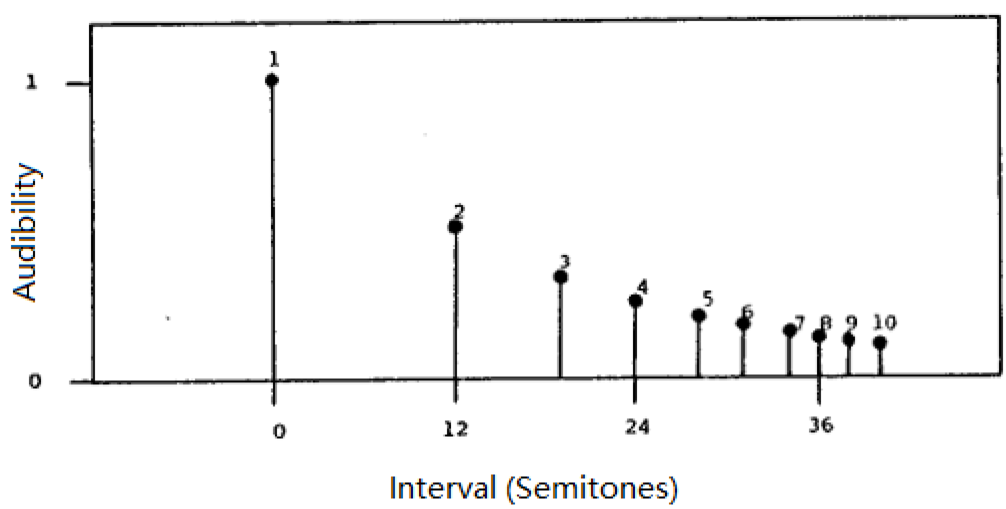


Figure 33 Harmonic-template in Parncutt's approach (Parncutt, 1989)

Parncutt's model adopted this harmonic-template and used it in a (virtual) pitch determination algorithm (e.g. Jenson, 2008). A (virtual) pitch determination algorithm generates a number of candidate pitches and each pitch is associated with a predicted *weight* (representing the likelihood the entire sound is perceived as having this pitch). The computed *weight* of each pure-tone frequency component (*weight* (f_1)) is then used to denote the complex-tone audibility ($A_c(f_i)$) of the input sound.

$$A_c(f_i) = [weight\,(f_1)\,,\, weight\,(f_2)\,,\, weight\,(f_3)\,,\, \ldots\, weight\,(f_n)];$$

The *tonalness* corresponds to the complex tone with highest audibility divided by a scalar:

$$T = \mathrm{Max}\,(A_c(f))/6.2$$

where 6.2 is the scalar used to make the *tonalness* of a major triad (root position) roughly equals to 1.

To demonstrate the idea, let's assume the input sonority is a C_{maj} triad. Based on the pitch determination algorithm, pitches C, E, G are expected to have the highest *weight*. More specifically, the overtone series of both E, and G have a strong function to suggest C (i.e. many overtone frequencies in common) as the root pitch according to the pitch determination algorithm, therefore the audibility of the C note (divided by a constant scalar) represents the *Tonalness* of C_{maj}. When the input triad is $C_{min,}$ the note E^b functions considerably less to suggest C as the root pitch, therefore the audibility of the C, whilst still the highest audibility due to the fifth (G), is much weakened compared to C major. This results in a lower value of *Tonalness*, thus corresponding to a more dissonant perception.

Furthermore, Parncutt's model can also be used to analyse the dynamic properties between successive sonorities (such as chord progressions). The successive sonorities are analysed based on essentially two main concepts: the *pitch commonality* (the degree to which two sonorities share common pitch components) and *pitch distance* (the aggregate of pitch proximity between two sonorities). Parncutt's computational model provides an intuitive way to realize harmonic-template based theories. The model prediction result has also achieved a major success in accounting for the consonance of major and minor chord structures.

4.4.2 Hofmann-Engl Model

The Hofmann-Engl model (Hofmann-Engl, 2004) is built on the perception fact that for some musical intervals, the higher note tends to indicate the lower note to be the overall pitch. In particular, he identified six musical intervals with such functions; they are the unison, perfect fifth, major third, minor seventh, major second and major seventh intervals. These intervals have such special functions because of they appeared in the natural harmonic series of a fundamental frequency (see the illustration in Fig.34).

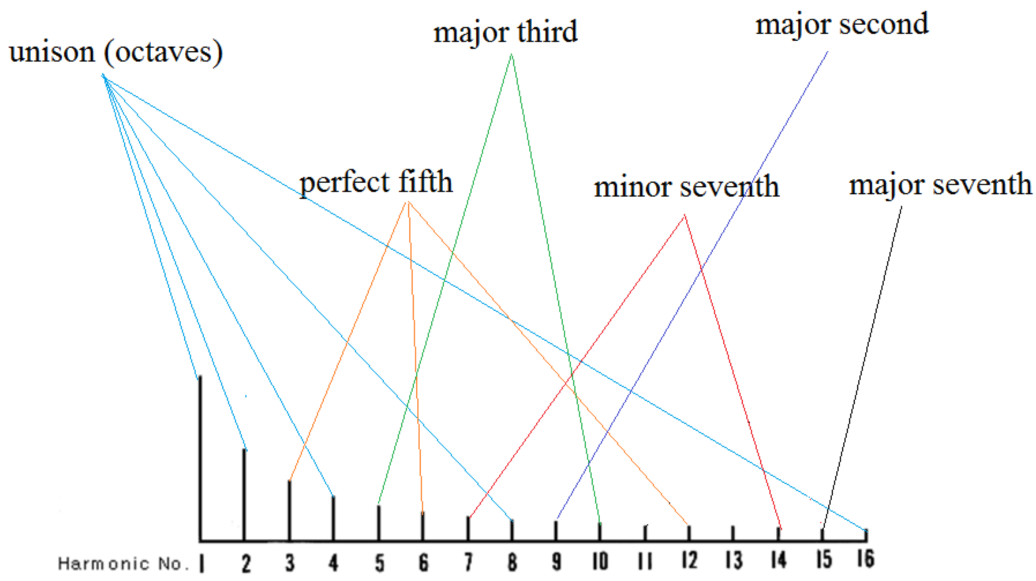


Figure 34 Special intervals in Hofmann-Engl model and natural harmonic overtone series

The above figure illustrated how the fundamental of higher notes naturally resides at the overtone series of the lower note for six special intervals used in Hofmann-Engl model.

However, the strength of such indication functions (S) is different for the six musical intervals, and it is modelled by:

$$S(c) = \frac{6^2 - c^2}{6}$$

Where the indexes of six intervals (c) are from 0 = unison (octaves), 1= perfect fifth, 2 = major third, 3 = minor seventh, 4 = major second, and 5 =major seventh.

Substituting c = 1 – 6, we can obtain: 6 *Hh3*, 5.83 *Hh*, 5 *Hh*, 4.5 *Hh*, 3.3 *Hh*,and 1.83 *Hh*. All other intervals are therefore assigned to 0 *Hh*. The setting of such values is initially tuned according to Hofmann-Engl's experience, and it is later proved to be quite in line with common perceptions. A unit of *Hh* is proposed for this concept and it is after the German scientist Hermann von Helmholtz. From the above formula, we can observe that the unison/octave interval has a strongest function to indicate its lower note as the tonal centre of the multi-tone structure, the perfect 5th; major 3rd, minor 7th, major 2nd and major 7th also have similar functions, but are gradually weakened.

The Hofmann-Engl model is a practical tool that is easy to use in music analysis, but such a simplified model ignores the role of the frequency accuracy of a partial. For example, assume a frequency centre f_0 = 400Hz, and a test partial (f_t) whose frequency is 605Hz. As 605Hz is approaching 600Hz this will be roughly treated as a perfect fifth interval, the rate of frequency error γ is computed by (605–600)/600 = 0.0083. When threshold of error δ is set at 0.01, and the weight of confirmation of the fifth interval (600Hz/400Hz) is 5.83 *Hh*, the justified $S(c)$ should be:

$$5.83 \times (1 - 0.0083/0.01) = 0.9911\ Hh$$

To predict perceived level of *CDC* for a given input chord structure, all twelve chromatic pitch classes were assumed as the hypothetical tonal centres and their corresponding strength of root at particular pitch $S(p)$ are calculated.

$$S(p) = \frac{\sum_{i=1}^{n} Si(\mathrm{c})}{\mathrm{n} \cdot \mathrm{g}} \cdot \sqrt{1/\mathrm{i}}$$

Where n refers to the number of tone components in a chord; the term $\sum_{i=1}^{n} Si(\mathrm{c})$ describes the total strength that all tone component indicating pitch p as the tonal centre. This value is further scaled by n g where g is a constant scalar (g = 6 *Hh*) and an extra term $\sqrt{1/\mathrm{i}}$ is used to represent the impact of the position of that tone component (the lower position of a tone component has a higher weight in the determination of tonal centre, similar to Parncutt's approach).

Once the strength of tonal centre $S(p)$ has been computed over all 12 pitch classes, the overall *CDC* tonal consonance, known as the *Sonance* $S(ch)$ concept in Hofmann-Engl model, can be estimated. The fundamental principles to compute *Sonance* are two:

i) The stronger the strongest $S(p)$ is, the higher the degree of *Sonance*.

$$S(\mathrm{ch}) \propto \max(S(p));$$

ii) The more ambiguous the tonal centre, the less degree of *Sonance*.

$$\mathrm{S(ch)} \propto 1/ \sum_{j=1}^{n} Sj(p)$$

where $\propto$ refers to a strict proportional mathematical relationship

To conclude, the Hofmann-Engl model provides a very straightforward method to estimate the tonal consonance. Similar to Parncutt's model, the basic theoretical assumption is that tonal consonance concept mainly determined by the strength of the input sound spectrum matches to a particular harmonic template. Comparing to Parncutt's model, the Hofmann-Engl model considers an extra factor that the number and strength of other noticeable complex pitches will actually decrease the degree of tonal consonance (*Sonance*). This is perceivable as the increment of the numbers and strength of other noticeable pitches will weaken the pitch perception of the tonal centre, resulting in an ambiguous complex tone sensation (tonal dissonance).

4.4.3 Relative Periodicity

The estimation of *tonal dissonance* in Stolzenburg's model can be viewed as a temporal version of the pattern matching algorithm: In frequency based analysis, input sonorities with a better match to the harmonic-template are considered to be more consonant (thus higher value of *harmonicity*); when the same thinking is applied to time-domain analysis, the task is shifted from 'finding the harmonic-template that best fit the input sonority' to essentially 'finding the fundamental periodicity of the input waveform.' At the centre of this algorithm is therefore a *periodicity detection* approach (Stolzenburg, 2012)

Using the *periodicity detection* algorithm, each component note or frequency partial is expressed in ratio form in reference to the lowest note/frequency partial of the input sonority. This computation method requires that the input frequency intervals to be expressed in the form of simplest integer fractions. Therefore for random frequency intervals a Stern-Brocot tree (Hayes, 2008) is used to approximate the frequency ratios under a certain error-tolerance threshold[23]; for diatonic intervals, just-tuned

[23] This value is set at 1% in Stolzenburg's approach.

frequency ratios are used. The computational method can be summarised in following main procedures:

For an input sonority with frequency component: $\{f_0, f_1, f_2 \ldots f_n\}$, the model firstly converts the input into ratio forms with respect to the lowest frequency:

$$f_0: \{f_0/ f_0, f_1/ f_0, f_2/ f_0, \ldots f_n / f_0\}.$$

Next, the frequency ratios are further converted into fractions of simplest integers:

$$\{f_0/ f_0, f_1/ f_0, f_2/ f_0, \ldots f_n / f_0\} = \text{or} \approx \{a_1/ b_1, a_2/ b_2, a_3/ b_3, \ldots a_{n-1} / b_{n-1}\}$$

where a_n and b_n are simple integers.

The *relative periodicity* (h) is computed by the least common multiple (*lcm*) function of b_n, as only the denominators b_n matter to the fundamental periodicity.

$$h = lcm\ (b_1, b_2, b_3 \ldots b_{n-1})$$

The least common multiple function essentially computes the least periodicity that is shared by all partial/tone components, as it is hypothesised that *relative periodicity* (h) is inversely related to the *harmonicity* (Juan, 2006).

For example, for a major triad (in root position):

$$\{a_1/ b_1, a_2/ b_2, a_3/ b_3\} = \{1/1, 5/4, 3/2\}$$

$$h = lcm\ (1,4,2) = 4$$

For a minor triad (root position):

$$\{a_1/ b_1, a_2/ b_2, a_3/ b_3\} = \{1/1, 6/5, 3/2\}$$

$$h = lcm\ (1,5,2) = 10$$

Therefore, the major triad is deemed more consonant (as expected).

4.4.4 The Predicted Result of Harmonic-Template-Based Models

Compared to beats-based models, the harmonic-template-based models produce similar rankings for musical dyads, but they generally surpass beats-based models in the prediction of musical triads (Table 11).

Triad Class	Intervallic Structure	Empirical Ranking	Harmonic-template based models		
			Parncutt	H-Engl	Stolzenburg
Major	4-3	1	1	2	2
	3-5	2	6	3	3
	5-4	3	3	1	1
Minor	3-4	4	4	8	4
	4-5	5	6	9	5
	5-3	6	10	10	7
Diminished	3-3	7	9	5	9
	3-6	8	5	4	8
	6-3	9	8	6	6
Augmented	4-4	10	2	7	10

Table 11: Model predictions for common triads

The intervallic structure shows the intervals of the lower note to middle note and middle to higher note in semitones; the empirical ranking is obtained from (Roberts, 1986), 1–10 with 1 being the most consonant; the model predictions correspond to (Hofmann-Engl, 2012; Stolzenburg, 2012) and data for H-Engl is calculated from web-based java application: *Harmony Analyser 3.2* (http://www.chameleongroup.org.uk/software/piano.html)

One of the theoretical advantages for the harmonic-template model is that it considers the information of the root whereas this information is not used in beats-based models.

However, the harmonic-template-based models, do not consider the effect of sensory beats. Moreover, harmonic template-based models are unable to differentiate the dissonance effect of augmented and diminished triads, and this is illustrated in Fig.35.

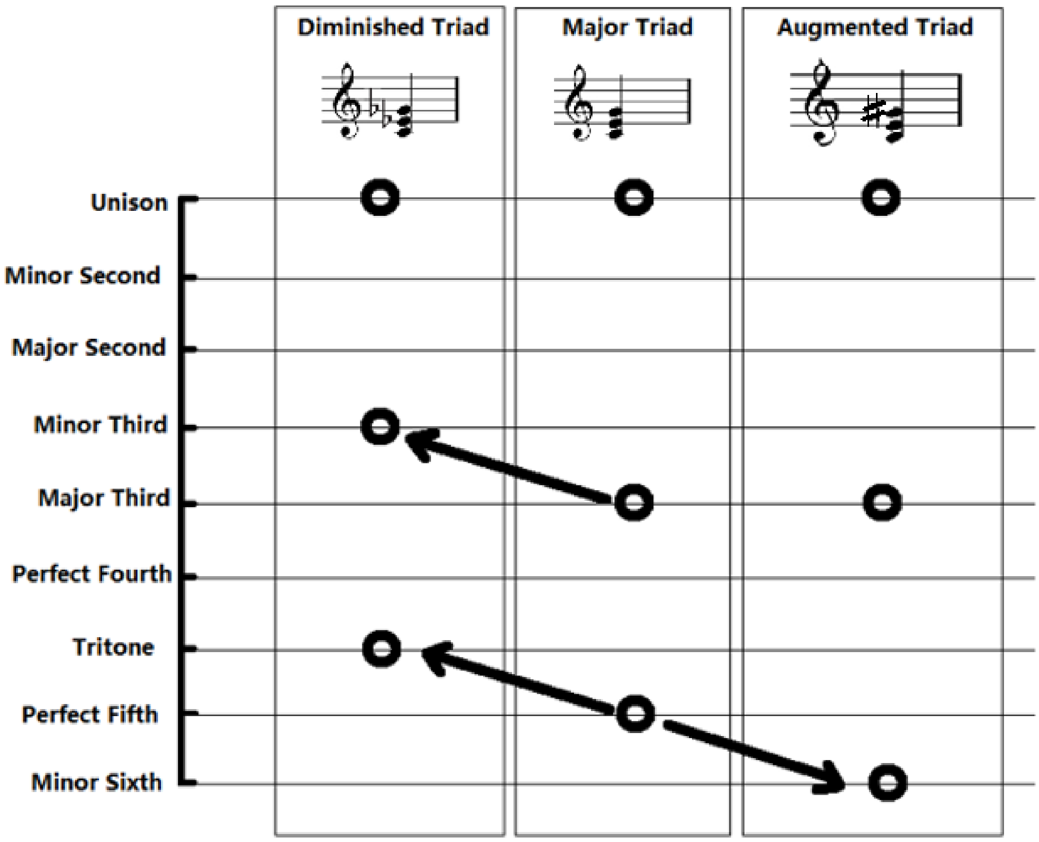


Figure 35 shift of notes from major triad to diminished and augmented triad

Under harmonic-template based models, the algorithm essentially determines how well an input sound spectrum matches the harmonic pattern, akin to a pattern recognition algorithm. For example, a major triad is composed of a frequency ratio of 4:5:6 (which are the 4^{th} 5^{th} and 6^{th} harmonics of a fundamental) and thus is expected to be a more consonance concept. The computational method of harmonic-template based models implies that the more the tone component was altered from the major triadic structures, the less degree of consonance is expected. Following this thinking, the diminished triad should be more dissonant than the augmented triad, as both the third and fifth are shifted downwards from the major structure for the diminished triad, whereas only the fifth tone is shifted upwards for the case of augmented triad. But empirical studies such as (Roberts, 1986) and (Cook, 2006) have shown that augmented triads are more dissonant than diminished triads.

The beats-based model also falls into difficulty distinguishing the minor triad from diminished triads, which forced Hofmann-Engl to conclude:

"*the algorithm failed where minor thirds have been involved (listeners overestimate the sonance*[24] *of minor thirds).*"

4.5 Comparison between Sensory Beats and Tonal Consonance

The relationships between *sensory beats* and *tonal consonance* can be summarized in following aspects:

Sensory dissonance can be viewed as a lower-level dissonance concept which mainly reflects the physiological limitation of the inner ears; and it is usually studied by the physiological functions. On the other hand, the *tonal consonance* is relatively a higher-level perception concept that is mainly associated with the brain functions; and it is usually analysed under psychological principles (e.g. Gestalt psychology).

Under beats-based analysis, the concept of *sensory dissonance* is actively defined by the level of beat /roughness effect contained within a sound stimulus, whereas the tonal consonance concepts are passively defined by the 'absence' of beat sensations. Conversely, under harmonic-template based theories, the *tonal consonance* concept is explicitly defined by a harmonic spaced acoustic template, but the dissonance concept is weakly defined as those sound spectra which do not generally match the acoustic harmonic templates.

In tonal music applications, the harmonic-template based theories consider an extra piece of information – the *tonal centre*, where in beats-based theories, such information is not used. The tonal *CDC* are not limited to 'how concordant between each pair of musical notes' (Tenney, 1988; beats-based theoretical thread), but also means how strong the chord structure suggests a clear tonal centre (harmonic-template based theoretical thread). For this reason, the harmonic-template based theories are more apt for the analysis of tonal musical entities, such as musical chords,

[24] Sonance, a concept that measures the level of consonance under Hofmann- Engl model (Hofmann-Engl, 2006)

or structures of harmonic tones (Temperley, 2004). And in contrast, the perception principle of *sensory dissonance* can be applied to virtually any sound.

The *tonal consonance* and *sensory dissonance* are not entirely inversely-related, in that, a lower level of beat effect does not necessarily correspond to a higher level of tonal consonance (take a suspended 4th triad composed of pure-tone partials for instance); however, a sonority with a higher degree of tonal consonance usually corresponds to a lower level of beat sensations.[25]

In harmony analysis for simultaneous chords, it can be concluded that sensory beats are more prominent whenever 1–2 semitones are observed within the chord structure (Cook, 1999). When 1–2 semitones intervals are absent from the chord structure (such as for tertiary chords), the sensory dissonance effect becomes less significant. In Fig.36, it can be observed that the sensory dissonance of 1–2 semitones intervals is significantly higher than any other intervals. Due to this, the beats-based theories are generally inadequate in the analysis of the harmony of tertiary chords.

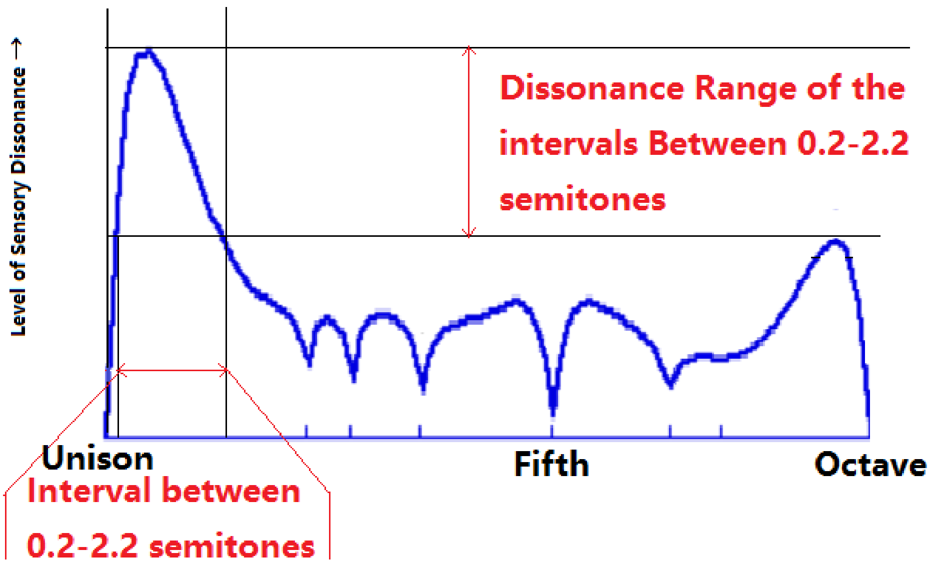


Figure 36 Dissonance effect of 0.2–2.2 semitones

The computation is based on Sethare's numerical approach (Sethare, 1999)

[25] Harmonic tones usually have the beating effect occurred at above the 6th harmonics, where according the beats-based the model, the beating effect is significantly weakened by the lower settings of the amplitudes for higher harmonics of musical tones.

In harmonic-template based theory, the fifth interval is particularly important in determining the harmony of musical chords, as it establishes the root perception at its lower note (Parncutt, 1989). For this reason, it can be used to explain why the major and minor triads (root position) are more consonant than other triadic structures. Furthermore, as the major third interval also has a function to indicate its lower note as the tonal centre (root), harmonic-template based theory also explains why major triads are perceived as more consonant than the minor structures. However, the harmonic-template based theories run into difficulties to interpret the perceptual rankings for harmony for the following cases:

- Consonance and dissonance perception for chords at both root position and inversions are approximately the same (Roberts, 1986).
- Augmented chords are perceived to be more dissonant than the diminished triads (Carlos, 1987)

The theoretical reasons for the above-mentioned shortcomings are this: harmony-template based analysis relies on the information of the root; in the cases the chords are inverted, or for augmented and diminished chords, the root perceptions are relatively weak. Furthermore, the harmonic-template theories are highly sensitive to identifying the consonance concepts, but are weaker in identifying and distinguishing the dissonances in musical chords. This is one of the main intensions for this research: to further analyse and distinguish the perception of tonal dissonance concepts[26] and attempts to use the proposed tonal dissonance concepts: *ambiguity dissonance*, *gloom dissonance* and *tension dissonance* to interpret the perception phenomena those harmonic-template theories fail at (such as the perception of inverted, augmented, diminished and other atonal chord structures).

[26] In this research, the tonal dissonance concept is further categorised into three different functions: ambiguity dissonance, gloom dissonance and tension dissonance.

4.6 Summary

In this chapter, the numerical methods of building the psychoacoustic models according to physiological and psychological hypotheses have been reviewed. In particular, the models for beats-based and harmonic-template based theories are covered in sections 4.3 and 4.4.

The inputs to these models are generally acoustic quantities, such as the amplitude and frequencies of the component partials. The model outputs are a one-dimensional psychological measurement: under beat-based models, the amount of *sensory beats* is inverse-related to the consonance concept; and under harmonic-template based models, the degree of *harmonicity* is proportional to the tonal consonance concept.

The predicted results have shown that both beat-based and harmonic-template-based models have been successful in accounting for dyadic harmony; however, none are able to account fully for triadic harmony. There are also some theoretical problems associated to both beats and harmonic-template based models. In the next two chapters (Chapter 5 and Chapter 6), some modifications are made to the beats-based and harmonic-template based models accordingly, aiming at resolving some of their theoretical issues and achieving better prediction results in line with empirical observations.

Chapter 5 Sensory Dissonance

Reviewed physiological and psychological theories pertaining to music harmony have implied that the semantic meanings of the consonance and dissonance concept (*CDC*) can be multi-fold. The primary objective of this research model is to analyse and predict the consonance and dissonance perception responses with respect to each type of psychophysical mechanisms. The model output thus consists of a number of dissonance concepts (where the consonance is passively defined as the lack of dissonance).

This chapter introduces the first type of dissonance concept in a multi-dimensional dissonance based analysis – *sensory dissonance*. As reviewed previously, the psychoacoustic studies correlating acoustics with sensory dissonance are mainly postulated on the estimation of the *beats effect.* However, the inconsistent prediction results of beat-based computational models have raised many theoretical issues (see section 4.3). In this chapter, a theoretical attempt is made to enhance the beats-based computational model by incorporating the features of *secondary beats effect*. Moreover, in line with the pitch-based analytical thread, the model predicts the amount of sensory dissonance effect with respect to each pitch component. The chapter is structured as follows:

Section 5.1 reviews the theoretical problem of previous computational models in relation to the *secondary beats effect*, from which the concepts of *primary* and *secondary beats effect* are introduced. In Section 5.2, an experimental study is conducted to obtain a quantitative measure of the *secondary beats effect*. In section 5.3, the model input configuration is presented. The model input configuration presented in this session is not only for the analysis of sensory beats, but also for the entire analytical model (multi-dimensional harmonic analysis). The computational method for sensory beats is presented in section 5.4. And in section 5.5 the model estimation result will be compared to previous analytical models. Section 5.6 summarizes this chapter.

5.1 Secondary Beats

Since the 1960's, many analytical models have been developed to account for the auditory perception of consonance and dissonance. One of the main theoretical threads for dissonance perception is the sensation of auditory beats. Sensory beats are perceived as an 'unsettling' and 'disturbing' sound that is usually observed between two pure-tone interactions (see section 4.3.1). According to *tonotopical theory*, the beats effects are attributed to the physiology of the inner ear. The threshold of frequency band that beat effect occurs is considered to be the critical bandwidth (*CB*), therefore when the frequency separations of two pure-tone partials (f_1, f_2) exceed *CB*, the amount of beats sensation (*B*) is expected to be zero:

$$B\ (f_1, f_2) = 0 \text{ when } |\ f_1 - f_2\ | > CB$$

Within *CB*, the amount of beats sensation is characterized by the pure-tone dissonance curves. A typical dissonance curve has a zero value when the frequency separation ($|f_1 - f_2|$) is zero, and increases dramatically to a maximum at $|\ f_1 - f_2\ | \approx 25\%\ CB$; and slopes down to zero at $|\ f_1 - f_2\ | = CB$. A typical curve can be viewed in Fig.37.

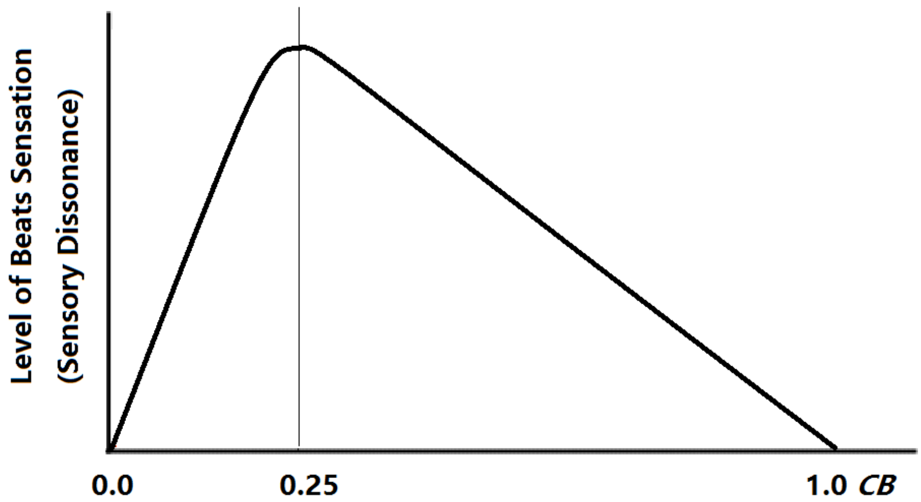


Figure 37 A typical dissonance curve

However, later experimental studies (Hindemith, 1984 and Benade, 1976) had discovered that the beats sensation also occurs beyond the critical bandwidth. More specifically, the beats sensation generally becomes noticeable when the frequency ratios of two pure-tone partials (f_1, f_2) approaches integer frequency ratios, that is:

$B(f_1, f_2) \neq 0$ when $f_1/f_2 = m/n \pm \Delta$

where m, n are small positive integers, and Δ is the frequency difference that is usually less than 30Hz (Parncutt and Hair, 2011).

Therefore, conventional dissonance curves only consider one special case of sensory beats that is when m=n=1; known as *primary beats*. To distinguish from primary beats, other types of sensory beats are known as *secondary beats*. Compare to the *secondary beats*, *primary beats* usually have a stronger sensory effect that can be easily noticed.

The phenomenon of *secondary beats* cannot be explained by the *tonotopical theory*, because they do not require the pure-tone interval to be less than *CB*. For example, when m=1 and n=2, this corresponds to an interval of octave, which is significantly larger than the critical bandwidth (*CB* usually correspond to is 2–5 music semitones depending on absolute frequency range).

Helmholtz had noticed this phenomenon and attempted to interpret the *secondary beat effect* with the non-linear distortions of the ear transfer functions. In particular, he attributed the *secondary beats* to *combination tones*. *Combination tones* are a phenomenon that when two pure-tone partials are played together, a third virtual tone can be noticed, with its frequency equal to $| f_1 \pm f_2 |$. To understand the non-linear distortion mechanism, suppose our ear has a squaring (non-linear) transfer function, combination tones can be derived from following demonstration:

Transfer function (Y) of the input pure tones:

$$Y_1(t) = \sin^2(f_1 t);$$

$$Y_2(t) = \sin^2(f_2 t);$$

The transfer function of simultaneous pure-tone pair is:

$$\begin{aligned} Y(t)^2 &= [\sin(f_1 t) + \sin(f_2 t)]^2 \\ &= \sin^2(f_1 t) + \sin^2(f_2 t) + 2\sin(f_1 t)\sin(f_2 t) \\ &= Y_1(t) + Y_2(t) + \cos((f_1 - f_2)t) - \cos((f_1 + f_2)t) \end{aligned}$$

Therefore, extra terms $\cos((f_1 - f_2)t)$ and $\cos((f_1 + f_2)t)$ are introduced, corresponding to the combination tone frequencies $| f_1 \pm f_2 |$. Helmholtz hypothesized that the virtual combination tones are able to interfere with f_1 and f_2 and therefore produce the secondary beats.

