Temporal Regularity in Audition



Vani Rajendran
University College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy
Michaelmas 2017

To my family:

Appa, Amma, Dhevi, Sakthi, Thatha.

Acknowledgements

Where to even begin...

I would first like to thank the Wellcome Trust for this incredible opportunity and privilege, as well as the Clarendon Fund. I am also deeply grateful to my supervisors Jan, Andy, and Ben W for their steadfast support and encouragement throughout, starting from when I joined the group as a bright-eyed MSc student 5 years ago. I would also like to thank my past and present colleagues from the Auditory Neuroscience Group for being a fun and inspiring bunch to work with. In particular, I must thank Nicol for his ability to come up with brilliant analyses, spot every possible flaw in anything, and be just downright infuriating when it comes to writing manuscripts together. And Dan for teaching me how to play Piano Man, among other things. And Ben W for being awesome (or crazy?) enough to embark on the adventure that Methodicate has been and continues to be. I also thank my family and friends who stood by me even through the rougher bits. There are too many names to list but if we've shared a laugh or a tear, I am thinking of you. If we've made music together, I am also thinking of you.

And finally, last but not least, I thank you, Ben. This is not the place to list the ways in which you have supported me through the years and the ways you have challenged me to reach higher than I ever thought I could. Without you there would neither have been a beginning nor an end.

Abstract

Sound, by its very nature, is a temporal phenomenon. Everything from the perception of pitch to the sense of closure at the end of a symphony, relies on the brain's ability to integrate information over time. Ultimately it is perception that enables the richness of our interactions with the world around us, and it is a remarkable feat that the brain can quickly and accurately sift through the flood of information entering the ears to construct a coherent yet dynamic internal representation of the external world. Underlying this feat in part is the brain's ability to rapidly detect temporal patterns over timescales that span orders of magnitude, from sub-milliseconds to tens of seconds. Of particular interest is the ability to detect rhythms, or sound patterns in the range of hundreds of milliseconds to seconds. This timescale is particularly fascinating because it is critical to the perception of rhythm and beat in music, an ability that comes surprisingly naturally to us despite its neural underpinnings being far from understood. It is not just musical beat perception that remains mysterious; even the brain's mechanisms for detecting simple temporal patterns at this timescale are still not known. The work in this thesis therefore explores two broad questions that are key to understanding temporal processing at the rhythm timescale: 1- what are the perceptual consequences of rhythmic temporal regularity in sound? and 2- how does the brain detect temporal regularities at the rhythm timescale? I combine human psychoacoustics, rodent electrophysiology, and computational

modelling to demonstrate that rhythmic sound patterns are easier to detect than arrhythmic ones, that adaptation in the auditory system may be a mechanism by which information about temporal structure at the rhythm timescale is encoded and made available to higher structures, and that low-level auditory processing may play a substantial part in shaping complex percepts such as where and how clearly we feel the beat in music.

Contents

1 General Introduction **1**

2 Predictable Timing Facilitates the Detection of Repeating Sound Patterns **6**

2.1 Introduction 8

2.2 Methods 9

2.3 Results 12

2.3.1 Shorter noise targets are harder to detect 13

2.3.2 Substantial jitter impairs noise target detection. 13

2.3.3 What about natural sounds? 15

2.3.4 Adequate sampling in stimulus set is crucial. 15

2.4 Discussion 16

2.4.1 Does repetition detection rely on synaptic memory traces of recent inputs? 17

2.4.2 Modulation filterbanks as an alternative mechanism. 17

2.4.3 Repetition of frozen noise targets results in distinct peaks in the modulation spectrum. 18

2.4.4 Could a modulation filterbank model account for previous findings? 21

2.5 Conclusions 22

3 Midbrain Adaptation May Set the Stage for the Perception of Musical Beat **23**

3.1 Introduction . . . 25

3.2 Results . . . 26

3.2.1 On-beat sounds evoke stronger neural responses on average than off-beat sounds . . . 26

3.2.2 A large on-beat neural emphasis is a defining feature of the perceived beat . . . 27

3.2.3 Strength of on-beat neural emphasis varies systematically with the profile of each recorded IC unit's firing pattern . . . 33

3.2.4 Adaptation may explain why neural responses are stronger on the beat . . . 35

3.2.5 Problems with the beat-related z-score metric used in Nozaradan, et al. (2012) . . . 36

3.3 Discussion . . . 41

3.4 Methods . . . 46

3.4.1 Sound Stimuli . . . 46

3.4.2 Electrophysiological recordings . . . 48

3.4.2.1 Surgical protocol . . . 48

3.4.2.2 Data preprocessing . . . 49

3.4.2.3 Clustering . . . 50

3.4.2.4 Psychoacoustics . . . 51

4 Responses in Primary Auditory Cortex May Shape Where and How Clearly the Beat is Felt in Real Music **53**

4.1 Introduction . . . 56

4.2 Results . . . 57

4.2.1 Firing rates in the rat auditory cortex are on average higher on the beat than off the beat . . . 57
4.2.2 Many units in the rat auditory cortex show a significant on-beat emphasis on a trial-by-trial basis . . . 61
4.2.3 Many units also show significant "amplification" of the on-beat emphasis relative to cochleagram activation on a trial-by-trial basis . . . 64
4.2.4 Cortical activity is more successful at "finding" the beat than cochleagram activation . . . 66
4.2.5 What about the songs that showed little to no on-beat neural emphasis? . . . 72
4.3 Discussion . . . 73
4.4 Methods . . . 79
4.4.1 Stimuli . . . 79
4.4.2 Surgical protocol . . . 79
4.4.3 Data analysis . . . 80
4.4.3.1 Data preprocessing . . . 80
4.4.3.2 Calculating the STRF using an LN model . . . 81

5 General Discussion 82

5.1 Music psychology and the natural bounds of human temporal processing 85
5.1.1 Timescales . . . 85
5.1.2 Beat - a perceptual accent . . . 86
5.1.3 Not all beats are created equal, nor is there always an accent: subjective rhythmisation . . . 88
5.1.4 To summarise . . . 89
5.2 The psychoacoustics of beat perception: sensorimotor synchronisation 90
5.2.1 When tapping along, we are usually early . . . 90

5.2.2 Beat period and phase may have distinct underlying representations . . . 91
5.2.3 How fast we can tap along depends on what we are tapping to 93
5.2.4 To summarise . . . 94
5.3 Neurophysiology of beat perception . . . 94
5.3.1 Strong beats differ physiologically from weak beats . . . 95
5.3.2 Entrainment of oscillatory activity to musical beat . . . 96
5.3.3 Brain areas involved in beat perception . . . 97
5.3.4 Model-based accounts of beat perception . . . 100
5.3.5 Beat processing in nonhuman species . . . 101
5.3.6 To summarise . . . 102
5.4 Predictable timing in auditory perception . . . 103
5.4.1 Temporal predictability in pattern detection . . . 103
5.4.2 Temporal predictability in auditory scene analysis . . . 105
5.4.3 Rhythms in speech perception . . . 106
5.4.4 To summarise . . . 108
5.5 Neural mechanisms of timing . . . 109
5.5.1 Dedicated timekeeping circuits? . . . 109
5.5.2 Is the passage of time implicit in neural responses? . . . 111
5.5.3 To summarise . . . 112
5.6 Conclusions and outlook . . . 113

Bibliography **116**

List of Figures

2.1 **Stimulus design**. A) For each of the parameter combinations shown for the REP conditions, 16 trials were generated. Target duration is displayed in milliseconds and as a percentage of the fixed inter-onset interval (IOI) of 500 ms. Jitter is shown as the standard deviation of the Gaussian distribution from which IOI durations are drawn in milliseconds, and as a percentage of the mean IOI of 500 ms. B) All stimuli were 8s in duration and began as non-repeating noise that then transitioned (if at all) at a random time between 3 and 4 seconds to a repeating section (or pink noise in the case of PINK). C) Schematic representations of the different stimulus conditions where non-repeating noise is shown in white, repeating noise targets are patterned, and pink noise is shown in gray. T500-J0 (target duration of 500 ms, no jitter, top) is closest to the type of stimuli used in previous noise learning studies. A target duration of 300 ms (T300) was used in the jittered (REP-J) context. Control conditions were RAND (non-repeating white noise) and PINK (non-repeating white noise that transitions to pink noise). N=16 trials for PINK and each of the 7 REP conditions, and N=64 trials for RAND. 11

2.2 **Detection performance under temporally regular and jittered conditions**. A) Fraction detected for REP-R conditions. Solid lines represent individual subjects, and boxplots summarise the data over the 17 subjects for a given stimulus condition. For display purposes only, a small amount of Gaussian noise (std=1.5%) has been added to the y-values of the solid lines in panels A-D to spread out overlapping data points. B) Fraction detected for REP-J conditions. C) Mean reaction times for REP-R conditions. D) Mean reaction times for REP-J. 14

2.3 **Inadequate sampling can lead to incorrect conclusions.**. A) Percent detected for jittered conditions, where each colour is a different subject. These data are based on an initial run of this experiment (with a different set of participants) where stimuli were not generated afresh for each subject. B) For this limited stimulus set, detection performance was best in the 2% jitter condition (Wilcoxon paired signed rank test, $* \; p < 0.05$, $** \; p < 0.01$). 16

2.4 **A modulation filterbank model would account for the effect of temporal jitter on detection performance**. A) Method used to quantify the peakiness of the modulation spectrum, shown for one example from REP-R (T300-J0), REP-J (T300-J100), and RAND. From left to right: the raw sound; cochleagram; modulation spectrum in each frequency channel; modulation spectrum summed over frequency channels. Note the vertical stripes at 2 Hz and its harmonics present in both REP stimuli, and their absence in RAND. The standard deviation of the summed modulation spectrum was calculated for each stimulus trial. B) Top: Fraction detected plotted against the standard deviation of the summed modulation spectrum. Each point represents the fraction detected, out of the nearest-neighboring 16 stimulus examples (regardless of subject) along the x-axis from within the same condition. Different colors represent the different conditions, as shown in the legend. Black line is the fit from a logistic regression. Bottom: Reaction time in seconds plotted against the standard deviation of the summed modulation spectrum. For display purposes, the average (clumped) reaction times of each 16 nearest neighbors along the x-axis from the same stimulus condition are plotted. The black line is a linear regression run on all (un-clumped) data. On the secondary y-axis is a histogram showing the distribution of all (un-clumped) standard deviation values for all stimuli within each condition. C) A histogram showing the distribution of standard deviation values calculated over 1 s intervals during RAND (black) and T500-J0 (blue), analogous to the stimuli used in Agus, et al. 2010. Note the larger standard deviation values for both conditions in panel C using a 1 s window compared to the 3 s window used in panel B. For our examples of RAND ($n = 1088$) and T500-J0 ($n = 272$), we see no overlap, suggesting that the modulation spectrum would contain enough information to detect repetition from a single repeat of a 500 ms frozen noise target. 20

3.1 **On-beat responses are on average stronger than off-beat responses.** A) An example units normalised peri-stimulus time histogram (PSTH) to stimulus pattern P2 at 1x tempo. The last three complete cycles of this pattern are shown together with the stimulus trace above. The coloured numbers indicate noise bursts, green being on-beat and blue being off-beat. Normalisation was done by dividing each unit's firing rate concatenated across all 7 stimuli by its standard deviation. B) Raster-style plot showing tap times of human listeners across trials and cycles of this stimulus. Each dot marks the timing of a tap, each row of dots shows tap responses for a single cycle of the stimulus pattern, and different colour dots distinguish the tap responses from different subjects. Beneath is a histogram of tap times pooled across subjects, cycles, trials for this stimulus. A clear, regular tapping pattern is present, which indicates a strong consensus among listeners to hear four beats in this 16-event sequence, timed on the 1st, 5th, 9th and 13th interval. C) Normalised firing rate in response to on-beat (left, in green) and off-beat (right, in blue) noise bursts for the highlighted cycle shown in panel A. To quantify a unit's mean on- and off-beat response, the firing rate over the first 40 ms of each sound-evoked response to on-beat and off-beat sounds (red boxes) across all 33 s of all stimuli was averaged (the average based on just the highlighted cycle are shown for demonstration purposes), and the mean of the result was taken. D) Scatter plot showing the mean on-beat response for each unit as a function of that unit's mean off-beat response. Most points fall well above the main diagonal, indicating that on average, on-beat sounds evoked stronger responses than off-beat sounds. . . . 28

3.2 **Raster-style plot of tap times for each stimulus.** Each dot marks the timing of a tap, each row of dots shows tap responses for a single cycle of the stimulus pattern, and different colour dots distinguish the tap responses from different subjects. Below (in grey) is a histogram of tap times pooled across subjects, cycles, trials for each stimulus. At the bottom of each panel is the stimulus pattern. 29

3.3 **On-beat firing rate versus off-beat firing rate for each stimulus.** All stimulus patterns except P3 show on-beat responses that are significantly larger than off-beat responses ($p < 0.0001$, Wilcoxon paired signed rank test, Bonferroni corrected, $N = 248$ units). 30

3.4 **Exploration of on-off for all possible combinations of beat periods beat positions.** A) Heatmaps represent on-off for all possible combinations of beat grouping (y-axis) and beat position (x-axis) for each of the 7 rhythmic sound patterns tested. Red Yellow and white cyan boxes mark the most and second most common beat interpretations reported by human listeners, respectively. Brackets indicate those groupings whose theoretical tapping frequencies are between 0.5-4 Hz, the range within which beat is typically perceived [London, 2012b]. B) Same as panel A, but based on population firing rates. C) All hypothetical on-off values for sound (grey) and the two most commonly perceived beat structures (red), pooled across the 7 stimuli. D) Same as panel C, but for population firing rates. 32

3.5 **Clustering analysis.**A) Heat map illustrating each units normalised firing rate in response to one cycle of pattern P2 at 1x tempo. Units are grouped by cluster. The stimulus pattern is plotted beneath each panel for reference. B) Each clusters mean peri-stimulus time histogram (PSTH), with the total number of units in each cluster (N) displayed to the left of each trace. C) Mean PSTH traces of multi-units (blue) and single units (red) in each cluster. The number of single units (n) is displayed to the left. D) Same as Figure 3.1D, but with units labelled by the cluster to which they belong. 34

3.6 **Firing rates estimated from exponential fits closely match measured on-off responses.** A) The average firing rate evoked by noise bursts preceded by varying amounts of silence is used to fit an exponential to each unit. Data and exponential fits for an example unit from each cluster are shown. B) Scatterplot shows a high correspondence between estimated on-off values based on exponential fits (x-axis) and the real data (y-axis). 36

3.7 **Heatmaps of on-off values by stimulus.** A) Heatmaps of *on-off* values for sound for all plausible beat structures for each of the 7 stimuli. Yellow boxes mark the most commonly tapped beat structure, and cyan boxes mark the second most commonly perceived beat structure for each stimulus. The histograms directly below show the two most commonly perceived beat structures in red, and all possible beat structures in grey. Note that *on-off* values were calculated over each stimulus as played, which results in small discrepancies between the *on-off* histograms for sound for the same pattern presented at different rates. This is due to the looped pattern being cut off in slightly different places at the end of the 33 s. If the final incomplete loop of each stimulus pattern is cut off artificially, the sound on-off histograms for P1 at all rates and P2 at all rates become identical as expected. B) Same as panel A, but for firing rates. C) Same as panel A, but for firing rates estimated by the exponential fits modelling adaptation. Note the close similarity between values in panel B and C. 37

3.8 **Why a frequency domain analysis can be misleading.** A) Manipulating the envelope of pattern P2 at 1x profoundly changes the peaks in the frequency spectrum and changes the beat-related z-score. The top row are the time domain signals with a normal (left), shortened (middle), and sawtooth (right) shape. Underneath is the result of performing a Fourier transform on the top signals. Beat-related frequencies are marked by red circles. Note that the relative size of all peaks changes with minor changes in shape of the time domain signal, and the the changes made do not necessarily reflect anything to do with beat processing. B) A tone (grey) with a 16 Hz amplitude modulation (red envelope) would result in a positive beat-related z-score though there is no meaningful beat present. C) A negative beat score can be obtained, even though a clear 1 Hz beat is present in the signal. . . . 38

3.9 **An illustration of the rhythmic patterns P1-P3, from top to bottom.** The slowest tempo (1x) is illustrated for each pattern where each event was 200 ms in duration. In the experiment, P1 was played at tempi of 1x, 2x, and 4x, P2 at 1x, 2x, and 3x; and P3 at 1x. 47

3.10 **Principal Component Analysis for hierarchical clustering.** A) The first five principal components based on clustering normalised firing rates. B) The elbow method was applied to decide that five principal components would be used for clustering. C) The elbow method applied again to decide that 8 clusters adequately explained variance in the data. One of the 8 clusters contained only 1 unit, and this cluster was discarded from further analysis, yielding the 7 clusters referred to throughout. 50

4.1 **Beat-triggered firing rates by stimulus.** Each colour represents the firing rate of one of the 98 cortical units that make up this dataset. 58

4.2 **Beat-triggered cochleagram activation by stimulus.** Each colour represents one of the 27 1/3-octave spaced frequency channels in the cochleagram, measured in arbitrary units. 59

4.3 **On-beat and off-beat firing rates by stimulus**. The x-axis is mean off-beat firing rate, y-axis is mean on-beat firing rate. Each dot represents one unit (98 in total). Musical excerpts for which neurons showed significantly higher on-beat firing rates than off-beat firing rates are indicated by a red diagonal. Those that showed no significant difference between on-beat and off-beat firing rates are indicated by a blue diagonal. 62

4.4 **On-beat and off-beat cochleagram activation by stimulus**. The x-axis is mean off-beat cochleagram activation in arbitrary units, y-axis is mean on-beat cochleagram activation in arbitrary units. Each dot represents one of the 27 1/3-octave spaced frequency channels in the cochleagram. Musical excerpts for which cochleagram activation was significantly higher on the beat than off the beat are indicated by a red diagonal. Those that showed no significant difference between on-beat and off-beat cochleagram activation are indicated by a blue diagonal. 63

4.5 **A grand average comparison of on-beat and off-beat cochleagram and neuronal firing rates**. (Left) Average on-beat vs. off-beat cochleagram activation across all musical excerpts for each of the 27 frequency bands with 1/3-octave spacing spanning the range from 50 Hz to approximately 20 kHz, measured in arbitrary units. (Right) Average on-beat vs. off-beat firing rate across all musical excerpts for each of the 98 single and multi-units. 64

4.6 **Units with a significant on-beat emphasis on a trial-by-trial basis**. Rows are units and columns are songs. Coloured boxes represent unit/song combinations where on-beat responses were significantly larger than off-beat responses on a trial-by-trial basis. 65

4.7 **Units that show a significant trial-by-trial amplification of on-beat emphasis relative to that present in the cochleagram**. Rows are units and columns are songs. Coloured boxes represent unit/song combinations where on:off-beat for neural activity was significantly larger than on:off-beat cochleagram activation at that unit's BF on a trial-by-trial basis. 67

4.8 **Comparison of how the consensus beat ranks against other plausible beat structures based on cortical activity and cochleagram activation.** (Left) Median rank across trials for the consensus beat structure based on on:off-beat values for all plausible beat structures. Ranks were calculated for all unit/song combinations where cortical activity showed significant amplification of on:off-beat values over that of the cochleagram. (Right) Consensus beat rank based on cochleagram activation in the frequency band corresponding to that unit's BF. 70

4.9 **Difference in consensus beat ranks based on cortical activity and based on cochleagram activation.** Values shown are the consensus beat's rank based on cortical activity subtracted from the median rank across trials based on cochleagram activation. Positive values mean that cortical activity would have been more effective than cochleagram activation at distinguishing the consensus beat out of a pool of plausible beat structures based on a higher on:off-beat value. . 71

