THE NEURAL BASIS OF MUSICAL CONSONANCE

A thesis submitted to The University of Manchester for the degree of

Doctor of Philosophy

Faculty of Medical and Human Sciences

2014

OLIVER BONES

SCHOOL OF PSYCHOLOGICAL SCIENCES
## Table of Contents

Tables ............................................................................................................................ 9

Figures ............................................................................................................................. 9

Abbreviations .................................................................................................................... 12

Abstract ........................................................................................................................... 13

Declaration ........................................................................................................................ 14

Copyright statement ........................................................................................................ 14

Acknowledgments ............................................................................................................ 16

Chapter 1. Literature Review .......................................................................................... 17

1.1 Introduction ................................................................................................................ 17

1.2 Pitch ............................................................................................................................. 19

1.2.1 Pitch defined .......................................................................................................... 19

1.2.2 Cochlear representation of complex tones ............................................................ 20

1.2.3 Neural representation of complex tones ............................................................... 23

1.2.4 Pitch theories ........................................................................................................ 27

1.2.4.1 Place theories and temporal theories ................................................................. 27

1.2.4.2 Pattern recognition models .............................................................................. 31

1.2.4.3 Modern temporal theories .............................................................................. 33

1.2.4.4 Summary .......................................................................................................... 35
4.4 Materials and Methods ................................................................. 124

4.4.1 Participants ................................................................................ 124

4.4.2 Behavioural testing ................................................................. 125

4.4.2.1 Stimuli .................................................................................. 125

4.4.2.2 Procedure ............................................................................. 127

4.4.2.3 Analysis ............................................................................... 128

4.4.3 Electrophysiology ................................................................. 128

4.4.3.1 Stimuli and recording procedure ........................................ 128

4.4.3.2 Analysis ............................................................................... 130

4.5 Results .......................................................................................... 133

4.5.1 Individual differences in consonance preference and the harmonicity of phase locking ................................................................. 133

4.5.2 Behavioural ratings for diotic and dichotic stimuli ...................... 136

4.5.3 Harmonic salience of the FFR for diotic and dichotic stimuli ........ 138

4.6 Discussion ....................................................................................... 142

4.6.1 Individual differences in consonance preference are related to individual differences in neural temporal coding ................................................................. 142

4.6.2 Dotic presentation results in both stronger consonance preference and stronger neural representation of harmonicity for consonant intervals ....... 143
4.6.3 Harmonicity of the FFR is enhanced in the diotic condition due to the addition of components produced by cochlear interactions

4.6.4 Conclusions

4.7 Acknowledgements

4.8 References

Chapter 5. Subcortical representation of musical dyads: individual differences and neural generators

5.1 Abstract

5.2 Keywords

5.3 Introduction

5.4 Methods

5.4.1 Participants

5.4.2 Psychophysical testing

5.4.2.1 Stimuli

5.4.2.2 Procedure

5.4.2.3 Analysis

5.4.3 Frequency-following response

5.4.3.1 Stimuli and recording protocol

5.4.3.2 Analysis
5.4.4 Auditory model........................................................................................................167

5.5. Results..........................................................................................................................168

5.5.1 The effect of high-pass masking noise on pleasantness ratings of low-frequency dyads .................................................................................................................................168

5.5.2 The effect of high-pass masking noise on the FFR to low-frequency dyads ..................................................................................................................................................170

5.5.3 The relation between NCI and individual preference for consonance for non-musicians ..............................................................................................................................................174

5.5.4 Comparison of the effects of high-pass masking noise on the FFR and a model of the inner hair cell receptor-potential ........................................................................................................177

5.6 Discussion ......................................................................................................................181

5.6.1 Neural harmonic salience of the FFR and musical consonance perception of non-musicians ............................................................................................................................................181

5.6.2 The FFR to low-frequency musical dyads is likely to be partly generated in the basal region of the cochlea .................................................................................................................183

5.6.3 Conclusions ................................................................................................................185

5.7 References .....................................................................................................................186

Chapter 6. Losing the music: Aging affects the perception and subcortical neural representation of musical harmony .........................................................................................................................190

6.1 Abstract ..........................................................................................................................191
6.2 Introduction .................................................................................................................. 192
6.3 Methods ....................................................................................................................... 194
   6.3.1 Participants ............................................................................................................. 194
   6.3.2 Behavioural stimuli .............................................................................................. 196
   6.3.3 Behavioural procedure .......................................................................................... 197
   6.3.4 The frequency-following response (FFR): recording protocol and stimuli .............. 199
   6.3.5 FFR analysis ........................................................................................................ 200
6.4 Results ......................................................................................................................... 201
   6.4.1 The effect of age on consonance perception ......................................................... 201
   6.4.2 The effect of age on neural harmonic salience ..................................................... 207
6.5 Discussion .................................................................................................................... 209
   6.5.1 Age is associated with a decline in the distinction between categories of musical dyad ........................................................................................................................................ 209
   6.5.2 Age is associated with weaker representation of harmonic structure by neural temporal coding ......................................................................................................................... 211
6.6 References .................................................................................................................. 214

Chapter 7. Discussion and Conclusions ........................................................................... 219
Bibliography ...................................................................................................................... 226
Tables

Table 3.1 Interval frequency ratios. 93
Table 4.1 The intervals of the dyads used in behavioural and electrophysiological testing. 126
Table 6.1 Musical dyads. 195
Table 6.2 The eight low notes used to create dyads. 196

Figures

Figure 1.1 Post-stimulus time histograms of auditory nerve spiking (Brugge et al., 1969, p. 388, Fig. 1). 24
Figure 1.2 Inner hair cell receptor potentials in response to different frequency pure tones (Palmer and Russell, 1986, p. 9, Fig. 9). 25
Figure 1.3 A schematic representation of temporal pitch theory. 36
Figure 1.4 Inter-spike interval histograms of cat auditory nerve fibres in response to equal amplitude complex tones with missing F0s (Cedolin and Delgutte, 2005, p.355, Fig. 9). 55
Figure 3.1 Complex and simple frequency ratios. 86
Figure 3.2 Monaural interactions between notes of a consonant dyad. 89
Figure 3.3 Calculating consonance preference and a neural consonance index. 95
Figure 3.4 An example of an FFR stimulus. 97
Figure 3.5 Calculating the harmonic salience of the FFR. 101
Figure 3.6 Spectra of the FFR to the Perfect 5th dyad for a single participant. 103
Figure 3.7 The same FFR represented by a power spectrum, phase- 107
locking value, and magnitude-squared coherence.

Figure 3.8 A comparison between period template analysis and the method described in the protocol.

Figure 4.1 Sections of two modulated waveforms.

Figure 4.2 Monaural interactions between the harmonics of a consonant musical dyad.

Figure 4.3 Schematic representation of a consonant interval, the Perfect 5th, and a dissonant interval, the Tritone.

Figure 4.4 Pleasantness ratings and harmonic salience of the FFR.

Figure 4.5 Consonance preference scores as a function of NCI and musical experience.

Figure 4.6 Waveforms and spectra of the FFR to the Perfect 5th.

Figure 4.7 Sections of two modulated waveforms.

Figure 4.8 Monaural interactions between the harmonics of a consonant musical dyad.

Figure 4.9 Schematic representation of a consonant interval, the Perfect 5th, and a dissonant interval, the Tritone.

Figure 4.10 Pleasantness ratings and harmonic salience of the FFR.

Figure 4.11 Consonance preference scores as a function of NCI and musical experience.

Figure 4.12 Waveforms and spectra of the FFR to the Perfect 5th.

Figure 4.13 The combined spectra of the root note and the interval note of the Tritone and the Perfect 5th.

Figure 4.14 Pleasantness ratings and consonance preference scores.

Figure 4.15 The average spectra of the FFR to the Perfect 5th and the Tritone, with No Noise, Level 1 noise, and Level 2 noise.

Figure 4.16 Neural harmonic salience and NCI scores.

Figure 4.17 Consonance preference of young normal hearing non-musicians as a function of NCI for each noise level.

Figure 4.18 Average spectra of the FFRRAW, FFRADD, and FFRSUB to the Perfect 5th from the current study and Bones et al. (2014).

Figure 4.19 Consonance preference as a function of NCI calculated from FFRRAW, FFRADD, and FFRSUB.

Figure 4.20 The average power of the harmonic peaks and neural noise in the spectrum of the FFR to the Perfect 5th for each Noise Level.

Figure 4.21 The output of LF model and HF model of the IHC RP in response to the Perfect 5th stimulus with No Noise and Level 2 noise.
Figure 5.10  The average power of the harmonic peaks and neural noise in the output of the IHC RP models for No Noise and Level 2 noise.

Figure 6.1  Each dyad is created by combining a low note and a high note.

Figure 6.2  The distribution of participant ages.

Figure 6.3  Pleasantness ratings for musical dyads and affective voices by each age group.

Figure 6.4  Consonance preference and happiness preference as a function of age.

Figure 6.5  Averaged waveforms and spectra of the FFR to the Tritone and the Perfect 5th.

Figure 6.6  Neural harmonic salience of the FFR to each dyad and NCI scores for both age groups.

Word count: 58,596
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABR</td>
<td>Auditory brainstem response</td>
</tr>
<tr>
<td>AC</td>
<td>Alternating current</td>
</tr>
<tr>
<td>ACF</td>
<td>Autocorrelation function</td>
</tr>
<tr>
<td>AM</td>
<td>Amplitude modulation</td>
</tr>
<tr>
<td>AN</td>
<td>Auditory nerve</td>
</tr>
<tr>
<td>BM</td>
<td>Basilar membrane</td>
</tr>
<tr>
<td>CF</td>
<td>Characteristic frequency</td>
</tr>
<tr>
<td>DC</td>
<td>Direct current</td>
</tr>
<tr>
<td>EEG</td>
<td>Electroencephalography</td>
</tr>
<tr>
<td>ERP</td>
<td>Event-related potential</td>
</tr>
<tr>
<td>F0</td>
<td>Fundamental frequency</td>
</tr>
<tr>
<td>F0DL</td>
<td>Fundamental frequency difference limen</td>
</tr>
<tr>
<td>FDL</td>
<td>Frequency difference limen</td>
</tr>
<tr>
<td>FFR</td>
<td>Frequency-following response</td>
</tr>
<tr>
<td>FFT</td>
<td>Fast-Fourier transform</td>
</tr>
<tr>
<td>HP</td>
<td>Huggins pitch</td>
</tr>
<tr>
<td>Hz</td>
<td>Hertz</td>
</tr>
<tr>
<td>IC</td>
<td>Inferior colliculus</td>
</tr>
<tr>
<td>IHC</td>
<td>Inner hair cell</td>
</tr>
<tr>
<td>ISI</td>
<td>Inter-spike interval</td>
</tr>
<tr>
<td>ISIH</td>
<td>Inter-spike interval histogram</td>
</tr>
<tr>
<td>NCI</td>
<td>Neural consonance index</td>
</tr>
<tr>
<td>OAE</td>
<td>Otoacoustic emission</td>
</tr>
<tr>
<td>OHC</td>
<td>Outer hair cell</td>
</tr>
<tr>
<td>RP</td>
<td>Receptor potential</td>
</tr>
<tr>
<td>SGC</td>
<td>Spiral ganglion cells</td>
</tr>
<tr>
<td>SR</td>
<td>Spontaneous rate</td>
</tr>
<tr>
<td>TFS</td>
<td>Temporal fine structure</td>
</tr>
</tbody>
</table>
Abstract

The University of Manchester
Oliver Bones
Doctor of Philosophy
The Neural Basis of Musical Consonance
August 2014

Three studies were designed to determine the relation between subcortical neural temporal coding and the perception of musical consonance. Consonance describes the pleasing perception of resolution and stability that occurs when musical notes with simple frequency ratios are combined. Recent work suggests that consonance is likely to be driven by the perception of ‘harmonicity’, i.e. the extent to which the frequency components of the combined spectrum of two or more notes share a common fundamental frequency and therefore resemble a single complex tone (McDermott et al, 2010, *Curr Biol*). The publication in Chapter 3 is a paper describing a method for measuring the harmonicity of neural phase locking represented by the frequency-following response (FFR). The FFR is a scalp-recorded auditory evoked potential, generated by neural phase locking and named from the characteristic peaks in the waveform with periods corresponding to the frequencies present in the fine structure and envelope of the stimulus. The studies in Chapters 4 and 5 demonstrate that this method predicts individual differences in the perception of consonance in young normal-hearing listeners, both with and without musical experience. The results of the study in Chapter 4 also demonstrate that phase locking to distortion products resulting from monaural cochlear interactions which enhance the harmonicity of the FFR may also increase the perceived pleasantness of consonant combinations of notes. The results of the study in Chapter 5 suggest that the FFR to two-note chords consisting of frequencies below 2500 Hz is likely to be generated in part by a basal region of the cochlea tuned to above this frequency range. The results of this study also demonstrate that the effects of high-frequency masking noise can be accounted for by a model of a saturating inner hair-cell receptor potential. Finally, the study in Chapter 6 demonstrates that age is related to a decline in the distinction between the representation of the harmonicity of consonant and dissonant dyads in the FFR, concurrent with a decline in the perceptual distinction between the pleasantness of consonant and dissonant dyads. Overall, the results of the studies in this thesis provide evidence that consonance perception can be explained in part by subcortical neural temporal coding, and that age-related declines in temporal coding may underlie a decline in the perception of consonance.
Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright statement

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. The page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, and designs, trade marks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot
and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=487), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University’s policy on Presentation of Theses.
Acknowledgments

I would like to express my gratitude for the generosity and patience with which my supervisors Prof. Christopher J. Plack and Dr. Kathryn Hopkins have given me their time throughout my PhD: it scarcely needs saying that this research would not have been possible without their support, and I am forever grateful for having had the opportunity to work with them. In addition, the assorted wisdom and intellect of Prof. Ravi Krishnan, Dr. Fred Marmel, Dr. Richard Baker, Keith Wilbraham, and Andrew King has also been instrumental in the completion of this work: thank you.

I would also like to thank my parents for their unfailing support for my various pursuits over the years: somehow it all seems to have made sense eventually. Most of all, I would like to thank my wife Sarah, without the support of whom the completion of this work – and so much else besides – would not have been possible.
Chapter 1. Literature Review

1.1 Introduction

‘Music’ can be understood as the name given to ‘structured sounds produced directly or indirectly by humans… [which] convey emotions or produce enjoyment’ (McDermott & Hauser, 2005). Despite there being no obvious evolutionary benefit from participation with music (active or passive), it ranks alongside sex and food in terms of its perceived importance to producing pleasure (Dube & Le Bel, 2003). Regions of the brain anterior and posterior to Heschl’s gyrus and in the superior and middle temporal gyri respond specifically to music (Fedorenko, McDermott, Norman-Haignere, & Kanwisher, 2012), and depressed sensitivity to musical reward is not associated with depression of sensitivity to other types of reward, such as monetary reward (Mas-Herrero, Zatorre, Rodriguez-Fornells, & Marco-Pallares, 2014). It would appear therefore that the brain has neural reward pathways specifically designated to music.

Within Western music discrete subsets of musical notes called scales are used to create harmony (simultaneous musical notes) and melody (consecutive musical notes). Western music is conventionally tuned to the equal temperament scale, whereby the continuum of possible musical notes is split into octaves (a doubling in frequency), and into 12 notes within each octave. Adjacent musical notes within an octave differ in frequency by approximately 6%, a step size called a ‘semitone’. The chromatic scale
contains all 12 notes of the octave, whereas the diatonic (major and minor) scale uses a subset of seven notes.

The use of scales is an almost universal feature of music (McDermott & Hauser, 2005). Scales have an oft-repeated note called the ‘tonic’ around which the other notes of the scale are orientated according to musical ‘key’. In musical theory the notes of the scale have a tonal hierarchy determined by their stability with the tonic, the order of which correlates with listeners’ expectations to hear that note within a melody in that key (Krumhansl & Kessler, 1982). The resolution or non-resolution of the expectations created by a sense of tonal hierarchy help to produce the feelings of tension and release that are evoked by listening to music. This sense of hierarchy is likely to be culturally determined: very young infants are able to discriminate melody changes which are in key equally well as those which are out of key, whereas adults are better able to discriminate melody changes which violate musical key; as infants grow older their ability to discriminate becomes progressively more similar to adults, presumably due to extended exposure to the musical rules in their culture resulting in the development of their tonal expectations (for a review see McDermott & Hauser, 2005).

Combinations of notes have an ‘interval’, named after their orientation within the scale. For example, the Perfect 5th interval is the distance between a lower note and a note five places higher in the diatonic scale, and seven piano keys higher on the keyboard. This thesis is concerned with ‘consonance’ – the pleasant perception of resolution produced by combining musical notes with particular intervals. The following literature review will therefore necessarily also cover theories of pitch, and the principle
physiological mechanisms involved in the perception of pitch. Chapter 2 provides an overview of the work contained within this thesis within the context of the literature reviewed here.

1.2 Pitch

1.2.1 Pitch defined

The perception of pitch is fundamental to the perception of music. Pitch can be considered as that perception of sound necessary for the discrimination of musical notes; that is, if a sound can be used to construct a melody it can be said to produce pitch (Plack & Oxenham, 2005). Of greater importance to music than absolute pitch is relative pitch, i.e. the relationship between pitches. This appears to be the case in all music cultures, and there is evidence that infants are sensitive to relative changes in pitch but not the transposition of whole melodies which retain relative pitch relationships (for a review see McDermott & Hauser, 2005). The physical quality of sound most intimately related to pitch is repetition rate, rather than audio frequency per se.

Both pure tones and complex tones produce pitch. A pure tone is a simple sinusoidal waveform whilst a harmonic complex is a tone consisting of a number of sinusoidal components with frequencies which are all integer multiples of a low frequency. This low frequency is described as the fundamental frequency (F0), which is usually the same as the rate of the overall periodic fluctuation of the complex – the ‘temporal envelope’. The individual sinusoidal components of a complex tone are variously described
as ‘harmonics’, ‘harmonic components’, ‘frequency components’ or historically ‘partials’ (partial tones). Hereafter the term ‘harmonics’ is used.

1.2.2 Cochlear representation of complex tones

The cochlea is a spiral shell-shaped structure encased within the temporal bone of the skull. It is divided by the scala media along its length into two perilymph filled chambers, the scala vestibuli and the scala tympani. At the apical end of the cochlea there is a small opening—the helicotrema—between the scala media and the walls of the cochlea, allowing for the flow of perilymph between the scala vestibuli and the scala tympani. The scala media is filled with endolymph and is itself bounded by two membranes joined at the helicotrema. The basilar membrane (BM) separates the scala media from the scala tympani, whilst Reissner’s membrane separates it from the scala vestibuli.

When sound enters the ear the perilymph of the cochlea is set in motion via energy transferred to the oval window, a membrane between the middle ear and the scala vestibuli. The motion of perilymph within the cochlea results in displacement of the BM, creating a waveform with peaks which travel from the base towards the apex. The shape of the BM waveform is determined by mechanical properties which vary along its length. At the basal end of the cochlea the BM is narrow and stiff, whilst at the apical end it is wide and flaccid. As a consequence high-frequency pure tones produce a vibration pattern with maximum displacement towards the basal end of the cochlea, whilst low frequency pure tones produce displacement along the length of the BM with maximum displacement occurring towards the apical
end. This tonotopic arrangement means that each location along the BM has a characteristic frequency (CF) at which it maximally responds with an approximately sinusoidal variation in amplitude at a frequency approximately equal to the CF (Moore, 2003).

In ears in which the cochlea is functioning normally, the input required to maintain a constant velocity at a particular place on the BM rapidly increases as the frequency of the input is varied away from the CF. The tuning curve (a plot of the input required to maintain a constant BM velocity as a function of frequency) of each place along the BM therefore has a sharp peak at the CF. In post-mortem ears the tuning of the BM is broader, with a higher level of input required to maintain the same velocity even at CF (Sellick, Patuzzi, & Johnstone, 1982). The response of the BM in the healthy cochlea as a function of input at a place tuned to the tone is shallow and nonlinear (Ruggero, Rich, Recio, Narayan, & Robles, 1997), with a gain of approximately 50 dB applied to low inputs (Yates, 1995). In post-mortem ears the response of the BM is linear with no gain. The sharpened tuning and the gain at low levels of input in healthy ears are the result of active amplification of the BM’s displacement at places tuned to the tone. This amplification is provided by the motor function of outer hair cells (OHCs; Brownell, 1983). The OHCs form part of the organ of Corti, a structure seated on the BM. Tiny hairs on their upper surface (stereocilia) are implanted in the tectorial membrane above the BM. As the BM moves up and down a shearing motion occurs between the organ of Corti and the tectorial membrane. Ion channels in tip links of the stereocilia are opened, causing depolarisation of the OHC. Upon depolarisation the protein prestin, found
within the membrane of OHCs, undergoes a change in its structural arrangement, resulting in a reduction in OHC length. The change in length of the OHCs applies additional force to the motion of the BM at the CF place and a sharpening of the frequency tuning of its response (Fettiplace & Hackney, 2006). The resulting frequency selectivity of the cochlear partition results in the auditory periphery behaving like a bank of band-pass filters (‘auditory filters’) with over-lapping pass bands (Fletcher, 1940). Auditory filter bandwidth increases with centre frequency: if $F$ is the centre frequency in kHz auditory filter bandwidth (Hz) can be estimated as $24.7 \times (4.37F + 1)$ (Glasberg & Moore, 1990). When subsequent harmonics of a complex tone excite distinct locations on the BM (i.e. more than one harmonic does not fall within the bandwidth of a single auditory filter) they are said to be ‘resolved’; when two harmonics are not separated out by the cochlea in this way they are said to be ‘unresolved’.

The shearing motion of the organ of Corti that occurs with motion of the BM also results in depolarisation of inner hair cells (IHCs). IHCs act as transducers, converting the mechanical energy of the BM into electrical (neural) energy. When the stereocilia of the IHCs are bent towards the basal body of the IHC, potassium ions flow into the IHC body, causing the voltage to change from $-40$ mV to $+40$ mV, and the IHC to therefore become depolarised. This change in intracellular potential causes voltage-gated calcium channels to open, allowing calcium ions to enter. The calcium ions accumulate in the vicinity of the synapse, and mediate the release of the neurotransmitter glutamate into the synaptic cleft at the base of the IHC,
causing the depolarisation of a corresponding dendrite of the auditory nerve (AN) which in turn generates an action potential in the AN fibre.

1.2.3 Neural representation of complex tones

Much of what is known about how pitch might be represented in the auditory system has come from animal studies. Typically these involve anesthetising the animal, surgically removing a portion of the cranium, and inserting a microelectrode into a neuron within the desired portion of the auditory pathway. The microelectrode measures changes in voltage across the cell membrane during action potentials.

The AN contains nerve fibres consisting of spiral ganglion cells (SGCs) which synapse on IHCs (type I fibres) and on OHCs (type II fibres). Each type I fibre receives its input from a particular place on the BM. As a consequence the AN is tonotopically arranged, with high CF fibres found on the outside of the AN and fibres with progressively lower CF found towards the centre. In this way frequency information is encoded in the AN by place.

In the absence of sound, AN fibres fire spontaneously at a rate dependant on the size and positioning of their synapse on the IHC: fibres with large synapses tend to have higher spontaneous rates (SR) than those with smaller synapses (Moore, 2003). When stimulated by sound, nerve fibres spike most frequently at a particular phase of the stimulus: whilst an individual AN fibre may not fire for many cycles of the stimulus (due to the cell’s refractory period), when it does fire it will tend to be during a specific phase of the stimulus (Pickles, 2008). Fig. 1.1 demonstrates the tendency for AN fibres to fire at a particular phase of the stimulus and the half-wave
rectification that occurs during the transduction process, and also the sensitivity of the neural response to the phase relationship of the stimulus harmonics. Half-wave rectification occurs because the stereocilia of the IHC are bent back and forth in synchrony with the motion of the BM, which is displaced upwards when the oval window moves outwards during the rarefaction phase of the stimulus. Depolarisation of IHCs only occurs when the stereocilia at their tip are sheared towards the scala tympani wall, whilst hyperpolarisation and therefore a reduction in spontaneous firing occurs with shearing away from the scala tympani wall (Brugge, Anderson, Hind, & Rose, 1969). The IHC receptor potential (RP) contains an alternating current (AC) component (see Fig.1.2), reflecting the opening and closing of ion channels as the stereocilia are bent back and forth, and which therefore follows the frequency of BM displacement (Russell & Sellick, 1978).

The tendency of neurons to fire at a particular phase of the BM response means that temporal fine structure (TFS) is encoded in the
resulting AN inter-spike intervals (ISIs), up to a frequency of approximately 5000 Hz (Johnson, 1980). The IHC RP also contains a DC component which follows the envelope of the tone (Russell & Sellick, 1978). At low frequencies the AC component dominates the response, but at high frequencies the direct current (DC) component dominates (Fig. 1.2). Some authors suggest that it is this which limits phase locking in the AN to approximately 5000 Hz:
the proportion of nerve fibre activity that is phase locked to the stimulus frequency relative to the overall mean discharge rate approximately corresponds to the size of the AC component, which may be limited at high frequencies by the capacitance of the IHC membrane, relative to the DC component (Palmer & Russell, 1986; Russell & Sellick, 1983).

BM tuning curves are reflected in the tuning curves of AN fibres: they have steep high-frequency sides but broad low frequency sides, such that fibres with high CFs respond to low-frequency pure tones at high levels; Rose et al. (1971) reported phase locking to an 85 dB SPL 500 Hz tone in squirrel monkey AN fibres with CFs of 1700 Hz. Geisler, Rhode, and Kennedy (1974) reported that a squirrel monkey nerve fibre with a CF of 7500 Hz responded at 100 spikes/s to a 90 dB SPL 500 Hz tone. In the same study a nerve fibre with a CF of 9100 Hz responded to a tone with the same frequency at 200 spikes/s, but failed to respond at all to a 9300 Hz tone.

Type I AN fibres terminate at the cochlear nucleus of the auditory brainstem. The cochlear nucleus serves as an entrance to the binaural pathway, since afferent (ascending) outputs lead to both the ipsilateral and contralateral superior olivary complex, the first stage of the auditory pathway at which binaural interaction occurs. Tonotopic organisation is preserved at the level of the cochlear nucleus, as is the temporal representation of frequency contained within ISIs of the AN. The retention of temporal information in the cochlear nucleus is mainly made possible by ‘primary-like’ units; neurons with large synapses and discharge rates which preserve the ISIs of the AN. Projections from all nuclei of the cochlear nucleus except the contralateral ventral nucleus synapse at the IC of the midbrain via a tract of
axons called the lateral lemniscus. The IC is also tonotopically arranged, and
demonstrates phase locking up to a frequency of approximately 2000 Hz
(Krishnan, 2007). The upper limit of phase locking reduces with ascension of
the auditory pathway. By the level of the primary auditory cortex the majority
of phase locking occurs only at 250 Hz or lower (Wallace, Shackleton, &
Palmer, 2002).

1.2.4 Pitch theories

1.2.4.1 Place theories and temporal theories

Ohm’s (1843, cited in de Cheveigné, 2004) law states that only sinusoidal
frequency components produce pitch, implying a one-to-one relationship
between sensation and perception. The pitch produced by a complex tone is
therefore understood to correspond strictly to the harmonics of the tone and,
crucially, to the F0 which must be present. Relative amplitudes of the
harmonics affect the overall timbre but not the pitch. This understanding of
pitch informed Helmholtz’s (1877/1954) influential description of the inner ear
as a bank of frequency resonators.

Ohm’s law was a matter of controversy which traversed the 19th and
the 20th century. At the heart of the place theory of pitch is the Fourier
transform, i.e. the understanding of the pitch of a complex tone as the sum of
the sensations produced by its harmonics. This understanding relies upon
the resolvability of harmonics in the cochlea. However, whilst resolved
harmonics produce a stronger response from regions of the cortex sensitive
to pitch (Norman-Haignere, Kanwisher, & McDermott, 2013), it is not clear
that pitch relies upon resolvability rather than low harmonic number per se
(e.g. see Bernstein & Oxenham, 2003). As de Cheveigné (2004) notes, a fundamental limitation to Ohm’s understanding of pitch is the nonlinear frequency response of the cochlea: the representation of a complex tone in the auditory system cannot simply be understood as the linear summation of representations of the constituent pure tones.

Seebeck (1841, cited in Plomp, 1967) investigated pitch by blowing air through holes in plates rotating at a constant velocity. When the spacing of the holes was regular, a tone was produced with a pitch equivalent to the distance between the holes, and even when the spacing of the holes varied a pitch was produced, so long as the spacing was regular (e.g. a,b,a,b). Seebeck’s work led him to understand pitch as being related to periodicity, rather than frequency per se. Ohm (1843, cited in de Cheveigné, 2004) insisted that a component must exist at the F0 for pitch to occur, suggesting that the disks used by Seebeck must produce a sinusoidal component at this frequency. The claim that the pitch of the tone produced by the rotating plates was stronger than the perception of the individual harmonics was dismissed as an ‘acoustical illusion’ (Ohm 1844, cited in Plomp, 1967). Helmholtz, whose model of pitch required stimulation in the region of the cochlea corresponding to the F0, attempted to rationalise Seebeck’s results by suggesting that, even in the absence of an F0 component in the stimulus, the ear’s nonlinearities produce distortion products at this frequency. However, this explanation was seriously challenged when it was demonstrated that what was now being referred to as ‘periodicity pitch’ still occurred when tones at the F0 were removed via phase-cancellation (Schouten, 1938), and even when the frequency region of the cochlea
corresponding to the F0 was masked by noise (Licklider, 1956). Schouten (1940) believed this was evidence for a ‘residual pitch’ produced by unresolved harmonics. Rather than exciting distinct places on the BM, unresolved harmonics (approximately above the 8th to 10th harmonic; Bernstein & Oxenham, 2003; Plomp, 1964) interact with one another and produce complex temporal firing patterns in the resulting neural excitation. This new residue theory stated that, rather than there being a one-to-one relationship between sensory input and perception, the F0 corresponding to the perception of pitch is instead coded in the envelope at the output of the cochlea.