Theorists such as (Wegel and Lane, 1924) preferred to use *aural harmonics* to explain the secondary beats effect. *Aural harmonics* is a phenomenon that when a sine tone is played (over certain threshold loudness), a series of additional virtual harmonic tones can be heard. *Aural harmonics* is also a phenomenon that can be explained by the non-linear processing of the inner ear. *Aural harmonics* correctly predicts the *secondary beat* effect, but it cannot explain that in most case the aural harmonics themselves cannot be heard (Lawrence and Yantis, 1956).

The thread of using non-linear processing of the inner ear to explain *secondary beats* has been abandoned in more recent years. One of the major reasons is that virtual phenomena such as *combination tones* and *aural harmonics* require a high loudness level, whereas *secondary beats* can be heard at relatively low levels of loudness. Moreover, the phenomenon of binaural beats[27] suggests that beats sensation is a higher-level perception phenomenon which cannot be explained by the inner-ear functions. (Warren, 2008) used electroencephalogram (EEG) techniques to read brain activity and concluded that the beats effect originated at the inferior colliculus of the midbrain and the superior olivary complex of the brainstem.

In addition, (Plomp and Smoorenburg, 1968) provides a time-domain explanation to how secondary beats can be perceived by the periodicity patterns of the nerve impulses and this is illustrated in Fig.38. f_1 and f_2 are related such that: $f_1 = 4f_2 + \Delta$. As f_1 contains the impulse trains not only at a frequency of f_1 but also at its sub-harmonic frequencies : $f_1/2$, $f_1/3$, $f_1/4$ … f_1/n, the $f_1/4$ pulse train will highly interfere with f_2 and therefore produce the beat sensation.

[27] Binaural beats is a beats effect produced by the pure-tone interferences between the left and right inner ears.

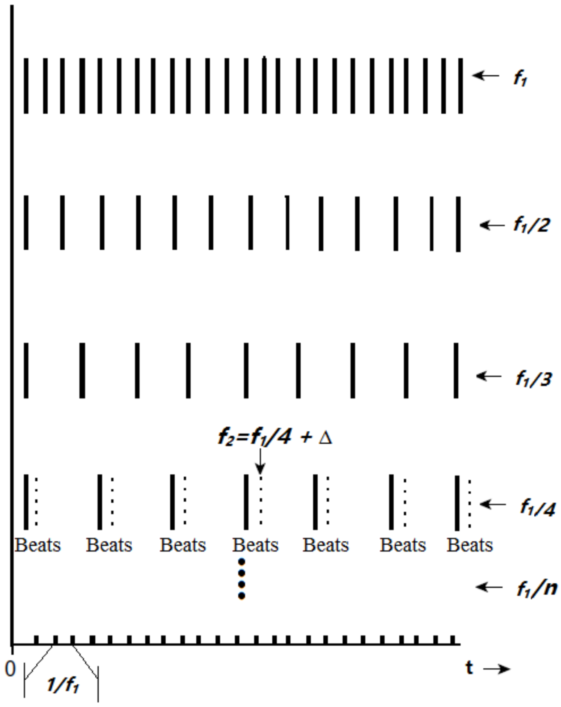


Figure 38 A Time domain analysis of secondary beats

The phenomenon of secondary beats implies that a single pure tone is able to interact with a range of frequencies beyond one critical bandwidth. This effect is however ignored in previous analytical models such as (Plomp and Levelt, 1965) and (Sethare, 1999). To build a model that incorporates secondary beats effect, we must obtain a theoretical pure-tone dissonance curve. Previous research has discovered that the frequency range where secondary beats occurs are $m/n \pm \Delta$ (m, n are positive integers) (Lawrence and Yantis, 1956), but the degree of sensory dissonance has not been explicitly studied. In the next section, an experiment is conducted to explore the pure-tone dissonance curve beyond one critical bandwidth.

5.2 An Experimental Study of Dissonance Curve

A dissonance curve depicts the perceived level of dissonance as a function of pure-tone frequency intervals. According to *tonotopical theory*, the primary beats effect occurs only within one critical bandwidth (*CB*, an interval that is usually less than five semitones for musical notes, see Fig.39). However, the phenomenon of secondary beats means that the beating effect can also be noticed when the pure-tone intervals exceed one *CB*. In this particular test, the perceived degree of sensory dissonance for frequency intervals slightly above an octave (frequency ratio: 2:1) and tritave (frequency ratio: 3:1) were studied based on participants' perception responses.

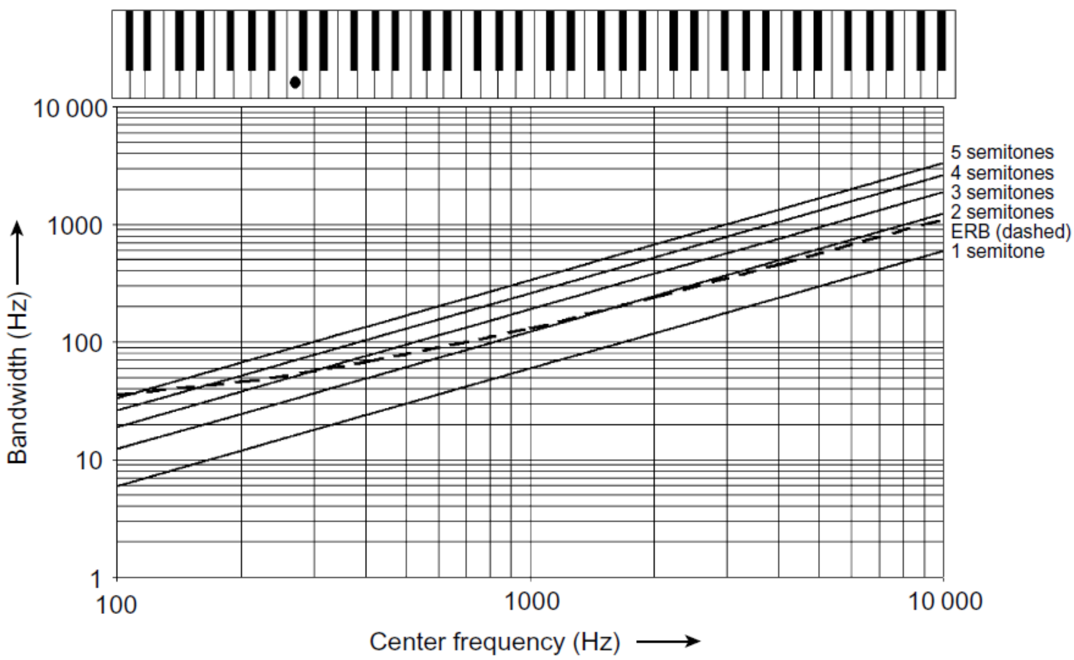


Figure 39 Equivalent rectangular bandwidth (ERB) as a function of centre frequency (Howard, and Angus, 2009)

To synthesise the pure-tone intervals, a bass tone was fixed at 170Hz, and the frequencies of a higher tone categorized into two classes (see Table 12). The first class contains ten frequencies ranging from 340Hz to 385Hz which are used to construct the pure-tone intervals slightly above octave (correspond to audio file #1–10). The second class contains another ten frequencies (correspond to audio file #11–20) but ranging from 510Hz to 555Hz which are used to build the intervals slightly above a tritave (frequency ratio: 3:1). Therefore, 20 audio files were synthesised (see Table 12) in this particular test. As a fixed bass tone (170Hz) is used, the frequency dependencies of the dissonance curve are not studied in this particular test.

Audio #	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Frequency (Hz)	170& 340	170& 345	170& 350	170& 355	170& 360	170& 365	170& 370	170& 375	170& 380	170& 385
Audio #	11	12	13	14	15	16	17	18	19	20
Frequency (Hz)	170& 510	170& 515	170& 520	170& 525	170& 530	170& 535	170& 540	170& 545	170& 550	170& 555

Table 12 Frequency components of the sound files

Participants were asked to listen and rate their perceived level of dissonance for each of the pure-tone pairs. Their rated perception rankings were then used to construct the numerical model of dissonance curve.

Prior to the study, ethical approval was granted from the University of York's Physical Sciences Ethics Committee.

5.2.1 Participants' Profile

To minimize the influence of the cultural and musical background, participants were chosen from both eastern and western culture backgrounds and none of them had undergone any professional musical training. The participants consist of 12 students from the Electronics Department, University of York (UK); and 16 students from the Electrical Engineering Department, Nanjing University (China). The average age of the participants was 21.1 years.

5.2.2 Test Procedure

Prior to the test, a brief introduction was given to each participant, including the objective and the overview of the test. If participants agreed to take part, they signed the consent page and proceeded to the test.

In this test participants heard synthesised sounds via earphones. A computer application interface was designed to present the test to the participants and also to collect the experimental results from participants. The entire test consisted of three pages.

On the first page (see Fig.40), participants heard three audio excerpts. The frequency component of each audio file could be modified by sliding the horizontal bar. Participants were asked to adjust the horizontal bar to a point where the maximum dissonance is perceived. They could then proceed to the next session by clicking the button 'save and proceed'.

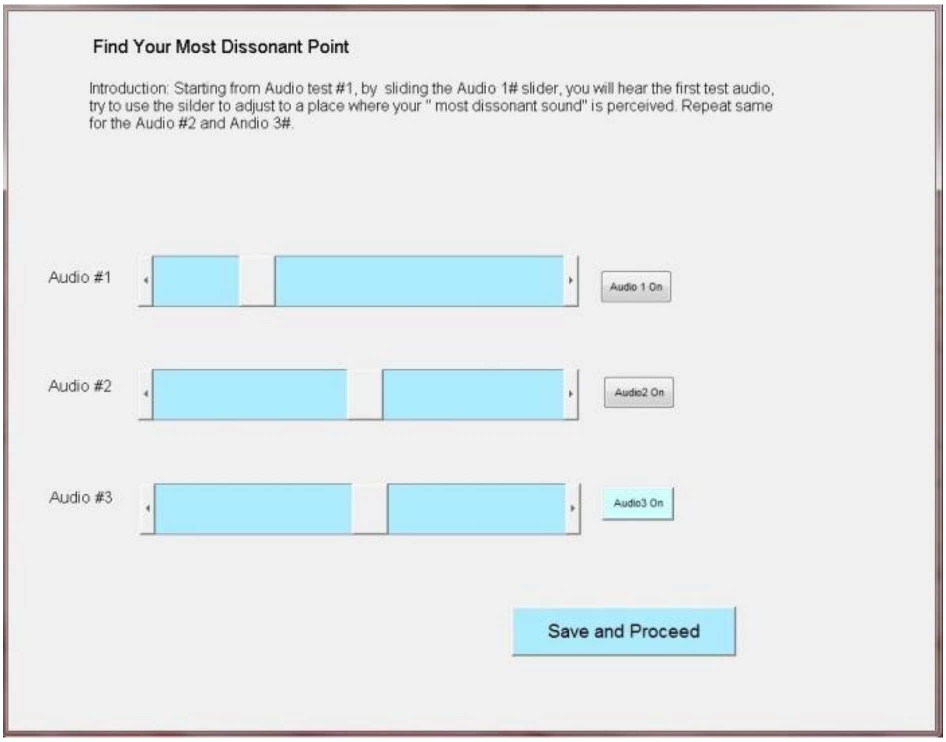


Figure 40 Page #1 of GUI for the sensory dissonance test

On the second page (see Fig.41), participants saw two tests. For each test, the participant heard ten audio files (S1–S10) by clicking respective buttons. They were asked to rate their perceived level of dissonance by sliding the vertical bars (to the top is the most dissonant and to the bottom is the least dissonant). The leftmost sliding bar is fixed at the top (S_0), it serves as a reference of the most dissonant sound; during the test, participants were free to listen and refer to this sound in order to fully utilize the scale of vertical bar.

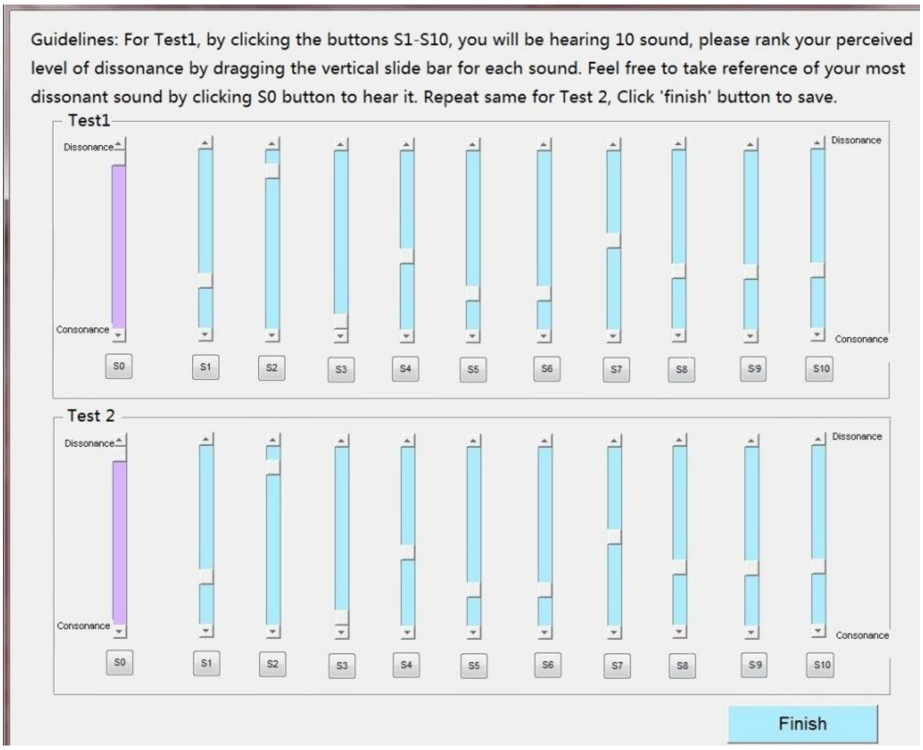


Figure 41 Page #2 of GUI for the sensory dissonance test

On the last page (see Fig.42), participants were asked to compare three specific sounds (sound #1–3) and rate the level of dissonance using the sliding horizontal bars. The sound could be activated by either pressing the button (sound #1–3) or sliding the horizontal bar. The entire test was completely when the ‘finish’ button of this page was clicked.

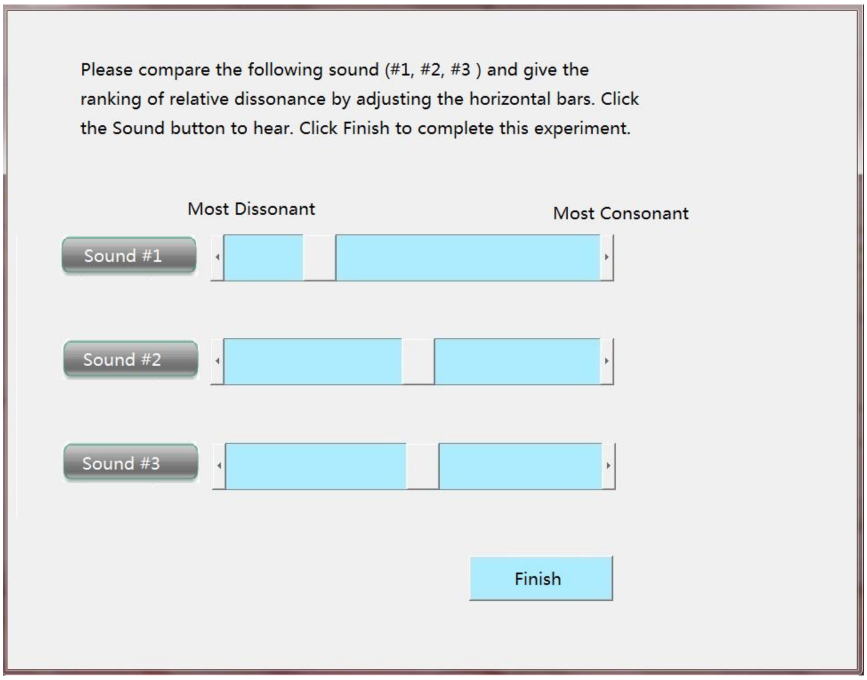


Figure 42 Page #3 of GUI for the sensory dissonance test

5.2.3 Analysis of the Test

Among the three pages of tests, the second page was the main test used to the construct the dissonance curve. The ten sound files in test 1 correspond to audio file #1–10 (340–385Hz +170Hz pure-tone pairs), and the ten sound files in test 2 correspond to audio file #11–20 (510–555Hz +170Hz pure-tone pairs). When the page is loaded, audio files #1–10 are randomly assigned to S1–S10 in test #1; and the audio files #11–20 are randomly assigned to S1–S10 in test #2. The left-most sound in test 1 and test 2 were obtained from the first page, corresponding to the most dissonant sound set by each participant.

In the first page, Audio #1, 2 and 3 are composed of a 170Hz pure tone and another pure-tone depending on the position of sliding bars:

Sliding Bar Audio #1: 170–230Hz

Sliding Bar Audio #2: 340–385Hz

Sliding Bar Audio #3: 510–555Hz

The frequency resolution of each sliding bar is set to 1Hz. The test in page one serves two purposes. The first purpose is for data validation: Audio #1 on page one corresponds to interval range of primary beats. According to *tonotopical theory,* the maximum of dissonance happens at about 25% of critical bandwidth, which is roughly at 7% from the left of horizontal bar. Therefore, participants were expected to set this value at around 7% from the left of horizontal bar. If participants set this value less than 3% or more than 12%, their perception data in page 2 was excluded from the final analysis. The second purpose is to note the participant-specified most dissonant pure-tone intervals for the test in the next page (the S0 audio in test 1 and test 2).

Page 3 compares the three most dissonant pure-tone pairs obtained from the first page. This is an across-comparison between the effects of primary beats, secondary beats slightly above octave, and secondary beats slightly above tritave. The relative value can be used to scale the level of sensory dissonance between primary beats, and secondary beats effects (octave and tritave).

Result from page 2&3 were recorded and combined for statistical analysis.

5.2.4 Results

23 out of 28 subjects passed the data validation process from the test in page 1. The experimental results for test1 and test 2 on GUI page 2 are shown in Fig.43 and Fig.44. The mini horizontal bars represent the rated averaged level of dissonance of the 23 participants. The vertical scale 10 represents the highest level of sensory dissonance that has been assigned by participants. A polynomial regression curve is also fitted to demonstrate the trend of the dissonance curve from discrete data points.

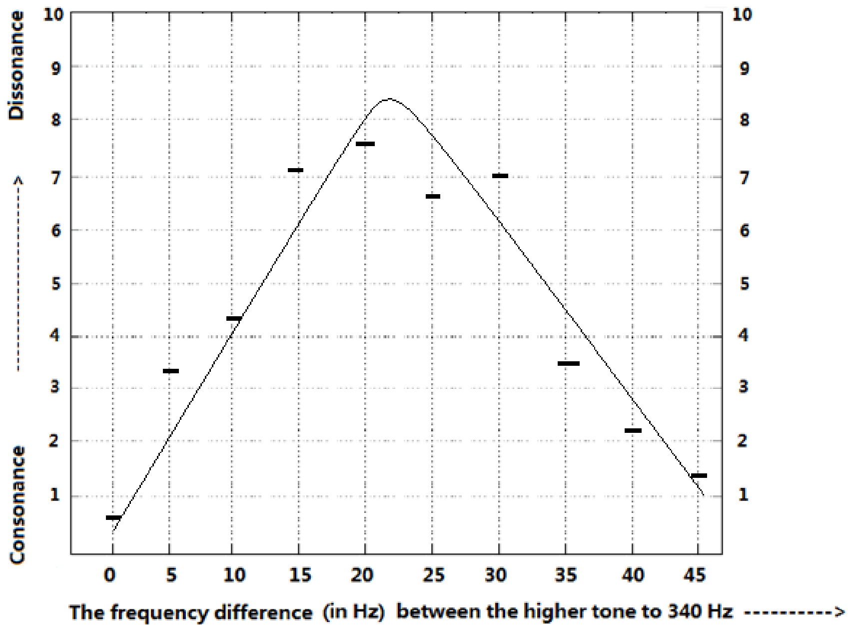


Figure 43 Experimental result for intervals above one octave

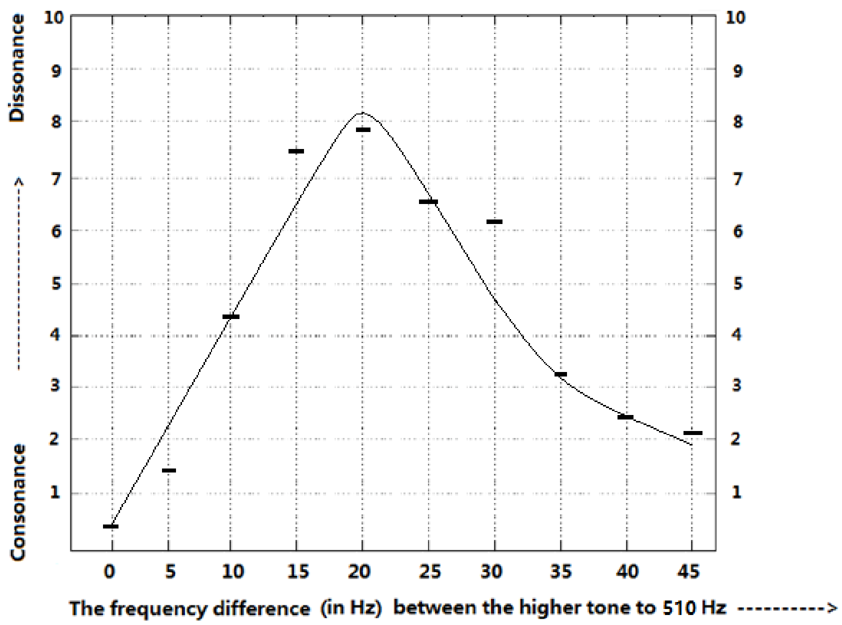


Figure 44 Experimental result for intervals above one tritave

The test results from test page 1 and page 3 are combined and shown in Fig.45 below.

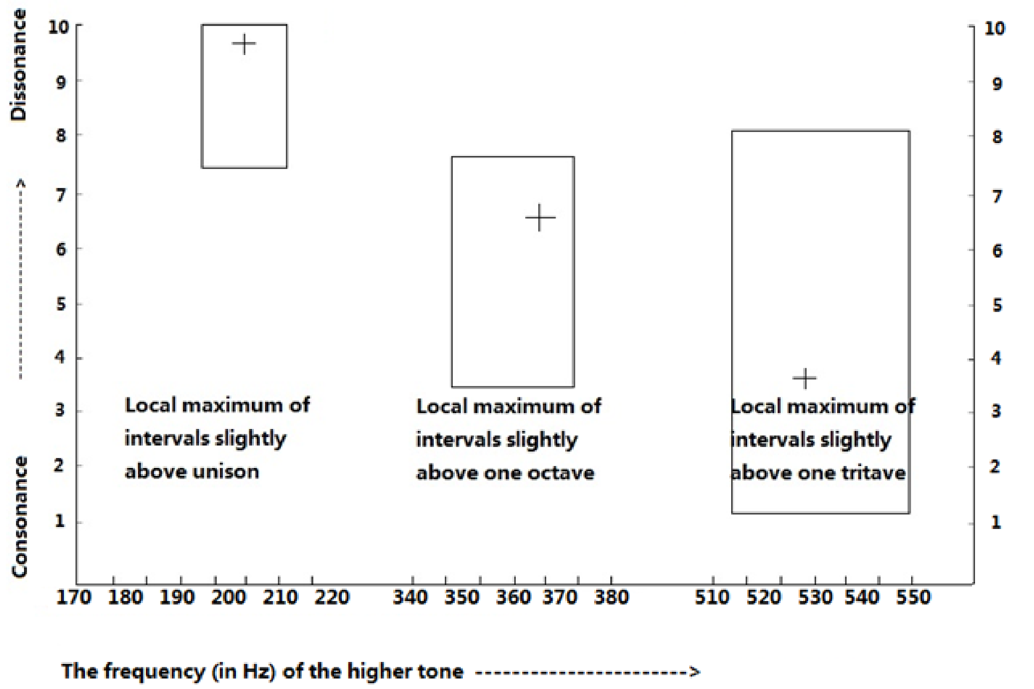


Figure 45 Cross comparison between local maxima

In Figure 45, the local maximum of sensory dissonance between primary beats, secondary beats slightly above an octave and secondary beats slightly above a tritave are compared. The rectangular box shows the boundary of both the interval of maximum dissonance (obtained from test in test page 1) as well as the relative strength of the sensory dissonance (obtained from test in application page 3). The cross represents the numerical centre of the measurements: treating the maximum sensory dissonance effect of primary beats as 10, the maximum dissonance sensory effect of primary beats, secondary beats of intervals slightly above octave and tritave are measured at 9.83, 6.78 and 3.65 respectively.

Lastly, the cultural background of the participants does not demonstrate any significant influence on the experimental result in all aspects.

5.2.5 Conclusion of the Experimental Test

Based on the experimental results from Fig.44 and Fig.45, a dissonance curve similar to that of primary beats (see Fig.37) can be observed for the intervals slightly above octave and tritave:

- At the octave and tritave interval, a minimum sensory dissonance is observed;
- The level of sensory dissonance starts to increase as the frequency separations are increased;
- A local maximum is reached at the frequency separation roughly between 10-30Hz above octave and tritave intervals;
- And further frequency separation decreases the level of sensory dissonance effect.

Across comparing the sensory dissonance effect between primary beats and secondary beats, we may observe that the sensory dissonance effect is inversely related to the complexity of frequency ratios:

Primary beats happen slightly above unison interval (frequency ratio m:n = 1:1) and have the highest level of sensory dissonance effect.

Secondary beats happen slightly above octave interval (frequency ratio m:n = 2:1) and have the medium level of sensory dissonance effect.

Secondary beats happen slightly above tritave interval (frequency ratio m:n = 3:1) and have the lowest level of sensory dissonance effect.

The experimental results verified the fact that distant partials may also introduce beats sensations (secondary beats effect). The phenomenon of secondary beats also implies that the harmonic timbres are in nature more consonant than the non-harmonic timbers. This is because the partial components of harmonic timbre have simple integer ratios with respect to each other, therefore the least sensory beats are introduced; whereas for non-harmonic timbres, greater level of sensory dissonance is expected due to the secondary beats effect. Such implications are in line with the Harmonic Relation Theory proposed by Jean Philippe Rameau and Wilhelm Wundt (Rameau and Wundt, 1721) in accounting for the music harmonies.

5.3 Estimation of Sensory Dissonance

In the experimental test (section 5.2), it was shown that the dissonance curves at the intervals slightly above octave and tritave (frequency ratio m:n = 2:1 and 3:1) are similar to primary beats, and the sensory effect generally decreases when the frequency ratio m: n gets more complex. Therefore, in our model, the dissonance curve for secondary beats effect is derived from the primary beats.

According to Sethere's numerical model (Sethare, 1999) of primary beats, the level of sensory dissonance d for two pure tone components (f_1 , f_2) can be model by (assuming the two pure-tone partials both have unit amplitude):

$$d\,(f_1, f_2) = e^{-3.5s(f_1 - f_2)} - e^{-5.75s(f_1 - f_2)}$$

where s depends on the bass frequency(f_1) of the pure-tone pair:

$$s = 0.24 / (0.021 f_1 + 19)$$

An example of such dissonance curve d (f_1, f_2) (f_1 = 260Hz) can be visualized in Fig.46:

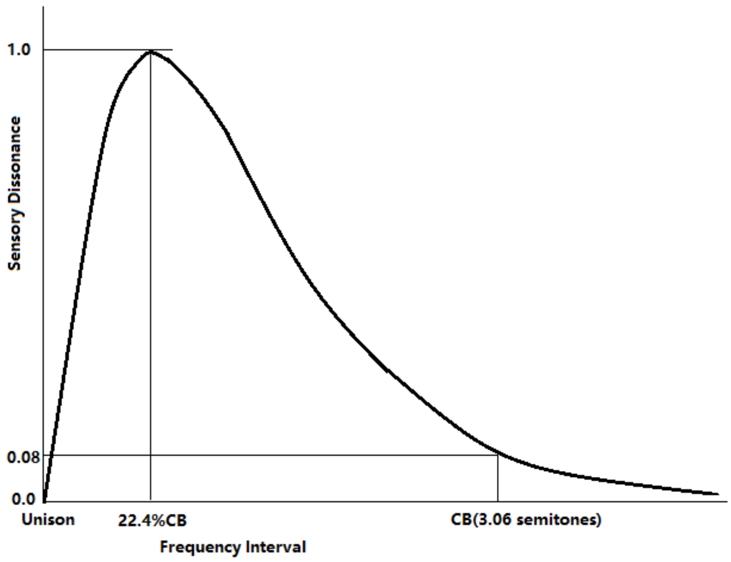


Figure 46 Sethere's dissonance curve with base frequency (f_1=260Hz).

For this particular plot, the maximum sensory dissonance occurs at the interval of 22.4% CB (critical bandwidth) and less than 0.08 unit of sensory dissonance is modelled for frequency intervals larger than CB. Such curve is quite in line with the *tonotopical theory*.

According to the hypothesis that secondary beats effect is inversely related to the complexity of integer ratios, the secondary beats (d_{2nd}) can be modelled by dividing primary beats $d(f_1\ f_1/m,\ f_2/n)$ by the product of m and n (therefore the increment of either m or n will decrease the overall beats sensation):

$$d_{2nd}\ (f_1, f_2, m, n) = \mathrm{d}(f_1\ f_1/m, f_2/n)\ /m\ n$$

where d_{2nd} estimates the secondary beats effect introduced by the integer ratio m/n. When $m = n = 1$, the above formula simply estimates the primary beats.

The sensory dissonance curve thus can be extended to the frequency intervals beyond tritave, resulting in a multi-peaked dissonance curve. Fig.47 demonstrated such a dissonance curve considering the secondary beats effect at the frequency intervals slightly above octave (m=1, n=2) and tritave (m=1, n=3).