4.10 **Tapping precision as a function of on-beat neural emphasis.** The x-axis is a simple count of the number of units (out of the 98 recorded from in this dataset) that showed significant on-beat cortical gain relative to cochleagram activation on a trial-by-trial (see Figure 4.7). The y-axis is the variance around consensus tap times as a fraction of the beat period. This was calculated by first pooling all taps across all annotators for a given song, then computing the distance between every tap and its nearest consensus tap time, and then normalising by the consensus beat period for that song so that distances were expressed as a fraction of the beat period. Finally, the variance was taken across these normalised distances to arrive at the tap variance for each song. Each data point in the figure represents one song. 74

5.1 **Timescales in auditory perception.** A) The beat perceived depends on the tempo at which a musical rhythm is played. In this simple, recognizable example rhythm, notes represent sound events and those with a single stem are quarter notes, notes with the attached stem are eighth notes, and the remaining symbol is a quarter rest (silence). The basic unit of time here is the quarter note; a quarter rest is the same duration as a quarter note, and each eighth note is half the duration of the quarter note. Tempo is conventionally specified in beats per minute, so for the slow tempo (in red), there would be 75 quarter notes per minute, and each quarter note is therefore 800 ms in duration. The fast tempo (in blue) is twice the speed of the slow tempo. In both cases, the beat may be comfortably perceived at 800 ms intervals or 1.25 Hz (filled circles), but depending on the tempo this may coincide with different events in the music. The alternation of strong (solid lines) and weak beats (dotted lines) are illustrated for each tempo. Syncopation (green triangle), or when a beat is felt where there is silence, is very common in music. B) This schematic illustrates the timescales over which common auditory events unfold. Time is on a log scale from small intervals (fast rates) to large intervals (slow rates), with values shown in milliseconds and in Hz. The indifference interval is marked in purple; shorter intervals are temps courts, longer intervals are temps longs. 87

5.2 **Illustration of a selection of perturbations used to study period and phase correction in sensorimotor synchronisation.** The x-axis represents time, and here a temporal grid representing 600 ms intervals is marked by the vertical dotted lines. Circles represent clicks to which a listener would align their taps, and blue circles mark sounds whose timing would be a departure from the isochronous condition where there is no perturbation. In the Isochronous condition (top), a click is played every 600 ms. In Anomaly, a single click in the sequence is manipulated such that the IOIs that flank it are too long and too short by 100 ms, allowing the remainder of the sequence to remain unchanged. In Phase Shift, a single IOI is lengthened by 100 ms but this time is not gained back, resulting in a phase shift of 100 ms that persists for all remaining clicks, even though their IOI remains 600 ms. In Tempo Change, the IOI changes abruptly from 600 ms to 500 ms. This would be perceived as a faster tempo and would require and adjustment in the period of taps. 92

Chapter 1

General Introduction

An appreciation of music is only one of many consequences of an ability to detect auditory rhythms. To perceive a temporal pattern is to make a prediction about what should happen in the future, and a practical advantage of this is the ability to quickly parse a complex acoustic scene such as a noisy pub into distinct perceptual objects [Winkler et al., 2009a]. Temporal coherence, or the near-simultaneous to simultaneous presentation of different sounds, is a powerful cue that binds sounds such that they are perceived as a single sound source [Turgeon et al., 2002, Turgeon et al., 2005, Shamma et al., 2011]. Across time, the predictability of features such as location, pitch, loudness, and timbre also play a pivotal role in auditory scene analysis [Bendixen, 2014]. Specifically, predictability across time influences whether temporally disjointed sounds that comprise a scene are perceived as integrated or segregated [Bendixen et al., 2010, Andreou et al., 2011, Rajendran et al., 2013]. Predictive coding based on temporal and other feature regularities may also account for the stability of auditory objects [Pressnitzer and Hupe, 2006, Denham and Winkler, 2006, Chait et al., 2007, Winkler et al., 2012].

Speech is perhaps the most pervasive and critical context in which we rely on our ability to derive meaning from complex temporal patterns. The intelligibility of

speech has been shown to correlate with the brains ability to follow the speech envelope [Ahissar et al., 2001, Peelle and Davis, 2012], particularly in the 4-8 Hz range [Luo and Poeppel, 2007]. This range corresponds to the syllable rate of speech production [Greenberg et al., 1999] and dominates the temporal modulations in the speech envelope [Chi et al., 1999, Chandrasekaran et al., 2009, Elliott and Theunissen, 2009]. The syllabic rate is nearly an order of magnitude slower than fine structure elements such as formants (30-50 Hz), and a few-fold faster than intonation contours that are typical of phrasal units (1-2 Hz). Content at all of these timescales is parsed concurrently to extract meaning. Interestingly, temporal manipulations to speech more drastically impair intelligibility [Adank and Janse, 2009] than extreme spectral manipulations do [Shannon et al., 1995].

Current theories suggest that auditory object formation and speech perception may rely on attention-driven oscillatory mechanisms that entrain to temporally structured sound features [Lakatos et al., 2008, Schroeder and Lakatos, 2009, Ahissar et al., 2001, Peelle and Davis, 2012, Luo and Poeppel, 2007], as well as basic sensory neural mechanisms [Fishman et al., 2012]. Though the question of how stable perceptual objects are formed from potentially noisy and conflicting information is still an open one, the final representation of an auditory object is remarkably distinct and robust, even if it overlaps spectrotemporally with the unattended background [Ding and Simon, 2012, Ding and Simon, 2013].

Musical beat perception is an interesting case where the temporal regularities detected are more abstract yet compellingly clear perceptually. Beat, or the steady pulse that emerges from music that one can tap or dance to (typically periodic in the Western music tradition), is not a physical property of the sound itself, but is a perceptual phenomenon that is widely considered to be a cognitive skill, though its species-specificity and domain-specificity are topics of current debate [Honing, 2012]. Our ability to move to the beat, though natural, is not a trivial feat. It requires

an ability to use sensory input to plan motor actions in anticipation of future sensory input [Repp, 2005, McAuley et al., 2006, Repp and Su, 2013], and yet we do this spontaneously [Van Dyck et al., 2015]. Even in the absence of movement, there is a spontaneous tendency to internally impose a rhythm where there is none, for example by hearing identical ticks of a clock as tick-tock-tick-tock [Bolton, 1894, Brochard et al., 2003, van Noorden and Moelants, 1999, Bååth, 2015].

Musical rhythms are particularly engaging to a vast array of brain areas that include auditory and cognitive areas, as well motor areas such as the basal ganglia, supplementary motor area, striatum, cerebellum, sensorimotor cortex, and premotor cortex [Parsons, 2001, Grahn and Brett, 2007, Zatorre et al., 2007, Chen et al., 2008, Grahn, 2009, Teki et al., 2011]. Engagement of motor-related areas appears to be automatic since it is observed consistently even when listeners are instructed not to make overt movements [Chen et al., 2008]. Again, entrainment of oscillations appears to be central to musical rhythm perception. Cortical oscillations have also been shown to reflect metrical structure; this is noteworthy because it suggests that oscillations, in addition to entraining to the rate of individual events in a rhythmic sequence [Doelling and Poeppel, 2015], are also able to entrain to higher-level temporal regularities, but the precise mechanism behind this is still unknown. Neural resonance theory models the interactions between a sensory and motor layer of neurons as a dynamical system, and the results resemble both neurophysiological and behavioral aspects of beat perception [Large et al., 2015]. Neural resonance theory is compatible with the dynamic attending theory, which postulates oscillatory fluctuations in attention [Large and Jones, 1999]. These theoretical models are consistent with physiological evidence that entrainment of oscillations may provide a mechanism for sensory selection [Schroeder and Lakatos, 2009, Lakatos et al., 2013] and for communication between different neuronal groups [Fries, 2015].

Stepping back from all of this, there is a fair amount of evidence supporting

the notion that entrainment of oscillations is central to the brains ability to process temporally predictable sounds in the rhythm range. A key unanswered question is how different oscillatory populations of neurons come to be entrained to different time-varying sound features. In other words, how are temporal patterns at the rhythm timescale detected in the first place? Given that all sounds enter the ear and ascend the auditory pathway, entrainment processes must at some point be sensory-driven. It is therefore not unreasonable to imagine that some relevant precursors to perception must be observable in the auditory system, that they emerge as a result of bottom-up auditory processing, and that they might provide clues as to how entrainment process are set into motion.

The big picture question, therefore, is how auditory processing drives the higher-level processing of information that ultimately enables perception and interaction. What this thesis strives to do is to add crucial missing data points to the questions of how temporal structure influences the detection of sound patterns, and how low-level sensory processing influences the representation of temporally structured sound such that it may serve as a relevant precursor to the perception of musical beat. Musical beat perception is particularly tricky since nonhuman species, apart from isolated exceptions [Patel et al., 2009, Cook et al., 2013, Rouse et al., 2016], have shown little compelling evidence of being able to perceive and synchronise to the beat as precisely as humans [Geiser et al., 2014]. However, the cross-species methods developed here leverage the possibility that bottom-up sensory processing might to a large degree be similar across mammalian species, and use this to address the current methodological and conceptual gaps between human and animal studies on the topic of rhythm [Rajendran et al., 2017b].

In Chapter 2, I show that hidden sound targets become easier to detect if they appear at regular intervals and demonstrate that a modulation filterbank model could account for this result. In Chapter 3, I show that midbrain adaptation may be shaping

the representation of rhythmic sounds in a manner that influences where the beat is felt in musical rhythms. In Chapter 4, I show that firing rates in the auditory cortex are similarly stronger on the beat than off the beat in real music, and that this may influence not only where the beat is felt, but also how precisely listeners can tap to the beat. This thesis concludes with a general discussion in Chapter 5 that surveys the wider interdisciplinary context within which this work exists and provides an outlook for future work.

Chapter 2

Predictable Timing Facilitates the Detection of Repeating Sound Patterns

Most of the work contained in this chapter has been published in:

Rajendran, V. G., Harper, N. S., Abdel-Latif, K. H., & Schnupp, J. W. (2016). Rhythm facilitates the detection of repeating sound patterns. Frontiers in neuroscience, 10.

I came up with the study design and data analysis with input from NSH and JWHS, who were also involved in revising the manuscript. KHAL collected a set of initial pilot data for this study. The final data acquisition, data analysis, and writing are my own.

Abstract

Humans show a remarkable ability to quickly and reliably detect repeating sound patterns of arbitrary complexity. However, the studies that have shown this have only explored conditions where repeating target sounds were presented at precisely regular (isochronous) time intervals, raising the question of whether temporal regularity may be aiding in the detection of a repeating pattern. In this study, I presented human listeners with white noise stimuli that either contained a frozen segment of noise that repeated in a temporally regular or jittered manner, or did not contain any repetition at all. Subjects were instructed to respond as soon as they detected any repetition in the stimulus. I found that pattern detection was best when repeated targets occurred in a temporally regular manner, suggesting that temporal regularity plays a facilitative role in pattern detection. I contextualise these findings through an analysis of temporal jitter present in natural sounds, and then demonstrate that a modulation filterbank model could account for these results.

2.1 Introduction

Beneficial to survival in a complex and ever-changing acoustic environment is the ability to quickly identify relevant sounds that comprise the scene. One useful strategy is to detect recurring patterns over time, as these are often salient and suggestive of animate sound sources. Consider footsteps: steps on gravel sound nothing like steps through grass or through puddles, yet all of these very disparate sounds are easily recognised as the sound of footsteps if they occur in a rhythmic, repeating pattern. To recognise rhythmic patterns, the brain needs to search for recurrences of arbitrary and potentially complex sounds over timescales ranging from fractions of a second to tens of seconds.

Studies of auditory pattern detection often employ Gaussian white noise stimuli because they are spectrally broadband and devoid of prior meaning to listeners. Humans exhibit an impressive capacity to rapidly form recognition memories of frozen white noise tokens [Kaernbach, 2004], and the memories traces have been shown to last for weeks [Agus et al., 2010]. While human sensitivity to arbitrary repeating patterns has been well documented [Kaernbach, 2004, Chait et al., 2007, Agus et al., 2010], the question of how repetition is detected in the first place remains poorly understood. Previous noise learning studies have only explored conditions where repeating noise tokens were presented at precisely regular (isochronous) time intervals, and it is unclear whether such regularity is necessary or helpful for pattern detection. If sensory memory alone is responsible for pattern detection, then whether the sounds occur at regular or irregular intervals should have no effect on pattern detection.

However, the experiments described here reveal that detection performance does decline with increasing temporal irregularity, indicating that a sensitivity to slow temporal modulations or entrainment to the rhythmic structure of incoming sounds

might facilitate pattern detection.

2.2 Methods

The experimental methodology was approved by the local Ethical Review Committee of the Experimental Psychology Department of the University of Oxford, and conforms to the ethical standards in the 1964 Convention of Helsinki.

In order to investigate to what extent temporal regularity might facilitate pattern detection, human subjects were asked to detect repeating noise patterns played over headphones. I generated frozen noise targets and manipulated their regularity by embedding them in non-frozen filler noise of varying length. In this manner, pattern detection was probed in a temporally regular (REP-R) and temporally irregular or jittered (REP-J) context. REP-R stimuli were designed to measure how the detectability of a target depended on its duration relative to a fixed inter-onset interval (IOI). REP-J stimuli were designed to measure how the detectability of a target depended on the variability of IOI. Background false alarm detection rates were measured with a control stimulus of totally non-repeating noise (RAND). As a further control to test the subjects' ability to report changes in the quality of the noise stimuli rapidly and reliably, we also incorporated a fourth stimulus type in the experiment (PINK), in which the spectrum of the noise changed from white (flat amplitude spectrum) to pink (1/f amplitude spectrum). MATLAB was used for stimulus generation, response collection and data analysis.

All stimuli were 8 s in duration, and either remained non-repeating noise throughout (RAND condition), or started with non-repeating noise for a variable (3 to 4 s) duration before transitioning to alternating noise targets and fillers (REP conditions) or to pink noise (PINK condition). The repeating section contained exactly 8 repeats of a single noise target embedded in noise fillers with the duration and jitter

parameter combinations shown in Figure 2.1. REP-R stimuli had target durations T of either 500, 400, 300, or 200 ms, and filler durations F=500-T ms to yield a constant inter-onset interval (IOI) of 500 ms. For the REP-J stimuli, T was fixed at 300 ms, and F was drawn independently from a Gaussian distribution (truncated at zero) with a mean of 200 ms and a standard deviation J of either 10, 50, or 100 ms. Thus, REP-J stimuli had normally-distributed random IOIs with a mean of 500 ms, while REP-R had a fixed IOI of 500 ms, corresponding to a repetition rate of approximately 2 Hz for all REP stimuli. Expressed as percentages, REP-R explored the detection of targets that were 100, 80, 60, or 40% of the IOI, and REP-J probed temporal jitter levels of 2, 10, or 20%, when quantified as the standard deviation of the IOI as a fraction of its mean.

Subjects first underwent an instructional period during which the task was explained and examples of each stimulus type were played until the subjects reported that they could hear the repeating pattern on at least one occasion in both the REP-R and REP-J contexts. Subjects were told that the experiment would consist of four blocks, that stimuli in each block would come one after another with a short silence (approx. 3 s) between stimuli, and that they would be given a break between each 9 min block. Subjects were instructed to press a button as soon as they detected repetition or a transition in the sound.

Each data collection block contained 4 trials of each of the seven REP conditions, 16 RAND trials, and 4 PINK trials, all randomly interleaved. For each subject, over all four blocks, this amounted to 16 trials of each of the seven REP conditions, 64 RAND trials, and 16 trials of PINK. Importantly, the stimulus for each trial was generated from a different random seed and was therefore unique, with its own target, fillers, and set of jittered intervals. This eliminated the possible confound of longer-term memory effects across multiple trials [Agus et al., 2010]. Additionally, in order to reduce the likelihood that any trend found could be explained by the

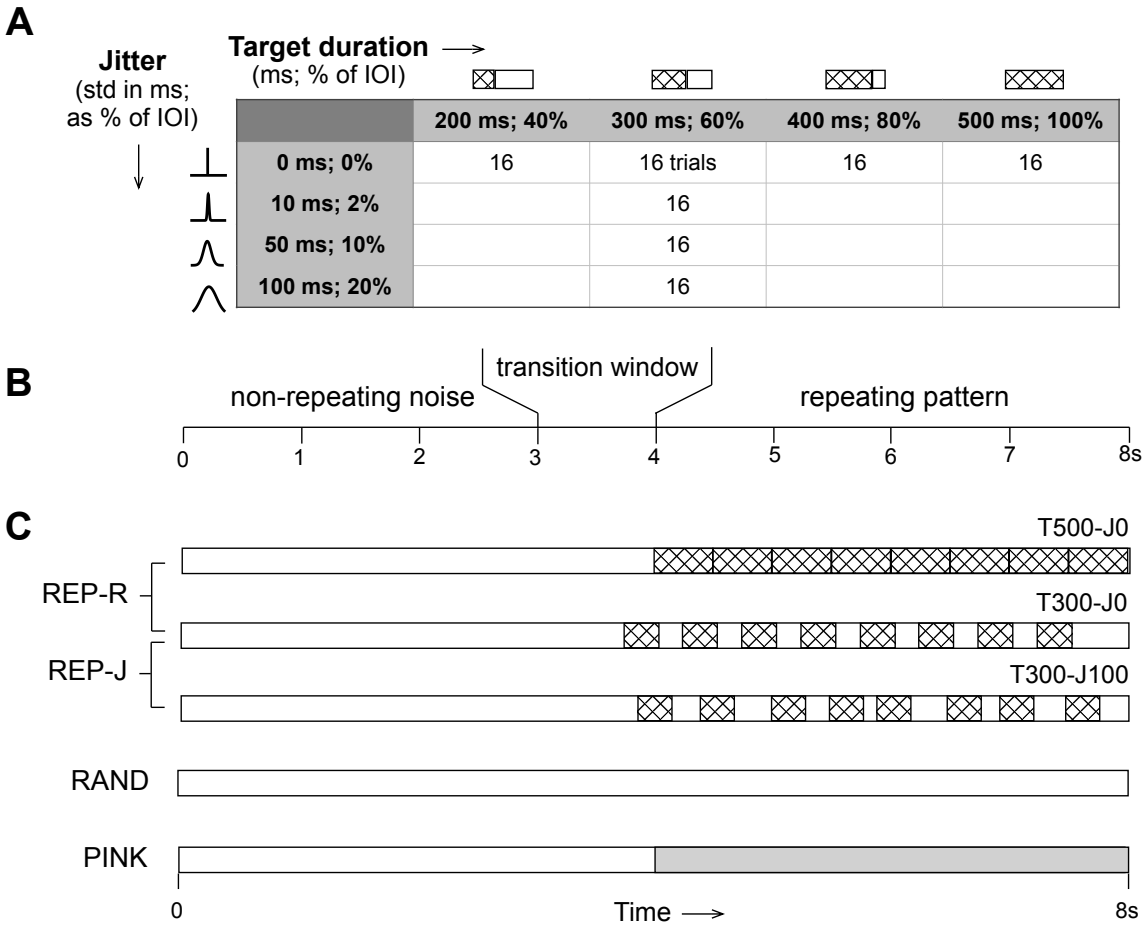


Figure 2.1: **Stimulus design**. A) For each of the parameter combinations shown for the REP conditions, 16 trials were generated. Target duration is displayed in milliseconds and as a percentage of the fixed inter-onset interval (IOI) of 500 ms. Jitter is shown as the standard deviation of the Gaussian distribution from which IOI durations are drawn in milliseconds, and as a percentage of the mean IOI of 500 ms. B) All stimuli were 8s in duration and began as non-repeating noise that then transitioned (if at all) at a random time between 3 and 4 seconds to a repeating section (or pink noise in the case of PINK). C) Schematic representations of the different stimulus conditions where non-repeating noise is shown in white, repeating noise targets are patterned, and pink noise is shown in gray. T500-J0 (target duration of 500 ms, no jitter, top) is closest to the type of stimuli used in previous noise learning studies. A target duration of 300 ms (T300) was used in the jittered (REP-J) context. Control conditions were RAND (non-repeating white noise) and PINK (non-repeating white noise that transitions to pink noise). N=16 trials for PINK and each of the 7 REP conditions, and N=64 trials for RAND.

particular noise stimuli that make up a single stimulus set, a different stimulus set was independently generated for each subject.