Plomp (1967) found evidence for this theory using complex tones altered so that the F0s were 10% lower in frequency and the second harmonic and above were 10% higher in frequency. Participants were asked to determine whether the altered complex tones or unaltered complex tones were higher in pitch. For F0s up to 700 Hz the altered complexes were perceived as being higher in pitch, suggesting that the envelope produced by the upper harmonics dominated the perception of pitch, rather than the F0. Plomp found a transition region between 700 and 1400 Hz where pitch became more ambiguous. Above 1400 Hz the pitch of the altered complexes followed the F0, with judgements indicating that the altered complexes had a lower pitch than the unaltered complexes. Plomp interpreted these results as evidence that periodicity pitch occurred for F0s up to 700 Hz, but that the auditory system is unable to represent pitch this way for F0s above 1400 Hz.

Houtsma and Goldstein (1972) demonstrated that complexes containing only two successive harmonics of a missing F0 can be used to
discriminate musical intervals and musical melodies. Participants were able to discriminate pitch even when the two harmonics were presented to different ears. Since, contrary to Schouten (1940), the F0 of the complex was not represented in the vibration pattern of either cochlea the F0 must have been determined from the combined information. More recently Gockel, Carlyon, and Plack (2011) demonstrated that listeners were able to discriminate pitch by synthetically analysing stimuli consisting of a narrowband noise and a Huggins pitch (HP) stimulus, for the perception of which binaural processing is necessary. These results indicate that any mechanism responsible for the pitch produced by the two harmonics (Houtsma & Goldstein, 1972) and the two types of sound (Gockel et al., 2011) must exist beyond the point of binaural interaction in the auditory pathway, i.e. the medial superior olive.

For pitch to be encoded temporally the information pertaining to pitch must be represented in neural phase locking. The upper limit of phase locking of the AN in cats is approximately 5000 Hz (Johnson, 1980) and approximately 3500 Hz in guinea pigs (Palmer & Russell, 1986). This is also approximately the upper limit of octave judgements, i.e. the highest frequency at which listeners are accurately able to adjust a tone so that it is one octave above a reference tone (Ward, 1954), and the frequency at which frequency discrimination judgements (frequency discrimination limens; FDLs) rapidly deteriorate (Moore, 1973). Since auditory filter shape is sensitive to level (Glasberg & Moore, 1990) a pitch mechanism dependent on place information might also be expected to be sensitive to level. However, Moore and Glasberg (1989) found that random variations in signal level only
disrupted FDLs at frequencies above 4000 Hz, i.e. above the approximate frequency at which it is assumed temporal coding is no longer possible.

However, more recently Oxenham, Micheyl, Keebler, Loper, and Santurette (2011) demonstrated that listeners can reliably match the pitch of a pure tone to a preceding complex with an F0 of 1200 Hz but only containing the 6th harmonic (7200 Hz) and above. In a second task listeners identified whether the second of two four note melodies was the same as the first or whether one of the notes was lower in pitch. When the first melody was constructed using pure tones between 1000 and 2000 Hz, and the second melody was constructed using complexes with F0s in the same frequency range but filtered between 7500 and 16 000 Hz, participants were able to perform the task correctly 95% of the time. The task was then repeated, but with the harmonics of the complex tone in the second melody all shifted by the same random amount so that they no longer had a harmonic relation to a common F0; the envelope produced by the harmonics remained the same, due to the retention of the frequency spacing between harmonics. This time participants performed at chance level. This suggests that performance in the previous condition had not been based on envelope cues; rather, participants must have discriminated pitch using the harmonics with frequencies well above the assumed upper limit of phase locking.

1.2.4.2 Pattern recognition models

Goldstein (1973) modelled pitch as the output of a central processor which matches signals to harmonic templates. The input to the processor is provided by a Fourier transform performed in the cochlea, as suggested by Helmholtz. In this model pitch is a ‘multi-modal probabilistic function’:
harmonic complexes have a ‘principle pitch mode’ at their true period, but also other pitch modes corresponding to other periods. Limitations in cochlear frequency resolution due to auditory filter bandwidth and errors due to stochastic representation mean that the frequency judgement of each component has a probability distribution, sometimes leading to ambiguities in pitch judgement. Ambiguous pitch judgement manifests as the perceived pitch corresponding to a pitch mode other than the principle mode (see also Goldstein, Gerson, Srulovicz, & Furst, 1978). Although this was originally conceived as a place model, Srulovicz and Goldstein (1983) later demonstrated that the input to the central processor could also be coded temporally.

Terhardt (1974) proposed two types of pitch: an ‘analytic pitch’ which adhered to Ohm’s law of a one-to-one match between sensory inputs and perception; and a ‘virtual pitch’, derived by a central processor using learned rules activated by the lowest frequency cue present. The pitch templates are learned via exposure to the harmonic structure of speech. Initially the harmonics contained within vowel sounds are perceived as being separate and as having distinct pitches. After repeated exposure however the harmonic relation is learned. Once learned harmonic templates are activated by analytic (place) pitch components and used to identify the lowest common sub-harmonic as the ‘virtual’ pitch (Terhardt, 1979). Shamma and Klein (2000) went further, suggesting that harmonic templates could arise via exposure to any stimuli, regardless of the harmonic content. In this model harmonic templates are formed as a result of the harmonic distortion that occurs with IHC half-wave rectification. Harmonically related frequencies
become associated due to coincidence of phase locking. Although an appealing theory, the proposed ‘coincidence matrix’ would need to exist prior to the IC in order to account for frequencies above approximately 2000 Hz: as discussed, Gockel et al. (2011) demonstrated that any pitch mechanism is likely to exist beyond the level of binaural interaction, i.e. at least at the level of the IC.

1.2.4.3 Modern temporal theories

At the heart of modern temporal pitch theory is the autocorrelation function (ACF). An autocorrelation is a measure of the extent to which a function correlates with itself (‘auto’) as it is shifted. The ACF of a sine wave with a period $P$ will therefore be equal to 1 when it is shifted by a ‘lag’ equal to $P$. Licklider’s (1951) ‘duplex theory’ suggests that the auditory system makes use of operations available in both the frequency domain and the time domain, with the cochlea performing a frequency analysis and the ‘neural part of the system’ performing autocorrelation analyses on ‘nerve impulses’ in order to extract pitch. The result is an array of ACFs; each corresponding to a different frequency channel. This concept forms the basis for much of contemporary temporal pitch theory. Meddis and Hewitt (1991) extended this representation of the auditory system by adding an operation which sums ACFs across frequency channels to produce a summary ACF. This model was able to predict pitch for a range of stimuli including complexes with a missing F0 (see also the similar pitch model based upon the duplex theory proposed by Moore, 2003).

Cariani and Delgutte (1996a) measured AN fibre responses to complex tones in cats. They recorded the ISIs between both successive and
non-successive spikes (all-order ISIs) in ISI histograms (ISIHs). By summing information across nerve fibres they were able to assess the population level temporal response. Consistent with previous work, they found that for a range of complex tones the dominant all-order interval of the pooled response corresponded to the F0. For time-varying stimuli, such as single vowel sounds, they demonstrated that the pitch contour was matched by the changes in the dominant ISI over time. Perhaps most importantly however, they found a qualitative agreement between the strength of the pitch of the stimuli used and the salience of the representation of pitch in the pooled AN response, determined as the ratio of the number intervals corresponding to pitch to the number of other intervals. Stimuli which produce a strong sense of pitch such as formant vowels and complex tones produced a strong representation of the F0, whereas stimuli containing only unresolved harmonics which produce a weak sense of pitch produced a weak representation of the F0. Cedolin and Delgutte (2005) also used all-order pooled ISIHs of cat AN fibre responses to complexes with missing F0s, over a wider range than that used by Cariani and Delgutte (1996a). Rather than take the dominant interval mode as the pitch estimate the authors used a ‘periodic template’ measure, calculating the number of intervals inside bins placed at a period and its integer multiples. This method accurately predicted the missing F0 of complexes up to a frequency of 1300 Hz. Above this frequency the third harmonic of a 1300 Hz F0 is 3900 Hz and therefore approximately at the upper limit of phase locking. Therefore for complexes with a missing 1300 Hz F0 only the second harmonic was represented temporally and the pitch became ambiguous.
1.2.4.4 Summary

Fig. 1.3 displays a schematic summary of the main processes involved in modern temporal pitch theory. Note that temporal information pertaining to pitch is present even in frequency channels well above the frequencies of the stimulus. Temporal models such as this demonstrate that, in many cases, there is sufficient temporal information available in the auditory periphery for a pitch mechanism. However, pitch estimates based on temporal coding in AN cat fibres are most accurate at low-frequency F0s, whereas at high-frequencies place information provide a better estimate of pitch (i.e. rate-place profiles: the CF of nerve fibres with the highest spike-rate; Cedolin & Delgutte, 2005; Larsen et al., 2008). Considered together, it seems likely from the literature that pitch perception involves use of both place and temporal information, possibly integrated in such a way that place information at an early stage serves to prime temporal processes at a later stage (e.g. see McLachlan et al, 2009). That pitch perception exists at frequencies well above the assumed upper-limit of phase locking (Oxenham et al., 2011), and that the combination of binaural harmonics to form a pitch percept may occur at a level where phase locking is assumed to be restricted to 2000 Hz (Gockel, Carlyon, & Plack, 2011) suggests that a purely temporal model of pitch is insufficient. It may be that temporal information is transformed to place information at an early stage of the auditory pathway. However, that the pitch of stimuli with no long-term spectral information such as amplitude-modulated (AM) noise can be estimated by temporal coding of the modulation rate (Cariani & Delgutte, 1996a) is a strong indicator that, for
this stimulus at least, temporal coding is more than simply an epiphenomenon of IHC transduction.

1.2.5 The frequency-following response

Electroencephalography (EEG) is a method for recording electrical potentials in the brain via electrodes placed on the scalp. Event-related potentials (ERPs) refer to responses to a specific stimulus (event). Measurements
made this way are often referred to as ‘far-field’, due to their non-proximity to the potential. Although the generation of far-field potentials are not as well understood as near-field measurements, such as those performed in animal studies, they are thought to represent the start and end of neural propagation and features of anatomical organisation. For example, potentials measured this way can reflect sharp bends in the neural tract (Deupree & Jewett, 1988) and changes in the conductivity of the medium (Møller, 2007).

A common ERP used in studies of hearing is the auditory brainstem response (ABR). The ABR represents the onset response from neurons in the auditory pathway up to and including the brainstem, and reflects pre-attentive processing (Burkard & Don, 2007). The ABR is usually measured using a transient click stimulus with duration of less than 100 µs, and when tone-bursts are used rapid stimulus onset usually results in a larger response, i.e. the ABR is an onset response to the first few cycles of the stimulus (Suzuki & Horiuchi, 1981). The frequency-following response (FFR) represents activity in the same portion of the auditory pathway. However, whereas the five characteristic peaks of the ABR are assumed to represent individual generators at progressive stages of the auditory pathway, the name of the FFR describes the characteristic peaks in the response at periods corresponding to the stimulus frequency. As opposed to the transient response represented by the ABR, the FFR represents sustained neural activity, phase locked to the cycles of the stimulus waveform, reflecting both the temporal envelope and the fine-structure (Moushegian, Rupert, & Stillman, 1973). FFRs recorded using a ‘horizontal’ montage (active and reference electrodes placed on the mastoids behind each ear) are believed
to represent mainly contributions from the AN (latency approximately 3 ms) whilst FFRs recorded using a ‘vertical’ montage (active electrode at either Cz or high-forehead and reference electrode on the seventh cervical vertebra) are believed to represent mainly contributions from the IC (latency approximately 6 ms; Galbraith, Bagasan, & Sulahian, 2001; Smith, Marsh, & Brown, 1975).

The amplitude of the FFR decreases gradually over time, possibly as a result of neural adaptation, and also with increasing frequency, possibly due to a reduction in phase locking. The upper frequency limit of the response in general is believed to reflect the upper limit of phase locking of the generation site, i.e. vertical montage recordings typically have an upper frequency limit of approximately 2000 Hz, reflecting the upper limit of phase locking in the IC, whereas more robust high-frequency FFRs are possible from horizontal montage recordings (Krishnan, 2007).

The cochlear generation site of the FFR is determined by the interaction of multiple factors. At low intensities, the FFR to low-frequency tones is likely to be generated at or close to the place on the BM tuned to that frequency (Huis in't Veld, Osterhammel, & Terkildsen, 1977; Moushegian, Rupert, & Stillman, 1978; Yamada, Yamane, & Kodera, 1977). However, at high intensities the generation site is likely to shift in a basalward direction (Gardi & Merzenich, 1979). As the response of neurons corresponding to the place on the BM tuned to the tone becomes saturated, neurons corresponding to higher frequencies become activated (de Boer, 1977): as demonstrated by Kiang and Moxon (1974), the low frequency sides of high-frequency nerve fibre tuning curves are broad, so that at high
intensities they respond to frequencies well below their CF. In addition, neural synchrony degrades rapidly beyond the basal region of the cochlea (Kiang & Moxon, 1974): it is possible therefore that the synchrony of phase locking necessary to generate the FFR requires activation of basal nerve fibres, even in the case of low-frequency tones (Dau, 2003; Janssen, Steinhoff, & Böhnke, 1991, cited in Dau 2003).

In an early example of the FFR being used to assess temporal pitch theory, Smith, Marsh, Greenberg, and Brown (1978) recorded responses to a 365 Hz pure tone and a complex tone containing harmonics 2–5 of a missing 365 Hz F0. They noted that in both cases the majority of the energy of the FFR was at 365 Hz. Moreover, using narrow band noise to mask the region of the cochlea tuned to 365 Hz did not affect the 365 Hz energy of the FFR. Smith et al. related their results to the psychophysical phenomenon of the pitch of a missing F0 persisting even when the frequency region corresponding to the pitch is masked. They identified this as evidence for pitch being determined by the periodicity of the stimulus waveform, and for the encoding of pitch by phase locking at the level of the brainstem. Enticing though this possibility is, Hall (1979) directly challenged these conclusions by measuring the FFR to three stimuli which all had the same pitch: a high amplitude 800 Hz tone with low amplitude 600 and 1000 Hz sidebands; an 800 Hz carrier tone equal in amplitude to 600 and 1000 Hz sidebands; and 600 and 1000 Hz sidebands, with the 800 Hz carrier tone absent. In each case the pitch corresponded to a 200 Hz missing F0. In the case of the high amplitude carrier and low amplitude sidebands the stimulus had a 200 Hz envelope which was clearly identifiable in the spectrum of the FFR. In the
case of the equal amplitude carrier and sidebands the stimulus contained an
envelope periodicity of both 200 and 400 Hz, both of which were represented
equally in the FFR. In the case of the stimulus containing only 600 and 1000
Hz sidebands the stimulus envelope was 400 Hz (corresponding to the
spacing between the two components). This frequency rather than the
frequency of the missing F0 was identifiable in the FFR, demonstrating that
the FFR is sensitive to temporal fluctuations of the stimulus rather than pitch
*per se*.

More recently Gockel, Carlyon, Mehta, and Plack (2011) used a
technique introduced by Goblick and Pfeiffer (1969) to alternatively enhance
and suppress the FFR to TFS and envelope. By subtracting the FFR to the
stimulus presented with the onset polarity inverted 180 degrees from the
FFR to the stimulus presented in the original polarity, the envelope (which is
in the same phase in both polarities) is suppressed, whilst the response to
the TFS of individual harmonics (which is in inverted phase) is enhanced.
Alternatively, by adding the responses to the two polarities the response to
the envelope is enhanced whilst the response to the TFS is reduced. As
expected, when harmonics 2-4 of a complex tone with a missing 244 Hz F0
were presented monaurally, the addition FFR contained a peak in the
spectrum corresponding to the F0. In the subtracted FFR the spectrum
contained peaks corresponding to the individual harmonics but not the F0.
When only the 2nd and 4th harmonics were presented monaurally the
subtraction FFR again contained peaks only at the harmonics. The addition
FFR spectrum contained a peak at the frequency of the 2nd harmonic, i.e. the
envelope frequency. When only the 3rd harmonic was presented monaurally
the addition FFR spectrum did not contain any identifiable peaks, whilst the subtraction waveform contained the single harmonic. However, when the same harmonics were presented dichotically, with the 2\textsuperscript{nd} and 4\textsuperscript{th} harmonic presented to the left ear and the 3\textsuperscript{rd} harmonic to the right, the subtraction FFR waveform spectrum contained peaks at each harmonic whilst crucially the addition FFR spectrum contained a peak corresponding to the envelope frequency in the monaural 2\textsuperscript{nd} and 4\textsuperscript{th} harmonic condition, not the missing F0 corresponding to the pitch of the complex. This final condition demonstrates that the FFR is likely to represent a linear summation of the temporal information coded in the two ANs.

These studies demonstrate that the FFR is likely to reflect the retention of temporal information at the level of the IC, as opposed to the extraction of a pitch percept \textit{per se}. However, the integrity of the temporal information at this stage of the auditory pathway may still be important to the perception of pitch. Marmel et al. (2013) measured strength of phase locking in participants with a range of ages and audiometric thresholds as the composite of the signal-to-noise ratio of the FFR at the stimulus frequency, and the strength of cross-correlation of the waveforms of the FFR and the stimulus. Individual strength of phase locking was found to correlate with FDLs, even when the effects of age and hearing threshold were accounted for. Hearing thresholds were also found to correlate with FDLs when the effects of age and the FFR were accounted for. Age was strongly correlated with both strength of phase locking and FDLs, but did not correlate with FDLs when the effect of strength of phase locking was accounted for. The results of this study are consistent with both hearing loss and age affecting pitch.
perception independently, and with the effect of age being mediated by a decline in the integrity of temporal coding (see Section 1.4).

Krishnan, Xu, Gandour, and Cariani (2005) used synthesised Mandarin speech sounds to explore differences in the FFR of English and native Mandarin speakers. Mandarin is a tone language which uses variations in F0 to give the same syllable a number of different meanings. They measured the FFR whilst presenting synthesised Mandarin speech stimuli with identical formants but varying F0 components. The periodicity of the F0 and harmonics of the Mandarin speech sound stimuli were better represented in the FFR of Mandarin speakers, and the contour of the stimulus F0 was more strongly correlated with the corresponding contour in the FFR over the duration of the stimulus. Krishnan et al. suggest that the extensive training in pitch involved in learning to speak a tonal language may enhance temporal coding. Wong, Skoe, Russo, Dees, and Kraus (2007) also investigated the relation between Mandarin lexical tone discrimination and the FFR, in English speaking musicians and non-musicians. Musicians were significantly better at tone discrimination and had FFR pitch contours that correlated with the stimulus significantly more strongly than the non-musician group, suggesting that representation of lexical and musical tones share a common subcortical basis (see also Bidelman, 2013). Exploring these findings further, Carcagno and Plack (2011) used stimuli with either a static F0 or a dynamic rising or falling F0 contour with excursions similar to the tones used by Krishnan et al. (2005). Participants trained for 10 hours over 27 days on a particular stimulus type. Following training, participants were better able to discriminate the stimuli, and had more robust representations
of the stimulus envelope in the FFR. Behavioural improvements were partly stimulus specific for stimulus type (dynamic or static) and for direction of the contour (up or down). Improvements to the representation of the stimulus in the FFR was also partly specific for stimulus type, but not for contour.

1.3 Consonance

1.3.1 Definition and overview

‘Consonance’ refers to a pleasing perception of stability and resolution which occurs when the interval between two (or more) simultaneously played musical notes is a simple frequency ratio. The frequency ratio of the interval called the Perfect 5\(^{th}\) is approximately 3:2. The Perfect 5\(^{th}\) is the most consonant interval after the unison (1:1; two identical notes) and the octave (2:1; the same notes, but one doubled in frequency, i.e. the interval between the 1\(^{st}\) and 2\(^{nd}\) harmonics of a harmonic series). ‘Dissonance’ refers to the perception of instability and lack of resolution which occurs when notes are combined in complex frequency ratios. The dissonant Tritone for example is the interval between a low note and a high note which is only one piano key lower than the Perfect 5\(^{th}\), but which has a ratio of \(\sqrt{2}:1\).

Whether the perception of consonance is innate or the product of cultural exposure is controversial. That the Perfect 5\(^{th}\) is found in nearly all musical systems may reflect a universal predisposition towards simple frequency ratios, as may the discovery of flute-like instruments dating back to 7000 BC on which the diatonic scale can be played (McDermott & Hauser, 2005). Some studies suggest that infants are better able to discriminate
frequency changes to consonant pure tone intervals than dissonant pure
tone intervals (Schellenberg & Trainor, 1996) and that dissonant intervals
cause infants distress (for a review see McDermott & Hauser, 2005). More
recent work however suggests that infants are more sensitive to a prior
period of familiarisation than consonance or dissonance per se, and that
infants are disposed to pay more attention to a melody following a three
minute period of familiarisation than to an unfamiliar melody regardless of its
relative consonance or dissonance (Plantinga & Trehub, 2014), thus
suggesting a cultural component to preference for consonance over
dissonance. Similarly, evidence for whether animals perceive consonance in
the same way as humans is ambiguous: whilst some studies have shown
that monkeys and birds can be trained to discriminate consonance and
dissonance (e.g. see Izumi, 2000), other work casts doubt over whether this
discrimination is evidence for a preference for consonance over dissonance.
McDermott and Hauser (2004) placed cotton-topped tamarin monkeys in a
Y-shaped maze with different sounds playing in each of the two branches of
the maze. The monkeys did not receive a reward for choosing either of the
branches, so repeated selection of one branch other the other was taken as
an indication of preference for the sound in that branch. Whilst the monkeys
reliably preferred food chirps over fear chirps, and quiet noise over loud
noise, they did not display a preference for consonant sounds over dissonant
sounds. McDermott & Hauser suggest that this indicates that preference for
consonance is a human trait not shared with monkeys.

1.3.2 Consonance theories
1.3.2.1 Beating

Early theories of consonance and dissonance in the West were inherited from the ancient Greeks. Pythagoras studied consonance and dissonance by plucking strings of varying lengths and found that simple integer vibration ratios tended to result in consonant sounds whilst complex ratios tended to result in dissonant sounds. Helmholtz (1877/1954) developed these ideas, conceiving musical intervals as the ‘co-existence of two systems of vibrations corresponding to the two tones’. Whereas Euler (1739, cited in Mickelson & Riemann, 1977) had understood the relation between frequency ratios and consonance as a consequence of the listener subconsciously counting vibrations, Helmholtz observed that when two slightly different frequency pure tones are played simultaneously, the ‘intensity of the tone will be alternatively greater and less in regular succession’ (Helmholtz, 1877/1954). He identified the source of this auditory ‘beating’ as the limited frequency resolution of the ear. Helmholtz incorporated Ohm’s law into his portrayal of the cochlea as a bank of resonators, responding only to particular frequencies and therefore operating as a physiological analogue to a Fourier transformer: different pitch is produced by the stimulation of nerve cells corresponding to different places along the cochlea. Dissonance is understood by Helmholtz as the unpleasant perception of ‘roughness’ that occurs when more than one harmonic stimulate a resonator simultaneously and the resulting beating is above approximately 20 Hz (Terhardt, 1974). Consonance is understood as the absence of this sensation when musical notes are combined in simple frequency relations. For example, an octave interval has a frequency ratio of 2:1, so that all the harmonics of the upper
note align exactly with the harmonics of the lower note and no beating occurs. The consonant Perfect 5\textsuperscript{th} interval has a frequency ratio of approximately 3:2; most of the harmonics of each note align and those that do not are far apart so again no beating occurs. In contrast, complex vibration ratios such as the dissonant Tritone (approximately $\sqrt{2}:1$) result in more harmonic components being closely spaced without exactly coinciding, leading to beating.

Helmholtz’s theory of consonance remained influential throughout the 20\textsuperscript{th} century. Plomp and Levelt (1965) instructed eight young normal hearing males to rate pairs of pure tones for consonance, described as meaning ‘beautiful’ and ‘euphonious’ if asked for a definition. Each of five sets of intervals had a different mean interval frequency. Consonance scores were then plotted as a function of interval for each mean frequency. In accordance with Helmholtz, for each set small intervals were rated as being least consonant. Similar results were reported by Terhardt (1977). Plomp and Levelt interpreted their findings as evidence that, for chords outside of a musical context at least, consonance is the absence of the roughness caused by beating.

However, Plomp and Levelt’s (1965) results were not entirely in accordance with Helmholtz (1877/1954): whereas Helmholtz believed the frequency difference for maximum roughness to occur to be invariant with tone frequency, Plomp and Levelt (1965) found the frequency difference corresponding to minimum and maximum consonance to vary with mean interval frequency. They attributed this to cochlear frequency selectivity.
Because the harmonics of a complex are all integer multiples of the F0, frequency spacing is linear. Auditory filter bandwidth however increases with centre frequency: as a result, the number of harmonics that fall within an auditory filter increases with increasing harmonic number. For example, a harmonic with a frequency of 2000 Hz is unresolved if it is the 20th harmonic of a complex with a 100 Hz F0, because the 19th and 21st harmonics with frequencies of 1900 and 2100 respectively fall within the same auditory filter. If it is the 2nd harmonic of a complex with a 1000 Hz F0 on the other hand it will be resolved. Plomp and Levelt noted that the pure tone interval corresponding to minimum consonance was approximately 25% of auditory filter bandwidth, and that the smallest frequency difference for which maximum consonance occurred corresponded to approximately 100% of auditory filter bandwidth and therefore varied with tone frequency. They extended their analysis to include complex tones. By normalising the data for pure tones so that dissonance could be estimated as a function of auditory filter bandwidth, they calculated an approximate total dissonance resulting from two complex tones as the sum of dissonance created by adjacent harmonics. Their computation accurately predicted the ranking of intervals by their consonance.

Kameoka and Kuriyagawa (1969a) found that the frequency difference which corresponded to maximum dissonance not only increased with frequency, as noted by Plomp and Levelt (1965), but also with stimulus level, as would be expected with a broadening of auditory filter bandwidth (Glasberg & Moore, 2000). Kameoka and Kuriyagawa (1969b) were able to fit computational estimates of dissonance to behavioural data by modelling
dissonance as a summation of psychological roughness. However, more recently McLachlan, Marco, Light, and Wilson (2013) provided a serious challenge to the summation of roughness model of dissonance by directly comparing participants’ dissonance scores for pairs of pure tones with scores for pairs of complex tones; whereas an algorithm for calculating dissonance based on Helmholtz’s beating theory predicted that pairs of complex tones would be more dissonant than pairs of pure tones (due to the summation of the roughness caused by multiple harmonics), participants rated pairs of complex tones and pairs of pure tones similarly.

1.3.2.2 Harmonicity

Terhardt (1977) described harmonicity as ‘the perceptual phenomena which are specific of musical sounds: Tonal affinity, compatibility of simultaneous and / or successive tones, fundamental-tone relation, and tonality’. An early example of this ‘tonal affinity’ theory of consonance is that proposed by Stumpf (1890); Stumpf asked participants to describe how many tones they could hear when they were presented pairs of tones. Participants made more errors, stating that there was only one tone present, when consonant combinations were presented. Stumpf believed that this ‘tonal fusion’ of consonant pairs of tones was the cause of the perception of consonance. DeWitt and Crowder (1987) repeated Stumpf’s (1890) study, using a preceding ‘warning’ tone that was one octave below the lower note of each pair of tones, intended to encourage ‘fusion’ by suggesting a harmonic series. Similar to Stumpf’s (1890) results, Dewitt and Crowder (1987) found that more erroneous single tone judgements were made for consonant intervals than dissonant intervals, and that response latency also increased
for consonant combinations of tones, assumed to indicate that the task was more difficult.

Terhardt (1974) suggested that the dissonance of combinations of pure tones could be explained by beating, but that this psychophysical understanding of dissonance was insufficient to explain musical consonance. He proposed the extension of his virtual pitch theory to this end. This theory suggested that harmonic relations were learned so as to create pitch templates. Indeed, the Perfect 5th and Perfect 4th intervals are present in complex tones as the distance between the 2nd and 3rd harmonics and between the 3rd and 4th harmonics respectively, meaning that the auditory system could form templates for these intervals through repeated exposure to almost any natural sound. Terhardt (1977) notes that when two complex tones with a simple frequency ratio are combined the harmonics act as if they are a single tone with a virtual pitch corresponding to a missing F0. Many years before this, Rameau (1971/1722) had suggested that consonant combinations of musical notes share a common tonal centre and that this represented a ‘natural law of physics’. However, whereas Rameau (1971/1722) had written of a theoretical fundamental bass note in his ‘Treatise on Harmony’, Terhardt argued that the sub-harmonic common to the combined harmonics of a chord is perceived as virtual pitch. Simple frequency ratios result in more harmonics sharing a sub-harmonic and therefore a stronger perception of virtual pitch: the perception of consonance is the sense of two or more notes ‘fusing’.