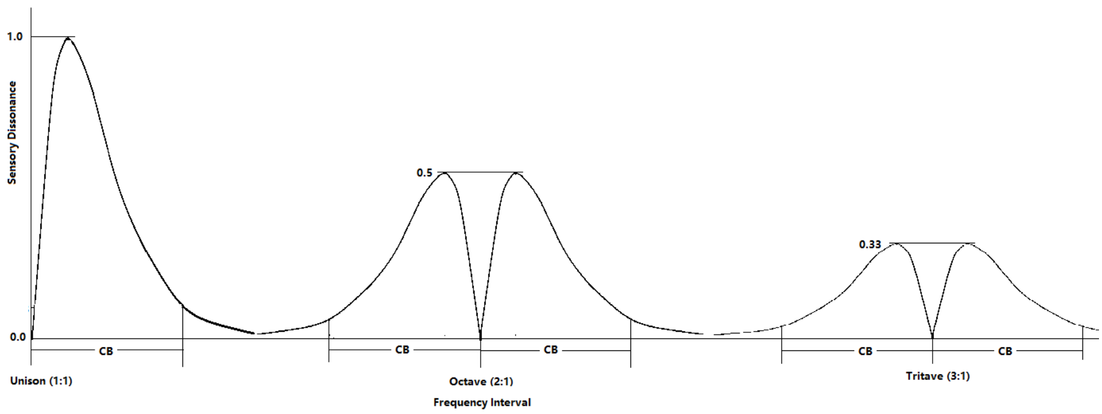


Figure 47 A multi-peaked theoretical dissonance curve with secondary beats effects

As the secondary dissonance effect is significantly weakened when n or m is increased, this model only consider the secondary beats effect at near the octave (m =1, n =2) and tritave intervals (m =1, n =3).

The sensory dissonance effect D_s for an input sound entity can be estimated by summing up the pure-tone sensory effect over its partial components:

$$D_s = \sum^{i} \sum^{j} d_{2nd}(f_i, f_j, m,n) \cdot A_{fi} \cdot A_{fj}$$

Where A_{fi} and A_{fj} are the audibility of two audible pure-tone components; a higher audibility contributes to a higher level of beats sensation. The term $\sum^{j} d_{2nd}(f_i, f_j, m,n) \cdot A_{fi} \cdot A_{fj}$ computes the sensory dissonance introduced between a frequency partial (f_i) and all the rest of partials (f_j), and such effect is further summed over all frequency partials (f_i) in order to obtain the overall level of sensory beats. Where m= 1 and n=1, 2, 3 accounts for the primary beats, and secondary beats occurring near the octave and tritave intervals respectively.

5.4 Model Prediction Results for Musical Dyads and Triads

In order to estimate the overall sensory dissonance effect for complex tone scenarios, a linear summation algorithm is applied:

$$D(m,n)= \sum^{j} \sum^{i} d_{2nd} (f_i,f_j,v_i,v_j,m,n)$$

Where d_{2nd} is the dissonance function of two pure-tone partials whose frequencies are f_i,f_j and amplitudes are v_i,v_j. The dissonance effect is summed over all possible combinations of pure-tone pairs within the sonority. The overall sensory dissonance D_s is therefore:

$$D_s = D\,(1,1) + D\,(1,2) + D\,(1,3)$$

Where $D\,(1,1)$, $D\,(1,2)$ and $D\,(1,3)$ correspond to primary, secondary beat effect near octave interval and secondary beat effect near tritave intervals.

However, it is worthy of note that using the summation of pure-tone sensory dissonance to estimate the overall sensory dissonance is quite a simplified approach. As indicated by (Huron, 1996), sensory dissonance is rarely observed in musical structures, and the summation of sensory dissonance effect on the pure-tone level may not be reflected for complex tone scenario. Moreover, the sensory effect between pure-tone interactions can also be masked by other sensory beats (Taylor et al., 1974). The complex function of the higher auditory pathway such as central auditory integration and brain activities may be much more complicated than linear summation can model.

Musical dyads and triads are usually composed of harmonic complex tones. In this study, harmonic tones are synthesised using a harmonic frequency spectrum (S_h):

$$S_h (f_0) = \sum^n 0.8^{(n-1)} sin(2n\pi f_0)$$

where f_0 is the fundamental frequency of the harmonic tone; n is the number of overtone partial (n = 1–15). The amplitude of the n^{th} partial is decayed by a factor of $0.8^{(n-1)}$.

The frequency spectrum of $S_h (f_0)$ can be visualized in Fig.48.

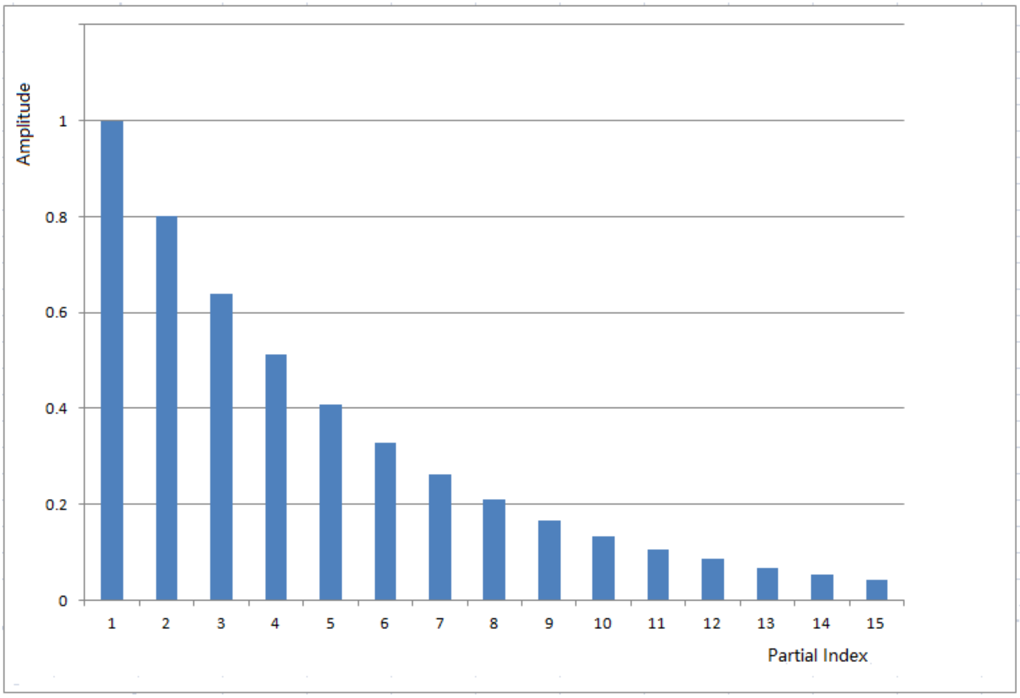


Figure 48 Frequency spectrum of S_h

In the following sections, S_h will be used to study the model predictions of musical dyads and triads.

5.4.1 Sensory Dissonance for Musical Dyads

To visualize the model prediction of the sensory dissonance for the musical dyads between the musical interval of Unison and Octave, a bass note at C_4 (fundamental frequency at 261Hz) is used, and the higher note has the fundamental frequency ranging from 261Hz to 522Hz (C_5). The Matlab simulation of predicted level of dissonance for these intervals gives a result below (see Fig.49).

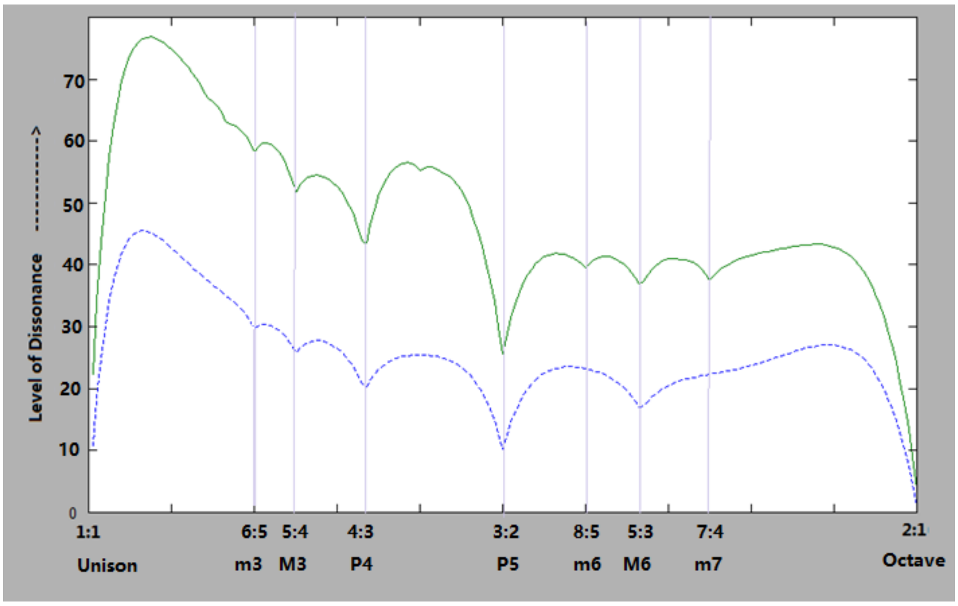


Figure 49 The model predictions for the dyadic intervals between unison and octave

The solid curve above is the model prediction that considers the secondary beats effect; and the dotted curve below is the prediction considering only the primary beats. The frequency resolutions of both curves are 1Hz. The frequency ratios of a few common music intervals are shown on the horizontal axis.

Based on the observation from Fig.49, the prediction result from present research model with secondary beats effect generally has a higher absolute value than previous prediction results which consider the primary beats only. However, in the prediction result of the present research model (with secondary beats effect), local minimums of the consonant intervals such as the fifth (frequency ratio 3:2) and fourth intervals (frequency ratio 4:3) are further sharpened. Furthermore, the present model is able to highlight a few extra local minimums that previous models are unable to discover, such as the dips occurring at frequency ratios of 8:5 and 7:4.

The prediction results from the present research model with secondary beats effect provide a correct prediction for the consonance of diatonic musical intervals such as fifth, fourth, major third/sixth. In addition, the present model is able to indicate the intervals of minor sixth (frequency ratios of 8:5) and minor seventh (frequency ratios of 7:4) where models based on primary beats were unable to predict under the same harmonic tone configuration (S_h).

5.4.2 Sensory Dissonance for Musical Triads

The model prediction results for major, minor, diminished and augmented triads and their inversions are studied in this section. Musical triads are synthesised with the bass note fixed at C_4 261Hz. The predicted rankings of sensory dissonance are presented in Table 13 in comparison to the empirical rankings as well as other sensory-beats based models.

Triad Classes	Intervallic Structure	Empirical Rankings	P&L	K&K	Sethare	Sensory Dissonance Model in This Research
Major	3-4	1	3	1	3	1
	5-3	2	8	8	5	6
	4-5	3	1	3	1	3
Minor	4-3	4	3	1	3	2
	5-4	5	1	3	1	5
	3-5	6	8	8	5	8
Diminished	3-3	7	10	3	9	4
	6-3	8	5	6	7	9
	3-6	9	5	6	7	7
Augmented	4-4	10	7	10	9	10

Table 13 Model predictions for common triads

The data of Plomp & Levelt (1965), Kameoka & Kuriyagawa (1969) and Sethares (1999)'s models are copied from Table 9 in Chapter 4. The last column is the prediction result of the sensory beats in this research. The inconsistent prediction results are highlighted in dark grey backgrounds

From Table 13, it can be observed that none of the models (including the model presented in this research) are able to fully predict the empirical rankings. The computational model incorporating the secondary beats effect has a slightly better prediction for the minor and augmented triads, but generally fails to reproduce the empirical trend of consonance of major > minor > diminished regardless of the inversions of triads.

5.5 Summary of Chapter

The secondary beats effect is a theory often ignored in previous sensory beats-based models. There are two main reasons for this: firstly, secondary beats effect is relatively weak when comparing to primary beats effect; secondly, previous analytical models have been influenced by the *tonotopical theory* where sensory beats were limited to the frequency intervals within a critical bandwidth (primary beats).

In this chapter, a theory enhancing sensory-beat based analysis by incorporating the secondary beats effect was proposed. To this end, an experimental listening test was conducted to obtain a quantitative measure of the secondary beats effect. Then a multi-peaked dissonance curve was theorized based on the experimental result. With the use of such multi-peaked dissonance curve, a summation algorithm was used to predict the sensory dissonance for complex tones such as musical dyads and triads.

The model prediction result is successful in accounting for the consonance of common musical intervals such as perfect fifth, perfect fourth and major third, but fails to produce a prediction result in line with the common perception trait of musical triads. A possible explanation of this failure is that music perception is perhaps more complex than auditory sensory response; higher-level psychological mechanisms may be involved for the perception of musical triads.

In the next chapter, another theoretical attempt is made from a different perspective – the tonal *CDC*. This approach is explored as accounting for the musical perception of musical triads where the sensory dissonance based models had failed.

Chapter 6 Ambiguity Dissonance

This chapter introduces the second type of dissonance concept in multi-dimensional-dissonance-based analysis – *ambiguity dissonance*. The meaning of *ambiguity dissonance* is opposite to the *tonal consonance* concept under harmonic-template based theories; it refers to musical entities with vague or ambiguous tonal centre perceptions. The perception mechanism of *ambiguity dissonance* is fundamentally different from that of *sensory dissonance*, as it reflects the psychological aspect of sound perception rather than the physiological functions of auditory pathway. For that reason, *ambiguity dissonance* is estimated based on the interaction of cognitive features rather than acoustical features. In our approach, *noticeable pitch/note components* and the *tonal centre concept* are the two main cognitive features involved; numerical methods extracting these features from acoustical input are introduced in section 6.1 and 6.2 respectively. With the use of these cognitive features, a numerical model is introduced to estimate the *ambiguity dissonance* effect (in section 6.3). The main points of the *ambiguity dissonance* model are summarized in section 6.4 of this chapter.

6.1 Noticeable Pitch Components

In cognitive-based music harmony theories, perceived consonance and dissonance are discussed based on tonal structures instead of acoustical features. Therefore, it is important to convert acoustic input into a set of audible tone components for psychological-based computational model.

For a musical entity, the noticeable pitch components usually correspond to its musical notes. However, the perceived level of salience of each note component can be different for each note component. In this section, a numerical model that estimates the perceived level of salience with respect to each audible pitch component is presented. This numerical method is designed not only for analysis of tonal musical structures but for any general sound object.

The psychoacoustic theories pertaining to pitch perceptions have been well established. The modelling of pitch perception for a multi-tone simultaneity is however not quite consistent in the literature, presumably due to the complexity between critical bandwidth and mutual masking effect (Parncutt, 1989). In this research, a three-step model is proposed to approximate this perception process (see system diagram in Fig.50):

(1) Audibility of tone components (section 6.1.1)

(2) Mutual masking effects (section 6.1.2), and

(3) Harmonic masking (section 6.1.3)

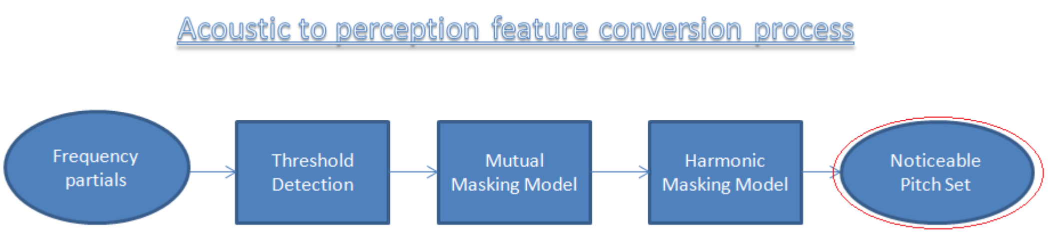


Figure 50 Acoustic to perception feature conversion diagram

6.1.1 Audibility of Tone Components

At this stage, the amplitude (A_m, in dB form) of each partial component (f_i) is subtracted by their corresponding hearing threshold in order to obtain the audibility of each pure-tone partial, $A(f_i)$:

$$A(f_i) = \max \{ A_m(f_i) - T_q(f_i), 0\}$$

Where $T_q(f_n)$ the threshold of hearing, fitted by the numerical formula of (Terhardt, 1979):

$$T_q(f) = 3.64 (f)^{-0.8} - 6.5\exp(0.6(f-3.3)^2) + 10^{-3}(f)^4$$

This function can be visualized in Fig.51 below:

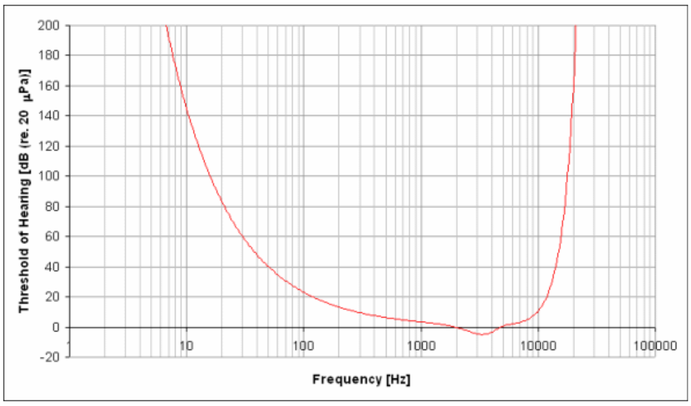


Figure 51 threshold of hearing as a function of frequency, taken from (Terhardt, 1979)

The max function is used to ensure the audibility is always equal or greater than zero.

6.1.2 Mutual Masking Effects

Once the audibility of each pure-tone partial is obtained, the mutual masking effect can be considered. The mutual masking effect can be viewed as a mutual inhibition between each pair of adjacent partials (when they fall within same critical bandwidth). As a result, the audibility of each partial component is decreased by a certain amount. In the literature, the modelling of the mutual masking effect is just as difficult as that for beat sensations (Parncutt, 1989). As a reduction mechanism, the mutual masking level of a particular partial, $ML(f_i)$ is modelled by summing up all the masking effects caused by other partials (f_j):

$$\mathrm{ML}(f_i) = \sum^{j} ml\,(f_j, f_i);$$

Where *ml* is the masking level between two pure-tone partials. According to Parncutt and Terhardt (1989), each pure tone component at a moderate loudness partially masks every other pure-tone components to a maximum frequency range of 3 critical bands, the masked level of partial A and partial B is modelled by:

$$ml\,(f_a, f_b) = \mathrm{k_M} \cdot |\, f_a - f_b \,|; \text{ (Terhardt et al. 1982)}$$

where $\mathrm{k_M}$ is the masking parameter that has a typical value of 25dB/cb (cb = critical bandwidth) according to (Parncutt, 1989).

Therefore, the level of audibility of partial n $\mathrm{A}(f_n)$ after the mutual masking effect is:

$$\mathrm{A}(f_i) = \max\,\{\mathrm{A}(f_i) - \mathrm{ML}(f_i), 0\}$$

Again, the max function here is to avoid possibility of negative audibility.

6.1.3 Tonal Masking Effects (Audibility of Complex-Tone Sensations)

Tonal masking effect refers to the perception phenomenon where the pitch of harmonically related pure-tone partials will not be perceived individually; instead, they will produce a ‘fused tone’. At this stage, the number of partials will be essentially converted into a number of audible pitches. It is foreseeable that the number of audible pitches will be significantly smaller than the number of pure-tone partials when the input sound is composed of musical tones.

To determine the noticeable pitches, the algorithm first sorts the audibility of pure-tone partials from highest audibility to lowest. Running from the first partial A (f_1), the one with highest audibility A), find pure tone partials that are ‘completely masked’ by the harmonics of A (f_1) by the ‘threshold of complete masking’ function (T_c):

$$T_c\,(f_{\mathrm{masker}}, f_{\mathrm{test}}) = \begin{cases} 1; \text{ if } n\, f_{\mathrm{masker}} = f_{\mathrm{test}} \text{ and } \mathrm{A'}_{\mathrm{masker}} \cdot \mathrm{k_P}\, /n - \mathrm{A'}_{\mathrm{test}} > 0; \\ 0; \text{ otherwise;} \end{cases}$$

Where $n\, f_{\mathrm{masker}} = f_{\mathrm{test}}$ checks if f_{test} is the one of the harmonic frequencies of f_{masker}, a tolerance threshold δ (δ =1%) is used to check this equality; the harmonic partial index n is running from 2 to 16. $f_{\mathrm{masker}} \cdot \mathrm{k_P}\, /n - \mathrm{A'}_{\mathrm{test}} > 0$ checks if the audibility of the masker is able to completely mask another partial at its n^{th} overtone frequency. k_P is the harmonic decrement parameter, taken at 5dB/harmonic herein.

If T_c =1, meaning the test partial f_{test} is completely masked by the masker partial f_{masker} in terms of the complex tone sensation, then the A(f_{test}) is removed from the data array A(f_{i}).

After the harmonic masking effect of all the partials has been considered, the remaining audibility of pure-tone partials (f_m) are considered to be the audible pitch components (complex tone sensations) $A_c(f_m)$. The fundamental frequencies of these audible pitch components are denoted by f_m.

6.1.4 Tonal Functions and Pitch Classes

In the analysis of music structures, it is particularly useful to classify pitch components $A_c(f_m)$ into *pitch classes*[28] (P_c), as a pitch component belonging to each *pitch class* shares a similar tonal harmonic function. The following procedures are involved in order to identify the pitch class of a particular pitch component:

1. Get the fundamental frequency of the tone component: f_i.

2. Find the frequency with the simplest integer ratio between f_i and the fundamental frequency of tonal centre f_{c}. To find the simplest integer ratio, we have to firstly define a range of ratios within the tolerance threshold δ (δ =1%), that is between $(1\text{-}\delta)fi/f_0$ and $(1+\delta)fi/f_c$. Two integer variables *m* and *n* will be used in a searching algorithm for the simplest integer ratio R_s. A pseudocode for this process is presented as following:

Main: if fi > f_{c}, $f_{\text{i}}/f_{\text{c}} = k_r$; else $f_{\text{c}}/f_{\text{i}} = k_r$ (this is to guarantee the constant k_r always >=1)

loop #1: for *m* = 1 to T_m (T_m is the largest number allowed in the algorithm T_m is set to 200 in this research)

loop #2: for *n* = 1 to T_m

[28] In music, a pitch class (p.c. or pc) is a set of all pitches that are a whole number of octaves apart

if (1- δ) k_r <n/m < (1+ δ) k_r return m and n; break; (this particular setting guarantee the numerator has the highest priority to have a small integer value)

If $f_i > f_c$, $R_s = n/m$, else $R_s = m/n$.

3. Find the corresponding pitch class P_c from the simplest integer ratio. In this particular approach, the pitch class (P_c) is characterized by a frequency ratio bounded between 1 (unison) and 2 (octave). Therefore the primary task is to convert R_s into a range between 1~2 based on the principle of octave equivalence:

If $R_s < 1$ then $R_s = 1/ R_s$;

Loop: for i = 0 to 10

If $1< R_s / 2^j <=2$, then $P_c = R_s / 2^j$;

An example may help to understand the above-mentioned algorithm. Assuming we want to compute the pitch class P_c for a tone component whose fundamental frequency is 228Hz, when the fundamental frequency of the tonal centre is 100Hz. According to the algorithm in step 2, the simplest frequency ratio R_s will be 16/7; and the pitch class P_c of this tone is therefore $16/ (7\times2^1) = 8/7$.

Note that a pitch class does not necessary correspond to one of the 12-tet musical intervals. For instance, the pitch class with frequency ratio 8:7 in the above example correspond to a music interval between major second and minor third. There are ideally an infinite number of pitch classes within an octave interval.

6.1.5 Multiplicity

The *multiplicity* (M) concept is simply the number of noticeable pitches of the input sonority:

$$M = \text{size}\ (f_m)/2;$$

In the analysis of music harmony, the *multiplicity* is also an important factor that we need to consider. In both beats and harmonic-template based approaches, it is observed that as the total number of tones increase, the predicted dissonance or equivalent concept will be inevitably increased. Ideally, silence or a single partial has the most consonant predictions as no beating or ambiguity of tonal centre is predicted. However, this is contradictory to the music observation where, under certain conditions, a musical chord consist of notes from different pitch classes sounds more consonant than a single tone (either harmonic complex tone or simply a pure tone) (Huron, 1996). Therefore, the *multiplicity* is identified as an extra impact factor in the harmony perception of musical chords. Usually, when two chords contain the same amount of *multiplicity*, their perceived level of consonance can be compared.

6.2 Tonal Centre Concept

The concept of *Tonal Centre* has been one of the most important cognitive features that can influence the consonance and dissonance perceptions in music harmony. As reviewed in section 3.3, many theories involve the discussion of acoustical and cognitive structure with respect to a tonal centre. For a simultaneous sound entity (characterized by frequency spectra), a frequency partial component or pitch component is generally considered as consonance when it is harmonically related to the tonal centre of that sound simultaneity.

Such observation can also be extended to non-simultaneous musical entities. For instance, in *tonal harmonic hierarchy,* the consonance and dissonance property of a specific tone/chord component is considered in relation to the tonic key provided by the tonal context. Cognitive music theories believe that the meanings of a musical entity are never absolute and fixed, but conveyed "*only through the meaning of a whole, through their tonal relations within the tonal structure*" (Langer, 1942).

Conventional psychoacoustic approaches tend to exclude the influence of tonal context for several reasons. First of all, music harmony is more commonly known as the 'vertical' aspect of music (Howie, 1976), meaning the study of the sound object should be limited to a specific time instant, therefore implying that the context should not be considered. Secondly, as mentioned in the previous section, current psychoacoustic models are typically supported by physiological insights. The physiological system implies the model predictions are limited to the auditory sensory response which is rather absolute and deterministic: the sensory responses only depend on the acoustic structure of the input sound stimulus. Therefore, the musical context becomes insignificant. Lastly, to simplify the model computations, numerical models typically study the output response by isolating one variable at a time. In order to study the perception of consonance and dissonance concept (*CDC)*, previous models often require the object sound stimulus to be a simple isolated entity, ideally independent from musical context.

Music perception is a relatively complex process; it involves both physiological sensory response and higher-level cognitive activities (Stevens, 1957). The tendency for conventional psychoacoustic models to focus solely on sensory aspects could therefore contribute to the limitations of previous model prediction results. In practice, musical chords are rarely used and perceived as isolated entities. Even when listeners were asked to rate the perceived consonance level for a set of single chords, comparisons and therefore context is involved (Arthurs and Timmers, 2016). It is generally hard for listeners to give a sensory-based judgement without comparing one chord to another. For this reason, we hypothesize that the interaction between sensory responses and pre-stored tonal music context is possibly one of the higher-level cognitive processes that is crucial in the process of music harmony perception. To implement this process, the next section introduces a psychoacoustic approach to model the key features of tonal music context, followed by a computational method that estimates the perceived harmony of the input chord structure in the subsequent section.

6.2.1 A Numerical Model for Tonal Context

Tonal context can be viewed as a simple causal system containing a set of noticeable pitch components[29] within a given time interval. Each pitch component is further associated with an *impact* coefficient, representing the extent to which that particular tone left an impression on the memory. According to (Huron, 1993), many factors could influence this impact coefficient, such as loudness, tone duration, masking effect, time elapsed, etc. The impact coefficient for the pitch components with the same pitch perception can be grouped into one by summing their impact coefficients. The entire tonal context can be thus modelled by the impact coefficient array as a function of musical notes.

In tonal music, the tonic of the musical key is one of the most important features of the tonal context. In reference to the tonic, the perception (known as the *harmonic functions*) of all other pitch components can be determined. Therefore, it is

[29] The psychoacoustic modelling of acoustic-to-noticeable-pitch conversion is designed according to Parncutt's model (Parncutt, 1983)

particularly important to study and define the tonic concept in psychoacoustic terms. In the literature, a high correlation can be found between the *key-finding algorithm* (Krumhansl, 1997) and *virtual pitch determination algorithm* (Terhardt, 1974), they share common features like: (1) they both contain a tonal centre concept: for a key-finding algorithm, this refers to the tonic key, for the pitch perception algorithm, it represents the pure tone sensation with highest pitch (virtual) salience; (2) pitch components which are 4 or 7 semitones above the tonal centre concept contribute to the establishment of the tonal center concept; (3) other pitch components generally contribute less to the key or virtual pitch perception. Therefore, in the present approach, the tonal centre concept (with fundamental frequency f_c) of a given musical context is estimated using an algorithm akin to Terhadt's virtual pitch determination algorithm, which consists of following major steps:

1. Listing all the pitch candidates present in the tonal context model.
2. For each candidate (object candidate),

   2.1 Find how well another pitch candidate (test candidate) supports the tonality of the object candidate by computing the *correlation coefficient* between candidate pitch component and test pitch component. Two main principles are involved in the computation of *correlation coefficient*:

   2.2.1: if the test candidate lies on one of the overtone series of the object candidate, a certain *correlation coefficient* is returned, where a higher harmonic index corresponds to lower *correlation coefficient*[30]; otherwise a zero *correlation coefficient* is returned. This computation method is based on the principles of virtual pitch perception models.

   2.2.2: if the test candidate matches the overtone series of a tone component from the object candidate's pitch class, a certain *correlation coefficient* value is also assigned. This part of the computation essentially distinguishes the tonal centre determination algorithm from the pitch perception models, according to the

[30] In this model, a ±8% frequency tolerance is used as the threshold of noticeable pitch difference.

hypothesis that pitch components from same pitch class share similar harmonic functions (Randel, 1990).

2.2 The *tonal weight* of this note candidate equals the sum of the total *correlation coefficients.*

3. The tonal centre concept is predicted to be the note candidate with the highest *total weight*.

A key difference between this algorithm and the virtual pitch algorithm is the use of pitch components as the fundamental element in calculating the *correlation coefficient*; whereas in typical virtual pitch determination algorithms, the *correlation coefficients* are calculated based on the interaction of pure-tone partials. Using cognitive features (noticeable pitch components) rather than solely acoustic features (pure-tone frequencies) this model attempts to predict the perception phenomenon on a psychological rather than physiological level.

In addition to the determination of a *tonal centre*, we also propose a concept of *tonal centre strength.* The *tonal centre strength* indicates the extent to which the tonal context can influence a particular musical entity. Drawing on the analogy of virtual pitch perception, the *tonal centre strength* is proportioned to its *tonal weight*, and inversely related to the number and *tonal weight* of other noticeable pitch components

For the analysis of an isolated sound object/musical entity with no external tonal context, the tonal centre and tonal strength can also be determined from its internal acoustic structure. In another word, the input sound stimuli itself serves as the tonal context for its pitch components. For example, the tonal centre of an isolated triad is fundamentally its root.

Pitch-based tonal functional analysis requires both tonal context as well as the pitch structure of the input sound stimuli. In the next section, a numerical method that converts the acoustic features of the input sound stimuli into a set of audible pitch component is introduced.