Stimuli were played through a TDT RM1 mobile processor (Tucker Davis Technologies, Alachue, FL, USA), and presented diotically at 50 dB SPL over Sennheiser HD 650 headphones (Wedemark, Germany). The TDT device delivered the stimuli and recorded button presses, allowing precise reaction times to be measured. Experiments were conducted in a double-walled sound-proof chamber.

Two performance measures were analyzed for all conditions tested: fraction detected and reaction time. The fraction detected is the proportion of trials (out of a total of 16 in each REP or PINK condition) during which the subject pressed the button to indicate detection. Reaction time was measured from the onset of the first target noise to the time of the button press. By dividing reaction time by 0.5 s, one can determine approximately how many noise targets had been presented before detection. Miss trials where repetition was present but not detected were excluded from reaction time calculation.

2.3 Results

Twenty-one paid participants aged 20 to 40 with normal hearing were recruited for this study. Three subjects were authors on this study, and 11 had some musical training. To ensure that subjects were performing the task correctly, subjects with a false alarm rate greater than 50%, calculated as the percentage of all RAND stimuli during which an erroneous detection was reported, were excluded from further analysis, leaving a final total of 17 subjects.

The population false alarm rate, calculated as the proportion of erroneous detections during RAND trials (out of a total of 64) averaged across all 17 subjects, was 17.9%. The average reaction time for PINK trials, was 588 ms.

2.3.1 Shorter noise targets are harder to detect

Figures 2.2A and 2.2C show REP-R detection performance and reaction times, respectively. A few subjects were near 100% detection for all REP-R stimuli, but the overall trend is for detection performance to increase with increasing target duration (Fig. 2.2A), and for reaction time to decrease (Fig. 2.2C). Relative to the shortest target duration (T200-J0: 200 ms target duration, 0 ms jitter), detection performance was significantly higher and reaction times significantly lower ($p < 0.001$ and $p < 0.01$ respectively, $n = 17$ subjects, Wilcoxon signed rank test, Holm-Bonferroni corrected) for all other target durations tested. A significant drop in reaction time was also observed between T300-J0 and T500-J0 ($p < 0.05$, $n = 17$, Wilcoxon signed rank test, Holm-Bonferroni corrected). Thus, as the duration of repeating targets makes up a larger proportion of the IOI, their repetition is more likely to be detected, and fewer target presentations are required for their detection.

2.3.2 Substantial jitter impairs noise target detection.

Figure 2.2B and 2.2D show individual detection rates and reaction times, respectively, for REP-J conditions. Subjects could detect the repeating pattern quite well for all jitter levels, but there are some systematic effects of jitter. We observed no statistically significant differences in reaction time between REP-J conditions ($p > 0.05$, $n = 17$, Wilcoxon signed rank test, Holm-Bonferroni corrected), but detection performance on T300-J100, the most jittered condition, was significantly worse than on T300-J0 and T300-J10 ($p < 0.05$, $n = 17$, Wilcoxon signed rank test, Holm-Bonferroni corrected). Thus, a substantial amount of jitter makes a repeating target more difficult to detect, but modest amounts of jitter appear to be well tolerated.

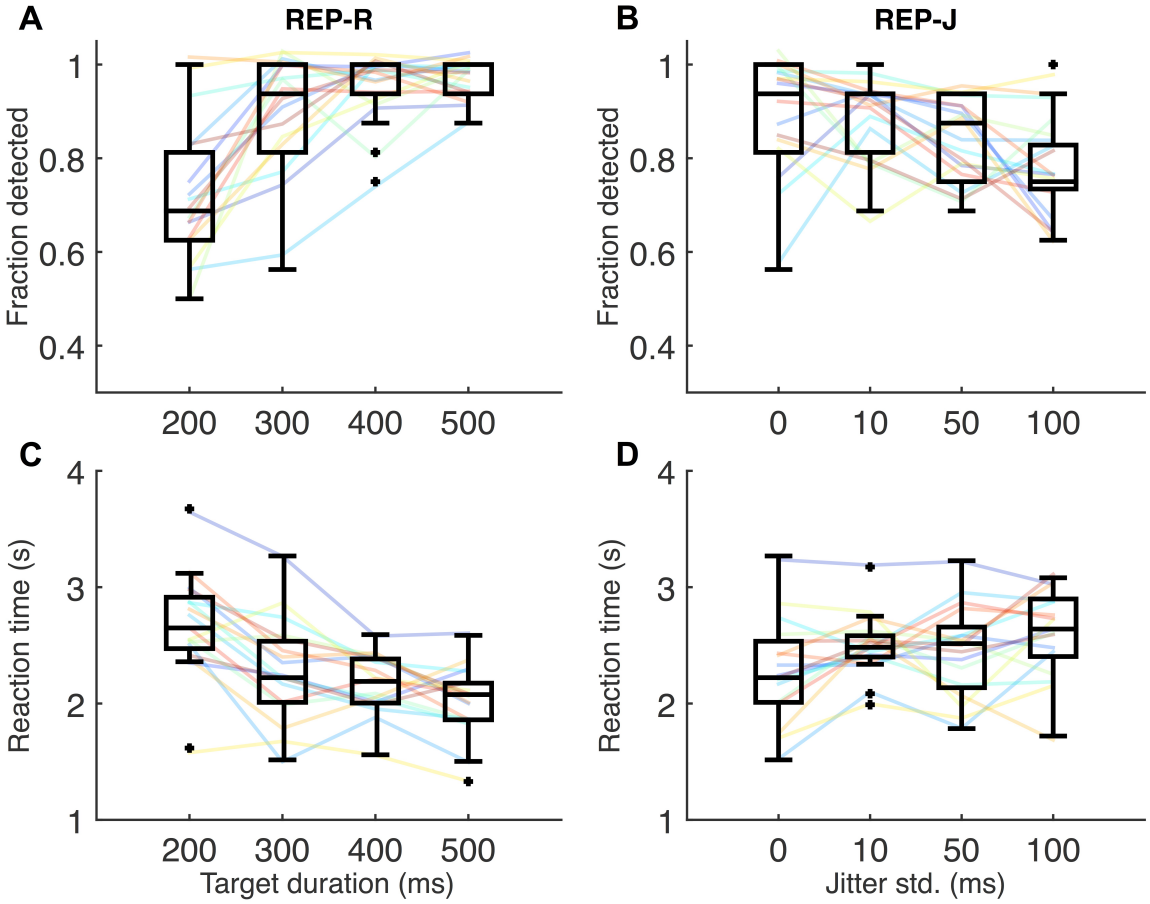


Figure 2.2: **Detection performance under temporally regular and jittered conditions**. A) Fraction detected for REP-R conditions. Solid lines represent individual subjects, and boxplots summarise the data over the 17 subjects for a given stimulus condition. For display purposes only, a small amount of Gaussian noise (std=1.5%) has been added to the y-values of the solid lines in panels A-D to spread out overlapping data points. B) Fraction detected for REP-J conditions. C) Mean reaction times for REP-R conditions. D) Mean reaction times for REP-J.

2.3.3 What about natural sounds?

This study of pattern-in-noise detection was motivated by the ecological need to detect rhythmic sounds in the natural world, such as detecting footsteps out of background noise. These experimental results so far indicate that pattern detection benefits from temporal regularity, and it would provide some context for these findings to explore how much temporal jitter is present in natural sounds. I therefore analyzed step interval data from normally walking healthy humans, compiled from three separate studies [Frenkel-Toledo et al., 2005, Hausdorff et al., 2007, Yogev et al., 2005], available on the PhysioNet database [Goldberger et al., 2000]. The dataset logged pressure sensor data recorded from underneath both feet, and I defined foot strikes as each time pressure under either foot crossed a threshold. Footstep intervals were calculated as the time between successive foot strikes. Since participants were pacing back and forth through a hallway, the need to turn around introduced some footstep intervals that were clear outliers from an otherwise tight distribution. Hence, as an outlier-proof measure of the jitter in step intervals, I calculated the median percentage deviation from the median step interval for each individual. These median deviations ranged from 1.6 to 6.6% across the 72 subjects, with a median of 1.9%. The entire range is less than the median deviation of the intermediate jitter condition (T300-J50) in this study of 6.7%. From this we can conclude that, at least for this class of rhythmic natural sounds, the amount of temporal jitter present would be too small to impair detection performance.

2.3.4 Adequate sampling in stimulus set is crucial.

In an initial run of this study, a single stimulus set was generated and then tested on all subjects. At first we were intrigued by a seemingly counterintuitive finding that a small amount of jitter (2% jitter condition) appeared to improve pattern detection

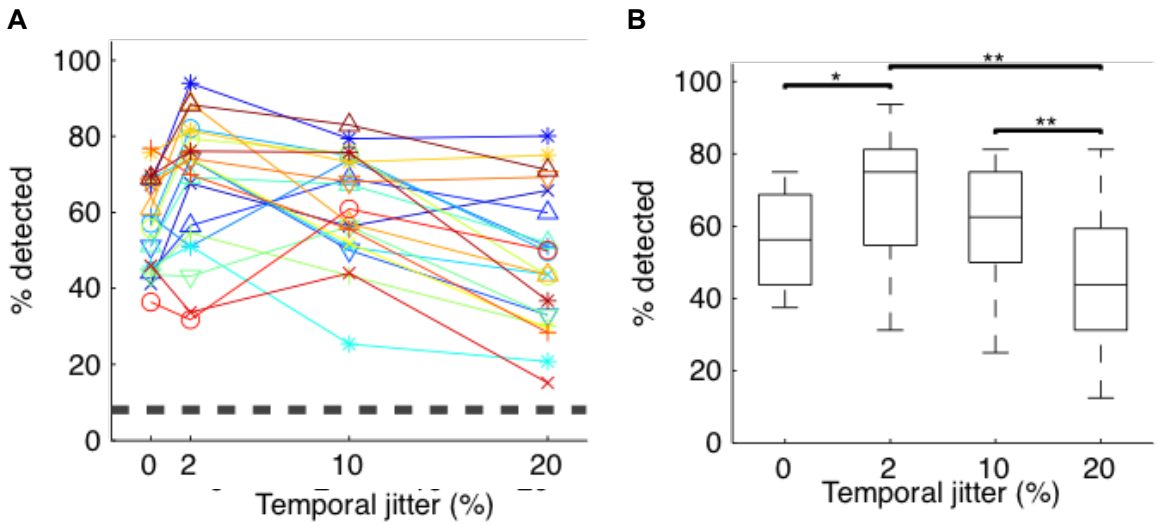


Figure 2.3: **Inadequate sampling can lead to incorrect conclusions.**. A) Percent detected for jittered conditions, where each colour is a different subject. These data are based on an initial run of this experiment (with a different set of participants) where stimuli were not generated afresh for each subject. B) For this limited stimulus set, detection performance was best in the 2% jitter condition (Wilcoxon paired signed rank test, $* \; p < 0.05$, $** \; p < 0.01$).

relative to perfectly isochronous presentation of noise patterns (see Figure 2.3). However, when I compared the stimulus set used against a larger sample of stimuli from each condition, we quickly noticed that the tested stimulus set, by chance, was not as representative of the underlying distribution as we had assumed. This was a clear mistake that meant that the study needed to be re-run with a larger stimulus sample, which did by generating a unique set of stimuli for each subject. It is therefore of utmost importance, especially in cases where stimuli are meant to represent some underlying distribution, that a sufficient number of stimuli are sampled. Not checking for this can lead to incorrect conclusions even if the analyses are sound.

2.4 Discussion

In this study, I discovered, firstly, that for a fixed inter-onset interval (IOI), a repeating target noise becomes easier to detect as its duration increases, and secondly, that

detection performance declines with substantial amounts of temporal jitter (more than the amount of jitter found in footsteps).

2.4.1 Does repetition detection rely on synaptic memory traces of recent inputs?

Agus, et al. (2010) suggested that memory traces that are presumably needed for repetition detection may involve a spike-timing dependent plasticity (STDP). Networks incorporating STDP have been shown to quickly learn to detect a repeating pattern of afferent spiking activity amidst otherwise stochastic firing [Masquelier et al., 2008]. This makes STDP an appealing candidate mechanism consistent with experimental observations made to date, with two possible caveats. First, subjects were able to recognise repetition with only two presentations of a frozen noise target [Agus et al., 2010], a performance that is so far unmatched by existing models of STDP. Secondly, a purely STDP based model would accurately detect noise targets equally well whether they arrive at regular intervals or not, which is in contrast to our finding that temporal regularity results in better detection performance. This does not rule out that STDP may have a role to play, but it does suggest that it alone does not account for all aspects of noise learning.

2.4.2 Modulation filterbanks as an alternative mechanism.

A mechanism that would potentially account for the timing aspect of our findings is a modulation filterbank, which is a set of neural filters tuned to different frequencies of modulation of the sound envelope (typically within a frequency band). Modulation filterbank models of the auditory system have shown good agreement with human psychoacoustic data on amplitude modulation detection [Dau et al., 1997a] and speech intelligibility [Jorgensen and Dau, 2011], and electrophysiological evidence for

modulation tuning exists at various levels of the auditory system [Joris et al., 2004, Schreiner and Urbas, 1986, Schreiner and Urbas, 1988, Kilgard and Merzenich, 1999]. It is possible that the brain relies at least in part on modulation filters to detect repetitions in noise, and that the performance decrease observed in the presence of jitter might be explained by the fact that jittered stimuli will drive modulation filters less strongly. I illustrate the plausibility of this idea through the following analysis.

2.4.3 Repetition of frozen noise targets results in distinct peaks in the modulation spectrum.

The modulation spectrum was calculated for each stimulus and the standard deviation of the modulation spectrum was used as a measure of its peakiness. A peaky modulation spectrum would indicate that some modulation filters are being driven more strongly than others, and the question is whether this correlated with detection performance using the method illustrated in Figure 2.4A. The first step was to transform each sound stimulus into a simple approximation of the activity pattern received by the auditory nerve by calculating a sounds log-scaled spectrogram ('cochleagram'). For each sound, the power spectrogram was taken using 20 ms Hanning windows, overlapping by 10 ms. The power across neighboring Fourier frequency components was aggregated using overlapping triangular windows comprising 43 frequency channels with center frequencies ranging from 150 Hz to 19,200 Hz (1/6 octave spacing). Then, the log was taken of the power in each time-frequency bin, and finally any values below a low threshold were set to that threshold. These calculations were performed using code adapted from melbank.m.

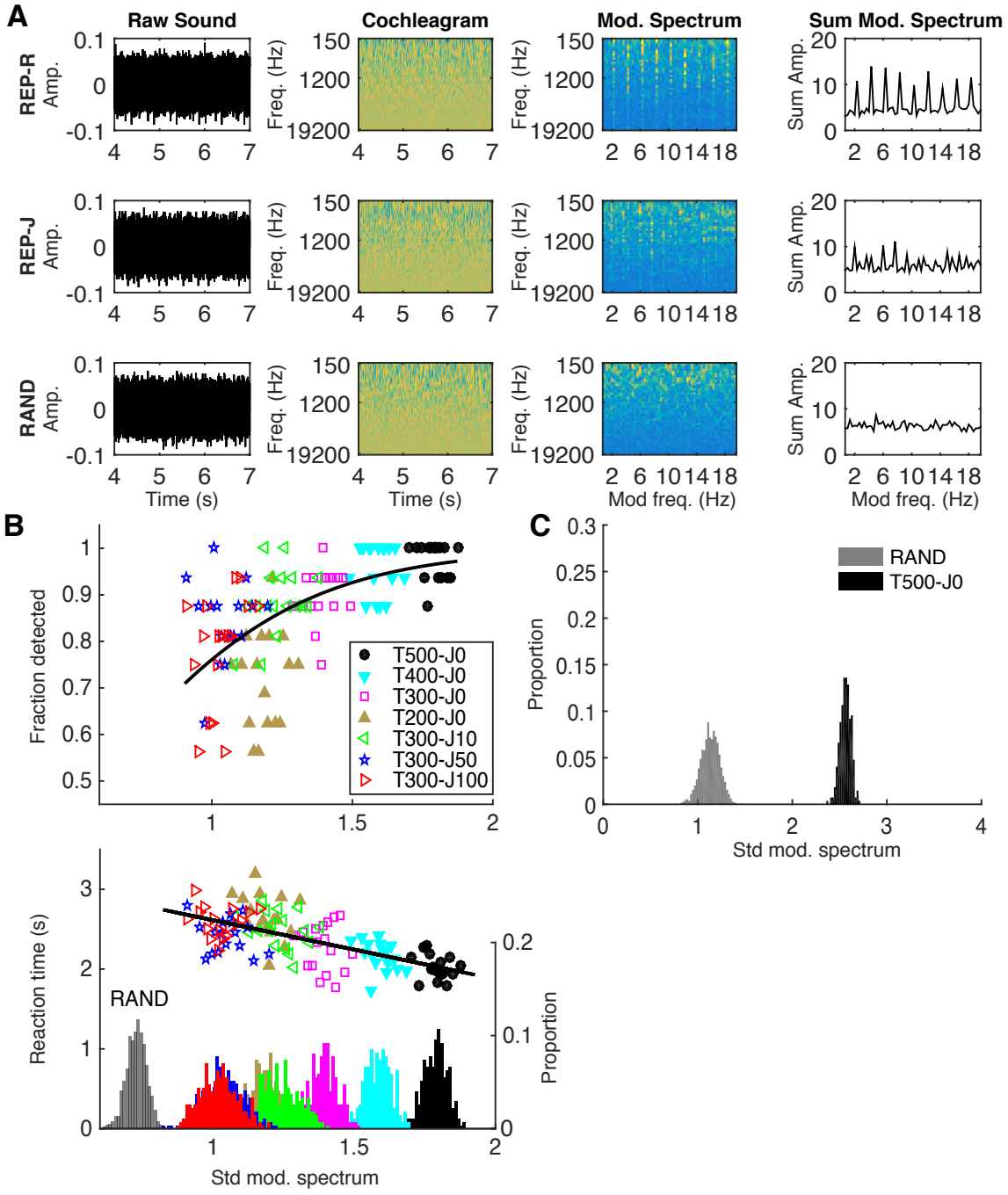

A
Raw Sound
Cochleagram
Mod. Spectrum
Sum Mod. Spectrum
REP-R
REP-J
RAND
Amp.
Freq. (Hz)
Sum Amp.
Time (s)
Mod freq. (Hz)
B
Fraction detected
T500-J0
T400-J0
T300-J0
T200-J0
T300-J10
T300-J50
T300-J100
Reaction time (s)
RAND
Proportion
Std mod. spectrum
C
RAND
T500-J0
Proportion
Std mod. spectrum

Figure 2.4: **A modulation filterbank model would account for the effect of temporal jitter on detection performance**. A) Method used to quantify the peakiness of the modulation spectrum, shown for one example from REP-R (T300-J0), REP-J (T300-J100), and RAND. From left to right: the raw sound; cochleagram; modulation spectrum in each frequency channel; modulation spectrum summed over frequency channels. Note the vertical stripes at 2 Hz and its harmonics present in both REP stimuli, and their absence in RAND. The standard deviation of the summed modulation spectrum was calculated for each stimulus trial. B) Top: Fraction detected plotted against the standard deviation of the summed modulation spectrum. Each point represents the fraction detected, out of the nearest-neighboring 16 stimulus examples (regardless of subject) along the x-axis from within the same condition. Different colors represent the different conditions, as shown in the legend. Black line is the fit from a logistic regression. Bottom: Reaction time in seconds plotted against the standard deviation of the summed modulation spectrum. For display purposes, the average (clumped) reaction times of each 16 nearest neighbors along the x-axis from the same stimulus condition are plotted. The black line is a linear regression run on all (un-clumped) data. On the secondary y-axis is a histogram showing the distribution of all (un-clumped) standard deviation values for all stimuli within each condition. C) A histogram showing the distribution of standard deviation values calculated over 1 s intervals during RAND (black) and T500-J0 (blue), analogous to the stimuli used in Agus, et al. 2010. Note the larger standard deviation values for both conditions in panel C using a 1 s window compared to the 3 s window used in panel B. For our examples of RAND ($n = 1088$) and T500-J0 ($n = 272$), we see no overlap, suggesting that the modulation spectrum would contain enough information to detect repetition from a single repeat of a 500 ms frozen noise target.