Mathews and Pierce (1980) attempted to test both the harmonicity theory of consonance and Helmholtz’s (1877/1954) beating theory by
creating two-note chords (dyads) using complex tones with harmonics for which the frequency spacing had been ‘stretched’ so that they no longer bore a harmonic relation with one another, but also did not violate Helmholtz’s (1877/1954) criteria for avoiding roughness. ‘Cadence’ describes the perceived suitability of a consonant chord as the final chord of sequence. Chords which were rated as having high cadence when not stretched were rated as having low cadence when stretched. Moreover, it made little difference to their perceived cadence when the closely spaced harmonics of non-stretched dissonant dyads were removed so that no beating could occur. However, in a separate task they found that both musicians and non-musicians were able to make key judgements above chance level using the stretched dyads, indicating that a sense of key was not entirely lost when the harmonic relation between musical notes was disturbed.

More recently, McDermott, Lehr, and Oxenham (2010) provided a compelling case for the importance of harmonicity to musical consonance, utilising individual differences in preference. Participants rated dyads for their pleasantness, and also complexes which were either harmonic or which contained harmonics shifted up in frequency or ‘jittered’ in alternate directions and were therefore inharmonic. Participants also rated pairs of closely spaced pure tones which were either presented to both ears or to separate ears so that beating could not occur. Each individual’s average dyad rating was z-scored in order to remove individual variation in use of the rating scale. Individual consonance preference scores were then calculated by subtracting their mean ratings for the most dissonant dyads from their mean ratings for the most consonant dyads. Individual preference for
harmonicity and for stimuli which did not produce beating was calculated by subtracting ratings for inharmonic stimuli from harmonic stimuli, and by subtracting ratings for beating stimuli from non-beating stimuli respectively. By recruiting a large cohort of more than 250 participants, McDermott et al. were able to demonstrate reliable individual differences, i.e. a range of individual scores which closely correlated across trials. Moreover, although participants preferred non-beating stimuli over beating stimuli in general, and stimuli which were harmonic over stimuli which were inharmonic in general, only preference for harmonicity significantly correlated with preference for consonance; preference for non-beating stimuli did not correlate with preference for consonance. This study provides convincing evidence for the importance of harmonicity to the perception of consonance. The study also found that musical experience correlated with both preference for harmonicity and consonance, but not preference for non-beating stimuli. This could be evidence for the perception of consonance being in part culturally determined via the learning of harmonic relationships in music.

Congenital amusia is a neurological deficit in pitch processing, one symptom of which is an inability to recognise musical melodies. Using the same methodology as McDermott et al. (2010), Cousineau, McDermott, and Peretz (2012) found that amusics tended to rate all intervals as being similarly pleasant, with a large variation in individual ratings. Amusics also demonstrated less preference for harmonicity than controls, and performed poorly at harmonicity discrimination tasks. However, amusics were able to discriminate beating and demonstrated preference for stimuli lacking beating similar to controls. Amusics also demonstrated normal preference ratings for
affective voices, demonstrating that their lack of preference for consonant
dyads was not driven by a general deficit in affective labelling.

A way in which simple harmonic ratios could be important to the
auditory system can be seen in Shapira Lots’s (2008) coupled oscillator
model. The model consists of synchronised oscillators (representing
periodically firing neurons) with firing patterns with fixed periods, with the
input to each neuron consisting of the sum of an external input and an input
from the other neuron. A feature of such a system is that changes in the
intrinsic oscillator frequency for which the synchronised ratio remains
constant varies for different ratios of intrinsic frequencies. Wider intervals for
which synchronisation can be maintained can be considered as
corresponding to more stable intrinsic frequencies. The ordering by stability
of frequency ratios in this model match the Western music theory ordering of
musical intervals, arranged by consonance.

Stumpf’s (1890) concept of tonal fusion has found expression in
models of how consonance might be extracted from temporal information in
the auditory pathway based on neuronal periodicity detection. Ebeling (2008)
modeled consonance as the perceptual consequence of a mechanism that
calculates the coincidence of neural spiking in the AN that occurs when two
tones form a musical interval. Resolvable harmonics are represented in the
periodicity of neural ISIs, with consonant musical intervals producing
stronger representations of harmonically related periods. In this analysis
therefore, consonance is understood as the perceptual consequence of the
harmonic structure of tones combined in simple frequency ratios.
More recently however McLachlan et al. (2013) challenged Stumpf’s (1890) tonal fusion theory. Participants were instructed to adjust a probe tone so that the pitch matched a target harmonic of a pair of complex tones: McLachlan et al. predicted that if consonance were driven by tonal fusion the task should be more difficult when tone pairs were consonant than when they were dissonant. They found however that pitch matching errors occurred more frequently for dissonant pairs of tones than consonant pairs of tones. McLachlan et al. proposed an alternative model of consonance, based upon McLachlan’s (2011) pitch model: spectral pitch cues from processing in the early stages of the auditory pathway are used to prime a periodicity detection mechanism, as in McLachlan’s (2011) pitch model. This process determines one pitch of a chord; the remaining pitches are determined by activating chord templates and directing attentional processes. In this model, dissonance is the ‘cognitive incongruity’ that occurs when there is a mismatch between the pitches of a chord and the tuning of the neurons which have been primed. McLachlan et al. suggest that musicians perceive consonant chords as being more consonant than do non-musicians (McDermott et al., 2010; McLachlan et al., 2013) due to greater familiarity with the pitches of chords regularly used in music and therefore more accurate neural priming.

1.3.3 Consonance and temporal coding

Tramo et al. (2001) measured the response of AN fibres in cats to consonant and dissonant musical dyads. The dominant period of each dyad’s waveform was identified using an ACF, and was found to correspond in each case to
Rameau’s (1722/1971) fundamental bass note, and the F0 of Terhardt’s (1974) virtual pitch. The dominant interval in the all-order ISIHs calculated from AN responses to each dyad was found to correspond to the dominant period of the dyad.

Tramo et al.’s (2001) identification of the fundamental bass note in the AN response can be compared to later work by Cedolin and Delgutte (2005) identifying the missing F0 of complex tones in the ISIs of the AN. Fig. 1.4 displays data from Cedolin and Delgutte (2005, p.355, Fig. 9). Traditionally the method for identification of pitch from ISIHs had been to either identify the largest or earliest peak. However, Cedolin and Delgutte note that in the case of higher F0s the earliest peak is systematically delayed with respect to the F0 and often lower in amplitude compared to other peaks, due to neural refractoriness (Fig. 1.4B). Even in the case of the lower F0 (Fig. 1.4A), the peaks of the ISI are similar in height. To overcome this difficulty in objective identification of the F0 corresponding to pitch, Cedolin and Delgutte used a periodic template method which made use of all the peaks in the ISI distribution: each template selected intervals within bins at a given ISI and its integer multiples; pitch was estimated as the F0 corresponding to the interval which maximized the signal-to-noise ratio of intervals within the template to the mean distribution of intervals (‘pitch template contrast ratio’; Fig. 1.4C-D).

Bidelman and Krishnan (2009) used a similar analysis in an FFR study of consonance. They recorded the FFR to nine dyads, six consonant and three dissonant, and calculated the ACF for the mean FFR to each dyad. In order to use the same period template analysis as Cedolin and Delgutte (2005), Bidelman and Krishnan first rectified the ACF of each dyad, so that
all negative values were positive. They then weighted each ACF with a decaying exponential so that longer periods contributed less, and finally applied a period template analysis in the same way as Cedolin and Delgutte (2005). Bidelman and Krishnan (2009) found that the dominant period in their analysis of the FFR to each dyad corresponded to the F0 of the root note, and that the magnitude of this measure (‘pitch salience’) correlated with mean consonance ratings.

It is perhaps surprising that the dominant period in the FFR to each dyad was found to be the F0 of the root note, since Tramo et al. (2001) identified the dominant period in the ISI histogram of cat AN responses to dyads as the fundamental bass note - which was the dominant period identified in each dyad. This may be a consequence of the weighting function.
used by Bidelman and Krishnan (2009) to give precedence to shorter periods; had they not weighted the ACF in this way lower frequencies corresponding to the F0s identified by Tramo et al (2001) may have been found to be most salient (see further discussion in Chapter 2). It should also be noted that the ACF of the FFR is not an exact analogue to the pooled all-order ISI histogram, and that differences in pitch estimate may be due to what they represent: the latter is a pooled measure of the all-order intervals of spikes within each measured fibre, the rate of which is limited by the neural refractory period; the FFR measures all intervals between all the spikes of all the neurons responsible for its generation.

Lee, Skoe, Kraus, and Ashley (2009) analysed the frequency spectrum of the FFR to musical dyads, recorded from musicians and non-musicians. Compared to non-musicians, the FFR of musicians had greater amplitude at the F0 of the upper note and also at frequencies corresponding to distortion products produced by interactions between the harmonics of the two notes. Although the purpose of the study was to examine differences between musicians and non-musicians, it is noteworthy that in the case of the consonant dyad (the Major 6th) the distortion products served to enhance the harmonicity of the FFR, due to the simple frequency ratio of the two notes. Lerud, Almonte, Kim, and Large (2014) suggest that the larger distortion products in the musicians’ FFR reported by Lee et al. (2009) could be the result of musicians having greater synaptic efficiency and therefore processing that is more nonlinear.

1.4 Age and temporal coding
1.4.1 Introduction

Hearing loss is typically associated with the loss or dysfunction of cochlear hair cells. This can occur as a result of presbyacusis (age-related hearing loss), exposure to loud sounds, exposure to ototoxic drugs such as aminoglycoside antibiotics, and hypoxia (oxygen deprivation). OHCs are particularly susceptible to damage. Healthy OHCs amplify the BM response to low level signals (Ruggero & Rich, 1991), providing approximately 50 dB of gain (Yates, 1995). Loss or damage of OHCs therefore causes quiet sounds to become less audible. Loss of sensitivity can also occur as a result of IHC damage: if all of the IHCs in a given frequency range are destroyed or damaged, frequencies in the resulting ‘dead region’ are likely to only be heard when excitation of the BM spreads to regions where IHCs remain intact. OHCs are also responsible for the frequency selectivity of the cochlea: OHC loss therefore also causes the BM response to become less frequency selective (Ruggero & Rich, 1991; Ruggero et al., 1997).

Hearing loss is also associated with a decline in ability to make use of TFS, although with some variability (Hopkins, Moore, & Stone, 2008). Moore and Peters (1992) found no evidence for a relation between auditory filter bandwidth and the FDL or F0DL, but found that some older listeners had larger F0DLs despite having clinically normal hearing. More recently Hopkins and Moore (2011) used methods designed to specifically assess participants’ ability to use TFS. In one task participants were asked to discriminate between normal harmonic complexes and complexes containing harmonics which were all shifted by the same amount. Since the shifted complexes
retain the same envelope as the non-shifted complexes, when both types of complex are band-pass filtered to only contain unresolved harmonics, discrimination depends upon use of the TFS cues (Moore & Sek, 2009). The second task required participants to discriminate low-frequency pure tones with inter-aural phase differences (Hopkins & Moore, 2010). Despite performance at the TFS tasks correlating with audiometric thresholds, Hopkins and Moore (2011) found that older listeners with normal audiometric thresholds and frequency selectivity performed worse on TFS tasks than younger listeners with normal audiometric thresholds. This is important, as it suggests a ‘hidden’ component to hearing loss independent of audiometric thresholds but associated with age.

More recently a number of studies have demonstrated a decline in the temporal coding represented by the FFR with age (Anderson, Parbery-Clark, White-Schwoch, & Kraus, 2012; Clinard & Tremblay, 2013; Marmel et al., 2013; Ruggles, Bharadwaj, & Shinn-Cunningham, 2012). A small number have tested the relation between age, pitch perception, and the FFR: Clinard, Tremblay, and Krishnan (2010) found that age predicted an increase in FDLs, but did not find a relation between age and the amplitude or phase coherence of the FFR at the stimulus frequency. However, Marmel et al. (2013) found that age predicted larger FDLs and also that the representation of the stimulus in the FFR predicted performance. When the effect of the FFR was controlled for, the correlation between age and performance was no longer significant, suggesting that the effect of age on pitch perception may be mediated through the strength of temporal coding represented by the FFR.
1.4.2 Neural inhibition

Walton, Simon, and Frisina (2002) studied the effects of aging on mice by measuring phase locking of single neurons in the IC in response to AM noise. They found that neurons in older mice had a higher spike count in general, but that compared to younger mice they had a lower spike count in response to AM noise at their best modulation frequencies relative to unmodulated noise. In younger mice neurons were found to phase lock better to higher modulation frequencies, and a larger proportion of neurons had high best modulation frequencies. Older neurons were found to fire with less accuracy cycle-by-cycle, and types of unit which responded with rapid onset in young mice responded with a greater latency (Walton, Frisina, Ison, & O'Neill, 1997) and with a longer recovery period (Walton, Frisina, & O'Neill, 1998) in older mice (see also Frisina & Walton, 2006).

A loss of neural spectro-temporal precision with age is consistent with a loss of inhibition (Caspary, Schatteman, & Hughes, 2005). Neurons receive excitatory glutamatergic inputs and inhibitory glycinergic and / or GABAergic inputs. During spontaneous or evoked activity the ratio of excitatory and inhibitory inputs remains constant, defining the neuron’s spectro-temporal response characteristics (Gourevitch, Edeline, Ocelli, & Eggermont, 2014). If GABA receptors are blocked by an agonist, the neural frequency response broadens (Caspary, Palombi, & Hughes, 2002) and phase locking declines (Klug et al., 2002; Koch & Grothe, 1998). Importantly, Milbrandt, Albin, and Caspary (1994) found that GABA receptor binding in the IC was significantly reduced in 26 month old rats compared to 3 month old rats, whilst the cross-
sectional area remained the same. This provides strong evidence that the
decline in temporal processing associated with age may be in part related to
a change in temporal neural response properties, due to a decline in the
sharpening provided by GABAergic inhibition.

1.4.3 Cochlear neuropathy

By examining temporal bones from deceased listeners aged from 0 to 100
years with normal hair cell populations, Makary, Shin, Kujawa, Liberman, and
Merchant (2011) found that SGC population varied from approximately 33.6k
(0-10 years) to 22.4k (91-100 years), with a loss of approximately 100 cells
per year. Animal studies indicate that even the loss of a large portion of the
AN (50-90 %) does not significantly affect auditory thresholds (Schuknecht &
Woellner, 1953), suggesting that a large number of afferent fibres are instead
required for the accurate representation of complex stimuli, and for
challenging listening conditions.

Sergeyenko, Lall, Liberman, and Kujawa (2013) studied the effects of
aging in mice. They found that both ABR and otoacoustic emission (OAE)
thresholds (used to determine integrity of OHC function) at high and low
frequencies became progressively more elevated with age up to
approximately 96 weeks. Above this age thresholds at all frequencies rapidly
became more elevated. Differences between ABRs and OAEs were found at
supra-threshold levels of stimulation: the amplitude of wave I of the ABR
decayed from 64 weeks onwards, whereas the amplitude of OAEs,
representing OHC function, began to decline only after 144 weeks.
Immunostaining revealed that it is likely that the reduction in wave I of the
ABR was caused by either decoupling of the IHC from the AN, or a loss of AN fibres: there was little evidence of IHC loss in older mice (~80% survival), whereas SGC loss was high (~40% survival). Presynaptic ribbons were associated with a glutamate receptor patch in 95% of cases, with ribbon count reducing with age from 4 weeks old and continuing throughout the life span.

Deafferentation and synaptic damage may also occur as a result of noise trauma. Glutamate, the neurotransmitter released by IHCs, is neurotoxic in large quantities. Excessive glutamate release disrupts afferent dendrites, and repeated or prolonged excitotoxicity results in calcium influx and SGC death (Pujol & Puel, 1999). Temporary loss of hearing sensitivity following noise trauma may be the result of glutamate excitotoxicity and the uncoupling of SGC dendrites from IHCs, with a threshold recovery time-course corresponding to synaptic recovery (Puel, Ruel, Gervais d’Aldin, & Pujol, 1998). Puel, Pujol, Tribillac, Ladrech, and Eybalin (1994) found that when a glutamate agonist was applied locally to SGCs in guinea pigs the damage caused was comparable to that seen following noise trauma. When a glutamate antagonist was applied its effect was to protect against noise damage (Puel et al., 1998).

Animal studies demonstrate that although threshold shifts resulting from noise trauma may be temporary, the associated synaptic damage and consequent deafferentation may persist. Kujawa and Liberman (2009) found that after mice had been exposed to an 8-16 kHz octave band of 100 dB SPL noise for two hours, thresholds were elevated by 40 dB at 32 kHz 24 hours post-exposure. By two weeks post-exposure hearing and OAE thresholds
had returned to normal, suggesting normal cochlear function. Although ABR wave I thresholds were also normal, at supra-threshold levels the amplitude only recovered to 40% of its pre-exposure amplitude, suggesting lasting damage to the AN. 24 hours post-exposure the number of synaptic ribbons per IHC was found to be reduced by approximately half, and ribbons which remained were swollen and displaced. The number of SGCs remained constant up to two weeks post-exposure but then decreased to approximately 50% pre-exposure numbers by 100 weeks post-exposure, compared to approximately 10% in controls. The level of noise used in the study, whilst substantial, is not atypical of that encountered by humans in regular life. It is possible therefore that undetected noise-induced deafferentation may occur over the lifespan.

That ABR thresholds can return to normal levels despite the loss of AN fibres in these animal studies is likely to be due to the loss being selective for high-threshold low-SR fibres (Furman, Kujawa, & Liberman, 2013; Schmiedt, Mills, & Boettcher, 1996). This could particularly impact temporal processing, since low SR fibres in the auditory periphery have been shown to phase lock better than high SR fibres (Joris, Carney, Smith, & Yin, 1994), and to make a relatively larger contribution to steady-state responses (Muller & Robertson, 1991; Rhode & Smith, 1985; Taberner & Liberman, 2005).

1.5 References


Hopkins, K., Moore, B. C. J., & Stone, M. A. (2008). Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine
structure information in speech. Journal of the Acoustical Society of America, 123, 1140-1153.


Chapter 2. Thesis overview

2.1 Overview

The work contained within this thesis comprises studies concerned with addressing outstanding issues in the area of neural temporal coding and the perception of musical consonance. In particular, the work examines the relation between temporal coding of harmonicity, as represented by the frequency-following response (FFR), and individual differences in the perception of musical consonance, as represented by preference for consonance over dissonance.

Despite the results of McDermott, Lehr, and Oxenham (2010) being arguably some of the most compelling in the musical consonance literature, no FFR study of musical consonance to date has directly tested the harmonicity theory, or specifically examined the representation of harmonicity in the FFR. As discussed in Chapter 1, Bidelman and Krishnan (2009) used Cedolin and Delgutte’s (2005) ‘periodic template’ analysis to measure the ‘pitch salience’ of the FFR to musical dyads, i.e. the strength of a period as measured by an autocorrelation function (ACF) of the FFR and its integer multiples relative to other periods. The results of this study are arguably in contradiction with the results of McDermott et al. (2010), since they suggest the representation of the period corresponding to the low-note F0 in the FFR, rather than the relevant fundamental bass note per se, to be a neural correlate for consonance in the case of all dyads (see also Bidelman, Gandour, & Krishnan, 2011a, 2011b; Bidelman & Krishnan, 2011; Krishnan,
Arguably, however, there is little evidence to suggest that this particular period is an appropriate predictor of consonance, beyond being a general index for the strength of the FFR. Furthermore, the analysis requires that the ACF of the FFR is rectified in order to render the ACF more like an inter-spike interval histogram (ISIH) and therefore an approximate analogue for the analysis by Cedolin and Delgutte (2005): this is problematic since the frequency content is altered by making negative values positive in this way. Chapter 3 describes an alternative method for examining the FFR, intended to specifically measure the extent to which the harmonicity of musical dyads are represented in the FFR. (The work in this chapter was prepared for submission as a methodological paper, the requirements for which were such that some parts are stylistically unusual, e.g. the protocol section is written in the imperative tense.)

A further, arguably more fundamental issue concerning FFR studies of consonance to date is that of how informative the FFR is as a measure, as presented in the literature. Consonant combinations of notes produce waveforms that are more periodic than dissonant combinations. The retention of this periodicity in the FFR is to be expected and not novel in itself, beyond demonstrating that the FFR is sensitive to this difference between consonant and dissonant waveforms. The demonstration that an FFR measure is sensitive to individual differences in the perception of consonance would make a more compelling case for the role of temporal coding. Chapter 4 is a publication of a study designed to test the hypothesis that the method described in Chapter 3 predicts individual differences in the
perception of consonance in young normal hearing listeners with a range of musical experience.

The analysis of the FFR throughout this thesis was performed in the frequency-domain. One advantage of this compared to time-domain analysis is that the relative contribution of different frequencies is made apparent. As discussed in Chapter 1, Lee, Skoe, Kraus, and Ashley (2009) noted that distortion products in the FFR to consonant dyads served to enhance the harmonicity. By comparing diotic and dichotic presentation of dyads, the study in Chapter 4 also examined the effect that distortion products resulting from monaural interactions had on the harmonicity of the FFR, and the effect that this had on the perception of consonance.

As discussed in Chapter 1, recent work demonstrates that the FFR may be sensitive to age-related deficits in temporal coding (e.g. Anderson, Parbery-Clark, White-Schwoch, & Kraus, 2012). However, despite evidence that the perception of consonance is associated with temporal coding there has been no work to date demonstrating a relation between age and the perception of consonance. Tufts, Molis, & Leek (2005) found that hearing impaired listeners who were also older than a normal hearing control group perceived consonant dyads as less distinct from dissonant dyads than did the control group. Bidelman and Heinz (2011) modelled the results of Tufts et al. (2005) using the reported audiometric data, and demonstrated that the results could be accounted for by temporal coding at the level of the AN. However, the association between a change in the perception of consonance and an age-related decline in temporal coding per se currently remains hypothetical.
The number of people over the age of 65 years in the UK is forecast to double by 2071 (Government Office for Science, 2008). If age is associated with altered music perception, it could have important implications for public health, since listening to music is linked with many dimensions of well-being in the elderly (Laukka, 2007): music is associated with countering depression and with self-expression in mental illness (Iwasaki, Coyle, & Shank, 2010); levels of the antibody immunoglobulin A have been shown to be responsive to music, as have levels of the stress hormone cortisol (for a review see Fancourt, Ockelford, & Belai, 2014); music has been described as a forum in which the elderly are able to interact with others (Hays & Minichiello, 2005a), and is associated with identity and well-being (Hays & Minichiello, 2005b), and with relaxation (Gabrielsson, 2002). Advances in understanding how music is represented in the auditory system and how this may become degraded with age may shed light on why music is beneficial in old age, how the benefits of listening to music might be optimised, and be informative for future work on improving quality of life for the elderly through listening to music.

A potential confound in studying the effect of age on temporal coding is the difficulty in disentangling this effect from the effect of age-related hearing loss. In particular, it can be challenging to find age-matched participants with clinically normal hearing at high frequencies. A possible solution to this problem is to use high-pass filtered masking noise, so that neither young nor old participants are able to use high-frequency hearing. However, as discussed in Chapter 1, recent work suggests that the FFR to even low-frequency stimuli may be generated in the basal portion of the
The study in Chapter 5 was designed to test two hypotheses: that the findings reported in Chapter 4 would be replicated with young normal hearing non-musicians; and that the FFR to low-frequency dyads was generated in part by the basal portion of the cochlea.

Finally, the study in Chapter 6 was designed to test the hypothesis that aging is associated with a decline in temporal coding, and therefore the temporal representation of harmonicity, and that this would correspond with a decline in the perception of consonance. Measures were taken to control for the possibility of age-related changes in general affect, and in aptitude to perform the task.

In carrying out this work, the decision was taken for papers to be written and submitted to peer-reviewed journals as the results were analysed. For this reason permission was sought and provided by the Faculty of Medical and Health Sciences Graduate Office for this thesis to be presented in the alternative thesis format. This format necessitates some repetition, as permitted by The University of Manchester in the alternative thesis format guidelines.

The work presented in Chapter 4 is co-authored by the PhD supervisors Prof. Christopher J. Plack and Dr. Kathryn Hopkins, and Prof. Ananthanarayan Krishnan of Purdue University. All other papers are co-authored by Prof. Christopher J. Plack. PhD student Oliver Bones is lead author on all papers, and undertook the experimental design, and the data collection and analysis in all studies.
2.2 References


Chapter 3. A method for measuring the harmonic structure of subcortical neural phase locking and its role in the perception of musical consonance

Oliver Bones

Christopher J. Plack

School of Psychological Sciences, The University of Manchester,
Manchester Academic Health Science Centre, Manchester, UK, M13 9PL

This chapter was prepared for publication as a paper. It is under review at the Journal of Visualized Experiments.

Publication 1
3.1 Keywords
Consonance; harmonicity; frequency-following response;
electroencephalography; temporal coding; phase locking; brainstem.

3.2 Short abstract
This protocol describes a method for calculating a neural correlate for the perception of consonance. The method involves measuring the power of the subcortical response which represents the harmonic structure of two-note musical chords.

3.3 Long abstract
Consonance describes the pleasing perception of stability and resolution that occurs when two or more musical notes with simple frequency ratios are combined to create a chord. The simple frequency ratio of consonant combinations of notes means that they have a combined frequency spectrum which contains components that are integer multiples of a common fundamental frequency, and which therefore resemble a single harmonic series. The extent to which the spectrum of the combination resembles a harmonic series (the harmonicity of the chord) predicts how consonant it is. The harmonicity of a musical chord is represented in the temporal firing patterns of populations of neurons in the brainstem which phase lock to the frequency content of the stimulus. Sustained neural activity in the brainstem is measurable as the scalp-recorded electrophysiological frequency-following response. This protocol describes a method for measuring individual
differences in the representation of the harmonicity of musical two-note
chords in the brainstem. The best fitting harmonic series of a dyad is found
using a harmonic sieve analysis. The strength of the representation of the
best fitting harmonic series in the frequency-following response is found by
calculating the power that contributes to the harmonic series relative to the
power that does not. The separate contributions of the harmonics and the
temporal envelope of the cochlear response are assessed by either adding
or subtracting the neural response to alternating stimulus polarities. These
methods can be used to investigate the extent to which temporal coding of
harmonicity accounts for variation in the perception of musical consonance.

3.4 Introduction

When two or more musical notes with simple frequency ratios are combined
to create a musical chord they produce a pleasing perception of resolution
and stability (consonance), whereas complex ratios produce an unpleasant
perception of instability and of being unresolved (dissonance). Why simple
ratios produce this perception has been a matter of debate since at least the
time of Pythagoras, who explored consonance using simultaneously plucked
strings of varying length.

Musical notes are examples of harmonic complex tones, consisting of
frequencies at integer multiples (harmonics) of a fundamental frequency
(F0). When a complex tone enters the ear the lower harmonics cause the
basilar membrane (BM) to vibrate at specific, discrete locations. Harmonics
that are separated out this way are said to be resolved. When two harmonic
complexes with a complex frequency ratio enter the ear, due to the irregular spacing of the combined spectrum (Fig. 3.1A) often more than one harmonic excites each place on the BM. When harmonics are unresolved in this way they interact to produce a complex waveform that repeats at a rate corresponding to the spacing of the harmonics. If the resulting rate of this beating is between approximately 15 and 300 Hz (Zwicker & Fastl, 1990) it produces an unpleasant sensation called roughness. For many years the dominant theory has been that simple frequency ratios result in the perception of consonance because the combined spectrum contains harmonics which are either perfectly coincident or sufficiently spaced apart to be resolved from one another, so that beating does not occur (Helmholtz, 1877/1954; Plomp & Levelt, 1965; see Fig. 3.1B).

However, an alternative theory of consonance is that it is the consequence of simple frequency ratios resulting in a combined spectrum which closely resembles a single harmonic series. This harmonicity may drive consonance by activating pattern recognition schemas (Terhardt, 1974, 1977, 1979, 1984) used in pitch perception, causing the combined notes to perceptually fuse into a single entity (Stumpf, 1890). This fusion may result from the way in which the combined notes are temporally coded in the peripheral auditory system. Neurons in the auditory nerve tend to fire at a particular phase of the BM’s vibration pattern, so that each resolved harmonic is temporally coded in the temporal fine structure (TFS) of inter-spike intervals (ISIs). The BM vibration modulation rate (the envelope) is also represented this way. The harmonic relation between the frequency components of a musical chord is therefore coded in the ISIs of the auditory
nerve. This has led to models of consonance as the perceptual consequence of a system which is more stable when neural firing is synchronized (e.g. Shapira Lots & Stone, 2008), or as the consequence of a neuronal periodicity detection mechanism (e.g. Ebeling, 2008).