6.3 Tonal Consonance and Ambiguity Dissonance

For an input musical entity, conventional psychoacoustic models predict the **overall** level of consonance/ dissonance. For instances, in sensory beats based models (Sethare, 1999; Plomp and Levelt, 1965), the overall level of dissonance is determined by the total amount of roughness sensations contained in the input sound stimuli; and in harmonic-template based models (Parncutt, 1989; Hofmann-Engl, 2006), the overall consonance is estimated based on how well the input sound stimuli matches to a harmonic spectra (*Corps Sonore*[31]). However, in tonal music, it is typically observed that the perceived 'quality' of a musical entity is usually contributed by specific note component(s) rather than the entire structure. For example, the middle note in major, minor triads (root position) determines the major minor tonality; and the highest note in diminished and augmented triads (root position) determines the main quality of diminished and augmented chords. For this reason, the proposed analytical model attempts to estimate the perception responses with respect to each audible pitch components; this approach enables more harmonic details to be visualized at note level (Gabrielsson and Juslin, 2003). The overall consonance and dissonance perception of an input musical structure can be estimated based on the harmonic properties of all its component pitches. In the following sections, the estimation of the *tonal consonance* property and *ambiguity dissonance* property with respect to each audible pitch component will be introduced.

6.3.1 Tonal Consonance

In reference to a tonal centre, some pitch classes are 'supporting' the perception of a tonal centre, and these pitch classes are viewed as having the *tonal consonance* function. In the Hofmann-Engl model, these pitch classes are the intervals of unison

[31] *Coprs sonore* is an idealized harmonic sound spectrum representing sonority with the highest tonal consonance under Jean-Philippe Rameau's Harmonic Relation Theory (Rameau and Wundt, 1721).

(frequency ratio 1:1), perfect fifth (3:2), major third (5:4), minor seventh (7:4), major second (9:8), and major seventh (15:8). The corresponding strength these interval contribute to the root perception are: 6 *Hh*, 5.83 *Hh* , 5 *Hh* , 4.5 *Hh* , 3.3 *Hh* ,and 1.83 *Hh* (Hofmann-Engl, 2006). The six special intervals from Hofmann-Engl model are obtained directly by looking at the relationships between intervals and harmonics (see Table 14).

Harmonic Index					12tET Interval	Error (Cent)
1	2	4	8	16	unison (octave)	0
				9	major second	4
			5	10	major third	−14
		3	6	12	perfect fifth	2
			7	14	minor seventh	−31
				15	major seventh	45

Table 14 Intervals and Harmonic Indexes

The above table listed the 12-tet equal temperament (12tET) intervals and their corresponding index of harmonic series. The unison, major third, perfect fifth and minor seventh intervals appeared at least two times in the first 16 harmonics, whereas the major second and major seventh appear only once, and at a relatively higher harmonic index (9^{th} and 15^{th}). Therefore, the tonal consonance functions of the major second and major seventh intervals are basically negligible. In the case of the minor seventh, although appearing twice (at the 7^{th} and 14^{th} harmonics), the corresponding error is relatively too high to be considered as one of the overtone series according to the virtual pitch determination algorithm (1% tolerance threshold$\approx$ 17.3 Cents < 31 Cents). The pitch class of unisons (octaves), perfect fifth, and major third intervals are thus the only pitch classes that qualify as having tonal consonance functions.

From the perspective of the virtual pitch determination algorithm, with respect to a tonal centre frequency (f_c,), there are two types of frequency partials that are harmonically related to the tonal centre, the first type is simply the frequencies at its harmonics, and the second type includes the frequencies that are octaves related to one of the harmonic overtone series. Perception-wise, the frequency components belonging to these two types are highly merged with the tonal centre, and therefore they are considered as the frequency partials of tonal consonance (f_{tc}), mathematically, it takes a general form of:

$$f_{tc} = n\, 2^m f_c \ (\mathrm{n} \in \mathrm{N},\ \mathrm{m} \in \mathrm{Z})$$

Where n is the harmonic overtone index of tonal centre frequency f_c and 2^m represents the octave related frequencies. In practice, small integer values of n and m are preferred, as the increase of either *n* or m would decrease the *weight* of the harmonic relationship between that frequency and f_c (see section 6.2). In this computational approach, $m = \pm 4, \pm 3, \pm 2, \pm 1, 0$ and $n = 1\text{–}6$ were used. Using such parameters results in the acoustic template shown in Fig.52:

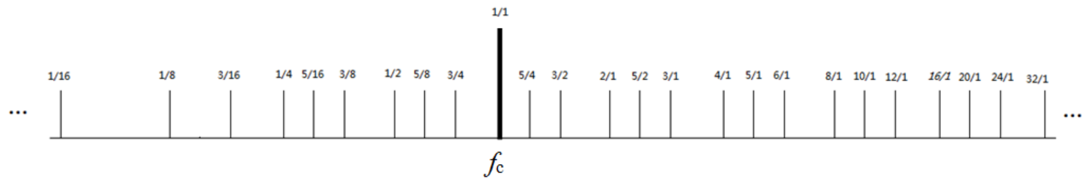


Figure 52 A spectral template of consonance frequency components (f_{tc})

Where on the horizontal axis is a ratio of f_{tc} over the fundamental frequency of tonal centre (f_c): f_{tc}/ f_c

As we may observe, with respect to tonal center frequency f_c, only three pitch classes are observed in the acoustic template of consonance:

1) Root pitch class, featured by the frequency ratios of: $2^m f_c$ ($n \in \mathrm{N}$, m$\in$ Z)

2) Fifth pitch class, featured by the frequency ratios of: $3\, 2^m f_c$ ($n \in \mathrm{N}$, m$\in$ Z), and

3) Major third pitch class, featured by the frequency ratios of: $5\, 2^m f_c$ ($n \in \mathrm{N}$, m$\in$ Z).

It can be noticed that the pitch classes of unison (octaves), major third and perfect fifth form the major structure. The major structure has a relatively clearer perception of tonal centre, and this can be used to interpret why major structure sound more stable and consonant than other chord structures. According to Shamma and Klein (2000), perceiving major triads as consonance is not due to enculturation, but founded on psychological and physiological principles.

To determine the degree of the tonal consonance function of an audible pitch component [f_i , A'_i], the algorithm checks if its fundamental frequency (f_i) is equal or nearing to one of the tonal consonance frequency partials f_{tc}, if they are equal, this pitch component has a strong tonal consonance function, if they are near, this pitch component has a weakened tonal consonance function, if the f_i is nowhere near to any one of the tonal consonance frequency partials f_{tc}, this pitch component has no tonal consonance function.

To achieve such comparisons, the algorithm firstly finds the nearest tonal consonance frequency partials $f_{tc\text{-}nearest}$ with respect to f_i :

$$f_{tc\text{-}nearest} : \min|\, f_{tc} - f_i \,| = f_{tc\text{-}nearest} \in f_{\mathrm{tc}};$$

The reason for comparing a note to its nearest consonant frequency partial is an assumption that people tend to perceive a raised or lowered pitch property in reference to the closest note in a major triadic structure. For instance, a pitch that is slightly above a major third (within a few cents) will be considered as a sharp tone, but not a flattened tone from the perfect fifth interval.

Next, $f_{tc\text{-}nearest}$ compared to f_i to obtain the level of tonal consonance $\mathrm{C_t}(f_i)$ function of pitch component f_i by:

$$\begin{cases} \mathrm{C_t}\,(f_i) = A'_i; \text{ if } f_i = f_{tc\text{-}nearest} \\ \mathrm{C_t}\,(f_i) = A'_i\,(1 - |f_{tc\text{-}nearest} - f_i\,| / \,\delta\, f_{tc\text{-}nearest}\,); \text{ if } (1-\delta)\, f_{tc\text{-}nearest} < f_i < (1+\delta)\, f_{tc\text{-}nearest} \\ \mathrm{C_t}\,(f_i) = 0; \text{ else} \end{cases}$$

Where δ is the *Coincidence coefficient*, $\delta = 0.01$ (Stolzenburg, 2012).

The $\mathrm{C_t}\,(f_i)$ function can be visualized in Fig.53.

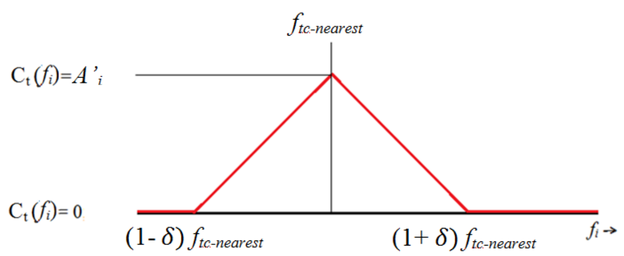


Figure 53 Tonal Consonance Effect

As the boundary between consonances is clearly defined by the use of the Coincidence coefficient (δ), under a qualified approach, the model is very sensitive to the tuning systems involved. Table 15 demonstrates that under several tuning systems, the intervals of unison, major third, fifth, and octave remain as the consonant intervals.

12- tet Intervals	$f_{tc\text{-}nearest}$ / f_c	Equal- Temp.		Pythagorean		Kirnberger III		Rational tuning		Just tuning	
		Exact Ratio	Error(%)	Exact Ratio	Error(%)	Exact Ratio	Error(%)	Exact Ratio	Error(%)	Exact Ratio	Error(%)
unison	1/1	1.000	**0.00**	1/1	**0.00**	1/1	**0.00**	1/1	**0.00**	1/1	**0.00**
minor second	1/1	1.059	0.06	256/243	0.05	25/24	0.04	16/15	0.07	16/15	0.07
major second	1/1	1.122	0.12	9/8	0.13	9/8	0.13	9/8	0.13	9/8	0.13
minor third	5/4	1.189	0.05	32/27	0.05	6/5	0.04	6/5	0.04	6/5	0.04
major third	5/4	1.260	**0.01**	81/64	**0.01**	5/4	**0.00**	5/4	**0.00**	5/4	**0.00**
perfect fourth	5/4	1.335	0.07	4/3	0.07	4/3	0.07	4/3	0.07	4/3	0.07
tritone	3/2	1.414	0.06	729/512	0.05	45/32	0.06	17/12	0.06	7/5	0.07
perfect fifth	3/2	1.498	**0.00**	3/2	**0.00**	3/2	**0.00**	3/2	**0.00**	3/2	**0.00**
minor sixth	3/2	1.587	0.06	128/81	0.05	25/16	0.04	8/5	0.07	8/5	0.07
major sixth	3/2	1.682	0.12	27/16	0.13	5/3	0.11	5/3	0.11	5/3	0.11
minor seventh	2/1	1.782	0.11	16/9	0.11	16/9	0.11	16/9	0.11	9/5	0.10
major seventh	2/1	1.888	0.06	243/128	0.05	15/8	0.06	15/8	0.06	15/8	0.06
octave	2/1	2.000	**0.00**	2/1	**0.00**	2/1	**0.00**	2/1	**0.00**	2/1	**0.00**

Table 15 Diatonic frequency ratios and their nearest harmonic-template ratios.

In this table, the nearest harmonic-template ratios for unison, minor and major second are considered to be 1/1; the nearest frequency ratios for minor, major third and perfect fourth are considered to be 5/4; the nearest frequency ratios for the tritone, perfect fifth, minor and major sixth are considered to be 3/2; and the nearest frequency ratios for a minor, major seventh and octave are considered to be 2/1. The highlighted (bold) cells represent those frequency ratios with an error less than 0.016.

Based on the observation from Table 15, when the cut-off error is fixed at 0.01, the major third interval under equal-temperament and Pythagorean tuning systems will not be considered as having tonal consonance function.

6.3.2 Ambiguity Dissonance

A*mbiguity dissonance* is the opposite concept to tonal consonance. It refers to a pitch component that does not support the perception of a tonal centre, but tends to support a different pitch class as the competing tonal centre. Within a chord structure, pitch components with *ambiguity dissonance* function throw the root perception into confusion; and such confusion is viewed as a type of tonal dissonance effect.

The pitch class with the strongest *ambiguity dissonance* effect would be the *subdominant* pitch class (P_c =4/3). This is because the pitch component from this pitch class tends to make the most salient pitch class of the sonority – the tonal centre pitch class to support it as a competing tonal centre. Similarly, the *submediant* pitch class will also work together with the tonal centre pitch class to suggest a virtual tonal centre at the *subdominant* pitch class (see illustration in Fig.54).

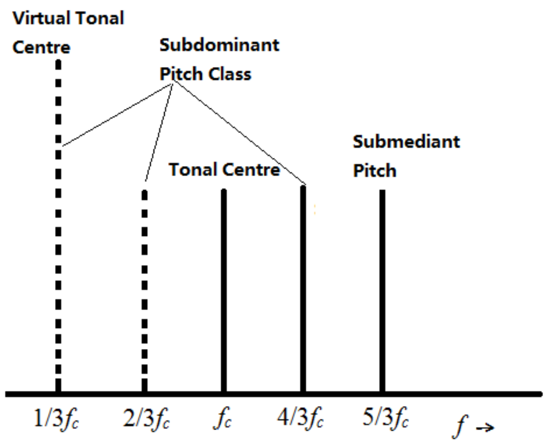


Figure 54 Sub-dominant function and virtual tonal centre

Where the fundamental frequencies of the tonal centre (f_c) subdominant pitch ($4/3f_c$) and submediant pitch ($5/3f_c$) are the 3rd, 4th and 5th harmonics of a virtual tonal centre at $1/3\ f_c$.

In the computational analysis, the amount of *ambiguity dissonance* effect D_a of a pitch component $[f_i\ ,\ A'_i]$ is estimated by:

(1) If the fundamental frequency of this pitch component belongs to the tonal centre pitch class, then this pitch component has no competing effect:

$D_a(f_i) = 0$; if $(1-\delta)\, 2^m f_c < f_i < (1+\delta)\, 2^m f_c$

Where f_c is the fundamental frequency of the tonal centre and 2^m represents its octave related frequencies. ($m = \pm 4, \pm 3, \pm 2, \pm 1, 0$)

(2) Else determine the virtual pitch *weight* (w) of frequency f_i according to the virtual pitch determination algorithm (see section 6.1 in this chapter):

$D_a(f_i) = w(f_i) / A'_i$; else

*Note: the virtual pitch *weight* is normalized by the audibility (A'_i) of the pitch component.

Based on such an algorithm, the pitch class of a fourth, major six and minor six scale degrees (P_c= 4/3, 5/3, 8/5) have the highest level of ambiguity dissonance.

6.4 Summary

According to the thesis hypothesis, the concept of a tonal centre plays an important role in the perception of tonal music harmony, and a series of tonal harmony properties were implemented and modelled in this PhD study. These tonal harmony functions include: *tonal consonance; ambiguity dissonance, gloom dissonance* and *tension dissonance*.

The tonal centre concept is the most salient pitch sensation of a sound object; it can be determined using a virtual pitch determination algorithm. For musical chords, the tonal centre concept corresponds to the root of chord.

A particular pitch component is considered as having a *tonal consonance* function (C_t) when its fundamental frequency is either one of the harmonic overtone series of the fundamental frequency of tonal centre OR octave-related to one of the harmonic overtone series of the fundamental frequency of tonal centre. The pitch component from the first, major third and fifth scale degrees have strong tonal consonance function.

A particular pitch component is considered as having *ambiguity dissonance* function (D_a) when its presence evokes competing tonal centre(s). The competing tonal centres increase the perception effort to identify the tonal centre concept, which makes it a tonal dissonance concept. The pitch component from the fourth, major and minor sixth scale degrees have strong ambiguity dissonance function.

The remaining tonal dissonance functions, *gloom* and *tension* will be introduced, discussed and modelled in the next chapter.

Chapter 7 Gloom and Tension Dissonance

The limitations of previous psychoacoustic models imply that the perception of music harmony is perhaps beyond the definitions of *sensory dissonance*[32] and *tonal consonance*[33]. In this chapter, a theoretical attempt is made which tries to incorporate two types of music emotions – gloom and tension – into the analysis of music harmony; with the goal of achieving a better prediction result in line with empirical observations. The structure of this chapter is as follows:

Section 7.1 reviews the principal dimensions of music emotions, from which the two uncorrelated dissonant emotions – gloom and tension are introduced.

Section 7.2 looks for the musical features associated to gloom and tension emotions – the falling and rising pitch contours.

Section 7.3 provides some theoretical insights into why people tend to associate falling and rising pitch contours with gloom and tension emotions.

Section 7.4 introduces a novel analytical approach which makes use of the rising and falling pitch properties for the analysis of tension and gloom dissonances. The reasoning behind such an approach is also discussed in this section.

Section 7.5 presents a qualitative study to see if listeners are able to consistently associate gloom and tension emotion to the rising and falling pitch properties accordingly. The theory proposed in section 7.4 is tested in this experimental study.

Section 7.6 presents a comparative test between gloom and tension emotions in order to get a statistical conclusion on whether the two types of emotions are significantly

[32] Sensory dissonance, measured by the amount of beats effect, is a physiological based dissonance concept under beats related theories.

[33] Tonal consonance, measured by the degree of how a sound matches harmonic template spectrum (*harmonicity*), is a consonance concept which happened at higher level than sensory dissonance under harmonic template related theories.

different from each other in terms of perceived degree of overall dissonance effect. Moreover, a comparative study is made between gloom and tension dissonances.

Section 7.7 summarises this chapter.

7.1 Gloom and Tension Emotion

Music emotion seems complex and ineffable; as American philosopher Susanne Langer described, "Human feelings are much more congruent with musical forms than with forms of language; music can reveal the nature of feelings with a detail and truth that language cannot approach" (Langer, 1942). However, many researchers (such as Tomkins, 1963) have postulated the existence of some 'basic' non-verbal musical emotions, a concept akin to the primary colours in vision. Based on their arguments, 'basic' musical emotions are underpinned by the 'biologically pre-programmed' psychoacoustical responses (known as *essentic forms*) coded at the sub-cortical areas in brain. Mixtures of such basic emotions result in hundreds of linguistic descriptors.

In the literature, much research has been carried out to probe and find proper descriptors that best denote the 'basic' emotions. The methodology typically involves an empirical investigation of participant's dissimilar judgements between emotional descriptors, followed by a multivariate data analysis. Relevant techniques including: cluster analysis, factor analysis, multidimensional scaling analysis, correlation analysis, and correspondence analysis. Employing such statistical methods, several conclusions have been drawn since the 1960's.

According to (Juslin and Sloboda, 2010), Kleinen incorporated factor analysis and identified two principal dimensions of music emotions: the first is related to the positive or negative valence of emotional state (the first principal dimension of music emotion), with the representative descriptors of 'cheerful – serious'; the second is associated with the level of excitement, with the descriptors of 'powerful – tender' (the second principal dimension of music emotion).

Wedin (Wedin, 1972) examined the music-related descriptors and drew a similar conclusion. There are also two principal emotional dimensions identified: the first is similar to Kleinen's first dimension, only slightly adjusted to another set of descriptors, 'gaiety – gloom'; and the second is described by 'tension – relaxation'.

With the principal dimensions defined, the meaning of other emotional descriptors can be located on a two-dimensional space (see Fig.55). The first and second principal dimensions of music emotions can also be found in Hevner's adjective circle (Fig.56).

In her perception study, the music emotional descriptors were categorised into eight clusters.

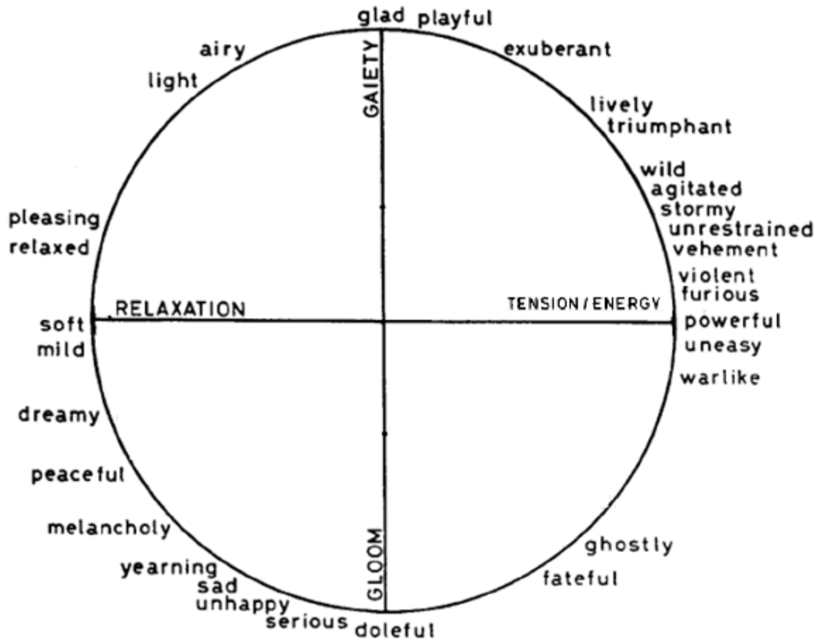


Figure 55 Wedin's two-dimensional plane of emotional descriptors

Figure taken from (Wedin, 1972)

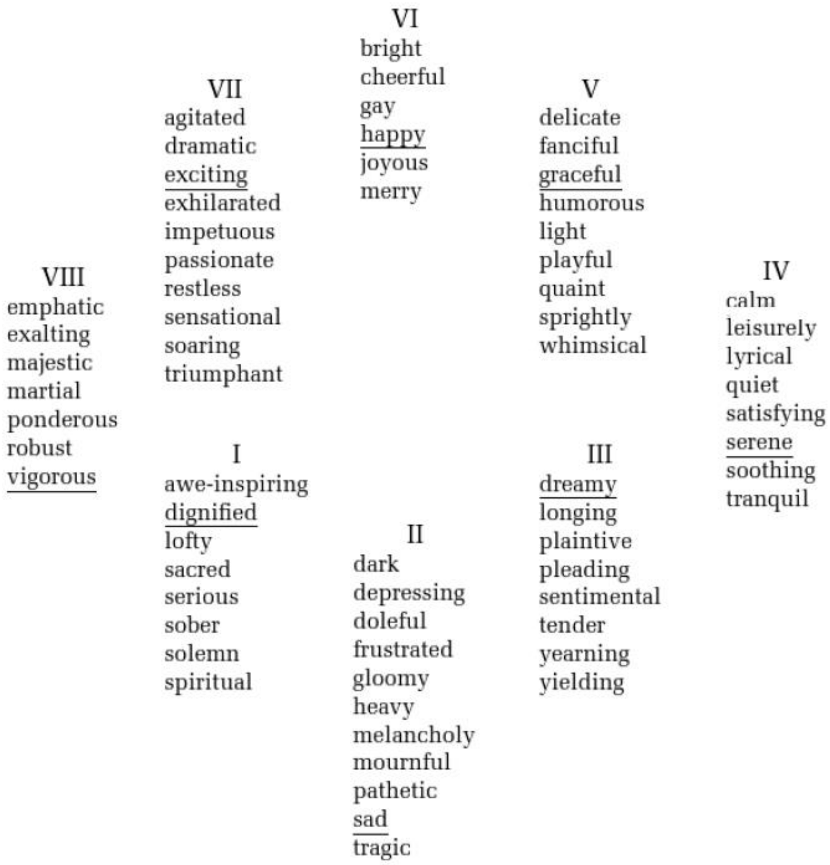


Figure 56 Hevner's adjective circle.

The representative descriptors for each cluster are underlined. This figure is taken from (Hevner, 1936)

Moreover, a close relationship is observed between: (1) the 'gaiety – gloom' axis and the first dimension of psychological affects (the valance[34]), and (2) the 'tension – relaxation' axis and the second dimension of psychological affects (the arousal[35]). Therefore, music emotions and psychological affects have been linked together and viewed as correlated concepts.

In the present research, Wedin's descriptions of principal dimensions were used ('gaiety – gloom'; and 'tension – relaxation') to represent two fundamental dimensions of music emotions. In western music literature, both gaiety and relaxation emotions have been related to musical consonance concepts; and tension and gloom

[34] Valence is the subjective positive-to-negative evaluation of an experienced state.

[35] Arousal is objectively measurable as the activation of the sympathetic nervous system.

emotions on the other hand convey musical dissonance. The double meaning of consonance (gaiety & relaxation) and dissonance (gloom & tension) implies a two-dimensional nature which cannot be measured by the one-dimensional concept such as a degree of sensory dissonance or tonal consonance. To incorporate two-dimensional musical emotions into harmony perception analysis, we must firstly identify the corresponding musical/ acoustical features. In the next section, the kind of musical features which may evoke gloom and tension emotions are considered in more detail.

7.2 Musical Features for Gloom and Tension Emotions

It has been acknowledged that music can be used as a communication tool for affective meanings. Musicians know how to 'encode' affective meanings with the use of musical features, and the audiences know how to 'decode' them. The musical features used can be roughly categorised into two groups: 1) the use of musical notes (pitch structures) originated by music composers; 2) the techniques used by music performers, which involve mainly dynamic features such as timing, loudness variation, tempo, and vibrato. As the current research interest is on the perception of music harmony, the former aspect is of most interest: the use of musical tones.

In order to extract and analyse gaiety & relaxation and gloom & tension emotions in music harmony, the first thing is to identify the corresponding musical/ acoustical features for music emotions. In Scherer's experimental study (Scherer and Oshinsky, 1977), participants were asked to assign a set of emotional descriptors to a given set of musical variables. A high correlation was found between the 'gloom' descriptor to descending melody/ pitch contours. (Madsen and Fredrickson, 1993) and (Krumhansl, 1997) on the other hand concluded that ascending melodies and/or increasing sound density are the key musical features which evoke a 'tension' emotion.

Associating ascending and descending melodic contours to the tension and gloom emotions are also observed in *melodic expectation theory* (Margulis, 2007; see section 3.2.4). Under *melodic expectation theory*, it has been generally concluded that:

(1) people tend to have a psychological expectation (in terms of pitch, tempo and loudness) in music perception;

(2) music variables exceeding such expectation (such as higher pitch, ascending conjunct motions[36], or higher loudness level) generally introduce psychological tension;

(3) when music variables are below expectation (such as lower pitch, descending conjunct motions, or lower loudness level), emotions such as gloom or sadness can be perceived;

(4) the melodic progression from higher level sound features to the ‘expected’ level leads to an emotion of ‘relaxation’; and

(5) the melodic progression from lower level sound features to the ‘expected’ level invokes a psychological experience of ‘happiness’ and ‘pleasantness’ (gaiety emotion).

We may also notice that gloom and tension emotions do not correspond to a particular **isolated** musical entity/symbol, but is associated with the ‘motions’ *between* musical entities. This is in line with the hypothesis that music harmony perception is not determined solely by the internal structure of the sound stimuli.

7.3 Possible Interpretations for Gloom and Tension Emotions

One of the possible interpretations of why people tend to associate rising/falling pitch contours to the tension/gloom emotions can be found in Ohala’s biological theory – *the frequency code of sound symbolism* (Ohala, 1983). Ohala reported that animals of larger physical size tend to create lower pitched voiced sounds and smaller sized animals tend to voice at a higher pitch. During the biological evolution process,

[36] Conjunct motions refer to step-wise melodic motions, typically with a step-size of one or two semitones. This term contrasts with disjunct melodic motions, also known as skip-wise motions, with an intervallic jump of more than two semitones.

lower-pitched voices have become a sound symbol of ‘heavy’ and ‘inhibition’. Individuals tend to employ a lower-pitched voice to express their sadness and annoyance (an emotion akin to gloom), and use higher-pitch voices to signify their weakness and nervousness, corresponding to a psychological tension emotion. In a separate study (Chen, 2002), it was found that listeners from different cultures and age groups also tended to map ascending melodies or rising pitch to a visual impression of shrinking size, and descending melodies or falling pitch to growing physical size, which appears to verify Ohala’s theory from another angle.

From a psychological point of view, it was proposed that people are inherently driven to maintain certain levels of psychological arousal (*arousal theory of motivation*, (Cheery, 2009). As illustrated by Cheery: ‘*if our* (arousal) *levels drop too low we might seek stimulation by going out to a nightclub with friends. If these levels become too elevated and we become overstimulated, we might be motivated to select a relaxing activity such as going for a walk or taking a nap*’. According to such a theory, an under-stimulated arousal level is associated with the emotions of depression and gloom; whereas an over-stimulated arousal level implies strain or tension.

Physiologically, there is a standard auditory range of human comfort (Lucker et al., 1978). Auditory perception features such as pitch, loudness, and tempo fall under such ranges, essentially corresponding to a sensory consonance concept. In response to external sound stimuli, an increase in pitch, loudness and tempo causes additional neural firings which activates the sympathetic nerve functions that lead to psychological tension or excitement; conversely, a decrease in pitch, loudness or tempo leads to a parasympathetic mechanism which is manifested in the experience of a gloom emotion (Krumhansl, 1997).

7.4 Analysing Gloom and Tension Dissonances

Following the theoretical thoughts that the harmony perception of musical chord is referential (in reference to tonal context) but not absolute, the perception gloom and tension dissonances are influenced by the presence of other chord structures. For example, when listeners hear a single minor triad, they may not be able to give a consonance/dissonance perception ranking to it as there are no other chord structures to compare it with. When a major and diminished chord is provided, they are able to rate the major triad as more consonant than the minor triad, and the latter is in turn more consonant than the diminished triad. During this comparison process (such as a chord progression), the rising and falling pitch properties can be noticed (see illustration in Fig.57). The current research model proposes that the tension and gloom dissonance functions can be assigned to rising and falling proprieties.

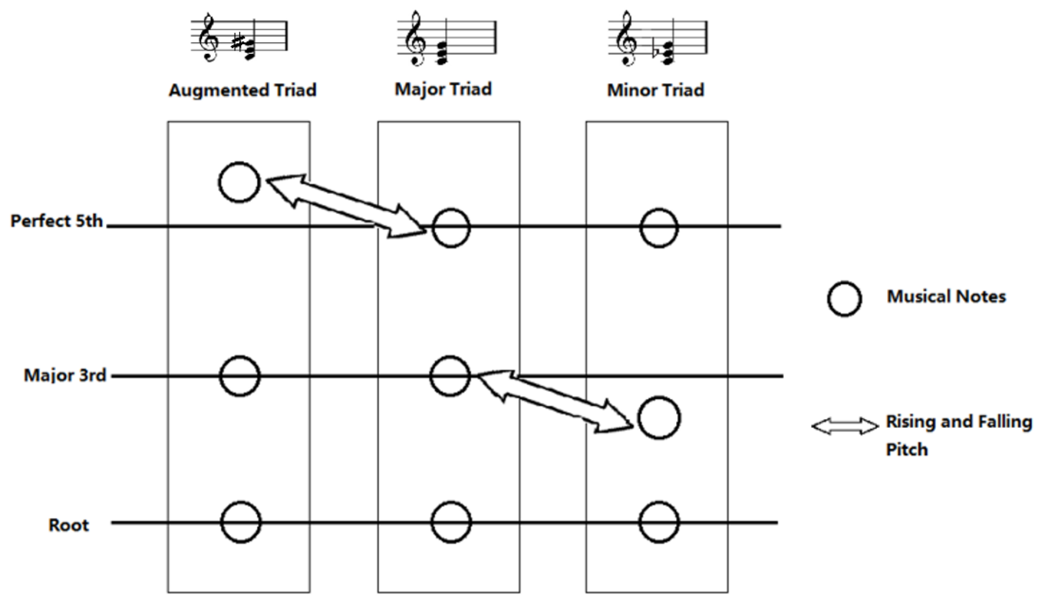


Figure 57 Illustration of falling and rising pitch properties

To determine the gloom and tension functions of a particular pitch component, we must firstly identify the rising and falling pitch properties associated with it. Under a specific tonal context, the *tonal consonance pitch components* were usually identified as the most consonance concepts. It is therefore hypothesized that the perception of a pitch/tone has to be compared with one of the *tonal consonance pitch components*(f_{tc}): if the fundamental frequency of a pitch component is slightly higher than the fundamental frequency of a *tonal consonance pitch component*, it is modelled with the tension dissonance function (rising pitch property); If the fundamental frequency of a pitch component is slightly higher than the fundamental frequency of a tonal consonance pitch component, it is modelled with gloom dissonance function (falling pitch property).