The cochleagram was calculated over a 3 s window starting 4 s into the sound, by which time frozen noise targets must have ensued in all REP conditions. The magnitude spectrum of the activity in each frequency channel was calculated and then summed across frequency channels to get the overall modulation spectrum of the sound. We then calculated the standard deviation of the modulation spectrum (up to 20 Hz) to quantify how much it deviated from a flat modulation spectrum. Gaussian white noise without repeating frozen noise targets (our RAND condition) should have a flat modulation spectrum and small standard deviation, while isochronously presented targets (our REP-R conditions) should introduce significant peaks in the modulation spectrum, increasing its standard deviation (Fig. 2.4A, rightmost column). As illustrated in Figure 2.4B, REP stimuli with peakier modulation spectra

were more reliably and more quickly detected by our subjects. Peakiness of the modulation spectrum correlates significantly with detection ($p = 0.03$, $n = 119$, 7 conditions by 17 subjects, logistic regression), and with faster reaction times ($p < 10^{-20}$, $n = 1649$, Pearson correlation).

2.4.4 Could a modulation filterbank model account for previous findings?

As mentioned earlier, Agus et al. (2010) demonstrated that human listeners could detect repetition in a 1 s long stimulus where a 500 ms noise token was played only twice. A question worth asking is whether peakiness in the modulation spectrum could still be helpful even when there are so few repeats. Figure 2.4C shows that the standard deviation of the modulation spectrum calculated from 1 s (two period) segments taken from our analogous T500-J0 stimuli do indeed differ substantially from equivalent 1 s segments from non-repeating RAND stimuli, suggesting that peaks in the modulation frequency domain could have provided a useful cue in the task by Agus, et al. (2010). However, modulation filters alone would not account for their observation that noise memory traces could have surprisingly long lasting effects. Thus, both modulation filter-like mechanisms and long term plasticity are likely to be required to fully account for our ability to detect patterns in noise.

Further work is needed to confirm whether modulation filters indeed underlie the results reported here, as well as in other related psychoacoustic studies. For example, timing predictability is also an important cue during auditory scene analysis [Bendixen, 2014], and different patterns of activity across frequency channels in the modulation frequency domain could be involved in the tendency for temporally jitter to cause streams to segregate [Andreou et al., 2011, Rajendran et al., 2013]. An additional consideration is the evidence for the oscillatory nature of temporal attention and its effect on task performance, which has been studied both in the visual

[Correa et al., 2006, Lakatos et al., 2008] and auditory [Henry and Obleser, 2012, Lakatos, 2005, Jaramillo and Zador, 2011, Lakatos et al., 2013, Jones et al., 2002, Lawrance et al., 2014] domain.

2.5 Conclusions

The results in this chapter demonstrate that the ability to detect a repeating pattern is affected by the regularity of timing with which repeated sounds are presented. Specifically, at a presentation rate of 2 Hz, applying a temporal jitter of 20% to the onsets of the noise targets significantly hindered their detection. In light of these findings, an analysis of the amount of jitter present in footsteps would suggest that it is likely too small to be detrimental to detection. Finally, I demonstrated that aspects of detection performance in this study and in other noise pattern detection studies can be well accounted for by the hypothesis that the auditory system uses low frequency modulation filters to detect rhythmic patterns. All together, we conclude that temporal regularity aids in detecting subtle structure in sound.

Chapter 3

Midbrain Adaptation May Set the Stage for the Perception of Musical Beat

Most of the work contained in this chapter has been published in:

Rajendran, V. G., Harper, N. S., Garcia-Lazaro, J. A., Lesica, N. A., & Schnupp, J. W. (2017). Midbrain adaptation may set the stage for the perception of musical beat. In Proc. R. Soc. B, Vol. 284, No. 1866. The Royal Society.

Study design and data analysis were designed together with input from NSH and JWHS, who were also involved in revising the manuscript. JAGL performed the electrophysiological recordings. The data analysis and writing are my own.

Abstract

The ability to spontaneously feel a beat in music is a phenomenon widely believed to be unique to humans. Though beat perception involves the coordinated engagement of sensory, motor, and cognitive processes in humans, the contribution of low-level auditory processing to the activation of these networks in a beat-specific manner is poorly understood. Here, I present evidence from a rodent model that midbrain pre-processing of sounds may already be shaping where the beat is ultimately felt. For the tested set of musical rhythms, on-beat sounds on average evoked higher firing rates than off-beat sounds, and this difference was a defining feature of the set of beat interpretations most commonly perceived by human listeners over others. Basic firing rate adaptation provided a sufficient explanation for these results. These findings suggest that midbrain adaptation, by encoding the temporal context of sounds, creates points of neural emphasis that may influence the perceptual emergence of a beat.

3.1 Introduction

When listening to a rhythmic sound such as music, one can often find and tap along with a steady (isochronous) beat. In principle, many possible interpretations of beat could exist, but in practice, only very few of them tend to be chosen by listeners. What could be the neurophysiological determinant for where in a rhythmic sound the beat is felt?

At the highest level, human studies have revealed beat-specific entrainment of cortical oscillations [Fujioka et al., 2009, Iversen et al., 2009, Fujioka et al., 2015, Snyder and Large, 2 Zanto et al., 2006]. A promising candidate mechanism for the entrainment of these oscillations may be the cortico-basal ganglia-thalamocortical loop [Parent and Hazrati, 1995], of which the basal ganglia are thought to be particularly important for beat perception [Grahn, 2009, Geiser et al., 2012, Bartolo and Merchant, 2015] and time perception [Meck et al., 2008, Teki et al., 2012]. If the dynamics of these circuits serve to select an interpretation of beat out of the many theoretically possible ones, then beat-relevant precursors may already be present as a result of low-level auditory processing. However, the influence of low-level auditory processing on beat perception has not yet been studied. We hypothesised that adaptive processes in the brainstem, which are common across mammalian species [Dean et al., 2005, Dean et al., 2008, Ingham and McAlpine, 2004, Willmore et al., 2016, Bulkin and Groh, 2011] regardless of whether they can synchronise their movements to a rhythmic stimulus [Geiser et al., 2014], may strongly influence which beat interpretations are chosen over others. We tested this hypothesis by investigating the correspondence between brainstem processing in a rodent model and beat perception in humans.

Seven rhythmic patterns [Nozaradan et al., 2012] constructed from identical broadband noise bursts were played to anesthetised gerbils while recording from the central nucleus of the inferior colliculus, the major midbrain relay through which ascending

information from the auditory nerve passes on its way to the cortex. The same rhythms were also played to human listeners who simply tapped along to the beat they perceived. While the influential study by Nozaradan, et al. (2012) provided the inspiration for this work, the frequency domain method used in that study does not yield a dependable measure of beat entrainment (see Figure 3.8). Here, I present a novel time domain analysis that avoids the pitfalls and ambiguities in the interpretation of frequency domain-based analyses (see also [Zhou et al., 2016]). On average, neural activity recorded from the midbrain was stronger on the beat than off the beat, and this asymmetry was a defining feature of the beat interpretations chosen out of all possible ones. Furthermore, adaptation provided a parsimonious account for these results. Together, these findings strongly suggest that midbrain adaptation may already be shaping where the beat is ultimately perceived in musical rhythms.

3.2 Results

3.2.1 On-beat sounds evoke stronger neural responses on average than off-beat sounds

The underlying hypothesis is that low-level stimulus processing might already create a neural emphasis that accompanies the perceptual emphasis felt on the beat. If this is the case, then it could be observable as higher firing rates on the beat than off the beat already at the level of the the IC under anaesthesia.

Neural activity was recorded extracellularly from 29 single units and 220 multi-units originating from 149 recording sites in the central nucleus of the inferior colliculus (IC) of 4 gerbils in response to 7 rhythmic sound patterns. The chosen rhythmic sound patterns were adapted from a previous human electroencephalography (EEG) study [Nozaradan et al., 2012] and are described in detail in the Methods. Each sound

pattern consisted of a repeated sequence of 12 or 16 equal duration events separated by brief silent gaps, where events were either silent intervals or identical broadband noise bursts. Beat locations for each rhythm were determined based on the most common tapping pattern across 13 human listeners (Figure 3.1B; see Figure 3.2 for tapping data for all stimuli). Noise bursts were classified as either being on-beat if taps were aimed at their onset, or off-beat otherwise. The mean firing rate during the first 40 ms of all on-beat and off-beat noise bursts was calculated for each unit (Figure 3.1C). A time window of 40 ms was chosen because it was the longest time window that allowed comparison across all seven stimuli since it corresponded to the duration of the noise burst in the fastest pattern. Figure 3.1D illustrates each unit's average on-beat and off-beat response across all sound events and all rhythmic patterns (see Figure 3.3 for on-beat versus off-beat responses for each stimulus individually). On average, neural responses were significantly stronger to on-beat sounds than to off-beat sounds, despite all sound events being acoustically identical noise bursts (Figure 3.1D, $p < 10^{-6}$, Wilcoxon paired signed rank test, $n = 248$ units).

3.2.2 A large on-beat neural emphasis is a defining feature of the perceived beat

Next, I explored whether a large neural emphasis on the beat might explain why some beat structures were more commonly perceived than other possible ones. The hypothesis here is that sound events that evoke a particularly strong neural response are more likely to be perceived as being on-beat, which would support the possibility that systematic differences in evoked response strength at low levels of the auditory pathway might predetermine both the grouping and phase of the beat interpretation that is ultimately selected at higher levels.

A beat structure as I define it here consists of a beat period and a beat position. For a given rhythmic pattern, the beat period is determined by the integer number

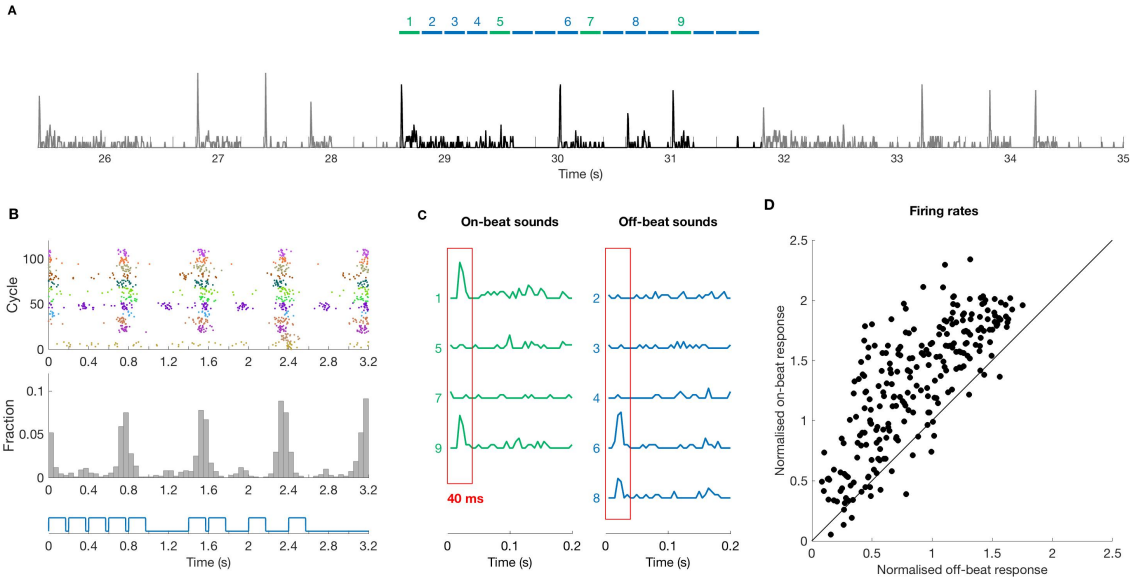


Figure 3.1: **On-beat responses are on average stronger than off-beat responses.** A) An example units normalised peri-stimulus time histogram (PSTH) to stimulus pattern P2 at 1x tempo. The last three complete cycles of this pattern are shown together with the stimulus trace above. The coloured numbers indicate noise bursts, green being on-beat and blue being off-beat. Normalisation was done by dividing each unit's firing rate concatenated across all 7 stimuli by its standard deviation. B) Raster-style plot showing tap times of human listeners across trials and cycles of this stimulus. Each dot marks the timing of a tap, each row of dots shows tap responses for a single cycle of the stimulus pattern, and different colour dots distinguish the tap responses from different subjects. Beneath is a histogram of tap times pooled across subjects, cycles, trials for this stimulus. A clear, regular tapping pattern is present, which indicates a strong consensus among listeners to hear four beats in this 16-event sequence, timed on the 1st, 5th, 9th and 13th interval. C) Normalised firing rate in response to on-beat (left, in green) and off-beat (right, in blue) noise bursts for the highlighted cycle shown in panel A. To quantify a unit's mean on- and off-beat response, the firing rate over the first 40 ms of each sound-evoked response to on-beat and off-beat sounds (red boxes) across all 33 s of all stimuli was averaged (the average based on just the highlighted cycle are shown for demonstration purposes), and the mean of the result was taken. D) Scatter plot showing the mean on-beat response for each unit as a function of that unit's mean off-beat response. Most points fall well above the main diagonal, indicating that on average, on-beat sounds evoked stronger responses than off-beat sounds.

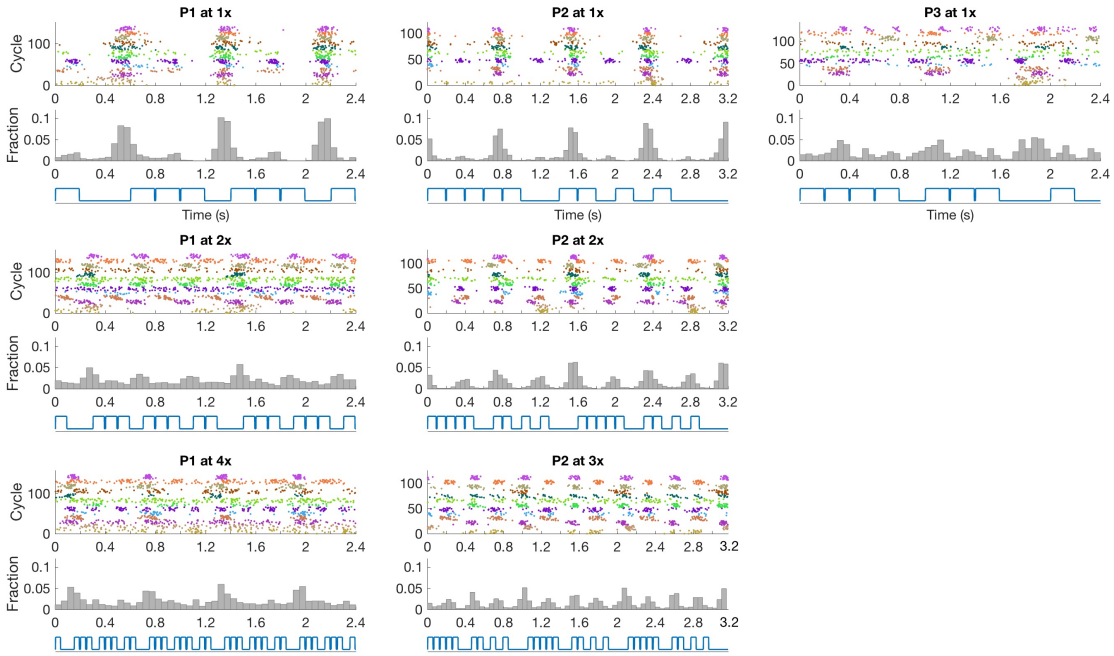


Figure 3.2: **Raster-style plot of tap times for each stimulus.** Each dot marks the timing of a tap, each row of dots shows tap responses for a single cycle of the stimulus pattern, and different colour dots distinguish the tap responses from different subjects. Below (in grey) is a histogram of tap times pooled across subjects, cycles, trials for each stimulus. At the bottom of each panel is the stimulus pattern.

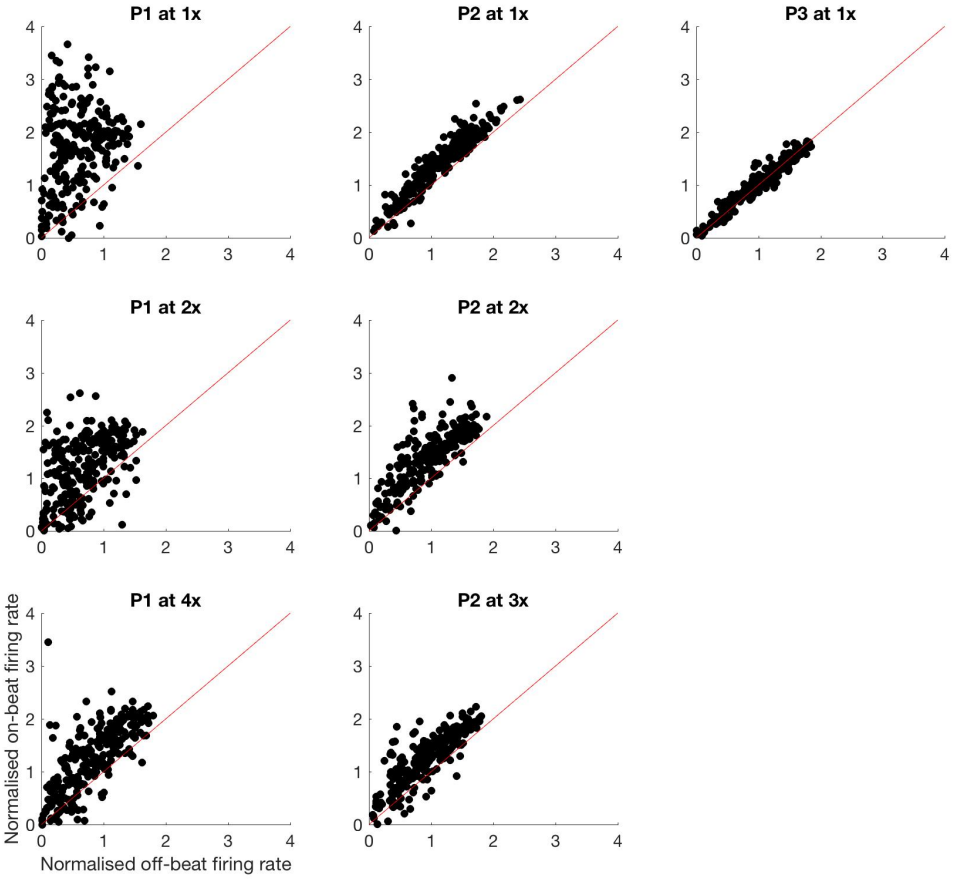


Figure 3.3: **On-beat firing rate versus off-beat firing rate for each stimulus.** All stimulus patterns except P3 show on-beat responses that are significantly larger than off-beat responses ($p < 0.0001$, Wilcoxon paired signed rank test, Bonferroni corrected, $N = 248$ units).

of events (noise bursts or silent intervals) that listeners grouped perceptually while tapping. The beat position refers the temporal frame or phase of the tap intervals, and is set by the time points that the listeners report as on-beat with their taps. Beat periods were considered hypothetically possible if they divided up the rhythmic pattern into an integer number of equal length intervals. Beat periods that were not an integer fraction of the number of events in the pattern would have put the beat on different places in successive repetitions of the pattern, and none of my subjects exhibited such a beat precession. Thus, for rhythms P1 and P3, both 12-event patterns, possible beat periods consisted of tapping a beat once every 2, 3, 4, 6, or 12 events, while for the 16 event patterns, possible beat periods would be 2, 4, 8 or 16 events long. Note that I do not consider a beat period of 1, where all events would be on-beat with none off the beat. Possible beat phases refer to the N possible positions at which the beat could start for a beat period that contains N events. 12-event patterns therefore had 2+3+4+6+12 = 27 possible beat structures, while the 16-event pattern 2+4+8+16 = 30 possible beat structures.