Recent work provides compelling evidence that it is harmonicity rather than the absence of beating which drives the perception of consonance. In a large cohort study (N = 265) McDermott, Lehr, and Oxenham (2010) tested participants’ preference for consonant dyads (two note chords) over dissonant dyads, demonstrating significant and reliable individual differences. Participants were also tested for their preference for harmonic over inharmonic complexes and non-beating over beating complexes. Preference for harmonicity explained approximately 45% of variance in preference for consonance, whereas preference for non-beating explained less than 5%. The present protocol describes methodology designed to
determine if these individual differences in preference for consonance over
dissonance can be explained by differences in the representation of
harmonicity in neural temporal coding.

Sustained phase locking by subcortical populations of neurons is
measurable as the scalp-recorded electrophysiological frequency-following
response (FFR). This method has made it possible to test the relation
between consonance and the temporal information at this level of the
auditory pathway (Bidelman & Krishnan, 2009, 2011). A common approach
is to use an autocorrelation function (ACF) of the FFR waveform to identify
the dominant period (or periods), as an approximation of methods used
elsewhere to analyse temporal coding in single neurons (Cedolin & Delgutte,
2005). An ACF is the correlation of a waveform with a delayed version of
itself as a function of delay. Waveforms that repeat will therefore correlate
strongly at a delay of the repetition period, and at multiples of that period.
The regular spacing of the frequency components of consonant dyads
means that they have regular, periodic waveforms. Using the correlation at a
delay corresponding to the period of the F0 of the lower note of a dyad as a
measure of neural pitch salience these studies demonstrate that the ranking
of dyads by this index corresponds to their ranking according to consonance
(Bidelman & Krishnan, 2009, 2011).

Studies such as these demonstrate that the temporal information in
the regular waveforms produced by consonant combinations of notes is
retained in the temporal firing patterns of neurons at the level of the inferior
colliculus. However, a limitation to the methodology employed in these earlier
studies is that it is assumed that the period corresponding to the F0 (and
multiples of this period) best represents the temporal information responsible for consonance. However, if the perception of consonance is driven by harmonicity it is likely that the temporal coding of this information is a more accurate correlate. A second limitation is that by performing analysis in the time-domain distortion products arising from nonlinear processing of musical dyads (Bones, Hopkins, Krishnan, & Plack, 2014; Lee, Skoe, Kraus, & Ashley, 2009; Lerud, Almonte, Kim, & Large, 2014) and their role in contributing to the perception of consonance are not made apparent.

This protocol describes an alternative method for recording and analysing the FFR to musical dyads, sensitive to individual (Bones et al., 2014; Bones & Plack, 2014a) and group (Bones & Plack, 2014b) differences in preference for consonance. The method assumes that the temporal information available in the auditory brainstem that best predicts the perception of consonance is the phase locking that specifically relates to the harmonicity of the dyad. One feature of this hypothesis is that it predicts that consonant combinations of notes presented to the same ear (diotic) will be perceived as being more consonant than notes presented to different ears (dichotic), due to phase locking to distortion products resulting from interactions on the BM between regularly spaced harmonics which enhance the overall harmonicity of the neural response (Lee et al., 2009; see Fig. 3.2). The protocol therefore also describes a procedure for identifying monaural distortion products and their role in the perception of consonance, and for assessing whether these frequency components are present in the TFS or the envelope. This is achieved by recording the FFR to stimuli that are alternately presented in their original onset polarity and with an onset polarity
that is inverted by 180 degrees. By *adding* the averaged response to one polarity to the averaged response to the other polarity, the response to the BM envelope and even-order nonlinearities are enhanced and the response to the TFS and odd-order nonlinearities are suppressed. Conversely, by *subtracting* the response to one polarity to the response from the other polarity the response to the TFS and odd-order nonlinearities are enhanced, and the response to the envelope and even-order nonlinearities are suppressed (Goblick & Pfeiffer, 1969). The protocol therefore describes the analysis of $\text{FFR}_{\text{ADD}}$, $\text{FFR}_{\text{SUB}}$, and $\text{FFR}_{\text{RAW}}$. 

![Diagram](image_url)
3.5 Protocol

All participants provided written informed consent in compliance with a research protocol approved by The University of Manchester Research Ethics Committee.

3.5.1. Behavioural stimuli

1.1) Create four 2000 ms duration root notes with F0s of 110.00 Hz (A), 130.81 Hz (C), 155.56 Hz (D#), and 185.00 Hz (F#). Create each note by adding together sine waves with frequencies which are integer multiples of each F0, up to the 20th harmonic, i.e.

\[ x(t) = \sum_{k=1}^{20} \sin(2\pi kFt) \]

where \( x(t) \) is amplitude as a function of time, \( k \) is harmonic number, \( F \) is F0, and \( t \) is time.

1.2) Calculate the level of each harmonic of the root note required for an overall level of 80 dB SPL when all harmonics are equal amplitude, i.e.

\[ x_k = 80 - 10\log_{10}(\frac{\omega_c}{F}) \]

where \( x_k \) is the level of each harmonic in dB SPL, \( F \) is the F0, and \( \omega_c \) is the low-pass filter cut-off frequency in Hz (2000; see step 1.12).
Note: Step 1.3 is optional. It applies to signal generation in software such as Matlab.

1.3) Calibrate the level of the root note at the output of the headphones by measuring the output to a 1000 Hz pure tone scaled between -1 and 1. Calculate the amplitude necessary to produce an overall level of 80 dB SPL at the output of the headphones, i.e.

\[ b = 10^{\frac{x_k - \text{max}(x)}{20}} \]

Where \( b \) is a scalar, \( \text{max}(x) \) is the maximum output of the headphones in dB SPL, and \( x_k \) is the level of each harmonic in dB SPL calculated in step 1.2. Multiply each value of the \( x(t) \) root note array by \( b \).

1.4) Ramp the first and last 10 ms of the root note on and off respectively, by multiplying the signal by a raised cosine-function, i.e.

\[
w(t) = \begin{cases} 
1 - \cos\left(\frac{\pi t}{T_{\text{ramp}}}\right) \times \frac{1}{2}, & 0 < t < T_{\text{ramp}} \\
x(t), & T_{\text{ramp}} < t < T - T_{\text{ramp}} \\
1 + \cos\left(\frac{\pi t}{T_{\text{ramp}}}\right) \times \frac{1}{2}, & T - T_{\text{ramp}} < t < T 
\end{cases}
\]

where \( w(t) \) is the window as a function of time, \( x(t) \) is the root note, \( t \) is time, \( T \) is the duration of the root note, and \( T_{\text{ramp}} \) is the duration of the ramp.
1.5) Create a diotic version of each root note with the same signal in left and right channels, and a dichotic version of each root note with the signal in the right channel only.

1.6) Repeat steps 1.1. – 1.5 to create six diotic interval notes with the same signal in left and right channels, and six dichotic interval notes with the signal in the left channel only. See Table 3.1 for the ratio of each interval note F0 to root note F0.

1.7) Add each diotic root note to each diotic interval note to create 24 diotic dyads. Reduce the power of each dyad by half, to correct for both notes being played to both ears, i.e.

\[ y(t) = \frac{x_{\text{root}}(t) + x_{\text{interval}}(t)}{\sqrt{2}} \]

where \( y(t) \) is a diotic dyad with both notes in both channels, and \( x_{\text{root}}(t) \) and \( x_{\text{interval}}(t) \) are root and interval notes respectively.

1.8) Pair each dichotic root note with each dichotic interval note to create 24 dichotic dyads, i.e.

\[ y_{\text{right}}(t) = x_{\text{root}}(t) \]

\[ y_{\text{left}}(t) = x_{\text{interval}}(t) \]

where \( y_{\text{right}}(t) \) and \( y_{\text{left}}(t) \) are the right and left channel of a dichotic dyad.
### Table 3.1: Interval frequency ratios.

<table>
<thead>
<tr>
<th>Interval</th>
<th>F0 ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minor 2nd</td>
<td>1.06</td>
</tr>
<tr>
<td>Perfect 4th</td>
<td>1.33</td>
</tr>
<tr>
<td>Augmented 4th / Tritone</td>
<td>1.41</td>
</tr>
<tr>
<td>Perfect 5th</td>
<td>1.50</td>
</tr>
<tr>
<td>Major 6th</td>
<td>1.68</td>
</tr>
<tr>
<td>Major 7th</td>
<td>1.89</td>
</tr>
</tbody>
</table>

Each interval corresponds to a different F0 ratio. For example, the ratio of the F0 of the higher note to the F0 of the lower note in a Minor 2nd interval is 1.06.

1.9) Create wide-band noise low-pass filtered at 2000 Hz with a Gaussian amplitude distribution, duration 2000 ms, and an overall level of 80 dB SPL.

1.10) Calculate the spectrum level of the noise required for an overall level of 80 dB SPL, i.e.

\[ x_s = 80 - 10 \log_{10}(\omega_c) \]

where \( x_s \) is noise spectrum level in dB SPL, and \( \omega_c \) is the low-pass filter cut-off frequency in Hz (2000; see step 1.13).

1.11) Combine each of the 48 dyads with the Gaussian noise so that the noise precedes the dyad, and the noise and the dyad are separated by a 500 ms silence.
1.12) Filter each stimulus (i.e. each combined noise and dyad) with a 256th order low-pass finite impulse-response filter, with a pass band between 0–2000 Hz and a stop band starting at 4000 Hz.

### 3.5.2. Collection of behavioural data

Note: This procedure was first described elsewhere (McDermott et al., 2010).

2.1) Seat participants in a sound attenuating booth.

2.2) Present stimuli in a randomised sequence via headphones.

2.3) Allow a practice run consisting of one of each interval.

2.4) Collect responses via a computer. Instruct participants to rate each stimulus for its pleasantness from -3 to +3, with -3 indicating very unpleasant and +3 indicating very pleasant.

2.5) Collect responses from four runs of the complete set of 48 stimuli.

2.6) For each participant, calculate average pleasantness ratings across all runs for each dyad in each ear condition (diotic and dichotic).
2.7) Z-score the average pleasantness ratings of each participant.

2.8) Calculate individual consonance preference (see Fig. 3.3A) in each ear condition by subtracting the average z-scored rating of the three dissonant dyads (Minor 2\textsuperscript{nd}, Tritone, Major 7\textsuperscript{th}) from the average z-scored rating of the three consonant dyads (Perfect 4\textsuperscript{th}, Perfect 5\textsuperscript{th}, Major 6\textsuperscript{th}).

3.5.3. FFR stimuli

3.1) Repeat steps 1.1–1.12, excluding steps 1.9–1.11, to create six diotic and six dichotic FFR dyads, with the following substitutions: use a single root note with a 130.81 Hz F0; set the note duration to 120 ms.
3.2) Create a second version of each dyad with an onset polarity that is inverted by 180 degrees (see Fig. 3.4), i.e.

\[-y(t) = -1 \times y(t)\]

where \(-y(t)\) is the dyad with onset polarity inverted by 180 degrees.

3.3) Combine the two versions of the dyad, separated by 120 ms silence (Fig. 3.4).

3.5.4. **Collection of FFR data**

4.1) Use high-forehead hairline (active electrode), C7 vertebra (reference electrode), and Fpz (ground) electrode sites. Clean each electrode site with skin-cleansing alcohol wipes.

4.2) Rub each electrode site with skin-preparation gel using cotton swabs.

4.3) Attach each electrode with electroencephalography (EEG) conductive paste and medical tape.
4.4) Ensure impedances are ≤ 3 kΩ. Check impedances after recording FFR to each dyad to ensure they are maintained at this level throughout the procedure.

4.5) Deliver stimuli via insert-headphones. Sit participants in a sound attenuating booth on a comfortable reclining chair, and instruct that they should avoid movement, that they should close their eyes and that they can sleep throughout the session.
Note: Step 4.6 is optional. Where it is not possible, first perform recordings with the insert tubing clamped shut so that the stimulus is inaudible. Ensure that the recording does not contain a stimulus artefact (Picton, 2010) above the level of the background EEG prior to continuing.

4.6) Position the headphone transducers (the part of the headphone which converts electrical energy into sound) outside of the booth to prevent stimulus artefacts from contaminating the recording.

4.7) Collect FFR in two sessions on two separate days. In one session collect the FFR to six dyads in a randomised order.

4.7.1) Present each stimulus containing a dyad alternately in its original polarity and with the onset polarity inverted by 180 degrees at a rate of 2.09/s.

4.7.2) Band-pass filter responses between 30 and 3000 Hz

4.7.3) Record 2000 responses to each stimulus polarity (4000 responses in total).
4.7.4) Compile the FFR as sub-averages of 10 responses to each stimulus polarity.

4.8) In a second session, repeat steps 4.1-4.7 using the remaining six dyads.

3.5.5. **Analysis of the FFR**

5.1) Delete any sub-average in which the amplitude of the response exceeds +/- 15 µV.

5.2) Average the remaining sub-averages together.

Note: Step 5.3 involves a method described elsewhere for selectively enhancing and suppressing the phase locked response to the BM envelope or TFS, and even and odd order distortion products (Goblick & Pfeiffer, 1969; Rickman, Chertoff, & Hecox, 1991).

5.3) For each participant, extract the response in the grand average waveform to each stimulus polarity.

5.4) Add the response to one polarity from the response to the other polarity to create an addition waveform (FFR_{ADD}).
5.5) Subtract the response to one polarity from the response to the other polarity to create a subtraction waveform (FFR\textsubscript{SUB}).

5.6) Perform a fast-Fourier transform to calculate the power spectrum of each dyad used in the FFR recording.

5.7) For F0s ranging from 30–1000 Hz in 0.01 Hz step-sizes, measure the ratio of the power inside 4 Hz wide bins placed at the F0 and integer multiples of the F0 to power outside the bins, i.e.

\[
HS = \frac{\sum_{k=1}^{n} x_k}{P - \sum_{k=1}^{n} x_k}
\]

where HS is salience of the harmonic series of a given F0; k is harmonic number; \(x_k\) is the power inside a bin placed at harmonic number \(k\); and \(P\) is the total power of the spectrum.

5.8) Calculate the harmonic salience of the FFR\textsubscript{ADD} and FFR\textsubscript{SUB} waveforms from steps 5.4 and 5.5 by repeating steps 5.6–5.7 using the FFR instead of the dyad, and using the F0 of the best fitting harmonic series of the corresponding dyad calculated in step 5.7 (see Fig. 3.5).
Perform a fast-Fourier transform on the response to the two polarities separately, and then average together to calculate the power spectrum of the raw FFR (FFR\textsubscript{RAW}).
5.10) Repeat step 5.7 using the FFR_{RAW} power spectrum and the F0 of the best fitting harmonic series to the corresponding dyad (step 5.7).

5.11) For each participant, for each FFR type (FFR_{ADD}, step 5.4; FFR_{SUB}, step 5.5; FFR_{RAW}, step 5.9) calculate a neural consonance index by subtracting the average harmonic salience score (step 5.7) for the three dissonant dyads (Minor 2^{nd}, Tritone, Major 7^{th}) from the average harmonic salience of the three consonant dyads (Perfect 4^{th}, Perfect 5^{th}, Major 6^{th}; see Fig. 3.3B).

3.6 Representative results
This protocol provides a method for recording diotic and dichotic versions of three different FFR types: FFR_{RAW}, FFR_{ADD}, and FFR_{SUB} (see Fig. 3.6). Distortion products in the FFR resulting from monaural interactions will be present in the diotic (Fig. 3.6A-C) but not dichotic FFR (Fig. 3.6D-F). The FFR spectra in Fig. 3.6 are representative examples from a single participant in response to the consonant dyad the Perfect 5^{th}. Note that the salience of the harmonic series in the diotic FFR_{RAW} spectrum (Fig. 3.6A) relative to the dichotic FFR_{RAW} spectrum (Fig. 3.6D) is likely to be the result of monaural interactions. Note also that harmonics present in the diotic FFR_{RAW} and FFR_{ADD} spectra (Fig. 3.6A-B) but not the FFR_{SUB} spectrum (Fig. 3.6C) are likely to have been produced by phase locking to the BM envelope, since this is suppressed in the FFR_{SUB}. Comparison between pleasantness ratings for diotic and dichotic dyads and the different FFR types may be informative of
Figure 3.6: Spectra of the FFR to the Perfect 5th dyad for a single participant. The FFR to diotic (left column) and dichotic presentations (right column) are shown, and as FFRRAW (top row), FFRADD (middle row), and FFRSUB (bottom row). Power is plotted in dB referenced to $10^{-16}$ V^2. The black circle in each plot indicates the frequency of the difference tone $F_2 - F_1$. 
the role of monaural interactions in the perception of consonance. Moreover, the extent to which individual NCI scores are predictive of individual consonance preference scores is informative of the role of temporal coding of harmonicity in the perception of consonance.

3.7 Discussion

This protocol describes a method for assessing the relation between the representation of harmonicity in subcortical neural temporal coding and the perception of consonance. There are a number of potentially confounding variables which should be considered and screened for as necessary when using this method, depending upon the research question. One such confounding variable is hearing loss, since elevated auditory thresholds will lead to a reduction in the sensation level of the stimulus and therefore the amplitude of the FFR (e.g. see Yamada, Yamane, & Kodera, 1977). Furthermore, elevated auditory thresholds are usually the consequence of damage to or loss of the cochlear outer hair cells (OHC; Ruggero & Rich, 1991) responsible for the tuning of the cochlea (Ruggero, Rich, Recio, Narayan, & Robles, 1997) and therefore resolving individual harmonics: the loss of this function is associated with a deficit in the ability to make use of TFS information (e.g. see Hopkins, Moore, & Stone, 2008), and may therefore also be associated with a reduction in the temporal coding of harmonicity by phase locking (e.g. see Bidelman & Heinz, 2011).

The stimuli used in the protocol are all low-pass filtered at approximately the upper limit of phase locking in the brainstem (2000 Hz;
However, the FFR is likely to be partly generated by nerve fibres with CFs above the frequency range of the stimulus (Bones & Plack, 2014a, Dau, 2003); accordingly, where possible participants should be matched in auditory thresholds above 2000 Hz. Where this is impractical (e.g. in age studies, since age is strongly related to hearing loss at high frequencies; Moore, 2007) hearing thresholds should be taken into account when interpreting the results.

Age may also affect the FFR independently of hearing loss (Anderson, Parbery-Clark, White-Schwoch, & Kraus, 2012), potentially due to a decline in neural temporal coding (Frisina & Walton, 2006; Walton, Simon, & Frisina, 2002) and deafferentation of the auditory nerve (Sergeyenko, Lall, Liberman, & Kujawa, 2013). The FFR is also sensitive to musical experience (Bidelman, Gandour, & Krishnan, 2011; Bones et al., 2014; Lee et al., 2009; Wong, Skoe, Russo, Dees, & Kraus, 2007) and tonal language experience (e.g., Mandarin; Krishnan, Xu, Gandour, & Cariani, 2005).

The FFR is a measure which lends itself to multiple methods of analysis: latency and amplitude of the waveform (Anderson et al., 2012), cross-correlation of the waveform with the stimulus (Marmel et al., 2013), ACFs (Bidelman & Krishnan, 2009; Carcagno & Plack, 2011; Gockel, Carlyon, Mehta, & Plack, 2011; Krishnan & Plack, 2011; Smalt, Krishnan, Bidelman, Ananthakrishnan, & Gandour, 2012), and signal-to-noise ratios (Krishnan et al., 2005; Marmel et al., 2013) are some of the more common. The harmonic salience procedure described here can be considered a signal-to-noise ratio analysis, with the signal assumed to be the frequency components which correspond to the best fitting harmonic series to the
stimulus, and noise assumed to be all other frequencies. The protocol describes analysis of the FFR power spectrum, i.e. the power of the response as a function of frequency. However, this type of signal-to-noise ratio analysis can be performed on a number of other frequency-domain measures. Fig. 3.7 displays three alternative methods of analysing the FFR in the frequency domain, recorded from a single participant in response to a Perfect 5th dyad with a 220 Hz root note F0. The normalized power spectrum (Fig. 3.7A) is calculated by first using a simple windowing algorithm to smooth the spectrum of the EEG recorded during the gap between the two stimulus polarities. This produces an estimate of the frequency content of the background EEG, which is then used to partially normalize the spectrum of the FFR (compare to Fig. 3.6A). Phase-locking value (PLV; Figure 3.7B) is a measure of phase-coherence between two signals (Lachaux, Rodriguez, Martinerie, & Varela, 1999), or, as here, between multiple responses to the same stimulus. This method requires that the FFR is compiled in individual responses or sub-averages (here N = 10). Magnitude-squared coherence (MSC, Dobie & Wilson, 1989; Fig. 3.7C) is another measure of coherence between sub-averages. It takes both amplitude and phase into consideration, and represents the ratio of the power of the averaged response to the average power in each sub-average. Future applications of this protocol may find these analyses more appropriate than the power spectrum.

The method currently used most widely to explore the relation between the FFR and musical consonance is a period template analysis of the ACF of the FFR (Bidelman et al., 2011; Bidelman & Heinz, 2011; Bidelman & Krishnan, 2009, 2011). This method uses an ACF of the FFR
that is rectified so that all negative values are converted to positive values.

For comparison, Fig. 3.8 displays the analysis of a Perfect 5th stimulus waveform with a 220 Hz root note F0 using both the period template analysis and the harmonic salience method of the protocol. The period template analysis uses bins placed at different periods and their integer multiples of the rectified ACF (Fig. 3.8A). Pitch salience (Fig. 3.8B) is estimated as the ratio of the ACF within the template to the ACF outside the template. The ACF is typically weighted so that longer periods contribute less to the analysis (Bidelman & Krishnan, 2009). However, for comparison this operation has been excluded. As described in the protocol, the harmonic salience measure is an analysis of the power spectrum (Fig. 3.8C). The
harmonic salience analysis of the Perfect 5\textsuperscript{th} waveform finds the F0 of the best fitting harmonic series to be 55 Hz i.e. two sub-harmonics below the root note F0, closely followed by 110 Hz (Fig. 3.8D). When the ACF weighting function is excluded from the analysis the period template analysis also finds the period corresponding to this frequency to be dominant (Figure 3.8C). Despite work using this method typically reporting the period corresponding
to the root note F0 as the dominant period (likely due to weighting of the ACF; Bidelman & Heinz, 2011; Bidelman & Krishnan, 2009), both measures find a shared sub-harmonic of the two notes to best represent the dyad in this comparison. It should be noted therefore that whilst the measure described in this protocol has the advantage of making the role of distortion products (and frequency components more generally) apparent, it is closely related to the ACF.

3.8 Acknowledgements

This work was supported by the Economic and Social Research Council [ES/J500094/1]

3.9 Disclosures

The authors have nothing to disclose

3.10 References


Chapter 4. Phase locked neural activity in the human brainstem predicts preference for musical consonance

Oliver Bones

Kathryn Hopkins

Ananthanarayan Krishnan

Christopher J. Plack

1 School of Psychological Sciences, The University of Manchester, Manchester Academic Health Science Centre, Manchester, UK, M13 9PL

2 Department of Speech, Language, and Hearing Sciences, Purdue University, West Lafayette, Indiana, USA, IN 47907

This chapter was published as a paper:

4.1 Abstract

When musical notes are combined to make a chord, the closeness of fit of the combined spectrum to a single harmonic series (the ‘harmonicity’ of the chord) predicts the perceived consonance (how pleasant and stable the chord sounds; McDermott, Lehr, & Oxenham, 2010). The distinction between consonance and dissonance is central to Western musical form. Harmonicity is represented in the temporal firing patterns of populations of brainstem neurons. The current study investigates the role of brainstem temporal coding of harmonicity in the perception of consonance. Individual preference for consonant over dissonant chords was measured using a rating scale for pairs of simultaneous notes. In order to investigate the effects of cochlear interactions, notes were presented in two ways: both notes to both ears or each note to different ears. The electrophysiological frequency-following response (FFR), reflecting sustained neural activity in the brainstem synchronised to the stimulus, was also measured. When both notes were presented to both ears the perceptual distinction between consonant and dissonant chords was stronger than when the notes were presented to different ears. In the condition in which both notes were presented to both ears additional low-frequency components, corresponding to difference tones resulting from nonlinear cochlear processing, were observable in the FFR, effectively enhancing the neural harmonicity of consonant chords but not dissonant chords. Suppressing the cochlear envelope component of the FFR also suppressed the additional frequency components. This suggests that, in the case of consonant chords, difference tones generated by interactions between notes in the cochlea enhance the perception of consonance.
Furthermore, individuals with a greater distinction between consonant and dissonant chords in the FFR to individual harmonics had a stronger preference for consonant over dissonant chords. Overall, the results provide compelling evidence for the role of neural temporal coding in the perception of consonance, and suggest that the representation of harmonicity in phase locked neural firing drives the perception of consonance.

4.2 Keywords
Musical consonance; harmonicity; individual differences; auditory brainstem; frequency-following response; pitch.

4.3 Introduction
When two musical notes are played together, such as when two piano keys are pressed simultaneously, the result is a type of chord called a ‘dyad.’ For example, the consonant dyad the Perfect 5\textsuperscript{th} is created by combining a lower note with a note that is seven keys higher on a piano. The term ‘consonance’ is used to describe combinations of notes which result in a pleasing perception of resolution and stability. In contrast ‘dissonance’ is used to describe combinations which produce an unpleasant perception of being unresolved and unstable. The fact that consonant combinations are deemed preferable to dissonant combinations (Plomp & Levelt, 1965; McDermott et al, 2010) contributes to a sense of musical key (Rameau, 1722/1971). Music is a universal human phenomenon (e.g. McDermott & Hauser, 2005); an understanding of the auditory processes involved in listening to music may
reveal which musical parameters are determined by innate factors, and uncover the mechanisms which are defective in those individuals with an impaired perception of the distinction between consonance and dissonance, and therefore an impaired enjoyment of music (e.g. those with sensorineural hearing loss; Tufts, Molis, & Leek, 2005).

An individual note produced by a musical instrument is an example of a complex tone. The spectrum of a complex tone contains a number of frequency components called ‘harmonics’ at integer multiples of the fundamental frequency \(F_0\). For example, the musical note ‘C’ has an \(F_0\) of 130.81 Hz and harmonics at 261.62, 392.43, 523.24, 654.05 Hz etc. The \(F_0\) is also the frequency of the overall repetition rate of the waveform. When a complex tone enters the cochlea the low-numbered harmonics are separated out (‘resolved’), with each harmonic exciting a distinct place on the basilar membrane (BM). Information about the frequency of resolved components is encoded in the auditory nerve by the tendency of auditory nerve fibres to synchronise their firing to the *temporal fine structure* (TFS) of the BM vibration (‘phase locking’; Brugge, Anderson, Hind, & Rose, 1969; Cariani & Delgutte, 1996; Rose, Hind, Anderson, & Brugge, 1971) so that the corresponding period of each resolved harmonic is represented in the inter-spike intervals (ISI) of the fibres innervating a place on the BM corresponding to that frequency. In addition, neurons will also tend to phase lock to the overall amplitude of the vibration over time (the *temporal envelope*) produced by the interactions of harmonics on the BM (Cariani & Delgutte, 1996; Hall, 1979). The dominant modulation rate of these interactions is equal to the
frequency spacing of the harmonics (the F0 in the case of harmonic complex
tones).

The scalp-recorded electrophysiological frequency-following response
(FFR) is a measure of neural phase locking in the brainstem (Moushegian,
Rupert, & Stillman, 1973). The FFR has been widely used to explore the
relation between temporal coding at this stage of the auditory pathway and
pitch perception (Anderson, Parbery-Clark, White-Schwoch, & Kraus, 2012;
Anderson, White-Schwoch, Parbery-Clark, & Kraus, 2013; Bidelman,
Gandour, & Krishnan, 2011a, 2011b; Bidelman & Heinz, 2011; Bidelman &
Krishnan, 2009; Carcagno & Plack, 2011; Clinard, Tremblay, & Krishnan,
2010; Gockel, Carlyon, Mehta, & Plack, 2011; Krishnan, Bidelman, &
Gandour, 2010; Krishnan, Bidelman, Smalt, Ananthakrishnan, & Gandour,
2012; Krishnan & Plack, 2011; Krishnan, Xu, Gandour, & Cariani, 2005;
Wong, Skoe, Russo, Dees, & Kraus, 2007). The FFR to a complex tone
contains phase locked activity to both the cochlear envelope and the TFS.
However, Goblick & Pfeiffer (1969) developed a method for selectively
enhancing the FFR to either of these components: the FFR is averaged
separately across trials with the stimulus presented in the original onset
polarity and across trials with the onset polarity inverted 180 degrees (see
Fig. 4.1). Adding the two averages together has the effect of suppressing the
FFR to the TFS whilst enhancing the response to the envelope; subtracting
the averages has the opposite effect of suppressing the response to the
envelope whilst enhancing the response to the TFS. This technique was
employed in the present study to explore the separate contributions made by
the neural representation of the temporal envelope and of the TFS of musical dyads in the perception of musical consonance.