As detailed in the previous chapter, the *frequency partials of tonal consonance* (f_{tc}) is mathematically given by:

$$f_{tc} = n\, 2^m f_c \ (n \in \mathrm{N},\ m \in \mathrm{Z})$$

Where n is the harmonic overtone index of tonal centre frequency f_c , and 2^m represents the octave related frequencies; n=1,2,3,4,5, 6 and $m = \pm 4, \pm 3, \pm 2, \pm 1, 0$.

Similar to the determination of tonal consonance function, for a given audible pitch component $A_c(f_i)$,the algorithm firstly find the nearest tonal consonance frequency partials $f_{tc\text{-}nearest}$ with respect to f_i :

$$f_{tc\text{-}nearest} : \min|\, f_{tc} - f_i \,| = f_{tc\text{-}nearest} \in f_{\mathrm{tc}};$$

Next, $f_{tc\text{-}nearest}$ compared to f_i to obtain the level of tension dissonance $D_{\mathrm{t}}(f_i)$ and gloom dissonance $D_{\mathrm{g}}(f_i)$:

$$\begin{cases}
\text{if } f_i > (1+\delta))f_{tc\text{-}nearest} \text{ : meaning the rising/higher pitch property,} \\
\qquad D_{\mathrm{t}}(f_i) = A'_i \cdot (f_i - f_{tc\text{-}nearest})/\, f_{tc\text{-}nearest}; \\
\qquad D_{\mathrm{g}}(f_i) = 0; \\
\qquad C_t(f_i) = 0; \\
\text{if } f_i < (1-\delta)f_{tc\text{-}nearest} \text{ : meaning the falling/lower pitch property,} \\
\qquad D_{\mathrm{t}}(f_i) = 0; \\
\qquad D_{\mathrm{g}}(f_i) = A'_i \cdot (f_{tc\text{-}nearest} - f_i)/\, f_{tc\text{-}nearest}; \\
\qquad C_t(f_i) = 0; \\
\text{else: meaning } f_i \text{ belongs to a tonal consonance frequency range.}
\end{cases}$$

$D_t(f_i) = 0;$

$D_g(f_i) = 0;$

$C_t(f_i) = A'_i \cdot (1 - |f_{tc\text{-}nearest} - f_i| / \delta\, f_{tc\text{-}nearest})$; (see section 6.4)

Where δ is the *Coincidence coefficient*, $\delta = 0.01$

From the above formulation, we can observe that the perceived level of tension and gloom dissonances is proportional to two factors: (1) the audibility of that tone component: A'_i; and (2) the intervallic distance between the fundamental frequency of that pitch component to its nearest tonal consonance frequency component: $| f_{tc\text{-}nearest} - f_i| / f_{tc\text{-}nearest}$. The further it is away from a tonal consonance frequency component, a stronger gloom/tension dissonance effect is modelled.

Taking the analysis of minor triad (root position) for example, according to the tonal centre determination algorithm, the lowest note is identified to be the root of chord. With respect to the root pitch, a series of consonance frequency components can be generated. These frequencies essentially refer to the fundamental frequencies of three pitch classes: the root pitch class (P_c=1/1), major third pitch class (P_c=5/4), and fifth pitch class (P_c=3/2). The lowest and highest note of the minor triad belongs to the root and fifth pitch class, therefore these two note components will be categorized with the tonal consonance function; the middle note, however, is one-semitone lower than its nearest tonal consonance pitch class – the major third pitch class, therefore the middle note will be considered as having gloom dissonance function.

According to the algorithm, the main harmonic function of 12 chromatic pitch classes can be illustrated in Fig.58 below.

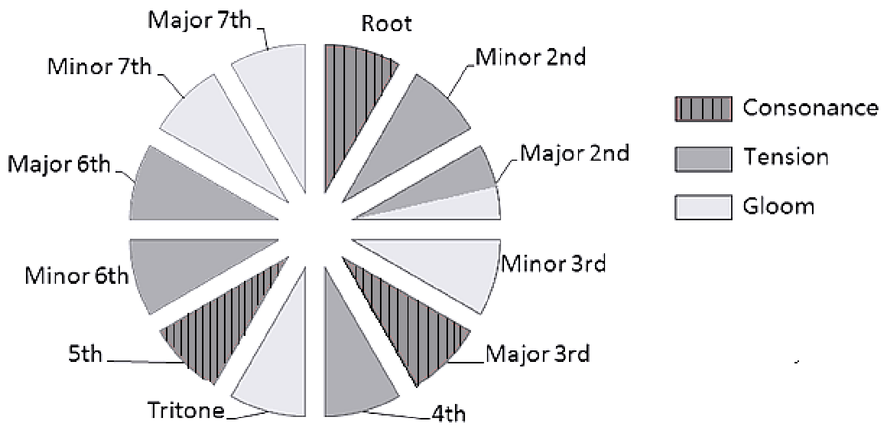


Figure 58 Functional classification of twelve pitch classes

Where the harmonic function of the major 2nd is both tension and gloom, as it can be viewed as two semitones higher than the root pitch class or two semitones lower than the major 3rd pitch class

To verify the theoretical thought proposed in this section, a listening test was conducted, and is introduced in the next section.

7.5 A Qualitative Test for Gloom and Tension Dissonance

7.5.1 Overview

In reference to a tonal centre, the tonal consonance concept has been given to pitches that belong to one of the three tonal consonance pitch classes: the root (I), major third (major III) and perfect fifth (V) pitch classes[37]. Other pitch classes can be generally viewed as tonal dissonance pitch classes. To estimate the overall dissonance level of a chord is to identify how many audible dissonant pitches and what types (gloom,

[37] Root pitch class is a set of octave-related pitches with frequency ratios of $\mathbf{2^n}$ over the fundamental frequency of tonal centre (root); Major third pitch class is a set of octave-related pitches with frequency ratios of $\mathbf{5x2^n}$ over the fundamental frequency of tonal centre; Perfect fifth pitch class is a set of octave-related pitches with frequency ratios of $\mathbf{3x2^n}$ over the fundamental frequency of tonal centre. $n \in \boldsymbol{Z}$

tension, ambiguity etc.) of dissonances that chord contains. In the listening test, major triadic chord structures serve as a reference point of the tonal consonance concept, where perceived harmony in term of gloom and tension dissonances of all the other chord structures can be compared and measured.

According to the hypothesis, when one of the pitches of the major triad is slightly raised, tension dissonance will be perceived; and when one of the pitches of major triad is slightly lowered, gloom dissonance will be perceived. If this hypothesis is correct, the following result is expected (Fig.59):

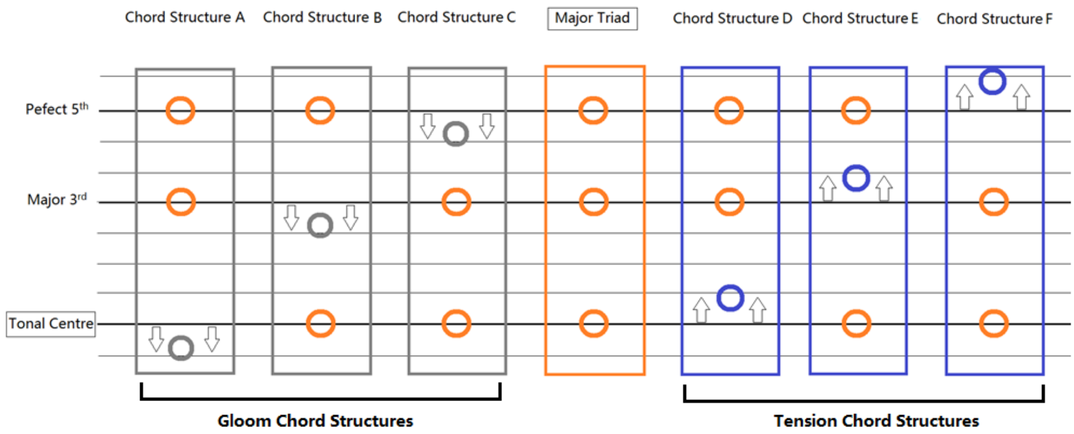


Figure 59 Expected experiment result for qualitative gloom and tension test

Chords A–C have one of their notes lowered from the major triad structure, therefore such chord structures are expected to have gloom dissonance when compared to the major triad; Chord D–F are expected to have tension dissonance as they have one of their notes raised from the major triadic structure.

The sound stimuli were divided into three sections.

The first section tests the upward and downward pitch contour at the bass note of the major triad. Thus listeners listened and compare chord structures A and D and to see if they would assign chord structure A to gloom and on the other hand, structure D as tension.

Similarly, the second section compares chord structures B and E, which considers the upward and downward pitch contour at the middle note of major triad.

The third section compares chord structures C and F, which considers the upward and downward pitch contour at the highest note of major triad. Three trials of experiments essentially tested the harmonic properties of pitches slightly above and below the one of the three pitch classes under tonal consonance concepts (root pitch class P_c=1/1, major III pitch class P_c=3/2 and V pitch class P_c=5/4).

The challenge of this experiment was to instil the root information to listeners before they hear the test chord structure. To achieve this, a major triad will be played to the listener one second before the onset of the main stimulus (dissonant chord structure). Therefore, in trial #1, listeners heard two sound samples in a random order: the first sound sample consisted of a major triad followed by chord structure A; the second sound sample consisted of a major triad followed by chord structure D. After they heard both samples, they were asked to comment on chord A and chord D in term of tension and gloom emotions. The same procedure applies for trials #2 and #3.

Three actions were taken to prevent the result being influenced by the enculturation:

- Choosing non-musically trained subjects from different cultural and geographical backgrounds,
- Using a synthesised timbre that is unfamiliar to subjects, (only eight harmonic overtones are synthesised in this test, see Fig.60)
- Using unconventional intervals in the test chord structures (see next section for more details). As illustrated in Fig.59, the rising and falling tones are not aligned with the grid of 12-tet equal-temperament.

Based on the experimental data collected, correlation analysis was used to examine the relationships between rising / falling pitches and tension / gloom dissonances.

7.5.2 Implementation

In this experiment, the fundamental frequency of the root was set at 240Hz, which is a frequency between B^b_3 and B_3 under ISO 16. The amount of pitch shift was set at 85 cents per note, which corresponds to a pitch interval that is slightly smaller than one semitone. Therefore, the reference major triad and chord structures A–F have the fundamental frequencies of (see Table 16):

	Maj. triad	***Chord A***	***Chord B***	***Chord C***	***Chord D***	***Chord E***	***Chord F***
bass	240	**229**	**252**	240	240	240	240
3rd	300	300	300	**286**	**315**	300	300
5th	360	360	360	360	360	***343***	***378***

Table 16 Fundamental frequencies of triadic structures used in the listening test

The modified frequency components (raised or lowered) are underlined in bold type

Based on the above table of frequencies, triads A–F were synthesised using PureData (see Fig.60).

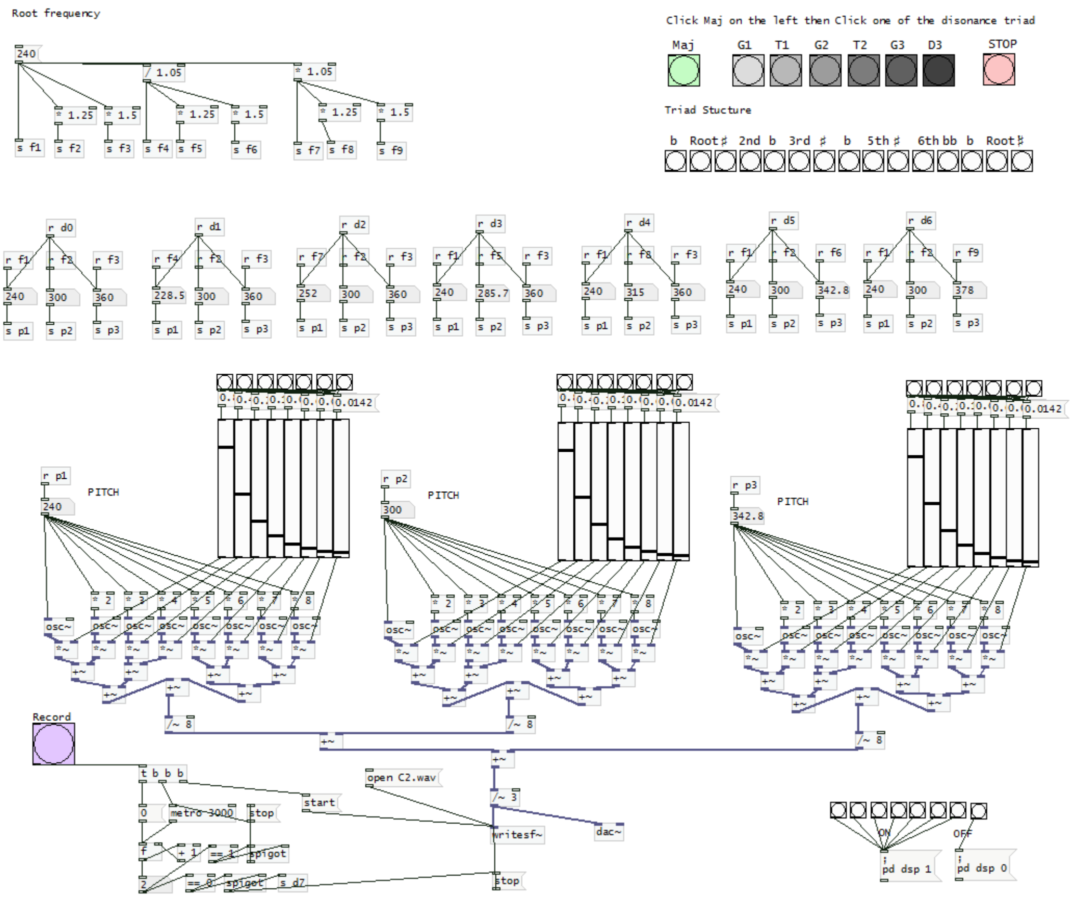


Figure 60 Triad synthesis using Puredata

Each harmonic tone consisted of eight harmonic partials, and 5dB/ partial was used to model the amplitude decay. Six audio files were prepared for this experiment:

- Audio #1A: 3s of the major triad, 2s of silence, followed by 3s of chord A.

- Audio #1B: 3s of the major triad, 2s of silence, followed by 3s of chord D.
- Audio #2A: 3s of the major triad, 2s of silence, followed by 3s of chord B.
- Audio #2B: 3s of the major triad, 2s of silence, followed by 3s of chord E.
- Audio #3A: 3s of the major triad, 2s of silence, followed by 3s of chord C.
- Audio #3B: 3s of the major triad, 2s of silence, followed by 3s of chord F.

(s: seconds)

An executable application was developed for listeners to complete this particular test. The application GUI was developed in Matlab, and compiled to a Windows application using Matlab's *deploytool*. At the opening page, listeners were shown a consent page (Fig.61), where the purpose, content, requirements, as well as possible risks of this experiment were introduced. By entering information of age and country of origin (defined by the most time the subject had spent before age 18) and clicking the 'understand and proceed' button, listeners agreed to participate in this listening test. The application would then proceed to the next page.

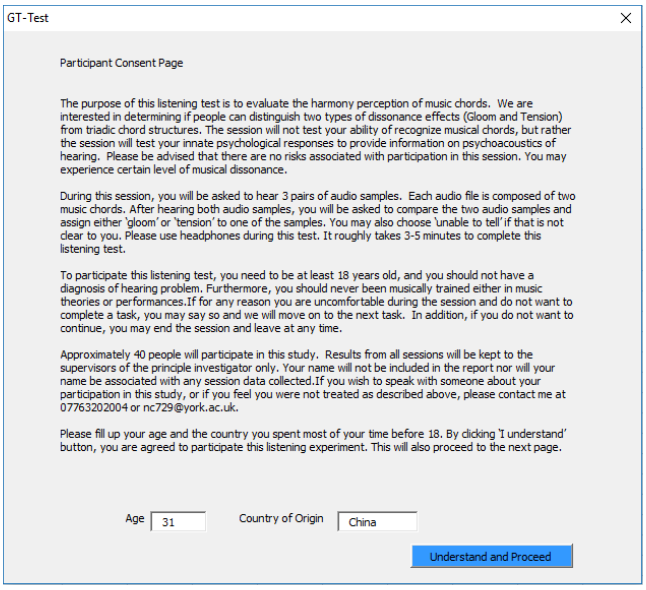


Figure 61 Consent page of the gloom-tension test

On the test page (Fig.62), three sub-tests were presented to the listeners. Upon opening the page, the audio material of Audio #1A and #1B was randomly assigned to Audio file A and Audio file B in test #1; similarly, Audio #2A, Audio #2B, Audio #3A, Audio #3B, were randomly assigned to the audio file A or B in test #2 and test #3.

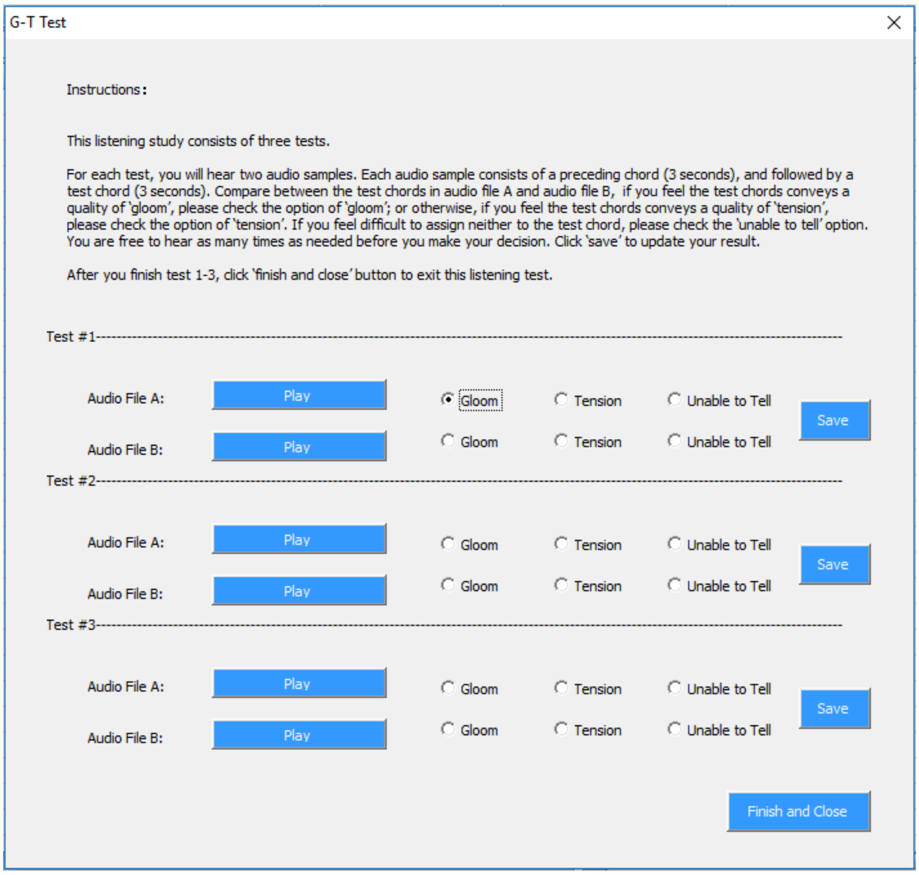


Figure 62 Test page of the gloom-tension test

During each test, listeners compared the dissonance chord structures and assigned either 'gloom' or 'tension' to the corresponding chord structures. They could also choose the option of 'unable to tell' when no obvious tension or gloom effect was perceived. The result text file was updated whenever users clicked the save button. After finishing the tests, the result text file was collected for further analysis. A sample result.txt file contains following information:

Age: 27

Country: China

Test 1A: Gloom

Test 1B: Tension

Test 2A: Gloom

Test 2B: Tension

Test 3A: Unable to tell

Test 3B: Unable to tell

7.5.3 Subject's profile

A total number of 42 subjects participated in this listening test. None of them had been musically trained or had any musical background. Participants were recruited by social network, 7 being students from the University of York, Department of Electronics.

Age-wise, 16 out of 42 subjects were between 18–30 years old; 12 subjects were aged 30–40; 8 subjects aged 40–50; and 6 subjects were over 50 years old (see Fig.63).

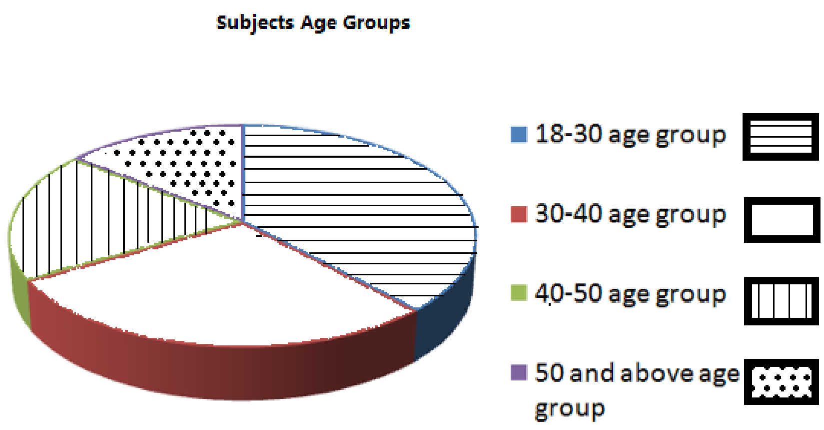


Figure 63 Age distribution of listener subjects

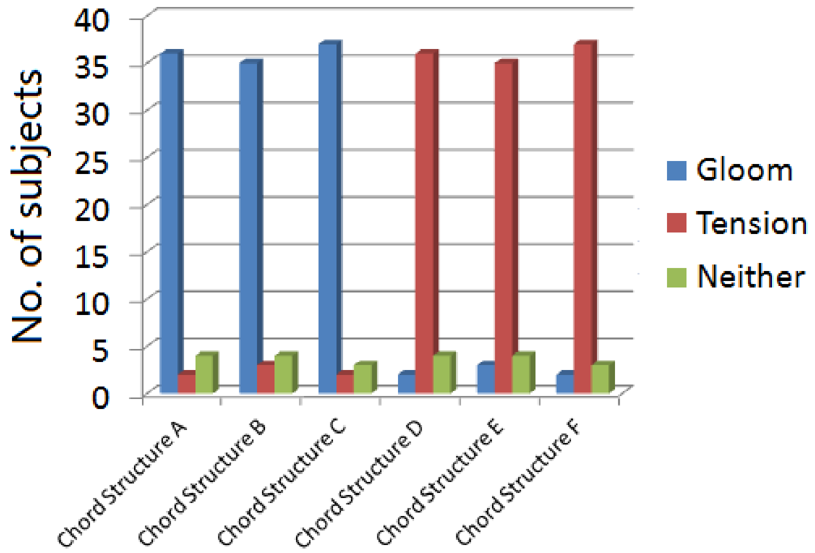
Regarding Country of Origin, 15 out of 42 subjects came from China; 8 came from the U.K; and the remaining subjects came from: Russia (1), Romania (3), Germany (1), Italy (1), Egypt (1), Cameroon (2), Japan (1), Singapore (2), Australia (2), United States (3) and Canada (2).

Figure 64 Subjects' geographical distribution

For some participants, this experiment was conducted remotely; only 11 out 42 subjects' tests took place at the Audio Lab, University of York. Prior to the study, ethical approval was granted from the University of York's Physical Sciences Ethics Committee.

7.5.4 Results and Discussions

All 42 subjects were able to complete the test, and the results are shown below in Fig.65:



No. of subjects	Gloom	Tension	Neither
Chord Structure A	36	2	4
Chord Structure B	35	3	4
Chord Structure C	37	2	3
Chord Structure D	2	36	4
Chord Structure E	3	35	4
Chord Structure F	2	37	3

Figure 65 Result of the Qualitative Test

Based on observation of the results, the majority (average 85.7%) of subjects tended to judge chord structures A–C as gloom; and chord structures D–F as tension. This is considered as a strong result which supports the original hypothesis.

3 subjects were unable to distinguish the gloom and tension dissonance throughout the tests. Two of them came from China and one from Romania. A possible reason is that the descriptor of gloom and tension were given in English, and not in their first language. Therefore, one adjustment was made for the subsequent quantitative test: the concept of gloom and tension was translated to the subject's first language.

Based on the observation of 42 subjects, the gloom and tension effects are not influenced by country of origin or age of subjects. 2 subjects had reverse predictions; they belonged to neither the same country of origin nor the same age group.

41 out of 42 subjects had consistent results from test #1 to test #3: meaning that if they assigned gloom to falling pitch, they would assign tension to rising pitch at the same time; if they assigned tension to falling pitch, they would assign gloom to rising

pitch at the same time; if they assigned ‘unable to tell’ to falling pitch, they would assign ‘unable to tell’ to rising pitch as well. One of the subjects changed his assignment strategy during the test: in test #1 and test #3, he assigned gloom to falling pitch; but in test #2, he assigned gloom to rising pitch.

To conclude, the result of this test showed that at the degree of 85 Cent pitch shifted from the root, middle and highest note of triadic structure, 85.7% subjects tend to judge a rising pitch as tension dissonance and falling pitch as gloom dissonance.

In the next section, we compare these two types of dissonances to see which one is more salient under a general dissonance concept.

7.6 Between Gloom and Tension Comparative Test

7.6.1 Overview

In the previous test, listeners clearly identified two types of dissonance in music chords, caused by a lowered and raised tone respectively from a major triadic structure. However, the strength of the general dissonance perceptions these two types of dissonance evoke is also of interest.

In this particular test, the comparison between gloom and tension is made under the same degree of pitch movement and at same level of audibility. The degree of pitch movement is defined by the pitch interval shifted from a pitch component of either I, major III, or V pitch classes. In the previous test, 85 Cents was used as the degree of pitch movement for both upward (tension) and downward (gloom) movements, and it will also be used in this experiment. The audibility describes the strength of a particular pitch being heard, its perception domain features modelled by mutual masking and harmonic masking processes (see section 6.4.3). As these two factors may potentially influence the overall dissonance effect of both gloom and tension, they are kept at the same level for this study.

This experiment also studies how gloom and tension dissonance is affected by the positions of pitch shift: at the root position, at the major third position and at the perfect fifth position. For instance, chord structures A and D in Fig.62 refer to a pitch shift at root position, noted as the first position (1^{st} pos.); chord structures B and E refer to a pitch shift at major third position (3^{rd} pos.); and chord structures C and F refer to a pitch shift at perfect fifth position (5^{th} pos). In this experiment, the same chord structures were studied (chords A– F). The perceived level of both gloom and tension dissonance can thus be compared to see if there is any difference at the three positions of the major triad.

In addition, this test also monitors whether the gloom and tension effect is influenced by absolute frequency. This is achieved by using different tonal centre frequencies (f_c). In the previous experiment, 240Hz was used as the tonal centre frequency. In this experiment, three fundamental frequencies were studied, at 100Hz, 240Hz, and 1100Hz; representing musical chords at bass, alto and treble pitch registers. Thus nine sets of experimental data were expected (see Table 17). Each data set contains listeners' dissonance rankings of both gloom and tension. If the results are consistent across all three fundamental frequencies, this suggests that the harmony perception of gloom and tension is based on intervallic pitch structures, independent of absolute frequency.

Statistical analysis was undertaken across each data group to obtain a quantitative measure of both gloom and tension dissonance.

7.6.2 Synthesizing Triads

As described in previous section, there were a total number of 21 triads synthesised in this experiment. The fundamental frequencies of these 21 triads are shown in Table 17.

Root	100Hz			240Hz			1100Hz		
	bass	**3rd**	**5th**	**bass**	**3rd**	**5th**	**bass**	**3rd**	**5th**
Major Triad	100	125	150	240	300	360	1100	1375	1650
		Chord.A1			*Chord.B1*			*Chord.C1*	
1st pos. down	***95.3***	125	150	***229***	300	360	***1048***	1375	1650
		Chord.A2			*Chord.B2*			*Chord.C2*	
1st pos. up	***105***	125	150	***252***	300	360	***1155***	1375	1650
		Chord.A3			*Chord.B3*			*Chord.C3*	
3rd pos. down	100	***119***	150	240	***286***	360	1100	***1310***	1650
		Chord.A4			*Chord.B4*			*Chord.C4*	
3rd pos. up	100	***131***	150	240	***315***	360	1100	***1444***	1650
		Chord.A5			*Chord.B5*			*Chord.C5*	
5th pos. down	100	125	***143***	240	300	***343***	1100	1375	***1571***
		Chord.A6			*Chord.B6*			*Chord.C6*	
5th pos. up	100	125	***158***	240	300	***378***	1100	1375	***1733***
		Chord.A7			*Chord.B7*			*Chord.C7*	

Table 17 Frequency table for gloom and tension comparative test

The shifted tones are underlined. Chord A2–A7 are the chord structures generated by shifting one of the pitch components 85 cents away from the major triadic structure (Chord A1); and the same applies to chord B series and chord C series

As the experiment requires the shifted tones to be synthesised at equal audibility, we need to convert audibility back to acoustic amplitude by reversing the computation of: harmonic masking → mutual masking → auditory sensitivity (see Section 5.6.2).

Triads were synthesised in Puredata (see Fig.60), by setting the tonal centre frequency to 100, 240, and 1100Hz respectively. Each harmonic tone consisted of eight harmonic partials, and 5dB/partial was used to model the amplitude decay. The amplitudes of fundamental partials were adjusted according to the calculation of equal audibility.

7.6.3 Subjects' profile

A total of 37 subjects participated in this listening test. Based on the conclusion from the previous test (section 6.1.4), we noted that gloom and tension effects are significantly different from each other, regardless of participants' age and country of origins. Therefore, this test did not require participants' age and culture background; no data was collected for demographical analysis. However, each participant had to satisfy the following conditions in order to take part in this test: 1), over 18 years old; 2) no hearing impairment and 3) not musically trained.

12 participants were from the University of York, Department of Electronics; 7 from York social network and these took the English version of the listening test. 18 participants were from the University of Nanjing, Department of Information Management; as their first languages are Mandarin Chinese, they took the Chinese version of the listening test. Prior to the study, ethical approval was granted from the University of York's Physical Sciences Ethics Committee.