Population neural activity was calculated as the average firing rate across all single and multi-units. The mean event-evoked population firing rate during the first 40 ms of each event (noise burst or silence) was classified as being on-beat or off-beat, where on-beat and off-beat were defined by each beat period and beat position combination for each possible beat structure. For comparison, the average sound content at on-beat and off-beat positions was calculated for each beat structure, counting a sound event as 1 and a silent interval as 0. The on-beat emphasis for sound content and population neural activity was calculated as the difference between its mean on-beat and off-beat values (*on-off*).

Figure 3.4A and 3.4B show the on-off values for all possible beat structures in the 7 stimulus patterns tested. Yellow and cyan boxes mark the most and second most commonly perceived beat structures, respectively. All *on-off* values for sound and

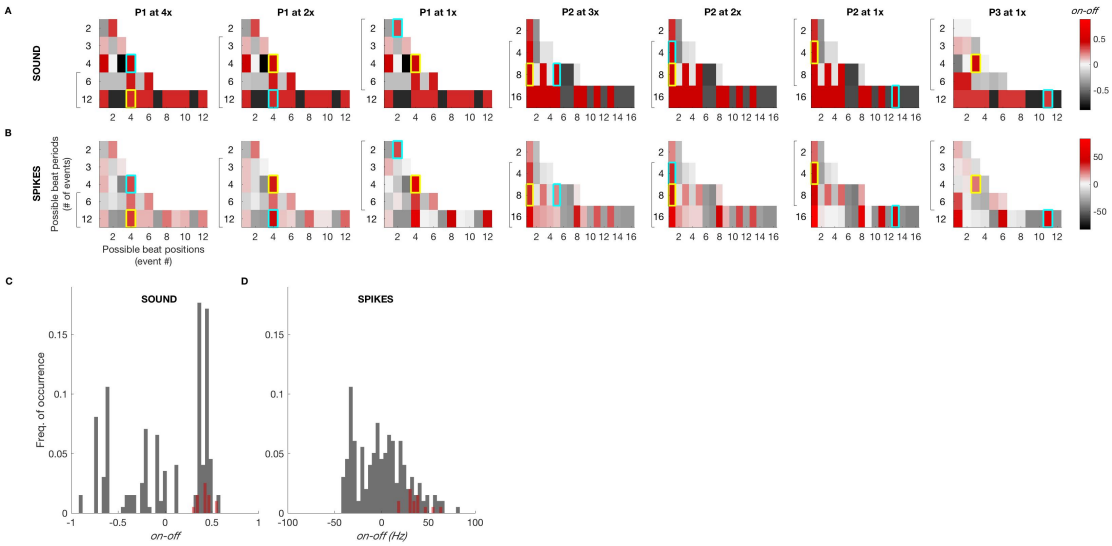


Figure 3.4: **Exploration of on-off for all possible combinations of beat periods beat positions.** A) Heatmaps represent on-off for all possible combinations of beat grouping (y-axis) and beat position (x-axis) for each of the 7 rhythmic sound patterns tested. Red Yellow and white cyan boxes mark the most and second most common beat interpretations reported by human listeners, respectively. Brackets indicate those groupings whose theoretical tapping frequencies are between 0.5-4 Hz, the range within which beat is typically perceived [London, 2012b]. B) Same as panel A, but based on population firing rates. C) All hypothetical on-off values for sound (grey) and the two most commonly perceived beat structures (red), pooled across the 7 stimuli. D) Same as panel C, but for population firing rates.

population neural activity, pooled across the 7 stimuli, are shown in the histograms in Figures 3.4C and 3.4D, respectively. The beat structures preferred by our listeners, shown in red, have significantly larger *on-off* values than the pool of all possible beat structures, shown in grey, for both sound and neural activity (firing rates: $p < 10^{-6}$; sound: $p < 10^{-4}$; Wilcoxon rank sum test, $N = 198$ possible beat structures and 14 perceived beat structures). It follows from a preference for beat structures with high on-beat sound content that on-beat positions would evoke relatively large on-beat neural responses. However, a consequence of neural processing is that relatively fewer possible beat structures overall resulted in high *on-off* values (Figure 3.4C-D). Quantitatively, the distribution over all *on-off* values for firing rates (Figure 3.4D) showed higher skewness, or a longer right tail, than the distribution of *on-off* for

sound content (Figure 3.4C; $p < 0.01$, paired sign test, $N = 7$ stimulus patterns).

Taken together, these findings would suggest that, out of the pool of all possible beat structures, those that are perceived by listeners tend to be those with a relatively large on-beat neural emphasis and relatively high underlying on-beat sound content, and that midbrain processing of sound further restricts the set of possible beat structures that later stages of the nervous system may select. Though commonly chosen beat structures show high *on-off* values for both sound content and population neural activity, it is the neural activity that better distinguishes from all candidate beat structures those that were actually perceived by listeners.

3.2.3 Strength of on-beat neural emphasis varies systematically with the profile of each recorded IC unit's firing pattern

A key advantage of extracellular recordings over noninvasive imaging techniques is the temporal and spatial resolution to observe response dynamics of single cells (single units) and small groups of cells (multi-units). Response patterns in the IC are known to be highly diverse, and to investigate the contribution of neurons with different response properties on any on-beat emphasis at the neural population level, I performed hierarchical clustering on normalised single and multi-unit firing rates in response to all seven stimulus patterns concatenated across time (see *Materials and Methods*). This resulted in 8 clusters containing our 249 single and multi-units (Figure 3.5B), of which one cluster was excluded since it contained only one unit. Previous studies in IC [Malmierca et al., 2008, Schnupp et al., 2015] suggest that single unit and multi-unit responses are comparable, and this would appear to be the case for our data too (Figure 3.5C). The seven clusters identified are not meant to provide a definitive or exhaustive categorisation of the response types that exist within IC, but simply

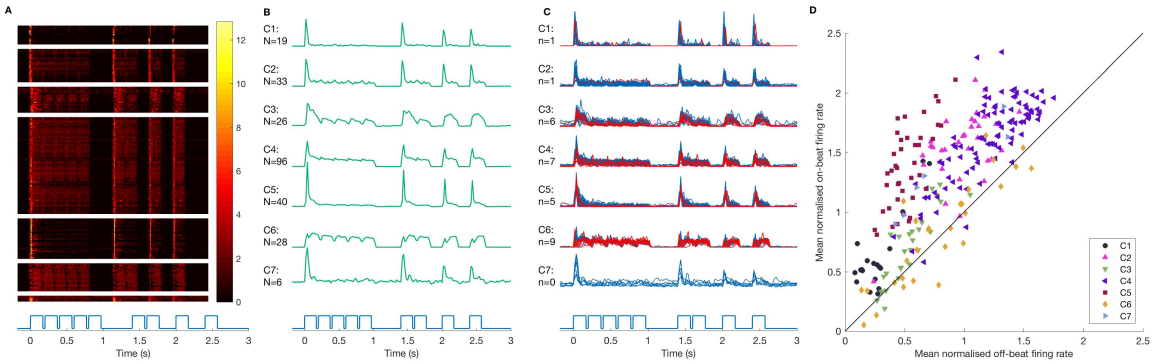


Figure 3.5: **Clustering analysis.**A) Heat map illustrating each units normalised firing rate in response to one cycle of pattern P2 at 1x tempo. Units are grouped by cluster. The stimulus pattern is plotted beneath each panel for reference. B) Each clusters mean peri-stimulus time histogram (PSTH), with the total number of units in each cluster (N) displayed to the left of each trace. C) Mean PSTH traces of multi-units (blue) and single units (red) in each cluster. The number of single units (n) is displayed to the left. D) Same as Figure 3.1D, but with units labelled by the cluster to which they belong.

provide a principled set of response profiles that are representative of this dataset and show characteristics of onset, on-sustained, and sustained firing patterns previously identified in the IC [Neuert et al., 2001]. Onset (C1, C5, C7), on-sustained (C2-C4), and sustained (C6) type units roughly account for 26%, 63%, and 11% of our sample, respectively.

To test whether the on-beat neural emphasis varied by response shape, the normalised on-beat and off-beat sound-evoked responses were re-examined from Figure 3.1D. Labelling units by cluster reveals that onset type units (C1, C5, C7) appear furthest from the diagonal and therefore shows the strongest emphasis to on-beat sounds (Figure 3.5D). Sustained type units (C6) appear near the diagonal, and on-sustained type units (C2-C4) fall somewhere in between.

3.2.4 Adaptation may explain why neural responses are stronger on the beat

A simple explanation for the evident beat processing occurring in the midbrain could be firing rate adaptation, which is the tendency for neural firing rates to decrease to a stimulus if it is prolonged or repeated at high rates. In the context of the tested patterns, adaptation would result in relatively weaker responses to sounds that are directly preceded by other sounds in the recent past, and in relatively stronger responses to sounds that are directly preceded by long periods of silences. This pattern is observable qualitatively (see Figure 3.5), so can firing rate adaptation explain what has been quantified so far as the contribution of midbrain activity towards beat processing?

Adaptation was quantified for each unit by fitting an exponential function to its sound-evoked firing rate as a function of the amount of silence immediately preceding the sounds. Silent intervals ranged from the 10 ms between consecutive noise bursts to the 3 s that separated each 33 s rhythmic pattern from the next. Data and exponential fits from an example unit from each cluster are shown in Figure 3.6A. When the population firing rate was calculated using each unit's exponential fit rather than the real data, the resulting *on-off* values for all candidate beat structures showed a very high correlation with the real data (Figure 3.6B), with an R^2 value of 0.987 (see Figure 3.7 for *on-off* heatmaps and histograms based on model estimates of firing rate). Adaptation is therefore likely to be a parsimonious account for why neural responses in the gerbil IC to otherwise identical sounds differ depending on temporal context, and the results (Figures 3.1 & 3.4) suggest that these differences may be of direct relevance for beat perception.

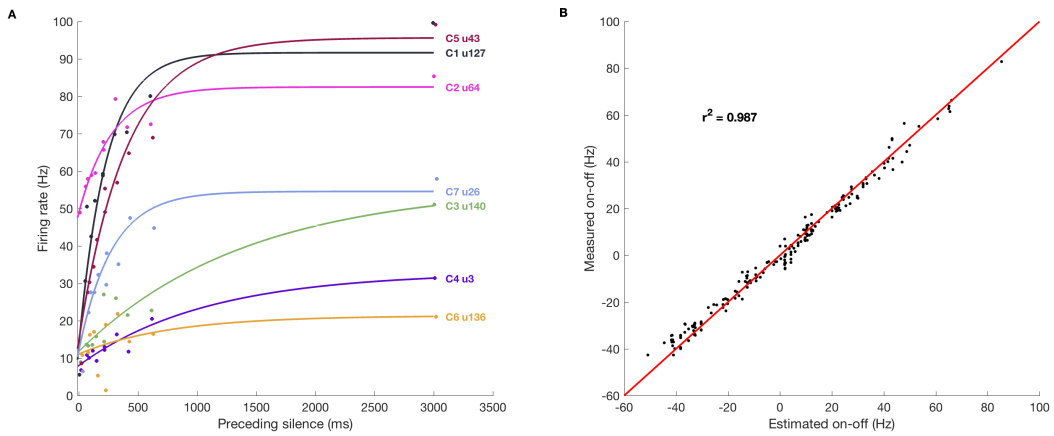


Figure 3.6: **Firing rates estimated from exponential fits closely match measured on-off responses.** A) The average firing rate evoked by noise bursts preceded by varying amounts of silence is used to fit an exponential to each unit. Data and exponential fits for an example unit from each cluster are shown. B) Scatterplot shows a high correspondence between estimated on-off values based on exponential fits (x-axis) and the real data (y-axis).

3.2.5 Problems with the beat-related z-score metric used in Nozaradan, et al. (2012)

Mean beat-related z-scores (referred to hereafter as MBRZs) were introduced in Nozaradan, et al. (2012) and used in a number of studies thereafter (e.g. [Chemin et al., 2014, Nozaradan et al., 2016]) to quantify beat entrainment in either stimulus waveform envelopes or in neural response measures such as EEG waveforms. Initially, I had planned to analyse the gerbil IC data using the same methods. I thereafter developed an alternative time domain method when I discovered that unfortunately, MBRZs lack the specificity required to make them a reliable, unambiguous metric of beat entrainment.

This lack of specificity arises because MBRZs are calculated from Fourier amplitudes of a number of frequencies that are harmonics of the given or suspected underlying beat frequency. The rationale behind this seems reasonable at first, as the MBRZ asks whether the amplitudes of frequency components that are beat related

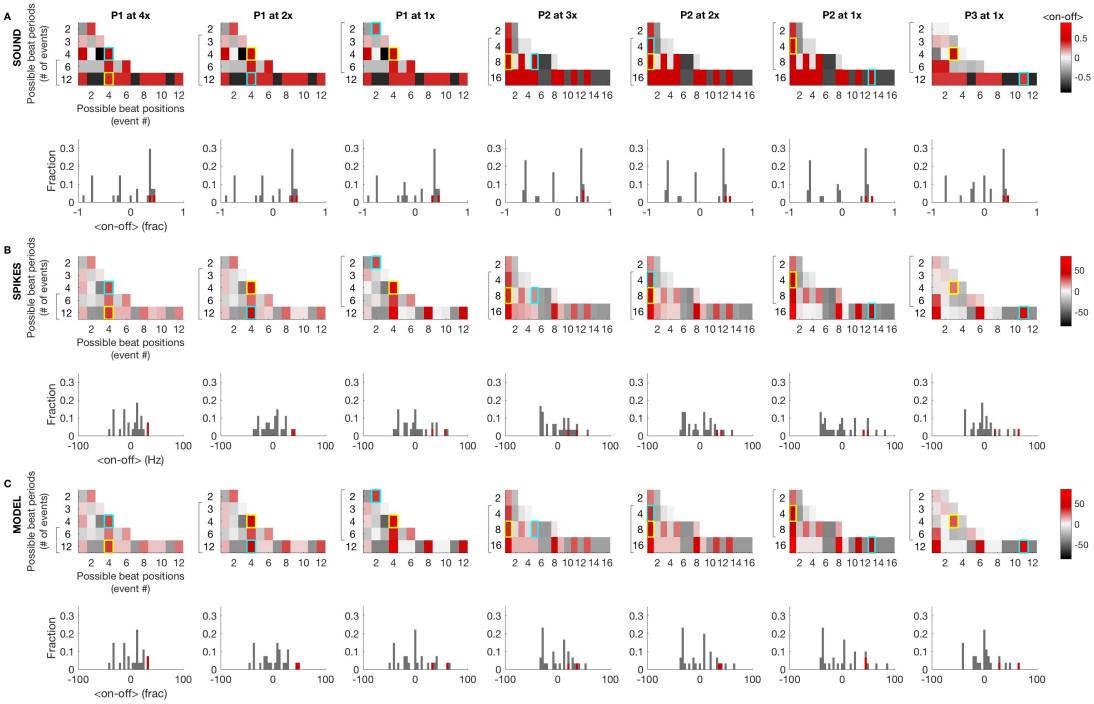


Figure 3.7: **Heatmaps of on-off values by stimulus.** A) Heatmaps of *on-off* values for sound for all plausible beat structures for each of the 7 stimuli. Yellow boxes mark the most commonly tapped beat structure, and cyan boxes mark the second most commonly perceived beat structure for each stimulus. The histograms directly below show the two most commonly perceived beat structures in red, and all possible beat structures in grey. Note that *on-off* values were calculated over each stimulus as played, which results in small discrepancies between the *on-off* histograms for sound for the same pattern presented at different rates. This is due to the looped pattern being cut off in slightly different places at the end of the 33 s. If the final incomplete loop of each stimulus pattern is cut off artificially, the sound on-off histograms for P1 at all rates and P2 at all rates become identical as expected. B) Same as panel A, but for firing rates. C) Same as panel A, but for firing rates estimated by the exponential fits modelling adaptation. Note the close similarity between values in panel B and C.

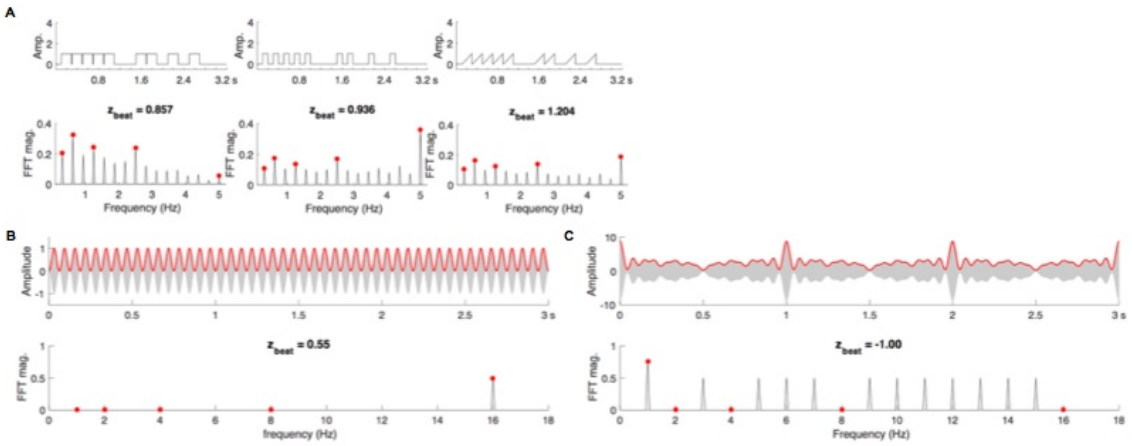


Figure 3.8: **Why a frequency domain analysis can be misleading.** A) Manipulating the envelope of pattern P2 at 1x profoundly changes the peaks in the frequency spectrum and changes the beat-related z-score. The top row are the time domain signals with a normal (left), shortened (middle), and sawtooth (right) shape. Underneath is the result of performing a Fourier transform on the top signals. Beat-related frequencies are marked by red circles. Note that the relative size of all peaks changes with minor changes in shape of the time domain signal, and the the changes made do not necessarily reflect anything to do with beat processing. B) A tone (grey) with a 16 Hz amplitude modulation (red envelope) would result in a positive beat-related z-score though there is no meaningful beat present. C) A negative beat score can be obtained, even though a clear 1 Hz beat is present in the signal.

are on average larger than the components that are not. However, the relative phases of these frequency components are not taken into account at all. This is a problem because harmonic frequency components with relatively modest amplitudes can interact to produce sizeable periodic transients in the time domain if all these components are in phase. Conversely, components with relatively large amplitudes may not produce much amplitude modulation at the common beat period if their phases do not favour constructive interference. By ignoring phase relationships, the MBRZ will miss potentially important periodic time domain features that can strongly affect the type of beat a signal may evoke when listened to. Additionally, changes in a signal's waveform that have no bearing on the signal's beat may change the MBRZ even if there is no obvious change in "beat strength." These problems will be illustrated below.