Explanations for the perception of consonance have been debated for many hundreds of years. Observing the behaviour of plucked strings, the ancient Greeks noted that when two notes are combined simple vibration ratios produce consonant sounds whilst complex ratios produce dissonant sounds. For example, the frequency ratio of the highly consonant Perfect 5th interval is approximately 3:2, whereas the ratio of the dissonant Tritone interval is approximately $\sqrt{2}:1$. A consequence of complex frequency ratios is that the combined frequency spectrum frequently contains harmonics which are closely spaced on the BM. Many modern accounts of consonance and dissonance have been informed by Helmholtz’s (1877/1954) theory of auditory ‘beating’ (e.g. Plomp, 1964; Plomp & Levelt, 1965; Rasch & Plomp,

Figure 4.1: Sections of two modulated waveforms. The frequency content of the two waveforms is the same but the second waveform (B) has an onset polarity inverted by 180 degrees with respect to the onset polarity of the first waveform (A). Phase locking to the TFS (black lines) of the two waveforms will therefore be in opposite phase. However, phase locking to the envelope (blue lines) of the two waveforms will be in the same phase.
This theory states that the perception of dissonance occurs when harmonics which are closely spaced on the BM interact with one another, causing amplitude modulation (beating) and a sensation of ‘roughness’. The auditory beating theory suggests that consonance is the perception that occurs in the absence of roughness. It is clear however that this is an insufficient explanation for musical consonance, e.g. at low F0s even consonant frequency ratios result in dyads with low-frequency components that are closely spaced enough for beating to occur (Terhardt, 1974).

The magnitude spectra of dyads with simple frequency ratios such as the Perfect 5th interval closely resemble a single harmonic series and therefore a single musical note. McDermott et al. (2010) demonstrated that the perception of consonance is likely to be driven by the closeness of fit of the combined harmonics from the two notes to a single harmonic series (the ‘harmonicity’; see also Cousineau, McDermott, & Peretz, 2012). Individuals with a stronger preference for harmonicity in non-musical tones also had a stronger preference for consonant over dissonant dyads. Individual ratings of the unpleasantness of beating on the other hand were not related to preference for consonance. Musical experience was found to significantly correlate with preference for both harmonicity and consonance, but not for absence of beating, further indicating the relation between harmonicity and music. The results of this study provide compelling evidence that it is harmonicity resulting from simple frequency ratios that drives the perception of consonance, rather than simply the absence of beating.

There are a number of reasons why harmonicity might be important to the perception of consonance. Mathematical models of sub-cortical ISI

1999).
processing have been used to explain psychophysical phenomena such as frequency discrimination (Meddis & Hewitt, 1991a, 1991b; Meddis & O'Mard, 1997) and the perception of consonance (Ebeling, 2008). In Ebeling's model it is the coincidence of neural firing when frequency components are harmonically related (and therefore have overlapping periods) that leads to the perception of consonance. In a bid to explain 'virtual pitch' (the pitch produced by a harmonic series consisting only of harmonics above a 'missing' F0) Terhardt (1974, 1979, 1984) proposed a harmonic template detection schema. In this schema pitch is determined by matching the combined frequency components of a sound to the best fitting harmonic series by finding the sub-harmonic (the low frequency component not present in the stimulus with a frequency of f/n where n is an integer) shared by the most harmonics present. Consonance is the perception of 'tonal affinity' when the combined spectra of two or more complexes have a strong 'fundamental-tone relation' (Terhardt, 1977). Similarly Stumpf (1890) conceived of consonance as the perceptual consequence of 'tonal fusion', i.e. the perception that occurs when two notes perceptually fuse into a single entity. One way in which two musical notes might fuse is by their harmonics closely resembling a single harmonic series with a single F0 and therefore a single pitch. A similar idea was held by 18th century German philosopher Rameau (1722/1971) who believed that the individual notes of a consonant chord were related to one another by a 'fundamental bass note,' i.e. a shared sub-harmonic. Models of consonance based on neuronal periodicity detection schemes (e.g. Ebeling, 2008) use information coded by the phase locking of neurons in the subcortical auditory pathway. Tramo, Cariani,
Delgutte & Braida (2001) observed that the dominant periods in consonant stimuli (corresponding to a shared sub-harmonic of the individual notes) were also the dominant periods in the pooled ISI histograms (ISIHs) recorded from the auditory nerve fibres of cats. Meanwhile, previous FFR studies have found that musical interval has a significant effect on the salience of the period corresponding to the F0 of the lower note of a dyad in phase locked brainstem activity, and that there was a strong correlation between the consonance of an interval and the average salience of this period (Bidelman & Krishnan, 2009, 2011). One aim of the present study was to test the hypothesis that the perception of consonance can be accounted for by the neural representation of harmonicity.

If consonance depends on the neural representation of harmonicity, it might be expected that perception would be dependent on the way the notes are presented to the two ears. When two notes of a consonant chord are presented to the same ear (or both notes to both ears; ‘diotically’), due to the regular spacing of the combined harmonics the interactions between the two notes on the BM produce temporal envelopes containing frequency components that are harmonically related to the components of the combined spectrum. For example, the F0s of the two notes of the C Perfect 5th dyad are 130.81 and 196.00 Hz. Therefore an envelope component of approximately 65 Hz will be present (see Fig. 4.2). The three frequencies are all approximately harmonics of a 65 Hz F0. Hence, phase locking to the envelope produced by harmonic interactions between the notes in a consonant chord may effectively reinforce the harmonicity of the neural representation of the combined spectrum. When musical notes enter
Figure 4.2: Monaural interactions between the harmonics of a consonant musical dyad (the Perfect 5th). Solid lines indicate root note harmonics; dotted lines represent interval note harmonics (A). The combined spectrum contains harmonics which are regularly spaced and often separated by a harmonic frequency difference. Where harmonics coincide (the third harmonic of the root note, and the second harmonic of the interval note) the harmonic from the interval note has been slightly off-set so that both are shown. Although resolved from one another at places on the BM tuned to their respective frequencies, harmonics interact within the cochlear auditory filter centred at a place on the BM tuned to the frequency between the two harmonic frequencies (indicated here by the downwards arrow; B). The resulting waveform has a TFS (solid line) and an envelope (dotted line; C) Neurons will tend to phase lock to the envelope modulation rate (equal to the difference between the interacting frequency components; $F_2 - F_1$) as well as to the TFS (D).
different ears (‘dichotically’), cochlear interaction between F0s is not possible and therefore enhancement of harmonicity in this way cannot occur.

The present study was designed to test the following hypotheses. 1) Individual differences in preference for consonant dyads over dissonant dyads are related to individual differences in the relative strength of the temporally coded neural representation of the harmonicity of these dyads. 2) Presenting notes to the same ear (‘diotic’ presentation) results in a stronger perception of consonance compared with presenting notes to different ears (‘dichotic’ presentation), due to the contribution of temporal interactions between the harmonics of the two notes on the BM that reinforce the harmonic series. 3) Temporal interactions between the harmonics of the two notes in the diotic case lead to a stronger neural representation of the harmonic series compared with the dichotic case.

4.4 Materials and Methods

4.4.1 Participants.
Thirty-two young normal-hearing participants with no history of neurological disorders, speech or language difficulties, or tinnitus completed behavioural testing. Four were removed from the data set as outliers. These participants had consonance preference scores more than two standard deviations below the mean due to inconsistent interval ratings across trials (the mean ratings for each interval were approximately equal). Of the remaining 28 participants (mean age, 22 years; range, 18-30 years; 18 females) 19 completed both behavioural testing and electrophysiological testing (mean age, 22 years;
range, 18-27 years; eight females). Hearing ability was assessed using pure-tone audiometry. Hearing thresholds for all participants were 20 dB HL or better for frequencies ranging from 500 to 8000 Hz.

Participants completed a questionnaire on their experience playing musical instruments. They indicated how many hours per week they played their instruments and the number of years that this applied to, with the option of indicating different periods of practice. For example, they could specify that they played their instrument for 10 hours per week between the ages of 10 and 14, and then for 2 hours per week from the age of 14 until the age of 20. Musical experience was then estimated by calculating the total number of hours practice for each participant. Musical experience ranged from 0 hours to 13000 hours.

4.4.2 Behavioural testing

4.4.2.1 Stimuli
Stimuli were dyads (two note chords) made up of a lower note (the ‘root’ note) and a higher note (the ‘interval’ note, i.e. the note that defines the distance between the two notes and therefore the name of the dyad). Four root notes were used for the consonance preference testing, all taken from the equal-temperament scale: A (110.00 Hz); C (130.81 Hz); D# (155.56 Hz); and F# (185.00 Hz). These were each combined with six interval notes to produce 24 dyads. Each interval note is defined by its ratio to the root note (see Table 4.1). The resulting complexes contain frequency components from two harmonic series. In the case of consonant intervals such as the Perfect 5th, the combined harmonics form a spectrum with frequency components which approximate a single harmonic series (see Fig. 4.3A). In
the case of dissonant intervals such as the Tritone, the combined harmonics are irregularly spaced and there is no clear harmonicity (see Fig. 4.3B).

Each dyad was low-pass filtered at 2000 Hz. For each note the harmonics in the pass-band of the filter were of equal amplitude, set so that the overall level of each note was 80 dB SPL for dichotic presentation and 77 dB SPL for diotic presentation (to correct for the fact that two notes were presented to each ear in the diotic case). Each dyad was 2000 ms in duration, including 10 ms raised-cosine onset and offset ramps.

Each dyad was preceded by wide-band Gaussian noise with a 2 s duration including 10 ms raised-cosine onset and offset ramps. The noise was low-pass filtered at 2000 Hz. A 500 ms silence separated the wide-band noise and the dyad. The purpose of the noise was to break up the sequence of dyads in order to prevent melodic structure from influencing responses (McDermott et al., 2010). All stimuli were generated digitally at a sampling rate of 24,414 Hz with 32-bit resolution. Stimuli were delivered via a 24-bit EMU 0202 USB audio device and Sennheiser HD 650 supra-aural headphones.

<table>
<thead>
<tr>
<th>Interval (semitones)</th>
<th>Interval (name)</th>
<th>F0 (Hz)</th>
<th>Equal temperament ratio to root note</th>
<th>F0 of best fitting harmonic series (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Minor 2nd</td>
<td>138.59</td>
<td>1.05946</td>
<td>46.20</td>
</tr>
<tr>
<td>5</td>
<td>Perfect 4th</td>
<td>174.61</td>
<td>1.33483</td>
<td>43.71</td>
</tr>
<tr>
<td>6</td>
<td>Augmented 4th / Tritone</td>
<td>185.00</td>
<td>1.41421</td>
<td>37.11</td>
</tr>
<tr>
<td>7</td>
<td>Perfect 5th</td>
<td>196.00</td>
<td>1.49831</td>
<td>32.70</td>
</tr>
<tr>
<td>9</td>
<td>Major 6th</td>
<td>220.00</td>
<td>1.68179</td>
<td>43.71</td>
</tr>
<tr>
<td>11</td>
<td>Major 7th</td>
<td>246.94</td>
<td>1.88775</td>
<td>43.58</td>
</tr>
</tbody>
</table>

Table 4.1: The intervals of the dyads used in behavioural and electrophysiological testing
4.4.2 Procedure

In order to investigate the effects of temporal interactions created by presenting two notes to the same ear, two conditions were tested: one in which both notes were presented to both ears (diotic condition), and one in which the root note was presented to the right ear, and the interval note was presented to the left ear (dichotic condition). Behavioural ratings and individual consonance preference were measured by following the methodology of McDermott et al. (2010). Participants were instructed to rate
each dyad for pleasantness using a scale from -3 (very unpleasant) to +3 (very pleasant). Each run consisted of 48 stimuli (24 dyads, diotic and dichotic) presented in a random order. Responses from four runs were recorded, with all runs performed consecutively on the same day. Prior to the test, participants completed a practice run consisting of one of each interval in order to familiarise them with the procedure and the range of intervals used. Participants were seated in a sound-attenuating booth and responded via a keyboard and a computer display visible through a window in the booth.

4.4.2.3 Analysis

Ratings for each interval were averaged across runs for each presentation condition (diotic and dichotic). In order to calculate consonance preference, averaged ratings were first z-scored for each individual in order to remove the influence of individual differences in the use of the scale (McDermott et al., 2010). Consonance preference was then calculated by subtracting each individual’s average z-scored rating of the three theoretically most dissonant intervals (the Minor 2\textsuperscript{nd}, the Tritone, and the Major 7\textsuperscript{th}) from the average z-scored rating of the three most consonant intervals (the Perfect 4\textsuperscript{th}, the Perfect 5\textsuperscript{th}, and the Major 6\textsuperscript{th}), as determined \textit{a priori} by Western tonal music tradition (e.g. Rameau, 1722/1971).

4.4.3 Electrophysiology

4.4.3.1 Stimuli and recording procedure

Stimuli were a subset of the dyads used for the behavioural measures. The root note was C (130.81 Hz), taken from the equal-temperament scale, with the same six interval notes used in the consonance preference test (see Table 4.1). Each dyad was presented in a diotic condition and a dichotic
condition, meaning that 12 dyads were presented to each participant in a randomised sequence. Each stimulus was 120 ms in duration, including 10 ms raised-cosine onset and offset ramps. Filtering and presentation level was the same as for the behavioural procedure. Each presentation window contained two stimuli separated by 120 ms silence. In order to use the method described by Goblick & Pfeiffer (1969) for enhancing the FFR to either the cochlear envelope or the TFS, the onset polarity of the second stimulus in the pair was inverted 180 degrees with respect to the onset polarity of the first stimulus (see Fig. 4.1). Presentations consisting of the two stimuli repeated at a rate of 2.09/s. FFR waveforms were averaged across 2000 presentations of each polarity.

Participants were seated in a comfortable reclining chair in a sound-attenuating booth and told that they could sleep. Stimuli were delivered via a TDT RP2.1 Enhanced Real Time Processor, HB7 Headphone Driver, and Etymotic ER30 transducers. The length of the ER30 tubing connecting the transducers to the ear tips made it possible to position the transducers outside of the recording booth, therefore preventing stimulus artefacts from affecting the recording. This is a particular concern in FFR recordings because the electrophysiological response and the stimulus share the same frequencies. Recordings contaminated by the transduction of stimulus harmonics can easily be mistaken for neural activity phase locked to that frequency. Tubing was visually inspected for kinks before each session.

The FFR was recorded using TDT BioSig software with high-pass filtering at 30 Hz, low-pass filtering at 3000 Hz, and a notch filter at 50 Hz to remove mains electrical noise. A vertical electrode montage was used, with
an active electrode at the high forehead hairline, a reference electrode at the seventh cervical vertebra, and a ground at Fpz (Bidelman & Krishnan, 2009, 2011; Krishnan & Plack, 2011; Krishnan et al, 2005). Impedances were maintained below 5 kΩ. As the BioSig software did not permit continuous recording, data was compiled online as 200 sub-averages of 10 responses (the smallest sized sub-average permitted by the equipment) to each stimulus polarity. Any sub-average of ten sweeps in which the peak amplitude exceeded 30 μV at any time during the waveform was considered an artefact and removed offline. Responses were digitally high-pass filtered at 45 Hz offline in order to attenuate frequencies containing cortical responses further.

4.4.3.2 Analysis

In order to determine the strength of harmonicity in the FFR, a measure of harmonic salience was derived in the following way. First the best fitting harmonic series was determined for the power spectrum of each stimulus. A fast-Fourier transform (FFT) was performed for each stimulus waveform. The power spectrum was then analysed by measuring the power inside of 4 Hz-wide bins placed at integer multiple frequencies of the F0. Harmonic salience was calculated as the ratio between the sum of power inside the bins and the sum of power outside the bins for F0s ranging from 30-1000 Hz in 0.01 Hz step-sizes. 30 Hz is considered to be the lower limit of musical pitch (Pressnitzer, Patterson, & Krumbholz, 2001), whilst the upper limit was chosen so as to be well above the F0s of all of the notes used. Table 4.1 contains a summary of the F0s that resulted in the highest harmonic salience (taken to be the F0s of the best fitting harmonic series) for each stimulus.
This information was then used to analyse the FFR. An FFT was performed for each FFR, and the harmonic sieve analysis described above performed with an F0 corresponding to the best fitting harmonic series to the stimulus. This salience measure was used to estimate the strength of harmonicity in the neural response. The routine was implemented in MATLAB using a script adapted from Bidelman and Krishnan (2009). The procedure described here is an adaptation of the ‘pitch salience’ measure described by the same authors, who analysed the autocorrelation function of the FFR to derive a measure of the strength of the periodicity corresponding to the root note F0. The analysis of the autocorrelation function of the FFR described by Bidelman and Krishnan (2009) measures the strength of a given period (1/F0) and its integer multiples (n/F0) in the FFR waveform. The present method differs from that used by Bidelman and Krishnan in two respects: 1) the analysis was performed on the power spectrum of the response; and 2) the best harmonic fit was found for each stimulus as described above rather than using the root note F0 to determine the harmonic sieve. This method was used rather than the method employed by Bidelman & Krishnan (2009) since the purpose of the study was to explore the effects of the neural representation of harmonicity, and the harmonic series of the root note’s F0 (or its sub-harmonics n/F0) did not always correspond to the best fitting harmonic series of the stimulus (see Table 4.1). In addition, performing analysis on the power spectrum in the current study made apparent the role of individual frequency components in supporting harmonicity.

In order to measure the harmonicity of the FFR to consonant dyads relative to dissonant intervals for each participant, a neural consonance
index (NCI) was calculated using a method similar to that used to calculate behavioural consonance preference: each individual’s average harmonic salience score for the three dissonant intervals was subtracted from their average harmonic salience score for the three most consonant intervals, determined a priori as in the behavioural analysis.

An aim of this study was to assess the separate contributions made by the neural representation of the cochlear envelope and TFS to the perception of consonance. In order to do this, analyses were performed on three FFR 'types.' First, the FFR containing both cochlear envelope and TFS components was analysed (FFR_{RAW}) by performing the harmonicity analysis described above on the mean power spectrum of the FFR to the direct polarity and inverted polarity. Where FFR_{RAW} spectra are plotted they are averages from the direct and inverted polarity response i.e. the spectra upon which analyses were performed. In the interests of simplicity, rather than plot two waveforms (the responses to both polarity stimuli) where FFR_{RAW} waveforms are plotted they are responses to the direct polarity stimulus. Recording responses to a direct polarity and to an inverted polarity version of each stimulus allowed analyses to be performed on a second FFR type with the response to the envelope suppressed and TFS enhanced. The second FFR type was created by subtracting the FFR to the inverted stimulus polarity from the FFR to the direct stimulus polarity (FFR_{SUB}). The inverted polarity stimulus contains harmonics that are in opposite phase to those in the direct polarity stimulus, but has an envelope that is in the same phase (see Fig. 4.1). By subtracting the responses to the two waveforms, the contribution of phase locking to the temporal envelope component is reduced.
and the contribution of phase locking to the TFS (phase locking to individual harmonics) is enhanced. Thirdly, by adding the FFR to the direct stimulus polarity and the inverted polarity (FFR_{ADD}) phase locking to the envelope is enhanced and phase locking to TFS is suppressed (Goblick & Pfeiffer, 1969). An FFT was performed for each FFR type and harmonic salience calculated as described above. The NCI calculated from FFR_{RAW}, FFR_{SUB} and FFR_{ADD} is hereafter referred to as NCI_{RAW}, NCI_{SUB} and NCI_{ADD} respectively.

### 4.5 Results

#### 4.5.1 Individual differences in consonance preference and the harmonicity of phase locking

Average dyad pleasantness ratings are displayed in Fig. 4.4A. The pattern of ratings for different intervals is consistent with previous studies involving normal-hearing listeners (Bidelman & Krishnan, 2009; McDermott et al., 2010): the Perfect 5th was rated most pleasant, and the Perfect 4th, the Major 6th, and the Tritone were rated progressively less pleasant. In the diotic condition the Minor 2nd is rated as least pleasant, whereas in the dichotic condition the Major 7th is rated as least pleasant. Note also the effect of presentation condition (diotic or dichotic). The effects of presentation condition in Fig. 4.4A-D are summarised in Fig. 4.4E and are discussed in more detail in Sections 4.5.2 and 4.5.3.

The average harmonic salience for each interval for FFR_{RAW}, FFR_{SUB} and FFR_{ADD} are plotted in Figs. 4.4B-D respectively. Note that the plots are similar to the plot of pleasantness ratings in Fig. 4.4A, with the consonant
intervals resulting in higher harmonic salience than the dissonant intervals in each case. Overall, harmonic salience scores were lower for the FFR_{SUB} data than in the other two FFR types.

Interval pleasantness ratings are plotted as a function of harmonic salience in Fig. 4.4F. There was a strong correlation between harmonic salience and behavioural pleasantness rating (paired by interval) for each presentation condition and FFR type (FFR_{RAW} diotic, $r_{s(4)} = 0.94$, $p<0.01$, dichotic, $r_{s(4)} = 0.83$, $p = 0.02$; FFR_{SUB} diotic, $r_{s(4)} = 0.94$, $p < 0.01$, dichotic, $r_{s(4)} = 0.94$, $p = 0.01$, FFR_{ADD} diotic, $r_{s(4)} = 0.94$, $p < 0.01$, dichotic $r_{s(4)} = 0.77$, $p = 0.04$). Due to data being non-normally distributed, all reported correlation coefficients are Spearman’s Rho.

Consonance preference and NCI scores were averaged across diotic and dichotic conditions for each participant and tested for correlation (see Fig. 4.5A). In Bonferroni corrected one-tailed tests consonance preference was found to significantly correlate with NCI_{SUB} ($r_{s(17)} = 0.49$, $p = 0.02$), but not NCI_{RAW} $r_{s(17)} = -0.04$, $p=0.44$) nor NCI_{ADD} ($r_{s(17)} = -0.16$, $p = 0.26$). NCI_{SUB} represents the extent to which the phase locked neural firing to the TFS of consonant dyads contained more salient harmonicity relative to dissonant dyads. To test this correlation further individual consonance preference scores were again correlated with NCI_{RAW}, this time partialing out NCI_{ADD} and thus the confounding effect of individual variance in temporal envelope coding. The correlation coefficient rose from -0.04 to 0.36, consistent with there being a relation between NCI_{RAW} and consonance preference, although the correlation was marginally non-significant ($r_{s(16)} = 0.36$, $p = 0.07$).
As expected, musical experience was significantly correlated with consonance preference ($r_{s}(26) = 0.64$, $p < 0.01$; Fig. 4.5B). In Bonferroni corrected one-tailed tests musical experience significantly correlated with NClSUB ($r_{s}(17) = 0.63$, $p < 0.01$), but not NClRAW ($r_{s}(17) = -0.26$, $p = 0.14$) nor NClADD ($r_{s}(17) = -0.13$, $p = 0.30$). However, as for the consonance preference measure, the correlation between musical experience and NClRAW became stronger when controlling for the effect of variation in NClADD, this time
becoming significant \( (r_{s(16)} = 0.45, p = 0.03) \). These results indicate that experience of playing a musical instrument is strongly associated with enhanced harmonicity of consonant dyads relative to dissonant dyads in the phase locking to TFS. When controlling for the effect of music experience using partial correlation, NCI\textsubscript{SUB} was not found to correlate significantly with consonance preference \( (r_{s(16)} = 0.23, p = 0.18) \), suggesting that the relation between representation of harmonicity in the FFR and consonance preference may be driven by a co-dependence of each of these variables on musical experience.

Fig. 4.5C displays the relation between musical experience and the representation of harmonicity in FFR\textsubscript{SUB} to dissonant and consonant dyads. Musical experience is only weakly associated with harmonicity of dissonant dyads \( (r_{s}=0.28, p=0.25) \) but is strongly associated with harmonicity of consonant dyads \( (r_{s}=0.52, p=0.02) \). The strong correlation between musical experience and NCI\textsubscript{SUB} (green circles, solid line) is driven by the relation between musical experience and the harmonicity of the FFR to consonant dyads.

4.5.2 Behavioural ratings for diotic and dichotic stimuli
A two-way repeated-measures ANOVA with interval and presentation condition (diotic or dichotic) as factors confirmed a main effect of musical interval on pleasantness rating \( (F_{(5, 135)} = 210.60, p < 0.01) \). There was also a significant effect of presentation condition on pleasantness rating \( (F_{(1, 27)} = 16.50, p < 0.01) \) and a significant interaction between presentation condition and interval \( (F_{(5, 135)} = 54.60, p < 0.01) \). In Bonferroni corrected pair-wise comparisons (\( t \)-tests) all consonant intervals were rated as being more
Figure 4.5: Consonance preference scores as a function of NCI and musical experience. Individual consonance preference scores plotted as a function of NCI calculated from each FFR type (A). Consonance preference (left axis) and NCI calculated from each FFR type (right axis) as a function of musical experience (B). Harmonic salience of the FFR\textsubscript{SUB} to consonant (HarmSal\textsubscript{CONS}) and dissonant (HarmSal\textsubscript{DISS}) dyads (left axis), and NCI\textsubscript{SUB} (right axis) as a function of musical experience.
pleasant in the diotic condition than in the dichotic condition (Perfect 4\textsuperscript{th}, t\textsubscript{(27)} = 8.60, p < 0.01; Perfect 5\textsuperscript{th}, t\textsubscript{(27)} = 8.80, p < 0.01; Major 6\textsuperscript{th}, t\textsubscript{(27)} = 4.60, p < 0.01). Of the dissonant intervals, only the Minor 2\textsuperscript{nd} was more highly rated in the dichotic condition than in the diotic condition (t\textsubscript{(27)} = 5.10, p < 0.01). In other words, the interaction between presentation condition and interval was mainly driven by the fact that the consonant intervals tended to be more highly rated in the diotic condition. Importantly, consonant intervals being rated as more pleasant in the diotic condition (see Fig. 4.4A) meant that consonance preference scores calculated from pleasantness ratings of diotic stimuli were greater than consonance preference scores calculated from ratings of dichotic stimuli (t\textsubscript{(27)} = 4.10, p < 0.01).

4.5.3 Harmonic salience of the FFR for diotic and dichotic stimuli

A two-way repeated-measures ANOVA of harmonic salience in the FFR\textsubscript{RAW} with interval and presentation condition as factors found a significant main effect of interval (F\textsubscript{(5, 90)} = 65.35, p < 0.01) which interacted with presentation condition (F\textsubscript{(5, 90)} = 12.08, p < 0.01). Presentation condition was not a significant main effect. To explore the interaction between interval and presentation condition further, Bonferroni corrected pair-wise comparisons of each interval were performed between diotic and dichotic conditions. The Perfect 5\textsuperscript{th} (t\textsubscript{(18)} = 3.19, p = 0.05) had greater harmonic salience in the diotic condition and the Major 7\textsuperscript{th} (t\textsubscript{(18)} = -5.92, p < 0.01) had greater harmonic salience in the dichotic condition. Differences between presentation conditions were not significant for other intervals.

As can be seen in Fig. 4.4E, for FFR\textsubscript{RAW} the difference between the mean harmonic salience of consonant and dissonant intervals was greater
for diotic presentation than it was for dichotic presentation. Accordingly, the mean NCI_{RAW} score calculated for the diotic condition was significantly greater than the mean NCI_{RAW} score calculated for the dichotic condition ($t_{18} = 3.47, p = 0.03$), indicating that the difference between harmonic salience of the FFR of consonant and dissonant intervals was greater in the diotic condition than in the dichotic condition.

Fig. 4.6 displays the diotic FFR_{RAW} (Fig. 4.6A), the dichotic FFR_{RAW} (Fig. 4.6B), and the diotic FFR_{SUB} (Fig. 4.6C) waveforms to the Perfect 5th dyad. The harmonics of the stimulus root (F0 = 130.81 Hz) and interval note (F0 = 196.00 Hz) are frequently separated by approximately 65 Hz. The FFR spectra are displayed in the middle row (Fig. 4.6D-F) in which harmonic series with an F0 of approximately 65 Hz are identifiable. The spectrum of the FFR_{RAW} for the diotic condition (Fig. 4.6D) contains a peak at 65 Hz that is much larger than in the dichotic condition (Fig. 4.6E), as confirmed by a Wilcoxon Signed Rank Test ($V = 177.00, p < 0.01, r = -0.86$). Note also the strong representation of the period corresponding to this frequency (approximately 15 ms) in the FFR waveform for the diotic condition (Fig. 4.6A). Likewise, the diotic FFR_{RAW} spectra to the other consonant intervals also contained significantly larger frequency components than the dichotic spectra at the difference tone between the F0s of the root and the interval note (Wilcoxon Signed Rank Tests, the Perfect 4th, $V = 145.00, p = 0.02, r = -$

---

1 This frequency is a C note one octave below the root note, and corresponds to the second harmonic of a 32.70 Hz harmonic series. In the harmonic sieve analysis of the Perfect 5th stimulus, 65.40 Hz was marginally less salient than 32.70 Hz. However this is due to the greater number of bins in sieves with lower f0s. The 32.70 Hz series fits well because the harmonics coincide with those of the 65.40 Hz series.
0.52, the Major 6th, $V = 159.00$, $p < 0.01$, $r = -0.66$), which in these cases correspond to the F0s of the best fitting harmonic series indicated in Table 4.1. The power of difference tones was not significantly different between diotic and dichotic presentation for other intervals (see Fig. 4.6G).

The peaks at approximately 325 and 715 Hz (the 5th and 11th harmonics of a 65 Hz F0) in the spectrum of the diotic FFRRAW (Fig 4.6D) are also not present in the stimulus (see Fig 4.3) nor the dichotic FFRRAW spectrum (Fig 4.6E). These presumably also result from interactions between harmonics in the two notes.

