TIMBRE PERCEPTION OF COCHLEAR IMPLANT USERS

by

Ilse Bernadette Labuschagne

Submitted in partial fulfilment of the requirements for the degree
Master of Engineering (Bioengineering)

in the

Department of Electrical, Electronic and Computer Engineering
Faculty of Engineering, Built Environment and Information Technology

UNIVERSITY OF PRETORIA

September 2011

SUMMARY

TIMBRE PERCEPTION OF COCHLEAR IMPLANT USERS

by

Ilse Bernadette Labuschagne

Supervisor: Prof J.J. Hanekom

Department: Electrical, Electronic and Computer Engineering

University: University of Pretoria

Degree: Master of Engineering (Bioengineering)

Keywords: Cochlear implant, Timbre, Brightness, Irregularity, Log rise-time, sustain/decay, Just-noticeable difference, synthesis.

The timbre perception of cochlear implantees (CI) is poor compared to normal hearing (NH) listeners. The cues that are normally transmitted to NH listeners may be less salient or even absent for CI users. From the literature, two spectral (brightness (T_b) and irregularity (IRR)) and two temporal timbre parameters (log rise-time (LRT) and sustain/decay (S/D) parameter (n)) have been identified as important timbre parameters. Each of these parameters was extracted for a set of thirteen instruments. Sounds could be resynthesized according to the specific timbre parameter set. The variation of loudness, pitch and perceived duration as functions of the timbre parameters were investigated to provide systematic balancing methods.

The just-noticeable differences (JNDs) were obtained for each of the parameters for thirteen instruments for NH listeners and a reduced instrument set of nine instruments for the CI users using a 1-up, 2-down, two-alternative, forced choice procedure. From the JNDs, predicted confusion matrices were constructed. From the confusion matrices, a feature information transmission analysis (FITA) indicated the salience of each of the parameters and NH and CI results could be compared.

OPSOMMING

TIMBREPERSEPSIE VAN KOGLEÊRE INPLANTINGGEBRUIKERS

deur

Ilse Bernadette Labuschagne

Studieleier:	Prof J. J. Hanekom
Departement:	Elektriese, Elektroniese en Rekenaaringenieurswese
Universiteit:	Universiteit van Pretoria
Graad:	Magister in Ingenieurswese (Bio-Ingenieurswese)
Sleutelwoorde:	Kogleêre implanting, Timbre, Helderheid, Ongelykheid, Logaritmiese stygingstyd, Volhouding/Versterwing, Net-waarneembare verskil, Sintese.

Die timbrepersepsie (toonkleurpersepsie) van kogleêre inplantinggebruikers (KI) is swak in vergelyking met normaalhorende (NH) luisteraars. Die inligting wat normaalweg na NH-luisteraars oorgedra word, is moontlik minder kenmerkend of selfs afwesig vir KI-gebruikers. Vanuit die literatuur is twee spektrale (helderheid (T_b) en ongelykheid(IRR)) parameters en twee temporale parameters (logaritmiese stygingstyd (LRT) en volhouding/versterwing (S/D) parameter(n)) geïdentifiseer as belangrike elemente van timbre. Elkeen van die parameters is uit dertien instrument opnames onttrek, wat dan as basis dien vir die hersintese van klanke. Die variasie van luidheid, toonhoogte en tydsduur as ‘n funksie van elkeen van die parameters is ondersoek om sistematiese balanseringsmetodes te ontwikkel.

Die net-waarneembare verskille van elkeen van die parameters vir dertien instrumente vir die NH-luisteraars en ‘n verkorte stel van nege instrumente vir die KI-gebruikers is bekom deur ‘n 1-op, 2-af, twee-alternatiewe, gedwonge keuse prosedure. Die net-waarneembare verskille is gebruik om verwarringsmatrikse saam te stel. ’n Analise van die eienskapinligting wat oorgedra word dui di belangrikheid van elkeen van die parameters aan en NH-data en KI-gebruiker data is vergelyk.

LIST OF ABBREVIATIONS

2AFC	Two-alternative, forced-choice
AM	Amplitude modulation
ANOVA	Analysis of variance
CI	Cochlear implant
eoa	End-of-attack
eor	End-of-release
FITA	Feature information transmission analysis
FM	Frequency modulation
IRR	Irregularity
JND	Just-noticeable difference
LRT	Logarithmic rise-time
NH	Normal hearing
RT	Rise-time
S/D	Sustain/decay
soa	Start-of-attack
sor	Start-of-release
T_b	Brightness

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION

1.1 BACKGROUND AND MOTIVATION

The purpose of this study was to evaluate the timbre perception of cochlear implant (CI) users. This was done by evaluating specific constituents of timbre using appropriate sound synthesis techniques.

Comparing present day CIs with the first experimental device implanted by Djourno and Eyriès in Paris is an indication of the progress made in the field of electrical stimulation of the auditory nerve. Although speech perception was almost non-existent with the early device, short speech samples from a small closed set could be identified. In addition, large frequency changes below 1 kHz and the presence of environmental sounds could be identified (Wilson and Dorman, 2008).

Successes in speech perception for CI users have led to demands to improve the quality of hearing through electrical stimulation. Verbal descriptions of electrically stimulated hearing include: "mechanical", "noise-like" and "unnatural" (Gfeller et al., 2003; Gfeller et al., 2005). Difficulties in music recognition and music enjoyment of CI users are a recurring point in the literature.

Although there are many descriptions of what music is, all of them are vague. There is, however, general agreement of the technical constituents of music. Studies involving music perception focus mainly on i) pitch perception (including melodic contours and harmony), ii) rhythm perception and iii) timbre perception (Limb, 2006a; Jackendoff and Lerdahl, 2006; Kong et al., 2004; Drennan and Rubinstein, 2008; McDermott, 2004). While rhythm and pitch are relatively simply defined, the description of timbre, as with the description of music, is again vague. The entry for timbre in New Grove Dictionary of Music and Musicians (2001) includes: "A term describing the tonal quality of sound" and "timbre is a more complex attribute than pitch or loudness, which can be represented by a one-dimensional scale; the perception of timbre is a synthesis of several factors". Of the three abovementioned elements of music, timbre and pitch discrimination appear to be the limiting factors in music perception tasks (Limb, 2006a; Kong et al., 2004).

The multidimensional nature of timbre makes a single definition of what timbre *is*, nearly impossible. However, attempts to identify the most important timbre cues have shown that brightness (T_b), irregularity (IRR) and logarithmic rise-time (LRT) are probably the perceptually most salient cues (Krimphoff et al., 1994; Jensen, 2001; Grey, 1977; McAdams et al., 1995; Caclin et al., 2005). Despite the complex nature of timbre, calculations of these dimensions involve simple equations.

The investigation of the perception of sound stimuli requires the careful consideration of perceptual processing. Results from multidimensional scaling techniques suggest that perceptual processing may occur along the timbre dimensions, and as such, obtaining just-noticeable differences (JNDs) of timbre constituents may be useful in identifying specific perceptual problems within timbre.

The stimuli used for timbre perception are usually either recordings that can be altered, or synthetic sounds recreated from timbre parameters. Although perceptual tests using altered recordings are valuable in the evaluation of perception in real-world situations, with such stimuli it is difficult to vary dimensions independently from one another. In contrast, the timbre parameter control of synthetic sounds is possible, but lacks the subtle complexities of timbre. In addition to this, many synthetic sounds, meant to represent a specific instrument, often differ significantly when compared to their real-world counterparts. An attempt to achieve the advantages of both types of stimuli requires one to investigate timbre analysis and synthesis methods. A mathematical synthesis method used for discrimination experiments allows the investigation of specific timbre constituents by allowing single timbre properties to be varied.

In conclusion, the accuracy of timbre parameters identified and used in timbre perception as a representation of instrumental sounds must be evaluated. An investigation of synthesis methods must produce a solution for systematic timbre parameter discrimination and JNDs of these parameters must found. Statistical analysis of the data can provide conclusions on the timbre perception of CI listeners compared to normal hearing (NH) listeners, differences or similarities within groups and the perception of specific timbre properties.

1.2 RESEARCH QUESTIONS

Examination of the literature produced important research questions. Timbre perception studies include investigation of spectral, temporal and spectro-temporal parameters (Krimphoff et al., 1994; Jensen, 2001; Grey, 1977; McAdams et al., 1995; Caclin et al., 2005). The first important question is which timbre parameters should be included in perceptual testing experiments. Such a timbre set must attempt to represent real-world instruments as accurately as possible. Secondly, the identification of a simple synthesis technique that attempts to accurately represent real-world sounds will contribute to finding meaningful results. Finally, it is necessary to evaluate how JNDs compare between CI listeners and NH participants.

1.3 APPROACH

Timbre parameters found from multidimensional scaling techniques, timbre parameters used during perception studies and sound analysis and synthesis literature was reviewed systematically in order to determine which parameters allow a good representation of the multidimensional quality of timbre. The prevalence of perceptual properties obtained from multidimensional studies and its use in synthesis methods provided good starting points for logical evaluations of these properties (section 3.1 and 3.2).

A study of synthesis methods indicated that trade-offs between instrument sound fidelity and timbre parameter control exists. An appropriate mathematical synthesis technique was chosen, which allowed the recreation of synthetic sounds that are completely defined by their parameters and that represent its real-world instrument counterparts as accurately as possible. Section 3.3 describes the mathematical development of the chosen synthesis method.

In preparation for experimental procedures to obtain JNDs, the stimuli generated using the synthesis method was balanced for loudness, pitch and perceived duration in order to eliminate these cues. Chapter 4 outlines the rationale and experimental procedure for the balancing procedures. The experimental procedure required NH listeners to balance the loudness, pitch and perceived duration of tones varying in timbre properties in three respective experiments. The data were analysed to produce a balancing equation generally applicable to the mathematical synthesis method.

JNDs for each timbre property were found using an adaptive two-alternative, forced-choice (2AFC) procedure. JNDs for each property were obtained for a variety of other timbre property values that represented actual instruments. The full experimental procedure can be found in section 5.3. Obtaining JNDs for each individual timbre property allowed an analysis of the performance of CI listeners compared to NH listeners. The investigation of individual timbre properties allowed identification of a CI user's perceptual difficulties.

1.4 OBJECTIVES

Salient timbre properties had to be identified. Timbre dimensions found in multidimensional scaling studies and timbre properties incorporated in existing synthesis models provided a useful starting point. These properties could be examined to produce a small, but representative set of timbre properties.

Recordings of instrumental sounds had to be obtained. The University of Iowa Electronic Music Studios (Fritts, 1997) instrument recording database provided a functional set of stimuli. The recordings had to be analysed with regard to the representative set of timbre properties of four spectro-temporal timbre parameters: T_b, IRR, LRT and sustain/decay (S/D).

A synthesis model had to be developed. The model had to use the analysed values of each instrument's timbre property set to recreate synthetic sounds based on the specific instrument. The necessity of complete timbre property control and rapid adjustments required for adaptive procedures during JND testing were two major considerations during the synthesis development process.

In preparation for discrimination experiments, synthetic sounds had to be balanced for loudness, pitch and perceived duration. An initial experiment with NH listeners had to determine the JNDs for each of the timbre parameters using a 2AFC with a 1-up, 2-down staircase procedure. The JNDs obtained for NH listeners were used to reduce the instrument set. This was beneficial for practical reasons concerning experimental time. With such a reduction, an acceptable distribution of instruments within the timbre space had to be maintained. The reduced set could be used to obtain the JNDs of CI listeners.

A confusion matrix created from the JNDs obtained for the two participant groups would serve as a timbre perception model. A feature information transmission analysis (FITA) allows the confusion matrices for each of the groups to be collapsed, allowing individual comparisons of each of the timbre dimension as well as indicating the amount of information transferred by each of the timbre parameters.

1.5 CONTRIBUTION

Using instrument recordings for timbre perception studies make it difficult to control cues that may be present in the recordings; however, the control offered by synthetic sounds may lack realism as an instrumental sound. The synthesis model used in this work attempts to represent instruments as realistically as possible, while being completely deterministic. Additionally, the synthesis model sets a mathematical foundation for further model improvements.

Perceptual loudness, pitch and perceived duration data provide simple and systematic methods for loudness, pitch and perceived duration balancing. Similar perceptual trends across NH participants also provide the opportunity to construct balance equations applicable to most listeners.

JNDs for T_b, IRR, LRT and S/D for NH and CI listeners provide information on the differences between groups. Analysis of the JNDs also provides important timbre perception data and an idea of the perceptual salience of each. Such information contributes to a better understanding of the timbre perception of CI users. Differences between NH and CI listeners allow areas to be identified where improvements may contribute to improved CI timbre perception.

The mathematical synthesis model and balance equations can be used to prepare stimuli for timbre perception studies. The JNDs for individual timbre parameters identifies problem areas for timbre perception.

In summary, the contributions of this study were i) the development of a mathematical synthesis model that allows complete control over all identified timbre parameters; ii) the investigation of the feasibility of pre-experimental loudness, pitch and perceived duration balancing; and iii) finding JNDs for the identified timbre parameters in order to establish possible problem areas for CI users.

1.6 OVERVIEW OF THE STUDY

The literature study in Chapter 2 provides background on the perception of CI users with regard to speech and music, specifically, pitch, rhythm and timbre. Synthesis methods from the literature are also discussed.

Chapter 3 provides a thorough investigation of the timbre parameters as described in the literature. It also provides mathematical analysis calculations and discussion on the salience of some of the parameters. This investigation served as the basis for choosing the timbre parameter set for the study. The timbre analysis procedure and calculations are described. The procedure description includes the isolation of harmonic partials and reasons for specific equations used during analysis. Timbre values for each of the instruments are obtained. The section on synthesis describes a mathematical model that recreates the spectral harmonics, taking brightness and irregularity into account. The temporal signal of the spectrum is then multiplied by a temporal envelope.

Chapter 4 outlines systematic loudness, perceived duration and pitch balancing procedures. Listeners were presented with a reference and test tone differing in one timbre parameters and were asked to adjust a slider bar to match either the loudness, pitch or perceived duration of the tones. Statistical analysis of the perceptual response data obtained from these experiments was used to produce a set of balance equations.

Chapter 5 outlines the layout of the discrimination experiments. Participants partook in an adaptive 2AFC procedure that yielded JNDs for each of the timbre dimensions of each of the instruments.

Chapter 6 provides an analysis of the results and constructions of confusion matrices using a model suggested by Svirsky (2000). Information transmission analysis estimates the amount of information transmitted by each of the parameters to each of the NH and CI listeners. Chapter 7 discusses the results of Chapter 6 with reference to the literature. The chosen timbre parameters are discussed along with other parameters from the literature. The meaning of the confusion matrices is discussed as well as the information transmission estimates.

CHAPTER 2 LITERATURE STUDY

2.1 MUSIC AND MUSIC PERCEPTION

Music perception studies usually investigate one or more of the following elements: i) pitch, ii) rhythm and iii) timbre. Although these elements form the basic building blocks of music, music is more than the sum of these three elements. Music is not the comprehension of a string of independent elements, but an arrangement of related elements (Peretz and Zatorre, 2005; Limb, 2006a). However, breaking up music into parts is a powerful tool for modelling music perception (Limb, 2006b).

2.2 MUSIC AND SPEECH

There are distinct modules for music and speech processing in the brain. Speech consist of rapid transitions (tens of milliseconds) and large frequency changes, while music usually consists of slower transitions with more precise timing and pitch (Zatorre et al., 2002; Limb, 2006b). A supporting study has shown that good speech perception is still possible with primarily temporal information after severe degradation of spectral content. Vowel and consonant recognition with only four spectral bands were well above chance and over 90% of words were correctly identified in simple sentence sets (Shannon et al., 1995). In contrast, it has been found that simple polyphonic melody recognition may require up to sixteen spectral bands and that musical enjoyment of more complex music may require up to 64 spectral bands (Shannon, 2005).

CI processors are optimised for speech intelligibility, which depends on cues that are considerably different to that of music, although studies have included correlations found between speech recognition scores and music related tasks (Gfeller et al., 1998; Galvin et al., 2007). Furthermore, training after implantation also focuses mainly on speech perception, while music perception is usually of secondary concern. Although the reasons for poor music task performance are numerous (of which limited temporal and spectral fine structure, limited number of frequency channels and lack of surviving hair-cells (McKay,

2005) are a few), music specific training appears to correlate with better performance in music recognition tasks (Gfeller et al., 2005; Galvin et al., 2007).

2.3 MUSIC PERCEPTION OF CI USERS

The difficulties of music perception by CI users are well documented (Limb, 2006a; McDermott, 2004; Drennan and Rubinstein, 2008). It is generally known that CI users fare significantly worse in music-related tasks compared to NH listeners.

It has been found that both appraisal (enjoyment) ratings and recognition scores are lower for CI users compared to NH listeners (Gfeller et al., 1998). A melodic pattern of four instruments were recorded (trumpet, clarinet, violin and piano). The stimuli represented four different families of instruments (brass, woodwind, string and percussion). The instrument recognition task presented the recordings and the participant had to choose which instrument was presented from a closed set of twelve instruments. The instrument appraisal task presented the recordings and the participant was asked to indicate enjoyment on a 100mm visual analogue scale representing high enjoyment on one end and low enjoyment on the other.

Scores indicated that CI users generally regarded instrumental sounds with lower appraisal compared to NH listeners, with appraisal scores of 47.31 compared to 57.03. The instrument recognition scores by Gfeller (2002) indicated that NH listeners identified the instruments correctly most of the time (ranging between 67.5% correct for violin to 100% correct for piano) and that confusion with other instruments was within the instrument family. In contrast, CI participant recognition scores varied between 20.2% correct (for clarinet) to 56% correct (for piano). The CI scores also indicated that instrument confusions did not necessarily fall within the same family of instruments.

Galvin et al. (2007) found that melodic contour identification varied greatly between CI participants. Melodic contours consisted of a five-note sequence consisting of rising, flat or falling pitches, or combinations of two of these contours. Larger intervals in pitch or melodic identification tasks increased scores. Furthermore, a significant correlation was found between CI users' melodic contour identification and vowel recognition performance. Finally, it was shown that training in musical tasks improved performance.

Gfeller et al. (2005) compared previously familiar melody recognition of CI users to NH listeners. They found that NH listeners accurately identified 54.7% of the melodies compared to 15.6% accurate identification by CI users. The difference was significant ($p<0.001$). The musical style of the test pieces also significantly influenced results. Pop and country items were more readily identified by CI users. This suggested a reliance on lyrics as an identification cue and was supported by correlations between correct identification of test items with lyrics and speech recognition.

Studies performed by Galvin et al. (2008, 2009a, 2009b) found that melodic contour recognition of the complex timbre of instrument sounds was significantly worse compared to the simpler timbre of a three-tone complex (a sound comprised of three harmonics). The CI user's reduced ability to recognise melodic contours in the presence of a masking instrument was also evident. The masking test may indicate a CI user's difficulty in separating the contours found in polyphonic melodies. A piano was used as the masker instrument and had a flat contour (repetition of the same note). The target melodic contour was varied between violin, organ and piano. The presence of a masking instrument resulted in significantly poorer melodic contour recognition by CI users, but not by NH listeners. No significant differences in results were found when comparing the violin and organ target contours for CI listeners, but results indicated significantly poorer results for the piano target. It is suggested that CI users make use of at least some timbre differences in extracting a target contour from competing melodic lines. This suggests that better timbre perception may lead to better melodic contour segregation in real-world music.

Rhythm perception studies have shown that the abilities of CI users to identify relatively simple stimuli compared well with NH listeners' abilities and that CI users relied heavily on temporal cues in music perception tasks (McDermott, 2004; Galvin et al., 2007; Kong et al., 2004).

It is evident from the literature that CI listeners have poor music perception compared to NH listeners. The evidence suggests that CI listeners rely heavily on temporal cues, such as rhythm and tempo, and speech cues, in the form of lyrics, during music listening. Poor instrument recognition of CI listeners points to limited discrimination of timbre properties. However, melodic stream segregation tasks involving two different instruments yielded better discrimination abilities compared to two streams of the same instrument, suggesting that CI listeners are able to make use of some timbre cues during listening tasks. Appraisal scores evidently suggest that CI listeners do not enjoy listening to music as much as NH listeners do.

2.4 TIMBRE PERCEPTION

Timbre is a multidimensional parameter of music. Several of these dimensions have been identified and labelled with a variety of descriptions. This makes comparison of the literature challenging. In order to gain insight into timbre perception of CI users, it is necessary to be knowledgeable about these dimensions.

In a study by McAdams et al. (1999), six timbre dimensions were defined:

i) Amplitude envelope smoothness (corresponding to shimmer (Jensen, 1999a)): an indication of microvariations in the amplitude of the sound.

ii) Amplitude envelope coherence (corresponding to spectral flux (Jensen, 2001)): an indication of how the spectral envelope changes over time. If the temporal envelopes of each of the harmonics have a similar shape, the amplitude envelope coherence is high.

iii) Spectral envelope smoothness (corresponding to irregularity (Krimphoff et al., 1994; Jensen, 1999a)): an indication of the differences in amplitude between subsequent harmonics.

iv) Frequency envelope smoothness (corresponding to jitter (Jensen, 1999a)): an indication of microvariations in the frequency of the sound.

v) Frequency envelope coherence (also known as stretched harmonics (Fletcher, 1971; Jensen, 1999a)): an indication of the spectral position of harmonic frequencies compared to the fundamental frequency.

vi) Frequency envelope flatness: a combination of frequency envelope coherence and frequency envelope smoothness.

Seven instruments were chosen to represent several branches of instrument types (see Figure 2.1). Instrument samples were modified to increase each of these timbral parameters (smoothing out microvariations and differences and increasing coherence), while keeping the other parameters constant. The effects of the changes on discrimination tasks were evaluated using a 2AFC discrimination task. Two pairs of sounds were presented. One pair consisted of the original samples while the other consisted of one original and one modified sample. The listener had to choose which pair contained two different sounds. The discrimination rates indicated that spectral smoothness and amplitude envelope coherence were the most salient features, followed by frequency flatness, frequency envelope coherence, frequency envelope smoothness and amplitude envelope smoothness.

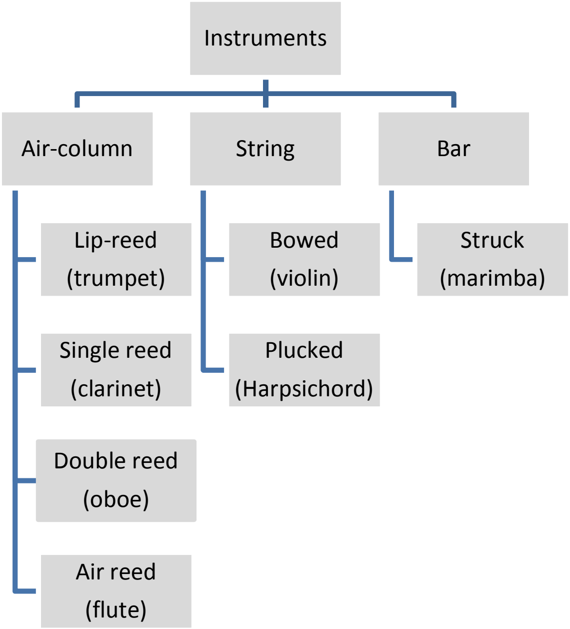

Figure 2.1. Instruments chosen by McAdams et al. (1999) to represent different families of instruments are classified according to their production methods.

It was observed that the salience of the timbre parameters differed for different instruments. For example, spectral smoothness was the most salient parameter for all instruments except the trumpet. This was due to the fact that the original spectrum of the trumpet was quite smooth. Figure 2.2 shows the discrimination rates for each of the instruments after timbre modifications.

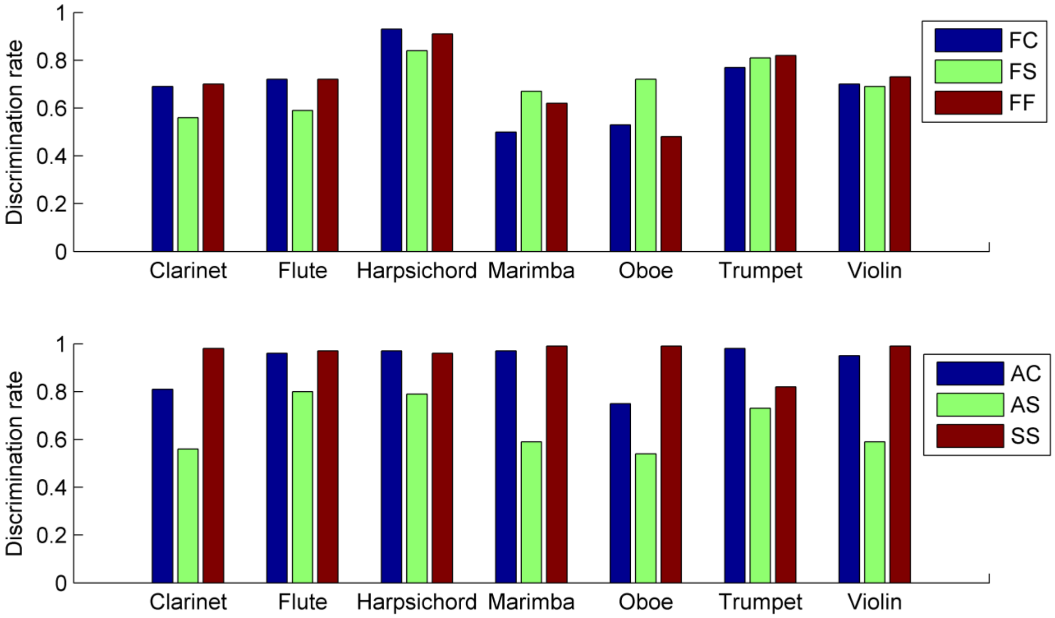

Figure 2.2. Discrimination rates for timbre modifications compared to original reference sounds as found by McAdams et al. (1999) are presented here as a graph. FC: Frequency Coherence, FS: Frequency Smoothness, FF: Frequency Flatness, AC: Amplitude Coherence, AS: Amplitude Smoothness, SS: Spectral Smoothness.

McAdams et al. (1995) investigated dissimilarity ratings of pairs of complex musical sounds, which showed that the spectral centroid and LRT were the most salient parameters, followed by attenuation of even harmonics (associated with spectral irregularity). Spectral flux was also a parameter used in identifying dissimilarity, but was not as salient a parameter as the previously mentioned parameters and it tended to be obscured if two other parameters concurrently varied. Since the spectrum only varied during an initial portion of the sound, it was suggested that spectral flux during a sustained segment might be more salient. In real-world sounds, however, spectral flux occurs in general only during the initial segments of notes (also refer to section 3.2).

Grey (1977) investigated three dimensions of timbre: i) spectral energy distribution, ii) synchronicity of the harmonics, and iii) presence of low-amplitude, high-frequency energy in spectral components during attack. The first two dimensions corresponded roughly to the spectral centroid or brightness and the spectral flux of a sound. The third dimension concerned itself with the attack portion of sounds and seemed to indicate instrument-type specificity. The importance of the spectral envelope contributing to the timbre was supported by studies from Gunawan and Sen (2008), Grey and Gordon, (1978), Horner et al. (2004) and Gabrielsson and Sjögren (1971).

In the literature, various important dimensions of timbre have been identified and used in perceptual studies, although timbre is not limited to these dimensions. Listeners readily perceive slight changes in spectral and temporal envelopes, frequency positions of harmonic components and spectral evolution (changes in spectral properties as a function of time). In instances where two dimensions of timbre are changed simultaneously, the relative salience of dimensions can be estimated. The literature provides information on the current knowledge of timbre perception and therefore provides a starting point for research into CI timbre perception.

2.5 TIMBRE PERCEPTION OF CI USERS

Recognition and appraisal scores for a small closed set of four instruments showed that CI users rated appraisal for two of the instruments significantly lower ($p<0.01$) than NH listeners. Recognition scores for CI listeners were also significantly lower for the entire instrument set compared to NH listeners and NH listeners showed greater consistency in their responses (Gfeller et al., 1998).

Supporting these findings, CI recognition of musical sounds from a closed set of sixteen instruments (McDermott and Looi, 2004) also indicated low recognition scores of 50% on average. Subjective quality ratings were also low and participants tended to assign higher quality ratings to those sounds that were more readily recognised.

The multidimensional scaling study by Kong et al. (2011) reported that CI listeners perceive timbre differently compared to NH listeners. The timbre perception of NH listeners were best characterised by a three dimensional timbre space of LRT, spectral centroid and spectral fine structure. The timbre perception of CI listeners also indicated one dimension correlating to LRT, while another dimension only weakly corresponded to the spectral envelope properties.

Timbre has also been found to be an important cue in the segregation of polyphonic melodies, which is helpful in perceiving and understanding music. CI users had greater difficulty in extracting a target melodic contour in the presence of masker instruments (Galvin et al., 2008, 2009b) compared to NH listeners. However, it was noted that some CI users were sensitive to target and masker timbre. CI participants 1 and 2 appeared to perform slightly better in their identification of the target instrument contour compared to the rest of the group and these were the participants who had more music experience before and after implantation.

The poor timbre perception of CI users is evident. Certain timbre properties are not as readily perceived by CI listeners compared to NH listeners, leading to poor performance in timbre perception tasks. Systematic investigation of timbre perception may contribute to a deeper understanding of the difficulties that CI users experience during music listening.

2.6 SYNTHESIS

Low CI user timbre recognition and appraisal scores undoubtedly points towards poor timbre perception, but specific problem areas remain unidentified. The multidimensional study by Kong et al. (2011) indicates that CI listeners perceive spectral cues different compared to temporal cues. Being able to compare JNDs of NH and CI listeners for temporal and spectral properties may identify a specific property or properties that limit CI users during timbre perception tasks. To obtain JNDs of specific timbre properties require specific type of stimulus.

Timbre perception studies use a variety of stimuli. Unmodified music excerpts and instrument recordings have been used in instrument recognition and appraisal tasks

(Gfeller et al., 1998; Gfeller et al., 2005; McDermott and Looi, 2004). Other studies made use of systematic modifications to the spectral and temporal properties of timbre to obtain dissimilarity ratings or discrimination measures (Grey and Gordon, 1978; Gunawan and Sen, 2008; Gabrielsson and Sjögren, 1972). Synthetic tones, as used by McAdams et al. (1995) and Grey (1977), have the potential to isolate any specific timbre property and is therefore an appropriate choice for investigating isolated timbre properties.

Important synthesis methods are discussed in the following sections. How the synthesized sounds are to be used and the advantages and disadvantages of each of the methods must be considered when choosing an appropriate synthesis method.

2.6.1 Waveguide synthesis

Physical models attempt to represent how a physical object (musical instrument) behaves in the presence of stimuli (breath, hammer or bow). An extensive knowledge of the physical properties of the instrument as well as the stimulus is required for the development of a physical model.

The Bilbao and Fitch (2006) piano model implemented in CSound[1] (Vercoe, 1985) uses the partial differential equation of a string with respect to time and one-dimensional space (displacement of the string). It is characterised by boundary conditions and initial conditions of both the string and hammer, the stiffness parameter (which is itself dependent on the physical properties of the string), the global decay rate of the string, the frequency dependent loss parameter of the string, the hammer mass and the hammer strike position on the string.

Variations of the physical parameters of the model produce variations in sound and influence the timbre parameters. Table 2.1 shows how brightness and irregularity change when the physical parameters of the piano change. Using the default parameters described in Table 2.1 in CSound with a pitch of 440 Hz (A4) and changing one physical parameter at a time for the ranges indicated, changes the brightness and irregularity as shown in the

[1] Csound is a programming language based on C and optimized for sound synthesis.

last two columns. The ranges were subjectively chosen to produces tones that still sound piano-like.

Table 2.1. Changing the physical parameters of the piano model of CSound change the brightness and irregularity.

Parameter	Default	Range	T_b extremes	IRR extremes
Number of strings	3	-	-	
Amount of detunedness [cents]	10	0-10	1.27-1.35	0.47-0.58
Stiffness [dimensionless]	2	0-5	1.32-1.37	0.50-0.51
30 dB decay time [s]	3	2-8	1.32-1.35	0.48-0.53
High frequency loss [dimensionless]	0.002	0-0.04	1.32-1.38	0.51-0.57
Boundary conditions	Pivotal	-	-	-
Piano hammer mass [g]	1	1-11	1.27-1.34	0.51-0.54
Natural frequency of hammer [Hz]	5000	1000-11000	1.23-1.60	0.39-0.97
Initial hammer position [m]	0.01	0-0.02	1.33-1.54	0.48-0.51
Normalised strike position along string	0.09	0-0.4	1.13-1.59	0.49-0.80
Normalised strike velocity	50	10-150	1.33-2.50	0.39-1.00

From Table 2.1 it is clear that changes in the physical parameters of the model cause changes in the timbre parameters, IRR and T_b. Although the quality of sounds produced with physical waveguide models is good, controlling one timbre parameter of a sound independently from another timbre parameter by adjusting a physical parameter may be difficult, if not impossible.

In-depth studies of physical models explain some of the basic timbre parameters for various instruments (Fletcher and Rossing, 1999). For example, evaluation of piano spectra shows that the physical hardness of the hammer has an effect on the brightness of the piano sound. Harder hammers tend to excite more high-frequency harmonics compared to softer hammers. Furthermore, a higher velocity of the hammer effectively produces a higher hardness. The contact time of the hammer on the string also produces differences in timbre. Hammers that are much lighter compared to the mass of the string have less contact time with the string and produce brighter sounds due to smaller attenuation of high frequency sounds. Many of the physical structures of pianos greatly influence its tone. Temporal decay times increase towards the higher pitch register of the piano. 60 dB decay times may vary from 0.2 s to 50 s from the higher to lower pitches. The fundamental frequency dominates the piano spectrum over most of the pitch range, but is weak in the two lowest octaves. The position of hammer strike affects the resulting spectrum. Depending on the hammer strike position along the string, some harmonics may be attenuated. A hammer striking at a fraction 1/ß of the length of the string produces attenuated harmonics at multiples of ß harmonics. Perception of dynamic expressions for the piano has more to do with the timbre of the piano rather than the actual sound level. “Louder” notes (player using more force to hit notes or opening of the piano lid) markedly increase the strength of higher harmonics, but produce relatively small changes in actual sound levels.

It is clear that the various physical properties of instruments influence timbre qualities. As an example, Figure 2.3 shows the waveguide model of a flute. Expressive elements are easiest to achieve with a physical model as the inputs and model parameters represent physical instrument sound production inputs and physical properties that can simply be changed to produce variations in sound. Including amplitude variations in the flow envelope of Figure 2.3, which represents breath pressure, will cause a flute sound with vibrato. Changing the bore delay compares physically to opening and closing the holes of the flute, which changes the length of the resonator – and thus the frequency of the note. Systematic control of individual timbre parameters would be difficult using such a waveguide model, since the relationship between physical and timbre parameters is not necessarily simple (also evident from Table 2.1). This synthesis method was therefore not considered suitable for the investigation of timbre parameter JNDs.

2.6.2 Resampling

Synthesis using resampling is also known as wavetable synthesis. Examination of an instrumental sound reveals that each instrument has a characteristic wave shape. This shape can be digitally stored (in a wavetable) and it is also possible to store a characteristic start and end segment for the specific instrument. In order to reproduce the note, the start segment is played, followed by a repetition of the characteristic waveshape, as required by the length of the reproduced sound and completed by the end segment. Different pitches can be produced by resampling the wavetable and replaying the sound at the original frequency.

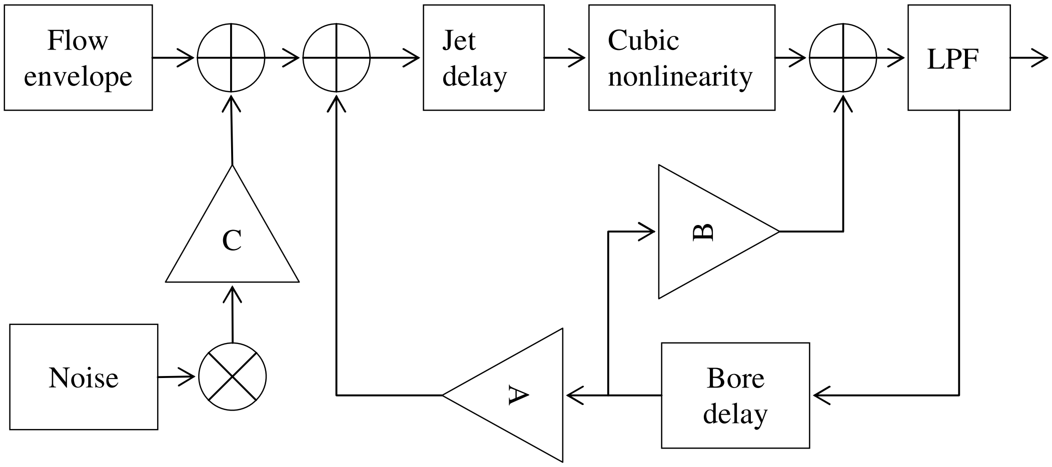

Figure 2.3. P. Cook's waveguide model of the flute (Cook, 1992).

Although resampling may work for a range of notes, instrument sound reproduction may use a few recordings of an instrument at a range of pitches and switching or interpolating between the samples during reproduction (Cook, 2002).

With present day digital storage space, resampling produces high fidelity sound using a simple method that is not computationally expensive, but the range of sounds that can be achieved is limited with respect to manipulation, for example expressive elements. This method is useful for recreating sound using limited controls, e.g. a midi keyboard (Jehan, 2001), but is limited with regard to experimental methods to obtain timbre parameter JNDs. Therefore, this method was not considered a viable option for experimentation.