To compute MBRZs, first the Fourier transform of the neural signal or the sound signal's Hilbert envelope is calculated, and the amplitudes of the first 12 to 16 harmonics of the beat frequencies are determined. The amplitudes of these harmonics are then classified as beat-related or beat-unrelated. Beat-related frequencies are harmonics of the beat frequency that correspond to plausible beat groupings. Plausible beat groupings were 1, 2, 4, 8, 16 for 16-event patterns, and 1, 2, 4, 12 for 12-event patterns. Beat-unrelated frequencies are the harmonics that do not correspond to a plausible beat grouping, e.g. 3, 4, 7 and 9 - 15 for a 16-event pattern. The z-scores for each of these amplitudes is then computed, using $z = (x - \mu)/\sigma$, where μ is the mean across both beat-related and beat-unrelated harmonics, and σ is their standard deviation. Finally, the average over the z-scores for the beat-related frequencies is calculated to yield the MBRZ.

In Nozaradan, et al. (2012), a larger MBRZ is taken as evidence for larger entrainment to beat or more beat-related activity, but as mentioned above, manipulations of a signal that have little to do with beat entrainment can also produce changes

in the MBRZ. In Figure 3.8A, three variants of the sound envelope for pattern P2 at 1x tempo are shown. The rhythmic pattern remains the same throughout, but the envelope of the constituent sound events is either a wide rectangle (left), a narrower rectangle (middle), or a sawtooth (right). Below each waveform is the result of performing a Fourier transform on the corresponding envelope (only the low frequency end is shown, as higher frequencies do not enter into MBRZ calculations). Beat-related frequencies are marked by red circles. Note that the relative amplitudes of the beat-related and beat-unrelated frequencies change even though the rhythmic pattern remains the same. Consequently, there are also changes in the MBRZ values (shown above the spectra), but these changes to the MBRZs do not reflect beat-related entrainment - an illustration of the lack of specificity of the MBRZ metric.

Panels B and C in Figure 3.8 illustrate the insensitivity of the MBRZ to important time domain features that arise from interactions between harmonics of the beat frequency. In Figure 3.8B, a signal envelope was created to have a positive MBRZ for a 1 Hz beat, even though it has no 1 Hz beat whatsoever. This was achieved simply by setting the amplitude of one of the highest beat-related frequencies to be large, and all other relevant harmonics of the beat frequency to be zero. In this case there can be no constructive interference to generate periodic features at the beat frequency. In Figure 3.8C, a signal envelope was created to have a negative MBRZ for a 1 Hz beat, but nevertheless has pronounced periodic transient features once every second, which, when listened to, produce a very salient 1 Hz beat. To create this envelope, the amplitude of the fundamental beat frequency was set to 0.75, all other beat related frequencies were to zero, and all non-beat related frequencies were set to 0.5. All frequency components are in cosine phase. Since most beat-related amplitudes are zero, they are on average smaller than the non-beat-related amplitudes, causing the MBRZ to be negative. But the fairly large single frequency component at the fundamental beat frequency interferes constructively with all the

supposedly non-beat-related frequency components to produce periodic transients at the beat frequency. This evokes a very salient beat at 1 Hz despite the negative MBRZ. According to Nozaradan, et al. (2012), the difference in MBRZ should be evidence that the signal in panel B has more 1 Hz beat than the signal in panel C, but in these examples the opposite is the case. Because the MBRZ does not reliably measure what it is supposed to measure, applying MBRZ analysis to the gerbil IC data would not have resulted in a meaningful, interpretable comparison with Nozaradan, et al. (2012).

3.3 Discussion

By using an approach that combines rodent electrophysiology and human psychoacoustics, I present evidence that low-level processing of rhythmic sounds may exert a more direct influence on beat perception than previously appreciated. Across the set of simple rhythmic sound patterns tested, the perceptual emphasis felt as a beat was accompanied by a subcortical neural emphasis. An evaluation of all hypothetically possible interpretations of beat for each rhythmic pattern revealed that a large on-beat neural emphasis may be a defining feature of the beat structures actually perceived by listeners, perhaps explaining the tendency for listeners to typically agree on where the beat is in a given rhythm. A diverse set of firing patterns contribute to the on-beat neural emphasis observed at the population level, and I showed that onset type cells in the IC show the largest on-beat neural emphasis compared to other cell types. All of these results could be sufficiently explained through a simple exponential fit that models adaptation.

An asymmetry in response strength to acoustically identical sounds comprising a rhythmic pattern has been previously reported in humans based on cortical EEG [Brochard et al., 2003, Potter et al., 2009, Todd and Lee, 2015]. The asymmetry in

responses observed in the human studies, however, reflected subjective accenting, which is a cognitive, attention-driven process. In contrast, our results, despite showing a similar emphasis on some sounds over others despite all sounds being identical, are not a reflection of high-level or cognitive processes since they can be well-described by a simple adaptation model. This strongly suggests that some pre-processing of sound that is relevant to beat perception is already occurring in the brainstem. Adaptation time constants in the IC are robust to a number of stimulus manipulations including duty cycle [Dean et al., 2008]; it would be interesting to find stimulus manipulations that do alter adaptation and investigate whether such manipulations also alter beat perception in a consistent manner. Adaptation has been characterised throughout the auditory system across a wide range of mammalian species, including humans (see [Pérez-González and Malmierca, 2014] for a review). This cross-species generality strongly suggests that the adaptive mechanisms described in this study are also present in humans during beat perception. Though the response patterns we see may be better described as a form of onset detection rather than beat-specific processing, the (average) physiology need not have corresponded with the chosen beat, and adaptation need not have described the physiology. Our results therefore implicate subcortical adaptation as a relevant stage in beat processing, which suggests that beat perception may to an extent be an emergent property of auditory processing, and one that is therefore not entirely culturally dependent.

A natural extension of this hypothesis is that, though rodents and other non-human species rarely exhibit the ability to synchronise their movements to rhythmic stimuli with the consistency or precision of that of humans [Geiser et al., 2014, Bartolo and Merchant, 2015, Hattori et al., 2013], they could nonetheless be able to perceive musical beat. The gradual audiomotor evolution hypothesis [Merchant and Honing, 2014] is consistent with this idea, suggesting that humans' superior movement-synchronisation ability may be due to stronger coupling between auditory and motor areas in humans

than in other animals. However, the idea that nonhuman animals can perceive beat remains to be tested experimentally, perhaps through tasks where animals are asked to discriminate, rather than synchronise their movements to, different rhythmic sound sequences.

The finding that bottom-up physiology may constrain the set of beat interpretations that can be perceived must be considered in the context of other well-known constraints on beat perception, including the 0.5-4 Hz frequency range within which a beat is typically perceived ([London, 2012a], p.28), and a general preference for binary (e.g. 2, 4) groupings over ternary (e.g. 3, 6) or other groupings ([London, 2012a], p.44)[Large, 2008]. Furthermore, the perceived beat and its neural signatures can be modulated at will by top-down mental imagery of beat structure [Nozaradan et al., 2011, Fujioka et al., 2014, Fujioka et al., 2015]. In light of these higher-level considerations, we postulate that the perceived beat may ultimately depend on both a sound's adapted representation in the brainstem, and the set of metrical templates that are common to the listener, with our findings providing evidence of a bias towards configurations that maximise the difference between average on-beat and off-beat responses. Interestingly, this hypothesis parallels the construction of many computational beat-detection algorithms, which typically consist of two stages: a driving function and a periodicity detector [McKinney et al., 2007, Böck, 2016]. The driving function, analogous to the subcortical representation observed here, is a processed version of the raw audio signal, and a range of beat detection algorithms employ onset detection to arrive at their driving function. The periodicity detector then extracts periodicities from the driving function and determines the most likely interpretation of the beat. Our results suggest that adaptive phenomena that are already present at early stages of the auditory pathway may play an important role in detecting, or perhaps rather, emphasising onsets, thereby shaping the driving function of the beat detector.

This work may also relate to other theoretical models of beat perception. The

rule-based model of Povel and Essens (1985) [Povel and Essens, 1985] states that perceptual accents, which are felt for sounds that differ in loudness or in frequency relative to their surroundings, can also arise purely through temporal context. Specifically, they posit that 1) an isolated sound will be perceived as accented, 2) the second of two similar sounds played sequentially will be perceived as accented, and 3) the first and last of three or greater similar sounds in a sequence will be perceptually accented. The location of perceptual accents within a rhythm (which may not be at isochronous intervals) determine the period and phase of the most likely periodic pulse. The adaptation mechanisms we observed here would place a neural emphasis on the first sound of any sequence and would thus not explain these empirical observations, which is a likely indication that adaptation is not the whole story. The empirical observations might reflect an intermediate stage between low-level representation and the fully-formed beat percept, and it would be an interesting follow-up to determine why, for example, the second of a group of two sounds is perceived as accented when the first would evoke higher firing rates subcortically. Other models suggest the importance of the attentional system [Large and Jones, 1999] and the motor system [Schroeder et al., 2010, Patel and Iversen, 2014, Morillon et al., 2015]. Neural resonance theory, an influential computational model that consists of a sensory and a motor layer of nonlinear oscillators whose interactions are modelled as a dynamical system, makes explicit predictions about neural activity and perception [Large et al., 2015, Tal et al., 2017].

This study represents an important first step towards understanding how low-level auditory processing drives beat perception. It also opens up a number of avenues for future exploration. Most interesting among them would be to trace beat processing through different structures in the brain to determine where and how beat-specific neural activity arises. The cortico-basal ganglia-thalamo-cortical loop [Parent and Hazrati, 1995] may be a particularly promising circuit to probe.

Given that the IC has direct and indirect projections to the thalamus, we speculate that the large periodic pushes we observed from on-beat positions interspersed with relatively quiescent off-beat intervals due to adaptation could play a critical role in coordinating the activity of the cortico-basal ganglia-thalamo-cortical loop in response to rhythmic sounds. This is consistent with human EEG findings that describe neural entrainment to rhythmic sound sequences in the auditory cortex [Nozaradan et al., 2011, Tal et al., 2017, Stupacher et al., 2017]. It has been shown that sporadic omissions of otherwise temporally regular sounds evoke strong omission responses in the human auditory cortex, but not in the human auditory brainstem [Lehmann et al., 2016]. The exact relationship between cortical entrainment measured using EEG and spiking responses in the cortex is still a crucial open question. The time domain methods developed here are ideal for cross-species exploration of beat processing in different brain areas and can easily be built upon for more complex stimuli such as music. Note that depending on the temporal and spatial resolution of the recording method used, neural response latencies may first need to be subtracted or deconvolved from the signal. Importantly, the time domain methods developed here also avoid the significant shortcomings of the frequency domain method used in the original Nozaradan, et al. (2012) EEG study (see Figure 3.8).

To conclude, I have shown that brainstem processing may restrict the range of possible beat interpretations for a given rhythmic sound pattern. Onset-type neural responses are particularly important for this type of processing, and this processing is very well captured by a model based on exponential adaptation. These results imply that a consequential part of beat perception may not be culturally determined but may be due to simple brainstem processes that are in-born or learnt from the low-level statistical characteristics of sensory input [Dean et al., 2005, Dean et al., 2008, Willmore et al., 2016]. This observation is a demonstration of how the nature of high-level brain processes is often biased in its characteristics by primitive, low-level

features of the nervous system.

3.4 Methods

3.4.1 Sound Stimuli

Seven out of nine sound patterns from Nozaradan, et. al (2012) [Nozaradan et al., 2012] were were recreated, but instead of pure tones, bursts of frozen pink (1/f) noise were used in order to increase the likelihood of driving neural activity across all recording sites, irrespective of frequency tuning. The patterns used by Nozaradan, et al. (2012) were inspired by the work of Povel and Essens (1985) and were designed to preferentially induce a beat percept based on a grouping of four events at the 1 tempo, with additional beat groupings possible based on subdivision or multiples of the preferred grouping. The seven patterns were constructed from three distinct base patterns (summarised in Figure 3.9), referred to throughout as P1, P2, and P3, played at different rates. P1 and P3 consisted of 12 events and P2 consisted of 16 events. At the slowest tempo, events were presented at a rate of 5 Hz, which meant that each event was 200 ms in duration and was either 200 ms of silence or a 190 ms burst of frozen pink noise followed by 10 ms of silence. At this 5 Hz presentation rate, P1 and P3, each consisting of twelve 200 ms events lasted 2.4s, while P2 lasted 3.2s. Each pattern was then looped continuously over 33 seconds.

P1 and P2 were additionally presented at faster event rates, or tempi. In the original study, this was done in order to investigate the effect that the tempo of a sound pattern had specifically on perception and on EEG responses since the shape of the sound envelope spectrum was relatively consistent for the same pattern played at different tempi. P1 was presented at 5 Hz, 10 Hz and 20 Hz, (1x, 2x and 4x the original tempo), P2 was presented at 5 Hz, 10 Hz and 15 Hz, (1x, 2x and 3x the original tempo), and P3 was presented at 5 Hz only (1x). The complete set of 7

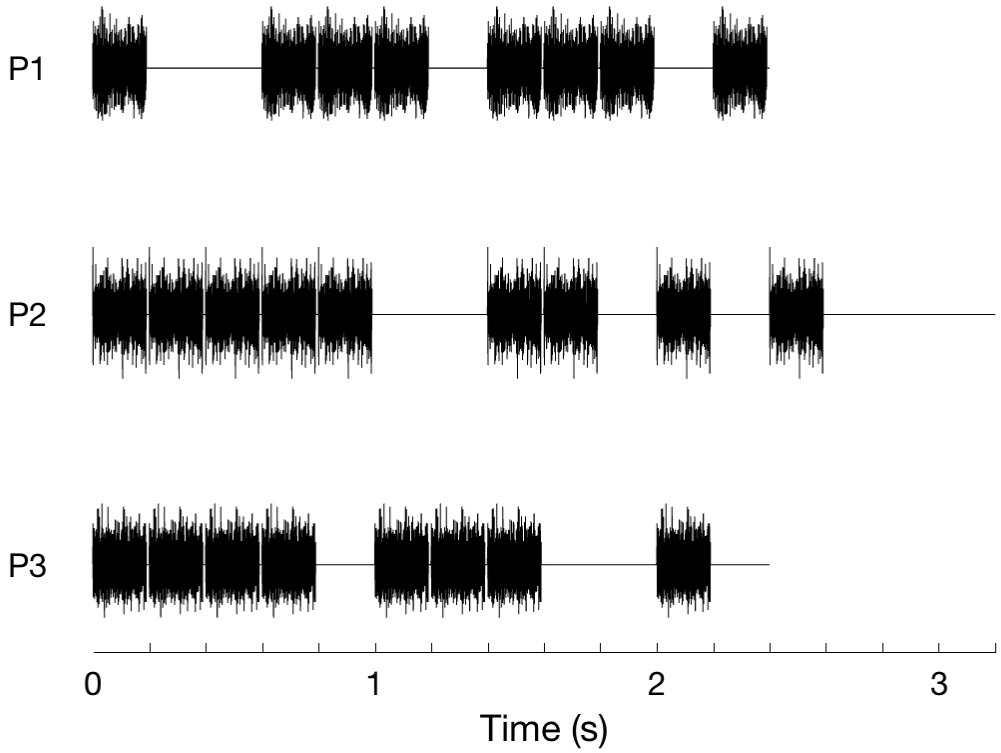


Figure 3.9: **An illustration of the rhythmic patterns P1-P3, from top to bottom.** The slowest tempo (1x) is illustrated for each pattern where each event was 200 ms in duration. In the experiment, P1 was played at tempi of 1x, 2x, and 4x, P2 at 1x, 2x, and 3x; and P3 at 1x.

patterns thus comprised P1 and P2 three different rates each plus P3 at an event rate of 5 Hz only. At the accelerated tempi, the original 190 ms pink noise token used for the noise events was truncated to 90 ms, 56 ms, or 40 ms for the 2x, 3x, and 4x conditions respectively, in each case followed by 10 ms of silence, while the silent intervals were shortened accordingly from 200 ms to 100 ms, 66 ms, and 50 ms for the 2x, 3x, and 4x conditions respectively. The tempi and patterns described here are the analogous to those used in Nozaradan, et al. (2012).

3.4.2 Electrophysiological recordings

3.4.2.1 Surgical protocol

The recording methods were identical to those used in Schnupp et al. (2015). All procedures were approved and licensed by the University College of London, London, UK (UCL) Animal Welfare and Ethical Review Body (AWERB) as well as the UK home office in accordance with governing legislation (ASPA 1986). Four male Mongolian gerbils weighing between 70 and 80 grams were anesthetised with an intraperitoneal injection of 0.65 ml per 100 g body weight of a mixture of five parts of ketamine (100 mg/ml), one part of xylazine (20 mg/ml), and 19 parts of physiological saline. To maintain anesthesia, the same solution was infused continuously during recording at a rate of approximately 2.1 μl/min. A craniotomy was performed centred on the lambdoid suture and extending 3.5 mm lateral from the midline on the right hand side. The visual cortex directly dorsal of the inferior colliculus was aspirated and the sinus was carefully retracted, exposing the IC. Pinnae were removed before the placement of headphones.

Recordings were made using a 64 channel silicon probe (Neuronexus Technologies, Ann Arbor, MI, USA) with 175 μm^2 recording sites arranged in a square grid pattern at 0.2 mm intervals along eight shanks with eight channels per shank. The probe was

inserted into the IC in a medio-lateral orientation for two of the four animals, and in a rostro-caudal orientation for the remaining two animals, in all cases aiming for the central nucleus of the IC.

The seven rhythmic sound stimuli were assembled into a block, with 3 seconds of silence separating each 33 s stimulus loop from the next. Each block was repeated 10 times at each penetration site. Stimuli were presented binaurally through headphones at 80 dB SPL. Sounds were presented with a sampling rate of 48828 Hz, and data were acquired at a sampling rate of 24414 Hz using a TDT system 3 recording setup (Tucker Davis Technologies).

3.4.2.2 Data preprocessing

This work made use of the Open Science Data Cloud (OSDC) [Grossman et al., 2013]. Raw voltage traces from the 64-channel recordings were obtained using custom-written Matlab (Mathworks) software. Offline spike sorting and clustering was done on the raw data using an automated expectation-maximization algorithm (Spikedetekt/Klustakwik) [Kadir et al., 2013], and clusters were manually sorted using Klustaviewa (Cortical Processing Lab, University College London). Firing rates for multi-units were calculated by binning spike times into 5 ms bins, which resulted in firing rate traces at an effective sampling rate of 200 Hz.

To determine whether multi-unit activity was reliably driven by our sound stimuli, a noise power to signal power cutoff of 40 was chosen [Linden, 2003]. Units that failed by this measure of repeatability were excluded from further analysis, leaving 249 reliably driven spiking units. Of those, 29 were identified as single units and the rest were deemed multi-units. All analysis was performed using custom-written Matlab code.