As can be seen in Fig. 4.6C, suppression of the cochlear envelope component resulted in the FFRSUB Perfect 5th waveform being lower in amplitude than the FFRRAW waveforms. The large frequency component corresponding to the envelope component that is seen in the FFRRAW spectra is significantly reduced in the FFRSUB spectra (Wilcoxon Signed Rank Tests, Perfect 5th, $V = 188.00$, $p < 0.01$, $r = -1.04$, Perfect 4th, $V = 178.00$, $p < 0.01$, $r = -0.80$, Major 6th, $V = 190.00$, $p < 0.01$, $r = -1.09$; Fig. 4.6F).

To assess whether the NCIRAW was enhanced in the diotic condition due to the contribution of phase locking to the envelope, NCIRAW and NCISUB scores were compared. A two-way repeated measures ANOVA of NCI with presentation condition and FFR type as factors found FFR type to be a main effect ($F_{(1,18)} = 30.83$, $p < 0.01$) and that the two factors significantly interacted ($F_{(1, 18)} = 47.27$, $p < 0.01$), indicating that the effect of presentation condition on NCI depended upon contributions from the response to the envelope being included.
For FFR\textsubscript{SUB}, the difference between the mean harmonic salience of consonant and dissonant dyads was greater for the dichotic presentation (Fig. 4.4E). This is discussed further in Section 4.6.3.
4.6 Discussion

4.6.1 Individual differences in consonance preference are related to individual differences in neural temporal coding

Large variation in fidelity of phase locking at the level of the brainstem as measured by the FFR has been shown to occur amongst even young and normal-hearing listeners previously (Ruggles, Bharadwaj, & Shinn-Cunningham, 2012; Ruggles & Shinn-Cunningham, 2011). Previous studies using the FFR have demonstrated a correlation between individual strength of phase locking and individual performance in tasks associated with pitch perception. Marmel, et al (2013) found strength of phase locking to correlate with performance in frequency discrimination even after variation explained by age and by hearing loss had been partialed out. Similar correlations have also been demonstrated for F0 discrimination tasks (Bidelman et al, 2011a; Krishnan et al, 2010; Krishnan et al, 2012). In the present study, individual consonance preference significantly correlated with NCI\textsubscript{SUB}. NCI is a measure of the salience of harmonicity in brainstem phase locking to consonant dyads relative to dissonant dyads. NCI\textsubscript{SUB} is calculated from phase locking mainly to TFS. These results suggest that, despite envelope components driving an increase in both harmonicity of neural coding and consonance preference in the diotic condition when compared to the dichotic condition (see Section 4.6.3), individual variation in the acuity of neural coding of harmonically relevant TFS (i.e. individual harmonics) differentiates individual consonance preference.

The study also demonstrates a relation between individual musical experience and both individual behavioural and physiological measures of
consonance. McDermott et al (2010) demonstrated a correlation between musical experience and both harmonicity preference and consonance preference. Previous studies have demonstrated an effect of musical training on F0 contours in the FFR (Bidelman et al, 2011b; Wong et al, 2007) and representation of musically relevant features of the spectrum of the FFR addition waveform (Lee, Skoe, Kraus, & Ashley, 2009). The results reported here suggest that musical experience results in enhanced phase locking to TFS but not to envelope.

4.6.2 Diotic presentation results in both stronger consonance preference and stronger neural representation of harmonicity for consonant intervals

The results here demonstrate that ear presentation condition (diotic or dichotic) impacts upon the perceived pleasantness of consonant intervals, with consonant intervals being perceived as more pleasant when they are presented diotically (both notes to both ears) than when they are presented dichotically (each note to a different ear). The increased pleasantness of consonant intervals in the diotic condition corresponds to a greater preference for consonant over dissonant dyads. To the authors’ knowledge this is the first time that this effect has been demonstrated.

Previous work has provided evidence for a relation between the perception of consonance and pitch-relevant temporal information at the level of the human brainstem (Bidelman & Heinz, 2011; Bidelman & Krishnan, 2009, 2011). This earlier work measured the integrity of phase locking to the stimulus by examining the period corresponding to the F0 of the root note. Similarly, here it is demonstrated that consonance is
accounted for by the strength of harmonicity of the FFR. Moreover, the present study demonstrates an effect of presentation condition on harmonicity of the FFR to consonant dyads, and provides evidence that this is due to the addition of harmonically relevant components resulting from interactions in the cochlea (see Section 4.6.3). The present study builds upon previous behavioural work demonstrating that harmonicity is likely the driver of the perception of consonance (McDermott et al., 2010) by providing evidence that phase locking to the frequency components of the combined spectrum of a musical dyad may be the physiological mechanism by which harmonicity is encoded in the auditory periphery.

4.6.3 Harmonicity of the FFR is enhanced in the diotic condition due to the addition of components produced by cochlear interactions

The results of the present study indicate that, whilst individual variation in phase locking to TFS differentiates individual preference for consonance, phase locking to the amplitude modulation of the BM response drives a general preference for diotic over dichotic musical dyads. In the case of consonant intervals, the frequency of the amplitude modulation (the temporal envelope) corresponds to the best fitting F0 of the combined harmonics of the two notes of the dyad.

Tramo et al (2001) used autocorrelation to demonstrate that, for consonant dyads, the dominant period in the stimulus was also represented in the pooled all-order ISIs from 100 cat auditory nerve fibres. The authors suggested that phase locking in the auditory nerve could therefore be a mechanism for extracting this frequency from consonant musical stimuli. In the present study it was hypothesised that this frequency component would
be greater in magnitude in the FFR to diotic dyads compared to dichotic
dyads, due to cochlear interactions between the harmonics of the two notes
producing additional frequency components corresponding to the best fitting
F0 (the sub-harmonic identified by Tramo et al). The results of this study
provide evidence supporting this hypothesis. The F0 component
representing the harmonic series of the combined spectrum of the two notes
of consonant dyads is dominant in the diotic FFR\textsubscript{RAW}. That these frequency
components are significantly reduced in the dichotic FFR\textsubscript{RAW} suggests that
they arise mainly from monaural processing, most likely interactions on the
BM.

Attempts have been made to estimate the power of propagated
cochlear distortion products previously, with evidence that the FFR to
complex stimuli contains frequency components much larger than would be
expected were they to have been generated in this way (Gockel, Farooq,
Muhammed, Plack & Carlyon, 2012). Hence it is likely that the components
corresponding to the envelope frequency arise from quadratic distortion in
the transduction of the interacting harmonics, coded in the output of high-
frequency channels (Dau, 2003; Geisler, Rhode, & Kennedy, 1974; Kiang &
Moxon, 1974). To test the hypothesis that additional frequency components
in the FFR in the present study represent the envelope produced by the
interaction of the two notes in the cochlea, the spectra of the FFR containing
phase locking to the envelope (FFR\textsubscript{RAW}) and of the FFR with phase locking
to the envelope suppressed (FFR\textsubscript{SUB}) were compared. It was found that
these components were not present in the diotic FFR\textsubscript{SUB}, suggesting that this
frequency component was produced by phase locking to the envelope of the cochlear response.

Phase locking to TFS is likely to be the primary mechanism for the coding of pitch (for review see Plack & Oxenham, 2005). However, Moore and Moore (2003) demonstrated that the pitch of complexes consisting of only unresolved harmonics is likely to be determined by envelope rate: when the spectral envelope (and therefore BM excitation pattern) was held constant for complexes in which the harmonics were all shifted upwards in frequency by the same amount (therefore retaining the same envelope rate), the perceived pitch of complexes containing resolved harmonics shifted in proportion to the shift in frequency. However the pitch of complexes containing only unresolved harmonics remained the same, suggesting that for these complexes pitch corresponded to the unchanged envelope rate.

Houtsma and Smurzynski (1990) demonstrated that performance in frequency discrimination and pitch identification tasks was better for complexes containing resolved harmonics than for complexes containing only unresolved harmonics. However, performance when using complexes containing only unresolved harmonics improved with increasing number of harmonics present. This presumably demonstrates that the representation of the F0 in the envelope produced by interactions on the BM is enhanced with increasing number of unresolved harmonics. This could indicate the existence of two mechanisms for pitch perception: a primary pattern matching mechanism (e.g. Goldstein, 1973; Terhardt, 1979) dependant on access to TFS information; and a secondary mechanism for deriving pitch information from the envelope information produced by unresolved
harmonics as suggested by Schouten (1940, cited in Houtsma & Smurzynski, 1990). The results of the current study suggest that this secondary mechanism could play an important role in ‘fusing’ the temporal information produced by the two notes of a musical dyad. As suggested by Tramo et al (2001), the additional low-frequency components present in the temporal coding of consonant dyads in the diotic condition correspond to the theoretical ‘fundamental bass’ note which Rameau (1722/1971) believed to relate the individual notes of a consonant dyad to one another. The present study demonstrates phase locking at a frequency corresponding to an F0 that defines the harmonicity of the dyad, suggesting that, when not artificially separated by dichotic presentation, phase locking to frequencies produced by cochlear interactions might serve as a mechanism for reinforcing the fusion of the two notes into a single ‘entity’ in the manner suggested by Stumpf (1890).

The fact that an additional low-frequency component is identifiable at all in the dichotic FFR for consonant intervals (Fig. 4.6E) is noteworthy. A component corresponding to the difference tone between the F0s of the two notes is significantly larger in the FFR_{raw} dichotic spectrum compared to the FFR_{sub} diotic spectrum for the Perfect 4^{th} (V = 155.00, p = 0.01, r = -0.62), Perfect 5^{th} (V = 155, p = 0.01, r = -0.62) and Major 6^{th} (V = 166.00, p < 0.01, r = -0.68). This suggests that the FFR may reflect interactions between harmonics at or after binaural integration in the superior olivary complex. Previous work has suggested that the FFR is not sensitive to such binaural interactions (Gockel, Carlyon, & Plack, 2011).
Interestingly, the results here demonstrate that when the response to the cochlear envelope is suppressed and the response to individual harmonics enhanced (FFR\textsubscript{SUB}), the harmonic salience of consonant dyads is significantly greater for dichotic dyads. Closer examination of spectra of the FFR\textsubscript{SUB} to diotic and dichotic dyads reveals the absolute magnitude of the harmonic frequencies of the consonant dyads to be lower in the diotic case, therefore reducing the harmonic salience score. This may be an effect of the nonlinearity of the cochlea, made more extreme by the subtraction routine: where harmonics imperfectly coincide (due to the use of the equal temperament scale, i.e. see equal temperament ratio to root note, Table 4.1) the response in the diotic case may be reduced compared to the dichotic case due to monaural neural suppression, i.e. the neural synchrony of an auditory nerve fibre’s response to a tone is suppressed by the addition of a second tone with a frequency slightly above or below the response region of the first tone (e.g. see Arthur, Pfeiffer & Suga, 1971). If this is the case, it would also be expected that frequency components resulting from central binaural interactions would be lower in magnitude in the diotic case.

4.6.4 Conclusions

Consonance preference for different musical intervals corresponded closely to the neural representation of harmonicity reflected in the FFR. Furthermore, individuals with a greater preference for consonance had a greater distinction between the representation of harmonicity in consonant and dissonant dyads in the FFR generated by phase locking to individual harmonics. When two notes of a consonant dyad were presented to the same ear, the dyad was perceived as being more consonant than when the two notes were presented
to separate ears. The FFR also revealed a stronger neural representation of harmonicity for consonant diotic dyads. When two notes were presented to the same ear, interactions between the harmonics of the two notes on the BM resulted in additional frequency components being present in the FFR. These components enhanced the harmonicity of the FFR, suggesting that this could be the physiological mechanism for the increased perception of consonance in the diotic condition. Overall, the results suggest that consonance preference depends in part on the sub-cortical neural temporal representation of harmonics and their cochlear interactions.

4.7 Acknowledgements

This work was supported by the Economic and Social Research Council UK (ESRC) and The Society for Education, Music and Psychology Research (SEMPRE).

4.8 References


Chapter 5. Subcortical representation of musical dyads: individual differences and neural generators

Oliver Bones

Christopher J. Plack

School of Psychological Sciences, The University of Manchester, Manchester Academic Health Science Centre, Manchester, UK, M13 9PL

This chapter was prepared for publication as a paper. It is currently under review at Hearing Research.

Publication 3
5.1 Abstract

When two notes are played simultaneously they form a musical dyad. The sensation of pleasantness, or ‘consonance’, of a dyad is likely driven by the harmonic relation of the frequency components of the combined spectrum of the two notes. Previous work has demonstrated a relation between individual preference for consonant over dissonant dyads and the strength of neural temporal coding of the harmonicity of consonant relative to dissonant dyads as measured using the electrophysiological ‘frequency-following response’ (FFR). However, this work also demonstrated that both these variables correlate strongly with musical experience. The current study was designed to determine whether the relation between consonance preference and neural temporal coding is maintained when controlling for musical experience. The results demonstrate that strength of neural coding of harmonicity is predictive of individual preference for consonance even for non-musicians. An additional purpose of the current study was to assess the cochlear generation site of the FFR to low-frequency dyads. By comparing the reduction in FFR strength when high-pass masking noise was added to the output of a model of the auditory periphery, the results provide evidence for the FFR to low-frequency dyads resulting in part from basal cochlear generators.

5.2 Keywords

Musical consonance; harmonicity; temporal coding; frequency-following response; neural generation site; auditory brainstem.
5.3 Introduction

‘Consonance’ refers to the pleasant and stable sensation produced when two or more musical notes are presented simultaneously in simple frequency ratios. McDermott, Lehr, and Oxenham (2010) demonstrated that individual preference for consonant over dissonant two-note musical chords (‘dyads’) correlated with preference for harmonicity (the closeness of fit of a series of frequency components to a harmonic series) over inharmonicity, suggesting that the perception of consonance is driven by the perceived harmonicity of the dyad.

Pitch may be represented in the brainstem due to the tendency of neurons to synchronize their firing to a particular phase of the basilar membrane (BM) vibration (‘phase locking’; Brugge, Anderson, Hind, & Rose, 1969; Rose, Hind, Anderson, & Brugge, 1971). The sustained phase locked response of populations of neurons at this stage of the auditory pathway can be measured as the frequency-following response (FFR), a scalp recorded auditory evoked potential which takes its name from the characteristic peaks in the waveform at periods corresponding to frequency components of the stimulus. Recent work suggests a relation between the integrity of the temporal coding represented by the FFR and pitch discrimination (Carcagno & Plack, 2011; Clinard, Tremblay, & Krishnan, 2010; Krishnan, Bidelman, Smalt, Ananthakrishnan, & Gandour, 2012; Marmel et al., 2013), musical experience (Bidelman, Gandour, & Krishnan, 2011a, 2011b; Wong, Skoe, Russo, Dees, & Kraus, 2007), tone language experience (Krishnan, Xu, Gandour, & Cariani, 2005), and the perception of musical consonance (Bidelman & Krishnan, 2009; Bones, Hopkins, Krishnan, & Plack, 2014).
Bones et al. (2014) demonstrated that individual differences in consonance perception for dyads could be predicted by the representation of harmonicity in the spectrum of the FFR. By subtracting the FFR to each stimulus presented in its original onset polarity from the FFR to the stimulus presented with the onset polarity inverted, phase locking to the cochlear envelope was suppressed whilst phase locking to temporal fine structure (TFS) was enhanced (Goblick & Pfeiffer, 1969). When the FFR was processed this way young normal hearing participants with a stronger representation of the harmonicity of consonant relative to dissonant dyads in the spectrum of the FFR (‘neural consonance index’; NCI) had a stronger preference for consonant over dissonant dyads. This suggests that temporal coding of the frequency components of the combined spectrum of two notes may be a mechanism for encoding the harmonicity, and consonance, of musical dyads. However, musical experience is also strongly associated with consonance preference (Bones et al., 2014; McDermott et al., 2010) and harmonicity preference (McDermott et al., 2010). Bones et al. (2014) found that the correlation between NCI and consonance preference did not remain significant when the effect of musical experience was controlled, suggesting that the relation between integrity of the representation of harmonicity in the brainstem and consonance preference could be driven by a codependence on musical experience. One purpose of the current study was to address the outstanding question of whether harmonicity of temporal coding in the brainstem as measured by the FFR predicts variation in consonance preference in individuals without musical experience. If so, this would support the hypothesis that consonance is associated with neural temporal coding,
rather than with some other aspect of processing that might co-vary with musical experience.

Since the upper limit of phase locking in the inferior colliculus is approximately 2000 Hz (Krishnan, 2007), the stimuli used to measure the FFR are typically below this frequency. Another issue in this area that is currently unresolved is that of which region of the cochlea is represented by the FFR to low-frequency dyads. The response to a low-frequency pure tone at low to moderate levels (approximately < 50 dB above hearing threshold; Krishnan, 2007) is likely to be generated in the region of the cochlea with characteristic frequency (CF) close to the frequency of the tone, and is measurable in listeners with high-frequency hearing loss (Moushegian, Rupert, & Stillman, 1978; Yamada, Yamane, & Kodera, 1977). However, the suprathreshold FFR to a low-frequency tone is reduced in amplitude by high-pass masking noise above the frequency of the tone (Bledsoe & Moushegian, 1980; Davis & Hirsh, 1976; Gardi & Merzenich, 1979; Huis in't Veld, Osterhammel, & Terkildsen, 1977). Gardi and Merzenich (1979) found that the response to a 500 Hz tone presented at a high level (100 dB SPL) was reduced in amplitude by approximately 50% by 60 dB SPL high-pass (2000 Hz) masking noise, suggesting that the FFR may have been at least partly generated basally to the region of the BM tuned to the tone. This reduction presumably represents the desynchronisation (Marsh, Smith, & Worden, 1972) of high-frequency neurons which in the absence of the masker had been phase locked to the stimulus frequency. More recently, it has been demonstrated that the FFR to a low-frequency tone is reduced in amplitude when preceded by a high-frequency tone, implying that neural
adaptation of neurons in a high-frequency channel can attenuate the response to a low-frequency tone (H. Gockel, personal communication, May 22, 2014). Again, this suggests that the FFR is generated basally to the region tuned to the tone. This can be partly understood as a consequence of the frequency response of the basal region of the BM, which is reflected in the tuning of auditory nerve (AN) fibers: high-frequency fibers have sharp peaks at their best frequencies and steep high-frequency sides but broad low-frequency tails (Geisler, Rhode, & Kennedy, 1974; Kiang & Moxon, 1974; Rose et al., 1971). High level tones such as the 100 dB SPL tone used by Gardi and Merzenich (1979) cause the BM response to broaden and the neural response of fibers tuned to the tone to become saturated. As a consequence fibers innervating the basal side of the BM response come to dominate the neural response (de Boer, 1977). Dau (2003) suggests that displacement of the basal region of the BM may be necessary to produce a well-defined FFR. Dau used an AN model (Heinz, Zhang, Bruce, & Carney, 2001) to simulate the FFR to a 300 Hz pure tone. When only 100-1500 Hz channels were included in the model the frequency of the tone was poorly represented by the simulated FFR and the waveform was low in amplitude up to the highest stimulus level used (100 dB SPL). However, when only 1500-10000 Hz channels were included the simulated FFR to stimuli of 80 dB SPL and above had a periodic response, increasing in amplitude as the tone was increased in level. In the high-frequency channel model the synchrony of the modelled response of the BM in the base of the cochlea (e.g. see Dau, Wegner, Mellert, & Kollmeier, 2000; Shore & Nuttall, 1985) meant that the spikes in the model's nerve fibers were well aligned. The
frequency of the stimulus was therefore well represented in the FFR simulated from the pooled response. In the low-frequency channel model the nerve fibers responded asynchronously, due to the asynchronous response of the BM at the apex of cochlea, resulting in a poorly defined FFR.

It is likely that the FFR to high-intensity low-frequency pure tones is at least partly generated basal to the region of the BM tuned to the tone. However, whilst the cochlear origin of the FFR for pure tones has been well researched, that for complex tones has not. A second purpose of the current study was to test the hypothesis that the FFR to 80 dB SPL low-frequency musical dyads originates from a portion of the cochlea basal to that tuned to the dyads. To address this, the effects of high-pass masking noise on the FFR were compared to the output of an auditory model.

5.4 Methods

5.4.1 Participants
All participants had normal hearing (absolute thresholds of 20 dB HL or better at octave frequencies between 500 and 8000 Hz). Thirteen (10 females) participated in the psychophysical section of the study (20–28 yrs, mean 23 yrs). Ten (7 females) participated in both the psychophysical and the electrophysiological part of the study (20–28 yrs, mean 24 yrs). All participants self-reported as having less than one year of experience learning to play a musical instrument, with that period having ceased at least five years prior to the study. All participants provided written informed consent in
compliance with a research protocol approved by the University of Manchester Research Ethics Committee.

5.4.2 Psychophysical testing

5.4.2.1 Stimuli

Stimuli were diotic musical dyads, consisting of a low ('root') note, and a high ('interval') note. Root notes were the eight notes above and including the musical note D3, all taken from the equal temperament scale: D (146.83 Hz), D# (155.56 Hz), E (164.81 Hz), F (174.61 Hz), F# (185.00 Hz), G (196.00 Hz), G# (207.65 Hz), and A (220 Hz). Each root note was used to create two types of dyad; a consonant Perfect 5\(^{th}\) and a dissonant Tritone. Dyads are named after the ratio of the fundamental frequencies (F0s) of the two notes (the size of the interval), approximately 3:2 in the case of the Perfect 5\(^{th}\) and \(\sqrt{2}:1\) in the case of the Tritone. The simple frequency ratio of the Perfect 5\(^{th}\) results in the combined spectrum of the two notes being harmonic, whilst the complex frequency ratio of the Tritone results in an inharmonic spectrum (Fig. 5.1A-B). Dyads were low-pass filtered at 2500 Hz. The harmonics of each note were equal amplitude so that the overall level of each dyad was 80 dB SPL. Each dyad was 2 s in duration including 10 ms raised-cosine onset and offset ramps.

For completeness, masking noise was used in the behavioural as well as the FFR section of the study. Each dyad was presented in three conditions: with no masking noise (No Noise); with Gaussian noise with a 32 dB spectrum level (Level 1), band-pass filtered between 2600 and 7000 Hz; and with the same masking noise with a 42 dB spectrum level (Level 2). The
The spectrum levels of the masking noise were chosen to be 15 and 5 dB respectively below the spectrum level of the dyads. When presented, masking noise had the same duration and ramps as the dyads.
Each stimulus was preceded by Gaussian noise with the same duration and ramps as the stimulus. This noise was low-pass filtered at 2500 Hz so that it had the same frequency range as the dyads. A 500 ms silence separated the noise and the stimulus. The purpose of the noise was to break up the sequence of the dyads to prevent any melodic structure from influencing responses (Bones et al., 2014; McDermott et al., 2010).

All stimuli were generated digitally in Matlab at a sampling rate of 24,414 Hz with a 32 bit resolution. Stimuli were delivered via a 24-bit E-MU 0202 USB audio device and Sennheiser HD 650 supra-aural headphones.

5.4.2.2 Procedure

Behavioural pleasantness ratings were measured by following the methodology of McDermott et al. (2010). Participants rated each dyad for pleasantness from -3 (very unpleasant) to +3 (very pleasant). Stimuli were presented in a random sequence. Participants were seated in a sound attenuating booth and made responses via a PC keyboard and a monitor placed outside the booth, visible through a window. Each run consisted of 48 presentations (eight root notes x two interval notes x three masking noise conditions). Each participant performed two runs consecutively on the same day, preceded by one practice trial of one of each interval in each masking noise condition.

5.4.2.3 Analysis

Ratings for each interval (Perfect 5th and Tritone) in each masking noise condition were averaged across runs and root notes, so that each participant’s rating for each interval in each masking noise condition were the
mean of 32 responses. Individual consonance preference for each masking noise condition was calculated using a version of the routine described by McDermott et al. (2010): averaged ratings were first z-scored for each individual in order to remove the influence of individual differences in the use of the scale. Consonance preference in each condition was then calculated by subtracting each individual’s z-scored rating of the Tritone (a dissonant interval) from their z-scored rating of the Perfect 5\textsuperscript{th} (a consonant interval).

5.4.3 Frequency-following response

5.4.3.1 Stimuli and recording protocol

Stimuli were a subset of those used in psychophysical testing, without the intervening noise: diotic Perfect 5\textsuperscript{th} and Tritone dyads with root notes set at A (220 Hz), with the same three masking noise conditions used in the psychophysical testing. There were therefore six stimuli in total (two dyads x three masking noise conditions). Stimuli were 120 ms in duration, including 10 ms raised-cosine onset and offset ramps.

Electrodes were positioned at the high forehead hairline (active), the seventh cervical vertebra (reference), and at Fpz (ground; Bidelman & Krishnan, 2009; Bones et al., 2014; Gockel, Carlyon, Mehta, & Plack, 2011; Gockel, Farooq, Muhammed, Plack, & Carlyon, 2012; Krishnan & Plack, 2011; Krishnan et al., 2005). Impedances were maintained below 3 kΩ. Participants were seated in a comfortable chair within a sound attenuating booth and told to remain as still as possible, with their arms, legs and neck straight, and told that they could sleep. Stimuli were delivered via a TDT RP2.1 Enhanced Real Time Processor, HB7 Headphone Driver, and Etymotic ER3A insert headphones. Prior to each recording, a test recording
was made with the headphone tubing clamped so that the participant could not hear the stimulus. Test recordings were then checked to ensure electromagnetic stimulus artefacts transmitted by the transducers were below the level of the background electroencephalography (EEG).

FFRs to each dyad were collected separately. Presentations consisted of two dyads separated by 150 ms (Fig. 5.1C-D). Presentations were made at a rate of 1.82/s. The onset polarity of the second of each pair of dyads was inverted 180 degrees relative to the first presentation (Fig. 5.1E-H). The acquisition window lasted for 447 ms from the onset of the first presentation. 4000 responses to each dyad were collected from each participant in total using TDT BioSig software. Responses were band-pass filtered online between 50 Hz and 3000 Hz. Data were compiled online as 200 sub-averages of 10 responses to each stimulus polarity. Any sub-average in which the peak amplitude exceeded +/-15 µV was assumed to be an artefact and rejected off-line. The artefact rejection threshold was chosen to achieve the lowest possible RMS (0.19 µV) in the portion of the response between stimulus polarities.

5.4.3.2 Analysis

A neural harmonic salience measure was used to determine how well the harmonicity of each dyad was represented in phase locking in the brainstem (Bones et al., 2014). The best fitting harmonic series to the power spectrum of each dyad was determined by finding the harmonic series for which the ratio of power inside 15 Hz wide bins placed at integer multiples of the F0 to power outside these bins was highest. Only F0s above 30 Hz were considered, since this was assumed to be the lower limit of pitch
The F0s of the best fitting harmonic series to the Perfect 5th and the Tritone were 55.00 and 44.25 Hz respectively. The salience of these harmonic series in the power spectrum of the FFR to each stimulus was then calculated in the same way, i.e. the ratio of power inside the bins placed at integer multiples of the F0 to power outside the bins. Because the FFR was recorded to each stimulus in alternating polarity, the spectra upon which analyses were performed were the mean spectra of the response to each polarity. This is referred to hereafter as FFR_{RAW}.

In order to investigate differences between the results of the current study and results reported previously (Bones et al., 2014), the same method for enhancing or suppressing phase locking to the envelope or TFS of the cochlear response was used: the mean FFR to the original polarity was added to the mean FFR to the inverted polarity to enhance the envelope response and suppress the TFS response (FFR_{ADD}); and the mean FFR to the original polarity was subtracted from the mean FFR to the inverted polarity to suppress the envelope response and enhance the TFS response (FFR_{SUB}; Goblick & Pfeiffer, 1969).

In an abbreviated version of the NCI measure described elsewhere (Bones et al., 2014) the strength of the representation of harmonicity in the FFR to the Perfect 5th relative to the Tritone was calculated as the neural harmonic salience of the former minus the neural harmonic salience of the latter. The NCI measure is a physiological analogue to the consonance preference measure (see section 5.4.2.3).
5.4.4 Auditory model

The FFR data were compared to the output of a simulation of the auditory periphery using the Development System for Auditory Modelling library (DSAM; http://dsam.org.uk/). The same Perfect 5th stimulus files that were used to measure the FFR were used as inputs to the model.

The first stage of the simulation was an outer/middle ear filter model. This consisted of two parallel cascades of first-order band-pass filters. The first cascade consisted of two filters with 3 dB down points at 4000 and 25000 Hz, the second consisted of three filters with 3 dB down points at 700 and 30000 Hz. The output was then converted to stapes velocity using the scalar $1.4 \times 10^{-10}$ (Sumner, O'Mard, Lopez-Poveda, & Meddis, 2003).

The BM was modelled using a dual resonance nonlinear (DRNL) filter (Meddis, O'Mard, & Lopez-Poveda, 2001). In order to determine whether the FFR was best accounted for by the apical region of the cochlea tuned to the dyad or by a basal region, two models were used: a low-frequency model (LF model) only included the output of 10 filter channels between 100 and 2500 Hz, spaced using the Greenwood CF spacing function; and a high-frequency model (HF model) only included the output of 10 cochlear filter channels between 2500 and 8000 Hz, spaced the same way.