2.6.3 Additive synthesis

Additive synthesis creates the original sound by adding sinusoids together as a result of analysis of the spectrum of a sound. From an alternative point of view, subtractive synthesis generates complex sounds rich in many harmonics. The use of filters achieves the desired response by attenuating the undesired components (De Poli, 1983).

Additive synthesis allows complete control over the spectrum. Analysis of the Fourier transform of a periodic wave shows which frequencies are most prominent in the spectrum. These frequencies change in absolute amplitude and in amplitude relative to each other for the duration of the note. The original signal can be estimated by adding together the main frequencies found through spectral analysis. The major problem with spectral synthesis is that each harmonic amplitude and phase must be defined in order to recreate high fidelity sounds (Jehan, 2001, De Poli, 1983). Specific timbre properties are therefore not necessarily explicitly defined.

It is possible to develop mathematical models that attempt to recreate instrument sounds with additive synthesis using only a limited set of specific timbre properties. The mathematical timbre model developed by Jensen (1999a) is an additive synthesis model that synthesizes sounds from a limited set of timbre properties. A mathematical approach to the additive synthesis method is advantageous for the adaptive experimental procedure, since individual characteristics adjustments require relatively simple calculations that can effortlessly be done by the computer running the experimental procedure. Due to these characteristics, a mathematical additive synthesis model was considered an appropriate synthesis model.

The detailed development of the mathematical approach to the additive synthesis method is described in section 3.3.

CHAPTER 3 METHODS I: ANALYSIS AND SYNTHESIS

Different synthesis methods enable different degrees of control over synthetic tones. The synthesis method described in this chapter allows complete control over chosen timbre parameters. The analysis and synthesis of tones involve three steps. First, an appropriate timbre parameter set must be defined. Second, instrument recordings must be analysed to calculate the timbre parameters associated with each instrument. Finally, the additive synthesis model must be developed and stimuli must be created using the model.

3.1 CONSIDERATION OF TIMBRE PARAMETERS

Simple models of timbre usually include at least the following three parameters or an equivalent or subsection thereof: i) spectral centroid; ii) spectral envelope, including the irregularity; iii) temporal envelope. The spectral centroid is a property associated with the spectral locus of a sound. Irregularity looks at the difference in amplitude between subsequent harmonic partials. Studies of the temporal properties usually consist of segments such as attack, S/D and release, but are sometimes limited to only the logarithm of the attack segment or rise-time (RT) (Krimphoff et al., 1994; Jensen, 2001; Grey, 1977; McAdams et al., 1995; Caclin et al., 2005). More sophisticated models may include additional timbre parameters, for example: shimmer (variations in the temporal amplitude envelope); jitter (variations in the frequency of harmonics); stretched harmonics (where each subsequent harmonic may not be a perfect integer multiple of the fundamental frequency); spectral flux (changes in spectral envelope from the start to end times of a sound); and noise (Jensen, 2001, 1999a). Shimmer and jitter are two different types of vibrato. Each is applicable to different instruments and can also be defined as expressive musical manipulations. Other examples of musical manipulations that can change timbral parameters are: legato (smooth, connected notes); staccato (detached notes); con sordino (using mutes, various types of which exist for each instrument); con pedale (using a pedal) in piano; or pizzicato (plucked string) for bowed instruments – to name just a few.

Expressive manipulations can also serve as instrument identification cues, since some manipulations are only associated with certain instruments.

Various important timbre properties from the literature are summarised for consideration in the mathematical synthesis method. Advantages and shortcomings are discussed where different usages of the same concept occur.

3.1.1 Brightness

Brightness is a perceptually linear psychoacoustic parameter associated with the spectral locus of a sound. The term brightness is an accurate verbal description of the timbre property. A trumpet is an example of an instrument with a high brightness, while the mellow sound of a French horn has a low brightness value. The following equation is used to determine the brightness of a sound.

$$Brightness = T_b = \frac{\sum_{k=1}^{N} k a_k}{\sum_{k=1}^{N} a_k} \tag{3.1}$$

a_k is the amplitude of the k^{th} harmonic and N is the total number of harmonics (Krimphoff et al., 1994; Caclin et al., 2005). Synonyms of brightness are spectral centroid, spectral centre of gravity and spectral locus. From equation (3.1) it is clear that brightness is a unitless parameter.

3.1.2 Logarithmic rise-time

The LRT is defined as:

$$LRT = \log(t_{max} - t_{0.1max}), \tag{3.2}$$

where t_{max} is the time where the maximum amplitude is reached, and $t_{0.1max}$ is the time where 10% of the maximum is reached (Krimphoff et al., 1994). The LRT is a description of the onset of the note and is measured in log(s). To produce a note on an oboe, for example, requires an almost explosive breath of air. The abrupt start of an oboe tone therefore has a low RT and LRT. In contrast the production of a bowed violin tone can grow smoothly from a quiet start to a maximum loudness and therefore has a long or high RT and LRT. The LRT may accurately describe the temporal onset of notes, but gives no further information regarding the remainder of the temporal envelope. A note can be sustained (flute or bowed string) or it may decay (piano or plucked string).

Consideration should also be given to real-world recordings during analysis when using equation (3.2), as the maximum amplitude of a sustained note is sometimes reached later during a sample, as shown in Figure 3.1. Defining the RT of this sample as the time taken for the amplitude to increase from 10% of the maximum to the maximum is inaccurate. Although different percentage values decrease the probability of such inaccuracies using this method, it may still not be accurate for all samples. For example, in Figure 3.1, percentages from 10% of the maximum to 90% of the maximum yield an RT of 0.7348 s, which is still much longer than a subjective estimation of the RT (approximately 0.1 s) would produce. In the case of Figure 3.1, a subjective estimation would produce the most accurate reading.

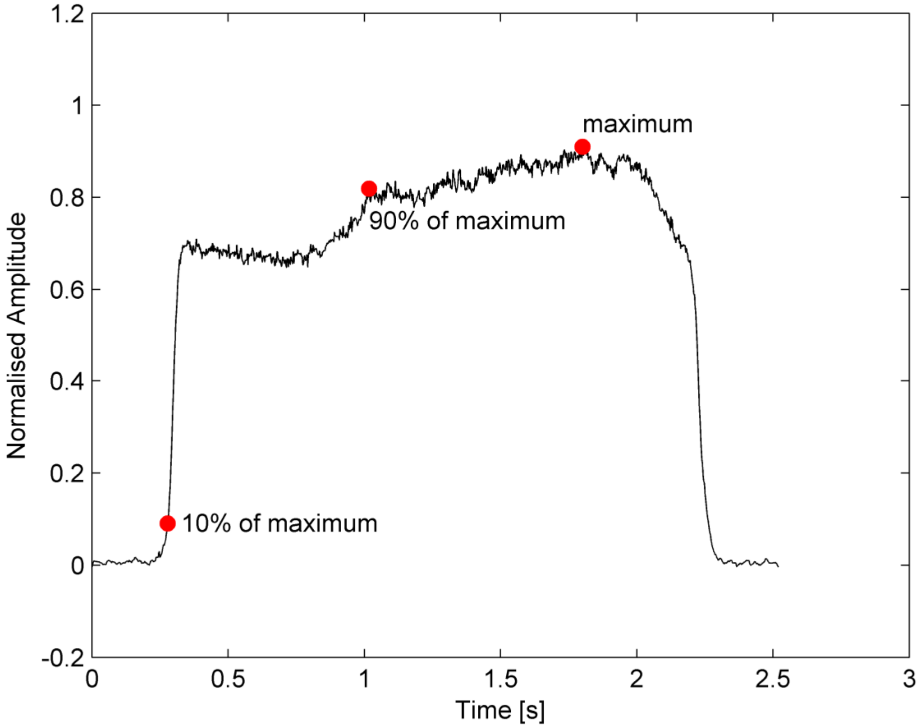

Figure 3.1. The RT found from this clarinet recording at 440 Hz is found as approximately 1.5 s, according to equation (3.2), although inspection reveals that it should be closer to 0.1 s.

3.1.3 Temporal envelope

Jensen (1999a, 1999b) described the entire temporal envelope, not only the LRT. To determine the attack and release of an envelope, the maximum and minimum of the derivative of the envelope was found. These respectively indicated where the slope increased and decreased the fastest. From these points, the derivative curves were followed backwards and forwards respectively until they were smaller and larger than the maximum and minimum of the derivative multiplied by some constant. The four points indicated the start-of-attack (soa), end-of-attack (eoa), start-of-release (sor) and end-of-release (eor). Jensen suggested constants of approximately 0.125 for the soa, eoa and eor and 0.4 for the sor.

To connect the four points, Jensen suggested two possible curves:

$$Curve = v_0 + (v_1 - v_0)(1 - (1 - x)^n)^{\frac{1}{n}} \tag{3.3}$$

$$\text{and } eCurve = v_0 + (v_1 - v_0)\frac{e^{nx} - 1}{e^n - 1}. \tag{3.4}$$

v_0 and v_1 were respectively the amplitudes of start- and end points of the segment. The value of n determined the shape of the curve. It must be noted that these equations are undefined when $n = 0$. When these equations are used, x, which represents the time axis, should be normalised. The segments connecting the four points were labelled as the attack, S/D and release segments and had the advantage of accurately describing both sustain and decay temporal envelopes. The Levenberg-Marquardt fitting algorithm provided a least-square error fit (Moré, 1977).

Even with certain assumptions that the amplitude at the soa and eor is zero and the amplitude at the eoa is one (for normalised amplitude envelopes), the envelope model still requires six values. The amplitude value at the sor will be different for sustain vs. decay envelopes and the start and end times and n-values for each of the three curve segments are required.

A less salient timbre parameter regarding the temporal envelope was the presence and extent of microvariations in the temporal amplitude envelope (McAdams et al., 1999) (see also section 3.1.7). Although microvariations will always be present in performed notes, the expressive element called vibrato is essentially a manipulation causing wanted microvariations in the temporal envelope for certain instruments such as the flute. In the literature, this property is also referred to as shimmer (Jensen, 1999a, 2001) and amplitude modulation (AM) (Eronen and Klapuri, 2000). Investigation of this less salient parameter shows that its perception is dependent on its own set of parameters, such as the frequency and strength of AM.

3.1.4 Irregularity

The irregularity of the spectrum indicates the difference in strength between adjacent harmonics. Krimphoff et al. (1994) defined the irregularity as:

$$Irregularity = \log\left(\sum_{k=2}^{N-1}\left|20\log(a_k) - 20\left(\frac{\log a_{k-1}+\log a_k+\log a_{k+1}}{3}\right)\right|\right). \qquad (3.5)$$

Alternatively, Jensen (1999a) defined irregularity as:

$$Irregularity = \frac{\sum_{k=1}^{N}(a_k - a_{k+1})^2}{\sum_{k=1}^{N} a_k{}^2}, \qquad (3.6)$$

where the $(N+1)^{th}$ harmonic amplitude was assumed to be zero. Irregularity has no unit.

Other similar timbre dimensions described the irregularity as the attenuation of even harmonics (Caclin et al., 2005). Instruments with high irregularity, such as the clarinet, can best be described as sounding “hollow” or “wooden”.

3.1.5 Spectral flux

While many timbre parameters can be described as being either a spectral or temporal property, spectral flux is a combination. Some synonyms for spectral flux are spectral evolution (Cusack and Roberts, 2004; Chowning, 1973), synchronicity (Grey, 1977) and amplitude envelope coherence (McAdams et al., 1999). Brass instruments are often described as having high spectral flux during the initial segments of the tone. Careful listening to brass instrument tones oftentimes reveal that the tone develops in time from a dull to a bright sound. Spectral flux is such a spectral change in time.

To determine how a spectrum changes as a function of time, a short-time frequency transform (STFT) may appear to be the simplest solution, but this is subject to limitations in accuracy due to the trade-off between frequency resolution and time resolution inherent in frequency transforms. Since musical notes are being analysed, inherent properties of quasi-harmonic notes and music can be used to achieve good spectral and temporal resolution. It is known that frequency transitions take longer in music than in speech (Zatorre et al., 2002). Furthermore, for quasi-harmonic notes, accurate predictions are possible of the positions of the largest spectral components that make up the sound. Applying bandpass filters to single harmonics for longer durations of the signal and calculating the fast-Fourier-transforms (FFTs) of the filter outputs allows high temporal resolution tracking of each of the harmonic components (Guillemain and Kronland-Martinet, 1996). The result is high temporal and spectral resolution. Helpful figures for this concept can be found in section 3.2.

The salience of spectral flux is much debated in the literature. Chowning’s (1973) frequency modulation (FM) synthesis contained the means for spectral evolution, since it was recognised that for certain instruments, these dynamic spectra may contribute to timbre. It was mentioned that the presence of spectral evolution may contribute greatly to reducing the synthetic or electronic sound properties of synthesised tones.

Multidimensional scaling techniques performed on similarity ratings of sixteen instruments showed a specific dimension related to spectral fluctuations throughout instrument tones as well as synchronicity of higher harmonics (Grey, 1977). Also, one of the three timbre dimensions found by McAdams et al. (1995) corresponded to spectral flux.

Caclin et al. (2005) found that varying spectral flux can be used in dissimilarity ratings. However, when spectral flux was covaried with the spectral centroid (brightness) and attack time, its influence on dissimilarity ratings decreased.

Analysis of the brightness and irregularity as a function of time also shows that these parameters do not vary much throughout the duration of an instrument tone – see section 3.2.

3.1.6 Inharmonicity

Some voiced instruments (instruments with discernible pitch) have harmonics at frequencies that are not integer multiples of the fundamental. Although most voiced instruments have harmonics at integer multiples of the fundamental, harmonics for tensioned strings occur at *approximately* integer multiples, as per the following equation (Fletcher, 1971; Fletcher and Rossing, 1999).

$$f_k = kf_0\sqrt{1 + \beta k^2} \tag{3.7}$$

This phenomenon is called stretched harmonics. Higher values of β produce sounds with a higher degree of stretched harmonics. A descriptive term for instruments with a high degree of stretched harmonics is “bell-like”.

For struck bars with i) both ends clamped, ii) one end clamped and one end free and iii) both ends hinged, the frequencies occur at:

$$f_1 = \frac{\pi\kappa}{8L^2}\sqrt{\frac{E}{\rho}}3.011^2; f_k = \frac{\pi\kappa}{8L^2}\sqrt{\frac{E}{\rho}}(2k+1)^2\ for\ k = 2,3,4 \ldots, \tag{3.8}$$

$$f_1 = \frac{\pi\kappa}{8L^2}\sqrt{\frac{E}{\rho}}1.194^2;\ f_2 = \frac{\pi\kappa}{8L^2}\sqrt{\frac{E}{\rho}}2.988^2;\ f_k = \frac{\pi\kappa}{8L^2}\sqrt{\frac{E}{\rho}}(2k-1)^2\ for\ k = 3,4,5 \ldots \tag{3.9}$$

and

$$f_k = \frac{\pi\kappa}{2L^2}\sqrt{\frac{E}{\rho}}k^2\ for\ k = 1,2,3 \ldots, \tag{3.10}$$

where κ is the radius of gyration and $\kappa = \frac{h}{\sqrt{12}}$ for a rectangular bar of height h, L is the length of the bar, E is Young’s modulus and ρ the density of the bar (Fletcher and Rossing, 1999). A glockenspiel is an example of a metallic bar instrument and a marimba is an example of a wooden bar instrument.

Although bells do not necessarily have stretched harmonics, they do contain harmonics at frequencies that are not integer multiples of the fundamental. For example, the first six harmonics ($f_1 - f_6$) of a church bell (Fletcher and Rossing, 1999) are $\{f_1 = f_0;\ f_2 = f_0 \times 2^{12/12};\ f_3 = f_0 \times 2^{15/12};\ f_4 = f_0 \times 2^{19/12};\ f_5 = f_0 \times 2^{24/12};\ f_6 = f_0 \times 2^{28/12}\}$.

3.1.7 Vibrato

Vibrato is a musical term that describes slight variations in amplitude or frequency controlled by the musician. Such controllable gestures are called expressions (Jensen, 2001). This may also serve as an identification cue, since an instrument's vibrato type is either AM or FM. For example, to achieve vibrato, a flute player varies breath pressure, resulting in amplitude variations (AM). This is also often more accurately identified as tremolo. On the other hand, vibrato in violin tones continuously varies string length, which varies pitch (FM). Some instruments rarely use vibrato (organ), while for others, vibrato is not possible (piano, marimba, etc.).

3.1.8 Detunedness

Some parameters are entirely instrument specific. The piano has one or two strings in its lower note register, while the higher register has three strings. The tuning of each of string pair or group of three is never perfect, leading to two or three very close frequencies constructively and destructively interfering to cause a pulsating amplitude envelope. This effect is also known as the occurrence of beats during the note duration.

3.1.9 Phase information

The phase information of the harmonics influences the temporal shape of the sound wave and therefore the timbre (Plomp and Steeneken, 1969). A measure of the phase information can be found as a peak factor and is defined as $(Amplitude\ range/2\sqrt{2}) \times RMS(y(t))$, where $y(t)$ is the sound wave (Schroeder and Strube, 1986). Figure 3.2 shows two waveforms consisting of the same harmonics, but differing in phase. The upper wave resulted from the addition of sine harmonics and has a higher peak factor compared to the lower wave, which is the result of the addition of alternating sines and cosines.

3.1.10 Summary

Parameters $T_{b,}$ IRR and LRT were chosen to include in the synthesis model. These parameters were chosen due to their prevalent presence in multidimensional studies. In order to more accurately describe a complete temporal envelope and discriminate between sustain and decay instruments, the S/D parameter was also included. Detunedness was disregarded here due to their absence in some instruments, while the arguable salience of flux also eliminated this parameter from the model. Vibrato was excluded since this timbre is an optional timbre parameter controlled by the performer.

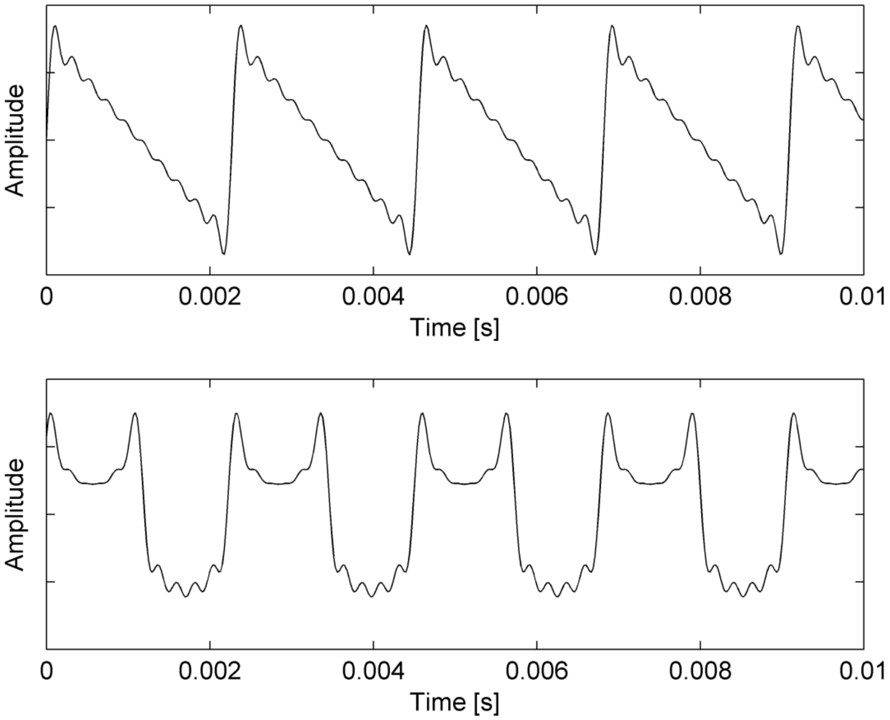

Figure 3.2. Two waveforms consisting of identical harmonics, but differing in phase. The upper waveform reflects of the addition of sinusoids, while the lower waveform reflects the addition of alternating sines and cosines.

3.2 ANALYSIS OF TIMBRE PARAMETERS OF INSTRUMENTS

The spectral parameters T_b and IRR and temporal parameters LRT and S/D were chosen to represent each instrument in timbre space. Flux was not included in the model due to its reduced salience when covaried with other parameters, such as brightness (section 3.1.5). Investigation of the spectral properties of instrument recordings as a function of time indicated little variation throughout the note (Figure 3.5) and supported the elimination of flux from the timbre set. Inharmonicity and detunedness were not included in the timbre set due to their presence in only a limited number of instruments, such as bar instruments, bells and instruments with stretched strings like the piano.

Instrument recordings from The University of Iowa Electronic Music Studios (Fritts, 1997) were analysed for the four chosen timbre parameters. The middle C's, (C4, 262 Hz) of the clarinet, French horn, oboe, plucked and bowed violin and cello, flute, piano, trumpet, tuba, saxophone and trombone were used. These instruments represented four instrument families. Section 3.3.4 confirms that these instruments also represent the timbre space well.

In order to find the brightness and irregularity of each of the instruments, the harmonics of each of the sounds were extracted. This was done by taking the FFT of the note recording in its entirety, since only the static spectrum was used. The static nature during the S/D parts of the note for each instrument was validated by placing narrowband filters across each of the harmonics and extracting the temporal envelopes of each harmonic. This allows investigation of the irregularity and brightness as a function of time. Figure 3.3 shows how the filters extract each harmonic individually. Figure 3.4 shows the output of the filters for a 262 Hz clarinet tone. Figure 3.5 indicates how the brightness and irregularity change during the S/D part of the note sample and that these parameters remain approximately constant. This supports the decision not to include flux to the timbre set (see section 3.1.5). The harmonics were obtained for the sound samples of each instrument. The brightness was calculated using equation (2.1).

Krimphoff et al. (1994) defined irregularity as harmonic component deviation from the local spectral envelope (equation (2.5)). However, for the analysis and resynthesis proposed here, Jensen's (1999a, 1999b) equation was used (equation (2.6)). Although the equation by Krimphoff et al. is more intuitive in explaining the concept of irregularity, Jensen's equation was used as it does not cause large variations in irregularity as the number of harmonics used varies (Figure 3.6 and Figure 3.7). Also, an analytical solution for a specific brightness and irregularity is possible during resynthesis for Jensen's equation.

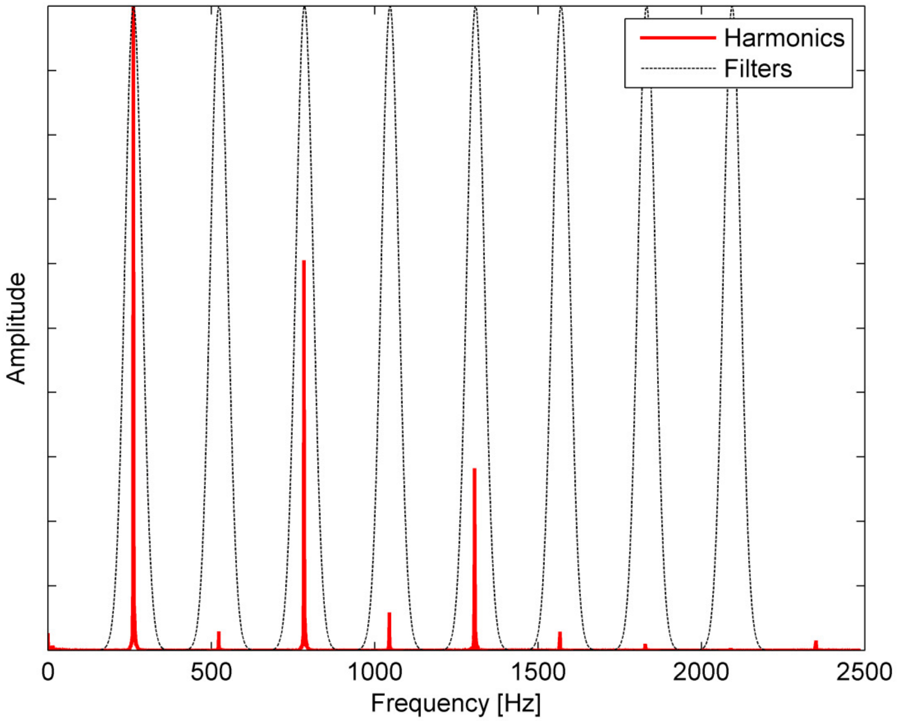

Figure 3.3. Narrowband filters placed across the first eight harmonics of a C4 (262 Hz) clarinet recording.

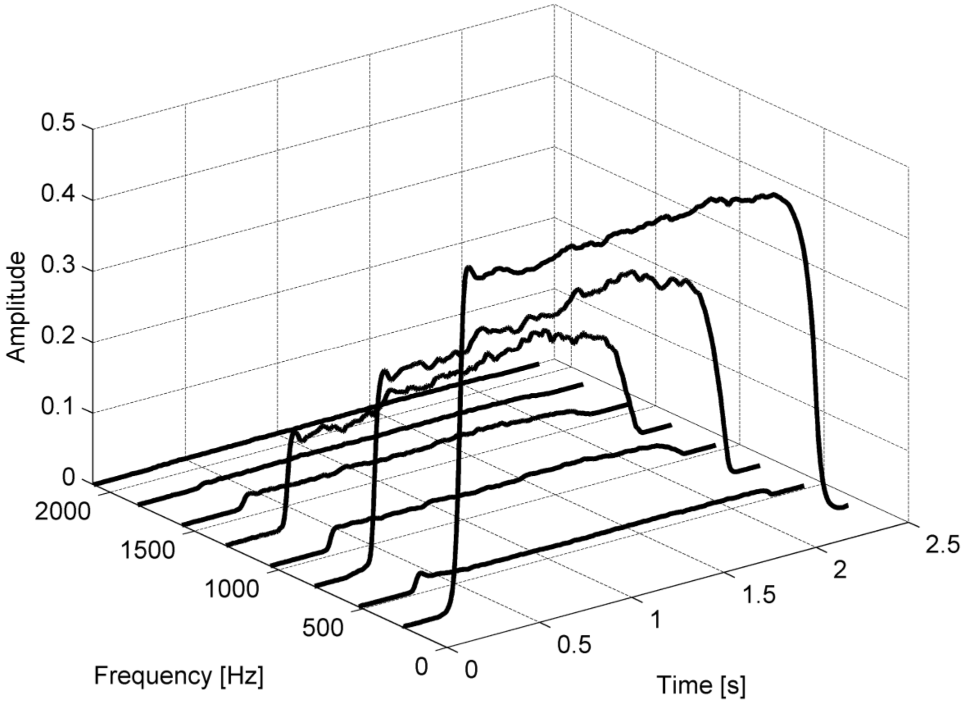

Figure 3.4. Temporal envelopes of the outputs of each of the filters for a C4 (262 Hz) clarinet recording.

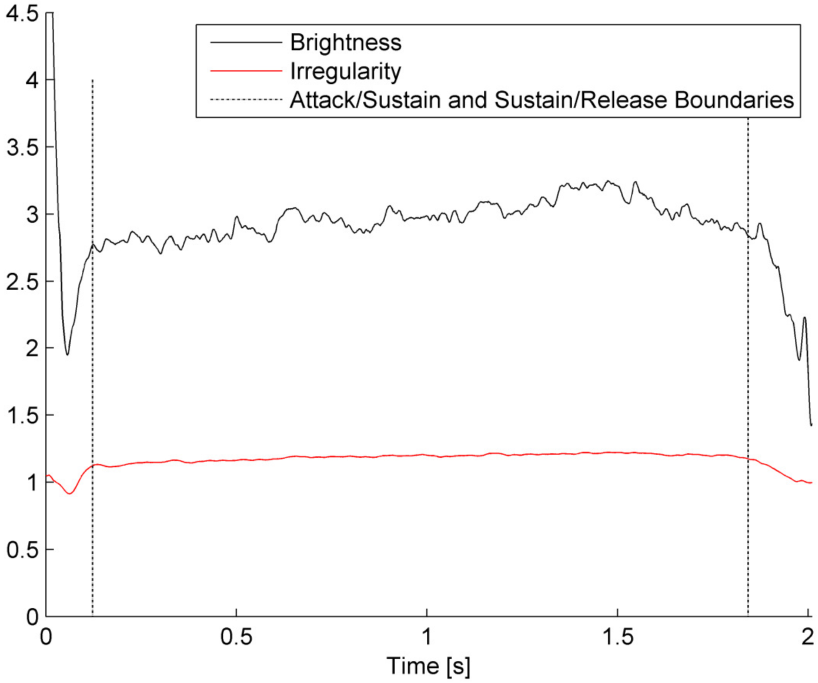

Figure 3.5. Brightness and irregularity are shown for the duration of a clarinet C4 (262 Hz) recording. The vertical lines separate the attack, S/D and release segments of the tone. The graph suggests that the brightness and irregularity remain approximately constant during the S/D part of the tone.

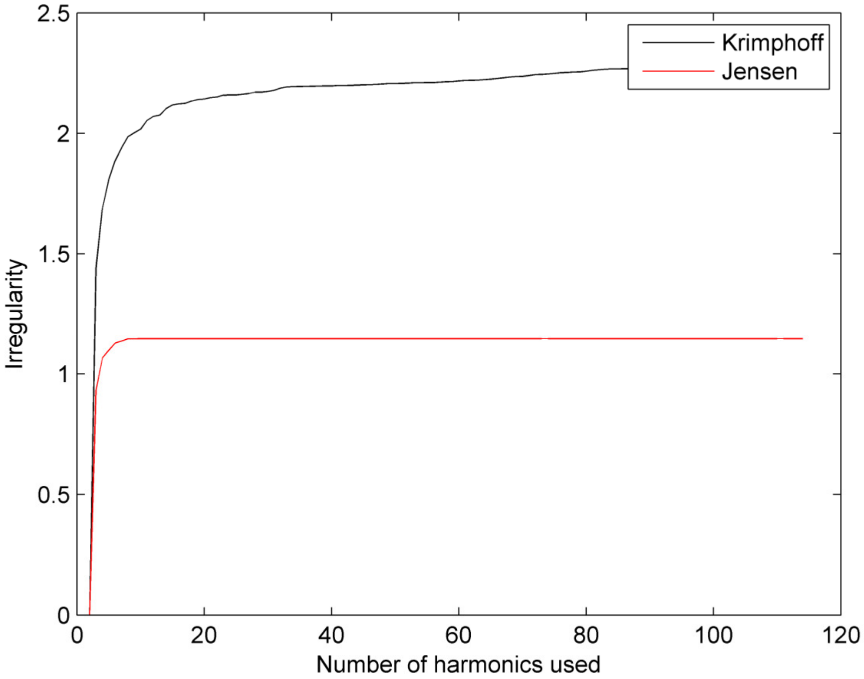

Figure 3.6. Irregularity as a function of the number of harmonics used with the Krimphoff et al. (1994) and Jensen equations (1999a, 1999b) for a C4 (262 Hz) clarinet tone.

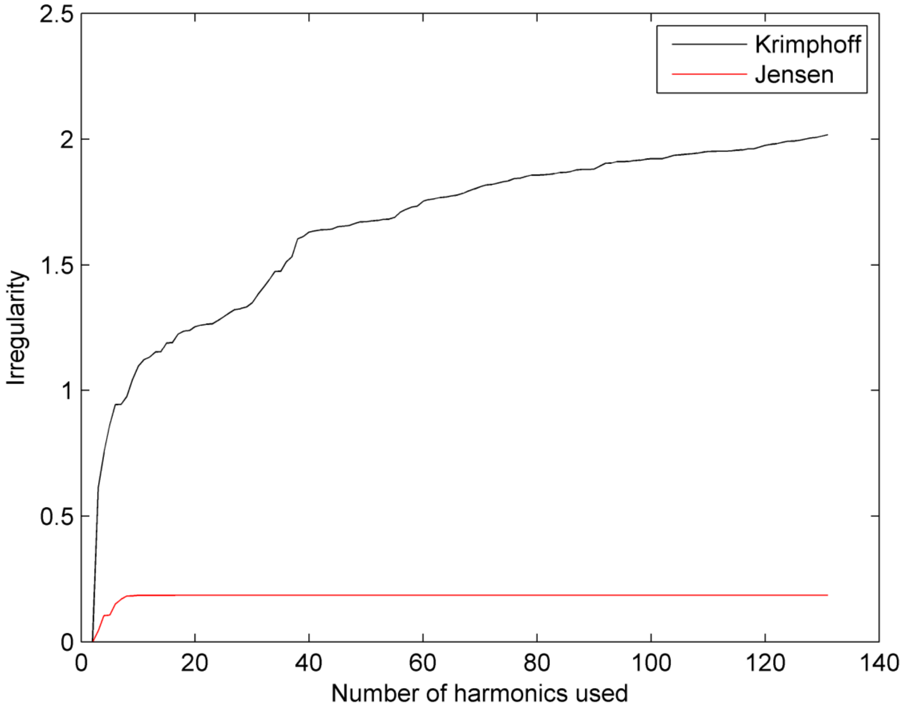

Figure 3.7. Irregularity as a function of the number of harmonics used using the Krimphoff et al. (1994) and Jensen equations (1999a, 1999b) for a C4 (262 Hz) trumpet tone.

The start- and end points of the temporal envelope were obtained as described by Jensen (1999a, 1999b). This method eliminates problems where the LRT, as described by Krimphoff et al. (1994) ($LRT = \log(t_{max} - t_{0.1max})$), falls short when real-world recordings reach their maximum amplitude long after the actual attack segment (see Figure 3.1).

The procedure based on Jensen's procedure (1999a, 1999b) is summarised as follows (refer to Figure 3.8):

i) The temporal envelope of the sound is extracted.

ii) The temporal envelope is smoothed by multiplying the FFT of the temporal envelope with that of the FFT of a very wide Gaussian.

iii) The maximum of the derivative of the first half of the temporal envelope is found. This indicates the highest slope of the temporal envelope and is defined as the attack.

iv) The minimum of the derivative of the second half of the temporal envelope is found. This indicates the largest negative slope of the temporal envelope and is defined as the release.

v) From the attack location, the derivative is followed backwards and forwards in time until the derivative becomes smaller than 0.125 times the value of the local maximum. From the release location, the derivative is followed backwards and forwards in time until the derivative becomes respectively larger than 0.4 times the local minimum value and 0.125 times the local minimum value. These four locations are then defined as the soa, eoa, sor and eor.

vi) Since the points obtained may be inaccurate, due to the smoothed nature of the envelope, (envelope smoothing decreases transient slopes), it is necessary to decrease the width of the Gaussian in steps. Microvariations increase with narrower Gaussians, increasing the chances of incorrectly obtaining the start and release of the sample. The Gaussian was therefore decreased until correlation between the smoothed and original envelope reached 99.9%.

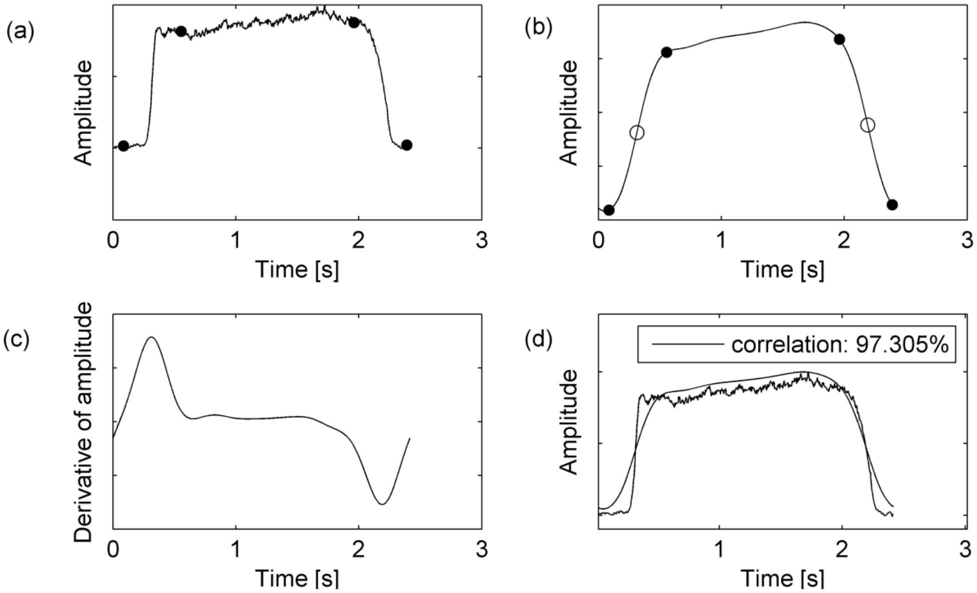

Figure 3.8. (a) The extracted envelope of the sample. (b) The resulting smoothed envelope when the inverse FFT of the FFT of the temporal envelope multiplied by the FFT of a wide Gaussian was taken. (c) The derivative of the smoothed envelope with the maximum and minimum corresponding to the attack and decay (unfilled circles in (b)).The locations soa, eoa, sor, eor are indicated by filled circles in (b). (d) Comparison of the smoothed and original temporal envelope shows a correlation of 97.3%. The process is repeated with a narrower Gaussian in step two until the smoothed envelope is a sufficient representation of the original temporal envelope.

The LRT can simply be found as the time from the soa to eoa.

$$LRT = \log(t_{eoa} - t_{soa}) \tag{3.11}$$

As discussed in section 3.1.2 and shown in Figure 3.1, this method produces less inaccurate readings compared to the method suggested by Krimphoff et al. (1994).

The number of temporal parameters can be reduced by combining Jensen's (1999a, 1999b) suggested S/D and release segments into one segment. Using equation (3.4) and setting $v_0 = 1$ and $v_1 = 0$, sustained or decaying segments can be represented with only one parameter value: n. The only constraint is that $n \neq 0$, although as $n \to 0$, equation (3.4) becomes linear. For values of $n > 0$, the instrument is sustained and when $n < 0$, the instrument is decaying. Figure 3.9 shows examples of various LRT and n.