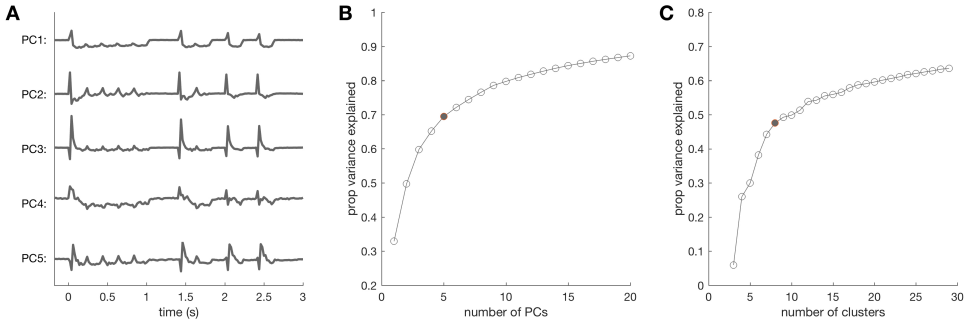


Figure 3.10: **Principal Component Analysis for hierarchical clustering.** A) The first five principal components based on clustering normalised firing rates. B) The elbow method was applied to decide that five principal components would be used for clustering. C) The elbow method applied again to decide that 8 clusters adequately explained variance in the data. One of the 8 clusters contained only 1 unit, and this cluster was discarded from further analysis, yielding the 7 clusters referred to throughout.

3.4.2.3 Clustering

To organise the variety of response patterns observed among IC units into representative groups, we clustered their period peri-stimulus time histograms (PSTHs) using a two-step process consisting of a principal component analysis (PCA) to reduce the dimensionality of the response patterns, followed by standard hierarchical clustering (see Figure 3.10). First, the PSTH of each unit in response to pattern P2, binned at 20 ms, was normalised to have unit power by dividing the PSTH by the standard deviation across all bins. To reduce the dimensionality of each response down from 160 time bins, the PSTH vectors were centered by subtracting their mean. Finally, PCA was performed using the Matlab function *princomp*. The elbow method was applied to determine a cut-off in the number of PCs beyond which the proportion of variance explained began to asymptote (Figure 3.10B). Five PCs (out of 249), which together accounted for 70% of the variance, was chosen as the cutoff. The first five principal component loadings for each unit were then subjected to hierarchical clustering using the Matlab function *cluster* with a Euclidean distance metric. To determine the number of clusters in the data, the percentage of response variability

accounted for by optimally splitting the dataset into between 3 and 30 clusters was calculated. The elbow method was applied again to choose 8 clusters as the point at which the variance in PSTH patterns explained by a model, which replaced each unit's PSTH with the average PSTH for the unit's cluster, began to level off with additional clusters (Figure 3.10C). Of the 8 clusters, one contained only 1 unit and was therefore excluded from further analysis, leaving a total 7 clusters consisting of 248 single and multi-units.

3.4.2.4 Psychoacoustics

The experimental methodology was approved by the local Ethical Review Committee of the Experimental Psychology Department of the University of Oxford, and conforms to the ethical standards in the 1964 Convention of Helsinki.

Fifteen paid participants aged 22 to 45 with normal hearing were recruited. Three subjects were authors on this study, and five had at least 3 years of musical training. Subjects were instructed to listen to the rhythms that would emerge from a masking noise that was ramped down over 3 seconds, to begin tapping with a finger on a hand held button once they had found the beat, and to continue tapping steadily until the rhythm stopped. Stimuli were played through a TDT RM1 mobile processor (Tucker Davis Technologies, Alachue, FL, USA), and presented diotically at 50 dB SPL over Sennheiser HD 650 headphones (Wedemark, Germany). The TDT device delivered the stimuli and recorded button presses, allowing precise tap times to be measured. Patterns were randomly interleaved and presented a total of three times over the course of the experiment. Two subjects whose tapping patterns were not isochronous were excluded from further analysis, leaving a total of 13 subjects.

Consensus beat frequency and phase were determined for each stimulus based on the most common tapping pattern recorded from our human participants. On-beat sounds were all sounds at a consensus tap location, and off-beat sounds were all other

sounds that did not coincide with a consensus tap location.

Chapter 4

Responses in Primary Auditory Cortex May Shape Where and How Clearly the Beat is Felt in Real Music

(Unpublished work)

Electrophysiological recordings, data analysis, and writing are my own. Jan Schnupp and Nicol Harper contributed to the design of the analyses presented here.

Abstract

Previously, I presented evidence from a rodent model that midbrain preprocessing of sounds may already be shaping where the beat is ultimately felt for a set of looped musical rhythms. Specifically, on-beat sounds on average evoked higher firing rates than off-beat sounds. This difference was a defining feature of the set of beat interpretations most commonly perceived by human listeners over others, and basic firing rate adaptation provided a sufficient explanation for these results, suggesting that sensory adaptation, by encoding the temporal context of sounds, creates points of neural emphasis that may influence the perceptual emergence of a beat. But do these findings extend to real music? Twenty 10-s long musical excerpts were played to 3 anaesthetised rats while recording extracellularly from the primary auditory cortex, and recordings were compared with beat annotations from 40 human listeners tapping along to the same excerpts. Consistent with the gerbil study, higher average firing rates were observed on the beat (during post-tap windows of 100 ms) than off the beat (all other periods of time), and this asymmetry was also present in the cochleagram (an approximation of spiking activity at the auditory nerve). Many units in the recorded population also showed a significant on-beat neural emphasis, as well as significant on-beat "gain" over any on-beat emphasis present in the cochleagram, for single trials on individual songs. A large on-beat emphasis in cortical activity was better able to distinguish the beat interpretation overwhelmingly chosen by listeners

out of a wider set of plausible beat structures. Finally, songs that were more "difficult" for listeners to tap to also had fewer units in the auditory cortex showing a significant on-beat neural emphasis. All together, these findings add evidence to the hypothesis that low-level auditory processing creates points of emphasis that may shape where and how clearly the beat is felt, not just for simple musical rhythms, but also in real music.

4.1 Introduction

In the previous chapter, I showed in the gerbil IC that a neural emphasis accompanies the perceptual emphasis felt as the beat in the context of a set of relatively simple musical rhythms [Rajendran et al., 2017a]. Specifically, I found that in response to simple musical rhythms, firing rates in the central nucleus of the inferior colliculus (IC) are on average larger on the beat than off the beat. This difference could be explained by firing rate adaptation, which would suggest that low-level auditory processing may be creating a predisposition to hear beats at certain locations in a musical rhythm.

Real music, however, is spectrotemporally much more complex, with different temporal patterns occurring simultaneously across different frequency bands (e.g. different parts being played by different instruments). Does the observation of stronger neural responses on the beat than off the beat extend to real music? If so, is it consistently present on single trials, since this is the information on the basis of which listeners are able to find the beat? Does the strength of the on-beat neural emphasis affect how clearly the beat is perceived?

The first 10 s of twenty musical excerpts, which spanned a range in tempo and musical genre, were played to three rats under anesthesia while recording from the auditory cortex. These musical excerpts were part of the MIREX 2006 dataset, which included tap annotations for each song from 40 different listeners. In all cases, a steady beat was reported by listeners within the first 10 s. Like in the gerbil IC in response to rhythmic patterns of noise bursts, population firing rates in the rat auditory cortex on average were higher on the beat than off the beat in real music. Many units also showed significant emphasis on the beat on a trial-by-trial basis for individual songs, and many of these units also showed significant on-beat cortical "gain" compared to the on-beat emphasis present in the cochleagram. For these neurons, a high on-beat emphasis was better able to distinguish the consensus beat interpretation out of a

pool of many plausible beat interpretations than any on-beat emphasis present in the cochleagram. These findings support those in the gerbil study and go further by suggesting that auditory processing creates points of emphasis that shape not only where, but also how clearly the beat is perceived.

4.2 Results

4.2.1 Firing rates in the rat auditory cortex are on average higher on the beat than off the beat

Neural activity from a total of 98 single and multi-units were analysed in response to twenty 10-s long musical excerpts taken from the MIREX 2006 dataset online, which included beat annotations made by 40 different listeners [Downie and Welge, 2017]. The most common tapping pattern for each excerpt was taken to be that excerpts consensus beat interpretation. Tap annotations from the 40 listeners were pooled and a simple peak finder was used to determine consensus tap times, and these were checked visually against all tap annotations to confirm that the consensus beat did reflect the majority opinion. Consensus tapping frequencies ranged from 0.7 Hz to 3.7 Hz, corresponding to beat periods of 1.43 s down to 273 ms.

Figure 4.1 provides a fairly raw glimpse at firing rates around beat times. The "beat-triggered averages" shown are based on calculating the average firing rate in a 200 ms window around each consensus tap time (100 ms before and 100 ms after) for each song. Of note are the differences between units and between songs. Song 8 for example shows high firing rates following consensus tap times, with little neural activity in the 100 ms preceding the beat, while Song 17 is an example of a song with no obvious difference between pre-beat and post-beat intervals.

For comparison with cortical firing rates, the cochleagram was calculated for each

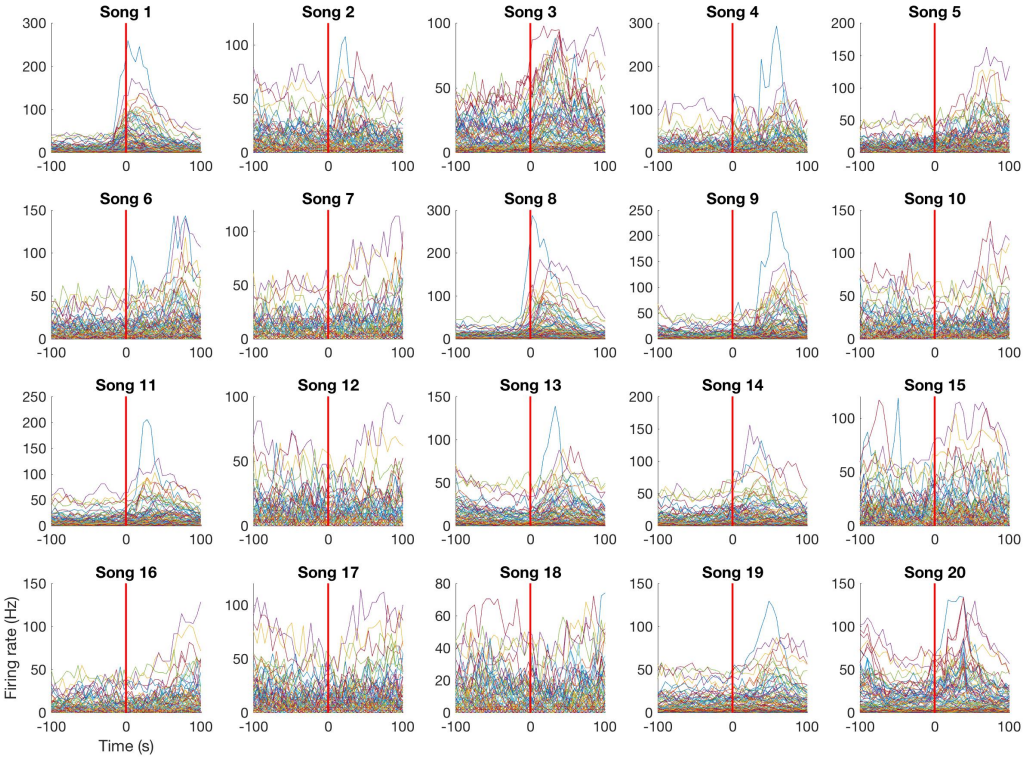


Figure 4.1: **Beat-triggered firing rates by stimulus.** Each colour represents the firing rate of one of the 98 cortical units that make up this dataset.

song. The cochleagram roughly approximates the spike rate at the auditory nerve and is the log of the aggregated magnitudes in log-spaced frequency regions of a spectrogram, set in arbitrary units (see Methods). Here, the cochleagram was calculated over 27 frequency bands with 1/3-octave spacing spanning the range from 50 Hz to approximately 20 kHz. Figure 4.2 shows beat-triggered cochleagram activation for each of the 27 frequency bands. Note that different frequency bands have different activation overall, but also that some stimuli show higher on-beat cochleagram activation across all frequencies (e.g. Song 1, Song 13), while others show higher on-beat cochleagram activation for only a subset of frequency bands (e.g. Song 8, Song 19), and some (e.g. Song 12, Song 19) show no obvious differences between pre-beat and post-beat intervals.

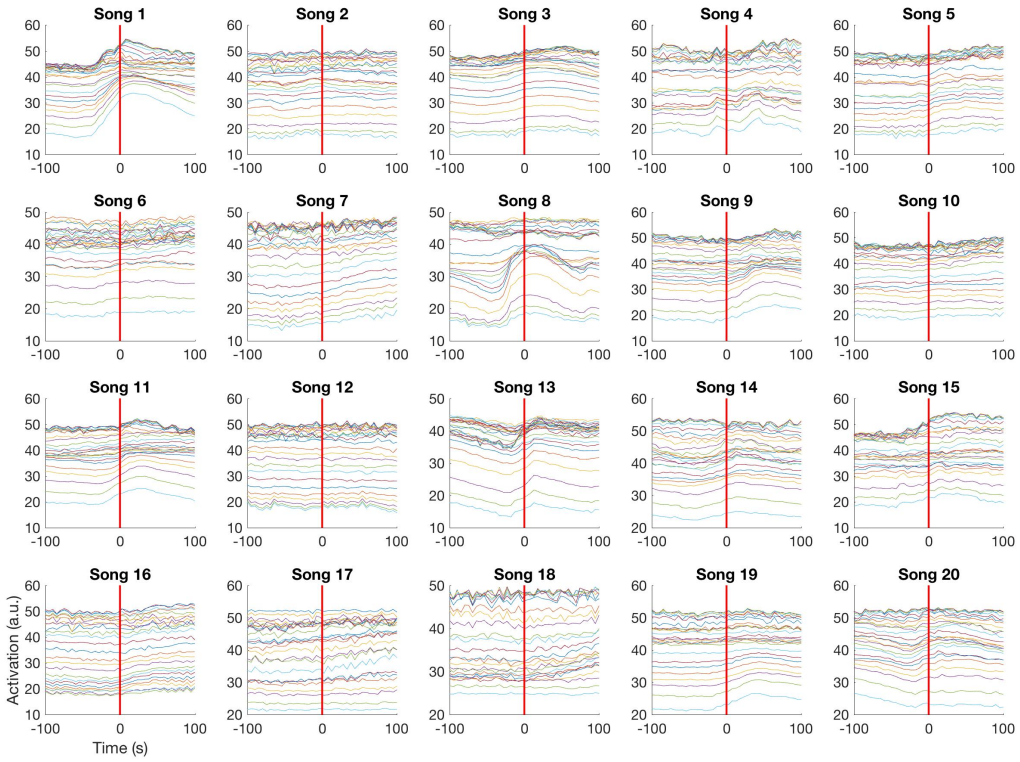


Figure 4.2: **Beat-triggered cochleagram activation by stimulus.** Each colour represents one of the 27 1/3-octave spaced frequency channels in the cochleagram, measured in arbitrary units.

Unlike the rhythms used in the gerbil IC study described in the previous chapter, the musical excerpts used here did not have discrete on and off states, which meant that it was not possible to compare neural responses to identical sounds that differed only in whether they were on or off the beat. An analogous definition of on-beat and off-beat was therefore defined as follows: the time window of 100 ms following each consensus tap was defined as on-beat, and all neural activity excluding these on-beat windows was defined as off-beat. I justify this definition on the basis that 1) the true perceived beat location is most likely after a listener taps due to the tendency for listeners to anticipate the beat with their movements by up to 100 ms [Repp, 2005], 2) defining off-beat activity as all neural activity that is not on the beat creates a reasonable degree of consistency with the gerbil study [Rajendran et al., 2017a], and 3) an interval of 100 ms is less than one half a beat cycle for the fastest beat period of 273 ms. To check that the results are not sensitively dependent on this particular definition on-beat and off-beat, all analyses were additionally run using time windows of 80 ms and 120 ms. The results were consistent with those presented here for a time window of 100 ms.

Figure 4.3 shows average on-beat neural responses on the y-axis plotted against average off-beat responses on the x-axis, for each unit and for each song. Like with the beat-triggered averages, the asymmetry between on-beat and off-beat neural activity varied by stimulus; in total, 15 out of 20 songs showed significantly higher average firing rates on the beat than off the beat ($p < 0.05$, Wilcoxon paired signed rank test, Bonferroni corrected for 20 stimuli, $N = 98$ units). For comparison, Figure 4.4 shows on-beat versus off-beat cochleagram activation for each frequency band and for each song. Here, 12 out of the 20 musical excerpts tested had higher cochleagram activation on the beat than off the beat ($p < 0.05$, Wilcoxon paired signed rank test, Bonferroni corrected for 20 stimuli, $N = 27$ frequency channels). Some songs (e.g. Song 1) show an on-beat emphasis both in cochleagram activation and in cortical

firing rates, while others (e.g. Song 6) show no on-beat emphasis in the cochleagram but do show significantly higher on-beat cortical firing rates, while others (e.g. Song 12) show no on-beat emphasis in either signal.

In the gerbil IC study, I found that neural activity was stronger on the beat than off the beat across the set of musical rhythms tested. To test whether this general finding holds true for real music, grand averages were calculated across songs for on-beat and off-beat neural activity. Cochleagram activation was also averaged across songs for comparison. Figure 4.5 shows that firing rates in the rat auditory cortex, in response to real music, were on average larger on the beat than off the beat ($p < 10^{-6}$, Wilcoxon paired signed rank test, $N = 98$ units), as was cochleagram activation ($p < 10^{-6}$, Wilcoxon paired signed rank test, $N = 27$ frequency channels).