The multi-channel output of the BM model was then used as the input to an inner hair cell receptor potential (IHC RP) simulation (Shamma, Chadwick, Wilbur, Morrish, & Rinzel, 1986) using parameters provided by Sumner, Lopez-Poveda, O'Mard, and Meddis (2002). The output was then averaged across frequency channels and used to simulate the FFR.
5.5. Results

5.5.1 The effect of high-pass masking noise on pleasantness ratings of low-frequency dyads

Pleasantness ratings for each interval in each masking noise condition are plotted in Fig. 5.2A. For both intervals variance in rating was greater with no masking noise (Tritone, SD = 1.31; Perfect 5th, SD = 1.23) than with Level 1 (Tritone, SD = 0.61; Perfect 5th, SD = 0.80) or Level 2 masking noise (Tritone, SD = 0.53; Perfect 5th, SD = 0.82). Interval had a significant effect on pleasantness rating ($F_{(1, 12)} = 15.96, p < 0.001, \eta^2 = 0.12$) with the Perfect 5th ($M = -0.75$) on average being rated as more pleasant than the Tritone ($M = -1.41; t_{(38)} = 5.85, p < 0.001$). Masking noise also had a significant effect ($\varepsilon = 0.57, F_{(1.14, 13.68)} = 40.48, p < 0.001, \eta^2 = 0.51$). In Bonferroni corrected t-tests ($\alpha = 0.017$) pleasantness ratings in the No Noise condition ($M = 0.13$) were significantly greater than Level 1 ($M = -1.30; t_{(25)} = 8.18, p < 0.001$) and Level 2 conditions ($M = -2.07; t_{(25)} = 8.84, p < 0.001$). Pleasantness ratings in the Level 1 ($M = -1.30$) condition were significantly greater than in the Level 2 condition ($M = -2.07; t_{(25)} = 7.05, p < 0.001$).
Figure 5.2: Pleasantness ratings and consonance preference scores. A, Pleasantness ratings as a function of noise level, grouped by interval. Dots indicate outliers (+/− 1.5 x IQR). B, Pleasantness ratings as a function of interval, grouped by noise level. C, Consonance preference as a function of noise level. Confidence intervals are 95%.
In a Greenhouse-Geisser corrected (ε) repeated-measures ANOVA the effect of masking noise significantly interacted with the effect of interval (ε = 0.79, $F_{(1.58, 18.96)} = 4.34, p = 0.04, \eta^2 = 0.01$; Fig. 5.2B). To explore the interaction further, consonance preference scores were calculated (Fig. 5.2C). In Bonferroni corrected paired t-tests ($\alpha = 0.017$) consonance preference in the No Noise condition ($M = 0.92$) was significantly greater than in the Level 2 condition ($M = 0.33; t_{(12)} = 2.48, p = 0.01$) but not the Level 1 condition ($M = 0.63; t_{(12)} = 1.40, p = 0.09$). Consonance preference scores in the Level 1 and Level 2 condition were not significantly different ($t_{(12)} = 2.07, p = 0.03$).

5.5.2 The effect of high-pass masking noise on the FFR to low-frequency dyads

Fig. 5.3 displays the average power spectra of the FFR\textsubscript{RAW} to the Perfect 5\textsuperscript{th} (Fig. 5.3A-C) and the Tritone (Fig 5.3D-F). The power spectrum of the FFR\textsubscript{RAW} to the consonant Perfect 5\textsuperscript{th} in the No Noise condition (Fig. 5.3A) contains clearly defined peaks, at frequencies both present and not present in the stimulus. As noted elsewhere, the distortion products in the spectrum serve to enhance the overall harmonicity (Bones et al., 2014; Lee, Skoe, Kraus, & Ashley, 2009). The FFR\textsubscript{RAW} to the dissonant Tritone in the same condition (Fig. 5.3D) also contains peaks in the spectrum corresponding to the frequency components present in the stimulus although they are not as clearly defined and are lower in amplitude than is the case in the FFR\textsubscript{RAW} to the Perfect 5\textsuperscript{th}.

With masking noise added to the dyads (Level 1, middle row; and Level 2, bottom row) the representation of the stimuli in the spectra of the
Figure 5.3: The average spectra of the FFR$_{\text{RAW}}$ to the Perfect 5th (A-C, left column) and the Tritone (D-F, right column), with No Noise (top row), Level 1 noise (middle row), and Level 2 noise (bottom row). Black circles indicate the frequency components present in the stimulus. Power is expressed as dB referenced to $10^{-16}$ V$^2$.

FFR$_{\text{RAW}}$ to the Perfect 5th and the Tritone are less well defined. In the case of the Perfect 5th the large peaks corresponding to frequencies in the stimulus
are noticeably reduced in amplitude, and many of the distortion products are no longer above the background EEG noise floor.

Neural harmonic salience scores are plotted in Fig. 5.4A. As for pleasantness ratings, for both intervals variance in neural harmonic salience was greater in the No Noise condition (Tritone, \(SD = 0.32\); Perfect 5\(^{th}\), \(SD = 1.24\)) than with Level 1 (Tritone, \(SD = 0.18\); Perfect 5\(^{th}\), \(SD = 0.37\)) and Level 2 masking noise (Tritone, \(SD = 0.19\); Perfect 5\(^{th}\), \(SD = 0.27\)). Musical interval had a significant effect on neural harmonic salience \((F_{(1, 9)} = 14.90, p < = 0.03, \eta^2 = 0.191)\) with the Perfect 5\(^{th}\) having greater neural harmonic salience on average \((M = 1.13)\) than the Tritone \((M = 0.61, t_{(29)} = 3.66, p < 0.001)\). Adding high-pass masking noise also had a significant effect \((\epsilon = 0.53, F_{(1,06, 9.54)} = 9.52, p = 0.01, \eta^2 = 0.234)\). In Bonferroni corrected paired \(t\)-tests \((\alpha = 0.017)\) neural harmonic salience in the No Noise condition \((M = 1.29)\) was greater than in the Level 1 \((M = 0.67; t_{(9)} = 3.05, p = 0.007)\) and Level 2 condition \((M = 0.65; t_{(19)} = 2.96, p = 0.008)\). Harmonic salience was not significantly different between the Level 1 and Level 2 condition \((t_{(19)} = 0.28, p = 0.79)\). Masking noise and interval interacted \((\epsilon = 0.54, F_{(1,08, 9.72)} = 13.26, p = 0.004, \eta^2 = 0.180; Fig. 5.4B)\). NCI is a measure of how much more salient the harmonicity of consonant dyads are than dissonant dyads. The large difference in salience between the Perfect 5\(^{th}\) and the Tritone in the No Noise condition but not the masking noise conditions seen in Fig 5.4B is reflected in the NCI measures \((Fig. 5.4C)\). In Bonferroni corrected paired \(t\)-tests \((\alpha = 0.017)\) NCI scores in the No Noise condition \((M = 1.24)\) were significantly greater than in the Level 1 \((M = 0.12; t_{(9)} = 3.96, p = 0.003)\) and Level 2 conditions \((M = 0.21, t_{(9)} = 2.96, p = 0.008)\). NCI in the Level 1
Figure 5.4: Neural harmonic salience and NCI scores. A, Neural harmonic salience as a function of noise level, grouped by interval. Dots indicate outliers (± 1.5 x IQR). B, Neural harmonic salience as a function of interval, grouped by noise level. C, NCI as a function of noise level. Confidence intervals are 95%.
condition was not significantly different to the Level 2 condition \(t_{(9)} = -1.24, \ p = 0.88\).

5.5.3 The relation between NCI and individual preference for consonance for non-musicians

Individual consonance preference of young normal hearing non-musicians is plotted as a function of NCI for each masking noise condition in Fig. 5.5. In a Spearman’s Rho correlation NCI predicted consonance preference in the No Noise condition, with individuals with larger NCI scores also having a greater preference for consonance over dissonance \(r_{s(8)} = 0.61, \ p = 0.03\). The reduction in variance of both neural harmonic salience and pleasantness ratings when masking noise was added to the stimulus can be seen in the clustering of data points in the Level 1 and Level 2 condition in the bottom left of the plot. Correlations between consonance preference and NCI in the Level 1 \(r_{s(8)} = 0.20, \ p = 0.29\) and Level 2 \(r_{s(8)} = -0.12, \ p = 0.63\) conditions were not significant.

Bones et al. (2014) found that NCI only predicted consonance preference when phase locking to the cochlear envelope was suppressed. In the current study however NCI predicted consonance preference without this component suppressed. To explore this further the spectra of the FFRs to the Perfect 5\textsuperscript{th} in each polarity were added and subtracted. Fig. 5.6A-C displays the spectrum of the FFR\textsubscript{RAW} (A), FFR\textsubscript{ADD} (the response to the envelope enhanced, the response to the TFS suppressed; B), and FFR\textsubscript{SUB} (the response to envelope suppressed, the response to the TFS enhanced; C) to the Perfect 5\textsuperscript{th}. Each FFR type to the Perfect 5\textsuperscript{th} from Bones et al. is displayed in Fig. 5.6D-F for comparison. Bones et al. used dyads with a
130.81 Hz root note F0. They found that when consonant dyads were presented diotically they produced large distortion products corresponding to the difference tone of the F0 of the two notes (F₂ – F₁). For the Perfect 5th the interval note F0 was 196 Hz. The large difference tone of approximately 65 Hz corresponded to the F0 of the harmonic series of the FFR, serving to enhance the overall harmonicity of the spectrum (Fig. 5.6D-E). Suppressing the response to the envelope also suppressed the difference tone (Fig. 5.6F), suggesting that phase locking at this frequency was to the cochlear envelope resulting from interaction of the two notes (Bones et al., 2014; Gockel et al., 2012). In the current study, despite the FFR to the No Noise Perfect 5th stimulus containing distortion products at most harmonic frequencies up to 2500 Hz, the difference tone corresponding to the F0 (110 Hz) is lower in magnitude (Fig. 5.6A). When the envelope component is
Figure 5.6: Average spectra of the FFR_{RAW} (top), FFR_{ADD} (middle), and FFR_{SUB} (top) to the Perfect 5th from the current study (A-C) and Bones et al. (2014; D-F). Power is expressed as dB referenced to $10^{-16} \text{ V}^2$. Black circles indicate frequencies present in the stimulus, arrows indicate the implied F0 of the harmonic series of the spectrum of the FFR.
enhanced (Fig. 5.6B) the difference tone is of a greater magnitude. When masking noise was added to the stimulus (Fig. 5.3B-C) the difference tone had a greater magnitude than in the No Noise condition, and was greater in magnitude than the other components of the spectrum.

Correlations between NCI for each FFR type and consonance preference are plotted in Fig. 5.7. Larger NCI scores were associated with greater consonance preference for all FFR types (Raw, $r_{s(8)} = 0.61, p = 0.03$; Add, $r_{s(8)} = 0.50, p = 0.05$; Sub, $r_{s(8)} = 0.61, p = 0.03$).

### 5.5.4 Comparison of the effects of high-pass masking noise on the FFR and a model of the inner hair cell receptor-potential

To explore whether masking noise reduced harmonic salience of the FFR to the Perfect 5th by reducing power in the spectrum at harmonic frequencies or by increasing power in the spectrum at non-harmonic frequencies (neural noise), a two-way repeated measures ANOVA of spectral power was performed. One factor of the ANOVA was frequency region with levels ‘harmonic peaks’ (i.e. the power inside the bins used to calculate harmonic salience) and ‘neural noise’ (i.e. the power of the spectrum between 0-2500 Hz outside of the bins used to calculate harmonic salience). Masking noise level was the other factor. Masking noise level was marginally non-significant after correcting for non-sphericity ($\epsilon = 0.52, F_{[1.04, 9.36]} = 4.11, p = 0.07, \eta^2 = 0.010$). Frequency region was a significant effect ($F_{[1, 19]} = 10.30, p = 0.01, \eta^2 = 0.213$). Moreover, frequency region interacted with masking noise level ($\epsilon = 0.51, F_{[1,02, 9.18]} = 5.41, p = 0.04, \eta^2 = 0.011$). The interaction between frequency and masking noise level can be seen in Fig. 5.8. In Bonferonni corrected $t$-tests ($\alpha = 0.017$) the difference in power of harmonic peaks in the
No Noise condition ($M = 2.51 \times 10^{-15} \text{ V}^2$) and both the Level 1 ($M = 0.89 \times 10^{-15} \text{ V}^2$; $t(9) = 2.19, p = 0.06$) and Level 2 condition ($M = 0.88 \times 10^{-15} \text{ V}^2$; $t(9) = 2.18, p = 0.06$) was not significant. Differences between neural noise in the No Noise ($M = 2.39 \times 10^{-16} \text{ V}^2$) condition and both the Level 1 ($M = 2.93 \times 10^{-16} \text{ V}^2$; $t(9) = -1.29, p = 0.23$) and Level 2 ($M = 2.93 \times 10^{-16} \text{ V}^2$; $t(9) = 0.95, p = 0.37$) condition were also not significant.

The output of the IHC RP model to the Perfect 5\textsuperscript{th} stimulus is displayed in Fig. 5.9. First consider the No Noise condition (Fig. 5.9A-D). The response of the LF model (containing frequency channels corresponding to the frequency content of the stimulus) contains a DC component (Fig. 5.9A), with frequency components corresponding to the stimulus and also distortion products (Fig. 5.9B). The output of the HF model, representing the response of IHCs with CFs higher than the frequency content of the stimulus is lower in

Figure 5.7: Consonance preference as a function of NCI calculated from FFR\textsubscript{RAW}, FFR\textsubscript{ADD}, and FFR\textsubscript{SUB}.
amplitude, with a smaller DC component (Fig. 5.9C). The HF model of the IHC RP also contains distortion products which serve to reinforce the harmonicity of the response (Fig. 5.9D), although they are lower in amplitude than those in the output of the LF model.

When Level 2 masking noise is added to the stimulus the response of the LF model is virtually identical to the No Noise condition (Fig. 5.9E-F). However, the output of the HF model to Level 2 masking noise (Fig. 5.9G) is markedly different to the output of the HF model to No Noise (Fig. 5.9C). The waveform has a DC component that is larger than the AC component, the harmonic frequency components are reduced in amplitude and the background noise floor is increased in amplitude (Fig. 5.9H).

The power of the harmonic peaks and neural noise floor in the output of each model in each condition is summarized in Fig. 5.10. In the LF model,
the mean power of harmonic peaks remains virtually constant between the
No Noise \((M = 3.49 \times 10^{-8} \text{ V}^2)\) and the Level 2 \((M = 3.46 \times 10^{-8} \text{ V}^2)\) conditions,
whereas, similar to the FFR data, the mean power of the harmonic peaks in the output of the HF model reduces from No Noise \((M = 0.99 \times 10^{-8} \text{ V}^2)\) to Level 2 masking noise \((M = 0.47 \times 10^{-8} \text{ V}^2)\). The power of the neural noise floor also remains virtually constant between No Noise \((M = 5.58 \times 10^{-10} \text{ V}^2)\) and Level 2 masking noise \((M = 5.68 \times 10^{-10} \text{ V}^2)\) in the LF model. However, the background noise floor of the HF model increases from No Noise \((M = 1.70 \times 10^{-10} \text{ V}^2)\) to Level 2 \((M = 4.75 \times 10^{-10} \text{ V}^2)\) masking noise.

5.6 Discussion

5.6.1 Neural harmonic salience of the FFR and musical consonance perception of non-musicians

The results of this study provide further evidence that the salience of subcortical temporal coding of the harmonicity of consonant dyads relative to dissonant dyads as represented by the FFR predicts individual preference for consonance over dissonance. One of the aims of the current study was to determine whether the relation between NCI (a measure of the salience of the harmonicity of the neural response to consonant relative to dissonant dyads) and consonance preference reported elsewhere for young normal hearing listeners with a range of musical experience (Bones et al., 2014) also occurs for young normal hearing listeners with no musical experience. The current study found a significant correlation between NCI and consonance preference for young normal hearing listeners with no musical experience. To the authors’ knowledge this is the first time that this has been demonstrated.
As has been shown previously (Bones et al., 2014; Lee et al., 2009), the FFR to a consonant dyad in the current study contained multiple distortion products which enhanced the overall harmonicity. Lee et al. (2009) found that distortion products had greater magnitude in the FFR of musicians compared to non-musicians. Lerud, Almonte, Kim, and Large (2014) suggest that this might be due to greater synaptic efficiency in musicians, leading to more nonlinear processing. The data presented here demonstrate that the nonlinearity of the auditory system of non-musicians also produces large distortion products which serve to enhance the harmonicity of the FFR to consonant dyads. Surprisingly however, the large difference tones in the FFR to consonant dyads reported by Bones et al. (2014) were not found in the No Noise condition of the current study. One possibility is that this is a consequence of the different F0s: the F0s of the root and interval note of the Perfect 5th dyad used in the current study and by Bones et al. were 220 and
440, and 130.81 and 196 Hz respectively. The smaller sample-size of the current study compared to the sample size of Bones et al. \(N = 10, 19\) may also be a contributing factor.

Bones et al. found that the FFR to consonant dyads presented diotically had greater harmonic salience than when presented dichotically, due to the difference tone produced by monaural interactions in the diotic condition. This resulted in larger NCI scores in the diotic condition, and coincided with higher pleasantness ratings for consonant dyads and therefore greater consonance preference. NCI only predicted consonance preference when the envelope and difference tone were suppressed (FFR$_{\text{SUB}}$). In the current study NCI calculated from FFR$_{\text{RAW}}$ was predictive of consonance preference. Whether suppression of the envelope component was not necessary due to the absence of the large difference tone previously reported will need further investigation.

The strong interaction between the effects of masking noise and interval on harmonic salience were not found for pleasantness ratings. Ratings for both the Tritone and the Perfect 5$^{\text{th}}$ declined incrementally with Level 1 and Level 2 masking noise, with both dyads being rated as being more pleasant in the Level 1 condition than in the Level 2 condition. It is likely that the reduction in pleasantness ratings when masking noise was added to the dyads is representative of the stimuli becoming increasingly unpleasant in a way unrelated to the perception of consonance per se.

5.6.2 The FFR to low-frequency musical dyads is likely to be partly generated in the basal region of the cochlea
The results of the current study suggest that the FFR to the low-frequency Perfect 5th dyad was partly generated by a region of the cochlea tuned to frequencies above the dyad. The addition of high-pass masking noise reduced the neural harmonic salience of the FFR. This effect was driven by a reduction in the amplitude of harmonic components of the FFR, rather than an overall increase in the background EEG noise floor, implying a nonlinear process.

With no masking noise added to the stimulus the spectra of both the LF and HF model output (Fig. 5.9B, F) are similar to the FFR spectrum in the same condition (Fig. 5.3A); both spectra contain frequencies of the stimulus and distortion products, indicating that processing in both the low- and the high-frequency pathway up to the output of the IHC is sufficiently nonlinear to generate additional harmonic frequencies. This is consistent with a saturating IHC response (Dallos, 1986; Patuzzi & Sellick, 1983), and demonstrates that even though the BM response to the low-frequency dyad at a place tuned to a higher-frequency is likely to be linear, the IHC response is sufficiently nonlinear to produce distortion products. It should be noted that distortion products in the FFR may also represent further nonlinear processing beyond the IHC (e.g. see Lins, Picton, Picton, Champagne, & Durieux-Smith, 1995, p.3059, Fig. 5.9).

Adding masking noise to the stimulus in the LF model has no effect because the frequency range of the masking noise is above the frequency range of the channels included in the model; in the HF model the masking noise has a clear effect on the IHC RP (Fig 5.6G-H). The harmonic peaks of the spectrum are reduced from a mean amplitude of 0.99 to 0.47 x 10^-8 V^2.
The ratio of this reduction in amplitude (0.47) was not significantly different to the mean ratio of the reduction of the harmonic peaks in the FFR spectra (0.55; \( r = -0.28, V = 18, p = 0.38 \)), consistent with the saturating response of IHCs in the basal portion of the cochlea accounting for the reduction in the amplitude of harmonic frequency components in the FFR to low-frequency dyads when high-pass masking noise is added.

The ratio of the increase in noise floor in the output of the HF model from No Noise to Level 2 masking noise (2.79) is greater than the mean ratio of the increase in background EEG noise of the FFR in the same condition (1.66). Although this difference is not statistically significant (\( r = -0.54, V = 45, p = 0.08 \)), it should be noted that the FFR represents the integration of the evoked response with spontaneous neural activity and the electrical potentials created by muscular activity, and therefore has a background noise floor even in the No Noise condition. It is possible therefore that the FFR is not sensitive to the increase in the noise floor of the IHC RP in response to Level 2 masking noise.

### 5.6.3 Conclusions

1) The salience of the harmonicity of consonant dyads relative to dissonant dyads in the FFR is predictive of individual consonance preference in young normal hearing non-musicians.

2) The harmonic salience of the FFR to low-frequency dyads is reduced by high-pass masking noise above the frequency range of the dyads.

3) The reduction in harmonic salience due to the addition of the noise is the result of a reduction in amplitude of the harmonic components of the FFR.
This can be accounted for by saturating high-frequency IHC RPs, suggesting that the FFR to low-frequency dyads is at least partly generated in the basal region of the cochlea.

5.7 References


Chapter 6. Losing the music: Aging affects the perception and subcortical neural representation of musical harmony

Oliver Bones

Christopher J. Plack

School of Psychological Sciences, The University of Manchester,
Manchester Academic Health Science Centre, Manchester, UK, M13 9PL

This chapter was prepared for publication as a paper. It is under review at the Journal of Neuroscience.

Publication 4
6.1 Abstract

When two musical notes with simple frequency ratios are played simultaneously the resulting musical chord is pleasing and evokes a sense of resolution or consonance. Complex frequency ratios, on the other hand, evoke feelings of tension or dissonance. Consonance and dissonance form the basis of harmony, a central component of Western music. In earlier work we provided evidence that consonance perception is based on neural temporal coding in the brainstem (Bones et al., 2014). Here we show, for listeners with clinically normal hearing, that aging is associated with a decline in both the perceptual distinction, and the distinctiveness of the neural representations, of different categories of two-note chords. Compared to younger listeners, older listeners rated consonant chords as less pleasant and dissonant chords as more pleasant, consistent with the neural representations of these chords as measured using the electrophysiological ‘frequency-following response.’ To control for a possible effect of age on general affect, participants also rated different categories of affective voice for pleasantness. Although older participants rated affective voices differently to younger participants, the effect of age interacted with type of stimulus (chord or voice) and the association between age and perception of musical consonance remained when controlling for the effect of variation in affect. The results of the study suggest that different mechanisms are responsible for the perceived pleasantness of musical chords and affective voices, and that age-related differences in consonance perception are related to differences in neural temporal coding.
6.2 Introduction

A musical note is an example of a complex tone. A complex tone consists of a fundamental frequency (F0) and a number of harmonics with frequencies at integer multiples of the F0. When a two-note chord (a ‘dyad’) is created using notes with simple F0 ratios the resulting perception of resolution and stability that occurs is described as ‘consonance’. The perception of consonance is likely driven by the extent to which the combined harmonics of a consonant dyad share a common F0 and therefore resemble a single complex tone (McDermott, Lehr, and Oxenham, 2010; see Fig. 6.1A-B). The harmonic structure of a dyad is temporally coded in the inferior colliculus (IC), the principle auditory midbrain nucleus, by the tendency for neurons to fire at a particular phase of the individual low numbered harmonics and the temporal envelope produced by the interaction of harmonics in the cochlea (Brugge et al., 1969; Bidelman and Krishnan, 2009; Lee et al., 2009; Bones et al., 2014). This neural temporal coding of harmonic structure predicts individual differences in the perception of consonance in young normal hearing listeners (Bones et al., 2014).

Aging has a detrimental effect on temporal processing, independently of age-related clinical hearing loss (Strouse et al., 1998; Clinard et al., 2010; Hopkins and Moore, 2011; Neher et al., 2012). This may be partly due to age-related changes in the response properties of neurons in the IC. Single units in old mice spike more frequently and code temporal fluctuations less accurately than young mice (Walton et al., 2002; Frisina and Walton, 2006), and units which respond with rapid onset in young mice respond with greater
latency (Walton et al., 1997) and with a longer recovery period (Walton et al., 1998). This age-related loss of temporal precision is consistent with a decline in neural inhibition (Caspary et al., 2005) and may be the consequence of down-regulation of the inhibitory neurotransmitter GABA (Milbrandt et al., 1994). GABA receptor agonists broaden the frequency response of IC neurons in chinchillas (Caspary et al., 2002) and have a deleterious effect on phase locking of IC neurons in bats (Koch and Grothe, 1998; Klug et al., 1998).
suggesting that normal spectro-temporal neural response properties may be dependent on the sharpening provided by GABAergic inhibition.

Age-related deficits in temporal processing may also be a consequence of deafferentation. Afferent auditory nerve spiral-ganglion cells become decoupled from the inner hair cells of the cochlea with normal aging (Makary et al., 2011; Sergeyenko et al., 2013) and following glutamate excitotoxicity resulting from noise-trauma (Puel et al., 1994). Elevation of hearing threshold following noise-trauma is usually temporary, with a time course corresponding to synaptic recovery time (Puel et al., 1998). However, animal models demonstrate that deafferentation persists even when auditory thresholds return to pre-noise levels (Kujawa and Liberman, 2009; Furman et al., 2013), suggesting that deafferentation may go undetected, and occur over the lifespan.

The present study was designed to test the hypothesis that the distinction in the representation of the harmonic structure of musical dyads by neural temporal coding of populations of neurons in the IC declines with age, and that this deficit is associated with a decline in the perception of consonance.

### 6.3 Methods

#### 6.3.1 Participants

All participants had clinically normal hearing thresholds (≤ 20 dB HL) at octave frequencies between 250 and 2000 Hz. No participants reported as having ‘perfect pitch’ (‘can you recreate a given musical note without an
external reference?)

), or as having a history of speech or language difficulties. Participants were non-musicians, having not received formal music training within the past five years. Where participants had received music training, an estimate of their total training in hours was logged and controlled.

Forty-four (27 female) participants aged 18-81 (mean = 38.0, SD = 20.5) participated in the behavioural section of the study. The distribution of age is displayed in Fig. 6.2. Group comparisons were made between 28 participants aged under 40 years (mean = 23.7 yrs, SD = 6.1 yrs) and 16 participants over 40 years (mean = 63.1 yrs, SD = 9.1 yrs). The mean hearing threshold between 250-2000 Hz in the young group was 4.66 dB HL (SD = 4.42 dB), the mean of the old group was 10.26 dB HL (SD = 3.62 dB). The mean musical experience of the young group was 434 hours (SD = 702
hours), the mean of the old group was 260 hours ($SD = 413$ hours). One participant withdrew from the old group prior to FFR testing.

6.3.2 Behavioural stimuli

Dyads were created by combining each of eight low notes with 11 high notes in the octave above, making 88 dyads in total. Each dyad is named after the ratio of the F0s of the low and the high notes (the ‘interval’; see Table 6.1). Low notes were the eight notes above D3 (see Table 6.2).

<table>
<thead>
<tr>
<th>Rank</th>
<th>FO ratio</th>
<th>Name</th>
<th>Abbreviation</th>
<th>Consonance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.498307</td>
<td>Perfect 5th</td>
<td>P5</td>
<td>Perfectly consonant</td>
</tr>
<tr>
<td>2</td>
<td>1.33484</td>
<td>Perfect 4th</td>
<td>P4</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.681793</td>
<td>Major 6th</td>
<td>M6</td>
<td>Imperfectly consonant</td>
</tr>
<tr>
<td>4</td>
<td>1.259921</td>
<td>Major 3rd</td>
<td>M3</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1.587401</td>
<td>Minor 6th</td>
<td>m6</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1.189207</td>
<td>Minor 3rd</td>
<td>m3</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1.781797</td>
<td>Minor 7th</td>
<td>m7</td>
<td>Dissonant</td>
</tr>
<tr>
<td>8</td>
<td>1.414214</td>
<td>Tritone</td>
<td>Tri</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>1.887749</td>
<td>Major 7th</td>
<td>M7</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1.122462</td>
<td>Major 2nd</td>
<td>M2</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>1.059463</td>
<td>Minor 2nd</td>
<td>m2</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.1: Musical dyads. Combining a low note with each of the 11 notes above it in the octave creates a different dyad, named after the ratio (‘interval’) of the two notes. Intervals can be categorized as ‘perfectly consonant’, ‘imperfectly consonant’, and ‘dissonant’. The hierarchical ranking of intervals contributes to a sense of musical ‘key’. Intervals are ranked here by consonance in accordance with Western music theory, and with the pleasantness ratings of the young group.

Each note consisted of equal amplitude harmonics with their levels set so that the overall level of each note was 72 dB SPL when low-pass filtered at 2200 Hz, and the overall level of each dyad was 75 dB SPL. Dyads had a 2000 ms duration plus 10 ms raised-cosine onset and offset ramps. Each dyad was preceded by Gaussian noise with the same duration and filtering. A 500 ms silence separated the noise and the dyad. The purpose of the
noise was to prevent randomly generated melodic sequences from influencing pleasantness judgments (McDermott et al., 2010; Bones et al., 2014).

In order to test for the possibility that older listeners rate dyads differently due to having a general lack of affect or an inability to perform the task rather than due to a decline in consonance perception, participants also rated affective voices for their perceived pleasantness (Cousineau et al., 2012). Voices were a subset of the Montreal Affective Voices (MAV) battery (Belin et al., 2008; vnl.psy.gla.ac.uk). The MAV battery is a collection of recordings of actors (male and female) expressing different emotions without using spoken words. Four affective voices (two male and two female) from the categories ‘fear’, ‘happiness’, ‘sadness’, and ‘surprise’ were used.