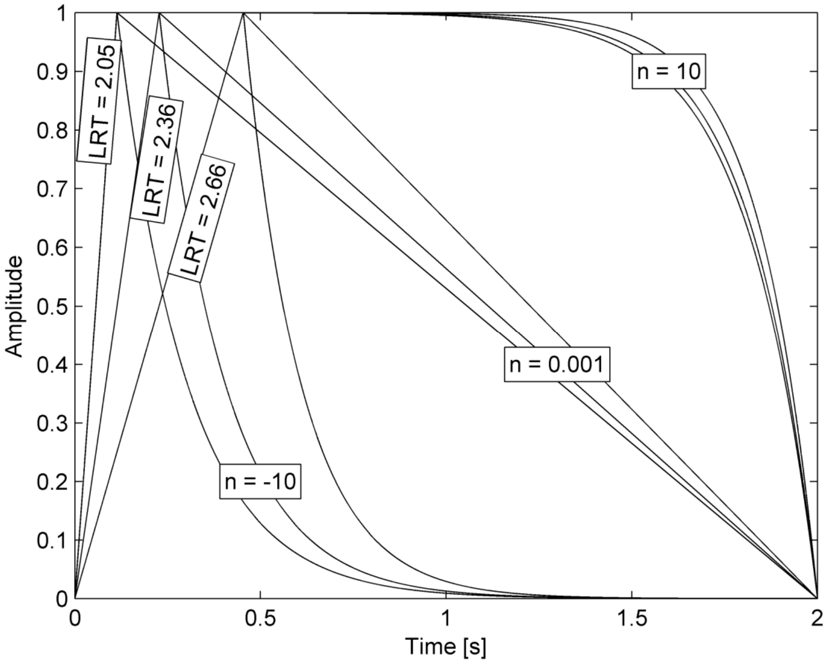

Figure 3.9. Examples of temporal envelopes with varying LRT (2.05, 2.36, 2.66) and n (-10, 0.001, 10).

The S/D segment was isolated using the eoa and eor values and normalised in time. Using the Levenberg-Marquardt numerical algorithm, equation (3.4) can be fit to instrument recording envelopes to find n.

Using equations (3.1), (3.6), (3.11) and (3.4) allows the calculation of T_b, IRR, LRT and S/D for the middle C's of the list of thirteen instruments of section 3.2. These instruments are representative of instrument families as well as being representative of the timbre space (see Figure 3.11 and Figure 3.12). The parameter values are summarised in Table 3.1.

Table 3.1. Timbre parameter values of the instrument set identified for use in experimental procedures.

Instrument nr.	Instrument name	T_b	IRR	LRT	S/D (n)
1	Clarinet	3.00	1.16	2.07	17.0
2	French Horn	2.41	0.192	1.56	18.2
3	Oboe	5.23	0.603	1.66	12.3
4	Violin (arco – bowed)	5.25	0.568	2.50	6.52
5	Violin (pizzicato – plucked)	2.58	0.562	1.22	-19.6
6	Flute	3.70	0.133	2.09	8.04
7	Piano	2.42	0.297	1.27	-7.49
8	Trumpet	5.41	0.185	1.53	14.8
9	Tuba	2.57	0.329	1.55	4.28
10	Cello (arco – bowed)	6.65	0.991	2.74	2.16
11	Cello (pizzicato – plucked)	1.88	0.920	1.18	-12.9
12	Saxophone	3.46	0.330	2.16	2.20
13	Trombone	4.18	0.227	2.00	8.69

3.3 ADDITIVE SYNTHESIS MODEL

Table 3.1 contains the analysed timbre parameter values of real-world recordings. The mathematical approach to the additive synthesis mentioned in section 2.6.3 is developed here and involves three steps. The synthesis method allows exact manipulation of the spectrum according to the spectral timbre parameter values; so first, a sound with a specific spectrum is synthesized according to the spectral properties T_b and IRR. This is followed by the recreation of the temporal envelope according to LRT and S/D. Finally, the temporal envelope is fitted to the sound from the first step.

3.3.1 Spectral model

To reduce the large set of harmonic component values, harmonic amplitudes were not stored individually, but estimations of the harmonics were attained using only the values for brightness and irregularity. In order to resynthesize either of the spectral parameters of brightness or irregularity, the remaining spectral parameter must also be taken into account. Recreating a specific brightness when the irregularity is not taken into account can be done by setting $a_k = B^{-k}$, where a_k is the amplitude of the k^{th} harmonic and $B = \frac{T_b}{T_b - 1}$ (Jensen, 1999a; Caclin et al., 2005).

The irregularity is implemented into the spectrum by multiplying the odd components with a factor x. This multiplication, however, influences the brightness so that B is not simply equal to $\frac{T_b}{T_b - 1}$. The solution of x and B must be solved simultaneously from equations (3.1) and (3.6) to obtain the desired value of IRR and T_b.

Assuming the number of components used, N, is even and that $N + 1$ is equal to zero, the irregularity when odd spectral components are multiplied by a factor x is:

$$IRR = \frac{(xB^{-1} - B^{-2})^2 + (B^{-2} - xB^{-3})^2 + \cdots + (xB^{-(N-1)} - B^{-N})^2 + (B^{-N} + 0)^2}{(xB^{-1})^2 + (B^{-2})^2 + \cdots + (B^{-N})^2 + 0} \tag{3.12}$$

The quadratic terms are computed and grouped.

$$IRR = \frac{[x^2B^{-2} + 2x^2B^{-6} + 2x^2B^{-10} + \cdots + 2x^2B^{-(2N-2)}] + [2B^{-4} + 2B^{-8} + \cdots + 2B^{-2N}]}{x^2B^{-2} + B^{-4} + x^2B^{-6} + \cdots x^2B^{-(2N-2)} + B^{-2N}} + \frac{[-2xB^{-3} + 2xB^{-5} + 2xB^{-7} + \cdots + 2xB^{-(2N-1)}]}{x^2B^{-2} + B^{-4} + x^2B^{-6} + \cdots x^2B^{-(2N-2)} + B^{-2N}} \tag{3.13}$$

Note that the first term of the first numerator group is not multiplied by two. We separate this term from the group and rewrite the group as a sum of terms. The final numerator group is also rewritten as two pairs of summations up to the upper limit of N instead of $2N$.

$$IRR = \frac{\left[x^2B^{-2} + 2x^2 \sum_{k=1}^{N/2-1} B^{-4k-2}\right] + \left[2\sum_{k=1}^{N/2} B^{-4k}\right] - \left[2x \sum_{k=1}^{N-1} (B^{-2k-1})\right]}{x^2 \sum_{k=1}^{N/2} B^{-4k+2} + \sum_{k=1}^{N/2} B^{-4k}} \tag{3.14}$$

All lower summation limits are changed to $k = 0$.

$$IRR = \frac{\left[x^2B^{-2} + 2x^2\sum_{k=0}^{N/2-1} B^{-4k-2} - 2x^2B^{-2}\right] + \left[2\sum_{k=0}^{N/2} B^{-4k} - 2\right] - 2x\sum_{k=0}^{N-1}(B^{-2k-1}) + 2xB^{-1}}{x^2\sum_{k=0}^{N/2} B^{-4k+2} - x^2B^2 + \sum_{k=0}^{N/2} B^{-4k} - 1} \tag{3.15}$$

It is known that $\sum_{k=0}^{\infty} y^k = \frac{1}{1-y}$ for $|y| < 1$. By definition, $B > 1$. Therefore, setting $y = \frac{1}{B^2}$ or $y = \frac{1}{B^4}$ yields $|y| < 1$ (in both cases). Hence $\sum_{k=0}^{\infty} B^{-2k} = \frac{1}{1-1/B^2} = \frac{B^2}{B^2-1}$ and $\sum_{k=0}^{\infty} B^{-4k} = \frac{B^4}{B^4-1}$.

The limit as the number of harmonics approach infinity is:

$$\lim_{N\to\infty} IRR = \frac{\left[x^2B^{-2} + 2x^2B^{-2}\left(\frac{B^4}{B^4-1}\right) - 2x^2B^{-2}\right] + \left[2\left(\frac{B^4}{B^4-1}\right) - 2\right] - \left[2xB^{-1}\left(\frac{B^2}{B^2-1}\right) + 2xB^{-1}\right]}{x^2B^2\left(\frac{B^4}{B^4-1}\right) - x^2B^2 + \left(\frac{B^4}{B^4-1}\right) - 1} \tag{3.16}$$

Multiplying the numerator and denominator by $B^2(B^4 - 1)$ yields:

$$\lim_{N\to\infty} IRR \\ = \frac{[x^2(B^4-1) + 2x^2B^4 - 2x^2(B^4-1)] + [2B^6 - 2B^2(B^4-1)] + [-2xB^3(B^2+1) + 2xB(B^4-1)]}{x^2B^8 - x^2B^4(B^4-1) + B^6 - B^2(B^4-1)} \tag{3.17}$$

After simplification, the irregularity is therefore defined as:

$$IRR = \frac{2B^2 - 2xB^3 + x^2B^4 + x^2 - 2xB}{x^2B^4 + B^2} \tag{3.18}$$

The brightness where odd spectral components are multiplied by a factor x is:

$$T_b = \frac{1xB^{-1} + 2B^{-2} + 3xB^{-3} + \cdots + (N-1)xB^{-(N-1)} + NB^{-N}}{xB^{-1} + B^{-2} + xB^{-3} + \cdots xB^{-(N-1)} + B^{-N}} \tag{3.19}$$

$$T_b = \frac{\sum_{k=1}^{N/2}(2k-1)xB^{-2k+1} + \sum_{k=1}^{N/2}(2k)B^{-2k}}{\sum_{k=1}^{N/2} xB^{-2k+1} + \sum_{k=1}^{N/2} B^{-2k}} \tag{3.20}$$

$$T_b = \frac{\sum_{k=1}^{N/2} 2kxB^{-2k+1} - \sum_{k=1}^{N/2} xB^{-2k+1} + \sum_{k=1}^{N/2} 2kB^{-2k}}{\sum_{k=1}^{N/2} xB^{-2k+1} + \sum_{k=1}^{N/2} B^{-2k}} \tag{3.21}$$

Taking the derivative of $\sum_{k=0}^{\infty} y^k = 1 + y + y^2 + y^3 + \cdots = \frac{1}{1-y}$ for $|y| < 1$ yields $\frac{d}{dy}\left(\frac{1}{1-y}\right) = \left(\frac{1}{1-y}\right)^2 = 1 + 2y + 3y^2 + 4y^3 + \cdots = \sum_{k=1}^{\infty} ky^{k-1}$ for $|y| < 1$. The lower limits of the terms with a factor k remain $k = 1$ and lower limits of the terms without a factor k are changed to $k = 0$.

$$T_b = \frac{2xB\sum_{k=1}^{N/2} k(B^{-2})^k - xB\sum_{k=0}^{\frac{N}{2}}(B^{-2})^k + xB + 2\sum_{k=1}^{N/2} k(B^{-2})^k}{xB\sum_{k=0}^{N/2}(B^{-2})^k - xB + \sum_{k=0}^{\frac{N}{2}}(B^{-2})^k - 1} \tag{3.22}$$

Rewriting the terms to suit the derivative form yields:

$$T_b = \frac{2xB^{-1}\sum_{k=1}^{N/2} k(B^{-2})^{k-1} - xB\sum_{k=0}^{\frac{N}{2}}(B^{-2})^k + xB + 2B^{-2}\sum_{k=1}^{N/2} k(B^{-2})^{k-1}}{xB\sum_{k=0}^{N/2}(B^{-2})^k - xB + \sum_{k=0}^{N/2}(B^{-2})^k - 1} \tag{3.23}$$

Taking the limit results in:

$$\lim_{N\to\infty} T_b = \frac{2xB^{-1}\frac{B^4}{(B^2-1)^2} - xB\frac{B^2}{B^2-1} + xB + 2B^{-2}\frac{B^4}{(B^2-1)^2}}{xB\frac{B^2}{B^2-1} - xB + \frac{B^2}{B^2-1} - 1} \tag{3.24}$$

Simplification yields:

$$\lim_{N\to\infty} T_b = \frac{2xB\left(\frac{B^2}{(B^2-1)^2}\right) - xB\left(\frac{B^2}{B^2-1}\right) + xB + 2\left(\frac{B^2}{(B^2-1)^2}\right)}{xB\left(\frac{B^2}{B^2-1}\right) - xB + \left(\frac{B^2}{B^2-1}\right) - 1} \tag{3.25}$$

The brightness is therefore defined as:

$$T_b = \frac{xB^3 + 2B^2 + xB}{xB^3 + B^2 - Bx - 1} \tag{3.26}$$

MATLAB's *solve* function can be used to obtain valid values of B and x for a given brightness and irregularity ($B \in \mathbb{R}, B > 1$ and $x \in \mathbb{R}, x > 0$) using equation (3.18) and equation (3.26). Valid solutions for T_b and IRR are shown in Figure 3.10. The brightness, by definition, must be larger than one and must be a real number. Irregularity, by definition, must be larger than zero, smaller than two and a real number. For example, Figure 3.10 indicates that a sound with $IRR = 6$ and $T_b = 2$ is not possible using this model. Since the odd components are multiplied by x, care must be taken when x-values are close to zero, since such values may yield a tone perceived as twice the intended frequency, as all odd frequencies are inaudible. For the perceived frequency, the irregularity would then be different to what was intended.

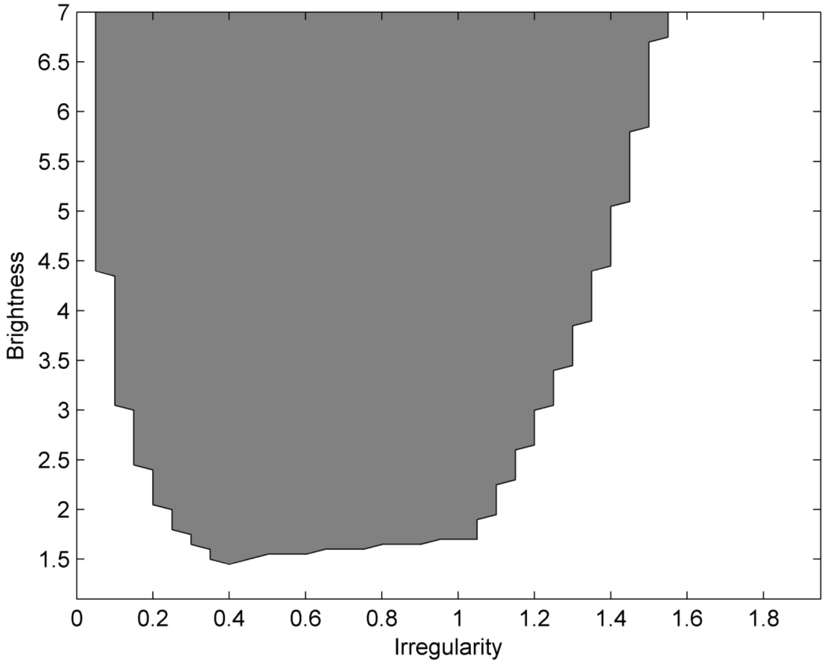

Figure 3.10. The grey area on the map indicates where valid brightness and irregularity values can be found.

The values of x and B are used to obtain the amplitudes of the harmonics components: a_k.

$$\begin{aligned} a_k &= B^{-k}\, for\ k = 2, 4, 6, 8, \ldots \\ a_k &= xB^{-k}\, for\ k = 1, 3, 5, 7, \ldots \end{aligned} \tag{3.27}$$

The tones are recreated by adding sinusoids of amplitudes a_k and frequencies at kf_0 (f_0 is the fundamental frequency) for $k = 1,2,3, \ldots N$, where N is the number of harmonics used.

$$s(t) = \sum_{k=1}^{N} a_k \sin\left(2\pi k f_0 t\right) \tag{3.28}$$

Unlike waveguide synthesis, the additive synthesis method creates only the spectral domain of the sound. Temporal parameters are also important contributions to timbre. To include the temporal dimensions, the temporal envelope can be created separately and multiplied by the signal obtained through additive synthesis.

3.3.2 Temporal envelope

The following procedure was used to recreate the temporal envelope. A line segment was used for the attack with $RT = 10^{LRT}$ in milliseconds (ms). Since n was found (section 3.1.3, equation (3.4)) for normalised time, the S/D segment is only recreated correctly for normalised time (1 s). In order to recreate the sound for any desired length, equation (3.4) can be used for $t = 0$, where the amplitude equals 1 to t of any desired duration, where the amplitude falls below zero for values of $t > 1$. This correctly recreates the shape of the envelope. The amplitude of the sound sample must then be normalised. This procedure allows the correct curve according to n for any desired duration. equation (3.4) was chosen to describe the S/D segment, as it seemed more appropriate due to the presence of the exponential function and its relevance to the physical world.

3.3.3 Timbre model

The sound wave, $s(t)$, of section 3.3.1 is multiplied by the temporal envelope to create the final sound resynthesized according to its analysed parameters.

3.3.4 Representative instruments

Some studies chose a set of representative instruments by classifying the instrument according to its family (McAdams et al., 1999). For NH listeners it has been found that confusion exists mostly within instrument family classes (Gfeller et al.; 2002). Table 3.2 shows family classifications of the chosen instrument set. The instruments used for the present study are representative of the woodwind, brass, percussive and string instrument families.

Since the purpose of this study was to examine the important dimensions of timbre, the chosen instruments also had to be representative of the entire four-dimensional space. Figure 3.11 and Figure 3.12 show the locations of the clarinet (clar), French horn (horn), oboe (oboe), bowed violin (violA), plucked violin (violP), flute (flut), piano (pian), trumpet (trum), tuba (tuba), bowed cello (cellA), plucked cello (cellP), saxophone (saxo) and trombone (trom) on the spatial representation plot of the four dimensions of timbre as two two-dimensional plots. The instruments are spread throughout the space and are not clustered together. In this way the instruments can be regarded as being a representative instrument set of the timbre space.

Table 3.2. The orchestral and production method classifications are shown here.

Instrument	**Instrument Family**	**Sound Production Method**
Horn	Brass	Lip reed
Trombone	Brass	Lip reed
Tuba	Brass	Lip reed
Trumpet	Brass	Lip reed
Clarinet	Woodwind	Single reed
Saxophone	Woodwind	Single reed
Oboe	Woodwind	Double reed
Bassoon	Woodwind	Double reed
Flute	Woodwind	Blow hole
Bowed (arco) violin	Strings	Bowed string
Bowed (arco) cell	Strings	Bowed string
Plucked (pizzicato) violin	Strings	Plucked string
Plucked (pizzicato) cello	Strings	Plucked string
Piano	Percussive	Struck string

3.4 SUMMARY

From the literature, timbre parameters were considered and four important parameters were chosen to be used in developing a synthesis method. The developed model could be used to synthesize stimuli with specific timbre parameters. The spectral envelope was recreated using the results from equations (3.18) and (3.26) in equations (3.27) and (3.28) and fitting a temporal envelope using equation (3.4) and (3.11) to the resulting sound.

It was confirmed that the instrument set was representative of instrument families and that the timbre space was well represented by the instruments. However, with psychoacoustic timbre studies, a necessary step is to equalise tones in loudness, perceived duration and pitch. The methodology is discussed in the following chapter.

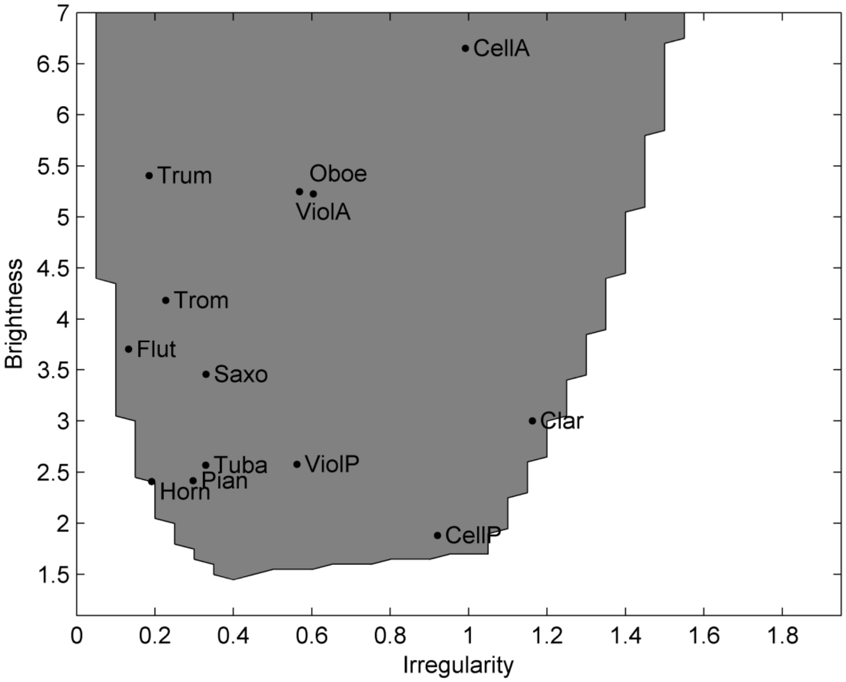

Figure 3.11. The two spectral timbre dimensions. The timbre space is four-dimensional.

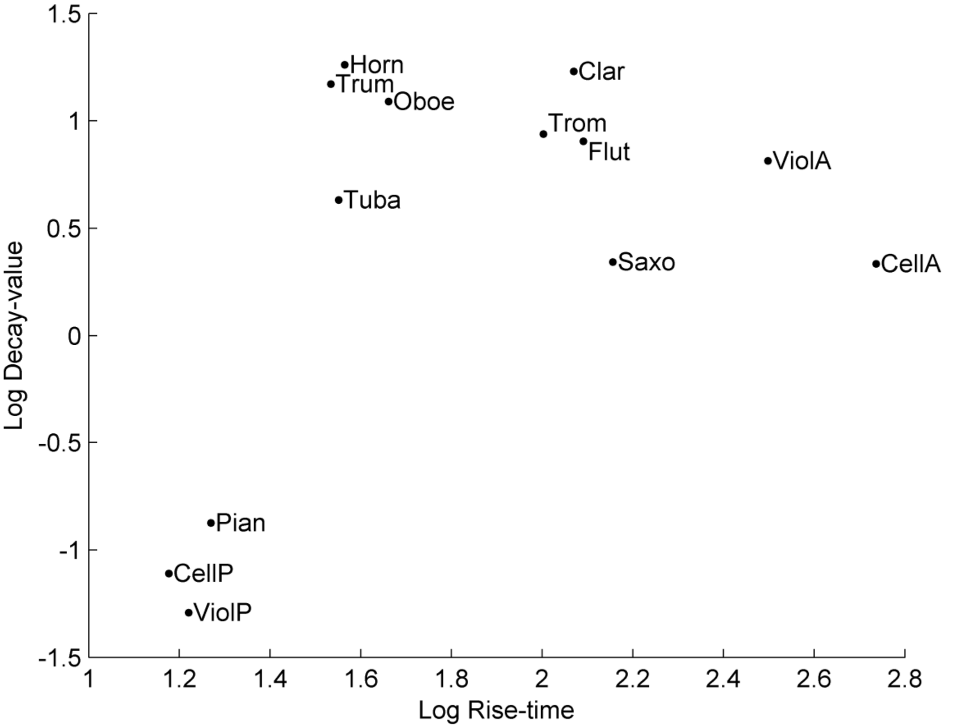

Figure 3.12. The two temporal timbre dimensions. The timbre space is four-dimensional.

CHAPTER 4 METHODS II: LOUDNESS, PITCH AND PERCEIVED DURATION BALANCING

Four important timbre parameters have been extracted from the representative set of instruments. A timbre model was developed in sections 3.3.1, 3.3.2 and 3.3.3 with the objective of resynthesizing sounds using only the four timbre values. Simultaneous solution of equation (3.18) and equation (3.26) allows the harmonic magnitudes of equation (3.27) to be found. A tone can then be synthesized according to equation (3.28) and fitted with a temporal envelope described in section 3.3.2. Table 3.1 contains the values of each instrument to be used in the model. This model is suitable for discrimination tasks, since each parameter can be varied individually while keeping all others constant. However, before continuing with discrimination experiments, stimuli had to be balanced for loudness, perceived duration and pitch (Grey, 1975). For example, when discriminating between two sounds of unequal T_b, timbre parameters IRR, LRT and S/D and non-timbre properties loudness, pitch and perceived duration, must be equal. Balancing is necessary to eliminate the possible influence of these cues on timbre discrimination results. The timbre parameters are controlled by the developed additive synthesis method, but in order to balance loudness, perceived duration and pitch as timbre parameters change require perceptual tests in which listeners adjust tone intensity, duration and frequency. The methods are described in this chapter.

Although studies usually balance stimuli for specific listeners, many studies make use of a pilot group of participants to balance loudness, perceived duration and pitch, while a different group of participants is used during the main experiments (Marozeau et al., 2003; McAdams et al., 1995, Grey, 1977).

In the present study, three of the listeners participating in the balancing procedures also participated in the discrimination procedures, although an entirely new participant group could also have been used. Experimental reference tones were recreated using the four timbre parameters and six participants were asked to compare test tones that varied in timbre parameters to the reference tones.

4.1 PARTICIPANTS

Six NH listeners participated in the balancing procedures. It was confirmed through audiometric screening that all participants had NH (pure tone thresholds ≤ 20 dB HTL for 250 Hz, 500 Hz, 1000 Hz, 2000 Hz, 4000 Hz and 8000 Hz). The three females and three males had an average age of 25, with all participants aged between 21 and 28.

4.2 STIMULI

Before testing commenced, the loudness of reference tones were set at 75% of the listener's loudness estimation curve (see Appendix B). All presented tones for balancing procedures were presented at this 75% point of the participant's loudness dynamic range, unless the listener preferred higher or lower levels (see Appendix section B). All tones were presented between 68.2 dBSPL and 76 dBSPL. All tones were sampled at 44.1 kHz and were presented in a soundproof booth through a KEF Q30 loudspeaker using an M-Audio Fasttrack Pro external soundcard. The test tones were compared to reference tones which were the original synthesized sounds according to the values of Table 3.1. The tones were of 2 s duration, except for the test tones of the perceived duration balancing procedure.

4.2.1 Loudness balancing

Each instrument's brightness and irregularity were varied independently across its allowable range in eight equal increments according to Figure 3.10. The upper brightness limits were set to two times the original brightness of the instrument in question. For LRT, values ranged from 1 to 3 (10 ms to 1000 ms) in eight linear steps. For S/D, eight logarithmically spaced n-values, ranging from a quarter to four times the original n-value of the instrument, were used (see Appendix section A). Each timbre parameter was varied individually from all others parameters for each of the instruments. Table 4.1, Table 4.2, Table 4.3 and Table 4.5 shows the test tone values consisting of the eight values used for T_b, IRR, LRT and n for each of the instruments. A total of 416 loudness adjustments were made for four timbre parameters, thirteen instruments and eight timbre intervals.

4.2.2 Perceived duration balancing

During the perceived duration balancing procedure, the LRT values were varied from half to twice the original LRT of the instrument in question, unless the upper limit exceeded an LRT of 3.1. These values are summarised in Table 4.4. Excluding the decay instruments, the values of Table 4.5 were used during perceived duration balancing for changes in n-values. The decay values were excluded, because the nature of these instruments do not allow duration changes. When necessary, the balancing responses of the sustain instruments (Clarinet, French horn, oboe, bowed violin, flute, trumpet, tuba, bowed cello, saxophone and trombone) were compared separately to the balancing responses of the decay instruments (plucked violin, piano and plucked cello). The motivation for the separation is evident from Figure 4.3 and Figure 4.6. It was assumed that spectral changes would not influence perceived duration. The eight increments of ten sustain instruments of two temporal parameter conditions totalled 160 perceived duration adjustments.

4.2.3 Pitch balancing

Table 4.1 and Table 4.2 indicate the values used for stimuli during the pitch balancing tasks. It was assumed that temporal parameter LRT and n would not change the pitch of the stimuli. A total of 208 pitch adjustments were made for the two spectral conditions, thirteen instruments and eight timbre variables.

Table 4.1. The incremental brightness values used for the balancing procedures.

Brightness								
Clarinet	2.35	2.87	3.39	3.92	4.44	4.96	5.48	6.00
French horn	2.11	2.50	2.88	3.27	3.66	4.04	4.43	4.82
Oboe	1.55	2.82	4.09	5.36	6.64	7.91	9.18	10.45
Bowed violin	1.54	2.82	4.10	5.38	6.66	7.94	9.22	10.50
Plucked violin	1.53	2.05	2.57	3.08	3.60	4.12	4.64	5.15
Flute	2.59	3.28	3.97	4.65	5.34	6.03	6.72	7.40
Piano	1.62	2.08	2.54	3.00	3.46	3.92	4.37	4.83
Trumpet	2.15	3.39	4.62	5.86	7.10	8.34	9.57	10.81
Tuba	1.51	2.03	2.55	3.07	3.58	4.10	4.62	5.14
Bowed cello	1.67	3.33	4.99	6.65	8.32	9.98	11.64	13.30
Plucked cello	1.65	1.95	2.25	2.56	2.86	3.16	3.46	3.76
Saxophone	1.51	2.28	3.05	3.82	4.60	5.37	6.14	6.91
Trombone	1.90	2.82	3.75	4.67	5.59	6.52	7.44	8.36

Table 4.2. The incremental irregularity values used for the balancing procedures.

Irregularity								
Clarinet	0.11	0.27	0.44	0.60	0.76	0.92	1.09	1.25
French horn	0.16	0.30	0.45	0.59	0.74	0.88	1.03	1.17
Oboe	0.04	0.24	0.45	0.65	0.85	1.05	1.26	1.46
Bowed violin	0.04	0.24	0.45	0.65	0.85	1.05	1.26	1.46
Plucked violin	0.14	0.29	0.44	0.59	0.74	0.89	1.04	1.19
Flute	0.07	0.25	0.43	0.61	0.79	0.88	1.03	1.17
Piano	0.16	0.30	0.45	0.59	0.74	0.88	1.03	1.17
Trumpet	0.04	0.24	0.45	0.65	0.86	1.06	1.27	1.47
Tuba	0.14	0.29	0.44	0.59	0.74	0.89	1.04	1.19
Bowed cello	0.03	0.25	0.46	0.68	0.89	1.11	1.32	1.54
Plucked cello	0.24	0.36	0.48	0.60	0.73	0.85	0.97	1.09
Saxophone	0.08	0.25	0.43	0.60	0.78	0.95	1.13	1.30
Trombone	0.06	0.25	0.43	0.62	0.81	1.00	1.18	1.37

Table 4.3. The incremental log rise-time values used for the loudness balancing procedures.

Log rise-time								
All instruments	1	1.29	1.57	1.86	2.14	2.43	2.71	3

Table 4.4. The incremental log rise-time values used for the perceived duration balancing procedures.

Log rise-time								
Clarinet	1.03	1.33	1.62	1.92	2.21	2.51	2.81	3.1
French horn	0.78	1.11	1.44	1.78	2.11	2.44	2.77	3.1
Oboe	0.83	1.16	1.48	1.80	2.13	2.45	2.78	3.1
Bowed violin	1.25	1.53	1.78	2.04	2.31	2.57	2.84	3.1
Flute	1.05	1.34	1.63	1.93	2.22	2.51	2.81	3.1
Trumpet	0.77	1.10	1.42	1.75	2.08	2.41	2.74	3.07
Tuba	0.78	1.11	1.44	1.77	2.10	2.44	2.77	3.1
Bowed cello	1.37	1.62	1.86	2.11	2.36	2.61	2.85	3.1
Saxophone	1.08	1.37	1.66	1.94	2.23	2.52	2.81	3.1
Trombone	1.00	1.30	1.60	1.90	2.20	2.50	2.80	3.1

Table 4.5. The incremental sustain/decay values used for the balancing procedures.

Sustain/decay								
Clarinet	4.25	6.32	9.39	13.95	20.72	30.80	45.76	68.00
French horn	4.55	6.76	10.05	14.94	22.20	32.99	49.02	72.84
Oboe	3.08	4.57	6.80	10.10	15.01	22.30	33.13	49.24
Bowed violin	1.63	2.42	3.60	5.34	7.94	11.8	17.54	26.06
Plucked violin	-4.90	-7.28	-10.81	-16.07	-23.87	-35.48	-52.72	-78.34
Flute	2.01	2.99	4.44	6.60	9.80	14.57	21.65	32.17
Piano	1.87	2.78	4.14	6.15	9.14	13.58	20.17	29.98
Trumpet	3.70	5.51	8.18	12.16	18.07	26.84	39.89	59.28
Tuba	1.07	1.59	2.36	3.51	5.22	7.75	11.52	17.12
Bowed cello	0.54	0.80	1.19	1.77	2.63	3.91	5.81	8.63
Plucked cello	-3.22	-4.78	-7.10	-10.55	-15.68	-23.3	-34.63	-51.46
Saxophone	0.55	0.82	1.22	1.81	2.68	3.99	5.92	8.80
Trombone	2.17	3.23	4.80	7.13	10.06	15.75	23.40	34.78

4.3 LOUDNESS BALANCING

4.3.1 Procedure

The stimuli had to be balanced for loudness. The experimental procedure was controlled with MATLAB software. The participant could compare the test tone with the reference tone as many times as needed by clicking on the reference and test tone buttons. The original intensity of the test tone was set equal to the intensity of the reference tone. A slider bar controlled the test tone's intensity, which could vary between +6 dB and -6 dB of the intensity of the reference tone. The slider step sizes were arranged linearly on the dB scale. After the participant was satisfied that the test tone loudness was equal to the reference tone, the result could be saved. The next tone pair was loaded.

The test involving a particular instrument was presented sequentially, with the eight brightness intervals (see Table 4.1) being presented first, followed by the eight irregularity intervals (see Table 4.2). The eight brightness and eight irregularity interval tones were presented in random order. After the stimuli for all instruments were presented, the same procedure was repeated for eight LRT and eight logarithmically spaced n-values for each of the instruments (see Table 4.3 and Table 4.5 respectively). Each test tone were only loaded and matched once.

4.3.2 Results

Caclin et al. (2005) loudness balanced stimuli according to its brightness. The amplitude of the test tone, T_i, was adjusted as per

$$20\log\left(\frac{A(T_i)}{A(T_o)}\right) = C\big(T_b(T_i) - T_b(T_o)\big) \qquad (4.1)$$

with $C = -1.9$, to approximately match the loudness of the reference tone, T_o. This resulted in a -1.9 dBSPL per change in brightness unit. This equation was used as the basis for all loudness balancing adjustments.

For each of the four parameters balanced (T_b, IRR, LRT, S/D), the responses of six participants were averaged for each of the instruments. Most of the preliminary one-way analyses of variance (ANOVAs) of the eight intensity adjustments of each of the six listeners for each instrument and each spectral parameter (T_b, IRR) indicated that the H_0 hypothesis ($H_0: \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_8$) should be rejected ($p < 0.05$). Post-hoc tests using Tukey's honestly significant difference revealed that differences in means occurred mainly at the extreme values suggesting upward or downward trends. Multiple one-way ANOVAs were used instead of two-way ANOVAs, since the parameter values across instruments were not equal (see Table 4.1, Table 4.2, Table 4.3, Table 4.5, Figure 4.1, Figure 4.2, Figure 4.7 and Figure 4.8) and parameter interactions were not of interest at this stage. Post-hoc ANOVA tests revealed that significant differences in listener response means occurred between the extreme values of the parameters. The preliminary one-way ANOVAs of the temporal parameter LRT and S/D for decay instruments also indicate similar significant differences in mean intensity adjustments. Once again post-hoc tests using Tukey's honestly significant difference revealed differences at the extremes.

Responses were adjusted across instruments in order to compare all responses in one model. Using robust linear regression, a curve was fit to the average response data for each of the instruments and each of the parameters. The means of these were averaged and outliers were tested for by defining the outlier boundaries as three times the standard deviation. In addition, linear regression was performed on the response data. Linear regression performed on the responses for all instruments also allowed the appropriate gradients for each of the parameters to be found as well as 95% confidence intervals of the gradients. The variance of the model was also found. The intercepts of the lines with the axes were not of interest.

Although an analysis of covariance (ANCOVA) indicate significant interaction effects between T_b and instrument ($F(12, 598) = 16.12$, $p < 0.05$) for loudness adjustments, regression lines provide a good fit through data points pooled across instruments. Using the mean of the gradient, loudness adjustments from six participants show that tones were adjusted on average by approximately 1 dBSPL per unit of brightness (gradient mean: 1.0 dBSPL, standard deviation: 0.31 dBSPL). No outliers were present. Linear regression (Table 4.6) indicates a gradient of -0.87 dBSPL per unit of brightness and an error standard deviation of 0.87. Using the equation proposed by Caclin et al. (2005), all tones could be adjusted using a coefficient of $C = -1$ or $C = -0.87$ in equation (4.1) according to Figure 4.1. The data in Figure 4.1(a) shows every individual response while Figure 4.1(b) are the averages across listeners for each of the eight increments of the thirteen instruments. Due to the range of values of Table 4.1, the data does not line up.

Table 4.6. Linear regression information from fitting a line segment to the loudness response data when changing brightness.