7.6.4 Test Procedure

Each participant received an executable application for this test. Participants were required to use headset/earphones to complete the test. On the opening page of the application GUI, the participant saw a consent page (see Fig.66), introducing the purpose, content, requirements, as well as possible risks of this experiment. By clicking the ‘I understand and proceed’ button, participants agreed to take part in this test. Chinese participants will use the Chinese version of the test (see Fig.66, right).

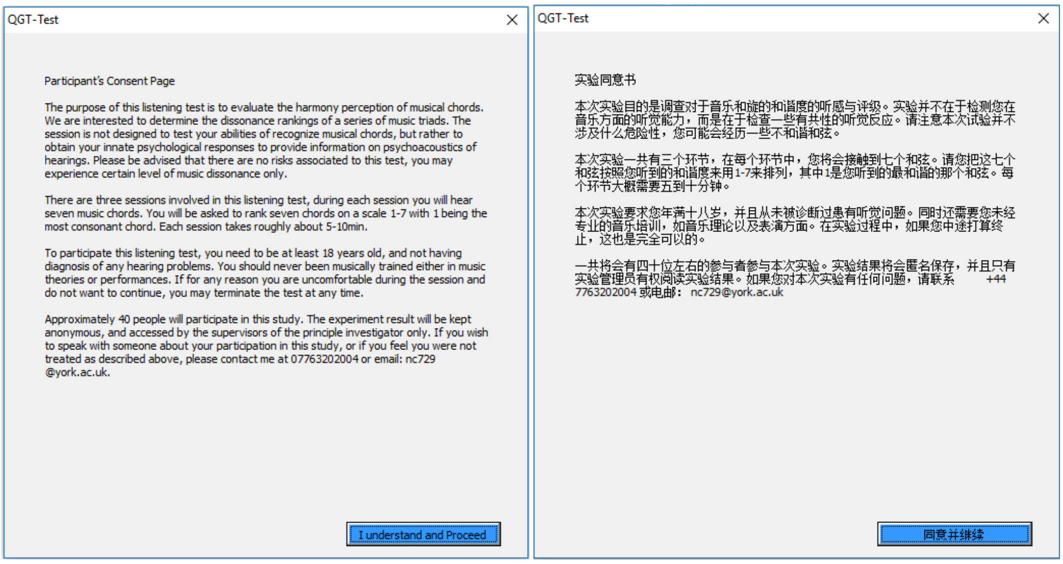


Figure 66 Bilingual consent page

The experiment contained three sections, corresponding to the test with tonal centre at 100Hz (session #1), 240Hz (session #2), and 1100Hz (session #3) respectively. For each session, participants saw an application interface as follows (Fig.67):

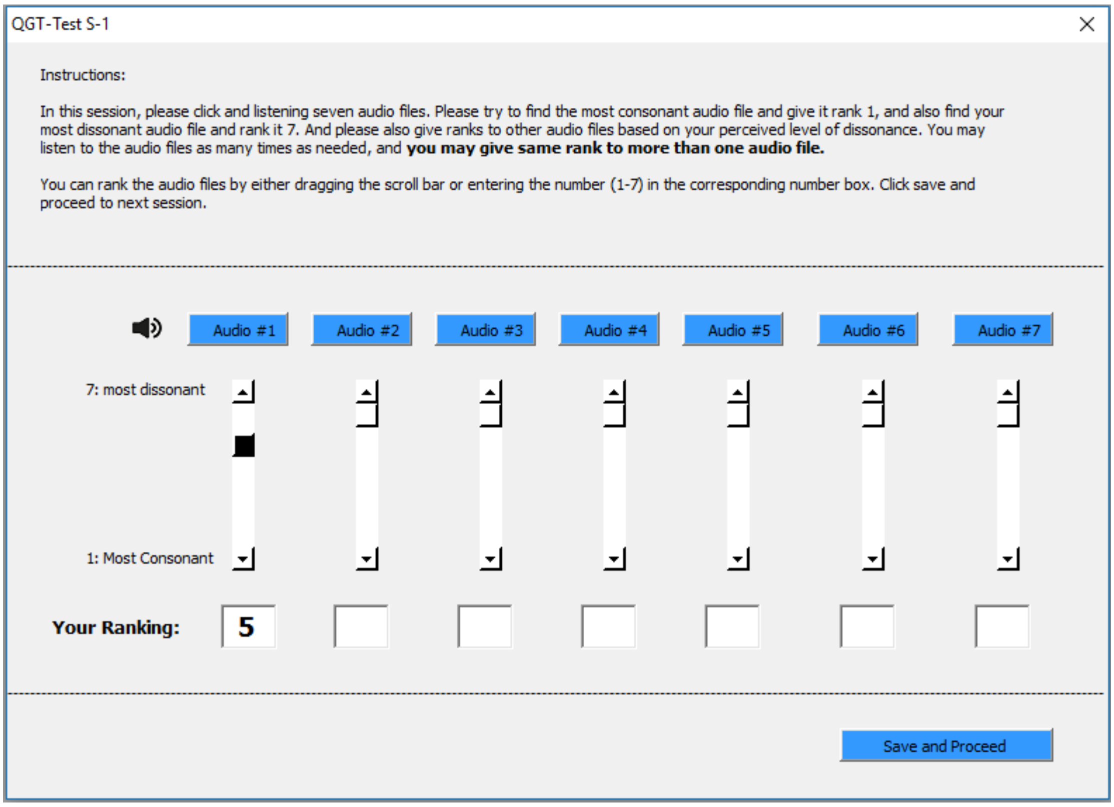


Figure 67 Test page for session #1

Upon opening this page, the synthesised chord A1–A7 was randomly assigned to audio file #1 – #7. Listeners were asked to find the most consonant and most dissonant audio file and assign them with ranking 1 and 7 respectively. They were also asked to rank the rest of audio files on a scale from one to seven. To avoid listeners having the impression that all seven scale ranks were to be used over the seven audio files, the instructions highlighted "you may give same rank to more than one audio file" in bold type. Scroll bars were adjusted to discrete values (1–7) only, and a change of scroll bar position also changed the ranking number in the corresponding number box. Similarly, typing in the ranking number into the text box also set the position of corresponding scrollbar.

Experiment session #2 and session #3 also have the same GUI interface; except that the seven audio files randomly loaded are chords B1–B7 and C1–C7 respectively. By the end of session #3, there is a click to finish button and thank you message box prompt. Five minutes of break were taken between experimental sessions.

7.6.5 Results and Analysis

Out of 37 participants, there was only one participant who did not think that the major triadic structure was the most consonant triad in session #2 of the test. All the other participants consider the major triad as the most consonant chord in all sessions.

The chord structures defined under the *gloom* dissonance were those with one of the tones shifted 85 cents *downwards* from the major triadic structure. These were chords: A2, A4, A6, B2, B4, B6, and C2, C4, C6. The chord structures defined under the *tension* dissonance were those with one of the tones shifted 85 cents *upwards* from the major triadic structure. These were chords: A3, A5, A7, B3, B5, B7, and C3, C5, C7. There were altogether 6 times where the gloom triads were determined as the most dissonant chord; in the remaining cases, tension chords were deemed as the most dissonant chord structure. The entire dissonance rankings of gloom and tension given by 37 participants are demonstrated in Fig.68.

The overall mean and standard deviation for the gloom dissonance group are $\bar{x}_g = 3.42$; $s_g = 1.203$; and the overall mean and standard deviation for the tension dissonance group are $\bar{x}_t = 5.954$; $s_t = 1.226$. The independent T-test shows that the two groups are significantly different from each other ($p<0.001$).

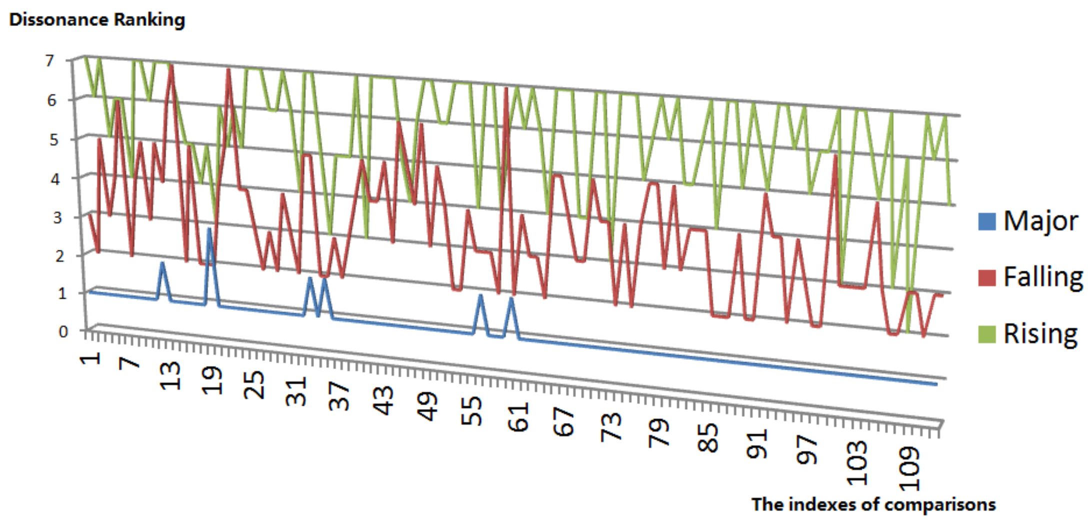


Figure 68 Plot of dissonance rankings

The two sample covariance f-test that compares gloom and tension data groups gives an f-value of 0.9473, which is higher than the critical value of 0.5 (significance level is set to 0.05; degree of freedom = 332 as 331 data points were involved). Such observation implies that the variance of the gloom chord structures has a significantly higher value than that of the tension chord structures.

Linear regression analysis with minimum square error gives a slope that D_g/D_t = 0.4048 (to consider the entire data point from all clusters A–C), where D_g is the dissonance salience of gloom; D_t is the dissonance salience of tension (see Fig.69). This ratio means that the overall perceived dissonance effect for gloom is 40.48% of the tension effect under the condition of equal audibility.

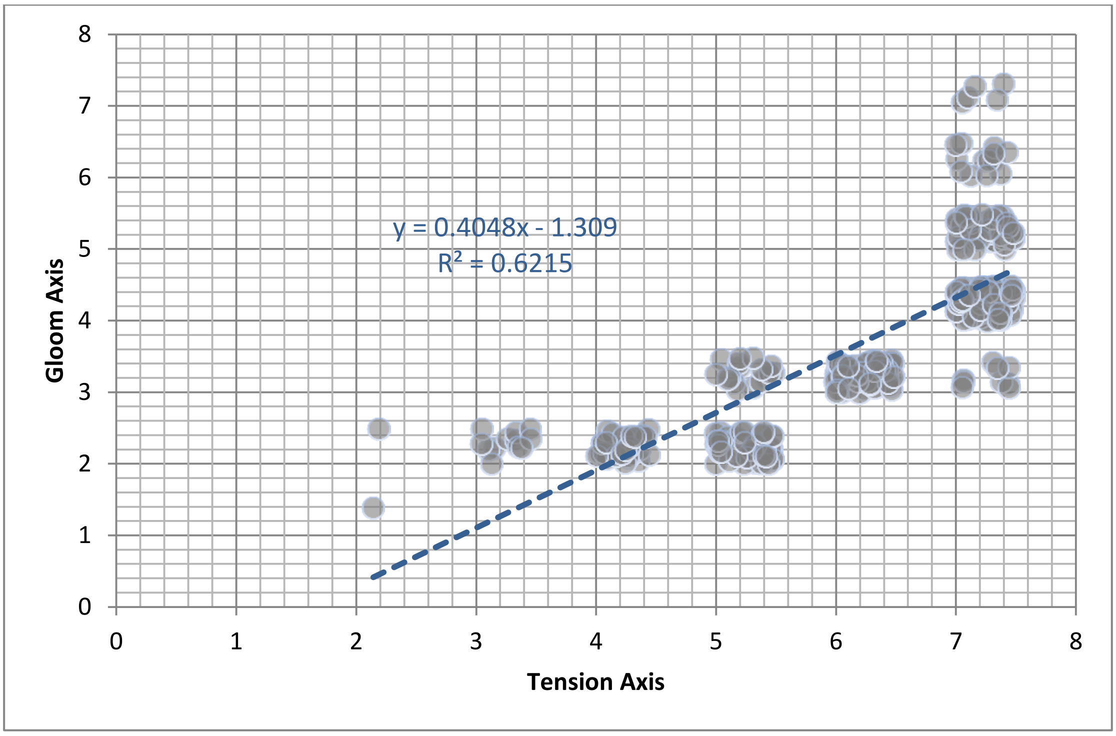


Figure 69 Linear regression analysis of gloom/tension.

7.6.5.1 Gloom and tension comparison at different tonal centre frequencies

One of the objectives of this experiment was to study the frequency dependency of the gloom and tension emotions, which can be viewed from the cross-session comparison result. In each session, there were three chord structures (#3, #5, #7) corresponding to the rising pitch properties and also three chord structures (#2, #4, #6) corresponding to the falling pitch properties. Therefore, the experimental data from chord structures #3, #5, #7 was combined across three positions to represent the tension emotion, and data from chord structures #2, #4, #6 were combined to represent the gloom emotion. The mean and standard deviation was calculated and is summarised in Table 18.

	Session 1 (100Hz)		Session 2 (240Hz)		Session 3(1100Hz)	
	Chord G	Chord T	Chord G	Chord T	Chord G	Chord T
Mean	3.541	5.792	3.278	6.067	3.189	5.948
SD	1.892	2.048	1.696	1.947	1.605	2.244

Table 18 Mean and Standard Deviation (SD) values for chords at different tonal centre frequencies

Chord G: chord structures with gloom emotion (chord structures #2, #4, #6); Chord T: chord structures with tension emotion (chord structures #3, #5, #7)

To test if the dissonance level of gloom and tension changes over tonal centre frequencies, a two sample T-test was performed to see if experimental data from the different sessions were significantly different from each other. The p-values were obtained and are summarised in Table 19.

p -value		Session 1 (100Hz)		Session 2 (240Hz)		Session 3(1100Hz)	
		Chord G	Chord T	Chord G	Chord T	Chord G	Chord T
Session 1 (100Hz)	Chord G	-	<0.001	0.5309	<0.001	0.391	<0.001
	Chord T	-	-	<0.001	0.5557	<0.001	0.7558
Session 2 (240Hz)	Chord G	-	-	-	<0.001	0.8173	<0.001
	Chord T	-	-	-	-	<0.001	0.8082
Session 3 (1100Hz)	Chord G	-	-	-	-	-	<0.001
	Chord T	-	-	-	-	-	-

Table 19 Two sample T-test across different tonal centre frequencies

Based on the result from Table 18, we may conclude that:

(1) The tension chord classes from the three experimental sessions are significantly different from any other gloom chord classes, meaning that gloom and tension dissonance can be distinguished by listeners regardless of the variable of tonal centre frequency. (When $p<0.01$, the null hypothesis H_0, that there is no difference between two data groups, is rejected).

(2) The tension chord classes between the three experimental sessions are NOT significantly different from each other, meaning that the perceived degree of dissonance does not vary much as a function of tonal centre frequencies. The same conclusion can also be made for the gloom chord classes. (When $p>>0.05$, the null hypothesis H_0, that there is no difference between two data groups, is accepted).

7.6.5.2 Gloom and tension comparison at different referential pitch classes

In major chord structures, there are three pitch classes involved, namely I, major III, and V. Correspondingly, they serve as three 'reference points' to determine the rising and falling pitch properties for other chord structures. A second objective of this experimental study was to observe any differences (in terms of perceived dissonance rankings) when the rising and falling pitch properties take place at each of these three positions. To perform a statistical analysis, the experimental data of chord structures #2 – #7 were merged across the three sessions and their mean and standard deviation were calculated (see Table 20).

	F/R at I pitch class		F/R at major III pitch class		F/R at V pitch class	
	Ch.S. #2	Ch.S. #3	Ch.S. #4	Ch.S. #5	Ch.S. #6	Ch.S. #7
Mean	3.414	5.854	3.36	6.0631	3.243	5.682
SD	1.249	0.942	1.883	1.647	1.374	1.485

Table 20 Two-sample T-test across different pitch classes (I, major III, V)

where Ch.S. is abbreviation for chord structure and F/R refers to falling/ rising pitch properties

Based on the experimental data from Table 19, we may observe that the highest standard deviations are at the major III pitch class (1.883 and 1.648) for both rising and falling pitch properties; and the lowest standard deviations are at the I pitch class (0.942 and 1.249). Such an observation means that listeners tend to have a better consistency for distinguishing gloom and tension emotions introduced at I pitch class and worst consistency at major III pitch class. A possible explanation for such a phenomenon is that I pitch class has the highest ‘stability’ in the tonal structure (highest weight in virtual pitch determination), therefore its rising and falling pitch properties becomes easier for listeners to identify. Conversely, the major III pitch class has the worst stability (lowest weight in virtual pitch determination) compared to I and V pitch classes, therefore its dissonance perception rankings are more scattered.

To further investigate whether the gloom and tension dissonances are significantly different in reference to I, major III, V pitch classes, a two-sample T-test was computed, based on the mean and standard deviation between two data groups. The p-values were obtained and are summarised in Table 21.

p -value		R/F at I pitch class		R/F at major III pitch class		R/F at V pitch class	
		Chord #2	Chord #3	Chord #4	Chord #5	Chord #6	Chord #7
R/F at I pitch class	Chord #2	-	<0.001	0.8848	<0.001	0.5616	<0.001
	Chord #3	-	-	<0.001	0.3252	<0.001	0.5538
R/F at major III pitch class	Chord #4	-	-	-	<0.001	0.753	<0.001
	Chord #5	-	-	-	-	<0.001	0.1902
R/F at V pitch class	Chord #6	-	-	-	-	-	<0.001
	Chord #7	-	-	-	-	-	-

Table 21 Mean and standard deviation for different chord structures

Where Ch.S. is the abbreviation for chord structure, and R/F means the rising and falling pitch properties

Based on the result from Table 20, we may conclude that:

(1) The tension chord classes from the three experimental sessions are significantly different from any other gloom chord classes. This means that gloom and tension dissonance can be distinguished by listeners regardless of the variable of the

referential pitch classes I, major III, and V. (When $p<0.01$, the null hypothesis H_0, that there is no difference between two data groups, is rejected).

(2) The tension chord structures between the three referential pitch classes are NOT significantly different from each other, meaning that the perceived degree of dissonance does not vary much as a function of referential pitch class. The same conclusion can also be drawn for the gloom chord classes. (When $p>>0.05$, the null hypothesis H_0, that there is no difference between two data groups, is accepted).

7.6.6 Conclusions

This listening experiment verifies that the perceived gloom emotion of musical chords is associated with falling pitch properties derived from the major chord structure; and the tension emotion is associated with rising pitch properties derived from the major chord structure. Furthermore, such observations do not change as a function of tonal centre frequency or the positions (I, major III, V pitch classes) of the major chord structure.

7.7 Summary

In this chapter, a theoretical approach that incorporates music emotions (gaiety-gloom, relaxation-tension) into the analysis of music harmony is presented. It is proposed that musical gloom and tension dissonances will be perceived when one of the notes from major triadic structure is slightly lowered or raised. If one of major triadic tone components is raised by one or two semitones, listeners perceive tension; otherwise if one of the major triadic tone components is lowered by one or two semitones, listeners perceive gloom. These theories are supported by the experimental data from listening tests.

Furthermore, as both gloom and tension emotions contribute to the overall dissonance of a chord structure, in a comparative experimental study, it was indicated that the tension emotions have a more significant role for the overall dissonance perception compared to gloom emotions: D_g: $D_t \approx 2:5$.

The proposed theoretical approach has also pointed out that the perception of music harmony should be multi-dimensionally natured, characterised by at least two uncorrelated psychological dissonances: gloom and tension emotions. The multi-dimensional theoretical analysis therefore presents a new perspective with which to interpret the empirical rankings of musical chords where previous analytical theories/models have generally failed. In next chapter, a detailed discussion of the multi-dimensional harmonic analysis will be included.

Chapter 8 Pitch-Based Multi-Dimensional Analysis

In previous chapters, the harmonic functions for each pitch component have been modelled, including *sensory dissonance* (in Chapt.5), *tonal consonance* and *ambiguity dissonance* (Chapt.6) and *gloom* and *tension dissonances* (in Chapt.7). In this chapter, these harmonic functions are integrated to analyse the perception of some common musical chords (triads and tetrads) resulting in a four-dimensional consonance and dissonance concept (*CDC*) model. The chapter is structured as follows:

Section 8.1 introduces the pitch-based distributed harmonic system for the analysis of a multi-tone structure. The distributed harmonic system refers to the predicted harmonic functions over all audible pitch components within the input sound stimuli. Compared to the overall consonance and dissonance measurement, the distributed harmonic system can provide more tonal harmonic details for the understanding of music harmony.

In section 8.2, a two-step analytical method is introduced that integrates the multi-dimensional distributed harmonic functions into a one-dimensional overall consonance and dissonance measurement.

In section 8.3, the empirical perception rankings of musical triads are used to train an algorithm to estimate the correlation coefficient between each type of dissonance function to the overall dissonance perception.

Section 8.4 applies the model derived in the previous sections to analyse the perception of musical tetrads.

Section 8.5 links the multi-dimensional harmonic features to the discussion of the 'quality' (or tone colour) of musical chords.

Section 8.6 summarizes this chapter.

8.1 Distributed Harmonic Functions

The concept of multi-dimensional harmonic analysis is developed based on the assumption that there is more than one type of psychological experience associated with the musical CDC. In previous chapters, the degrees of five types of harmonic functions have been estimated with respect to each noticeable pitch component. The model therefore outputs a distribution of harmonic functions. A typical output data structure is illustrated in Table 22.

	Noticeable Pitch Components	#1	#2	#3	#n
Harmonic Functions	Sensory Dissonance (D_s)	data	data	data	data
	Tonal Consonance (C_t)	data	data	data	data
	Ambiguity Dissonance (D_a)	data	data	data	data
	Gloom Dissonance (D_g)	data	data	data	data
	Tension Dissonance (D_t)	data	data	data	data

Table 22 Data structure of pitch based multi-dimensional harmonic analysis

The five harmonic functions can be analysed in four principal dimensions:

a. Sensory consonance to sensory dissonance
b. Tonal consonance to ambiguity dissonance
c. Tonal consonance to gloom dissonance
d. Tonal consonance to tension dissonance

The 1st dimension is the sensory consonance and dissonance dimension, measured through the beating effect introduced by a pitch component. Zero beats sensations imply an extreme sensory consonance concept. Differing from other harmonic functions, the sensory dissonance function does not require the information of tonal context, therefore it does not have tonal meaning.

The 2nd, 3rd and 4th dimensions are **tonal** consonance and dissonance dimensions. A common feature between these dimensions is the influence of the tonal centre concept.

All three types of tonal dissonance concepts (ambiguity, gloom and tension) are contrasted to the same tonal consonance concept.

By combining sensory and tonal consonance into a single consonance concept, the dissonance effect of each tone component can be measured in four directions: sensory, ambiguity, tension, and gloom dissonances. The pitch-based model estimations of sensory dissonance (D_s), ambiguity (D_a), tension (D_t), and gloom (D_g) dissonances have provided a distribution of dissonance effect over the frequency axis. A geometric representation is illustrated in Fig.70.

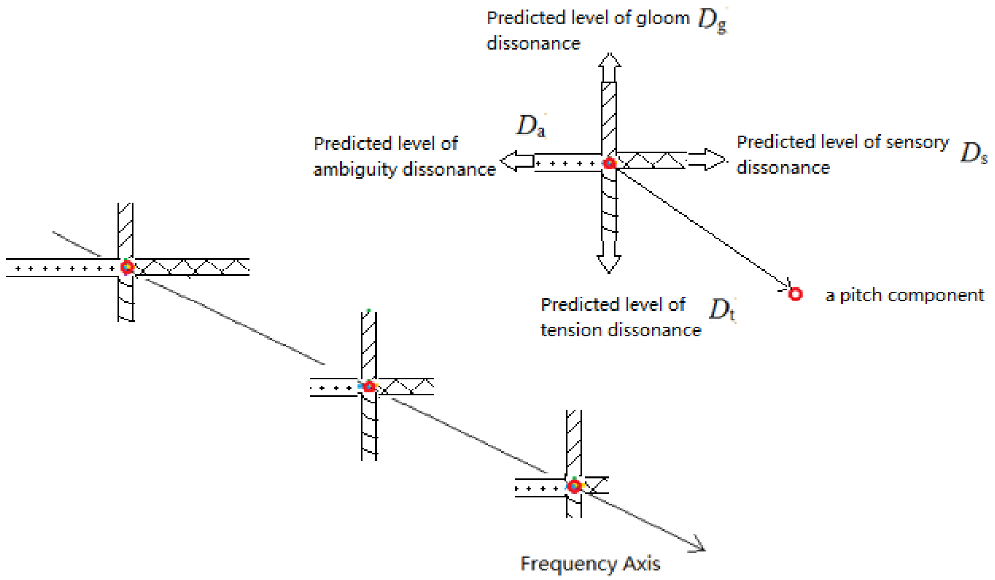


Figure 70 Geometrical representations of dissonance effects

In Fig.70, we may observe that for a sound entity with multiple noticeable pitch component (such as a musical chord), each pitch component has a unique dissonance property in term of predicted levels of *gloom, tension, and ambiguity* and *sensory*. The distribution of these dissonance properties may be used to denote the perceived 'quality' of a musical chord (or multi-tone structure)

8.2 Integrated Consonance and Dissonance Concept

Note so far the discussion of harmonic functions has referred to a particular pitch component, not the entire musical structure. To estimate the overall harmony of a musical entity, two tasks are involved:

(1) Estimating the overall harmonic effect with respect to each harmonic function; this is modelled by the summation according to the power law of psychological significance (β):

$$D_{(s/a/t/g)} = \Sigma\, [D_{(s/a/t/g)}\,(f_i)]^{\beta}$$

Where β is the significance factor, β =0.75 is used in this approach (Zwicker, 1970); and f_i is the fundamental frequency of an audible pitch component.

(2) Determines relative correlation coefficients (α) between each type of dissonance function with the overall dissonance concept, and uses a multivariable summation function to estimate the overall dissonance level:

$$D = \alpha_s\,(D_s)^{\beta} + \alpha_a\,(D_a)^{\beta} + \alpha_t\,(D_t)^{\beta} + \alpha_g\,(D_g)^{\beta}$$

Where α_s, α_a, α_t and α_g are the correlation coefficients of sensory dissonance (D_s), ambiguity (D_a), tension (D_t), and gloom (D_g) dissonances; correlation coefficients measure the degree to which a particular type of dissonance function contributes to the overall dissonance perception. A higher value of correlation coefficient means the dissonance effect of this particular type of dissonance is stronger than others. β is the psychological significance factor, β=0.75 according to (Zwicker, 1970).

The determination of which types of dissonance evoke a stronger dissonance effect as compared to others can be problematic. One way of obtaining a quantitative comparison result is using a listening perception test. In the previous chapter, one such experimental study was reported which compared the tension and gloom dissonance effects. However, the difficulty of conducting a perception test is that neither type of dissonance effect can be completely isolated; one type of dissonance function is usually accompanied by another for comparison. This research proposes a numerical approach combining empirical data from several perception studies, to study the overall harmony perception of a musical entity. The details of this numerical method are discussed in the next section.

8.3 Estimating Correlation Coefficients

To estimate the correlation coefficient (α), we may use a series of chord structures as the training data for the multivariable summation system:

$$D = \alpha_s (D_s)^{\beta} + \alpha_a (D_a)^{\beta} + \alpha_t (D_t)^{\beta} + \alpha_g (D_g)^{\beta}$$

The values of D_s, D_a, D_t, D_g, representing the normalized sensory, ambiguity, tension, and gloom dissonance effects, are obtained from model predictions, and the overall dissonance effects D are obtained from empirical rankings. The normalization algorithm is designed such that the musical structures with the highest level of each dissonance function were kept at one. Therefore, the value of D_s, D_a, D_t, D_g are bounded between 0~1.

In the literature, the empirical studies on the consonance/dissonance perception rankings for musical triads have made the following conclusions:

- Roberts (1986) studied the perception of four types of triadic structures and concluded that: the perceived level of consonance decreases in the order of: major > minor > diminished > augmented, and this perception order remains valid when these chords are at inverted positions.
- The empirical study of (Johnson-Laird et al., 2012) demonstrated that the perception of suspended 4th triads (and its inversions) is generally more consonant than diminished chords but less consonant than the minor chords.
- Cook's empirical study (Cook, 1999) showed that the triadic structures with one whole tone interval and one semitone interval from the fifth (what he referred as *mild dissonance*) are more dissonant than augmented triads, and the triadic structures with one semitone interval and one semitone interval from the fourth are perceived as the most dissonant chord structures (what he referred as *sharp dissonance*).

Based on the above-mentioned literatures, perception rankings of all triadic structures can be organized into a table (see Table 23).

Triads	Intervallic Structure			Empirical Dissonance Ranking
	Root Position	1st Inversion	2nd Inversion	
Major	4-3	3-5	5-4	1
Minor	3-4	4-5	5-3	2
Suspended 4^{th}	5-2	2-5	5-5	3
Diminished	3-3	3-6	6-3	4
Augmented	4-4	4-4	4-4	5
$Mild_1$	2-7	7-3	3-2	6
$Mild_2$	2-8	8-2	2-2	7
$Sharp_1$	1-6	6-5	5-1	8
$Sharp_2$	1-7	7-4	4-1	9

Table 23 empirical rankings of triadic structures

In Table 23, the column of intervallic structure refers to the number of semitones between the lowest and middle note – middle to highest note; and larger number of empirical dissonance ranking means it is a less consonant chord structure

An implied condition of Table 23 is that the perceived rankings of these chord structures are not influenced by their inversions. The concept of inverted chords is strongly associated with those chord structures with clear tonal centres (e.g. major and minor structures), such that when the chord structures are inverted, the perception of a root pitch class is retained. The chord inversion concept is however trivial for those unstable chord structures, as even at their 'root' position, the root is weakly perceived. In these cases, the lowest note serves as the root of chord. Therefore, the perception rankings presented in Table 23 are based on the root of a chord to be located as the lowest note for comparison, otherwise ambiguities are introduced for those chord structures with weak tonal centres.

In order to obtain α_s , α_a , α_t , α_g , for each chord structure, the amount of each type of dissonance is first predicted (D_s, D_s ,D_s, and D_s). With this data and empirical rankings of the overall dissonance D (Table 23), we can obtain a series of relationships (depending on the number of chord structures used) with four unknown parameters (α_s , α_a , α_t , α_g) to be determined.