6.3.3 Behavioural procedure

Behavioural methodology reported elsewhere (McDermott et al., 2010; Cousineau et al., 2012; Bones et al., 2014) was used to record individual pleasantness ratings for dyads and affective voices. Participants rated each

<table>
<thead>
<tr>
<th>Note</th>
<th>F0 (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D3</td>
<td>146.83</td>
</tr>
<tr>
<td>D#3</td>
<td>155.56</td>
</tr>
<tr>
<td>E3</td>
<td>164.81</td>
</tr>
<tr>
<td>F3</td>
<td>174.61</td>
</tr>
<tr>
<td>F#3</td>
<td>185</td>
</tr>
<tr>
<td>G3</td>
<td>196</td>
</tr>
<tr>
<td>G#3</td>
<td>207.65</td>
</tr>
<tr>
<td>A3</td>
<td>220</td>
</tr>
</tbody>
</table>

Table 6.2: The eight low notes used to create dyads. Each low note was combined with 11 intervals in Table 1.
sound for its pleasantness from -3 to 3, where -3 indicated very unpleasant and 3 indicated very pleasant. Individual differences in the perceptual distinction of different categories of musical dyad were measured by calculating consonance preference scores. This was calculated by first z-scoring each participant’s rating for each dyad in order to account for individual use of the ratings scale, and then subtracting the average z-scored rating of the five most theoretically dissonant intervals from the average z-scored rating of five most theoretically consonant dyads (McDermott et al., 2010; Cousineau et al., 2012), determined a priori by Western musical tradition (e.g. see Rameau, 1722/1971; Mickelson & Riemann, 1977), and reflected by the ratings of the young group in the present study. The ranking of dyads by consonance is displayed in Table 6.1. A happiness preference score was calculated for comparison with consonance preference by subtracting the mean of each participant’s z-scored rating of ‘fear’ and ‘sadness’ (the two lowest rated affective voices) from their z-scored rating of ‘happiness’ (the highest rated affective voice).

Participants sat in a sound-attenuating booth and made ratings via a keyboard, with a computer monitor outside the booth, visible through a window. Prior to testing, participants performed one practice block of one of each dyad and each affective voice. Following the practice block, participants performed two runs, each consisting of the 88 dyads and two of each affective voice category presented in a random sequence. The two runs were performed consecutively on the same day. Ratings for each dyad were averaged across root notes and runs, so that each participant’s dyad
pleasantness rating was the average of 16 responses, and each affective voice pleasantness rating was the average of eight responses.

6.3.4 The frequency-following response (FFR): recording protocol and stimuli

The FFR is a scalp-recorded event-related potential representing the sustained, phase locked response of populations of neurons in the auditory brainstem to a sustained stimulus (Moushegian et al., 1973). Periodic peaks in the waveform correspond to phase locking to individual frequency components in the stimulus and to the temporal envelope of the basilar membrane (BM) response resulting from the interaction of harmonics in the cochlea (for a review see Krishnan, 2007). In the present study a vertical electrode montage was used (Krishnan et al., 2005; Bidelman and Krishnan, 2009; Gockel et al., 2011; Krishnan and Plack, 2011; Gockel et al., 2012; Bones et al., 2014): an active electrode was positioned at the high forehead hairline; a reference electrode was positioned at the seventh cervical vertebra; and a ground electrode was positioned at Fpz. Impendences were checked at regular intervals and maintained below 3 kΩ throughout.

Participants were seated in a comfortable, reclining armchair within a sound-attenuating booth. They were instructed to keep their arms, legs, and neck straight, and to keep their eyes closed and to sleep if possible. Stimuli were delivered via a TDT RP2.1 Enhanced Real Time Processor and HB7 Headphone Driver, and Etymotic ER30 insert headphones. The length of the headphone tubing made it possible to position the transducers outside of the booth, preventing stimulus artefacts from contaminating the recording.
Two dyads were used to record the FFR. A low note A3 was combined with the high notes D#3 and E3 to create a Tritone and a Perfect 5\textsuperscript{th} dyad respectively. Dyads were 120 ms in duration including 10 ms onset and offset raised-cosine ramps, and were filtered and had the same level as in behavioural testing.

The FFR to each dyad was collected separately. Presentations consisted of two repetitions of the dyad separated by 150 ms silence (Fig. 6.1C-D). As in previous studies, the onset polarity of the second dyad in each presentation was inverted by 180 degrees relative to the first dyad (e.g. Gockel et al., 2011; Bones et al., 2014). This method allows for selectively enhancing the neural response to either the TFS or the envelope of the BM response by subtracting or adding the response to the two stimulus polarities respectively (Goblick & Pfeiffer, 1969). However, in the current study this technique was not used: spectra are the average response to the two stimulus polarities.

Presentations were made at a rate of 1.82/s. Responses were recorded during an acquisition window with 447 ms duration from the presentation onset. Responses were filtered online between 50 and 3000 Hz. 4000 responses to each dyad were collected from each participant, compiled online in sub-averages of 10 acquisitions. Any sub-average in which the peak amplitude exceeded +/- 30 \(\mu\)V was assumed to contain an artefact and rejected off-line.

6.3.5 FFR analysis
To determine how well the harmonic structure of each dyad was represented by neural temporal coding, we used the procedure described in a previous article (Bones et al., 2014). First, the F0s of the best fitting harmonic series to the power spectra of the two dyads were calculated. The salience of harmonic series with F0s ranging from 30-1000 Hz were analysed by measuring the power inside 15 Hz wide bins placed at integer multiples of the F0. The fit of each harmonic series was calculated as the ratio between the sum of power inside the bins to the sum of power outside the bins. The F0s of the best fitting harmonic series were 44.25 and 55 Hz for the Tritone and the Perfect 5th respectively. Second, the representation of the harmonic series with these F0s in the power spectrum of the FFR (hereafter referred to as ‘neural harmonic salience’) were calculated in the same way: the ratio between power inside and outside 15 Hz wide bins placed at integer multiples of the F0 of the best fitting harmonic series to the dyad.

The Neural Consonance Index (Bones et al., 2014; NCI) is a measure of the harmonic salience of the FFR to consonant dyads relative to dissonant dyads, and is a neural analogue of the consonance preference measure. Each participant’s NCI score was calculated by subtracting the neural harmonic salience of the FFR to the Tritone (a dissonant dyad) from the neural harmonic salience of the Perfect 5th (a consonant dyad).

6.4 Results

6.4.1 The effect of age on consonance perception
Pleasantness ratings by the two groups for each dyad are displayed in Fig. 6.3A. The average standard-deviation of the rating for each dyad in the young group ($M = 0.93$) was not significantly different to the old group ($M = 0.97; \ t = 0.799, \ p = 0.43$). Musical dyad had a large effect on pleasantness ratings ($\epsilon = 0.32, F_{(3.2, 134.4)} = 91.11, \ p < 0.001, \ \eta^2 = 0.388$). Pleasantness ratings by the young group were consistent with data reported elsewhere (Plomp & Levelt, 1965; Tufts, Molis, & Leek, 2005; Bidelman & Krishnan, 2009; McDermott et al., 2010; Bones et al., 2014), and in agreement with
theoretical denominations (Table 6.1): the perfectly consonant dyads (Perfect 4th and Perfect 5th) were rated as being most pleasant, followed by the imperfectly consonant dyads (Minor 3rd, Major 3rd, Minor 6th, and Major 6th), followed by the dissonant dyads (Minor 2nd, Major 2nd, Tritone, Minor 7th, and Major 7th). However, age was also an effect ($F_{(1, 42)} = 5.58$, $p = 0.02$, $\eta^2 = 0.086$) which interacted with the effect of musical dyad ($\epsilon = 0.32$, $F_{(3.2, 134.4)} = 11.35$, $p < 0.001$, $\eta^2 = 0.388$) in that the relative ratings of the intervals depended on age group. For example, the older group rated the Major 3rd as more pleasant than the Perfect 4th. The difference in rating of the consonant Perfect 4th and the dissonant Tritone by the young group was much greater than in the old group. Likewise the difference in rating of the perfectly consonant Perfect 5th and the imperfectly consonant Minor 6th was greater in the young group than in the old group.

The effect of dyad category (see Table 1; $\epsilon = 0.85$, $F_{(1.7, 71.4)} = 114.13$, $p < 0.001$, $\eta^2 = 0.384$) interacted with the effect of age ($\epsilon = 0.85$, $F_{(1.7, 71.4)} = 18.97$, $p < 0.001$, $\eta^2 = 0.094$): in Bonferroni corrected ($\alpha = 0.017$) Welch two-sample $t$-tests and Wilcoxon rank sum tests old listeners rated dissonant dyads as more pleasant than did young listeners ($t_{(21.41)} = 3.47$, $p = 0.002$). Other pairwise comparisons between category were not significant with the corrected $\alpha$ (perfect consonance, $W = 269$, $p = 0.278$, $r = -0.16$; imperfect consonance, $t_{(27.5)} = 2.34$, $p = 0.027$). In Bonferroni corrected ($\alpha = 0.005$) Welch two-sample $t$-tests and Wilcoxon rank sum tests, the Minor 2nd ($t_{(20.3)} = 347$, $p = 0.002$), Tritone ($t_{(21.7)} = 4.63$, $p < 0.001$), and Minor 7th ($t_{(27.4)} = 3.08$, $p = 0.004$) were rated as significantly more pleasant by the old group. Other pairwise comparisons were not significant with the corrected $\alpha$ (Major 2nd,
Affective voice data were missing from four participants, two from the old group and two from the young group. Affective voice category ($\varepsilon = 0.62$, $F_{(1.9,70.3)} = 198.73$, $p < 0.001$, $\eta^2 = 0.666$; Fig. 6.3B) and age ($F_{(1,38)} = 4.98$, $p = 0.03$, $\eta^2 = 0.075$) had a significant effect on ratings of affective voice, but there was no interaction ($\varepsilon = 0.62$, $F_{(1.9,70.3)} = 0.98$, $p = 0.37$).

Consonance preference and happiness preference scores for each age group are displayed in Fig. 6.3C. The interaction between age and dyad category is reflected in consonance preference scores; the young group had greater consonance preference scores ($M = 1.51$) than the old group ($M = 1.20$; $t_{(20.1)} = 2.86$, $p = 0.007$). To check for reliability, consonance preference scores of the two groups calculated from the two test runs were tested for correlation. Scores for both groups significantly correlated across runs (young, $r_s(26) = 0.59$, $p < 0.001$; old, $r_s(14) = 0.39$, $p = 0.001$).

The young group had greater happiness preference scores ($M = 2.05$) than the old group ($M = 1.95$; $W = 261$, $p = 0.04$, $r = -0.32$). To test the effect of age on preference ratings further, an ANOVA of preference scores was performed with factors age and type (consonance and happiness). Age ($F_{1,38} = 17.7$, $p < 0.001$, $\eta^2 = 0.220$) and type ($F_{1,38} = 197.91$, $p < 0.001$, $\eta^2 = 0.673$) were significant effects. The two effects interacted ($F_{(1,38)} = 9.21$, $p = 0.007$).
0.004, $\eta^2 = 0.087$; Fig. 6.3D), in that the effect of age was greater for consonance preference scores than for happiness preference scores.

Consonance preference as a function of age is displayed in Fig. 6.4A. Age was significantly negatively correlated with consonance preference ($r_{st(42)} = -0.36, p = 0.02$). Although all participants were non-musicians, the estimate of the total time spent playing a musical instrument in hours was used to control for effects of musical experience (McDermott et al., 2010; Bones et al., 2014). The partial correlation controlling for musical experience was also significant ($r_{st(41)} = -0.44, p = 0.003$).
Happiness preference as a function of age is displayed in Fig. 6.4B. Age was also negatively correlated with happiness preference ($r_{s(38)} = -0.34$, $p = 0.035$). To test whether the relation between age and consonance preference was driven by a decline in general affect, consonance preference
was partially correlated with age, controlling for happiness preference. This correlation was also significant ($r_{s(37)} = -0.49$, $p = 0.002$).

### 6.4.2 The effect of age on neural harmonic salience

Fig. 6.5 displays the average FFR waveforms (Fig. 6.5A-B) and spectra (Fig. 6.5C-D) to the Tritone and the Perfect 5th. Note that the regularity of the Perfect 5th stimulus waveform (Fig. 6.1D) is reflected in the FFR. The regular periodicity of the waveform is also reflected in the regularly spaced harmonics in the spectrum which form a harmonic series (Fig. 6.5D). As reported previously (Lee et al., 2009; Bones et al., 2014), the spectrum of the FFR to the Perfect 5th contains frequency components that are not present in the stimulus but which serve to enhance the harmonicity of the response, including the implied 110 Hz F0 of the harmonic series. These distortion products are a consequence of the regularly spaced frequency components of the stimulus and the nonlinearities which exist at all stages of the auditory pathway (Lins et al., 1995; Krishnan, 1999; Lee et al., 2009; Gockel et al., 2012; Smalt et al., 2012; Bones et al., 2014).

For both dyads the FFR of the young group is greater in amplitude than that of the old group. The frequency content of the FFR of each dyad is the same for both groups, but the peaks in the spectra in the young group are greater in amplitude, including the peaks corresponding to distortion products in the case of the Perfect 5th.

The neural harmonic salience of the FFR is plotted in Fig. 6.6A. Musical dyad ($F_{(1, 41)} = 64.24$, $p < 0.001$, $\eta^2 = 0.282$) was a significant effect, with the neural harmonic salience of the FFR to the Perfect 5th being greater
Figure 6.6: Neural harmonic salience of the FFR to each dyad (A) and NCI scores (B) for both age groups. Error bars indicate 95% confidence intervals (A). Age in years predicted a decline in NCI (C), and NCI predicted greater consonance preference (D).
than that to the Tritone. Age was also a significant effect \((F_{(1, 41)} = 10.73, \rho = 0.002, \eta^2 = 0.164)\). The two effects interacted \((F_{(1, 41)} = 11.81, \rho = 0.001, \eta^2 = 0.067)\). This can be understood by considering that the lower amplitude of the FFR in the old group corresponded to a relatively greater reduction in harmonic salience of the FFR to the Perfect 5th than the Tritone, due to the more harmonic structure of the spectrum. Note, however, that unlike the behavioural ratings of the Perfect 5th and the Tritone, the effect of age and harmonic salience was not a cross-over interaction: were there a simple linear relation between harmonic salience and behavioural ratings it might be predicted from the FFR data that the old group would rate both the Tritone and the Perfect 5th as less pleasant than the young group. The difference in relative harmonic salience of the two dyads between groups meant that the young group had a significantly greater NCI score \((M = 1.31)\) than the old group \((M = 0.41; t_{(41.0)} = 4.13, \rho < 0.001; \text{Fig. 6B})\).

NCI as a function of age is plotted in Fig. 6.6C. Age was significantly negatively correlated with NCI \((r_{(41)} = -0.46, \rho = 0.002)\). Partial correlation controlling for musical experience was also significant \((r_{(40)} = -0.45, \rho = 0.003)\). Fig. 6.6D displays consonance preference as a function of NCI for the whole sample. NCI significantly correlated with consonance preference \((r_{(41)} = 0.32, \rho = 0.036)\), and when controlling for musical experience \((r_{(40)} = 0.34, \rho = 0.029)\).

### 6.5 Discussion

#### 6.5.1 Age is associated with a decline in the distinction between categories of musical dyad
In the present study 44 participants rated musical dyads and affective voices for their pleasantness. The purpose of the study was to test the hypothesis that age would be associated with a decline in the perceptual distinction between consonance and dissonant musical dyads, due to a decline in temporal coding. An affective voice rating task was used to control for differences between the two age groups in use of the scale, aptitude for the task, and general affect. Both age groups made similar use of the scale and did not demonstrate differences in aptitude for the task.

Comparison between the young and the old group revealed a difference in the perceived pleasantness of musical dyads. The effect of group interacted with the effect that different combinations of notes had on the perceived pleasantness of musical dyads; in general, the old group distinguished between categories of dyad less than did the young group, and rated dissonant combinations as more pleasant. This difference in relative rating of different categories of musical dyad can be expressed as a difference in consonance preference. On average the young group had greater consonance preference than the old group, and age in years predicted a decline in consonance preference. The old age group also rated the positive affective voice (‘happiness’) as slightly less pleasant than did the young group, relative to the other categories of voice. However, the effect of age on preference interacted with stimulus type; the effect of age was greater for consonance preference than happiness preference. Moreover, the correlation between age and consonance preference remained significant when the effect of happiness preference was controlled – indeed the effect size increased. The results therefore suggest that different
mechanisms were responsible for the perceived pleasantness of musical dyads and affective voices. They suggest further that the consonance preference results are not due to general flattening of affect with age.

The distinction between perfect consonance, imperfect consonance, and dissonance is central to Western music (Rameau, 1722/1971), and it is the hierarchical arrangement of notes by their ratio to a ‘tonic’ that determines musical ‘key’ and listeners’ expectations to hear a particular note in a melodic context (Krumhansl and Kessler, 1982). The resolution or non-resolution of tension evoked by sense of key contributes to the emotional response evoked by music. Engagement with music in old age provides significant benefit to cognitive, social, and emotional well-being (for a review see Creech et al., 2013), and different aspects of music-use are associated with different dimensions of well-being (Laukka, 2007). Further work is needed to determine whether a reduction in consonance preference with age is associated with a decline in listening to and engaging with music, and what the repercussions of this are for quality of life.

6.5.2 Age is associated with weaker representation of harmonic structure by neural temporal coding

Tufts et al. (2005) reported that four listeners with mild to moderate hearing loss (mean age = 69) rated Perfect 4th and Perfect 5th dyads as being less consonant relative to other dyads compared to four normal hearing listeners (mean age = 50). The loss of outer hair cell function responsible for elevated auditory thresholds also results in a decline in the tuning of the cochlear response (Ruggero & Rich, 1991; Ruggero et al., 1997) necessary for resolving individual harmonics; Tufts et al. (2005) therefore interpreted the
results of the study as being a consequence of impaired cochlear frequency selectivity. Bidelman and Heinz (2011) modelled the response of the auditory nerve (AN; Zilany et al., 2009) to musical dyads in normal hearing and mild to moderately hearing impaired listeners using the audiometric data reported by Tufts et al. (2005). Bidelman and Heinz (2011) used a pooled autocorrelation function calculated from post-stimulus time histograms of AN spikes to determine the salience of temporal information relevant to each dyad. The contrast between consonant and dissonant dyads in the output of the normal hearing model was greater than in that of the hearing impaired model, suggesting that the consequences of hearing impairment for the perception of musical consonance can be accounted for by temporal coding at the level of the AN.

Similar behavioural results to those reported by Tufts et al. (2005) were found in the present study, in which hearing thresholds were controlled: all participants had hearing thresholds ≤ 20 dB HL at 250, 500, 1000 and 2000 Hz, and therefore had clinically normal hearing at the frequency range of the dyads used, but not above this frequency. At high stimulus levels the FFR to musical dyads is likely to be generated by regions of the cochlea tuned to frequencies above the stimulus (Gardi & Merzenich, 1979; Dau, 2003; Bones & Plack, 2014). However, since the response of the cochlea at places tuned above the frequency of the stimulus is linear and unaffected by outer hair cell gain (Ruggero et al., 1997) it is unlikely that the outer hair cell loss likely responsible for elevated hearing thresholds at these frequencies (Moore, 2007) in older participants would significantly affect generation of the FFR. However, since hearing loss at high frequencies is highly correlated
with age, partial correlation of consonance preference and age controlling for
differences in hearing threshold at high frequencies was not meaningful: as
such an effect of hearing loss cannot be fully ruled out.

We argue however that the results of the present study are evidence
for an age-related decline in neural temporal coding. There is a growing body
of evidence demonstrating age-related deficits in temporal processing
independently of hearing loss (Strouse et al., 1998; Clinard et al., 2010;
Hopkins and Moore, 2011; Anderson et al., 2012; Neher et al., 2012;
Ruggles et al., 2012; Marmel et al., 2013). Clinard et al. (2010) found that
age predicted pitch discrimination and magnitude of the FFR at the stimulus
frequency, but did not find a significant correlation between the two
measures. Marmel et al. (2013) also found that age predicted both pitch
discrimination and the representation of the stimulus in the FFR, and that the
FFR predicted behavioural performance. Importantly, the two measures
remained significantly correlated when controlling for the effect of hearing
loss.

It has been demonstrated previously that the representation of the
harmonic structure of musical dyads in the neural temporal coding
represented by the FFR is predictive of their pleasantness, and that
individual differences in the distinction between categories of dyad
correspond to individual differences in the representation of the harmonic
structure of dyads by neural coding (Bones et al., 2014). The present study
demonstrates that age is associated with a decline in the representation of
the harmonic structure of dyads in neural temporal coding, and that this may
correspond to a decline in the perceptual distinction of categories of musical dyads.

6.6 References


Chapter 7. Discussion and Conclusions

The purpose of the studies described within this thesis was to explore the relation between neural temporal coding at the level of the brainstem and the perception of musical consonance. Specifically, the work was designed to determine whether the temporal coding of harmonicity is predictive of the perception of consonance. To this end, Chapter 3 describes a method for determining how well the best fitting harmonic series to a musical dyad is represented in the spectrum of the FFR, and for calculating a neural consonance index (NCI). NCI is a measure of the harmonicity of the FFR to consonant dyads relative to dissonant dyads, and is intended to be a physiological analogue of the consonance preference measure described by McDermott, Lehr, and Oxenham (2010).

The study in Chapter 4 was designed to test three hypotheses: i) that this measure is sensitive to individual differences in consonance preference in young normal hearing listeners; ii) that when dyads are presented diotically, monaural interactions produce distortion products which, in the case of consonant dyads, enhance the harmonicity of the FFR; and iii) that this enhanced harmonicity results in consonant dyads being perceived as more pleasant when presented diotically than when presented dichotically.

The results demonstrate that NCI scores are sensitive to individual differences in preference for consonance. That this was the case provides evidence for the role of neural temporal coding in the perception of consonance, and in the perception of pitch more generally. That a measure of the representation of harmonicity in the FFR predicts individual differences
in this way provides support for the harmonicity theory of consonance, and suggests that the harmonicity that drives the perception of consonance may be encoded by neural phase locking. A systematic comparison of FFR indices would be required to determine whether harmonic salience is the FFR measure most sensitive to consonance perception: however, the results of the study in Chapter 4 suggested that this measure was sufficiently sensitive for its continued use in subsequent studies.

The results of the study in Chapter 4 also demonstrate that harmonic salience calculated from the FFR to diotically presented consonant dyads was greater than that calculated from dichotically presented consonant dyads, due to the addition of distortion products at harmonic frequencies. The enhanced harmonicity of the FFR in the diotic condition corresponded with consonant dyads being rated as more pleasant in this condition, adding further support to the hypothesis that harmonicity of the FFR is a predictor of the perception of consonance. Suppressing the FFR to the envelope and even-order distortion products removed the large frequency component corresponding to the difference tone of the two F0s, indicating that the distortion products resulting from diotic presentation are even-order, and possibly coded in the cochlear envelope rate. These results suggest that the ‘fusion’ that occurs when combing musical notes with simple frequency ratios may be driven by phase locking to additional distortion products which enhance the overall harmonicity. Although the distortion product at the missing F0 of the spectrum of the FFR to the Perfect 5th was not seen in Chapter 5, harmonic distortion products were seen throughout the rest of the spectrum in all of the studies.
A question left unanswered by the study in Chapter 4 was that of whether the relation between consonance preference and NCI was driven by a shared dependence on factors relating to music experience. Chapter 5 was therefore designed to test the hypothesis that NCI predicted consonance preference in normal hearing non-musicians. NCI was found to significantly correlate with consonance preference, implying a relation between temporal coding and the perception of consonance that is independent of musical experience.

The criteria for inclusion as a ‘non-musician’ in this study was less than one year formal music training, with that period having ceased over five years previous. These are similar criteria as used elsewhere (e.g. no formal training within five years; Norman-Haignere, Kanwisher, & McDermott, 2013; no more than one year formal training; Itoh, Suwazono, & Nakada, 2010; no more than three years formal training; Bidelman & Krishnan, 2009; Wong, Skoe, Russo, Dees, & Kraus 2007; no more than six years formal training, with training having ceased more than ten years previous; Song, Skoe, Banai, & Kraus, 2011). However, it should be noted that these criteria are not unambiguous: for example, the decision was taken to exclude an individual with experience of creating music using MIDI and music production software, despite this not constituting formal training. Similarly, it was decided that disk jockeys should be excluded. It should also be noted that these criteria control for experience of playing a musical instrument, but not exposure to music per se: no measure of time spent listening to music was taken. If musical experience affects consonance preference and the FFR due to the listener attending to musical sounds, then attentive music-listening may also have an
effect, in the absence of formal music training *per se*. Future work will be needed to determine whether this is the case; however, it is possible that the relation between the FFR and consonance preference reported in Chapters 4 and 5 is due to a shared dependence on time spent *listening* to music.

The FFR addition and subtraction procedure used in Chapter 4 was also used in Chapter 5, in order that the results could be compared. Whereas the relation between NCI and consonance preference in Chapter 4 was only found for the subtraction FFR, in Chapter 5 the relation was found for the FFR as recorded. In the interests of statistical power, analysis in the subsequent study was only performed on the FFR as recorded (FFR$_{\text{RAW}}$).

The study in Chapter 5 provides evidence for the FFR to low-frequency musical dyads being generated in part by the basal portion of the cochlea, and demonstrates that the effect of high-pass filtered masking noise on the FFR can be accounted for by a model of the IHC RP. That the high-pass filtered masking noise resulted in a reduction in FFR amplitude and harmonic salience demonstrated that the use of the masking noise in order to control for high-frequency hearing loss in the subsequent study was not viable.

The study in Chapter 6 was designed to test the following hypotheses: i) that older listeners would have a weaker representation of the harmonicity of dyads in the FFR, and that this would result in ii) older listeners having a weaker perceptual distinction between consonance and dissonance. The results demonstrated that, as shown by Anderson, Parbery-Clark, White-Schwoch, and Kraus (2012), the amplitude of the FFR was lower in older listeners, resulting in weaker harmonic salience scores in general and lower NCI scores. Note that, despite NCI predicting consonance preference over
the whole sample, these results do not provide evidence for NCI being predictive of consonance preference; rather, they demonstrate that NCI, a measure that was found to be predictive of consonance preference in Chapters 4 and 5, declines with age concurrent with a decline in consonance preference.

Further work is required to determine what are the implications for an age-related decline in consonance preference. Western music is constructed around principles of harmony – a central aspect of which is the distinction between consonance and dissonance. It would seem likely therefore that a loss of this distinction would be associated with a loss of the enjoyment of music. However, some caution should be exercised in interpreting causal relationships from the correlations contained within this thesis: for example, since both consonance preference and NCI are strongly associated with musical experience (Chapter 4), it is possible – albeit unlikely perhaps – that the effects of age (Chapter 6) are mediated through an age-related decline in engagement with music. Another area for possible future work is to determine whether auditory training or music-based therapy can improve temporal coding – and alter consonance preference. As discussed in Chapter 1, the benefits of music therapy are well documented: it is possible that some of the benefits of listening to music in old age are mediated through improving temporal coding (e.g. speech-in-noise discrimination; Anderson, White-Schwoch, Parbery-Clark, & Kraus, 2013). Caution should also be exercised in interpreting what the NCI measure used in this thesis – and the FFR more generally – represents. For example, since the stimuli used throughout are necessarily low-frequency it is unclear how the results
relate to the consonance of high-frequency dyads which presumably cannot be encoded temporally in the same way.

In summary, the conclusions to be drawn from the work contained within this thesis are:

i) The FFR to low-frequency dyads is partly generated by the high-frequency, basal region of the cochlea.

ii) The effects of high-pass filtered masking noise on the FFR to low-frequency musical dyads can be accounted for by a model of the IHC RP.

iii) For low-frequency musical dyads, the representation of the harmonicity of musical dyads by neural temporal coding predicts their perceived consonance.

iv) The NCI measure derived from the representation of the harmonicity of consonant and dissonant low-frequency dyads is sensitive to individual differences in preference for consonance.

v) Musical experience is associated with both consonance preference and NCI.

vi) The distinction between low-frequency consonant and dissonant dyads, in terms of harmonic salience of the FFR, declines with age.

vii) The perceptual distinction between low-frequency consonant and dissonant dyads also declines with age.

Taken together the research presented in this thesis provides evidence for the role of temporal coding in music perception, and describes a method for
investigating the causes and consequences of individual differences in this area.
Bibliography


Bledsoe, S. C., Jr., & Moushegian, G. (1980). The 500 Hz frequency-
following potential in kangaroo rat: an evaluation with noise masking. Electroencephalography and Clinical Neurophysiology, 48, 654-663.


Neuroscience, 25, 10952-10959.


neural synchronization in the anteroventral cochlear nucleus. I. Responses to tones at the characteristic frequency. Journal of Neurophysiology, 71, 1022-1036.


Electroencephalography and Clinical Neurophysiology, 32, 63-74.


evoked potentials: basic principles and clinical application (pp. 336-354): Lippincott Williams & Wilkins.


Ohm, G. S. (1843). Ueber die Definition des Tones, nebst daran geknüpfter


Russell, I. J., & Sellick, P. M. (1978). Intracellular studies of hair cells in the


Smalt, C. J., Krishnan, A., Bidelman, G. M., Ananthakrishnan, S., & Gandour,