Gradient [95% confidence intervals]	**R^2**	**F-statistic**	**p-value**	**Error variance**
-0.873 [-0.806, -0.940]	0.868	F(1, 102)=670	1.3×10^{-46}	0.754

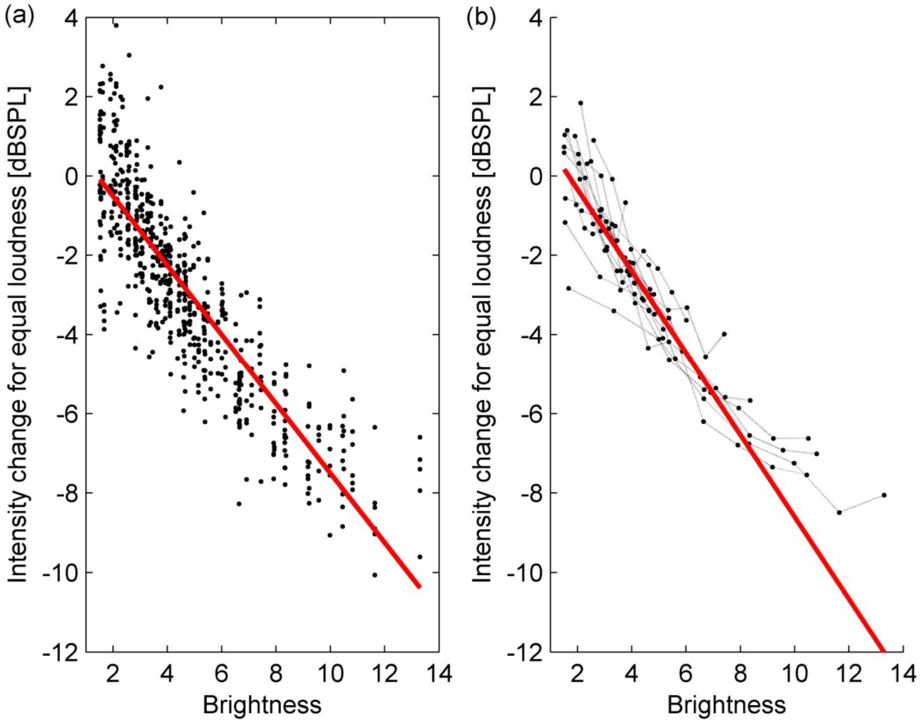

Figure 4.1. Each dot in (a) represents an individual response of one of the six participants for eight values of brightness for each of the thirteen instruments. The gradient found using linear regression of the entire dataset is indicated in red. The average across the participant responses for eight values of brightness for each of the thirteen instruments are shown in (b). The gradient for a set of data associated with an instrument was found using robust linear regression and the average of the gradients is presented by the red line.

It is observed that loudness does not increase indefinitely towards higher brightness values, but a ceiling effect is observed. Adjustments were made for brightness values up to 8, after which no further adjustment was made. This step is especially critical during the 2AFC procedure for obtaining JNDs to prevent the test tones from becoming too soft or inaudible and influencing responses.

Using the same methods for changes in irregularity show no significant interaction effects between IRR and instrument (F(12, 598) = 1.12, p > 0.05) for loudness adjustments. Adjustments of approximately 1.5 dBSPL per unit of irregularity (gradient mean: 1.5 dBSPL, standard deviation: 0.44 dBSPL) were made, resulting in an adjustment equation of $20\log\left(\frac{A(T_i)}{A(T_o)}\right) = C\big(IRR(T_i) - IRR(T_o)\big)$, with $C = -1.5$ according to Figure 4.2. The data in Figure 4.2(a) and (b) are respectively the individual intensity adjustments and the averages across listeners for each of the eight increments of the thirteen instruments. Due to the chosen increment values indicated in Table 4.2, the irregularity values do not line up. No outliers were present. Linear regression (Table 4.7) indicates a gradient of -1.5 dBSPL per unit of irregularity, with an error standard deviation of 0.36 dBSPL. This corresponds to $C = -1.5$.

Table 4.7. Linear regression information from fitting a line segment to the loudness response data when changing irregularity.

Gradient [95% confidence intervals]	R^2	F-statistic	p-value	Error variance
-1.46 [1.29, 1.64]	0.727	F(1, 102)=272	1.6×10^{-30}	0.132

Changes in the LRT for sustained instruments yield different results compared to decaying instruments. Figure 4.3 shows that as the LRTs increased (slower attack), sustained instruments were perceived as decreasing in loudness, but for decaying instruments, increasing LRTs were perceived as louder tones.

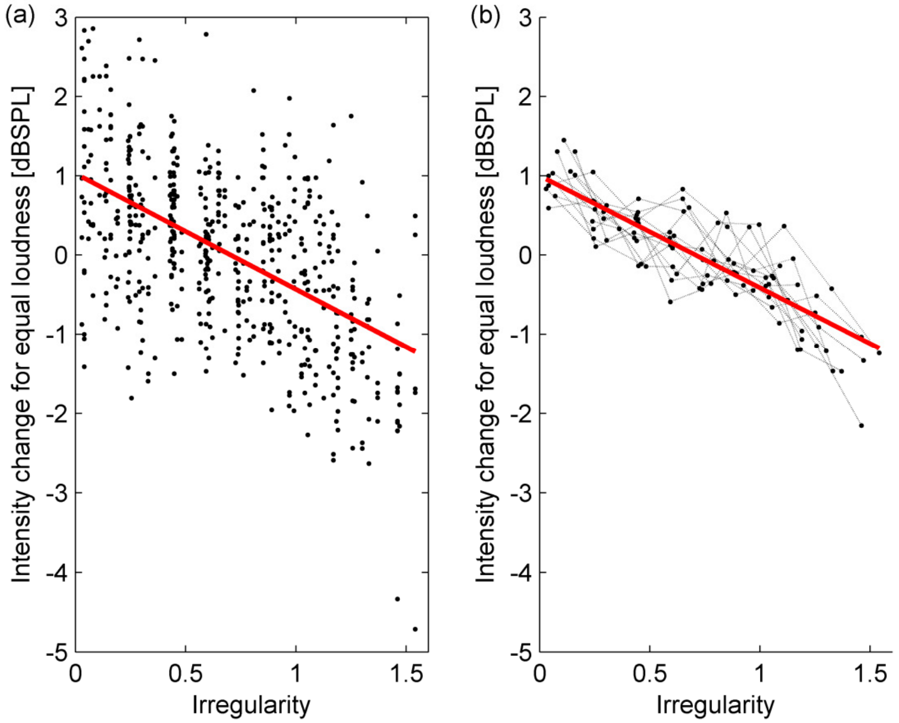

Figure 4.2. The left panel (a) shows every individual response for each of the six participants for eight values of irregularity for each of the thirteen instruments. The gradient found using linear regression is used to construct the red line segment. The panel on the right (b) shows the average response across the six participants for eight values of irregularity for each of the thirteen instruments. The average of the gradients is shown in red.

The average of the gradient means suggests adjustments for sustained instruments should be made as per $20\log\left(\frac{A(T_i)}{A(T_o)}\right) = C\big(LRT(T_i) - LRT(T_o)\big)$ with coefficient $C = 0.42$ (gradient mean: 0.42 dBSPL, standard deviation: 0.23 dBSPL). Linear regression (Table 4.8) indicates a gradient of 0.5 dBSPL per unit of LRT, with an error standard deviation of 0.26, corresponding to $C = 0.5$. Standard deviations are large compared to the differences in means.

Since no significant interaction effects are observed (ANCOVA) between LRT and loudness adjustments for decaying instruments (F(2, 138) = 0.35, p > 0.05), adjustments are made as per $20\log\left(\frac{A(T_i)}{A(T_o)}\right) = C\big(LRT(T_i) - LRT(T_o)\big)$ with $C = -0.97$ (gradient mean: 0.97 dBSPL, standard deviation: 0.18 dBSPL). Linear regression (Table 4.8) indicates a gradient of 0.98 dBSPL per unit of LRT and an error standard deviation of 0.37 dBSPL. This corresponds to adjustments with $C = -0.98$ (see Figure 4.4 and Figure 4.5). The panels on the right-hand side are the averages across listeners for each of the eight increments of respectively the ten sustain and three decay instruments. The panels on the left-hand side show each individual response for the ten sustain (Figure 4.4) and three decay instruments (Figure 4.5).

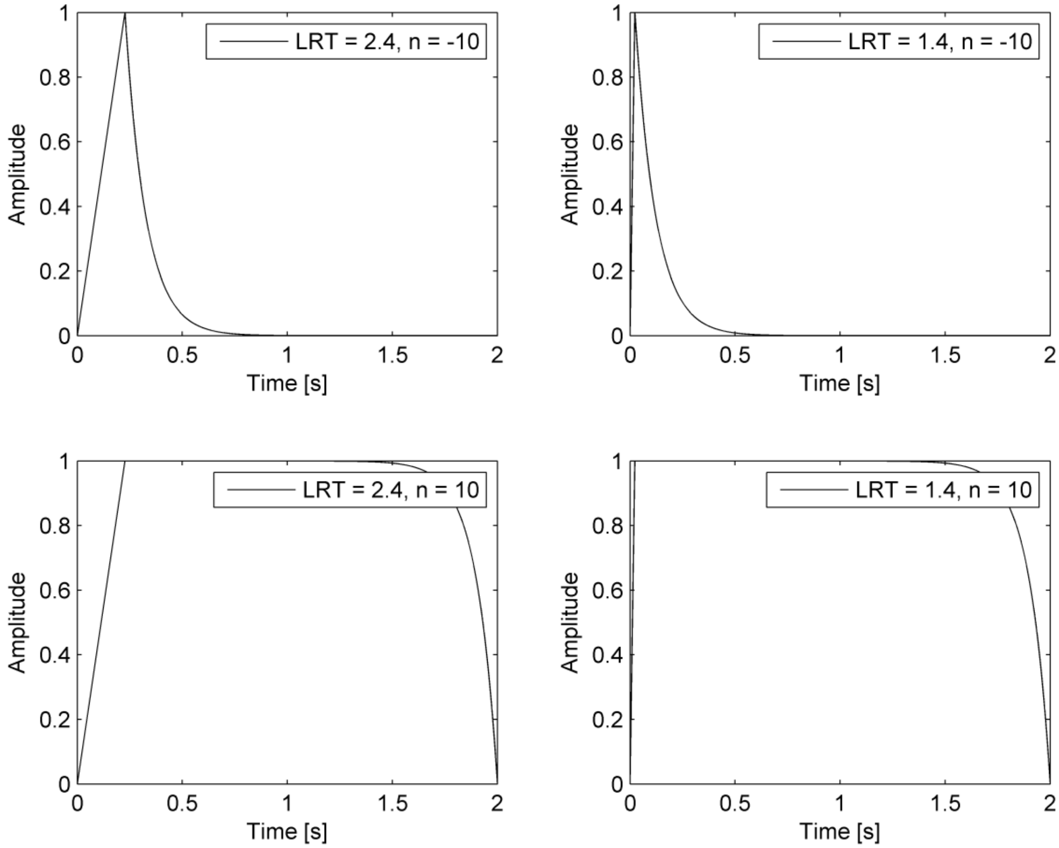

Figure 4.3. Examination of the temporal envelope gives an indication of why loudness changes as the temporal envelope changes. For decaying instruments an increase in LRT yields a perceptually longer duration tone, since the decay rate is fixed, which may also result in a perceptually louder tone. For sustained instruments, an increase in LRT yields a perceptually shorter duration tone, which may result in a perceptually quieter tone.

Table 4.8. Linear regression information from fitting a line segment to the loudness response data when changing LRT for sustain (first pair of rows) and decay (second pair of rows).

Gradient [95% confidence intervals]	**R^2**	**F-statistic**	**p-value**	**Error variance**
0.498 [0.588, 0.409]	0.613	F(1, 78)=124	9.3×10^{-18}	0.0688
Gradient [95% confidence intervals]	**R^2**		**p-value**	**Error variance**
-0.984 [-0.743, -1.23]	0.764	F(1, 22)=71.4	2.4×10^{-8}	0.139

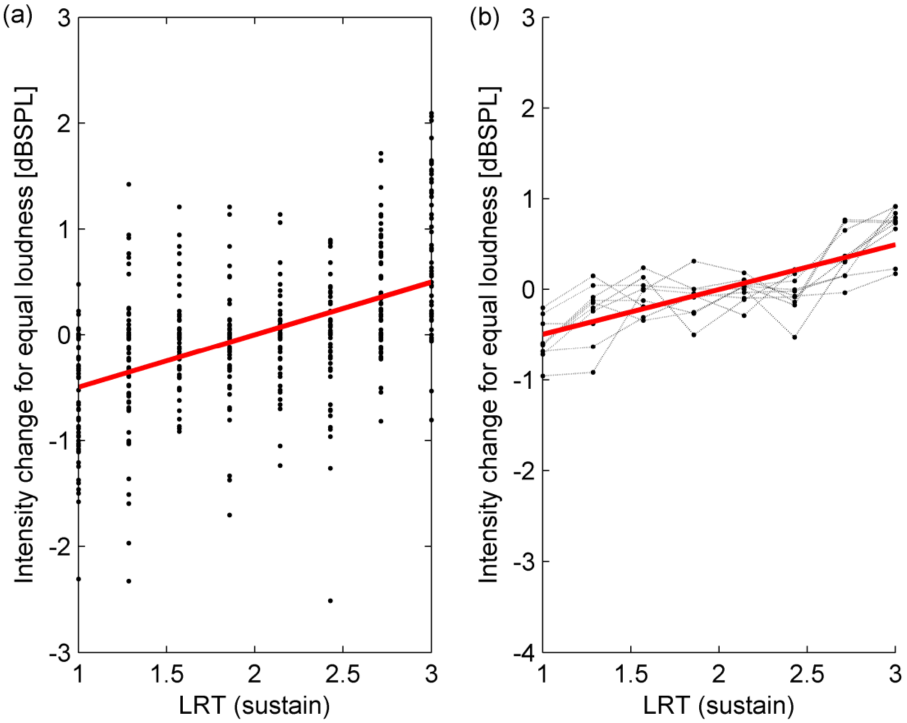

Figure 4.4. This figure is applicable only to sustain type instruments. Every individual responses of each of the six participants for eight values of LRT for each of the ten sustain instruments are represented by a dot in (a). The red line was constructed using results obtained during linear regression. The average of the intensity adjustments of the six participants for eight values of LRT for the ten sustain instruments is shown in (b). The average of the gradients of the ten data sets in (b) is found using robust linear regression and is indicated by the red line.

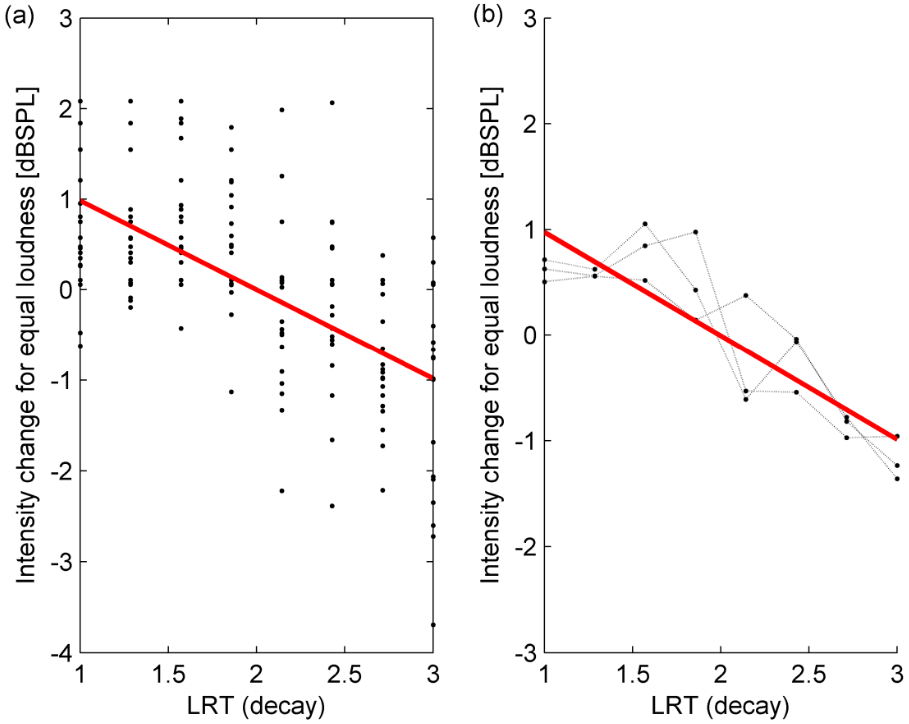

Figure 4.5. This figure is applicable to the three decay instruments. Each dot in panel (a) shows the individual responses of each of the six participants for eight values of n for each of the three decay instruments. Linear regression shows the gradient of the entire dataset in red. Panel (b) displays the average balancing responses across participants for eight values of LRT. In panel (b), the gradient for each set of data is found using robust linear regression. The average of the gradients found is presented by the red line.

Similar to the changes in LRT, changes in the n-value for sustained instruments yield different results compared to decaying instruments. Figure 4.6 shows that as the absolute value of n increased (more abrupt decays), sustained instruments were perceived as increasing in loudness, while decaying instruments were perceived as decreasing in loudness.

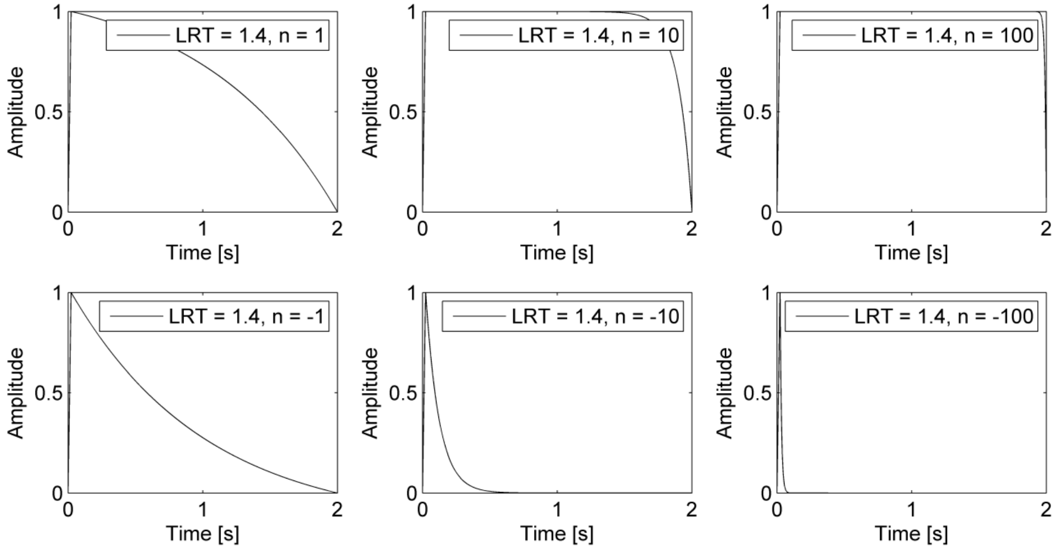

Figure 4.6. Examination of the temporal envelope gives an indication of why loudness changes for varying n-values. For decaying instruments, an increase in the absolute value of n yields a perceptually shorter duration tone, which results in a perceptually quieter tone. For sustained instruments, an increase in the absolute value of n yields a perceptually longer duration tone, which results in a perceptually louder tone.

Although significant differences in psychometric function slopes were observed (F(9, 460) = 3.71, p < 0.05), adjustments for sustained instruments are made as per $20\log\left(\frac{A(T_i)}{A(T_o)}\right) = C(\log(n(T_i)) - \log(n(T_o)))$ with $C = -0.17$ (gradient mean: -0.17 dBSPL, standard deviation: 0.15 dBSPL). Linear regression (Table 4.9) indicates a gradient of -0.088 dBSPL per unit of $\log(n)$ and an error standard deviation 0.14 dBSPL. This corresponds to adjustments with $C = 0.088$. Standard deviations are large compared to the differences in means.

For decaying instruments, an ANCOVA also show significant slope differences (F(2, 138) = 18.6, p < 0.05), but adjustments are made as per $20\log\left(\frac{A(T_i)}{A(T_o)}\right) = C(\log|n(T_i)| - \log|n(T_o)|)$ with $C = 2.8$ (gradient mean: -2.8 dBSPL, standard deviation: 0.46 dBSPL) due to good regression fits. Linear regression (Table 4.9) indicates a gradient of -2.3 dBSPL per unit of $\log(n)$ and an error standard deviation 0.52 dBSPL. This corresponds to adjustments of with $C = 2.3$ (Figure 4.7 and Figure 4.8). It is clear from Table 4.5 why the values are spread across the log(n) axis. The leftmost data of Figure 4.7 and Figure 4.8

respectively shows the individual responses for the ten sustain instruments and the three decay instruments. The rightmost panels of each of these figures show the averages across listeners for each of the eight increments of respectively the ten sustain and three decay instruments.

Table 4.9. Linear regression information is found by fitting a line segment to the loudness response data when changing n for sustain instruments (first pair of rows) and decay (second pair of rows) instruments.

Gradient [95% confidence intervals]	**R^2**	**F-statistic**	**p-value**	**Error variance**
-0.0876 [-0.0248, -0.150]	0.090	F(1, 78)=7.71	6.9×10^{-3}	0.0183
Gradient [95% confidence intervals]	**R^2**		**p-value**	**Error variance**
-2.34 [-1.83, -2.86]	0.804	F(1, 22)=90.2	3.0×10^{-9}	0.269

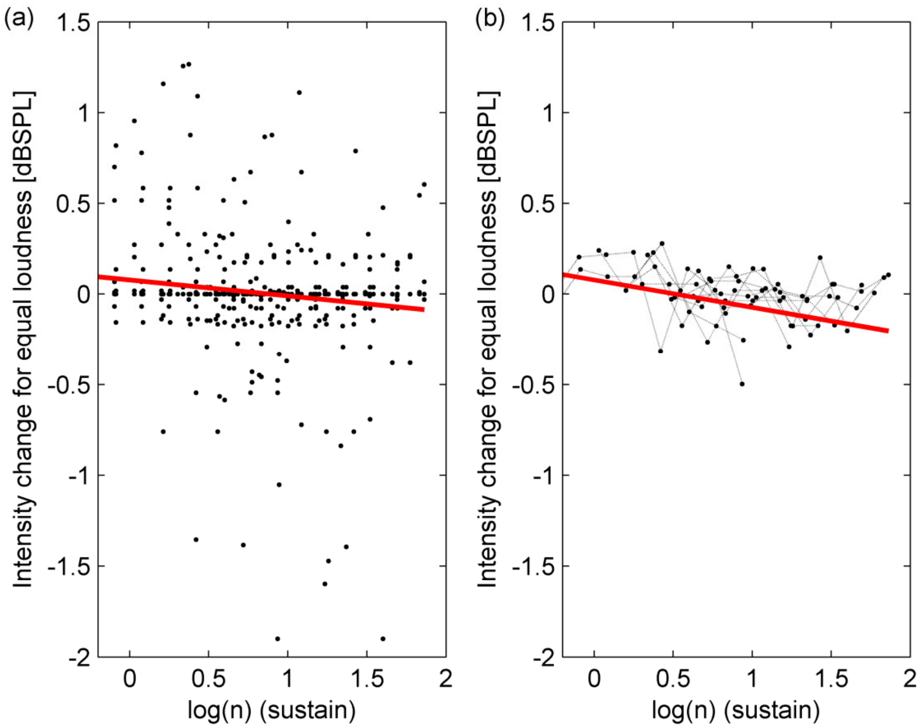

Figure 4.7. This figure is applicable to the ten sustain instruments. The dots on the left (a) are every individual response for each of the six participants and eight values of n for each of the instruments. The red line shows the gradient of the entire dataset obtained through linear regression. The data on the right are the averages of the adjustments made by the participants for each of the eight increments for each instrument. The gradient for each dataset associated with a particular instrument was found using robust linear regression and the average of these gradients is represented by the red line.

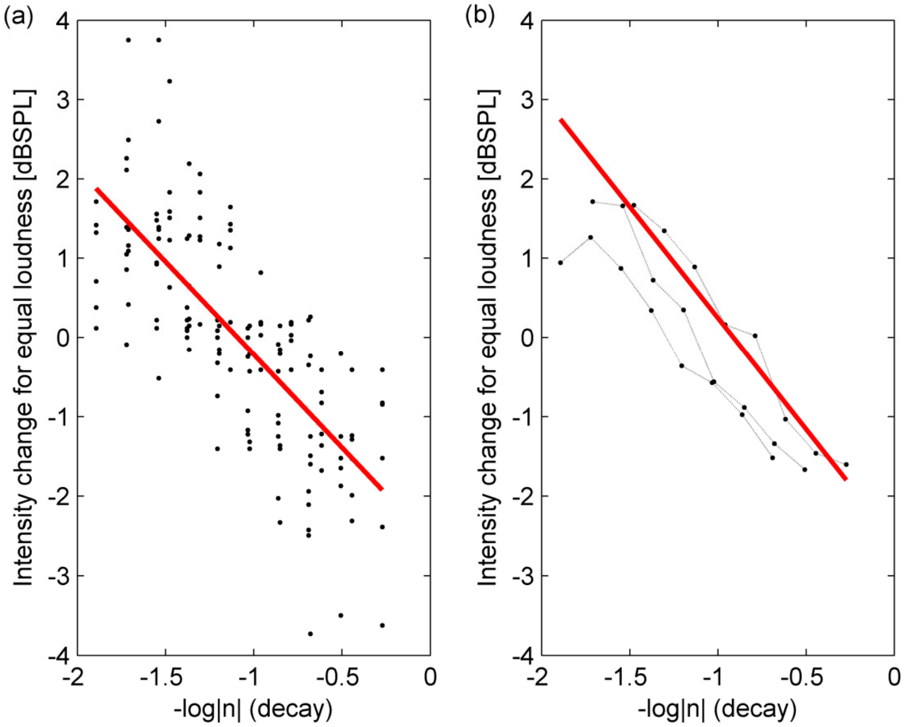

Figure 4.8. The data in this figure is applicable to the three decay instruments. Every data point in the leftmost panel (a) contains individual responses of each of the six participants for eight values of n and for each of the instruments. The The gradient of the red line was obtained through linear regression of the response data. The data on the right are the averages of responses across participants for each of the eight n-values for each instrument. The gradient of the average responses across n-values for each instrument was found using robust linear regression and the average of these gradients is represented by the red line.

Table 4.10 contains the values (C_{param}) in dBSPL per unit of timbre parameter used for adjustments.

Table 4.10. Coefficients used for adjustment of each of the parameters.

$\boldsymbol{param}$	T_b	IRR	LRT (sustain)	LRT (decay)	Log(n) (sustain)	Log|(n)| (decay)
$\boldsymbol{C_{param}}$	-1	-1.5	0.5	-1	0	2.5

$$A(T_i) = A(T_r) \times 10^{(C_{param}/20)\times(param(T_i)-param(T_r))} \tag{4.2}$$

$A(T_r)$ is the amplitude of the reference tone and $param(T_i)$ and $param(T_r)$ the timbre parameters of the test and reference tone respectively. For example, comparing a test tone with brightness $T_b = 6$ with a reference tone with brightness $T_b = 3$ and normalised amplitude requires that the amplitude of the test tone be $A(T_i) = 1 \times 10^{(-1/20)\times(6-3)} \approx 0.708$. The value found for T_b using linear regression suffered from a large standard deviation. The gradient value is approximately half of that used by Caclin et al. (2005). The loudness was adjusted as -1 dBSPL per T_b unit. Changes in loudness for log(n) for sustain instruments showed a very small gradient with a large standard deviation.

4.4 PERCEIVED DURATION BALANCING

4.4.1 Procedure

The stimuli had to be balanced for perceived duration. The experimental procedure was similar to that of the loudness balancing procedure. Each test tone had the same brightness and irregularity values as the reference tone. The original total duration of the test tone was set equal to the total duration of the reference tone (2 s). A slider bar controlled the test tone’s total duration, which could vary between +0.5 s and -0.5 s of the duration of the reference tone, i.e. each test tone was adjustable between 1.5 s and 2.5 s. The slider step sizes were arranged linearly on the time scale. Clicking on the reference and test tone buttons presented the sounds and participants were allowed to compare the test and reference tone as many times as they liked to attempt to match the perceived duration of the two tones as closely as possible. After the participant was satisfied that the test tone perceived duration was equal to that of the reference, the result could be saved and the next tone pair was loaded.

The test involving each instrument was presented sequentially. The eight LRT intervals were presented first, followed by the eight logarithmically spaced n-value intervals (see Table 4.4 and Table 4.5 respectively). Within the S/D and LRT presentations, the changes in values were presented in random order. Only one matching was done for each of the test tones.

4.4.2 Results

Changes in the LRT for sustained instruments yielded different results compared to decaying instruments. Figure 4.3 shows that as the LRT increased (slower attack), sustained instruments were perceived as shorter in duration, but for decaying instruments, an increasing LRT was perceived as longer in duration.

Similar to the changes in LRT, changes in the n-value for sustained instruments yielded different results compared to decaying instruments. Figure 4.6 shows that as the absolute value of n increased (more abrupt decays), sustained instruments were perceived as longer in duration, while decaying instruments were perceived as decreasing in duration.

It should be noted that for decay instruments, adjustment of the duration of decay type instruments is impossible, due to the nature of the sound. Sustained instruments, such as violins and clarinets, can be adjusted in duration. A player can produce short or long notes. A struck or plucked string, such as a piano or violin, always has the same perceived duration. An increase in the duration of the note (such as depressing the pedal, which lifts the piano mutes from the strings) yields a tone of longer duration, but it invariably changes the decay rate of the note (changes the n-value). These tones could therefore not be balanced for perceived duration, since the perceived duration is in actual fact one of the temporal timbre parameters.

The mean of the gradients shows that changes in the LRT of sustain instruments appear to create perceptible duration changes of 26 ms per unit LRT. Linear regression (Table 4.11) indicated a value of 30 ms per unit LRT (standard deviation: 37 ms) (see Figure 4.9). The left panel of Figure 4.9 show each listener's individual response for each of the eight increments for the ten sustain instruments, while the right panel shows the averages of listener responses for each of the eight instruments for the ten sustain instruments.

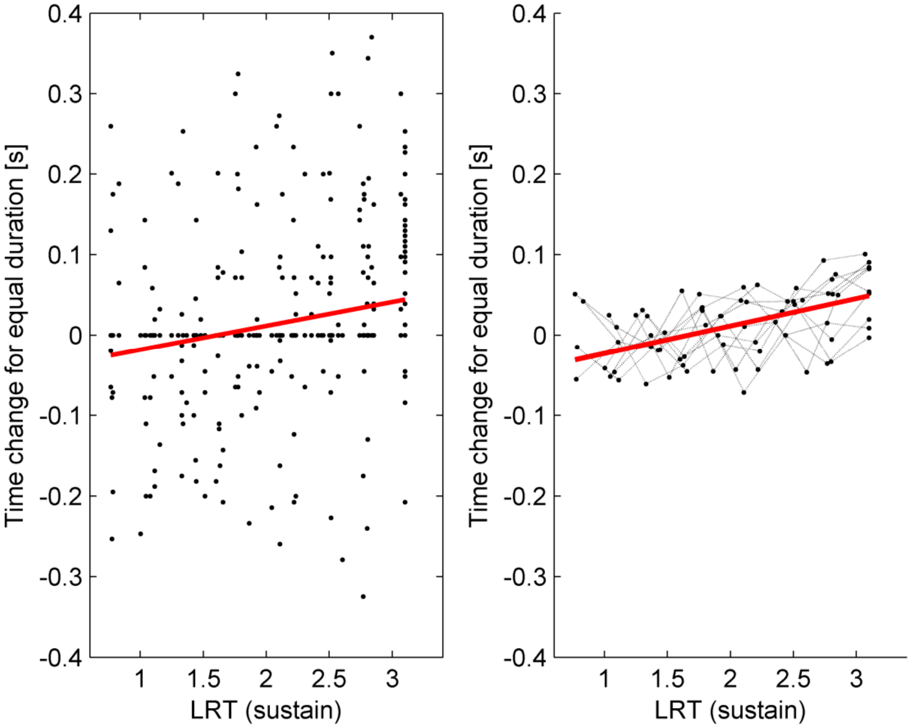

Figure 4.9. The data in this figure was acquired for sustain instruments. Individual responses of each of the six participants for eight values of LRT for each of the ten sustain instruments are shown as a data point in (a). The gradient obtained through linear regression is shown in red. The average responses of perception of six participants for eight values of LRT for each of the sustain instruments is shown in (b). The gradient for each set of data is found using robust linear regression and is represented by the red line segment.

Table 4.11. Linear regression information from fitting a line segment to the perceived duration response data when changing LRT.

Gradient [95% confidence intervals]	**R^2**	**F-statistic**	**p-value**	**Error variance**
-0.0295 [-0.0413, -0.0177]	0.241	F(1, 78)=24.75	3.80×10^{-6}	0.0014

Changes in individual gradients for n-values do not appear to follow any specific trend. Linear regression (Table 4.12) indicates no gradient and yields a very small R^2 value, indicating a bad linear fit and a large variance compared to the gradient (see Figure 4.10). Figure 4.10 shows each listener’s individual response for each of the eight instruments for the ten sustain instruments and the average responses across listeners for each of the eight instruments for the ten sustain instruments on respectively the left and right panels.

Table 4.12. Linear regression information from fitting a line segment to the loudness response data when changing n.

Gradient [95% confidence intervals]	R^2	F-statistic	p-value	Error variance
0.000 [-0.0005, 0.0006]	4.26×10^{-6}	F(1, 78)=0.033	0.856	0.0015

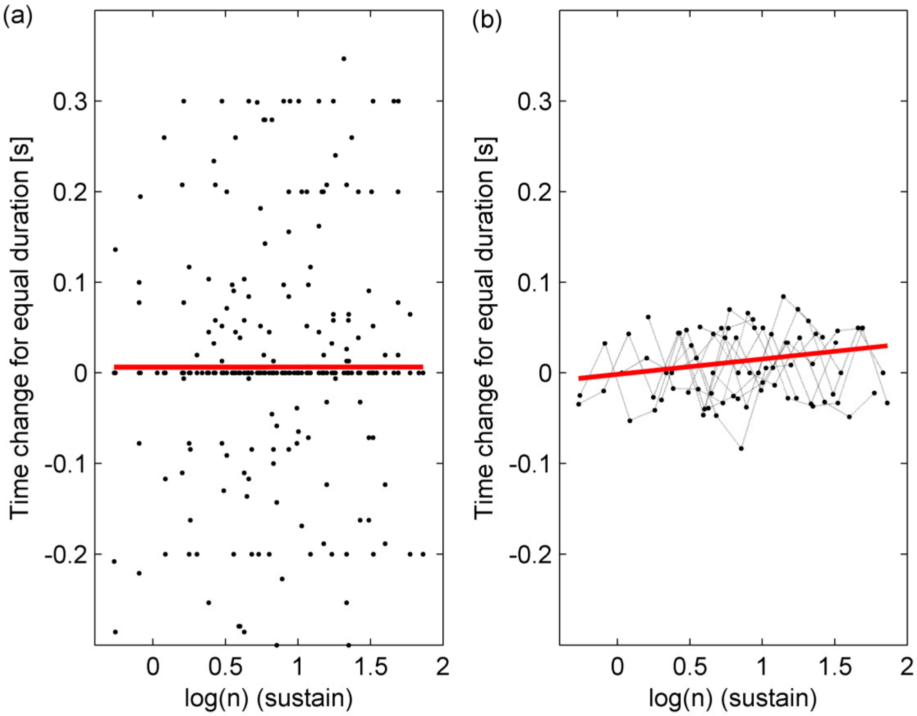

Figure 4.10. The data in this figure was acquired for decay instruments. Individual balancing adjustments from the six participants are shown in (a).Each dot represents a single adjustment by a participant. The red line indicates the gradients of the dataset and was obtained using linear regression. Panel (b) displays the average of the adjustments of the six participants. The average of the gradients for the responses associated with a particular instrument is found using robust linear regression and these average gradients is represented by the red line. The gradients do not follow any specific trend and adjustments fall within 0.1 s

The large standard deviation could be due to the fact that the stimuli used here were 2 s in duration. Due to the long duration of the tone, changes in duration that might have been more perceptible with shorter tones may be difficult to perceive. Caclin et al. (2005) used a linear time adaption for their stimuli, which range between 615 ms and 800 ms. For longer sound durations, it was found (Abel, 1972) that the discrimination ΔT between two tones of duration T and T+ΔT increased as T increased. At approximately T = 1 s, the discrimination ability was $\Delta T = 50$ ms. The results of the perceived duration balancing task appear to point towards the fact that perceived duration is probably not used as a cue for stimuli with a total duration of 2 s.

4.5 PITCH BALANCING

4.5.1 Procedure

The stimuli had to be balanced for pitch. As changes in timbre parameters T_b and IRR occur, tones with equal frequencies may be perceived as being different in pitch. The experimental procedure was controlled by MATLAB software. Clicking on the reference and test sound buttons presented each of the tones. Once again, the participant was allowed to compare the test tone and reference tone as many times as needed to perform the task. The original intensity of the test tone was set equal to the intensity of the 262 Hz reference tone. A slider bar controlled the test tone pitch, which was originally set randomly between approximately 255 Hz and 269 Hz and this could then be adjusted by approximately -10 Hz to 10 Hz. The slider step sizes were arranged linearly on the frequency scale. After the participant was satisfied that the test tone pitch was equal to the reference tone, the result could be saved. The next tone pair was loaded and could be listened to by clicking on the reference and test tone buttons.

The test involving each instrument was presented sequentially, with the eight brightness intervals being presented first, followed by the eight irregularity intervals (see Table 4.1 and Table 4.2 respectively). Within the brightness and irregularity tests, the changes in brightness and irregularity were presented in random order. Only one matching for each test tone was done.