In the implementation stage, four types of dissonance functions with respect to each chord structure in Table 23, were synthesised for each triad in Matlab, with the standard harmonic overtone series (see Fig.47 in the sensory dissonance test, Chapter 5). 12 overtone series are included, and the tonal centre frequency (f_c) is selected at 261Hz (corresponding to note C_4) in this simulation. The model estimation results are presented in Table 24–26 below, the strongest harmonic function of each chord structure is highlighted in bold type.

Triads (Root Position)	**D_s**	**D_a**	**D_t**	**D_g**
Major	**0.013**	0.048	0.000	0.000
Minor	0.162	0.117	0.067	**0.673**
Suspended 4th	0.608	**0.813**	0.169	0.353
Diminished	0.506	0.215	0.112	**1.000**
Augmented	0.374	0.451	**0.749**	0.227
Mild1	**0.894**	0.153	0.271	0.093
Mild2	**0.787**	0.395	0.638	0.182
Sharp1	**1.000**	0.416	0.592	0.103
Sharp2	**0.770**	0.436	0.683	0.362

Table 24 Model simulation result for triadic structure (root position)

Strongest harmonic function highlighted in bold, D_s= sensory dissonance, D_a= ambiguity dissonance, D_t= tension dissonance, D_g = gloom dissonance

Triads (1st Inversion)	D_s	D_a	D_t	D_g
Major	**0.226**	0.022	0.000	0.000
Minor	0.105	0.186	0.049	**0.663**
Suspended 4th	0.778	**1.000**	0.342	0.514
Diminished	0.346	0.403	0.104	**0.919**
Augmented	0.418	0.167	**0.575**	0.047
Mild1	**0.965**	0.29	0.152	0.147
Mild2	0.608	0.492	**0.846**	0.052
Sharp1	0.919	0.256	0.661	0.089
Sharp2	0.970	0.517	0.82	0.326

Table 25 Model simulation result for triadic structure (1st Inversion)

D_s= sensory dissonance, D_a= ambiguity dissonance, D_t= tension dissonance, D_g = gloom dissonance

Triads (2nd Inversion)	D_s	D_a	D_t	D_g
Major	**0.249**	0.059	0.000	0.000
Minor	0.393	0.277	0.046	**0.784**
Suspended 4th	0.697	**0.971**	0.172	0.477
Diminished	0.265	0.128	0.146	**0.836**
Augmented	0.560	0.191	**0.806**	0.079
Mild1	**0.851**	0.302	0.246	0.138
Mild2	0.929	0.438	**1.000**	0.083
Sharp1	**0.796**	0.378	0.793	0.452
Sharp2	**0.249**	0.399	0.130	0.560

Table 26 Model simulation result for triadic structure (2nd Inversion)

D_s= sensory dissonance, D_a= ambiguity dissonance, D_t= tension dissonance, D_g = gloom dissonance

From the model simulation result, a total number of 27 sets of training data (corresponding to 27 musical chords in Table 21) were used in the multivariable interpolation analysis, and the statistical analysis yielded: α_s =11.29, α_a =2.15, α_t =6.73, andα_g =1.81. Therefore, the overall dissonance rankings (D) can be numerically modelled by:

$$D = [11.29\,(D_s)^{\beta} + 2.15\,(D_a)^{\beta} + 6.73\,(D_t)^{\beta} + 1.81\,(D_g)^{\beta}] \,/\, C_t$$

Based on the model simulation result in Table 24–26, the following phenomenon can be observed:

(1) The tonal dissonance functions (ambiguity, gloom and tension) are not strongly influenced by the position of chords. This can be observed by comparing the degree of tonal dissonance effects between Table 24–26. One of the major reasons for this phenomenon is that the tonal centre frequency was fixed at 261Hz, not determined by the internal structure of each chord. Such an observation is in line with Robert's empirical study (Robert, 1986) that the perception rankings of major, minor, diminished and augmented chords are not influenced by their inversions.

A fixed tonal centre was used to reflect the practice in perception psychology that when listeners are asked to judge their perception of a series of chords, they tend to refer to the most stable chord structure – major triadic structure as a reference. A 'tonality' set up by the major triads infers the tonal centre information, which has a perceptual impact on the listener's judgement for other chord structures.

Another reason to use a fixed tonal centre in this simulation was mentioned earlier: the tonal centre perceptions for unstable structures are relatively weak. To determine the dissonance in reference to a tonal centre, therefore, this centre must be pre-defined.

(2) Based on the correlation coefficients obtained in this numerical study, it can be observed that the sensory dissonance (α_s =11.29) has a dominant role that is able to mask the presence of other types of dissonant functions in most cases. The most salient tonal dissonance function is identified as the tension dissonance function (α_t =6.73). The dissonance impact ratio between tension and gloom dissonance is 6.73/1.81 = 3.71 which is slightly higher than that obtained in the gloom-tension

comparative test presented in the previous chapter (≈2.5). However, it is clear that the tension dissonance effect is perceived to be stronger than the gloom dissonance in terms of general psychological dissonance in both studies. The ambiguity correlation coefficients (α_a =2.15) are determined in this particular model to be higher than gloom dissonance but less than the tension dissonance. But it is not significantly different from the gloom dissonance (α_g =1.81).

(3) Besides the major triads, the dissonance of all other chord structures was observed due to different types of dissonances.

For minor triads, the dissonance function with the highest prediction level (normalized) is the gloom dissonance in root, 1st inversion and 2nd inversion positions; and similar conclusions can be made for diminished triads. A primary reason for their gloom dissonance is probably the lowered notes (one semitone lower on the highest note from major triad to minor triad (root position), and for diminished triads, both the middle and highest note has been lowered from the major triadic structure.

For suspended 4th triads, a higher degree of ambiguity dissonance was observed. This is primarily due to the sub-dominant function of the 4th note in relation to the root pitch class, therefore introducing a strongly competing tonal centre against the tonal centre of the tonic (root). This is an expected result based on the theoretical analysis presented in Chapter 6.

The augmented triads have been associated with tension dissonances. This is primarily due to the raised highest note (root position) compared to the major triad.

The chord structures with a relatively high degree of sensory dissonance were generally perceived as the most dissonant chord structures: mild dissonance and sharp dissonance have the dissonance rankings from 6 to 9. And in contrast, the major, minor, diminished and augmented chords have a relatively low level of sensory dissonance. This is possibly one of requirements for the gloom, tension dissonance effect to be noticed (instead of being masked by the sensory dissonance effect).

8.4 Harmonic Analysis for the Seventh Chord

A musical tetrad is a four-note musical chord, and one of the most common tetrads seen in music harmony is the seventh chord. In the literature, few empirical tests have been conducted to obtain complete harmonic perception rankings for the tetrad structures. In this particular study, a slightly different approach is used, that is to analyse the dissonance functions with respect to each tone component within the tetrad structure. The merit of such an approach is the ability to visualize which and how specific note components contribute to the overall harmony perceptions.

Under the 12-tet equal temperament system, although each of the 12 pitch classes may have more than one predicted tonal harmonic function, very often, there is only one dominant type of harmonic function characterized by that pitch class. In the computational model, the dominant feature of each pitch class can be identified by comparing the normalized value of $D_a(P_c)$, $D_t(P_c)$, $D_g(P_c)$, and $C_t(P_c)$; where the normalization method is the same as in the previous section. The dominant harmonic function is assigned as:

Tonal function (P_c): max [$D_a(P_c)$, $D_t(P_c)$, $D_g(P_c)$,$C_t(P_c)$;]

Based on numerical estimations, the 12 pitch classes are featured with the following tonal harmonic functions (see Fig.71):

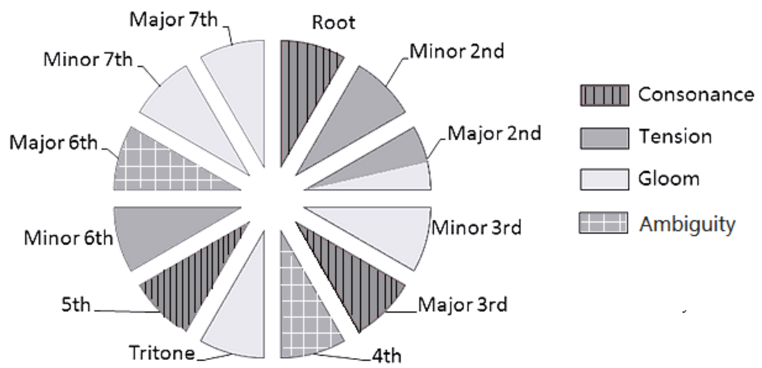


Figure 71 Main harmonic functions of each pitch class.

Based on the conclusions from Fig.71 we may analyse the harmonic function of musical tetrads. As discussed in chapter 2, the seventh chords have been classified into the following categories (Table 27):

Chord name	Intervals above the root			
	1st note	2nd note	3rd note	4th note
Major seventh	unison	major third	perfect fifth	major seventh
Major minor seventh	unison	major third	perfect fifth	minor seventh
Minor major seventh	unison	minor third	perfect fifth	major seventh
Minor seventh	unison	minor third	perfect fifth	minor seventh
Diminished major seventh	unison	minor third	diminished fifth	major seventh
Half-diminished seventh	unison	minor third	diminished fifth	minor seventh
Augmented major seventh	unison	major third	augmented fifth	major seventh
Augmented minor seventh	unison	major third	augmented fifth	minor seventh

Table 27 Common 7th chords and their structures

By then substituting the corresponding harmonic functions it is possible to obtain a distribution of harmonic functions for musical tetrads (See Table 28).

Chord name	Intervals above the root			
	1st note	2nd note	3rd note	4th note
Major seventh	consonance	consonance	consonance	gloom
Major minor seventh	consonance	consonance	consonance	gloom
Minor major seventh	consonance	gloom	consonance	gloom
Minor seventh	consonance	gloom	consonance	gloom
Diminished major seventh	consonance	gloom	gloom	gloom
Half-diminished seventh	consonance	gloom	gloom	gloom
Augmented major seventh	consonance	consonance	tension	gloom
Augmented minor seventh	consonance	consonance	tension	gloom

Table 28 Common 7th chords and their internal harmonic functions

It is interesting to note from Table 28 that the *gloom* dissonance appears quite commonly as opposed to the *tension* dissonance (seen only in augmented major 7^{th} chord). The tension chords are usually treated as a special chord, for example in classical music, and used sparingly to emote the listener in a particular way. The different application of these types of dissonance again may infer that *tension* has a more dominant dissonant effect than *gloom*, and their treatment within compositions in view of this hypothesis would be worthy of further investigation. On the whole, the pitch-based approach presented in this section provides a theoretical thread to analyse not only the consonance and dissonance perceptions, but also the perceived 'qualities' of musical chord structures.

8.5 Analysing the Quality of Musical Chords

Editing the quality of a chord is a main consideration in modern music composition. Musical intervals that were considered as dissonant in classical music theories were added to individualize the artist's musical styles. In this sense, the perception of a musical chord exceeds what CDC can cover.

It was mentioned in chapter 2 that each chord structure can be associated with a unique 'quality' within a tonal music structure. There are also many indications that the perception trait for these 'qualities' can be universally observed across different cultures. For example, the minor third interval was typically described by 'sad' and 'dark'; diminished chords are usually perceived as 'even darker', 'melancholic', and 'distressed'; and augmented chords are described by 'tense' and 'sharp' etc. Many music artists have drawn an analogy between 12 chromatic scales and 12 colour schemes, such as the colour wheel music theory[38]. Sam Winder (Winder, 2012) had promoted a colour mixing strategy to analyse the colour of chords (see Fig.72).

---

[38] Colour wheel theory has been viewed as a new science which correlates the perception of music entities to visual colour representations.

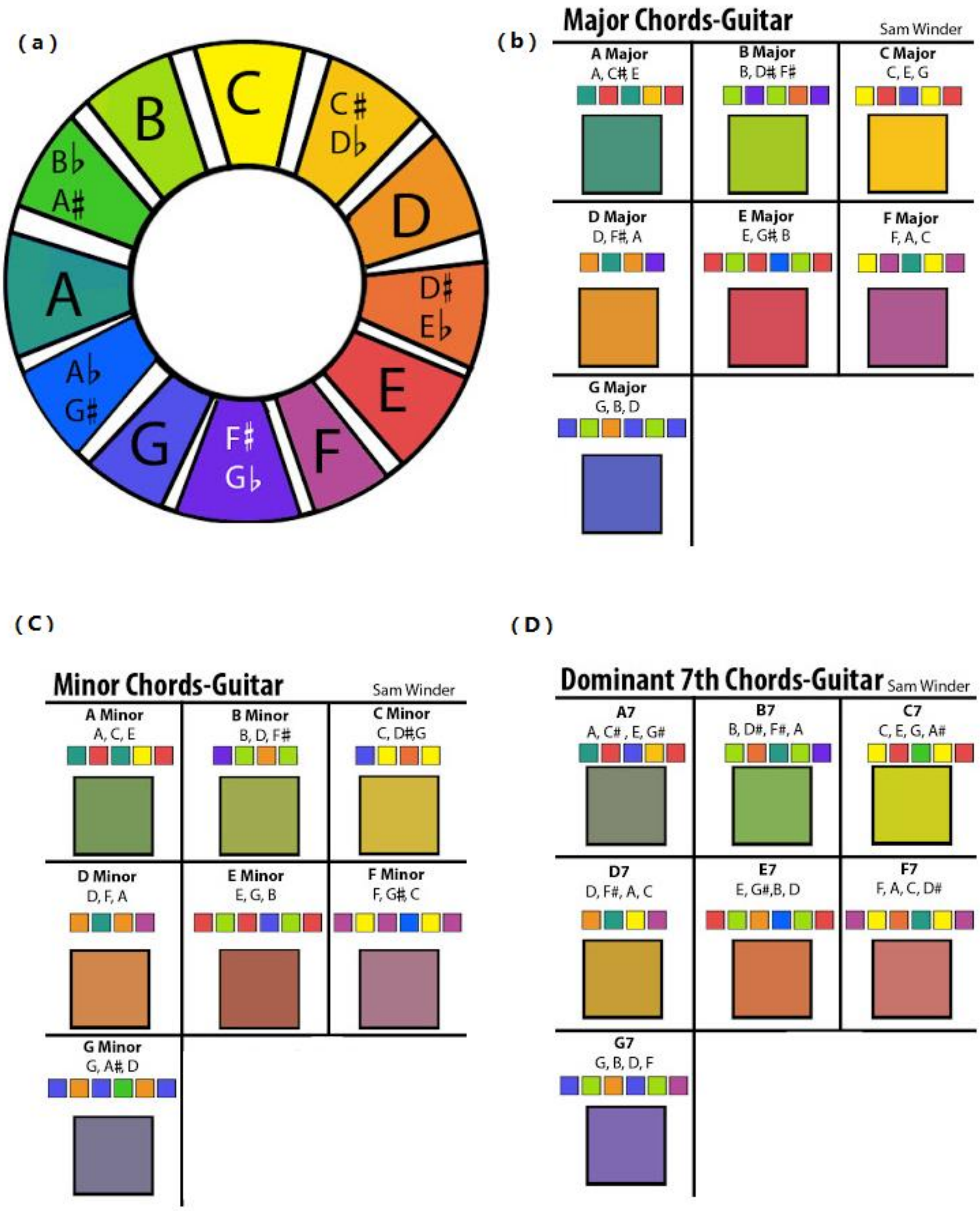


Figure 72 Winder's visual representation of musical chords (Winder, 2012)

(A) fundamental colour 'cues' for 12 pitch classes

(B) mixed colour for major chords

(C) mixed colour for minor chords

(D) mixed colour for dominant 7th chords

Figure reproduced from (Winder, 2013)

However, using a colour-mixing scheme to model the quality of a chord can be inaccurate for the reason that very often the quality of a chord cannot be simply mixed to create a new timbre as colours do. Also, colour perception is highly individual and heavily influenced by culture and entrainment (Johnson-Laird et al, 2012). The choice of using specific colours to denote a specific type of music emotion may therefore not be well-accepted across all demographic groups. In this research, we propose a pitch-based harmonic functional analysis to analyse the quality of a chord in terms of the degree of sensory dissonance, tonal consonance, ambiguity, gloom, and tension dissonances. For instance, the quality of the major chord structure is characterised by the tonal consonance ingredients; the quality of a minor chord structure is characterised by the gloom dissonance at the middle note (root position); and the qualities of diminished and augmented structure are characterized by the tension and gloom dissonances occurring at the highest note (root position).

8.6 Summary

In this chapter, a multi-dimensional harmonic analysis is introduced. The principal dimensions of *CDC* are identified to be:

(1) sensory consonance – sensory dissonance

(2) tonal consonance – ambiguity dissonance

(3) tonal consonance – tension dissonance

(4) tonal consonance – gloom dissonance

Making use of these principal dimensions, the empirical perception rankings of musical triads can be interpreted as follows:

The dissonance perceptions of minor and diminished triads are mainly attributed to the gloom dissonance; and diminished triads generally contain a stronger gloom dissonance effect, and are therefore are perceived as more dissonant than the minor triads.

The dissonance perceptions of augmented triads are more strongly influenced by the tension dissonance introduced by the augmented fifth interval; as the correlation coefficient for the tension dissonance is much larger than that of the gloom dissonance; the augmented triads are perceived to be more dissonant than the diminished triads.

The ambiguity dissonance is mainly apparent for the triads with competing tonal centres (see Section 6.2). The ambiguity effect is estimated to be stronger than the gloom dissonance effect, but weaker than the tension dissonance effect; therefore, this explains why suspended 4^{th} triads are perceived to be more consonant than augmented chords, but less consonant than diminished chords.

The impact of sensory dissonance is dominant; it is able to mask any other types of tonal dissonance when it becomes noticeable within a chord structure. In other words, a strong sensory dissonance effect is able to completely mask the tonal structure (tonal centre perception) and breaks the perception of tonal consonance and dissonance.

Chapter 9 Conclusion and Future Work

This research has identified four types of uncorrelated musical dissonance concepts (*sensory*, *ambiguity*, *gloom* and *tension*) and the lack of each type of dissonances is considered consonance. This research also indicates that the consonance and each type of the dissonance concepts correspond to a specific type of (1) musical entity/structure, (2) psychological experience and (3) psychophysical interpretation. This can be viewed in Tab 29.

Proposed Harmonic Concepts	Musical Entity	Experience/ Descriptors	Psychophysical Interpretations	Numerical Approaches
'Consonance'	Major triadic structure; resolution of dissonances; tone components confirming 'tonal centre';	Lack of dissonances; 'Tone-like', Clear tonal centre concepts; Relaxed, Resolved, Pleasant, Cheerful, Stable, Smooth, Tonal...Etc.	Tonal Fusion Theory; Harmonic Relation Theory (Rameau and Wundt, 1721); Frequency Ratio Theory (Sethare, 1999) ; Harmonic-template Theory (Terhardt, 1979); Melodic Expectation Theory (Margulis, 2007) Tonal Hierarchy (Krumhansl, 1990); etc.	The sum of all models below (Also known as *CDC*5) (Tenney, 1986)
'Sensory'	Simultaneities with 1-2 semitone intervals; Slightly separated adjacent pure-tone partials	Unsettling, Roughness, Annoying, Unconformable, Etc. *CDC*1&*CDC*2 (Tenney, 1988).	The camp of 'beats'/ critical band theories (Helmholtz, 1983), (Greenwood, 1961) (Plomp, 1999). Etc.	Dissonance Curves (Kaestner, 1909; Guthrie and Morrill, 1928; Kameoka, A. and Kuriyagawa, 1969; Plomp and Levelt, 1965); Sethare's approximation algorithm (Sethare, 1999); **Secondary Beats Model (proposed);**
'Ambiguity'	Sub-dominant functions; 4th and 6th scale degrees;	Ambiguous, Confused, Alternative. Etc. *CDC*3 (Tenney, 1988).	The camp of Harmonic-template-based (virtual pitch, virtual tonality) theories; (Rameau, 1721), (Terhardt, 1979), (Parncutt, 1989), (Hofmann-Engl, 2006). Etc.	Parncutt's 'Tonalness' concept (Parncutt, 1989); Hofmann-Engl model (Hofmann-Engl, 2006); Relative periodicity (Stolzenburg, 2012), etc. **Ambiguity Dissonance Model (proposed)**
'Gloom'	Minor triads and scale degrees; Diminished functions, chords, and scale degrees; Descending pitch contours	Sad, Serious, Doleful, Etc. *CDC*4 (Tenney, 1988). **Tested in this research;**	Melodic Expectation Theory (Margulis, 2007); Gestalt Psychology (Meyer, 1956); 1st principal dimension of musical emotion; **Descending pitch contours (proposed)**	**Gloom Dissonance Model (proposed)**
'Tension'	Augmented triads and tonal functions; Ascending pitch contours;	Stressful, Stormy, Furious, Etc. *CDC*4 (Tenney, 1988). **Tested in this research;**	Melodic Expectation Theory (Margulis, 2007); Gestalt Psychology (Meyer, 1956); 2nd principal dimension of musical emotion; **Ascending pitch contours (proposed)**	**Tension Dissonance Model (proposed)**

Table 29 Overview of the consonance and dissonance concepts.

Table 29 presents how Guthrie and Morrill musical theories, psychological experiences and psychophysical interpretations are categorized into five harmonic concepts. The bold type shows the research works proposed or achieved in this PhD study

On top of a comprehensive review of consonance and dissonance concepts, this research also proposed a numerical model to predict the perception of musical consonance and dissonance using the psychoacoustic approach. A psychoacoustic approach is one of the major methods employed to explore the perception mechanisms of music harmony whilst understanding of the biological functions of auditory pathway organs remains incomplete. Psychoacoustic studies propose new theoretical thoughts and perception theories to account for an auditory perception phenomenon, and Psychoacoustic models are built to validate these theories. In this chapter, the theoretical foundation for the entire thesis is summarized in section 9.1; and the main features of the research model are illustrated in section 9.2. Based on the empirical study and model simulation results, the key findings of this research are revisited in section 9.3. The thesis conclusion and future recommendations are presented in section 9.4 and 9.5.

9.1 Summary of Theoretical Approach

In this research, it is asserted that the semantic meaning of music consonance and dissonance concepts are underpinned by more than one type of perception mechanism, resulting in at least four types of dissonance concepts: *sensory, ambiguity, gloom* and *tension*.

The *sensory dissonance* concept corresponds to a 'rough' and 'unsettling' auditory sensation between pure-tone interactions that is described under the *theory of beats*. This phenomenon has been accounted for by the inner-ear functions – a camp of theories knowing as the *tonotopical theory*.

Ambiguity, gloom, and tension dissonances proposed in this research are in contrast to the *tonal consonance* concept, which arises from the prominent school of thought promoting the *harmonic-template theory.* Under this theory, the *tonal consonance concept* essentially refers to a set of harmonically-related tone components or a music structure that suggest a clear and unique 'holistic pitch' perception – known as the *tonal centre concept* in this research; the concept of *ambiguity, gloom,* and *tension* dissonances disorder this harmonic structure in different ways:

Ambiguity dissonance weakens the tonal centre perception by introducing highly competing secondary tonal centres.

Gloom dissonance corresponds to ‘sad’ emotions and psychological negative valance, and it is evoked by the pitch components that are lower (pitch-wise) than the tonal expectations set up by the *tonal centre concept.*

Tension dissonance literally means ‘tension’ and ‘anxiety’ emotions (psychological excitement), and is evoked by the pitch components that are higher (pitch-wise) than the tonal expectations set up by the *tonal centre concept.*

A*mbiguity, gloom*, and *tension* dissonances require the *tonal centre concept* to be established or identifiable, therefore they are grouped within the *tonal dissonance concepts*. As the *sensory dissonance* concept does not have such a requirement, a notable limitation is observed when attempting to use solely the *theory of beats* to account for music harmony perceptions. The discussion of music harmony usually revolves around the tonal centre concept (e.g. root of a musical chord, tonic of a musical mode) implying a better theoretical potential for the *harmonic-template theory*. However, *harmonic-template*-based approaches, measuring music harmony perception merely on the degree of *tonal consonance* fail to distinguish specific tonal dissonant structures (e.g. diminished, augmented, and suspended chord structures). As a remedy and theoretical breakthrough, this research further refines the tonal dissonance concept into three uncorrelated dissonance concepts: *Ambiguity, gloom, and tension* to account for the empirical rankings of music harmony where previous theories generally fail.

9.2 Summary of Modelling Method

This research has presented a straightforward modelling method for the four types of dissonance concepts as well as how to integrate four dissonance concepts into one model to predict overall harmony perception. The main features of this research model are summarised in sections 9.2.1–9.2.5.

9.2.1 Pitch-based Analytical Approach

The purpose of this research was to conceive a functional concept with respect to each pitch component for the harmonic analysis of a multi-tone sonority. According to the hypothesis, each pitch component may have at least one of five harmonic functions in the overall harmony perception of the chord. A tone component may have one or more harmonic functions but only one of them best represents its harmonic property.

A pitch-based analytical approach means the harmonic functions (such as the proposed concepts of *ambiguity*, *gloom*, and *tension*) are not analysed in reference to the entire input sonority; but with respect to each of its noticeable pitch components. As for the result, the model is able to predict a 'distribution' of harmonic functions with respect to each pitch component.

Moreover, the pitch-based analytical approach means the entire analytical approach is built on perception-based, rather than acoustic-based, theoretical analysis. The acoustic-based analytical approaches are more apt to study the biological-based functional responses (a bottom-up approach); whereas in Psychoacoustic modelling, direct psycho-physical relationships are established where the music harmony perception is mainly accounted for by the psychological based theories. For this reason, the computational analysis for this model is not based on acoustic partial interactions.

To convert the acoustic model input into a series of pitch components, three sequential pre- processing procedures were involved: the hearing threshold, mutual masking effect, and a *harmonic masking effect*. The *harmonic masking effect* is a new approach proposed in this research that mainly combines a set of harmonically related pure-tone sensations into one single pitch perception – the *fundamental*.

9.2.2 Modeling for Sensory Dissonance with Secondary Beats Effect

The concept of *sensory dissonance* has been extensively studied and modelled in the literature. The analytical method is essentially a two-step approach: the first step is to observe and measure the *beats* effect between two pure-tone partials, from which a pure-tone dissonance curve is plotted and modelled; the second step is to derive a summation method that integrates the beats effect over each possible combination of pure-tone pairs within a complex-tone sonority.

One of the main problems addressed in this approach is that the pure-tone dissonance curves obtained in previous models only consider the effect of *primary beats* and the *secondary beats effect* has been generally neglected. However, in the modelling approach presented in this thesis it is asserted that even the *secondary beats effect* is relatively weak compared to the *primary beats effect*, they should not be neglected, especially in a summation based analytical approach. To derive a more precise model for pure-tone dissonance curves, an empirical study was conducted in which such pure-tone dissonance curves were used with a convention summation algorithm to predict the perception of complex-tone scenarios.

9.2.3 Modeling for Tonal Consonance and Ambiguity Dissonance

The concept of tonal centre is crucial to the analysis of tonal music harmony, but this concept has not been completely modelled in previous *harmonic-template*-based approaches. In this research analysis, tonal centre feature extraction from a multi-complex-tone structure is viewed as similar to the pitch determination mechanism. A model based on Terhardt's virtual determination algorithm is used to estimate the root of the chord.

In reference to the *tonal centre concept*, two series of pitch components are modelled as having tonal consonance functions: the pitch components that are in the same pitch class as the tonal centre; and pitch components whose fundamentals lies on the overtone series of the tonal centre. Conversely, the pitch components with ambiguity dissonance properties are those that do not belong to the same pitch class as the tonal centre and their fundamental is a sub-harmonic frequency of the tonal centre. The pitch components on the sub-harmonic frequencies of the tonal centre tend to draw a large proportion of partial components to suggest themselves as the tonal centre; therefore, these pitch components tend to divert the tonal centre perception to a completely different pitch class. The aggregate sound containing such pitch components introduces a higher degree of confusion for listeners to identify the tonal centre, manifesting as the ambiguity dissonance concept.

9.2.4 Modeling for Gloom and Tension Dissonance

According to the theoretical proposal, the gloom and tension dissonances are identified by *falling and rising pitch properties*. In this research, the degree of *falling and rising pitch properties* are modelled based on the pitch distance from the object pitch component to its nearest tonal consonant pitch component: the closer it is to a tonal consonant pitch component, the smaller the degree of *falling / rising pitch properties*. The tonal consonant pitch component in this model refers to a series of virtual pitch components that are harmonically close related the tonal centre.

The rationale of using *falling / rising pitch properties* to analyse gloom and tension dissonance is that the perception of a multi-tone structure should not be analysed in isolation, but should be in a context related to the tonal structure. This is because very often the consonance and dissonance judgement of a particular music structure is based on comparisons with other chord structures; during such comparison processes listeners may notice rising and falling pitch contours, which essentially introduces the gloom and tension emotions. Another merit of this approach is that it presents a referential analytical method whereby the *tonal centre concept* is not bounded to the most dominant pitch component within a particular musical entity; rather, it can be any pre-defined pitch class for tonal music harmony analysis.

9.2.5 Integration of Dissonance Effects

Four types of dissonance concepts imply a multi-dimensional analysis for the prediction of overall consonance and dissonance for a music structure. In order to combine these concepts, the strength of each harmonic function (*tonal consonance, ambiguity, gloom, and tension* dissonances) is obtained by summing the corresponding effects across all pitch components; next, the overall dissonance is estimated by the summation of the strength of each dissonance function in contrast to the tonal consonance effect.

The weight contributing to the overall dissonance is presumably different for different dissonance concepts. Therefore, each type of dissonance function has to be multiplied by a dissonance salience coefficient before addition with other types of dissonances. A multivariable interpolation analysis is used to obtain these coefficients with the empirical data on the perception of musical triads.

9.3 Key Findings

The key findings of this PhD research are summarized on following sub-topics.

1. The meaning of consonance and dissonance concepts (*CDC*).

Based on the music literature, this thesis identifies two types of *CDC*. One is associated with how well two tone components 'fit in' with each other under the principle of *proportion theories*; the other describes how well the tone components suggest a unique tonal centre, under the *harmonic relation theory*. This thesis further relates *proportion theories* to the physiological *theory of beats* and *harmonic relation theory* to the *theory of harmonic template*. These two camps of theories correspond to the *sensory* and *ambiguity* dissonance concepts proposed in this research.

On top of these two types of *CDC*, the *gloom* and *tension* dissonances were proposed in music harmony analysis, corresponding to the *valence* and *arousals* dimensions of psychological affects. A total of four types of dissonance concepts are identified in this research.

2. Pure-tone dissonance curve with secondary beats effects.

The listener test for the secondary beats effect shows that the sensory dissonance effect generally decays when frequency ratios get more complex. The statistical measurements of primary beats (near frequency ratio 1:1), secondary beats above octave (near frequency ratio 2:1) and secondary beats above tritave (near frequency ratio of 3:1) have a ratio of: 9.83 : 6.78 : 3.65. In the numerical model, this ratio is approximately simplified by the ratio of 3: 2: 1. Using these dissonance curves has sharpened the prediction of consonance intervals of musical dyads; but still fails to predict the perception rankings for musical triads.