4.5.2 Results

Singh and Hirsh (1992) found that changes in spectral locus often influenced perceived pitch. The spectral locus changes in the study were large – at least one unit of brightness for every trial. However, when the fundamental frequency difference was increased past approximately 4 Hz, the pitch and timbre could be differentiated from each other. Russo and Thompson (2005) found strong influences of brightness on perceived interval size, although only a very large change in brightness was considered. Vurma and Ross (2007) found that timbre matching tasks for piano, oboe and voice yielded a range of in-tune ratings near the fundamental frequency. The participants were musically trained vocalists. The frequency ranges where the participants rated two tones as being in tune 75% of the time varied between a few cents to as many as 50 cents, with one cent being a one-hundredth part of a semitone and logarithmically spaced within the semitone. At 220 Hz this relates to a 6.5 Hz band of in-tune ratings. Pitch balancing appears to be a necessary step to eliminating its influence on timbre perception.

Pitch matching tasks using two different complex tones are known to be more difficult than matching tasks using pure tones or two identical complex tones (Hartman, 2005). The results of the pitch balancing task appear to support this view. In general, participants appeared to find the pitch balancing task more difficult than the loudness or perceived duration balancing task. Only two participants were able to perform the task, while other participants' results were not reproducible. 95% of the responses of the two participants fell within a ±0.8 Hz and ±1.5 Hz band respectively for brightness and within a ±1 Hz and ±1.5 Hz band respectively for irregularity as can be seen in Figure 4.11 and Figure 4.12. These values are much smaller than the perceptual discrimination in pure tone frequencies (Shower and Biddulph, 1931; Wier et al., 1977) (between approximately 0.5% and 1% for the sound levels used in this experiment). The responses followed no trend (gradients -0.017 and -0.34, standard deviations 0.71 and 0.61 for T_b and IRR respectively), suggesting that discrimination abilities are limited within the 95% confidence interval band. The discrimination abilities of pure tones at 262 Hz correspond approximately to the 95% confidence interval band. Although the gradient for changes in irregularity is higher, the IRR parameter range is smaller than for brightness. For the gradients found from linear regression, the frequency changes across the entire range for the specific instrument set is 0.21 Hz and 0.47 Hz for brightness and irregularity respectively. Figure 4.13 and Figure 4.14 show the individual balancing responses of the two individuals (NH2 and NH5) who were able to perform the task. Figure 4.13 shows the responses for changes in brightness and Figure 4.14 shows the responses of changes in the irregularity. Figure 4.11 and Figure 4.12 show the average response of the two individuals for respectively brightness and irregularity. Within a ±1.5 Hz band, responses appeared random. It was therefore deduced that although pitch may change with change in timbre, the changes are more salient when timbre changes are much larger. For these experiments it was therefore deemed unnecessary to balance pitch.

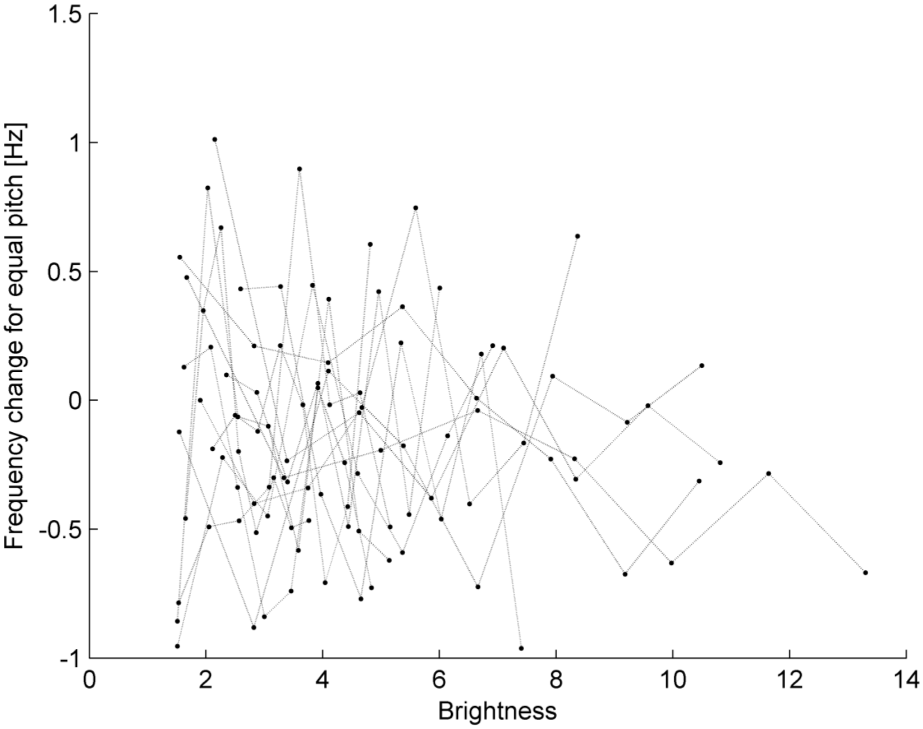

Figure 4.11. Average responses of perception of two participants who were able to perform the pitch balancing task are shown for eight values of brightness for each of the instruments. The gradient for each set of data is found using robust linear regression. The gradients do not follow any specific trend and the data fall within a 2 Hz band around the fundamental frequency.

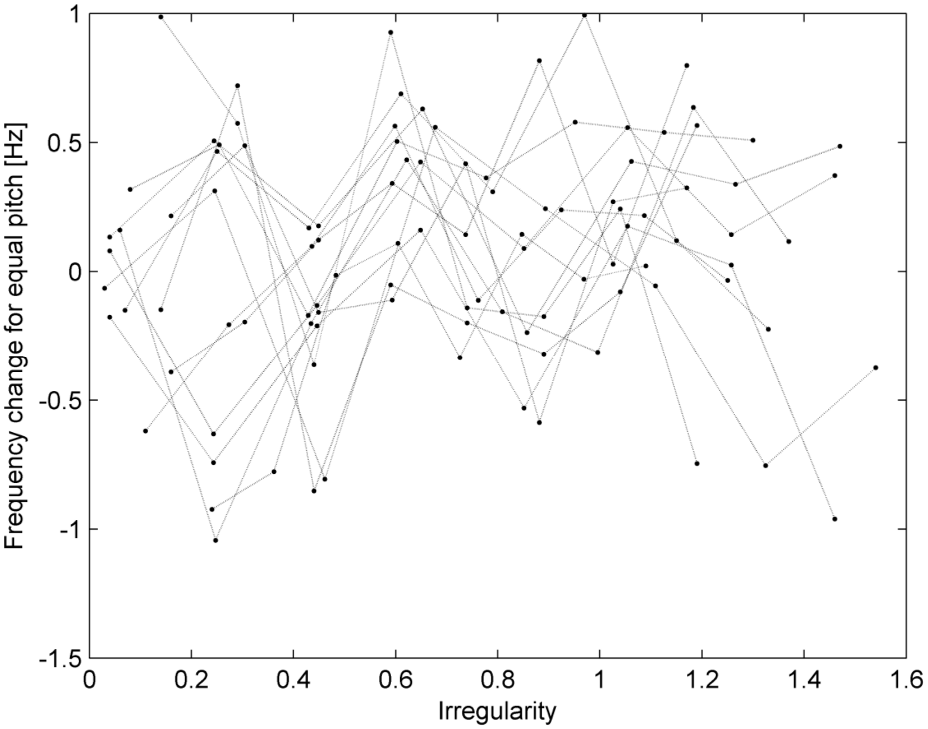

Figure 4.12. Average responses of perception of two participants who were able to perform the pitch balancing task are shown for eight values of irregularity for each of the instruments. The gradient for each set of data is found using robust linear regression. The gradients do not follow any specific trend and the data fall within a 2 Hz band around the fundamental frequency.

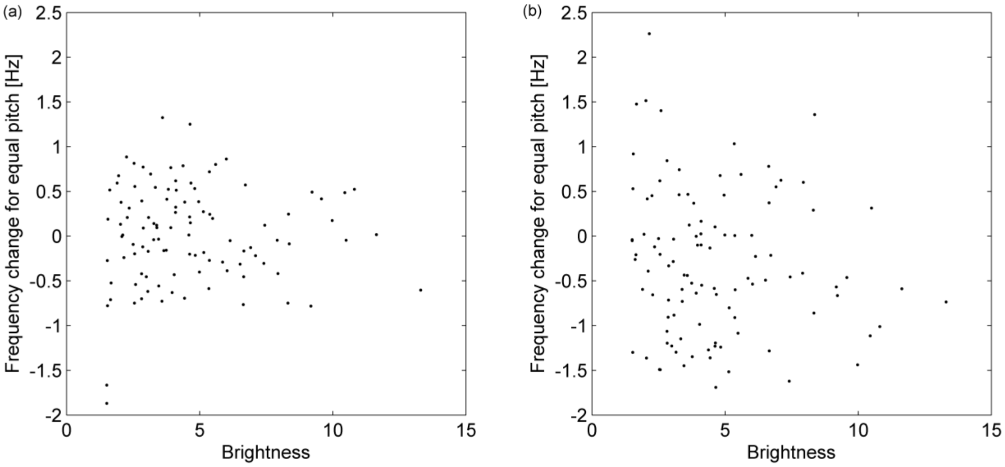

Figure 4.13. Individual responses for the two participants who were able to perform the pitch balancing task for eight changes in brightness for each of the instruments. The data on the left (a) and right (b) respectively correspond to the responses of NH2 and NH5. Responses appear random within a ±2 Hz band around the reference frequency.

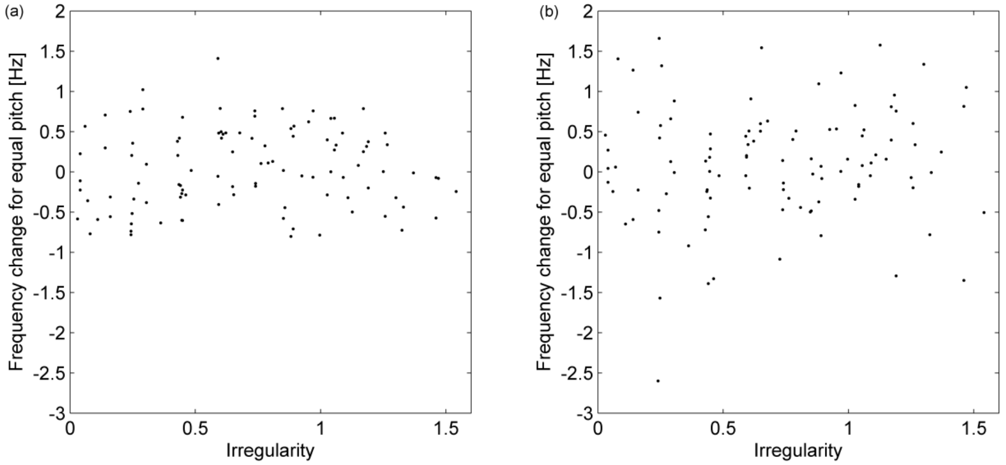

Figure 4.14. Individual responses for the two participants who were able to perform the pitch balancing task for eight changes in irregularity for each of the instruments. The data on the left (a) and right (b) respectively correspond to the responses of NH2 and NH5. Responses appear approximately random within a ± 1.5 Hz band around the reference frequency.

After large variances were observed, the participants whose results were not used here anecdotally reported difficulty with the task upon questioning. Even if their results were to be considered, the responses followed no trend and comparisons between participants also indicated no consistency. This could also be interpreted to mean that the participants' discrimination abilities were much worse than the two participants reported here. In such an unlikely event, if no changes were made to frequency, listeners would not be able to use pitch as a timbral cue. Caclin et al. (2005) gave no indication on how pitch is corrected for when timbre changes. The results appear to indicate that, in general, it is unnecessary to balance for pitch.

4.6 BALANCING PROCEDURE SUMMARY

In sections 4.3, 4.4 and 4.5 the balancing responses for loudness, perceived duration and pitch were collected during experimental procedures. Statistical analyses showed that for timbre changes in certain balancing procedures, trends with small standard deviations were observed, indicating consistency in responses across listeners. This consistency allows the construction of balancing equations applicable to the average listener.

From the results, only loudness balancing appeared necessary. The values of Table 4.10 used in equation (4.2) allow loudness balancing of tones for changes in timbre parameters. Trends were observed for perceived duration balancing responses, but a study of the literature indicated that the small duration changes associated with these trends would probably not be accessible as a cue for the relatively long duration tones used here. The pitch balancing responses of the two listeners who were able to perform the pitch balancing task did not indicate specific balancing trends and responses appeared random within pure tone discrimination thresholds. It was therefore not deemed necessary to balance pitch.

The constructed balancing equation (equation (4.2)) is useful in an adaptive 2AFC procedure. A simple calculation is needed to match the loudness of the synthesized test tone to the synthesized reference tone. Applying the balancing equation to synthesized tones eliminates possible non-timbre cues and allows discrimination experiments of individual timbre properties to proceed.

CHAPTER 5 METHODS III: DISCRIMINATION EXPERIMENTS

5.1 PARTICIPANTS

Six NH listeners participated in the discrimination experiments. Three of the participants from the balancing procedures were used in the discrimination experiment. All participants had NH (pure tone thresholds ≤ 20 dB HTL for 250 Hz, 500 Hz, 1000 Hz, 2000 Hz, 4000 Hz and 8000 Hz). The four females and two males had an average age of 25, with all participants aged between 24 and 28.

The average age of three female and three male CI participants was 31, with all participants aged between 21 and 58. Table 5.1 gives CI user demographic information. Where the CI user was implanted bilaterally, the information of the right ear is given first, followed by the left ear information.

5.2 STIMULI

For NH listeners, the thirteen instruments indicated in Table 3.1 were used.

To reduce the amount of testing time for CI participants, the instrument set was reduced from thirteen to nine instruments. The instrument set reduction procedure and results can be found in Appendix section C. Essentially, the reduction method uses the JNDs of NH listeners to define Gaussian distributed areas of confusion in the four-dimensional timbre space. The four-dimensional timbre space is scaled in order to see which instruments are perceptually close to one another.

All tones were presented between 68.2 dBSPL and 80 dBSPL. All tones were sampled at 44.1 kHz and were presented in a soundproof booth through a KEF Q30 loudspeaker using an M-Audio Fasttrack Pro external soundcard.

Table 5.1. CI participants' information regarding age, gender, implantation and deafness.

Participant number	Gender	Age	Processor	Implant	Strategy	Post-/Pre-lingual deafness	Nr of years implanted	Ear(s) implanted
S15	F	23	CP810, Freedom	Nucleus 24 Contour Advance, Nucleus 22 Series	ACE, SPEAK	Post	7, 19	Right, left
S24	F	21	Freedom	Freedom Contour Advance	ACE	Post	5	Right
S22	M	41	CP810	Freedom Contour Advance	ACE	Post	5	Right
S23	M	21	ESPrit 3G	Nucleus 22 Series	SPEAK	Pre	15	Right
S28	F	58	Freedom	Freedom Contour Advance	ACE	Post	5	Right
S26	M	22	CP810	Nucleus 24 Contour	ACE	Pre	9	Left

5.3 PROCEDURE

5.3.1 Brightness

Before commencement of the brightness discrimination test, a training session had to be completed. The session consisted of verbal explanations of the concept of brightness as well as examples with large differences in brightness. The listener was then presented with two repetitions of five pairs of tones and had to choose the brightest tone. A score of 9/10 for NH listeners and 8/10 for CI listeners was required in order to commence the brightness discrimination tests. The concept of brightness appeared intuitive and all NH participants scored at least 9/10 on their first attempt, while 5/6 CI participants scored at least 8/10 on their first attempt.

A 1-up, 2-down, 2AFC procedure was used to determine the JNDs of the brightness. For each instrument, a reference tone was created using the instruments' analysed parameters. For the test tones, the initial upper and lower brightness starting points were set at respectively twice the instrument's brightness and half the instrument's brightness and within model limits. A pair of tones consisting of the reference tone and test tone was presented in random order for each choice. The software matched the loudness of the test tones to the reference using equation (4.2) and Table 4.10. The order of the instruments was randomised for each listener to prevent the influence of practice and tiredness effects on the data. Since brightness is a perceptually simple parameter, the participants were asked to choose whether tone one or tone two was the brighter tone. If the participant chose correctly twice in a row, the difference in brightness between the reference and test tone was decreased with a factor of 1.4, but an incorrect choice immediately increased the difference with a factor of 1.4. The procedure was repeated for twelve reversals for each instrument, where a reversal is a change from increasing the parameter difference to decreasing the parameter difference or vice versa. The mean of the final eight reversal extremities was taken as the JND. No feedback was given as to the correct or incorrect nature of the participant's choice, but a bar indicated progress as the test proceeded.

5.3.2 Irregularity

Irregularity differs from brightness in that it is not a perceptually simple parameter. Listeners easily understand the concept of brightness and can determine which is the brighter of two stimuli. Irregularity is not as perceptually simple and verbal explanations and adjectives could not accurately supply information regarding tone quality. Listeners can identify that two tones with two different irregularities do not sound the same, but have difficulty in explaining how they are different. The irregularity discrimination test therefore consisted of a reference tone and two test tones. One of the test tones was the same as the reference, while the other differed in irregularity while keeping all other timbre parameters constant. The listener had to choose the tone that was different from the reference tone.

As with the brightness discrimination experiment, the irregularity discrimination experiment was preceded by training. A verbal explanation of what was expected of the listener and examples with large differences in irregularities were presented. The listener was then presented with two repetitions of the five pairs of tones and had to choose the tone that was different compared to the reference. A score of 9/10 for NH listeners and 8/10 for CI listeners was required in order to commence the brightness discrimination tests. All NH participants scored at least 9/10 on their first attempt, while 4/6 CI participants scored at least 8/10 on their first attempt.

Similar to the previous experiment, a 1-up, 2-down, 2AFC adaptive procedure with a factor of 1.4 was used to determine the JNDs of the irregularity. The reference tones were synthesized sounds with original instrument timbre parameters. The test tones' upper and lower irregularity starting points were set at twice the instrument's irregularity and half the instrument's irregularity respectively and within model limits. Three tones, consisting of the reference and two test tones (of which one was the same as the reference and the other was different), were presented in random order for each choice. The participants were asked to choose which tone was different compared to the reference, since spectral irregularity is a not a perceptually intuitive parameter and verbal descriptors are difficult to formulate. Loudness matching between the reference and test tones was based on equation (4.2) and Table 4.10 and was controlled by the software. The procedure was repeated for twelve reversals for each instrument, where a reversal is a change from increasing parameter difference to decreasing parameter difference or vice versa. The mean of the final eight reversals was taken as the JND. No feedback was given to the participant during the experiment, except for a bar that indicated the progress of the experiment.

5.3.3 Logarithmic rise-time

Training preceded the LRT discrimination test. Verbal explanations of the RT accompanied sound pair samples with large differences in RT to familiarise the participant with the concept. Two repetitions of five pairs of tones were presented and the participant had to choose the tone with the quickest RT. As before, scores of 9/10 for NH listeners and 8/10 for CI listeners was required in order to commence the LRT discrimination tests. As with brightness, the concept of RTs was intuitively understood and all NH participants scored at least 9/10 on their first attempt, while 5/6 CI participants scored at least 8/10 on their first attempt.

The same experimental procedure followed for the brightness discrimination procedure was followed here (1-up, 2-down, 2AFC adaptive procedure with a factor of 1.4, software controlled loudness matching, random order tone presentation, no feedback except a progress bar and twelve reversals of which the mean of the final eight reversals was taken as the mean). The upper and lower LRT starting points of the test tones were set at respectively four times and a quarter of the original LRT. Limits were imposed at $0 < LRT < 3$, corresponding to RTs of 1 ms and 1000 ms. Since LRT is a perceptually simple parameter, the participants were asked to choose which tone had the quickest attack or fastest RT.

5.3.4 Sustain/decay

As with the previous discrimination experiments, the concept of S/D had to be verbally explained together with examples of sound with large differences in the S/D parameter. The listener was presented with two repetitions of five pairs of tones and had to choose the tone with the quickest or most abrupt decay. A score of 9/10 for NH listeners and 8/10 for CI listeners was required in order to commence the decay discrimination tests. As with brightness and RT, the concept of S/D was intuitively understood and all NH participants scored at least 9/10 on their first attempt, while 5/6 CI participants scored at least 8/10 on their first attempt.

For the final S/D discrimination task, the same experimental procedure was used as with the brightness and LRT discrimination experiments. The test tones values were set at four times the instrument's n-value and a quarter of the instrument's n-value respectively for the upper and lower limits. Since the n-value is a perceptually simple parameter, the participants were asked to choose which tone had the most abrupt or quickest decay (corresponding to lower absolute values).

5.4 SUMMARY

The CI participant demographics and NH listener information has been provided along with a detailed description of the experimental procedure and stimuli used during the discrimination experiments. The data was collected and statistically analysed and the results compared to the literature in the following chapter.

CHAPTER 6 RESULTS: DISCRIMINATION EXPERIMENTS

The discrimination procedure allowed the JNDs for each timbre parameter of each instrument to be collected for the participants. The next step was to statistically analyse the data.

6.1 BRIGHTNESS

The brightness JNDs for each of the participants were found as the average of the JNDs approaching from higher values and the JNDs approaching from lower values. Figure 6.1 and Figure 6.2 show the JNDs of each of the listeners for each of the instruments. Average discrimination and standard deviations for each of the six NH listeners and each of the six CI listeners are given in Figure 6.3 and Figure 6.4. To allow a more detailed overview of NH participant performance, the scales of the graphs of NH and CI data do not match.

The average JND across all NH listeners and for all instruments is 0.049. Figure 6.5 shows that JNDs of all listeners increased as the brightness increased. The instruments with higher brightness values are instruments 3, 8, 10 and 13 (oboe, trumpet, bowed cello and trombone). Figure 6.1 shows that these instruments have slightly higher JNDs. An examination of Figure 6.2 shows that this is true for S24, S23 and S26, and to a lesser extent for S15. However, the JNDs of S22 and S28 are consistent across the range of brightness.

For the NH group, a two-way ANOVA indicated significant differences ($F(12, 65) = 2.81$, $p < 0.01$, $p = 0.0042$) in JNDs for different instruments, but no significant difference ($F(5, 72) = 0.32$, $p > 0.05$) in responses across participants. Post-hoc tests indicated that the bowed cello had a significantly higher mean compared to ten other instruments. The bowed cello did not have a significantly large JND compared to the oboe and violin, which also have high brightness values. The bowed cello had the highest brightness (6.65) in the entire instrument set, followed by the trumpet, bowed violin and oboe.

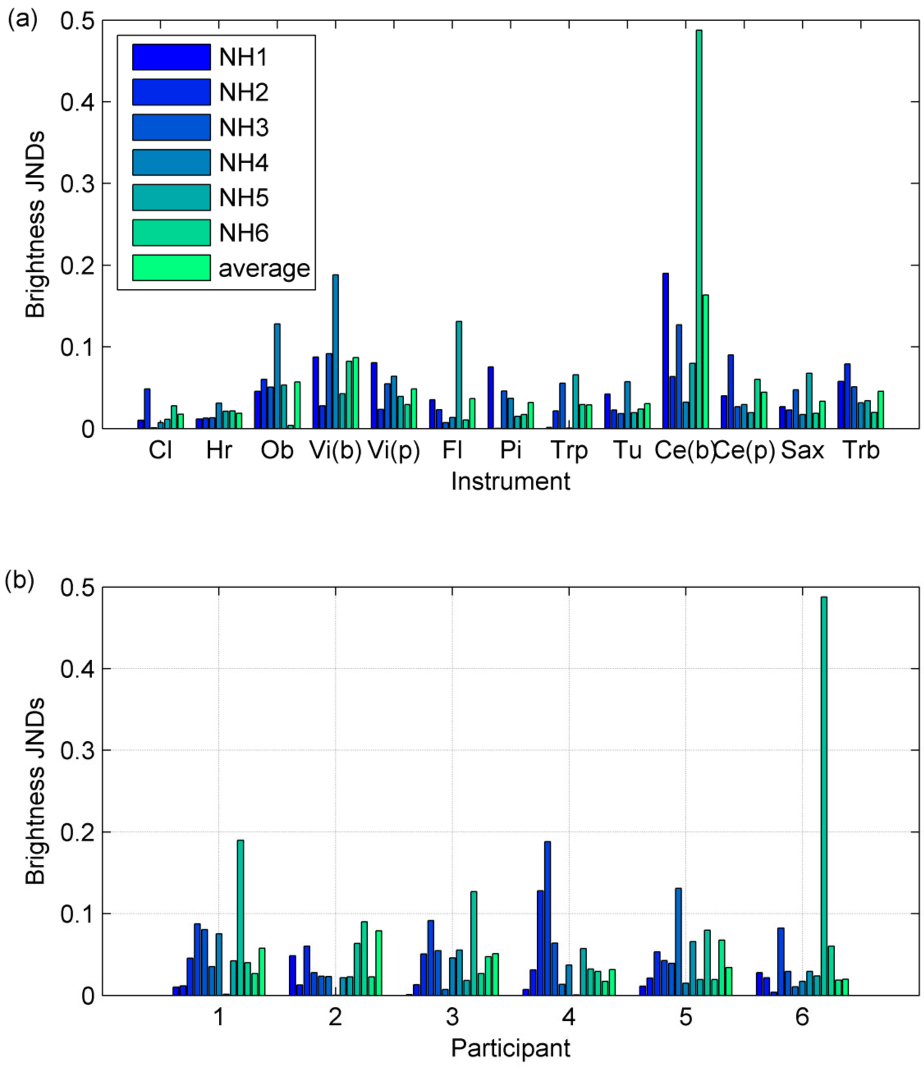

Figure 6.1. Brightness discrimination estimates for NH listeners. Brightness has no unit. Cl: Clarinet; Hr: French Horn; Ob: Oboe; Vi(b): Bowed violin; Vi(p): Plucked violin; Fl: Flute; Pi: Piano; Trp: Trumpet; Tu: Tuba; Ce(b): Bowed cello; Ce(p): Plucked cello; Sax: Saxophone; Trb: Trombone.

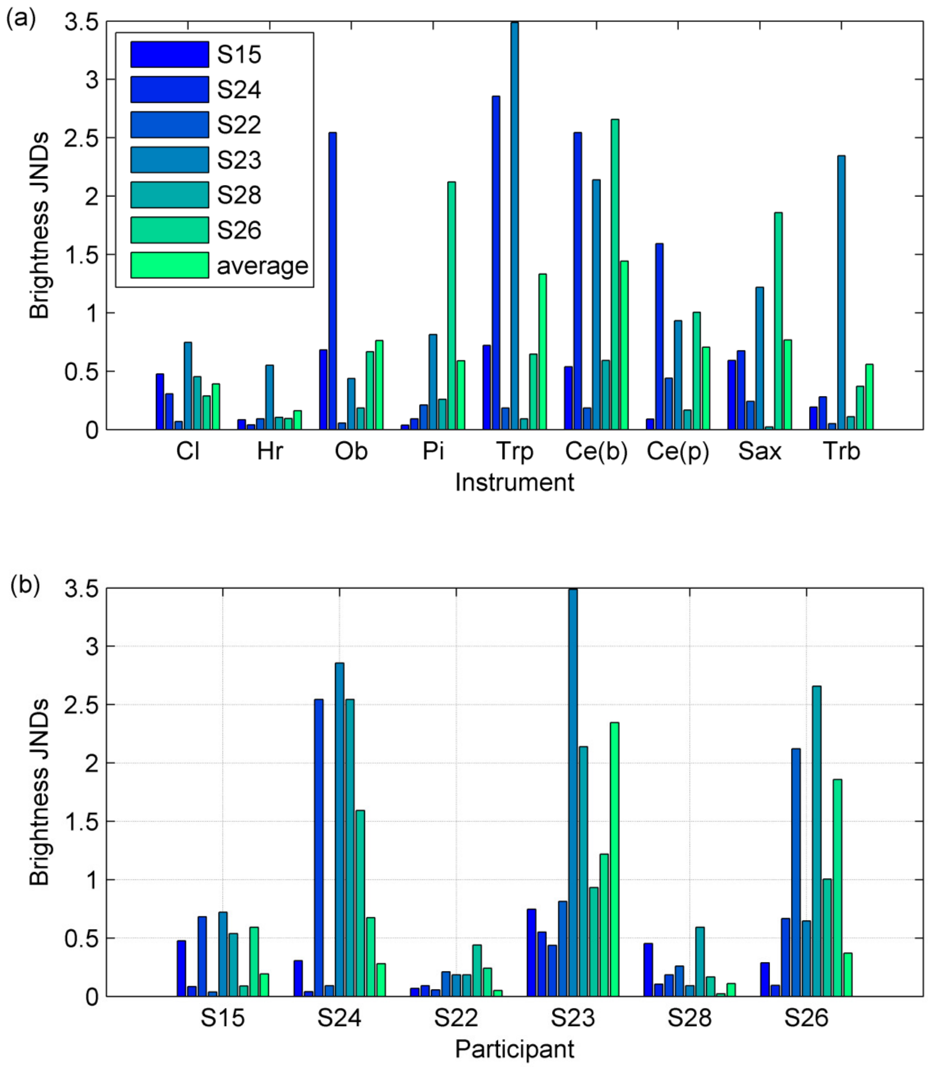

Figure 6.2. Brightness discrimination estimates for CI users. Brightness is a unitless dimension. Cl: Clarinet; Hr: French Horn; Ob: Oboe; Pi: Piano; Trp: Trumpet; Tu: Tuba; Ce(b): Bowed cello; Ce(p): Plucked cello; Sax: Saxophone; Trb: Trombone.

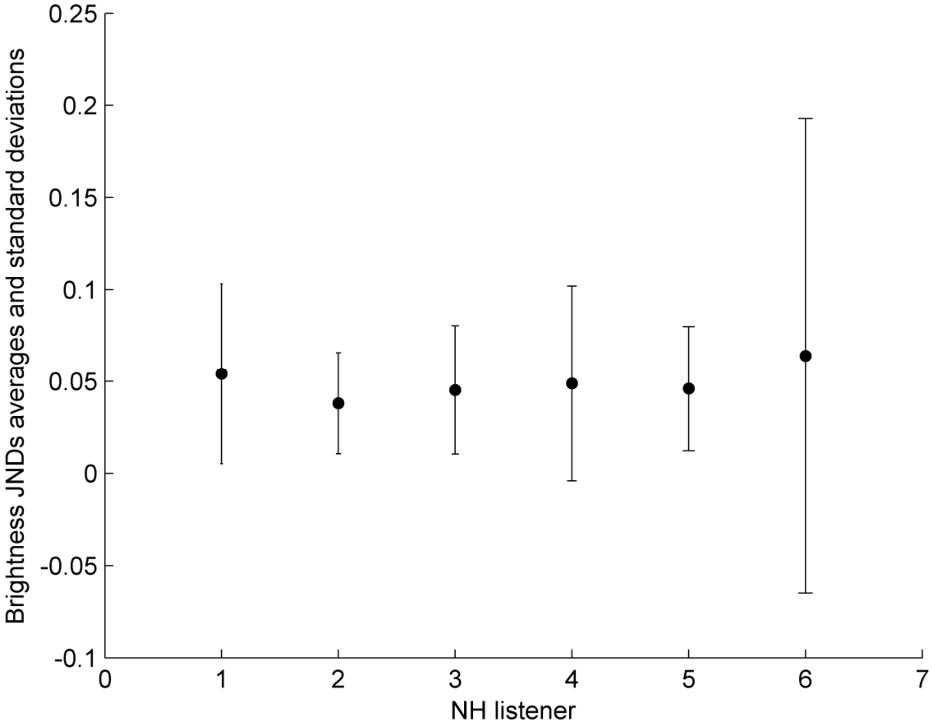

Figure 6.3. Averages of T_b JNDs and standard deviations of each of the six NH listeners.

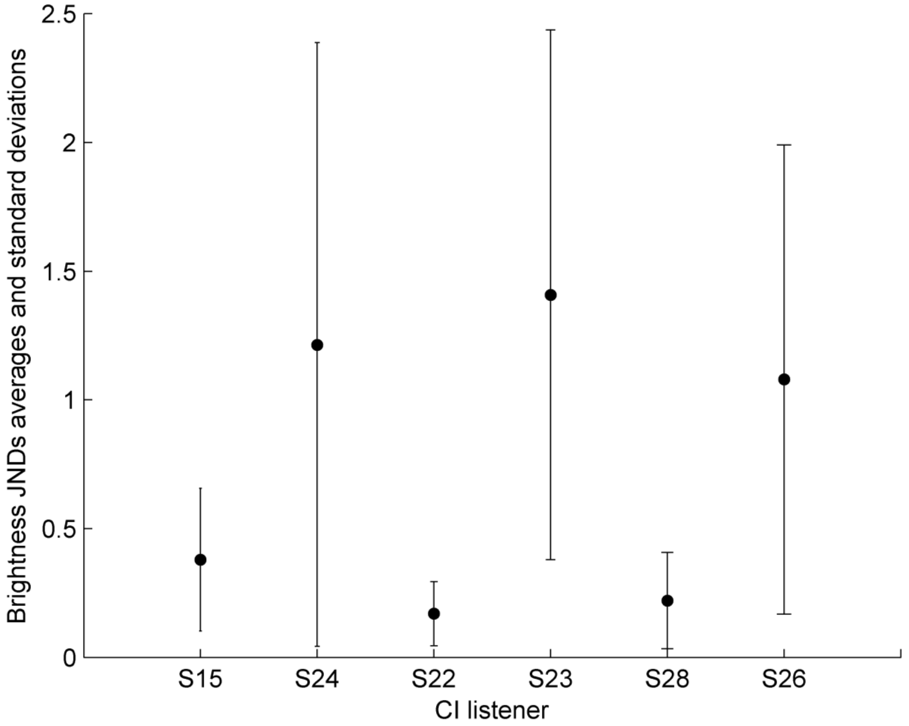

Figure 6.4. Averages of T_b JNDs and standard deviations of each of the six CI listeners.

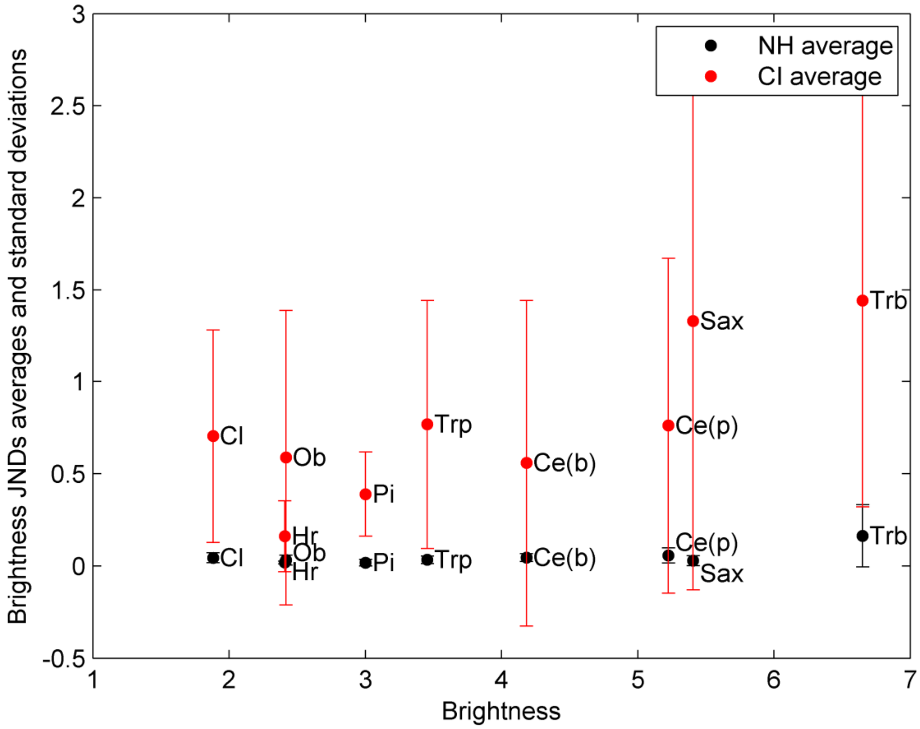

Figure 6.5. Average of T_b JNDs and standard deviations are displayed as a function of T_b. Towards the higher T_b values, JNDs of CI users tended to increase. Cl: Clarinet; Hr: French Horn; Ob: Oboe; Pi: Piano; Trp: Trumpet; Ce(b): Bowed cello; Ce(p): Plucked cello; Sax: Saxophone; Trb: Trombone

For the CI group, there was a significant difference between the JNDs of participants (F(5, 48) = 5.74, $p < 0.01$, $p = 0.0004$), but no difference across instruments (F(8, 45) = 2.12, $p > 0.05$, $p = 0.053$) due to the large standard deviations associated with the CI group. S22 and S28 performed significantly better than the two other participants (S23 and S24). S23 performed significantly worse than three other participants (S15, S22, S28). S24 performed significantly worse than two other participants (S22, S28). S26's JNDs were also higher compared to S15, S22 and S28, but not significantly so. Although not significant, investigation of the difference between instruments for the CI group also indicated that higher brightness values (instruments 10, 8, 12 and 3) were accompanied by larger JNDs.

To establish if a significant difference existed between the NH and CI group, a three-way ANOVA was performed. It showed that a significant difference in JNDs existed between the two groups (F(1, 106) = 35.97, $p < 0.01$, $p = 4.10\times10^{-8}$). No interaction effect between group and instrument was observed (F(8, 99) = 1.21, $p > 0.05$, $p = 0.3013$), indicating that NH listeners and CI users responded similarly to instruments.

6.2 IRREGULARITY

The JND for irregularity was found as the average for JNDs approaching from upper and lower values. Figure 6.6 and Figure 6.7 show the responses of the NH and CI listeners.

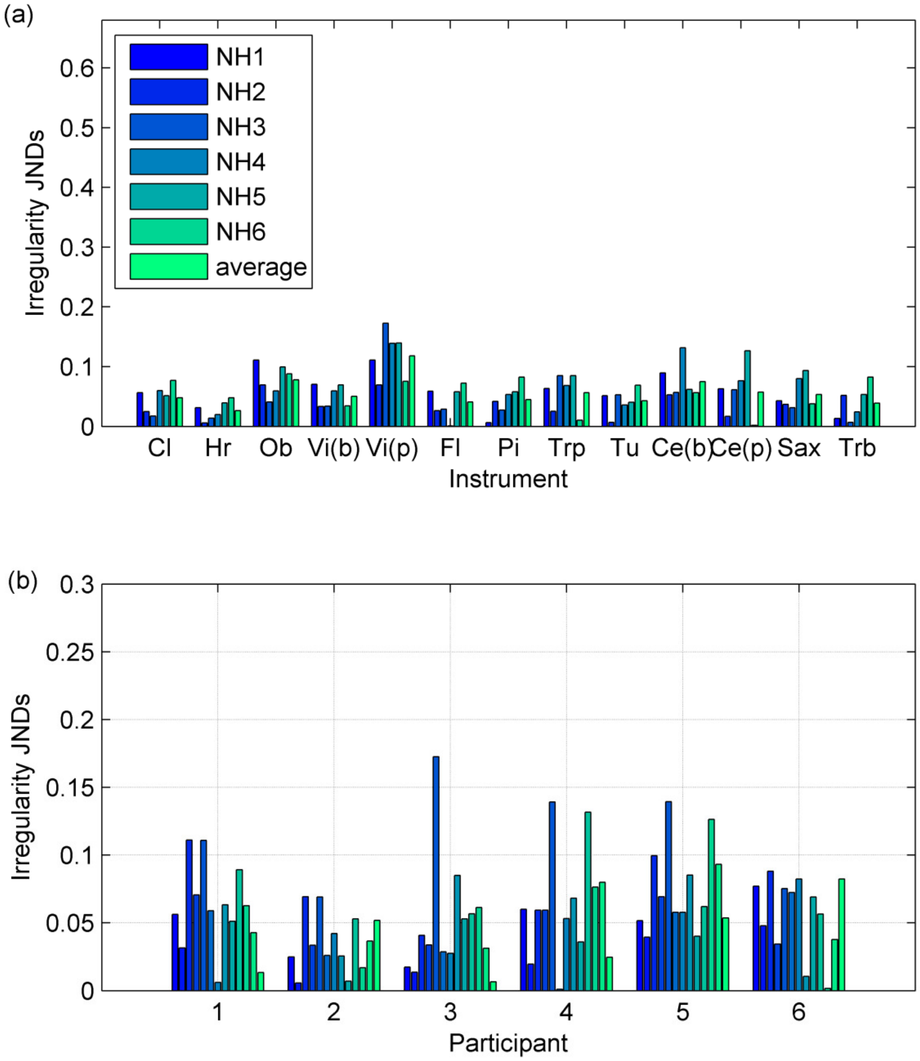

Figure 6.6. Dimensionless irregularity discrimination estimates for NH listeners. Cl: Clarinet; Hr: French Horn; Ob: Oboe; Vi(b): Bowed violin; Vi(p): Plucked violin; Fl: Flute; Pi: Piano; Trp: Trumpet; Tu: Tuba; Ce(b): Bowed cello; Ce(p): Plucked cello; Sax: Saxophone; Trb: Trombone.

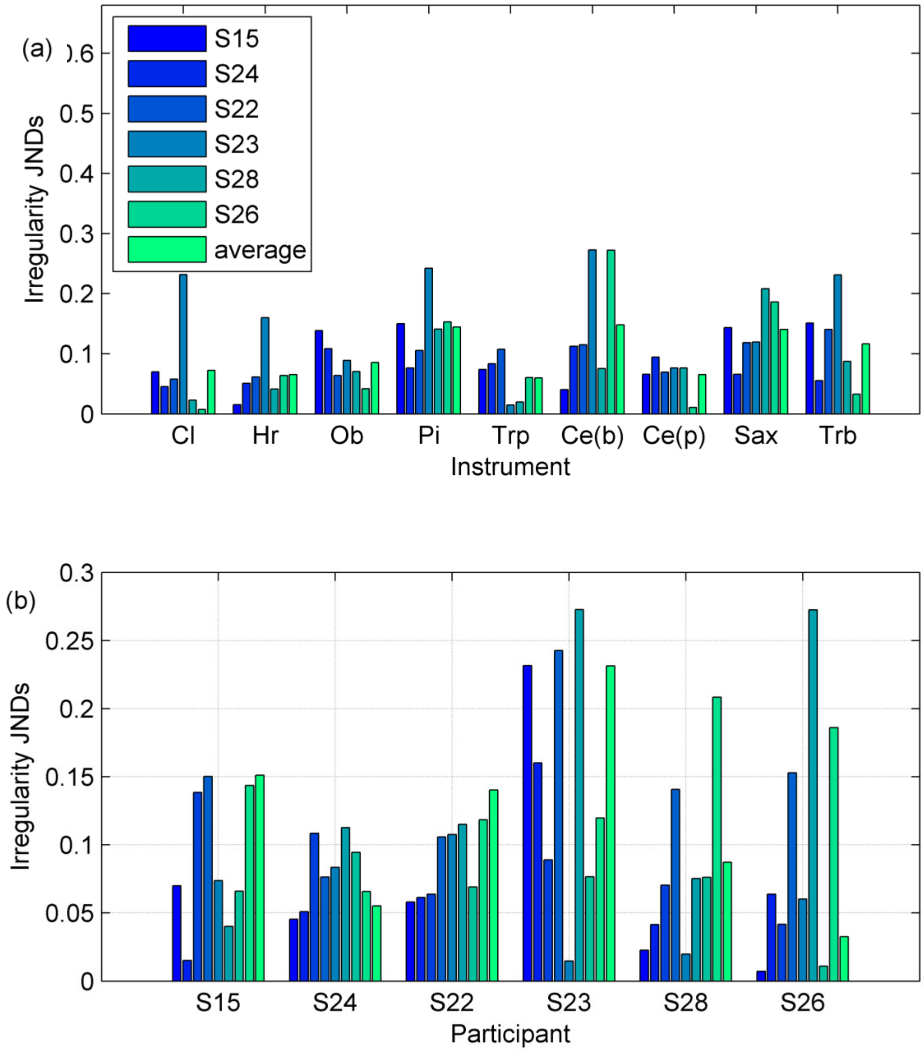

Figure 6.7. Irregularity discrimination estimates for CI users. Cl: Clarinet; Hr: French Horn; Ob: Oboe; Pi: Piano; Trp: Trumpet; Tu: Tuba; Ce(b): Bowed cello; Ce(p): Plucked cello; Sax: Saxophone; Trb: Trombone.

The average discrimination and standard deviations for each of the six NH listeners and each of the six CI listeners are given in Figure 6.8 and Figure 6.9.

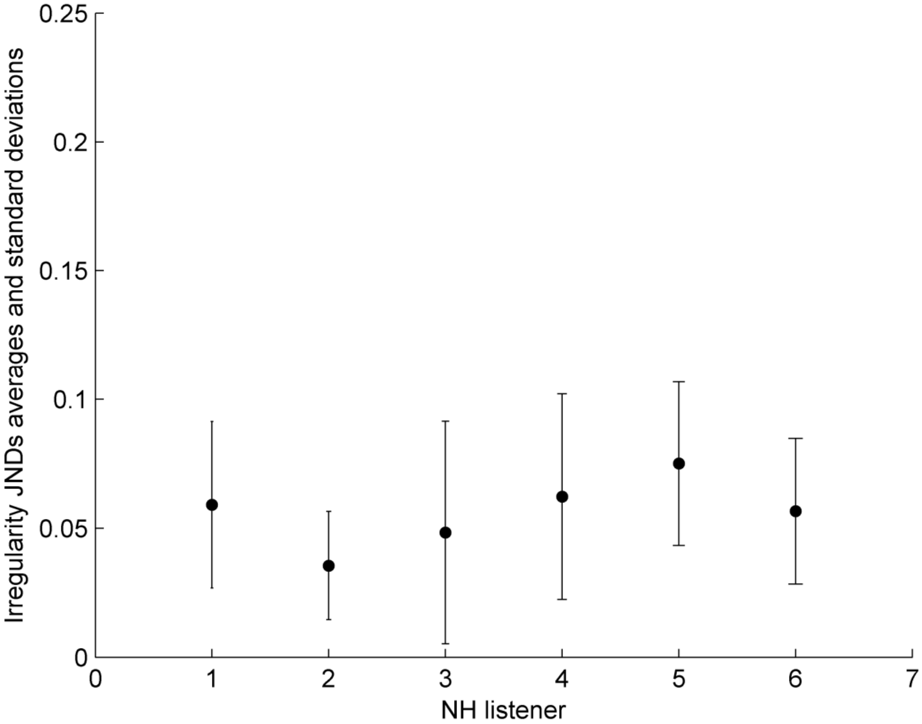

Figure 6.8. Average JNDs and standard deviations of each of the six NH listeners.

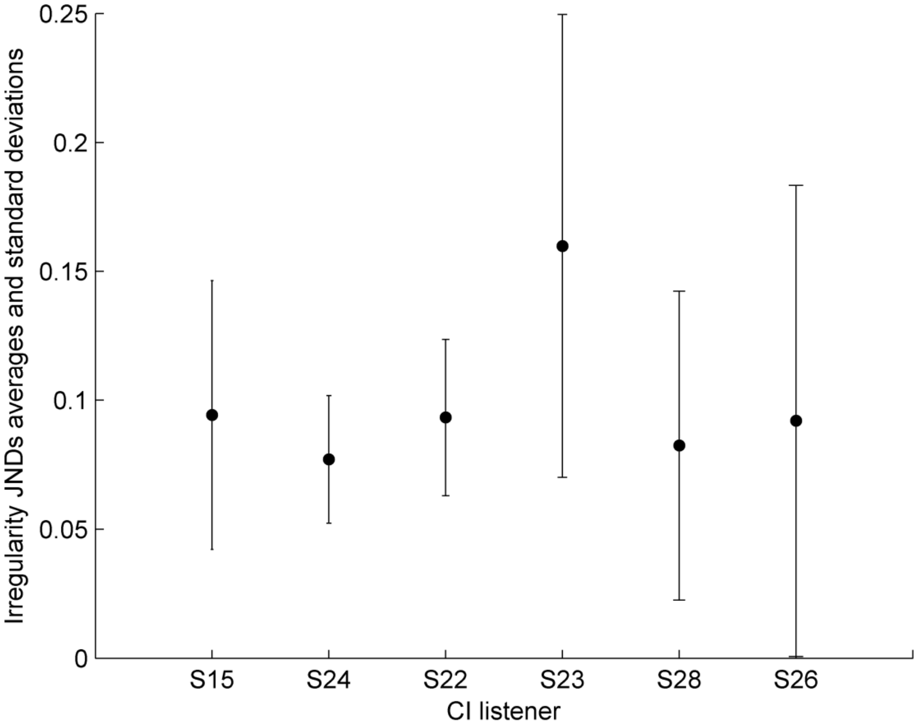

Figure 6.9. Average JNDs and standard deviations of each of the six CI listeners.

A two-way ANOVA performed on the NH listener data revealed significant differences between participants as well as instruments ($F(5, 72) = 3.31$, $p < 0.05$, $p = 0.0105$ and ($F(12, 65) = 4.64$, $p < 0.01$, $p = 2.8494\times10^{-5}$). Post-hoc tests revealed a significant difference in irregularity JND between NH2 and NH5, but that no other significant differences existed between other listeners. The decaying plucked violin note has the smallest value of n in the entire instrument set and is therefore the shortest note in the set. Both NH and CI listeners anecdotally mentioned increased difficulty in discriminating differences between a tone pair when the duration of the tones was shorter. The plucked violin note showed a significantly higher JND compared to all other instruments. Although not significant, the second and third highest JNDs were for the oboe and bowed cello, both of which have high brightness values ($T_{b,oboe}$, $T_{b,bowed\ cello} > 5$). This was followed by also relatively high JNDs for the short piano ($n_{piano} < 0$) and bright trumpet sounds ($T_{b,trumpet} > 5$). Figure 6.10 and Figure 6.11 respectively show the JNDs of the NH listeners as functions of the brightness and the decay parameter n. Results from the ANOVAs and these figures indicated that the irregularity of shorter sounds and sounds with higher brightness are more difficult to discriminate.

A two-way ANOVA for the CI users showed significant differences in JNDs for different participants ($F(5, 48) = 2.58$, $p < 0.05$, $p = 0.0406$). Post-hoc test revealed that S23 performed significantly worse compared to S24. Examination of the JNDs revealed that although not significant, the JNDs of S23 were higher than those of the other participants. An examination of the JNDs across instruments yielded no significant differences, but an investigation showed that the JNDs of the bright bowed cello and trombone and short plucked cello sounds were among the four highest JNDs obtained.

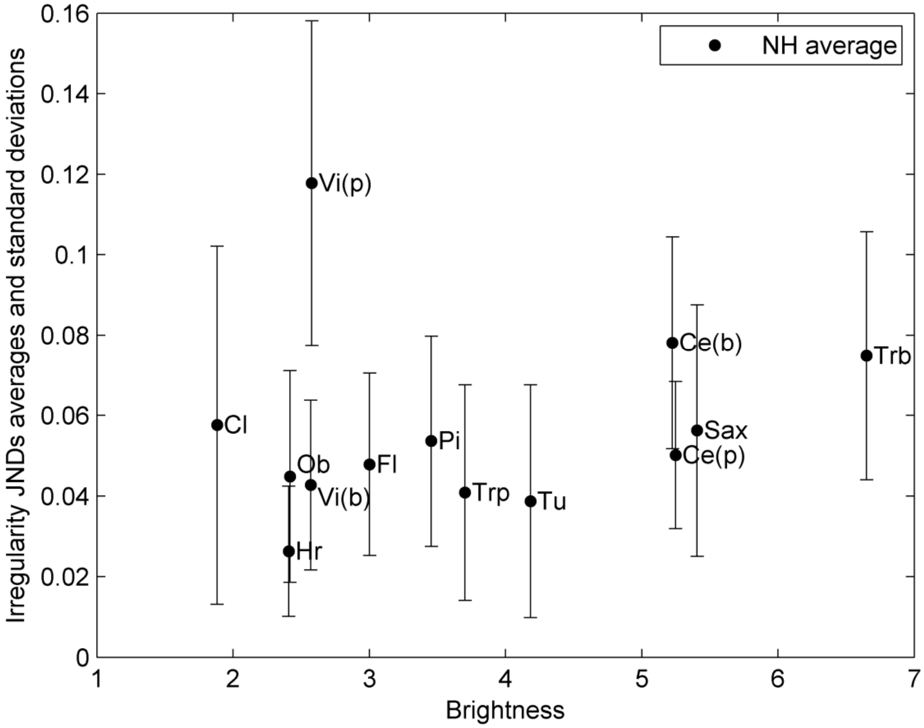

Figure 6.10. Average of IRR JNDs and standard deviations as a function of brightness for NH listeners. Note that JNDs tend to increase as a function of brightness. The outlier at the lower brightness of approximately 2.5 is the JND of the plucked violin note, with a large negative n-value. Cl: Clarinet; Hr: French Horn; Ob: Oboe; Vi(b): Bowed violin; Vi(p): Plucked violin; Fl: Flute; Pi: Piano; Trp: Trumpet; Tu: Tuba; Ce(b): Bowed cello; Ce(p): Plucked cello; Sax: Saxophone; Trb: Trombone.

A three-way ANOVA showed that a significant difference in JNDs existed between the two groups (F(1, 106) = 25.72, $p < 0.01$, $p = 2.10\times10^{-6}$). A significantly higher JND (F(8, 99) = 2.47, $p < 0.01$, $p = 0.0181$) was obtained for the bowed cello compared to the horn – instruments with respectively the highest brightness and second lowest brightness in the instrument set. Additionally, investigation of JNDs indicated, once again, that for all participants, a trend was observed where JNDs were higher for brighter as well as shorter sounds. An interaction between group and instrument was not observed (F(8, 99) = 1.89, $p > 0.05$, $p = 0.0709$), indicating that regardless of the group (NH or CI), instruments were responded to similarly.

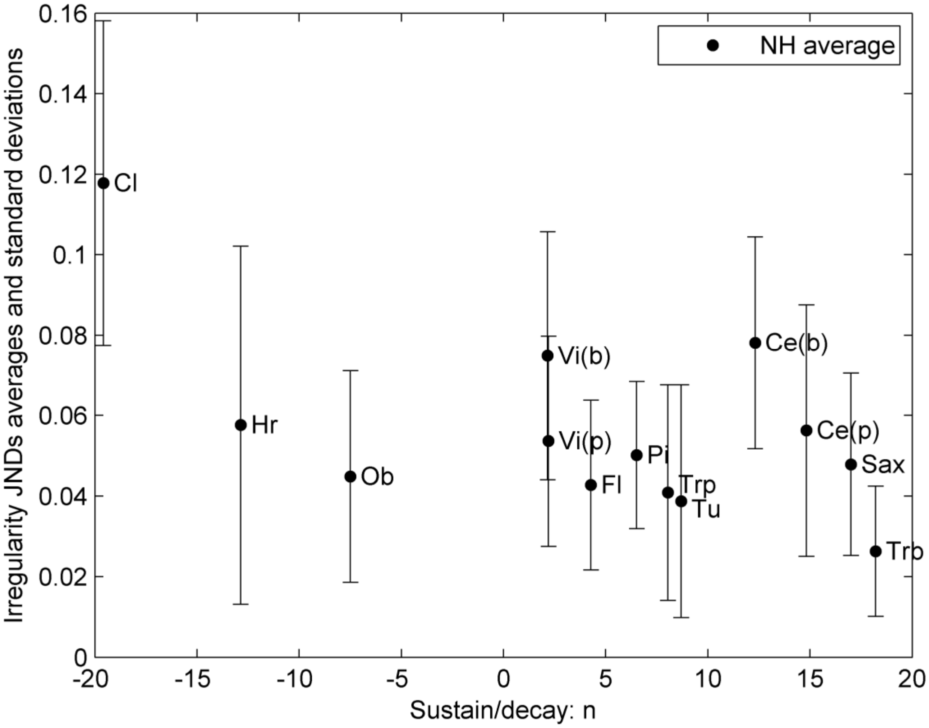

Figure 6.11. Average of IRR JNDs and standard deviations as a function of S/D values (n) for NH listeners. Note that JNDs tend to increase as the n-value becomes more negative. Cl: Clarinet; Hr: French Horn; Ob: Oboe; Vi(b): Bowed violin; Vi(p): Plucked violin; Fl: Flute; Pi: Piano; Trp: Trumpet; Tu: Tuba; Ce(b): Bowed cello; Ce(p): Plucked cello; Sax: Saxophone; Trb: Trombone.

6.3 LOGARITHMIC RISE-TIME

The LRT discrimination was found as the average of JNDs approaching from upper and lower values. JNDs for NH listeners and CI users are shown in Figure 6.12 and Figure 6.13. Different scales for NH and CI data is used to allow more details of the NH data to be visible.

Figure 6.14 and Figure 6.15 show the average JNDs and standard deviations of each of the listeners for NH and CI listeners. A two-way ANOVA performed on the NH listener data set showed significant differences between NH listeners ($F(5, 72) = 9.38$, $p < 0.05$, $p = 1.2157\times10^{-6}$), but no significant effect for instrument ($F(12, 65) = 1.20$, $p < 0.05$, $p = 0.3055$). A post-hoc test revealed that the JNDs of NH2 was significantly lower compared to three other listeners (NH1, NH4 and NH6) and that of NH5 was significantly lower than two other listeners (NH1 and NH6). A two-way ANOVA performed on the CI user data set showed significant differences between CI listeners (F(5, 48) = 12.66, $p < 0.01$, $p = 2.118\times10^{-7}$), as well as significant differences when comparing instruments (F(8, 45) = 2.84, $p < 0.05$, $p = 0.0135$). Post-hoc tests revealed that S23 performed significantly worse compared to the other CI users.

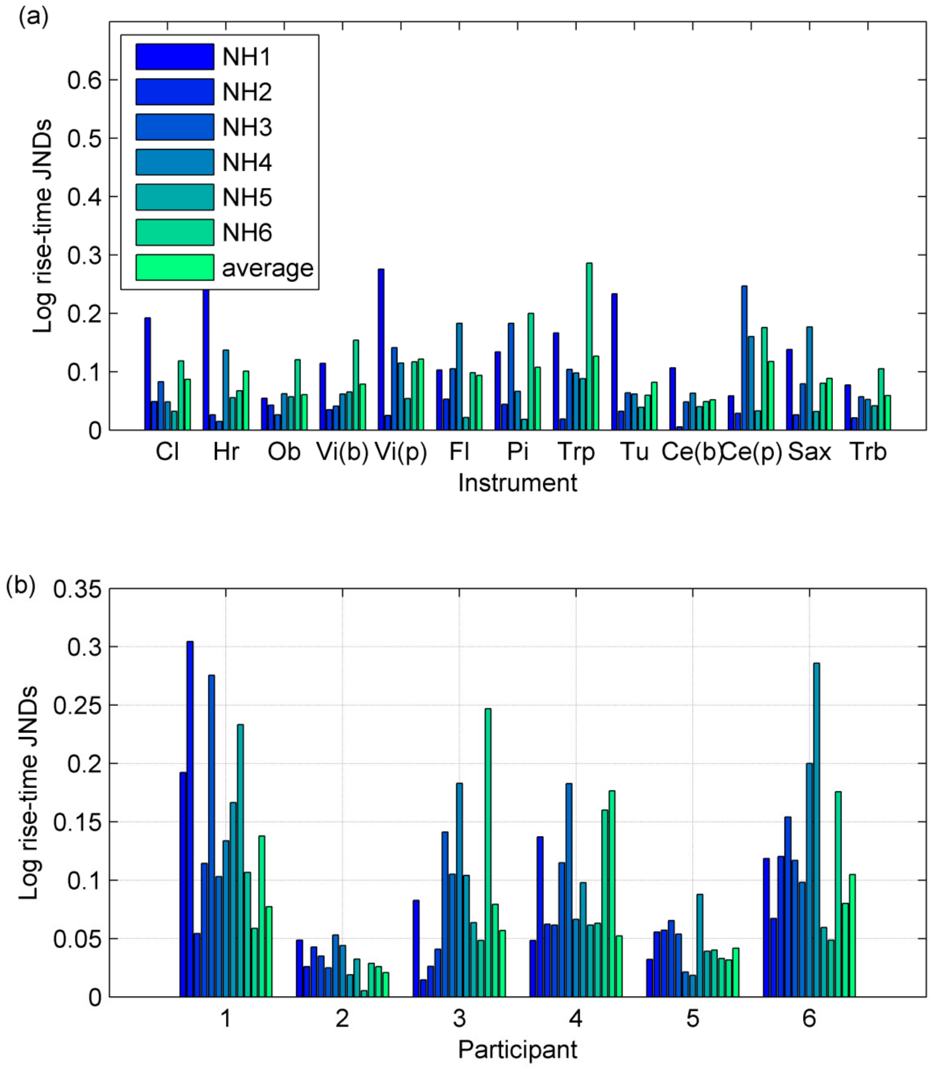

Figure 6.12. LRT discrimination estimates for NH listeners in log(s). Cl: Clarinet; Hr: French Horn; Ob: Oboe; Vi(b): Bowed violin; Vi(p): Plucked violin; Fl: Flute; Pi: Piano; Trp: Trumpet; Tu: Tuba; Ce(b): Bowed cello; Ce(p): Plucked cello; Sax: Saxophone; Trb: Trombone.

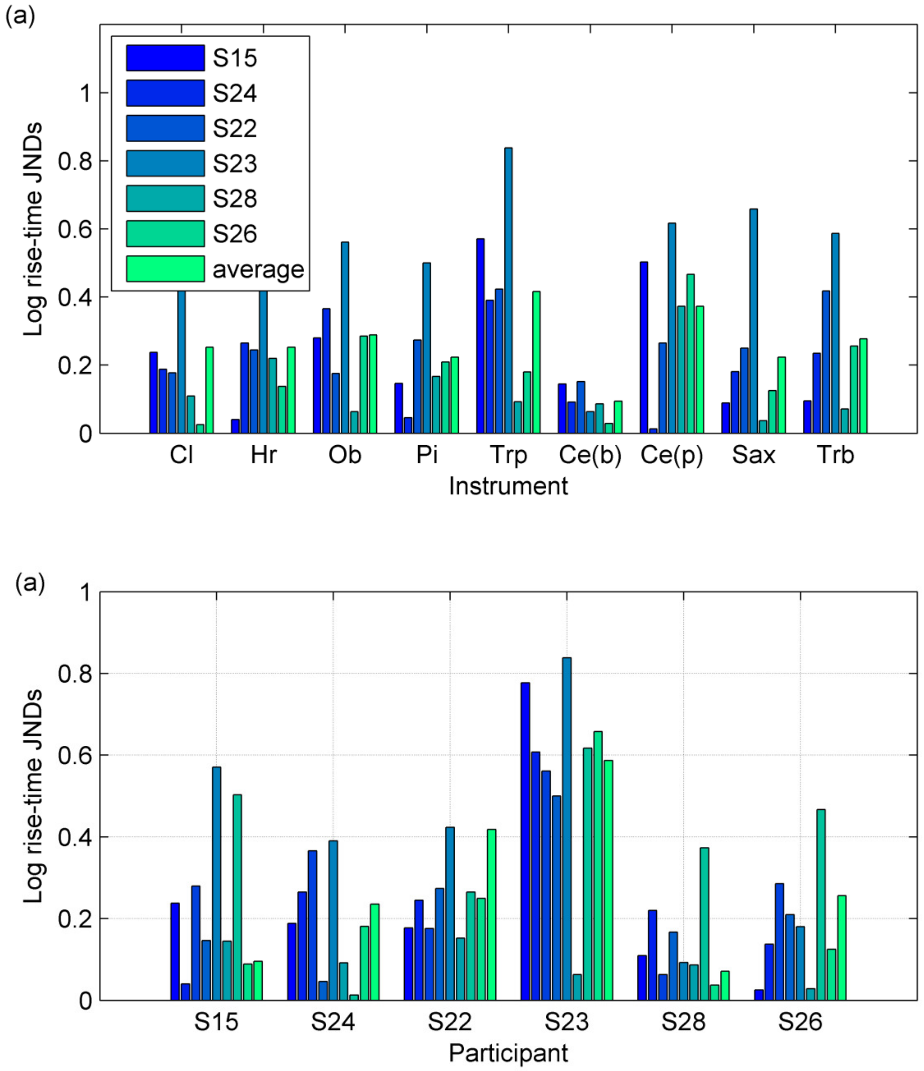

Figure 6.13. LRT discrimination estimates for CI listeners in log(s). Cl: Clarinet; Hr: French Horn; Ob: Oboe; Pi: Piano; Trp: Trumpet; Tu: Tuba; Ce(b): Bowed cello; Ce(p): Plucked cello; Sax: Saxophone; Trb: Trombone.

A three-way ANOVA revealed significant differences in performance between the CI and NH groups (F(1, 106) = 37.06, $p < 0.01$, $p = 2.76\times10^{-8}$). No significant differences across instruments were observed and no interaction effects were observed either (F(8, 99) = 1.63, $p > 0.05$, $p = 0.128$ and F(8, 99) = 0.0178, $p > 0.05$, $p = 0.6285$).

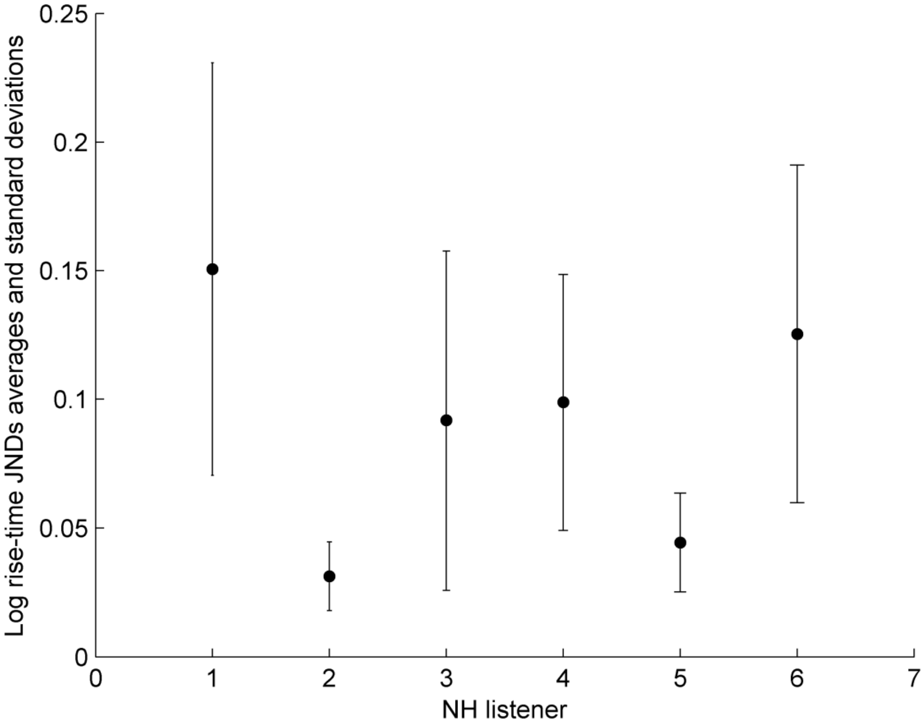

Figure 6.14. Average of LRT JNDs and standard deviations of each of the NH listeners in log(s).

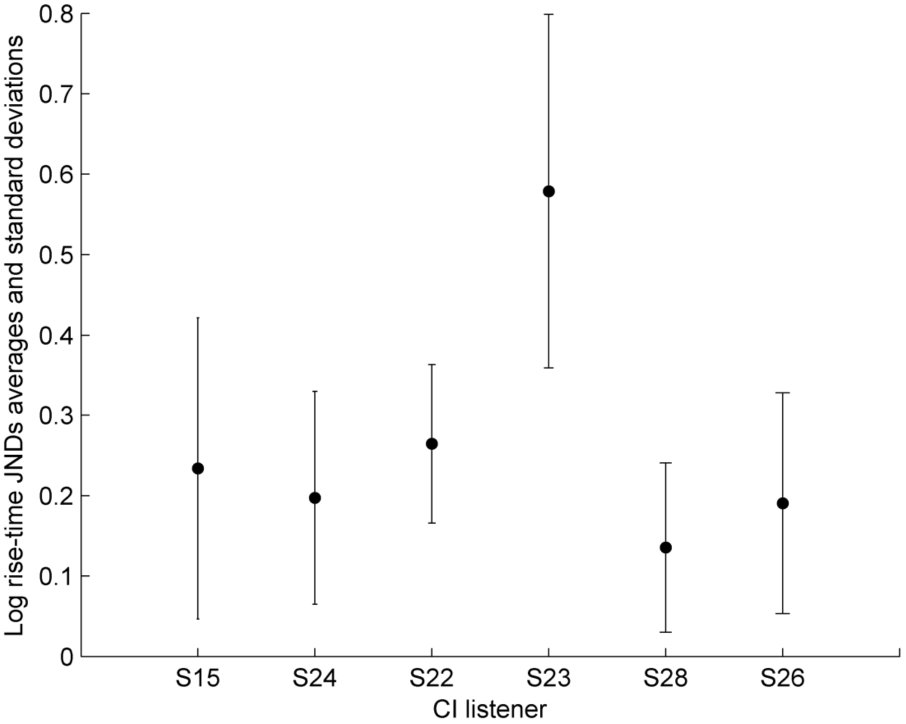

Figure 6.15. Average of LRT JNDs and standard deviations of each of the CI listeners in log(s).

6.4 SUSTAIN/DECAY

JNDs were calculated as the averages between the JNDs approaching from higher and lower values. Figure 6.16 and Figure 6.17 show the JNDs for NH and CI listeners.

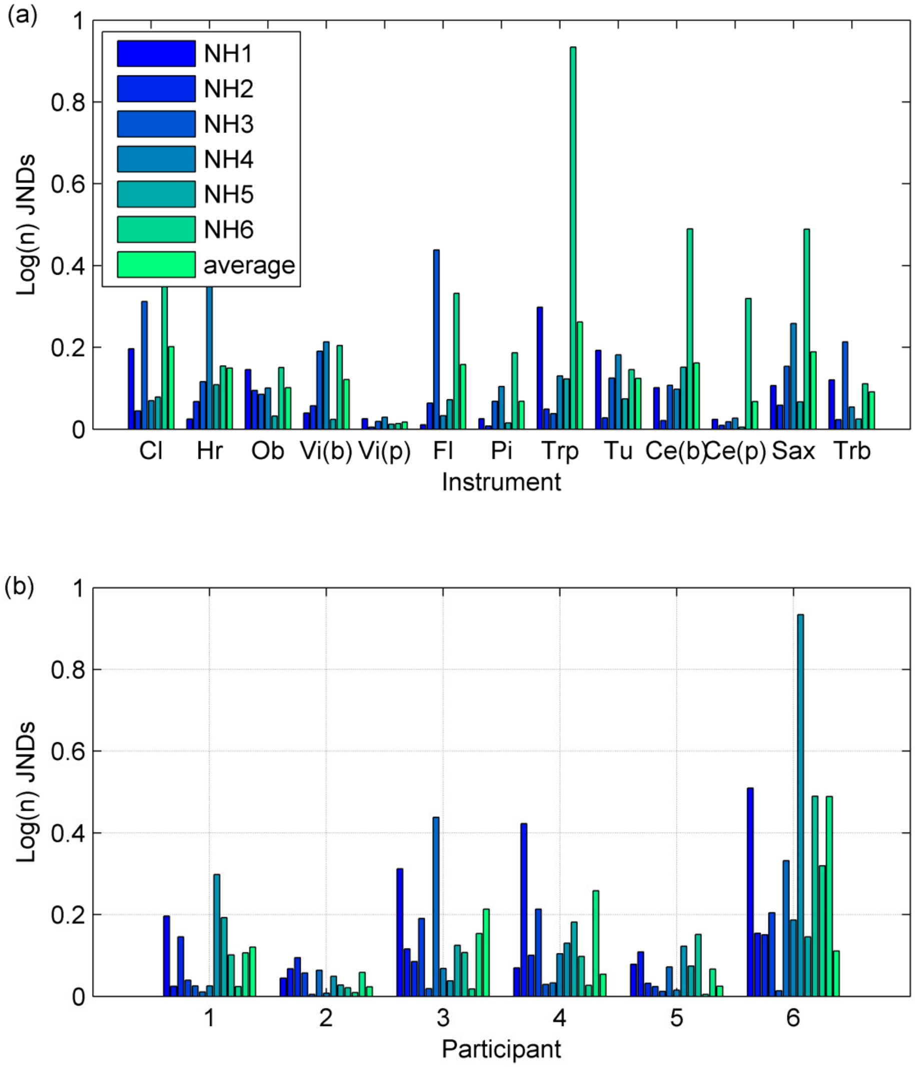

Figure 6.16. Log(n) discrimination estimates for NH listeners. LRT discrimination estimates for NH listeners. Cl: Clarinet; Hr: French Horn; Ob: Oboe; Vi(b): Bowed violin; Vi(p): Plucked violin; Fl: Flute; Pi: Piano; Trp: Trumpet; Tu: Tuba; Ce(b): Bowed cello; Ce(p): Plucked cello; Sax: Saxophone; Trb: Trombone.

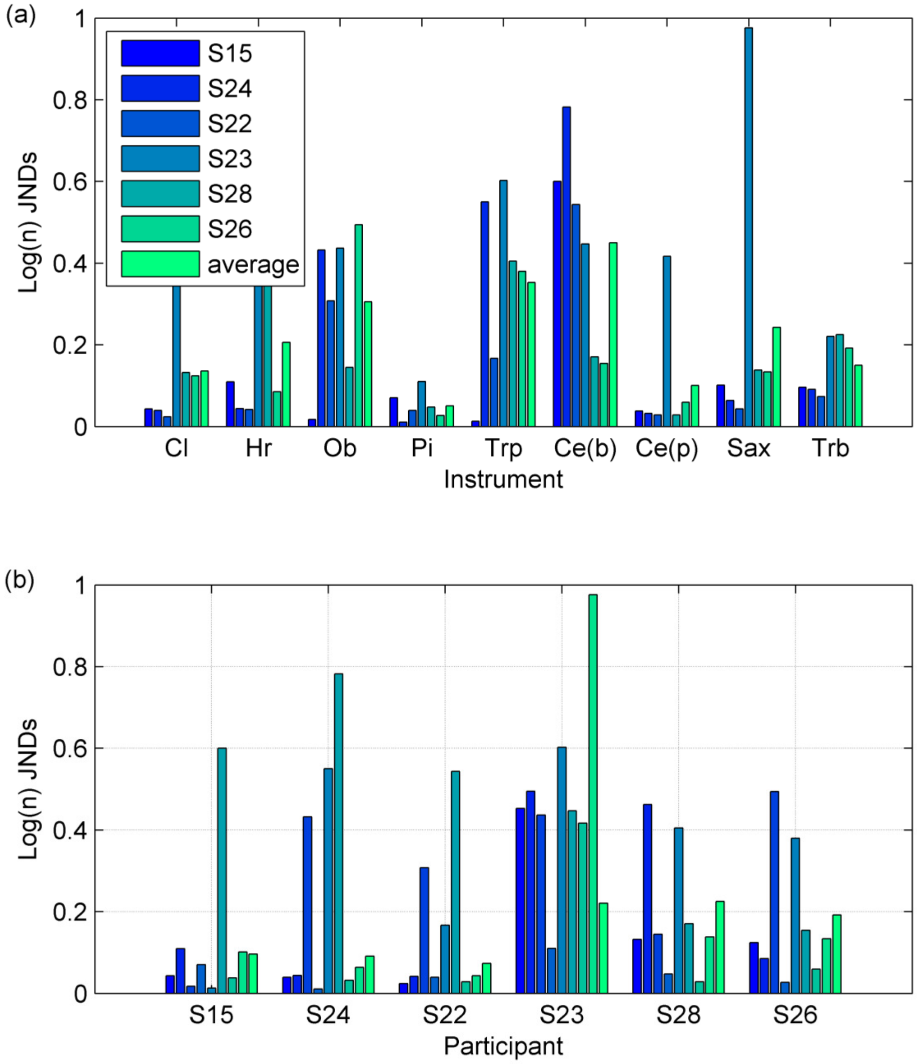

Figure 6.17. Log(n) discrimination estimates for CI users. Cl: Clarinet; Hr: French Horn; Ob: Oboe; Pi: Piano; Trp: Trumpet; Tu: Tuba; Ce(b): Bowed cello; Ce(p): Plucked cello; Sax: Saxophone; Trb: Trombone.