3. Experimental study for gloom and tension dissonances

In the experimental study presented in chapter 7, the majority of listeners (85.7%) tend to associate the rising pitch properties with tension emotions and falling pitch properties with the gloom emotion. The empirical data rejects the null hypothesis that the perceptions of sonorities with rising and falling pitch properties are not significantly different from each other. Moreover, in a frequency-dependent comparative test, it is found that the weight of rising pitch properties contributing to the overall dissonance perception is significantly larger than that of falling pitch properties ($p > 0.01$), and this result is independent from fundamental frequencies of the tonal centre concept (at least at 100, 240, 1100Hz). The ratio of gloom dissonance contributing to the overall dissonance over that of tension dissonance is approximately 2:5 (0.4048 in linear regression analysis).

4. Comparisons between sensory, ambiguity, gloom and tension dissonance effects.

In a numerical approach, the perception of overall dissonance combining sensory, ambiguity, gloom and tension dissonance effects are modelled by a multivariable summation function:

$$D = \alpha_s (D_{s)\,\beta} + \alpha_a (D_{a)\,\beta} + \alpha_t (D_{t)\,\beta} + \alpha_g (D_{g)\,\beta}$$

Where α_s, α_a, α_t and α_g are the correlation coefficients of sensory dissonance (D_s), ambiguity (D_a), tension (D_t), and gloom (D_g) dissonances; correlation coefficients measure the degree to which a particular type of dissonance function contributes to the overall dissonance perception. β is psychological significance factor, $\beta=0.75$.

The correlation coefficients were determined as: $\alpha_s = 11.29$, $\alpha_a = 2.15$, $\alpha_t = 6.73$, and $\alpha_g = 1.81$ respectively. These values correspond to the importance with which a particular type of dissonance influences the overall dissonance perception as: sensory > tension > ambiguity > gloom.

9.4 Validation of Hypothesis

In order to provide a more comprehensive interpretation of the consonance and dissonance perception of musical chord structures, this research thesis hypothesized that (cited from Chapter 1):

> *Current analytical models for musical consonance and dissonance can be improved by implementing a pitch-based multi-dimensional harmonic analysis.*

For the study of tertiary musical triads (where the sensory dissonance effect is less dominant), the augmented triads are predicted with highest tension dissonance; the diminished and minor triads contain a higher level of gloom dissonance (diminished > minor); and the major structures contain the lowest level of all types of dissonances. Based on the conclusion that the tension dissonance is more salient than gloom dissonance, the augmented triads are predicted as having the highest level of overall dissonance. The overall estimation (see Table 30) gives that the perception rankings of consonance is ordered such that: major > minor > diminished > augmented, which is a result completely in line with empirical observations (Roberts, 1986).

In addition, the model also predicts that the perception rankings for secondary triads (the mild and sharp dissonance triads) have the lowest ranking of consonance due to the impact of sensory dissonance, and the perception ranking for suspended 4^{th} triads is between minor triads and diminished triads, which is a result in line with previous empirical findings (Cook, 2006).

Therefore, the hypothesis is satisfied for musical triads and validated through empirical data. The model can also potentially be used to analyse consonance and dissonance based music perceptions for more complex musical entities, such as musical modes, chord progressions, and counterpoint.

Triad Classes	Inversions	Empirical Rankings	Beats based models				Harmonic-template based models			This Research	
			Helmholtz	**P&L**	**K&K**	**Sethare**	**Parncutt**	**H-Engl**	**Stolzenburg**	**Dominant Harmonic Function**	**Predicted Rankings**
Major	Root Position	**1**	3	3	1	3	1	2	2	Tonal consonance	**1**
	1st Inversion	**2**	6	8	8	5	6	3	3	Tonal consonance	**3**
	2ndInversion	**3**	1	1	3	1	3	1	1	Tonal consonance	**2**
Minor	Root Position	**4**	3	3	1	3	4	8	4	Gloom dissonance	**4**
	1st Inversion	**5**	1	1	3	1	6	9	5	Gloom dissonance	**5**
	2ndInversion	**6**	6	8	8	5	10	10	7	Gloom dissonance	**6**
Diminished	Root Position	**7**	10	10	3	9	9	5	9	Gloom dissonance	**7**
	1st Inversion	**8**	8	5	6	7	5	4	8	Gloom dissonance	**9**
	2ndInversion	**9**	8	5	6	7	8	6	6	Gloom dissonance	**8**
Augmented	Root Position	**10**	5	7	10	9	2	7	10	Tension Dissonance	**10**

Table 30 Model predictions for triadic structures

The empirical ranking is obtained from Roberts (1986), 1-10 with 1 being the most consonant; the model predictions correspond to (Plomp & Levelt, 1965; Kameoka & Kuriyagawa, 1969; Sethare, 1999; Parncutt, 1989; Hofmann-Engl, 2012; Stolzenburg, 2009) and this research model (last two columns). Data for P&L, K&K, Sethare and Parncutt cited from (Cook, 2006); data for H-Engl is calculated from *Harmony Analyzer 3.2*. (Hofmann-Engl, 2012)

9.5 Future Work

This research promotes a theoretical model that uses a pitch-based multi-dimensional method to analyse the perception of music harmony. This model can be further developed in the following ways:

1. A more precise modelling for the estimation of overall sensory beats.

As mentioned in section 5.4, the summation algorithm of the pure-tone dissonance effect for a complex tone scenario at present cannot be precisely modelled. Phenomena such as beats, masking beats and binaural-beats cannot apparently be explained by the inner ear physiology. The main limitation at the present stage is the unknown functional responses of higher-level auditory structures. In this way, the entire Psychoacoustic method can be viewed as an alternative technique to study the auditory perception phenomenon. More insight should be gained in order to build a physiological-based model, especially for the perception features such as overall loudness and sensory beats. This generally requires a technological break-through in neural activity monitoring.

2. An experimental design to test the ambiguity dissonance effect.

The concept of ambiguity dissonance proposed in this research is solely theoretical. It is developed on the one hand from Terhadt's virtual pitch concept; and on the other hand, on the empirical observation that the fourth scale degree is treated as dissonance under a tonal structure. The effect of ambiguity dissonance needs to be further measured from experimental tests. The challenge of designing this test is to isolate ambiguity dissonance from gloom and tension dissonance, as a tonal dissonance pitch component usually has more than one type of tonal dissonance function at the same time. One of the methods is probably to study the isolated sonority without the context of any tonal centres. When the tonal centre is missing from context, listeners have to determine the tonal centre from within the sonority. Therefore, the perception of gloom and tension becomes significantly weakened whereas the perceptions of ambiguity dissonance are not influenced.

3. A quantitative experimental study for gloom and tension dissonances.

The gloom and tension dissonance concepts are modelled on a qualitative level in this research model. The concepts of gloom and tension dissonances are distinguished by the falling and rising pitch properties; but the level of gloom and tension effects are roughly modelled as being proportional to the pitch distance from its nearest tonal consonance pitch component. Further experimental studies are needed in order to obtain a quantitative measure of how gloom and tension effects are changed as a function of pitch distance from its nearest consonant pitch component.

This research study was limited to the study of the perception of simultaneous musical chords only. However, the theoretical approach of the model formulation is not limited to simultaneous musical chords and can be applied to any musical structures, or even non-simultaneous musical entities. In particular, the theoretical study of the following topics would be of interest:

a. Apply this model to the analysis of larger sound aggregates, such as chord progressions and tonal systems.

One of the main potentials of this research model is to study consonance and dissonance perceptions in chord progressions. During chord progressions, proposed concepts of tonal centre, gloom and tension dissonance are expected to play more important roles than sensory and ambiguity dissonances. Moreover, the proposed model is expected to be able to analyse the harmony perceptions for musical modes, including the perception of each scale degree notes/chords in reference to the tonal centre concept – the tonic.

b. Apply this model to analyse the ‘timbre aspect’ of musical entities.

The ‘timbre aspect’ of a musical structure refers to the perceived ‘quality’. For a single tone component, this is simply its timbre; for a chord structure, this concept means the ‘colour’ of a chord. In modern music composing, the perception of music harmony is no longer limited to the *CDC*, musicians are more apt to edit the ‘colour’ of music harmony to personalize their music styles and enrich the music experiences. The multi-dimensional nature of *CDC* proposed in this research can be used as a theoretical clue to define the colour of musical entities (e.g. the colour of chord). Section 8.5 provides a general overview on how to make use of this analytical model to study the colour of chords. One premise of doing this research is a review study that summarized relationships between listener's psychological experiences (probably in term of identifiable descriptors) and their corresponding acoustic features; and this relationship should be immune from culture and personal factors.

References

Arthurs, Y. and Timmers, R. (2016). On the fluidity of consonance and dissonance: The influence of musical context. *Psychomusicology: Music, Mind, and Brain*, 26(1), pp.1-14.

Barry, M. (2008). *An explanation for the emergence of Jazz*, viewed 20 Jan 2016, < https://theoryofmusic.wordpress.com/2008/01/16/an-explanation-for-the-emergence-of-jazz-1956/ >.

Bartlett, S.F. (1934). *The problem of noise,* Cambridge University Press.

Bauer, B. and Torick, E. (1966). Researches in loudness measurement. *IEEE Transactions on Audio and Electroacoustics*, 14(3), pp.141-151.

Beck, J., Warren, R.M. and Warren, R.P. (1970). Helmholtz on Perception: Its Physiology and Development. *The American Journal of Psychology*, 83(3), p.448.

Benade, H. (1976). *Fundamentals of musical acoustics*, New York. Dover Publications Inc.

Benward, B. and Saker,M. (2003). *Music: In Theory and Practice,* Seventh Edition, New York, NY, USA: McGraw-Hill, Vol. I, p.77.

Bharucha, J. (1983). The representation of harmonic structure in music: Hierarchies of stability as a function of context. *Cognition*, 13(1), pp.63-102.

Bigand, E., Parncutt, R. and Lerdahl, F. (1996). Perception of musical tension in short chord sequences: The influence of harmonic function, sensory dissonance, horizontal motion, and musical training. *Perception & Psychophysics*, 58(1), pp.125-141.

Björklund, O. (2015). The design principles of an online listening test for investigating the perception of heavy metal harmony. *Master's thesis University of Helsinki,* Department of Philosophy, History, Culture, and Art Studies.

Boer, E. (1956). *On the "residue" in hearing*. 1st ed. 's-Gravenhage Excelsior, Amsterdam.

Bregman, A. (1994). *Auditory Scene Analysis: The Perceptual Organization of Sound.* MIT Press. pp.96-98.

Carlos, W. (1987). Tuning: At the Crossroads. *Computer Music Journal*, 11(1), p.29.

Cazden, N. (1945). Musical Consonance and Dissonance: a Cultural Criterion, *The Journal of Aesthetics and Art Criticism*, 4(1), pp.3-11.

Chen, Y. (2002) *One musical structure, two perception systems*, Musicology in China, Shanghai Conservatory, pp.109–116.

Cherry.K. (2016). *The Arousal Theory of Motivation*. Viewed on 12 December. 2016, <https://www.verywell.com/the-arousal-theory-of-motivation-2795380>

Cheveigné, A.D. (2011). *Pitch perception models- historical review*, CNRS – Ircam, Paris, France.

Cook, N. (1999). Explaining harmony: the roles of interval dissonance and chordal tension, *Annals of the New York Academy of Science*, 930, pp.382-385.

Cook, N. and Fujisawa, T. (2006). The Psychophysics of Harmony Perception: Harmony is a Three-Tone Phenomenon, *Empirical Musicology Review*, 1(2), pp.1-21.

Crocker, R.L. (1986). *A history of musical style*, New ed., Dover Publications.

Cross, C. and Goodwin, H. (1893). Some Considerations regarding Helmholtz's Theory of Consonance, *Proceedings of the American Academy of Arts and Sciences*. 27(3), pp.1-12.

Delone, R. (1975). Melody: Linear Aspects of Twentieth-Century Music. Wittlich, Gary (ed.) *Aspects of Twentieth-Century Music*, pp.270-301. Englewood Cliffs, New Jersey: Prentice-Hall.

Demers, J. (2010). Listening through the noise: the aesthetics of experimental electronic music. *Choice Reviews Online*, 48(07), pp.48-3764.

Deutsch, D. and Feroe, J. (1981). The internal representation of pitch sequences in tonal music. *Psychological Review*, 88(6), pp.503-522.

Dowling, W. and Fujitani, D. (1971). Contour, Interval, and Pitch Recognition in Memory for Melodies. *The Journal of the Acoustical Society of America*, 49(2B), pp.524-531.

Dowling, W.J. and Harwood, D.L. (1986). *Music Cognition.* New York, Academic Press.

Duncan, S. and Barrett, L. (2007). Affect is a form of cognition: A neurobiological analysis. *Cognition & Emotion*, 21(6), pp.1184-1211.

Ebeling M. (2009) Calculating Tonal Fusion by the Generalized Coincidence Function. In: Klouche T., Noll T. (eds) *Mathematics and Computation in Music*. Communications in Computer and Information Science, vol 37. Springer, Berlin, Heidelberg

Egermann, H., Pearce, M., Wiggins, G. and McAdams, S. (2013). Probabilistic models of expectation violation predict psychophysiological emotional responses to live concert music. *Cognitive, Affective, & Behavioral Neuroscience*, 13(3), pp.533-553.

Euler, L. (1766). Conjecture on the Reason for some Dissonances Generally Heard in Music. *Memoires de l'academie des sciences de Berlin 20*, 1766, pp. 165-173

Fletcher, H. and Munson, W. (1933). Loudness, Its Definition, Measurement and Calculation. *The Journal of the Acoustical Society of America*, 5(2), pp.82-108.

Freeman, J (2015*) Emotion, Feeling, Mood, Six seconds EQ network*, viewed on 17th Jun. 2015 < http://www.6seconds.org/2015/01/02/emotion-feeling-mood>

Gable, P. and Harmon-Jones, E. (2013). Does arousal per se account for the influence of appetitive stimuli on attentional scope and the late positive potential? *Psychophysiology*, 50(4), pp.344-350.

Gabor, D. (1947). Acoustical Quanta and the Theory of Hearing. *Nature*, 159(4044), pp.591-594.

Gabrielsson, A. (1995). Expressive intention and performance, In R. Steinberg (E.d), *Music and the mind machine*, Heidelberg, Germany: Springer, pp.35-47.

Gabrielsson, A. and Justin, P. (2003). Emotional expression in music, Oxford University Press, *Canadian Journal of Experimental Psychology*, 51, pp.336-352.

Glasberg, B. and Moore, B. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, 47(1-2), pp.103-138.

Goldstein, J. (1973). An optimum processor theory for the central formation of the pitch of complex tones. *The Journal of the Acoustical Society of America*, 54(6), pp.1496-1516.

Goodwin, A. (1980). An Acoustical Study of Individual Voices in Choral Blend. *Journal of Research in Music Education*, 28(2), pp.119-128.

Greenwood, D. (1961). Auditory Masking and the Critical Band. *The Journal of the Acoustical Society of America*, 33(4), pp.484-502. Grey, J.M. (1975). *An exploration of musical timbre,* Ph.D. dissertation, Department of Psychology Stanford University, viewed on 02 Feb. 2016 <https://ccrma.stanford.edu/papers/exploration-of-musical-timbre>

Grey, J.M. (1975). *An exploration of musical timbre*, Ph.D. dissertation, Department of Psychology Stanford University, viewed on 02 Feb. 2016 <https://ccrma.stanford.edu/papers/exploration-of-musical-timbre>

Grey, J.M. (1977). Multidimensional perceptual scaling of musical timbres. *The Journal of the Acoustical Society of America,* 61(5), pp.1270-1277.

Guthrie, E. and Morrill, H. (1928). The Fusion of Non-Musical Intervals. *The American Journal of Psychology*, 40(4), p.624.

Hayes, B. (2008) On the Teeth of Wheels. Theory in the Bedroom, and Other Mathematical Diversions. 26 (1), pp. 13-15.

Hevner, K. (1936). Experimental Studies of the Elements of Expression in Music. *The American Journal of Psychology*, 48(2), pp.246-266.

Hindemith, P. (1984) *The Craft of Musical Composition: Theoretical Part - Book 1*, New York, Schott & Co., Ltd., London, pp. 202-220.

Hofmann-Engl, L. (2004). *Virtual Pitch and its Application to Contemporary Harmonic analysis,* Online Publication, viewed on 04 July 2015, <www.chameleongroup.org.uk/research/link_virtual_analysis.html>

Holtzman, S.R. (1981). Using Generative Grammars for Music Composition. *Computer Music Journal*, 5(1), pp.51-64.

Honing, H. (2013). Structure and interpretation of rhythm in music. In Deutsch, D. (ed.), *Psychology of Music, 3rd edition,* London: Academic Press / Elsevier, pp.369-404.

Howard, D. and Angus, J. (2009). *Acoustics and Psychoacoustics*, 4th Edition, Oxford: Focal Press.

Howie, J. (1976) Acoustical Studies of Mandarin Vowels and Tones. *Cambridge University Press*, Cambridge, UK.

Huron, D. (1991). Tonal Consonance versus Tonal Fusion in Polyphonic Sonorities. *Music Perception: An Interdisciplinary Journal*, 9(2), pp.135-154.

Huron, D. (1993). *Consonance and Dissonance*, Index to Notes assembled by David Huron for the course Music 829B: Consonance and Dissonance, Comments and information was also contributed by Nandini Iyer, Bret Aarden, and Evelyn Hogland. Viewed on 06 Jun. 2014, <http://www.music-cog.ohio-state.edu/Huron/publications.html>

Huron, D (1996). The melodic arch in Western folksongs. *Computing in Musicology*, Vol. 10 pp. 3-23.

Hutchinson, W. and Knopoff, L. (1978). The acoustic component of western consonance. *Interface*, 7(1), pp.1-29.

Jenson, J. (2008). *A java Applet to calculate Virtual Pitch*, viewed on 03 Sep. 2014<http://jjensen.org/VirtualPitch.html#algo>

Johnson-Laird, P., Kang, O. and Leong, Y. (2012). On Musical Dissonance. *Music Perception: An Interdisciplinary Journal*, 30(1), pp.19-35.

Juan, S. (2006). *Harmonic duality : from interval ratios and pitch distance to spectra and sensory dissonance*, Leiden University Academy of Creative and Performing Arts, Faculty of Humanities, Leiden University.

Juslin, P. N. and Sloboda, J. A. (2010). *Handbook of music and emotion: Theory, research, applications*. New York: Oxford University Press.

Kaestner, G. (1909). Untersuchungen ber den Gefhlseindruck unanalysierter Zweiklange. *Psychologische Studien*, vol.4 pp.473–504. Reviewed by Kussi, T., 2007 in *Journal of New Music Research*. 36(2), pp.95 -104.

Kameoka, A. and Kuriyagawa, M. (1969). Consonance Theory Part I: Consonance of Dyads. *The Journal of the Acoustical Society of America*, 45(6), pp.1451-1459.

Kaplan, R. (1996). Harmonic Function in Chromatic Music: A Renewed Dualist Theory and an Account of Its Precedents Daniel Harrison. *Music Theory Spectrum*, 18(1), pp.124-133.

Károlyi, O. (1995). *Introducing modern music*. 1st ed. London: Penguin Books.

Kendall, R.A., Lipscomb, S.D., Cohen, A. and Tan, S.L. (2013). *The psychology of music in multimedia*. 1st ed. Oxford: Oxford University Press.

Kollmeier, B., Brand, T. and Meyer, B. (2008). *Perception of Speech and Sound*, Springer Handbook of Speech Processing, pp.61-82.

Krumhansl, C. (1997). An exploratory study of musical emotions and psychophysiology. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 51(4), pp.336-353.

Krumhansl, C. and Castellano, M. (1983). Dynamic processes in music perception. *Memory & Cognition*, 11(4), pp.325-334.

Kruskal, J. (1964). Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, 29(2), pp.28-46.

Langer, S. (1942). *Philosophy in a new key,* 3rd Revised edition. Harvard University Press.

Langner, G. (2015). *The Neural Code of Pitch and Harmony*. 1st ed. Cambridge University Press.

Lawrence, M. and Yantis, P. (1956). Thresholds of Overload in Normal and Pathological Ears. *Archives of Otolaryngology - Head and Neck Surgery*, 63(1), pp.67-77.

Licklider, J.O. (1968). Mechanisms of frequency discrimination, *The Journal of the Acoustical Society of America*, 44, 1533.

Linderman, J. (2014). *Songwriting technique: Harmonic quality of chords*. Viewed on 14 Jan. 2016, <http://www.songwritingmagazine.co.uk/tips-techniques/songwriting-technique-harmonic-quality-of-chords/15921 >

Lipps, T. (1995). *Consonance and dissonance in music*. 1st ed. San Marino, CA: Everett Books.

Lucker, J. Grzybmacher, N, and Ventry I. (1978). Effects of Instructional Set on the Comfortable Loudness Range. *Journal of Speech and Hearing Disorders*, Vol. 43, pp.168-172.

Lutfi, R. (1983). Additivity of simultaneous masking. *The Journal of the Acoustical Society of America*, 73(1), pp.262-267.

Madsen, C. and Fredrickson, W. (1993). The Experience of Musical Tension: A Replication of Nielsen's Research Using the Continuous Response Digital Interface. *Journal of Music Therapy*, 30(1), p.46

Margulis, E. (2007). Surprise and Listening Ahead: Analytic Engagements with Musical Tendencies. *Music Theory Spectrum*, 29(2), pp.197-217.

Mashinter, K. (2006). Calculating Sensory Dissonance: Some Discrepancies Arising from the Models of Kameoka & Kuriyagawa, and Hutchinson & Knopoff, *Empirical Musicology Review*, Vol.1, 65-84.

Meddis, R. and Hewitt, M. (1991). Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification. *The Journal of the Acoustical Society of America*, 89(6), pp.2866-2882.

Meyer, L.B. (1956). *Emotion and meaning in music*, Chicago: University of Chicago Press.

Moore, B. (1977, 1982). *An introduction to the psychology of hearing*, London, Academic Press.

Moore, B. (1985). Additivity of Simultaneous Masking. *The Journal of the Acoustical Society of America*. 73(262), pp.1-19

Narmour, E. (1977). *Beyond Schenkerism: The Need for Alternatives in Music Analysis.* Chicago: University of Chicago Press. 1977.

Ohala, J. (1983). Cross-Language Use of Pitch: An Ethological View. *Phonetica*, 40(1), pp.1-18.

Parncutt, R. (1987). *Sensory Bases of Harmony in Western music*, PhD thesis at the University of New England, Armidale NSW, Australia.

Parncutt, R. (1989). *Harmony: A Psychoacoustical Approach*. 1st ed. Berlin, Heidelberg: Springer Berlin Heidelberg.

Parncutt, R. and Hair, G. (2011). Consonance and dissonance in music theory and psychology: Disentangling dissonant dichotomies, *Journal of Interdisciplinary Music Studies*, 5 (2), pp.119-166.

Patterson, R., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C. and Allerhand, M. (1992). Complex sounds and auditory images, *Auditory physiology and perception*, Edited by Y. Cazals, K. Horner and L. Demany, Oxford, Pergamon Press, pp.429-446.

Peterson, A.P. and Gross, E.E. (1974). *Noise Measurement*, 7th Ed. Genrad, Concord.

Phillip, D. (2014) Relationships among Timbre Preference, Personality, Gender, and Music Instrument Selection, *Journal of Band Research*, 50 (1), pp.1-22.

Plomp, R. and Levelt, W. (1965). Tonal Consonance and Critical Bandwidth. *The Journal of the Acoustical Society of America*, 38(4), pp.548-560.

Plomp, R. and Mimpen, A. (1968). The Ear as a Frequency Analyzer. II. *The Journal of the Acoustical Society of America*, 43(4), pp.764-767.

Plomp, R. and Smoorenburg, G. (1970). *Frequency analysis and periodicity detection in hearing*. 1st ed. Sijthoff, the University of California.

Popper, K. and Eccles, J. (1977). *Objective Knowledge: An evolutionary Approach*, Oxford University Press, Oxford.

Pratt, R. and Doak, P. (1976). A subjective rating scale for timbre. *Journal of Sound and Vibration*, 45(3), pp. 317-328.

Punch, J., Joseph, A. and Rakerd, B. (2004). Most Comfortable and Uncomfortable Loudness Levels. *American Journal of Audiology*, 13(2), pp.144-157.

Rameau, J.P. and Wundt, W. (1721). Treatise on Harmony, translated, with introduction and notes, in Philip Gossett (eds), New York: Dover Publications, Inc. 1971.

Randel, D. (1990). *The New Harvard Dictionary of Music*. 1990 edition. Harvard University Press, 1990.

Rasch, R. and Plomp, R. (1999). *The perception of musical tones*, 2nd ed., Academic Press, San Diego, CA, pp.89-112.

Resnick, L. (1981). Psychophysical basis for consonant musical intervals. *American Journal of Physics*, 49(6), pp.579-580.

Roberts, L.A. (1986). Consonant judgments of musical chords by musicians and untrained listeners, *Acustica*, vol.62, pp.163-171.

Rossing, T. (2010). *The science of string instruments*. 1st ed. New York: Springer, pp.310-321.

Salzer, F. (1952). *Structural hearing: Tonal Coherence in Music,* vol. 2, New York, Dover Publications Inc. pp.163-170.

Scherer, K. and Oshinsky, J. (1977). Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion*, 1(4), pp.331-346.

Schouten, J.F. (1940). *The residue, a new component in subjective sound analysis,* Proceedings, Koninklijke Nederlandsche Akademie van Wetenschappen te Amsterdam, 43, p.356-365.

Schuijer, M. (2008). Analyzing Atonal Music: Pitch-Class Set Theory and Its Contexts. *Eastman Studies in Music 60*. Rochester: University of Rochester Press. pp.01-28.

Sethares, W. (1999). *Tuning, timbre, spectrum, scale*. 2nd ed. London: Springer.

Shamma, S. and Klein, D. (2000). The case of the missing pitch templates: How harmonic templates emerge in the early auditory system. *The Journal of the Acoustical Society of America*, 107(5), pp.2631-2644.

Schoenberg, A. and Stein, L. (1985). *Style and idea*. 1st ed. University of California Press.

Schneider, A. (1898*). 'Verschmelzung', tonal fusion, and consonance*: Carl Stumpf revisited. Viewed on 21 Feb. 2017, <https://link.springer.com/chapter/10.1007%2FBFb0034111>

Shepard, R. (1964, 1982). *Chroma and tonality*, MPAATE-GE 2623 Music Information Retrieval, New York University Lecture Notes. Viewed on 12 Apr. 2015, <http://www.nyu.edu/classes/bello/MIR_files/tonality.pdf>

Stevens, S. (1957). On the psychophysical law. *Psychological Review*, 64(3), pp.153-181.

Stolzenburg, F. (2012). Harmony perception by periodicity detection. *Journal of Mathematics and Music: Mathematical and Computational Approaches to Music Theory, Analysis, Composition and Performance,* 9(3) pp.215-238.

Taylor, M.M., Brenneman, H.A. and Smith, S.M. (1974). Masking beats with beats. *The Journal of the Acoustical Society of America*, 56(S1), pp.S37-S37.

Temperley, D. (2004). *The cognition of basic musical structures*. 1st ed. Cambridge, MA: The MIT Press.

Tenney, J. (1988) A History of Consonance and Dissonance. *Computer Music Journal*, 13(3), p.94.

Terhardt, E. (1970). Frequency analysis and periodicity detection in the sensations of roughness and periodicity pitch, In: *Frequency Analysis and Periodicity Detection in Hearing* (Plomp, R., Smoorenburg, G.F., eds.), Sijthoff, Leiden, pp.278-287.

Terhardt, E. (1974). Pitch, consonance, and harmony. *The Journal of the Acoustical Society of America*, 55(5), pp.1061-1069.

Terhardt, E. (1979). Calculating virtual pitch. *Hearing Research*, 1(2), pp.155-182.

Terhardt, E., Stoll, G. and Seewann, M. (1982). Pitch of complex signals according to virtual-pitch theory: Tests, examples, and predictions. *The Journal of the Acoustical Society of America*, 71(3), pp.671-678.

Ternsttrom, S., Sundberg, J. and Collden, A. (1983). *Articulatatory perturbation of pitch in singer deprived of auditory feedback*, Speech transmission Lab, Qt.Prog. Status Rep. 2-3, pp.44-155.

Tomkins, S. (1963). *Affect, imagery, consciousness.* Vol.2. The negative affects. New York, Springer Publishing Company, 1963.

Vancouver, G. (2008). *What is the Harmonic Table: how the C-thru Axis works*. Viewed on 04 Jan. 2017, <http://www.musicscienceguy.com/2008/12/what-is-the-harmonic-table-how-the-c-thru-axis-works.html>

Vos, J. (1986). Purity Ratings of Tempered Fifths and Major Thirds. *Music Perception: An Interdisciplinary Journal*, 3(3), pp.221-257.

Warren, R. (2008). *Auditory perception*. 1st ed. Cambridge, UK: Cambridge University Press.

Wedin, L. (1972). *Evaluation of a three-dimensional model of emotional expression in music*. 1st ed. Stockholm, Sweden: Psychological Laboratories, University of Stockholm, p.349.

Wegel, R.L. and Lane, C.I. (1924). The Auditory Masking of One Pure Tone by Another and its Probable Relation to the Dynamics of the Inner Ear. *Physical Review*, 23(2), pp.266-285.

Winder, S. (2012). *Color Music Charts*. Viewed on 03 August. 2015, <http://samwinder.blogspot.co.uk/2012/11/color-music-charts.html>

Wightman, F. (1973). The pattern-transformation model of pitch. *The Journal of the Acoustical Society of America*, 54(2), pp.407-416.

Wright, J.K. (1986). *Auditory Object Perception: Counterpoint in a New Context*. McGill University, Master's thesis.

Wright, J.K. and Bregman, A.S. (1987). Auditory stream segregation and the control of dissonance in polyphonic music. *Contemporary Music Review*, 2(1), pp.63-92.

Zhang, A. (2009). *The Framework of Music Theory as Represented with Groups*. Viewed on 20 Jan., 2017, <http://www.jstor.org/stable/10.5406/bulcouresmusedu.193.0007?seq=1#page_scan_tab_contents>

Ziębakowski, T. (2012). Combination Tones in the Model of Central Auditory Processing for Pitch Perception. *Archives of Acoustics*, 37(4), pp.571-582.

Zwicker, E. (1970) The limits of audibility of amplitude modulation and frequency modulation of a pure tone. *Acustica*, Vol.2, pp. 125-133.

Zwicker, E., Flottorp, G. and Stevens, S. (1957). Critical Band Width in Loudness Summation. *The Journal of the Acoustical Society of America*, 29(5), pp.548-557.