A two-way ANOVA for the NH group revealed significant differences between participants (F(5, 72) = 8.27, $p < 0.05$, $p = 5.4001\times10^{-6}$), but similar responses for instruments (F(12, 65) = 1.75, $p > 0.05$, $p = 0.0792$). NH6 performed significantly worse compared to the other listeners. A two-way ANOVA for the CI group revealed significant differences between participants F(5, 48) = 4.55, $p < 0.01$, $p = 0.00223$) and significant differences across instruments (F(8, 45) = 3.3, $p < 0.01$, p = 0.00551). S23 performed significantly worse compared to all other listeners, except S24. No other significant differences were observed in the CI group when S23 was excluded from the group. A significantly higher JND was obtained for the bowed cello compared to the piano and plucked cello. This may indicate that n-value differences in decay-type instruments are more readily differentiated. Although no significant difference (F(12, 65) = 1.75, $p > 0.05$, p = 0.0792) was observed across instruments for NH listeners, the average JNDs of the three decay-type instruments for all NH listeners were the three lowest JNDs observed across the instrument set, further supporting the idea that differences in decay-type instruments are easier to discriminate. No interaction effects were observed either (F(8, 99) = 0.0178, $p > 0.05$, p = 0.6285).

A three-way ANOVA revealed a significant difference between the NH and CI group (F(1, 106) = 4.70, $p < 0.05$, p = 0.0328) and a significant difference for different instruments was observed (F(8, 99) = 2.62, $p < 0.05$, p = 0.0127). No interaction effects between group and instrument were observed (F(8, 99) = 1, $p > 0.05$, p = 0.4437), indicating that the effect of the instrument on the JNDs does not differ between groups.

Figure 6.18 and Figure 6.19 indicate the average JNDs of log(n) values for each of the participants, along with their standard deviations.

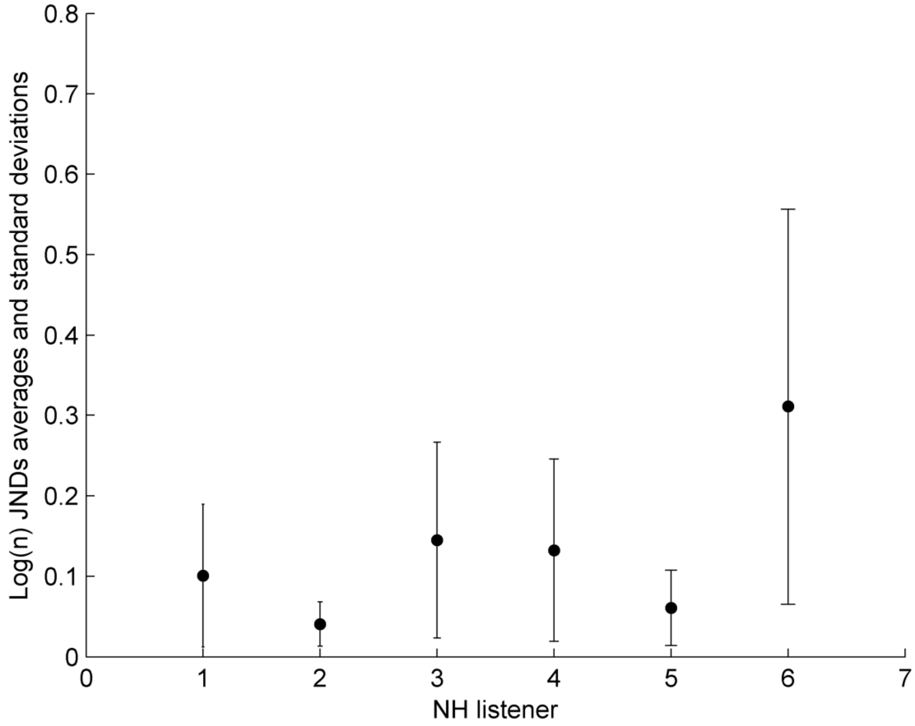

Figure 6.18. Average JNDs of log(n) and standard deviations of the NH listeners.

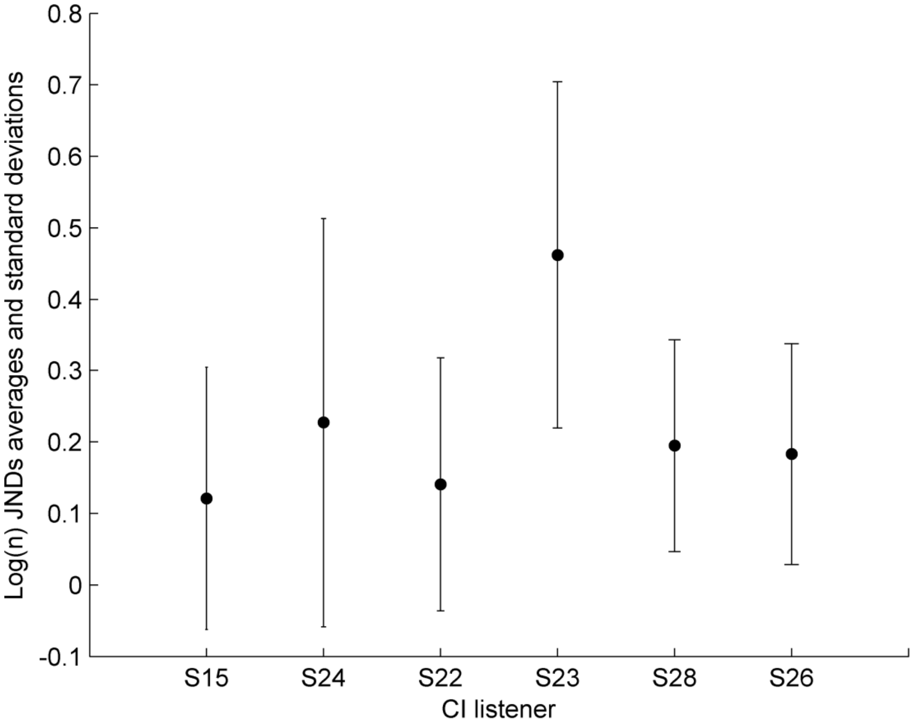

Figure 6.19. Average JNDs of log(n) and standard deviations of the CI listeners.

6.5 CONFUSION MATRICES

The JNDs obtained in the discrimination experiments are valuable data. Direct comparison of the JNDs of the NH group and the JNDs of the CI group allows preliminary problem areas within timbre to be identified. Specific observations such as the increase in brightness JNDs as brightness increases (Figure 6.5) is also evident from JNDs. Further processing of the JND data can be used to produce models that predict instrument confusions listeners are likely to make. In turn, the confusion matrices can be used to perform feature extractions to determine the extent to which certain timbre dimensions are perceived by listeners.

Svirsky (2000) proposed that phoneme recognition is based on the listener's discrimination along certain perceptual dimensions. Since, multidimensional studies propose that timbre recognition also employs discrimination along timbre certain dimensions, Svirsky's approach was applied to the dimensions of timbre (a concept explored and developed by Van Zyl (2008) and Burmeister (2008) and applied to timbre perception by Hugo (2009)). Each stimulus in the perceptual space was surrounded by a Gaussian distributed area of uncertainty, with standard deviations equal to the JNDs of the dimension. Svirsky suggested that confusion matrices could be constructed by integrating the product of the probability function of two instruments in each dimension. The Gaussian probability function S associated with stimulus E_i is given as:

$$\begin{aligned} S(E_i) &= S_i(x_1, x_2, \dots, x_m) \\ &= \frac{1}{JND_1 JND_2 \dots JND_m \left(\sqrt{2\pi}\right)^m} \\ &\times e^{-(x_1-T_{i1})^2/2JND_1^2} e^{-(x_2-T_{i2})^2/2JND_2^2} \dots e^{-(x_k-T_{im})^2/2JND_m^2}. \end{aligned} \tag{6.1}$$

Explained in terms of timbre, JND_i is the JND of timbre dimension i, and x_1 and T_{i1} are the values of timbre parameter i of two instrument tones. The confusion between these instrument tones in a confusion matrix is:

$$Cell_{ik} = \int S_i(x_1, x_2, \dots, x_m) dx_1 dx_2 \dots dx_m. \tag{6.2}$$

The results of these integrations can be seen in Table 6.1 and Table 6.2.

Table 6.1. Averages of the confusions of the six NH listeners indicate that almost no confusions between instruments exist.

	Clar	Horn	Oboe	Pian	Trum	CellA	CellP	Saxo	Trom
Clar	1	0	0	0	0	0	0	0	0
Horn	0	0.999999	0	6.66×10^{-7}	0	0	0	1.00×10^{-15}	0
Oboe	0	0	0.999998	0	1.66×10^{-6}	0	0	0	0
Pian	0	1.54×10^{-7}	0	1	0	0	1.17×10^{-11}	0	0
Trum	0	0	8.63×10^{-6}	0	0.999991	0	0	0	3.43×10^{-13}
CellA	0	0	0	0	0	1	0	0	0
CellP	0	0	0	1.18×10^{-10}	0	0	1	0	0
Saxo	0	0	0	0	0	0	0	1	2.84×10^{-13}
Trom	0	0	0	0	2.09×10^{-13}	0	0	1.39×10^{-13}	1

Table 6.2. Averages of the confusions of the CI listeners indicate that more confusions between instruments exist compared to NH listeners, although the confusions are still relatively small.

	Clar	Horn	Oboe	Pian	Trum	CellA	CellP	Saxo	Trom
Clar	9.97×10^{-1}	5.20×10^{-5}	8.32×10^{-4}	1.21×10^{-5}	4.56×10^{-7}	1.45×10^{-6}	3.05×10^{-7}	9.08×10^{-4}	9.92×10^{-4}
Horn	2.70×10^{-5}	9.63×10^{-1}	1.50×10^{-4}	2.46×10^{-4}	1.48×10^{-2}	5.41×10^{-12}	6.66×10^{-11}	8.47×10^{-3}	1.27×10^{-2}
Oboe	8.31×10^{-5}	1.17×10^{-3}	9.87×10^{-1}	1.06×10^{-5}	8.52×10^{-3}	4.95×10^{-5}	5.67×10^{-10}	1.13×10^{-3}	2.29×10^{-3}
Pian	7.01×10^{-5}	2.81×10^{-3}	5.56×10^{-4}	9.38×10^{-1}	3.48×10^{-3}	3.37×10^{-4}	1.83×10^{-2}	1.95×10^{-2}	1.74×10^{-2}
Trum	4.31×10^{-8}	5.40×10^{-3}	1.81×10^{-2}	5.64×10^{-5}	9.67×10^{-1}	6.16×10^{-13}	4.16×10^{-12}	1.54×10^{-3}	8.32×10^{-3}
CellA	1.46×10^{-4}	9.40×10^{-12}	4.12×10^{-4}	4.66×10^{-5}	5.98×10^{-12}	9.98×10^{-1}	8.03×10^{-5}	6.64×10^{-4}	1.88×10^{-4}
CellP	4.26×10^{-7}	1.79×10^{-10}	1.03×10^{-9}	4.42×10^{-3}	6.21×10^{-11}	1.40×10^{-4}	9.95×10^{-1}	4.25×10^{-5}	2.11×10^{-4}
Saxo	3.13×10^{-4}	5.63×10^{-3}	3.92×10^{-3}	1.12×10^{-3}	5.50×10^{-3}	2.66×10^{-4}	1.05×10^{-5}	9.72×10^{-1}	1.11×10^{-2}
Trom	1.11×10^{-3}	2.74×10^{-2}	2.36×10^{-2}	3.34×10^{-3}	4.11×10^{-2}	2.56×10^{-4}	1.69×10^{-4}	3.59×10^{-2}	8.67×10^{-1}

In order to establish the salience of each of the cues, an information transmission analysis was done, as done by Miller and Nicely (1955). The analyses were performed on the average predicted confusion matrices of both the NH and CI group.

6.6 FEATURE INFORMATION TRANSMISSION ANALYSIS

Performing a feature information transmission analysis (FITA) (Miller and Nicely, 1955) on the instrument confusions uncovers the amount of information transmitted to a listener by each of the timbre dimensions (Miller and Nicely, 1955). It is also a measure of the covariance between the input (instrument stimuli) and output (listener response). To analyse the amount of information transmitted through each of the timbre dimensions requires grouping of each of the instruments into distinct groups. Groupings were done as indicated in Table 6.3. Instruments with brightness values of $1 \leq T_b < 2.5$ were grouped as group 1, brightness values of $2.5 \leq T_b < 4$ were grouped as group 2, brightness values of $4 \leq T_b < 5.5$ were grouped as group 3 and brightness values of $T_b \geq 5.5$ were grouped as group 4. Irregularity values between $0 < \text{IRR} \leq 0.5$, $0.5 < \text{IRR} \leq 1$, $1 < \text{IRR} \leq 1.5$ were respectively grouped into group 1, group 2 and group 3. For the LRT values, group 1 contained instruments with $1 \leq \text{LRT} < 2$, while group 2 contained instruments with $\text{LRT} \geq 2$. The log(n) values were grouped as $\log n \leq -1$, $-1 < \log n \leq 0$, $0 < \log n \leq 1$ and $\log n \geq 1$, respectively falling into group 1, group 2, group 3 and group 4. The subjective division of instruments into these groups attempted to group instrument families together as far as possible.

Table 6.3. Partitioning of instruments into characteristic groups.

Parameter	Clar	Horn	Oboe	Pian	Trum	CellA	CellP	Saxo	Trom
T_b	2	1	3	1	3	4	1	2	3
IRR	3	1	2	1	1	2	2	1	1
LRT	2	1	1	1	1	2	1	2	2
N	3	3	3	1	3	2	1	2	3

The mutual information of x and y is

$$I(X,Y) = H(X) + H(Y) - H(X,Y), \tag{6.3}$$

where $H(X)$ and $H(Y)$ are the entropies of X $(x_1, \dots x_n \in X)$ and Y $(y_1, \dots y_n \in Y)$ and $H(X,Y)$ is the joint entropy of X and Y. The mutual information is a measure of the amount of information that can be obtained about one variable, say the input variable X, by observing another variable, say the output variable Y. To calculate the mutual information, the following equation is used:

$$I(X,Y) = -\sum_{x\in X} p(x)\log p(x) - \sum_{y\in Y} p(y)\log p(y) + \sum_{x,y} p(x,y)\log p(x,y), \tag{6.4}$$

where $p(x)$ and $p(y)$ are the probabilities of an input x and output y and $p(x,y)$ is the probability of output y given input x.

Figure 6.20 shows the percentage of information transmission for each of the parameters for NH listeners. For each of the NH listeners, the percentage of information transmitted is approximately equal. Figure 6.21 shows the percentage of information transmitted to each of the CI listeners. Information transmission varies more from one CI participant to the next. The information transmission appears high due to the fact that the confusion matrices obtained in section 6.5 indicate very little confusion across instruments. Even the collapsed confusion matrices show very little confusion across the categories of Table 6.3.

Comparison of Figure 6.20 and Figure 6.21 shows that the percentage of information transmitted to all of the CI listeners is less than that of the NH. A comparison of the amount of information transmitted to the two groups of listeners indicated that the information transfer through individual timbre parameters is similar. Across NH listeners, information transmission was almost identical, but the amount of information available to CI listeners varied across listeners.

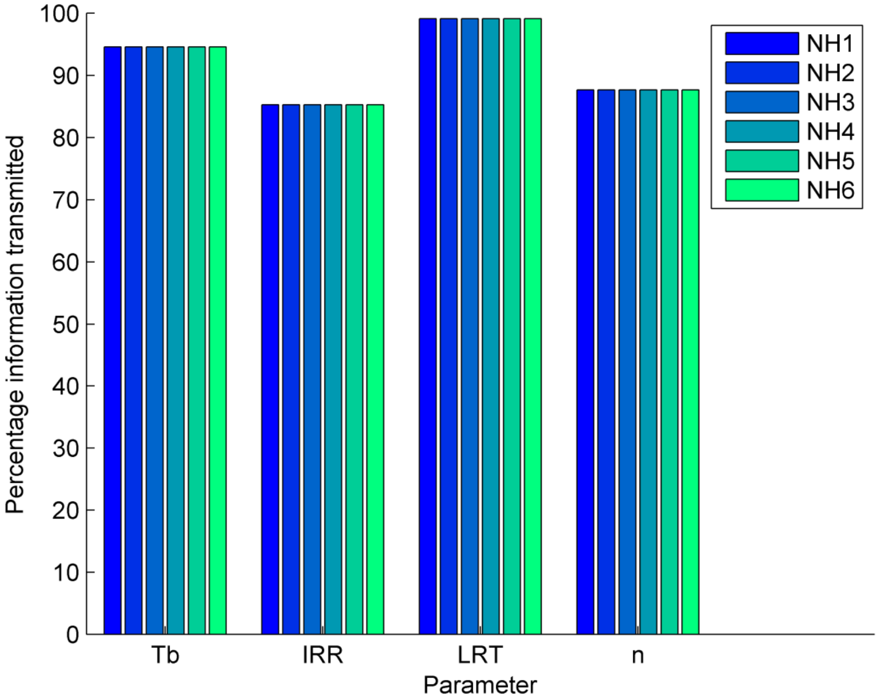

Figure 6.20. Percentage of information transmitted of each parameter for each of the NH listeners.

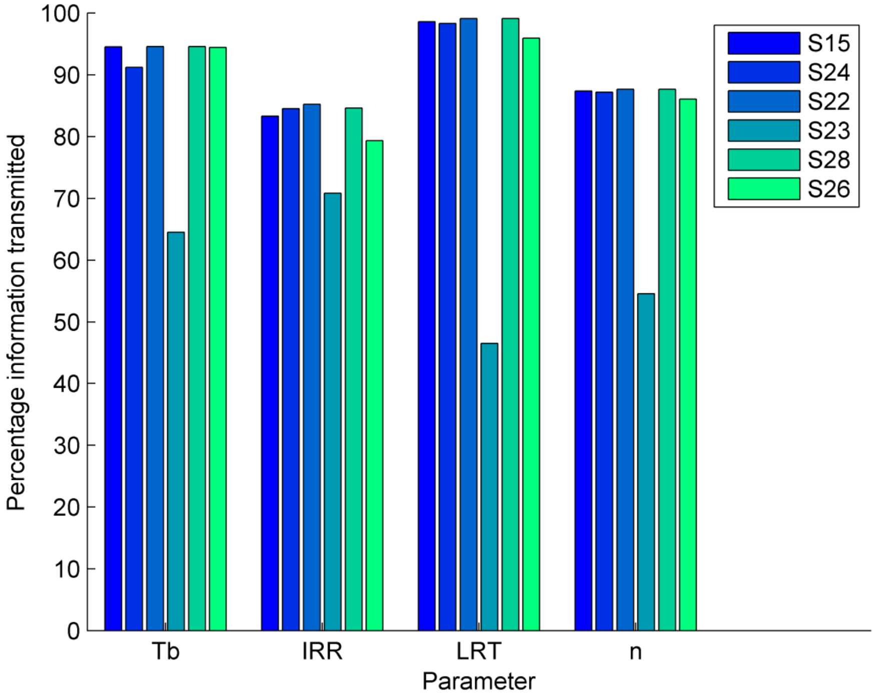

Figure 6.21. Percentage of information transmitted of each parameter for each of the CI listeners.

CHAPTER 7 DISCUSSION

The discussion includes thoughts on the choice of timbre parameters, their salience of the parameters and how these parameters contribute to producing good representation of real-world sounds. The importance of other timbre parameters not used here is also considered.

The JNDs obtained during the discrimination experiments are discussed with regard to the literature. The discussions of the JNDs relate results of previous timbre studies, followed by the results obtained here. Comparisons between the JNDs of T_b and LRT in the literature and the results obtained here were done, but logical comparisons of IRR and S/D JNDs obtained here to literature could not be done.

The confusion matrices are discussed with regard to the literature. An interpretation of the Svirsky's model applied to timbre recognition is provided and possible modifications are suggested to improve the model for such a timbre application. The FITA results are discussed with regard to the salience of the timbre dimensions.

7.1 TIMBRE PARAMETERS

The multidimensional nature of timbre results in a variety of timbre dimensions described, identified and tested in the literature. The literature study aimed to investigate the important parameters and draw together the various descriptions that are closely linked. A limited set of parameters was chosen in order to represent as closely as possible real-world sounds. The timbre dimensions chosen here appear to represent real-world instruments reasonably well: brightness and irregularity are established in the literature as perceptually salient timbre properties and the LRT and S/D is a good description of the temporal envelope in its entirety. Brightness is consistently defined in the literature and easily resynthesized. Irregularity is not as consistently defined, but the salience of spectral fine structure is evident. Although many sources concentrate on differences between even and odd harmonics, the spectral envelopes of most instruments are not that simple. It may be necessary in the future to redefine how irregularity is calculated and used to resynthesize tones. Investigation into the spectral envelopes of most instruments shows how the harmonics may fit the transfer function of a comb and lowpass filter, which is associated with parameters that could be transformed into brightness and irregularity. The possibility exists that synthetic tones based on these parameters will approximate real-world tones more accurately. The inclusion of phase information has an influence on the temporal waveshape and may also contribute to better instrument representations by synthetic sounds.

The temporal envelope discussed here consists of two parameters. The first parameter is the RT of tones, which has been researched and discussed extensively in the literature; the second parameter involves the decay time of the instruments in a decay curve that closely approximates that of real-world tones. It has the advantage that the parameter contains information about the nature of the instrument: whether it is a percussive attack that decays or whether it is a sustained instrument. It must be remembered that although the temporal envelope for sustain instruments does convey information about the nature of the instrument, the temporal parameters can be controlled to a certain degree in many instruments. A violinist, for example, can increase or decrease the rise and decay of a tone by changing the speed and pressure of the bow on the string.

Certain properties only present in specific instruments, such as detunedness in pianos, may be a salient feature for the specific instrument, even though the property is absent in most other instruments. Noisy components, which are not considered here, also accompany certain instruments, such as the breathy nature of the flute or the onset of bowed and plucked string instruments.

7.2 BRIGHTNESS JNDS

Brightness discrimination has been investigated previously using a variety of methods. Gunawan and Sen (2008) applied bandpass filters to instrument recordings that effectively changed the location of the spectral locus. They found that perceptual sensitivity was governed by the lower harmonics.

Emiroglu and Kollmeier (2008) morphed pairs of instruments, with a morphing parameter $\alpha = 0$ for one instrument and $\alpha = 1$ for the other. The morphing parameter was an indication of the change in timbre and JNDs were found for these changes. A French horn-trombone combination was chosen due to the large change in spectral centroid ($T_{b,horn} \approx 2.2$ and $T_{b,trombone} \approx 4.8$). The relationship between changes in the timbre parameter (α) and changes in the brightness was approximately linear (Emiroglu, 2007). Discrimination for NH listeners in quiet showed average discrimination abilities of $\alpha = 0.075$ corresponding to brightness discrimination of approximately 0.2.

Demany and Semal (1993) found discrimination for a centre frequency, F_c, of a spectral envelope, which corresponded approximately to the tone's spectral centroid. The average of three participants' discrimination abilities for approaches from brighter and less bright values was found to be 26.29 Hz. This value translates to a brightness of 0.066 for the fundamental frequency of 400 Hz.

The current work shows that an average JND of 0.049 was achieved by NH listeners. This is lower than results found by Emiroglu and Kollmeier (2008), but very close to that of Demany and Semal (1993). It is possible that the real-world sounds used by Emiroglu and Kollmeier, rather than the complex synthesized sounds used by Demany and Semal contributed to the increased JND. Table 7.1 contains a summary of results from brightness discrimination studies.

Table 7.1. Summary of brightness discrimination.

Study	Average brightness discrimination
Emiroglu and Kollmeier (2008)	0.2
Demany and Semal (1993)	0.066
Current study	0.049

A trend was observed in that brightness JNDs increased as brightness increased. A similar trend was also observed by Emiroglu (2007) and Gunawan and Sen (2008) noted that listeners were more sensitive to changes in the lower frequencies of a sound.

The significantly higher JNDs obtained by some of the CI listeners were mainly due to their inabilities to discriminate brightness associated with high brightness instruments. The differences in their brightness discrimination abilities of low and high brightness instruments also caused large standard deviations. This observation did not apply to all CI listeners, indicating listener-specific difficulties involving high frequency components.

7.3 IRREGULARITY JNDS

Trends indicated that irregularity JNDs became larger with increasing brightness as well as decreasing n-values. As n-values decrease, they are essentially perceived as shorter in duration. This is particularly noticeable for negative n-values. Anecdotal reports support the notion that spectral characteristics of shorter duration tones are more difficult to process, leading to larger JNDs. Although studies have investigated the perception of local spectral deviations, direct comparisons cannot be done to the irregularity JNDs obtained here. The implementation of irregularity in the current work is incorporated into the mathematical timbre model, while most of the literature manipulated individual harmonics (Caclin et al., 2005; Gabrielsson and Sjögren, 1971).

7.4 LOGARITHMIC RISE-TIME JNDS

In an experiment using varying linear rise and decay times with a steady state portion of 400 ms, discrimination thresholds of the RTs were estimated to be between approximately 1 ms and 17 ms for linear RTs of 1 ms to 100 ms. A repeat of the experiment using a 200 ms steady state portion showed discrimination values between 1 ms and 18 ms. The decay time discrimination was found to be between 1 ms and 17 ms and 1 ms and 20 ms for the two stimuli. Discrimination thresholds were generally found to increase as the RTs increased (Van Heuven and Van den Broecke, 1979).

A filtered sawtooth wave at 300 Hz with linear RTs estimated JNDs between 3 ms and 26 ms for individuals listening to linear RTs between 10 ms and 80 ms in 10 ms increments. The amplitude envelope consisted of a linear RT followed by an immediate linear decay with stimulus duration of 1 s (Kewley-Port and Pisoni, 1984).

Smurzyński and Houtsma (1989) used 1 kHz pure tones with linear RTs between 10 ms and 60 ms, 40 ms decay times and total duration of 256 ms to obtain discrimination thresholds of 1 ms to 4 ms.

Table 7.2 summarises the results of LRT studies. The current JNDs obtained during experimentation are lower than most data indicated in the table, but are comparable to the data found by Smurzyński and Houtsma (1989).

Table 7.2. Summary of approximate RT discrimination.

Study (year)	RT range (LRT)	Average RTs discrimination range (LRT)
Van Heuven and Van den Broecke (1979)	1 ms – 100 ms (0 – 2)	1 ms – 18 ms (0 – 1.26)
Kewley-Port and Pisoni (1984)	10 ms – 80 ms (1 – 1.90)	3 ms – 26 ms (0.477 – 1.41)
Smurzyński and Houtsma (1989)	10 ms – 60 ms (1 – 1.78)	1 ms – 4 ms (0 – 0.602)
Current study	15 ms – 550 ms (1.18 – 2.74)	1 ms – 2 ms (0.005 – 0.3)

The studies summarized in Table 7.2 found that the JND of the LRT increases as the LRT increases. However, this effect was not clearly observed in the experimental data.

7.5 SUSTAIN/DECAY JNDS

Although a direct comparison of the decay times found in the literature and the decay values that use the variable, n, could not be done, the times where the temporal envelope decayed from 90% of the maximum value to 10% of the maximum value for the various n-values were used for comparison. For decay type instruments, decay times of approximately 110 ms and 290 ms (corresponding to $n = -20$ to $n = -8$) were used and for sustain type instruments, decay times of approximately 120 ms to 960 ms (corresponding to $n = 2$ to $n = 18$) were used. Discrimination was found to vary between 2.7 ms and 62 ms. Van Heuven and Van den Broecke (1979) used linear decays, which correspond to $n \to 0$. Table 7.3 shows the summary of decay values in the literature.

Table 7.3. Summary of approximate decay time discrimination.

Study (year)	Decay time range	Average decay times discrimination range
Van Heuven and Van den Broecke (1979)	1 ms – 100 ms	1 ms – 20 ms
Current study	120 ms – 960 ms	2.7 ms – 62 ms

7.6 CONFUSION MATRICES

Application of Svirsky's model (2000) indicates that essentially no confusion will exist between instruments for NH listeners (Table 6.1). Confusion rates for CI listeners (Table 6.2) are also lower than one would expect, since experimental studies do not confirm such findings. Instruments are confused with one another, especially within instrument families (Gfeller et al., 2002). The recognition study by Gfeller et al. (2002) found 90.9% correct predictions for NH listeners and 46.6% correct for CI listeners, while the confusion matrices here suggest 100% correct prediction for NH listeners and 96.5% for CI listeners. Applied to instrument recognition or identification, Svirsky's model is not an accurate prediction. In its original form, it may be a better indication of whether participants will regard a pair of sounds as being the same or different.

Familiarity and expectations are known to influence perception (Kishon-Rabin et al., 2001; Puri and Wojciulik, 2008). Svirsky's model was proposed for phoneme recognition. Language is an everyday stimulus and the average person is quite familiar with linguistic stimuli. Musical stimuli are not necessarily encountered in everyday life. Compared to language, most people therefore have a limited knowledge of instruments. In general, CI users are even less familiar with instruments compared to NH listeners. It is possible that Svirsky's model, when applied to instrument recognition, may only be valid for participants with a comprehensive knowledge of instruments. A possible adaptation to the model for instrument identification prediction may be to multiply all JNDs with some constant value, since some of the timbre parameters within instrument families are similar. For example, brass instruments have similar irregularity and temporal envelopes, but a variety of brightness values. The single reed saxophone and clarinet also have similar temporal envelopes and similar brightness values, but very different irregularities.

7.7 FEATURE INFORMATION TRANSMISSION ANALYSIS

The brightness and LRT appear to carry the most information. The S/D appears to carry relatively little information, which may at first appear unexpected. The S/D parameter quite clearly distinguishes between two types of instruments: sustain or decay. However, within either of these categories, especially towards the extremes, very little information is transmitted compared to other parameters.

Irregularity also transmits less information compared to brightness. With the theoretical limits imposed on irregularity ($0 \leq \text{IRR} \leq 2$) and practical values used of approximately $0.1 < \text{IRR} \leq 1.5$ average JNDs of 0.056 and 0.100 are found for NH and CI listeners respectively. Theoretical limits of brightness ($T_b \geq 1$) and values used between approximately $1 < T_b < 13$ yield average JNDs of 0.049 and 0.74 for NH and CI listeners respectively. It is therefore reasonable to believe that brightness would be a more salient cue compared to irregularity, since a larger number of brightness steps can be discriminated. Visual inspection of Figure A2 in the Appendix section C supports this idea. The results from the FITA provides important information regarding each individual timbre dimension as well as the relative salience of the dimensions and should be held in consideration during future occasions where representative timbre sets are chosen.

CHAPTER 8 CONCLUSION

8.1 SUMMARY OF WORK

A literature study on general music and timbre perception and CI music and timbre perception provide a basis and motivation for a deeper investigation into the properties of timbre. Synthesis methods was summarised in the literature study since synthetic tones provided the control necessary for individual timbre property study. A description of the chosen synthesis method is found in section 3.3 and provides a solution to the research question regarding an applicable synthesis method. This method provides the control necessary for systematic discrimination testing. The parameters investigated here are salient contributions to music, although irregularity may not be as simple as it is defined in the literature. These parameters conclude the question of which parameters are necessary salient timbre properties to include in a synthesis model.

Balancing experiments (Chapter 4) indicated that loudness must be balanced for, but that pitch and perceived duration differences are negligible. Systematic balancing methods and equations are provided. The experimental procedure is described in Chapter 5. Statistical analyses of the discrimination results are given in Chapter 6. The method for the creation of instrument confusion matrices and the FITA performed on these matrices are also provided. Chapter 7 discusses the findings with regard to the literature.

8.2 FINDINGS

JNDs found for NH and CI listeners indicated that CI users performed significantly poorer for all parameters, but that CI listeners perceive temporal parameters better than spectral parameters. This finding that the temporal resolution of CI listeners is closer to that of NH listeners compared to spectral resolution is supported by the literature (Drennan and Rubinstein, 2008; Kong et al., 2004; Galvin et al., 2007; Limb, 2006a). Furthermore, trends were observed in that brightness and irregularity JNDs increased as brightness increased and that this effect was more noticeable for CI listeners. Shorter tones (more negative n-values) were also associated with increased difficulty in discrimination, resulting in larger JNDs. It must be noted that certain CI users obtained JNDs very close to NH listeners for different parameters. This appears to indicate that current processors can transmit information relatively well, but that limited timbre perception may be due to participant-specific factors. For example, during brightness discrimination tasks, S15, S22 and S28 performed comparable to NH listeners across the entire range of instruments, while S24 performed comparable to NH listeners only for instruments with low brightness values. Participant differences could possibly be ascribed to factors such as mapping or electrode insertion depth. CI participants performed reasonably well in the S/D segment discrimination, which is a temporal parameter. Average JNDs were only somewhat higher compared to NH listeners. S23 consistently obtained JNDs that were higher than other listeners and differences were often significant. The poor spectral resolution of higher frequencies and spectral fine structure may contribute to poor timbre perception.

Few interaction effects were observed for participant groups (NH and CI) or across instruments. This means that all listeners responded similarly to instruments, even though the CI group had much higher JNDs. The findings provide the result to the final research question regarding the JNDs of NH and CI listeners.

Svirsky's model (2000) of predicting phoneme recognition is a useful prediction model. Using the model as an instrument recognition model does not accurately portray the expected confusion matrix of instrument recognition. This may be due to less familiarity and knowledge of the stimulus set. The model used in the current context may be a better indication of whether a pair of sounds would be identified as two different sounds or instruments or a repetition of the same sound or instrument.

The FITA indicates that the LRT and the T_b transmitted the most of information and that the amount of information transmitted to certain CI users was comparable to that of the NH listeners. Other CI listeners received limited information. The irregularity transmitted less information than the LRT and T_b, and an alternative description of irregularity may improve information transmission, as well as improved the synthesis model's capacity to produce more realistic instrument sounds.

8.3 FUTURE WORK

Additional timbre parameters included in the synthesis methods may improve the fidelity of the experimental stimuli and provide information on how these timbre parameters are perceived.

A more systematic approach may be beneficial to establish the influence of different parameters on a specific timbre parameter. The experiment here used thirteen stimuli for NH listeners and nine for CI users, based on the parameters of existing instruments. An approach that varies parameters in increments across the entire timbre space, without necessarily representing any specific instrument, will allow better investigation of parameter interactions (for example the effect of tone length on spectral parameter discrimination or the effect of brightness on irregularity discrimination).

Improvement in spectral resolution and especially high frequency spectral resolution and participant specific limitations appear to be major limitations of timbre perception. Improvements in spectral resolution may be limited, but addressing participant specific limitation of high frequency spectral resolution may improve timbre perception.

Alternative prediction models or adaptations of Svirsky's current model, which incorporate limited knowledge and familiarity of the stimulus set, should be investigated in an attempt to produce a more accurate prediction model.

The present study reveals much about the perception of timbre and provides a basis to expand our knowledge. The future work topics may increase our understanding of timbre perception and provide valuable information in the pursuit of CI improvement.

APPENDIX

A. PSYCHOACOUSTIC CURVE OF SUSTAIN/DECAY PARAMETER N

It has been established that the psychoacoustic curves of T_b, IRR and LRT are approximately linear. The perception of the parameter n has not yet been investigated. Pure tones at C4 (262 Hz), with a sampling rate of 44.1 kHz, a LRT of 1 and a total duration of 2 s were presented to NH listeners. It was hypothesised that the log of the perception of the log of the n-value would be approximately linear. Tones with logarithmically spaced negative n-values of $n = -10^x$, with $x = [1.8, 1.6, 1.4 \ldots - 1.4, -1.6, -1.8]$ and logarithmically spaced positive n-values of $n = 10^x$, where $x = [-1.8, -1.6, -1.4 \ldots 1.4, 1.6, 1.8]$ were presented. These correspond to values between $n \approx -60$ to $n \approx 0$ and $n \approx 0$ to $n \approx 60$. Three sequential tones were presented, with the first tone corresponding to $n = -10^{-1.8}$ and the last tone corresponding to $n = 10^{1.8}$. The middle tone was the test tone and listeners were asked to point and click on a visual bar indicating how much the test tone corresponded to the first tone (represented by the left end of the bar) or the last tone (represented by the right end of the bar). Tones were presented at the 75% loudness level of the participants' psychoacoustic curve (see Appendix section B), unless listeners preferred higher or lower loudness levels. All tones were presented at loudness levels between 68.2 dBSPL and 76 dBSPL.

Figure A1 shows the average response of six NH listeners when presented with n-values. The logarithm of the response is found to be approximately linear. The parameter, n, will therefore be spaced logarithmically during tests.

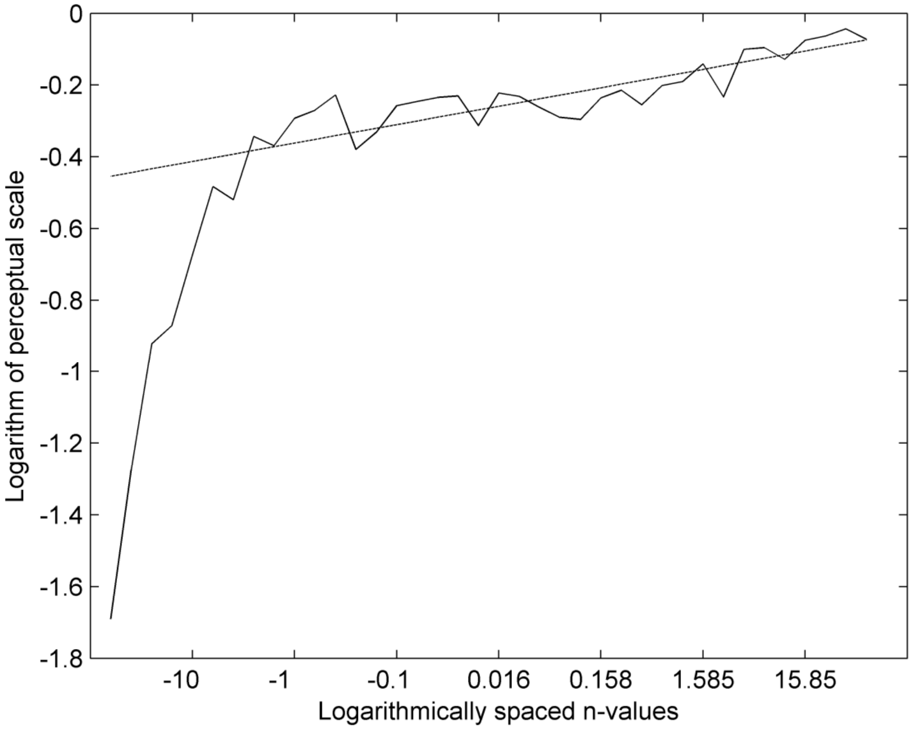

Figure A1. The logarithm of the average perceptual response of six listeners indicates that logarithmically spaced n-values are approximately a perceptually linear response.

B. LOUDNESS GROWTH CURVE PROCEDURE

The psychoacoustic loudness curve was obtained for each participant by presenting 262 Hz tones sampled at 44.1 kHz at intensities between 25 dBSPL and 95 dBSPL. At the start of the test, the maximum intensity (95 dBSPL) and minimum intensity (25 dBSPL) were presented to the listener as references. Thereafter, intensities were presented in 2 dBSPL increments between the two extreme values. Intensities were presented in random order and the participant was asked to assign a value between 0 and 100 according to the loudness of the sound. Each of the intensities was presented twice during the test and the listener had to respond by typing a value into a text box. The 75% point of the curve was then found and used as a starting intensity for balancing and discrimination tests. Before commencement of each of the tests, the participants were asked if the sounds presented at the 75% point of their loudness growth curve were at comfortable listening levels and adjustments were made if desired.

C. SET REDUCTION FOR CI USERS

To reduce the number of instruments used for CI participant testing, the JNDs were found for NH participants. Figure A2 and Figure A3 show how each instrument represents a point in the four-dimensional timbre space surrounded by an ellipse, which represents the JNDs of the specific instrument in the respective timbre dimensions. In order to reduce the set of instruments while retaining a good representation of the entire timbre space, it is necessary not only to examine the distances between instruments, but also the areas of confusion surrounding each instrument. For example, the absolute distance between the spectral properties of the oboe and violin is approximately equal to the distance between the piano and tuba. However, the spectral JNDs of the violin and oboe are greater than that of the piano and tuba and would probably be confused more often. It would therefore be more reasonable to eliminate one instrument in the oboe-violin pair rather than the piano-tuba pair in order to retain a good representation of the spectral timbre space.

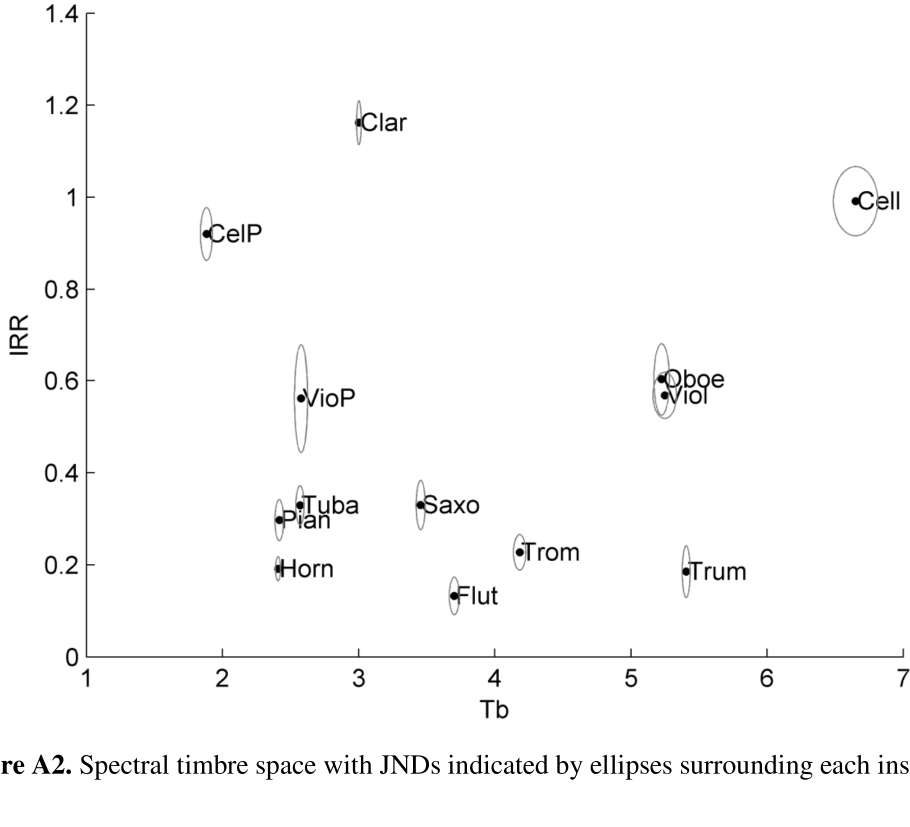

Figure A2. Spectral timbre space with JNDs indicated by ellipses surrounding each instrument.

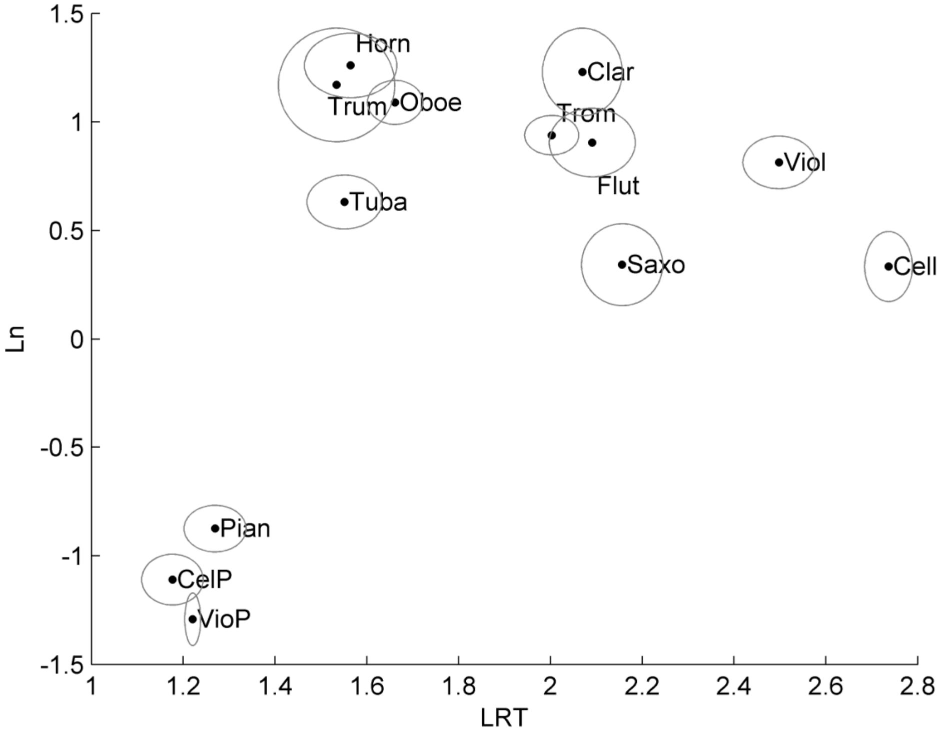

Figure A3. Temporal timbre space with JNDs indicated by ellipses surrounding each instrument.

In order to find the instruments that are close together in the timbre space, the following approach was used. A Gaussian envelope was placed along each axis with a mean equal to the position of the instrument in that dimension and a standard deviation equal to the JND (Svirsky, 2000). The degree of overlap for all possible pairs of instruments for each of the timbre parameters was calculated by taking the absolute of the difference of the two curves and finding the area under the resulting curve. The average of the four distances between each instrument pair was calculated and taken as the distance in timbre space. It was decided to reduce the dataset from thirteen to nine instruments. Even though the multiplication of all four dimensions to find the distance in timbre space appeared to be a more intuitive method, the average value of the four parameters was preferred in order to eliminate the weight a single dimension might contribute.

Table A1, Table A2, Table A3 and Table A4, are the distance tables associated with each of the four parameters. Table A5 and Table A6 respectively indicate the distances in the spectral and temporal domains and Table A7 is the distance table of the entire set of parameters. The distance tables are essentially normalised between 0 and 2, where 0 represents a complete overlap of instruments and 2 that the instruments do not overlap at all.

The bowed violin sound was timbrally very similar to that of the oboe, except for the LRT values in which a larger distance was apparent. The bowed violin sound was eliminated. The saxophone and trumpet respectively occupied approximately the same spectral and temporal timbre space as the tuba, so the tuba was eliminated from the set. Furthermore, the instrument set still contained three other brass instruments. The piano and pizzicato violin were paired and the pizzicato violin was eliminated from the set. Even though many other instruments were spectrally closer to the piano, the characteristic decay of both instruments supported the appropriateness of the pairing. The flute was eliminated due to the fact that the C4 recording was close to its lower limit of the instrument’s playing range and it has been noted that identification or instruments and their properties may be influenced by this (Grey, 1977). Its location in timbre space was well represented by the trumpet in the spectral domain and the clarinet and trombone in the temporal domain.

Table A1. Distance table of instruments with regard to brightness. Clarinet (Clar); French Horn (Horn); Oboe (Oboe) Arco (bowed) Violin (ViolA); Pizzicato (plucked) Violin (ViolP); Flute (Flut); Piano (Pian); Trumpet (Trum); Tuba (Tuba); Arco (bowed) Cello (CellA); Pizzicato (plucked) Cello (CellP); Saxophone (Saxo); Trombone (Trom).

T_b	Clar	Horn	Oboe	ViolA	ViolP	Flut	Pian	Trum	Tuba	CellA	CellP	Saxo	Trom
Clar	0	2	2	2	2	2	2	2	2	2	2	2	2
Horn	2	0	2	2	1.978	2	0.556	2	1.998	2	2	2	2
Oboe	2	2	0	0.455	2	2	2	1.933	2	2	2	2	2
ViolA	2	2	0.455	0	2	2	2	1.708	2	2	2	2	2
ViolP	2	1.978	2	2	0	2	1.912	2	0.46	2	2	2	2
Flut	2	2	2	2	2	0	2	2	2	2	2	1.999	2
Pian	2	0.556	2	2	1.912	2	0	2	1.972	2	2	2	2
Trum	2	2	1.933	1.708	2	2	2	0	2	2	2	2	2
Tuba	2	1.998	2	2	0.46	2	1.972	2	0	2	2	2	2
CellA	2	2	2	2	2	2	2	2	2	0	2	2	2
CellP	2	2	2	2	2	2	2	2	2	2	0	2	2
Saxo	2	2	2	2	2	1.999	2	2	2	2	2	0	2
Trom	2	2	2	2	2	2	2	2	2	2	2	2	0

Table A2. Distance table of instruments with regard to irregularity. Clarinet (Clar); French Horn (Horn); Oboe (Oboe) Arco (bowed) Violin (ViolA); Pizzicato (plucked) Violin (ViolP); Flute (Flut); Piano (Pian); Trumpet (Trum); Tuba (Tuba); Arco (bowed) Cello (CellA); Pizzicato (plucked) Cello (CellP); Saxophone (Saxo); Trombone (Trom).

IRR	Clar	Horn	Oboe	ViolA	ViolP	Flut	Pian	Trum	Tuba	CellA	CellP	Saxo	Trom
Clar	0	2	2	2	1.999	1.999	2	1.999	2	1.683	1.956	2	2
Horn	2	0	2	2	1.984	1.263	1.739	0.708	1.911	2	2	1.847	0.88
Oboe	2	2	0	0.569	0.49	1.999	1.976	1.996	1.956	1.978	1.961	1.926	1.998
ViolA	2	2	0.569	0	0.782	1.999	1.991	1.999	1.98	1.999	1.998	1.957	2
ViolP	1.999	1.984	0.49	0.782	0	1.988	1.817	1.942	1.744	1.95	1.923	1.678	1.944
Flut	1.999	1.263	1.999	1.999	1.988	0	1.89	0.853	1.962	1.999	1.999	1.926	1.527
Pian	2	1.739	1.976	1.991	1.817	1.89	0	1.47	0.57	2	2	0.54	1.201
Trum	1.999	0.708	1.996	1.999	1.942	0.853	1.47	0	1.713	1.999	1.999	1.625	0.738
Tuba	2	1.911	1.956	1.98	1.744	1.962	0.57	1.713	0	2	2	0.219	1.581
CellA	1.683	2	1.978	1.999	1.95	1.999	2	1.999	2	0	0.841	2	2
CellP	1.956	2	1.961	1.998	1.923	1.999	2	1.999	2	0.841	0	2	2
Saxo	2	1.847	1.926	1.957	1.678	1.926	0.54	1.625	0.219	2	2	0	1.48
Trom	2	0.88	1.998	2	1.944	1.527	1.201	0.738	1.581	2	2	1.48	0

Table A3. Distance table of instruments with regard to LRT. Clarinet (Clar); French Horn (Horn); Oboe (Oboe) Arco (bowed) Violin (ViolA); Pizzicato (plucked) Violin (ViolP); Flute (Flut); Piano (Pian); Trumpet (Trum); Tuba (Tuba); Arco (bowed) Cello (CellA); Pizzicato (plucked) Cello (CellP); Saxophone (Saxo); Trombone (Trom).

LRT	Clar	Horn	Oboe	ViolA	ViolP	Flut	Pian	Trum	Tuba	CellA	CellP	Saxo	Trom
Clar	0	1.986	1.989	1.981	2	0.198	2	1.976	1.996	2	2	0.756	0.761
Horn	1.986	0	0.969	2	1.759	1.986	1.687	0.294	0.228	2	1.851	1.996	1.988
Oboe	1.989	0.969	0	2	1.971	1.99	1.962	1.123	1.142	2	1.988	1.998	1.992
ViolA	1.981	2	2	0	2	1.963	2	2	2	1.869	2	1.919	1.999
ViolP	2	1.759	1.971	2	0	2	0.35	1.587	1.799	2	0.296	2	2
Flut	0.198	1.986	1.99	1.963	2	0	2	1.977	1.996	2	2	0.559	0.931
Pian	2	1.687	1.962	2	0.35	2	0	1.484	1.731	2	0.645	2	2
Trum	1.976	0.294	1.123	2	1.587	1.977	1.484	0	0.435	2	1.714	1.992	1.978
Tuba	1.996	0.228	1.142	2	1.799	1.996	1.731	0.435	0	2	1.883	1.999	1.997
CellA	2	2	2	1.869	2	2	2	2	2	0	2	2	2
CellP	2	1.851	1.988	2	0.296	2	0.645	1.714	1.883	2	0	2	2
Saxo	0.756	1.996	1.998	1.919	2	0.559	2	1.992	1.999	2	2	0	1.421
Trom	0.761	1.988	1.992	1.999	2	0.931	2	1.978	1.997	2	2	1.421	0

Table A4. Distance table of instruments with regard to decay. Clarinet (Clar); French Horn (Horn); Oboe (Oboe) Arco (bowed) Violin (ViolA); Pizzicato (plucked) Violin (ViolP); Flute (Flut); Piano (Pian); Trumpet (Trum); Tuba (Tuba); Arco (bowed) Cello (CellA); Pizzicato (plucked) Cello (CellP); Saxophone (Saxo); Trombone (Trom).

n	Clar	Horn	Oboe	ViolA	ViolP	Flut	Pian	Trum	Tuba	CellA	CellP	Saxo	Trom
Clar	0	0.314	0.887	1.621	2	1.275	2	0.309	1.872	1.973	2	1.954	1.426
Horn	0.314	0	1.037	1.804	2	1.504	2	0.607	1.957	1.994	2	1.987	1.652
Oboe	0.887	1.037	0	1.572	2	1.086	2	0.898	1.916	1.992	2	1.981	1.137
ViolA	1.621	1.804	1.572	0	2	0.55	2	1.366	1.084	1.823	2	1.75	0.911
ViolP	2	2	2	2	0	2	2	2	2	2	1.951	2	2
Flut	1.275	1.504	1.086	0.55	2	0	2	1.004	1.342	1.852	2	1.791	0.552
Pian	2	2	2	2	2	2	0	2	2	2	1.836	2	2
Trum	0.309	0.607	0.898	1.366	2	1.004	2	0	1.701	1.907	2	1.87	1.221
Tuba	1.872	1.957	1.916	1.084	2	1.342	2	1.701	0	1.411	2	1.31	1.698
CellA	1.973	1.994	1.992	1.823	2	1.852	2	1.907	1.411	0	2	0.155	1.968
CellP	2	2	2	2	1.951	2	1.836	2	2	2	0	2	2
Saxo	1.954	1.987	1.981	1.75	2	1.791	2	1.87	1.31	0.155	2	0	1.939
Trom	1.426	1.652	1.137	0.911	2	0.552	2	1.221	1.698	1.968	2	1.939	0

Table A5. Distance table of instruments with regard to the spectral domain. Clarinet (Clar); French Horn (Horn); Oboe (Oboe) Arco (bowed) Violin (ViolA); Pizzicato (plucked) Violin (ViolP); Flute (Flut); Piano (Pian); Trumpet (Trum); Tuba (Tuba); Arco (bowed) Cello (CellA); Pizzicato (plucked) Cello (CellP); Saxophone (Saxo); Trombone (Trom).

Spectral	Clar	Horn	Oboe	ViolA	ViolP	Flut	Pian	Trum	Tuba	CellA	CellP	Saxo	Trom
Clar	0	2	2	2	2	2	2	2	2	1.842	1.978	2	2
Horn	2	0	2	2	1.981	1.631	1.148	1.354	1.954	2	2	1.924	1.44
Oboe	2	2	0	0.512	1.245	2	1.988	1.964	1.978	1.989	1.981	1.963	1.999
ViolA	2	2	0.512	0	1.391	2	1.996	1.853	1.99	1.999	1.999	1.978	2
ViolP	2	1.981	1.245	1.391	0	1.994	1.864	1.971	1.102	1.975	1.962	1.839	1.972
Flut	2	1.631	2	2	1.994	0	1.945	1.426	1.981	2	2	1.963	1.764
Pian	2	1.148	1.988	1.996	1.864	1.945	0	1.735	1.271	2	2	1.27	1.6
Trum	2	1.354	1.964	1.853	1.971	1.426	1.735	0	1.856	2	2	1.812	1.369
Tuba	2	1.954	1.978	1.99	1.102	1.981	1.271	1.856	0	2	2	1.11	1.79
CellA	1.842	2	1.989	1.999	1.975	2	2	2	2	0	1.42	2	2
CellP	1.978	2	1.981	1.999	1.962	2	2	2	2	1.42	0	2	2
Saxo	2	1.924	1.963	1.978	1.839	1.963	1.27	1.812	1.11	2	2	0	1.74
Trom	2	1.44	1.999	2	1.972	1.764	1.6	1.369	1.79	2	2	1.74	0

Table A6. Distance table of instruments with regard to the temporal domain. Clarinet (Clar); French Horn (Horn); Oboe (Oboe) Arco (bowed) Violin (ViolA); Pizzicato (plucked) Violin (ViolP); Flute (Flut); Piano (Pian); Trumpet (Trum); Tuba (Tuba); Arco (bowed) Cello (CellA); Pizzicato (plucked) Cello (CellP); Saxophone (Saxo); Trombone (Trom).

Temporal	Clar	Horn	Oboe	ViolA	ViolP	Flut	Pian	Trum	Tuba	CellA	CellP	Saxo	Trom
Clar	0	1.15	1.438	1.801	2	0.736	2	1.142	1.934	1.986	2	1.355	1.093
Horn	1.15	0	1.003	1.902	1.88	1.745	1.844	0.45	1.093	1.997	1.925	1.992	1.82
Oboe	1.438	1.003	0	1.786	1.986	1.538	1.981	1.011	1.529	1.996	1.994	1.99	1.564
ViolA	1.801	1.902	1.786	0	2	1.257	2	1.683	1.542	1.846	2	1.834	1.455
ViolP	2	1.88	1.986	2	0	2	1.175	1.793	1.899	2	1.123	2	2
Flut	0.736	1.745	1.538	1.257	2	0	2	1.49	1.669	1.926	2	1.175	0.742
Pian	2	1.844	1.981	2	1.175	2	0	1.742	1.865	2	1.241	2	2
Trum	1.142	0.45	1.011	1.683	1.793	1.49	1.742	0	1.068	1.954	1.857	1.931	1.6
Tuba	1.934	1.093	1.529	1.542	1.899	1.669	1.865	1.068	0	1.706	1.942	1.655	1.848
CellA	1.986	1.997	1.996	1.846	2	1.926	2	1.954	1.706	0	2	1.078	1.984
CellP	2	1.925	1.994	2	1.123	2	1.241	1.857	1.942	2	0	2	2
Saxo	1.355	1.992	1.99	1.834	2	1.175	2	1.931	1.655	1.078	2	0	1.68
Trom	1.093	1.82	1.564	1.455	2	0.742	2	1.6	1.848	1.984	2	1.68	0

Table A7. Distance table of instruments with regard to all four timbre parameters. Clarinet (Clar); French Horn (Horn); Oboe (Oboe) Arco (bowed) Violin (ViolA); Pizzicato (plucked) Violin (ViolP); Flute (Flut); Piano (Pian); Trumpet (Trum); Tuba (Tuba); Arco (bowed) Cello (CellA); Pizzicato (plucked) Cello (CellP); Saxophone (Saxo); Trombone (Trom).

Complete	Clar	Horn	Oboe	ViolA	ViolP	Flut	Pian	Trum	Tuba	CellA	CellP	Saxo	Trom
Clar	0	1.575	1.719	1.9	2	1.368	2	1.571	1.967	1.914	1.989	1.677	1.547
Horn	1.575	0	1.501	1.951	1.93	1.688	1.496	0.902	1.524	1.999	1.963	1.958	1.63
Oboe	1.719	1.501	0	1.149	1.615	1.769	1.985	1.487	1.753	1.992	1.987	1.976	1.781
ViolA	1.9	1.951	1.149	0	1.695	1.628	1.998	1.768	1.766	1.923	1.999	1.906	1.728
ViolP	2	1.93	1.615	1.695	0	1.997	1.52	1.882	1.501	1.988	1.542	1.92	1.986
Flut	1.368	1.688	1.769	1.628	1.997	0	1.973	1.458	1.825	1.963	2	1.569	1.253
Pian	2	1.496	1.985	1.998	1.52	1.973	0	1.739	1.568	2	1.62	1.635	1.8
Trum	1.571	0.902	1.487	1.768	1.882	1.458	1.739	0	1.462	1.977	1.928	1.872	1.484
Tuba	1.967	1.524	1.753	1.766	1.501	1.825	1.568	1.462	0	1.853	1.971	1.382	1.819
CellA	1.914	1.999	1.992	1.923	1.988	1.963	2	1.977	1.853	0	1.71	1.539	1.992
CellP	1.989	1.963	1.987	1.999	1.542	2	1.62	1.928	1.971	1.71	0	2	2
Saxo	1.677	1.958	1.976	1.906	1.92	1.569	1.635	1.872	1.382	1.539	2	0	1.71
Trom	1.547	1.63	1.781	1.728	1.986	1.253	1.8	1.484	1.819	1.992	2	1.71	0

The reduced set consists of the clarinet, French horn, oboe, piano, trumpet, tuba, bowed and plucked cello, saxophone and trombone.

REFERENCES

Abel, S. M. (1972). Duration Discrimination of Noise and Tone Bursts, *Journal of the Acoustical Society of America* **51**(4):219-1223.

Bilbao, S. and Fitch, J. (2006).Prepared Piano Sound Synthesis, in *Proceedings of the 9th International Conference on Digital Audio Effects.* Montreal, Canada. pp. DAFX-77-DAFX-82.

Burmeister, B. (2008). *Cue estimation for vowel perception prediction in low signal-to-noise ratios,* MEng dissertation, University of Pretoria. Available from http://upetd.up.ac.za/thesis/available/etd-05132009-151828/.

Caclin, A., McAdams, S., Smith, B. K. and Winsberg, S. (2005). Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones, *Journal of the Acoustical Society of America* **118**(1):471-482.

Chowning, J. (1973). The Synthesis of Complex Audio Spectra by Means of Frequency Modulation, *Journal of the Audio Engineering Society* **21**(7):526-534.

Cook, P. (1992). A Meta-Wind-Instrument Physical Model, and a Meta-Controller for Real-Time Performance Control, in *Proceedings of the International Computer Music Conference.* San Jose, San Francisco, USA, pp. 273.

Cook, P. R. (2002). Sound Production and Modeling, *IEEE Computer Graphics and Applications* **22**(4):23-27.

Cusack, R. and Roberts, B. (2004).Effects of differences in the pattern of amplitude envelopes across harmonics on auditory stream segregation, *Hearing Research,* **193**:95-104.

De Poli, G. (1983). A Tutorial on Digital Sound Synthesis, *Computer Music Journal* **7**(4):8-26.

Demany, L. and Semal, C. (1993). Pitch versus Brightness of Timbre: Detecting Combined Shifts in Fundamental and Formant Frequency, *Music Perception* **11**(1):1-14.

Drennan, W. R. and Rubinstein, J. T. (2008). Music perception in cochlear implant users and its relationship with psychophysical capabilities, *Journal of Rehabilitation Research & Development* **45**(5):779-790.

Emiroglu, S. (2007). *Timbre perception and object separation with normal and impaired hearing*, Ph.D. dissertation, Carl-von-Ossietzky-Universität.

Emiroglu, S. and Kollmeier, B. (2008). Timbre discrimination in normal-hearing and hearing-impaired listeners under different noise conditions, *Brain Research* **1220**:199-207.

Eronen, A. and Klapuri, A. (2000). Musical Instrument Recognition using Cepstral Coefficients and Temporal Features, in *Proceedings of the 2000 IEEE International Conference on Acoustics, Speech and Signal Processing,* Istanbul, Turkey, pp. II753-II756.

Fletcher, H. (1971). Normal Vibration Frequencies of a Stiff Piano String, *Journal of the Acoustical Society of America* **52**(2):471-483.

Fletcher, N. H. and Rossing, T. D. (1999). *The Physics of Musical Instruments,* 2nd Ed. Springer-Verlag, New York. pp. 40-50, 56-67, 366-374, 383-384, 438, 453-455, 490-494, 546-548.

Fritts, L. (1997), *The University of Iowa Electronic Music Studios* [Online]. Available from: http://theremin.music.uiowa.edu/index.html [Accessed: 21 March, 2010].

Gabrielsson, A. and Sjögren, H. (1971). Detection of Amplitude Distortion in Flute and Clarinet Spectra, *Journal of the Acoustical Society of America* **52**(2):471-483.

Galvin III, J. J., Fu, Q. and Nogaki, G. (2007). Melodic Contour Identification by Cochlear Implant Listeners, *Ear and Hearing* **28**(3):302-319.

Galvin III, J. J., Fu, Q. and Oba, S. (2008). Effect of instrument timbre on melodic contour identification by cochlear implant users, *Journal of the Acoustical Society of America* **124**(4):EL189-EL195.

Galvin III, J. J., Fu, Q. and Oba, S. I. (2009a). Effect of competing instrument on melodic contour identification by cochlear implant users, *Journal of the Acoustical Society of America* **125**(3):EL98-EL103.

Galvin III, J. J., Fu, Q. and Shannon, R. V. (2009b). Melodic Contour Identification and Music Perception by Cochlear Implant Users, *The Neurosciences and Music III – Disorders and Plasticity: Annals of the New York Academy of Sciences***1169**:518-533.

Gfeller, K., Christ, A., Knutson, J., Witt, S. and Mehr, M. (2003). The Effects of familiarity and Complexity on the Appraisal of Complex Songs by Cochlear Implant Recipients and Normal Hearing Adults, *Journal of Music Therapy* **40**(2):78-112.

Gfeller, K., Knutson, J. F., Woodworth, G., Witt, S. and DeBus, B. (1998). Timbral Recognition and Appraisal by Adult Cochlear Implant Users and Normal-Hearing Adults, *Journal of the American Academy of Audiology* **9**(1):1-19.

Gfeller, K., Olzewski, C., Rychener, M., Sena, K., Knutson, J. F., Witt, S. and Macpherson, B. (2005). Recognition of "Real-World" Musical Excerpts by Cochlear Implant Recipients and Normal-Hearing Adults, *Ear and Hearing* **26**(3):237-250.

Gfeller, K., Witt, S., Woodworth, G., Mehr, M. A., and Knutson, J., (2002). Effects of

Frequency, Instrumental Family, and Cochlear Implant Type on Timbre Recognition and Appraisal, *Annals of Otology, Rhinology and Laryngology* **111**:349-356.

Grey, J. M. (1975). *An Exploration of Musical Timbre*, Ph.D. Dissertation, Stanford University.

Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres, *Journal of the Acoustical Society of America* **61**(5):1270-1277.

Grey, J. M. and Gordon, J. W. (1978). Perceptual effects of spectral modifications on musical timbres, *Journal of the Acoustical Society of America* **63**(5):1493-1500.

Guillemain, P. and Kronland-Martinet, R. (1996). Characterization of Acoustic Signals Through Continuous Linear Time-Frequency Representations, *Proceedings of the IEEE* **84**(4):561-585.

Gunawan, D. and Sen, D. (2008). Spectral envelope sensitivity of musical sounds, *Journal of the Acoustical Society of America* **123**(1):500-506.

Hartman, W. M. (2005). *Signals, Sound, and Sensation,* Springer, Michigan, pp. 297-298.

Horner, A., Beauchamp, J. and So, R. (2004). Detection of random alterations to time-varying musical instrument spectra, *Journal of the Acoustical Society of America* **116**(3):1800-1810.

Hugo, S. (2008). *Modelling and measurement of timbre perception in the electrically stimulated auditory system,* MEng dissertation, University of Pretoria. Available from http://upetd.up.ac.za/thesis/available/etd-10082010-163307/.

Jackendoff, R. and Lerdahl, F. (2006). The capacity for music: What is it, and what's special about it? *Cognition* **100**:33-72.

Jehan, T. (2001) *Perceptual Synthesis Engine: An Audio Driven Timbre Generator,* M.Sc. thesis, Massachusetts Institute of Technology.

Jensen, K. (1999a). *Timbre Models of Musical Sounds*, Ph.D. Dissertation, University of Copenhagen.

Jensen, K. (1999b). Envelope model of isolated musical sounds, in *Proceedings of the 2nd COST G-6 Workshop on Digital Audio Effects,* 9-11 December, 1999, Trondheim, Norway, pp.W99-1-W99-4.

Jensen, K. (2001). The Timbre Model, in *Workshop on Current Research Directions in Computer Music*.

Kewley-Port, D. and Pisoni, D. B. (1984). Identification and discrimination of rise-time: Is it categorical or noncategorical? *Journal of the Acoustical Society of America,* **75**(4):1168-1176.

Kishon-Rabin, L., Amit, O., Vexler, Y. and Zaltz, Y. (2001). Pitch discrimination: Are professional musicians better than non-musicians? *Journal of Basic and Clinical Physiology and Pharmacology* **12**(2):125-143.

Kong, Y., Cruz, R., Ackland Jones, J. and Zeng, F. (2004). Music Perception with Temporal Cues in Acoustic and Electric Hearing, *Ear and Hearing* **25**(2):173-185.

Krimphoff, J., McAdams, S. and Winsberg, S. (1994). Caractérisation du timbre des sons complexes. II Analysis acoustiques et quantification psychophysique, *Journal de Physicque* **4**(C5):625-628.

Limb, C. J. (2006a). Cochlear implant-mediated perception of music, *Hearing Science: Current Opinion in Otolaryngology and Head and Neck Surgery* **14**:337-340.

Limb, C. J. (2006b). Structural and Functional Neural Correlates of Music Perception, *The Anatomical Record* **288A**:435-446.

Marozeau, J., De Cheveigné, A., McAdams, S. and Winsberg, S. (2003). The dependency of timbre on fundamental frequency, *Journal of the Acoustical Society of America* **114**(5):2946-2957.

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G. and Krimphoff, J. (1995). Perceptual scaling of the synthesized musical timbre: Common dimensions, specificities and latent subject classes, *Psychological Research* **58**:177-192.

McAdams, S., Beauchamp, J. W. and Meneguzzi, S. (1999). Discrimination of musical instrument sounds resynthesized with simplified spectrotemporal parameters, *Journal of the Acoustical Society of America* **105**(2):882-897.

McDermott, H. J. (2004). Music perception with cochlear implants: A review, *Trends in Amplification* **8**(2):49-82.

McDermott, H. J. and Looi, V. (2004).Perception of complex signals, including musical sounds, with cochlear implants, *International Congress Series* **1273**:201-204.

McKay, C. M. (2005). Spectral processing in cochlear implants, *International Review of Neurobiology* **70**:473-509.

Miller, G. A. and Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants, *Journal of the Acoustical Society of America*, **27**(2):338-352.

Moré, J. J. (1977). The Levenberg-Marquardt algorithm: Implementation and theory. *Conference on Numerical Analysis,* 28 June-11 July, 1977, University of Dundee, Scotland.

New Grove Dictionary of Music and Musicians. 2nd Ed. 2001. New York: Grove.

Peretz, I. and Zatorre, R. J. (2005). Brain Organization for Music Processing, *Annual*

Review of Psychology **56**:89-114.

Plomp, R. and Steeneken, J. M. (1969). Effect of Phase on the Timbre of Complex Tones, *Journal of the Acoustical Society of America* **46**(2):409-421.

Puri, A. M. and Wojciulik, E. (2008). Expectation both helps and hinders object perception, *Vision Research* **48**:589-597.

Russo, F. A. and Thompson, W. F. (2005). An interval size illusion: The influence of timbre on the perceived size of melodic intervals, *Perception and Psychophysics* **67**(4):559-568.

Schroeder, M. R. and Strube, H. W. (1986). Flat-spectrum speech, *Journal of the Acoustical Society of America* **79**(5):1580-1583.

Shannon, R. V.,Zeng, F., Kamath, V., Wygonski, J. and Ekelid, M. (1995). Speech Recognition with Primarily Temporal Cues, *Science* **270**(5234):303-304.

Shannon, R. V. (2005). Speech and music have different requirements for spectral resolution, *International Review of Neurobiology* **70**:121-134.

Shower, E. G. and Biddulph, R. (1931). Differential pitch sensitivity of the ear, *Journal of the Acoustical Society of America* **3**(2):275-287.

Singh, P. G. and Hirsh, I. J. (1992). Influence of spectral locus and F0 changes on the pitch and timbre of complex tones, *Journal of the Acoustical Society of America* **92**(5):2650-2661.

Smurzyński, J. and Houtsma, A. J. M. (1989). Auditory discrimination of tone-pulse onsets, *Perception and Psychophysics* **45**(1):2-9.

Svirsky, M. A. (2000). Mathematical modeling of vowel perception by users of analog multichannel cochlear implants: Temporal and channel-amplitude cues, *Journal of the Acoustical Society of America* **107**(3):1521-1529.

Van Heuven, V. J. J. P. and Van den Broecke, (1979). M. P. R. Auditory discrimination of rise and decay times in tone and noise bursts, *Journal of the Acoustical Society of America* **66**(5):1308-1315.

Van Zyl, J. (2008), *Objective determination of vowel intelligibility of a Cochlear implant model*, MEng dissertation, University of Pretoria. Available from http://upetd.up.ac.za/thesis/available/etd-03082009-174318/.

Vercoe, B. L. (1985). *Csound* (5.11) [Online]. Available from: www.csound.com [Accessed: 1 April, 2010].

Vurma, A. and Ross, J. (2007). Timbre-Induced Pitch Deviations of Musical Sounds, *Journal of interdisciplinary music studies* **1**(1):33-50.

Wier, C. C., Jesteadt, W. and Green, D.M. (1977). Frequency discrimination as a function of frequency and sensation level, *Journal of the Acoustical Society of America* **61**:178-184.

Wilson, B. S. and Dorman, M. F. (2008). Cochlear Implants: A remarkable past and a brilliant future, *Hearing Research* **242**:3-21.

Zatorre, R. J., Belin, P. and Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech, *TRENDS in Cognitive Sciences* **6**(1):37-46.