4.2.2 Many units in the rat auditory cortex show a significant on-beat emphasis on a trial-by-trial basis

While the grand averages add evidence to the idea that the perception of beat is accompanied by a neural emphasis, the interesting question is whether this information is available on a trial-by-trial basis for each song, since this is ultimately the information on the basis of which a listener finds the beat. To explore whether on-beat neural activity is higher on a trial-by-trial basis, every unit's on-beat and off-beat neural response was calculated individually for the 12 trials of each song. Figure 4.6 shows the songs for which each unit showed significantly higher neural activity on the beat than off the beat on a trial-by-trial basis ($p < 0.05$, paired one-sided T-test, Bonferroni-corrected for 20 stimuli, $N = 12$ trials). A number of units (e.g. units 40-50) do not show significant on-beat emphasis for any of the songs, while others (e.g. units 1-10) show a significant on-beat emphasis for a number of songs. Songs also range from driving many units to show an on-beat emphasis (e.g. Song 1, 8, 11) to driving no on-beat emphasis in any units (e.g. Song 2, 10, 12, 13, 18). The songs

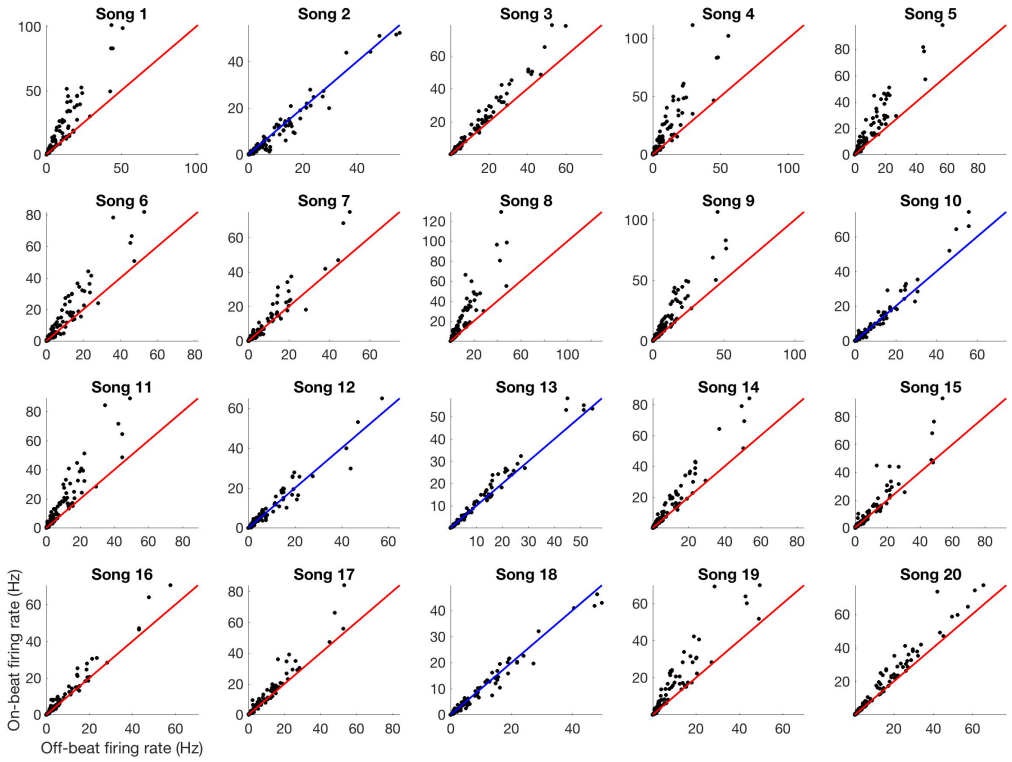


Figure 4.3: **On-beat and off-beat firing rates by stimulus**. The x-axis is mean off-beat firing rate, y-axis is mean on-beat firing rate. Each dot represents one unit (98 in total). Musical excerpts for which neurons showed significantly higher on-beat firing rates than off-beat firing rates are indicated by a red diagonal. Those that showed no significant difference between on-beat and off-beat firing rates are indicated by a blue diagonal.

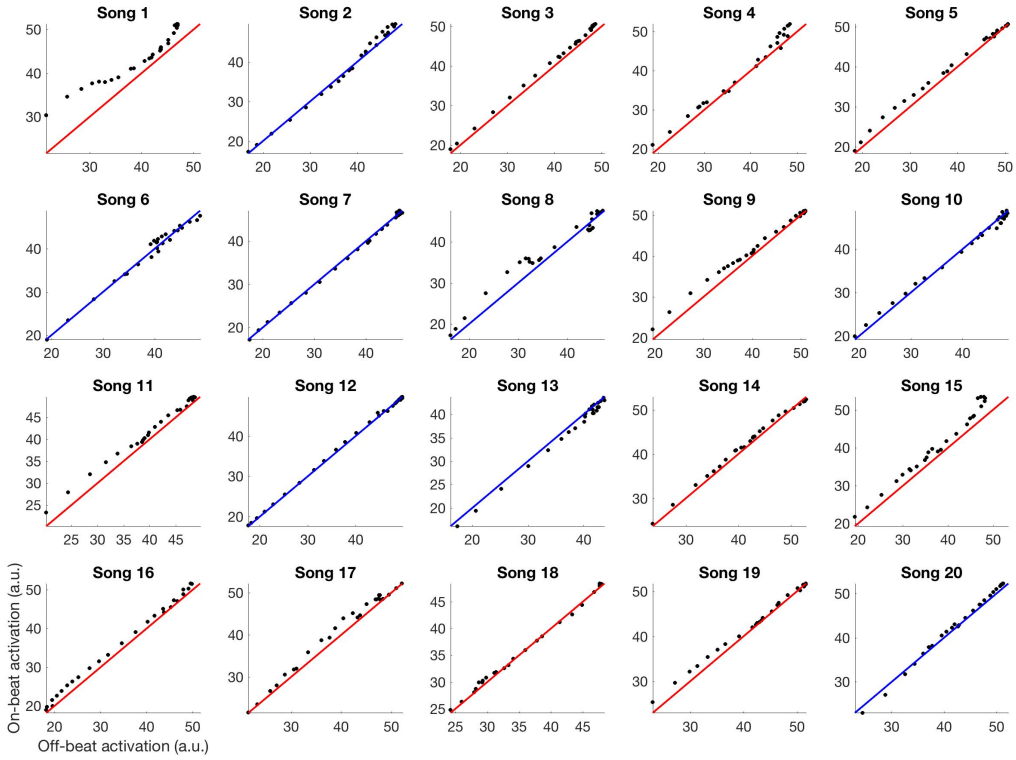


Figure 4.4: **On-beat and off-beat cochleagram activation by stimulus**. The x-axis is mean off-beat cochleagram activation in arbitrary units, y-axis is mean on-beat cochleagram activation in arbitrary units. Each dot represents one of the 27 1/3-octave spaced frequency channels in the cochleagram. Musical excerpts for which cochleagram activation was significantly higher on the beat than off the beat are indicated by a red diagonal. Those that showed no significant difference between on-beat and off-beat cochleagram activation are indicated by a blue diagonal.

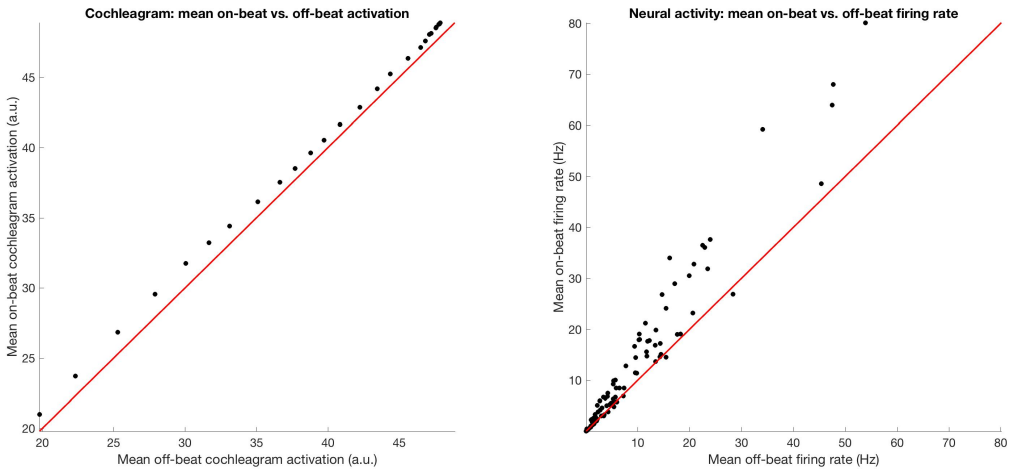


Figure 4.5: **A grand average comparison of on-beat and off-beat cochleagram and neuronal firing rates**. (Left) Average on-beat vs. off-beat cochleagram activation across all musical excerpts for each of the 27 frequency bands with 1/3-octave spacing spanning the range from 50 Hz to approximately 20 kHz, measured in arbitrary units. (Right) Average on-beat vs. off-beat firing rate across all musical excerpts for each of the 98 single and multi-units.

that fail to evoke much of a trial-by-trial on-beat emphasis match those in Figure 4.3 that show no significant on-beat neural emphasis across the population. While these results highlight the diversity in neural responses and in musical excerpts, the majority of songs do have at least a handful of units that show a significant on-beat emphasis.

4.2.3 Many units also show significant "amplification" of the on-beat emphasis relative to cochleagram activation on a trial-by-trial basis

The next question is whether the trial-by-trial on-beat emphasis observed in cortical firing rates result from additional processing by the auditory cortex and the anatomical stages leading up to it, or whether they would have already been present at the auditory nerve. To test this, on-beat neural activity would need to be compared to cochleagram activation at the frequency band that the neuron is best tuned to. A

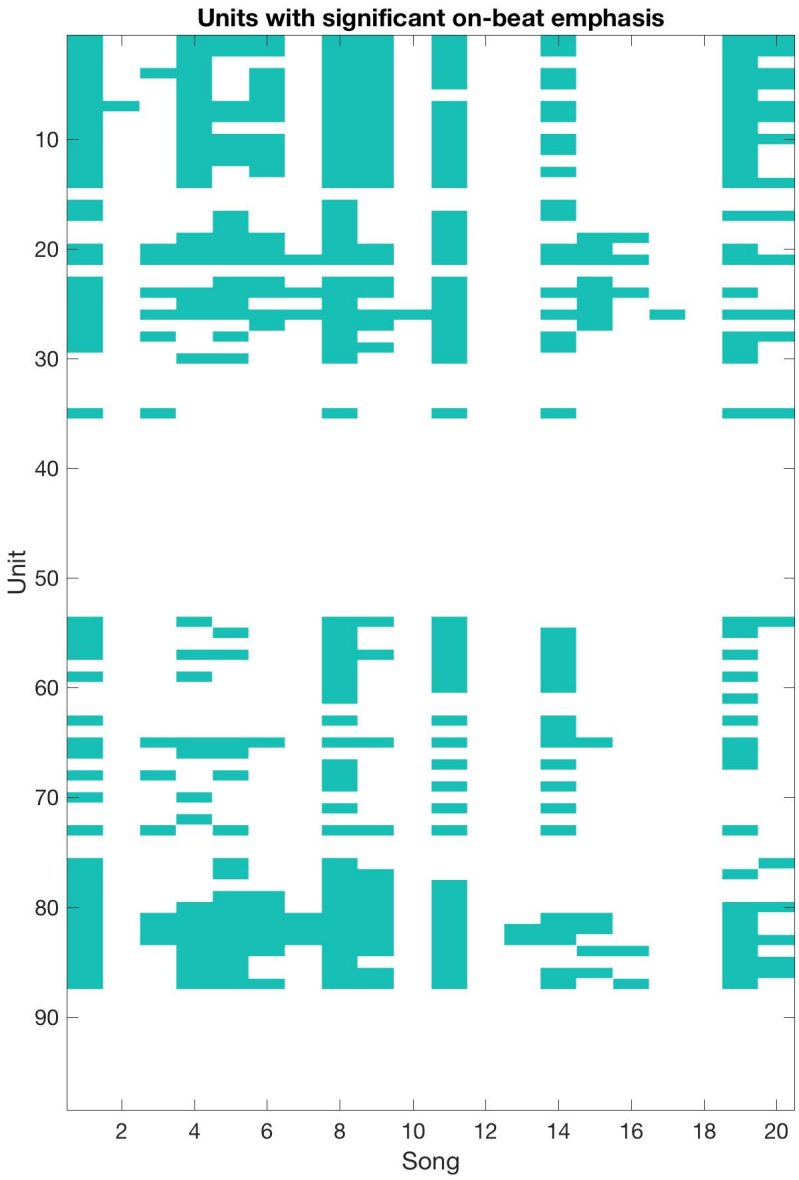


Figure 4.6: **Units with a significant on-beat emphasis on a trial-by-trial basis**. Rows are units and columns are songs. Coloured boxes represent unit/song combinations where on-beat responses were significantly larger than off-beat responses on a trial-by-trial basis.

linear-nonlinear (LN) model was fitted to each unit to estimate the mapping between cochleagram activation and firing rate, resulting in an estimation of its spectrotemporal receptive field (STRF; see Methods). The frequency band containing the largest coefficient was taken as that unit's "best frequency" (BF).

Each unit's on-beat neural activity was then compared to cochleagram activation in the frequency channel corresponding to that unit's BF. To allow for comparison between any on-beat emphasis in neural activity with that in cochleagram activation, on-beat activity in each signal was expressed as a percentage of that signal's off-beat activity. Figure 4.7 reveals those units that showed a trial-by-trial on-beat emphasis that was significantly larger than any on-beat emphasis already present in the cochleagram at that unit's BF ($p < 0.05$, one-sided T-test, Bonferroni-corrected for 20 stimuli, $N = 12$ trials). There is a fairly close correspondence between Figure 4.7 and Figure 4.6, suggesting that any on-beat emphasis present in cortical firing rates are an amplification of any on-beat emphasis that might have been present at the auditory nerve.

4.2.4 Cortical activity is more successful at "finding" the beat than cochleagram activation

In the gerbil study, I showed that a large asymmetry between on-beat and off-beat neural activity was possibly a feature that distinguished the beat structure tapped by listeners from a pool of other hypothetically plausible beat structures. To test whether a large neural emphasis is a distinguishing feature of the consensus beat in real music, a more general method to assemble a set of possible beat structures first needed to be developed. Unlike the rhythms in the gerbil study, the musical excerpts used here (and in general) are not composed of discrete on/off events that form a fixed pattern that loops, and thus cannot simply be broken down into the multiplicative factors of the total pattern length (e.g. 1 group of 16, 2 groups of 8, 4 groups of

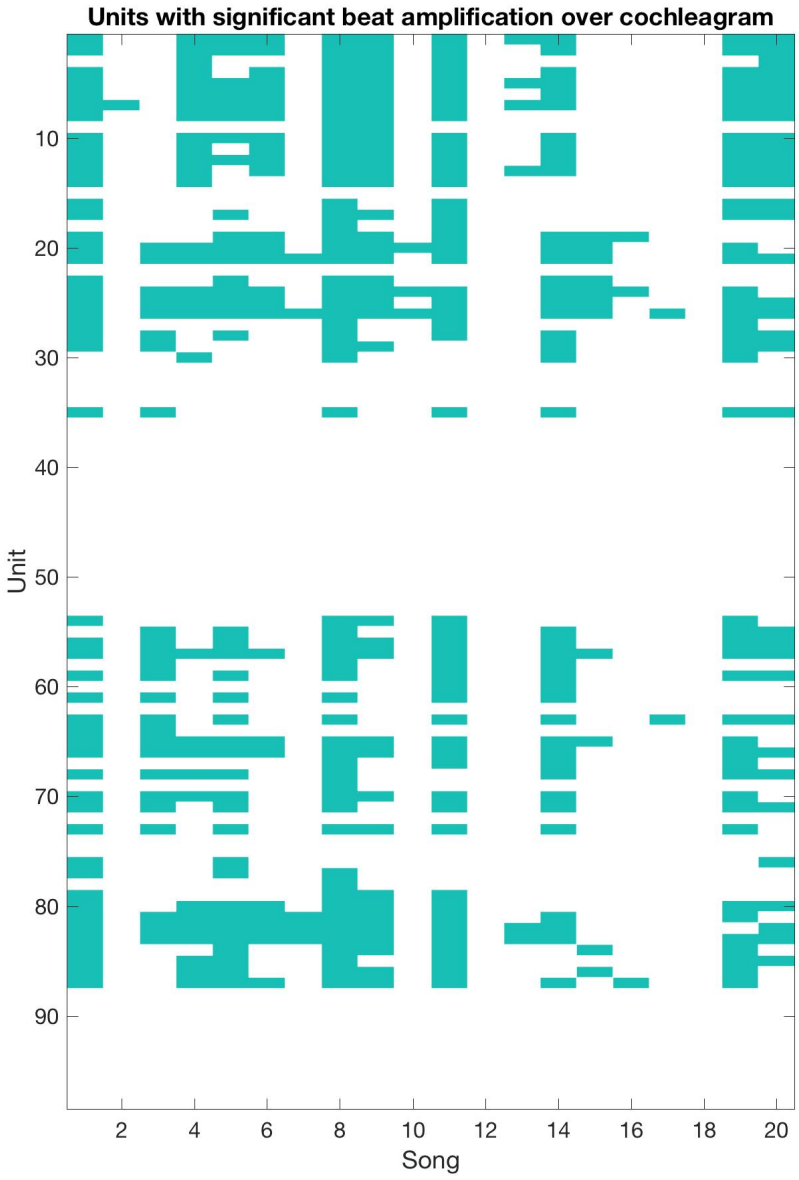


Figure 4.7: **Units that show a significant trial-by-trial amplification of on-beat emphasis relative to that present in the cochleagram**. Rows are units and columns are songs. Coloured boxes represent unit/song combinations where on:off-beat for neural activity was significantly larger than on:off-beat cochleagram activation at that unit's BF on a trial-by-trial basis.

4, or 8 groups of 2 for a pattern composed of 16 events). It is difficult to attain an exhaustive list of plausible beat structures without delving too deeply into the realm of music theory. The method described here therefore represents an attempt to strike a balance between what is simple and generalisable and what is exhaustive and music theoretically correct.

While the consensus beat interpretation was the overwhelming choice for all songs, alternative interpretations of the beat were always either one level slower (tapping every other consensus beat), or one level faster (tapping at 2x or 3x the speed of the consensus beat). Whether the fast tempo was a subdivision by 2 or 3 was determined by the unanimous judgment of myself and two other listeners with over 5 years of musical training. For all songs the consensus beat was best grouped by 2 for the slow tempo. These three levels - consensus, faster, and slower - represented the set of plausible beat periods for this analysis. Theoretical "tap times" in between consensus taps were estimated by simply dividing the interval between two consecutive consensus taps by 2 or 3 depending on the subdivision required for the fast tempo. This yielded 10 plausible "beat structures," or beat period and phase combinations, for each song that had a subdivision by 3 (Song 4, 5, 12, 15) and 7 plausible beat structures for each of the remaining songs that had a subdivision by 2.

Next, on-beat and off-beat activity was calculated for each plausible beat period and phase combination. As with the previous analysis, for each hypothetical beat period and phase combination, the "on-beat" value was the average signal waveform (firing rate or cochleagram activation) during the first 100 ms following the hypothetical tap, and the "off-beat" value was the average waveform excluding these windows. On-beat and off-beat neural activity was calculated on a trial-by-trial basis for each "beat-relevant" unit that, based on the previous analysis, showed significant "on-beat cortical gain" of the consensus beat over what was present in the cochleagram. This analysis is analagous to that shown in Figure 3.4 in the previous chapter, just that

here it is done on a trial-by-trial basis for each individual beat-relevant unit and its corresponding cochleagram activation at BF for each song.

If cortical activity serves to amplify any on-beat neural emphasis present at lower stages in auditory processing, and if this is at all relevant in influencing the beat structure actually perceived, then we should expect that the consensus beat structure should have one of the highest on:off-beat values out of the pool of possible beat structures for each song, and that the consensus beat should consistently rank higher based on cortical firing rates than on cochleagram activation. For each trial, the the on:off-beat values for each hypothetical beat structure were ranked in descending order such that the beat structure with the highest on:off-value was ranked #1, the second highest was ranked #2, and so on. Figure 4.8 shows the consensus beat's median rank across trials for each unit/song combination that showed a significant amplification of the on-beat emphasis in firing rates relative to the cochleagram. Visually, it is apparent that for a few songs (e.g. Song 4, 6, 7), the consensus beat ranked particularly poorly in the cochleagram but scored relatively better (i.e. had a lower rank) based on cortical activity.

A final summary is shown in Figure 4.9, which is a histogram of the difference between the consensus beat's rank based on cochleagram activation and cortical activity (Figure 4.8B - 4.8A, cochleagram rank minus cortical activity rank). If cortical activity ranked the consensus beat more favourably than cochleagram activation, this would result in a positive difference. For example, if for a particular unit/song combination the consensus beat ranked #1 out of all possible beat structures based on cortical activity and #4 based on cochleagram activation, the difference would be +3 in the histogram. While in most cases the ranks were the same between cortical activity and cochleagram activation, the overall distribution of these differences has a mean significantly greater than zero ($p < 0.0001$, one-sided Wilcoxon signed rank test). This suggests that, consistent with the gerbil study, a large on-beat emphasis

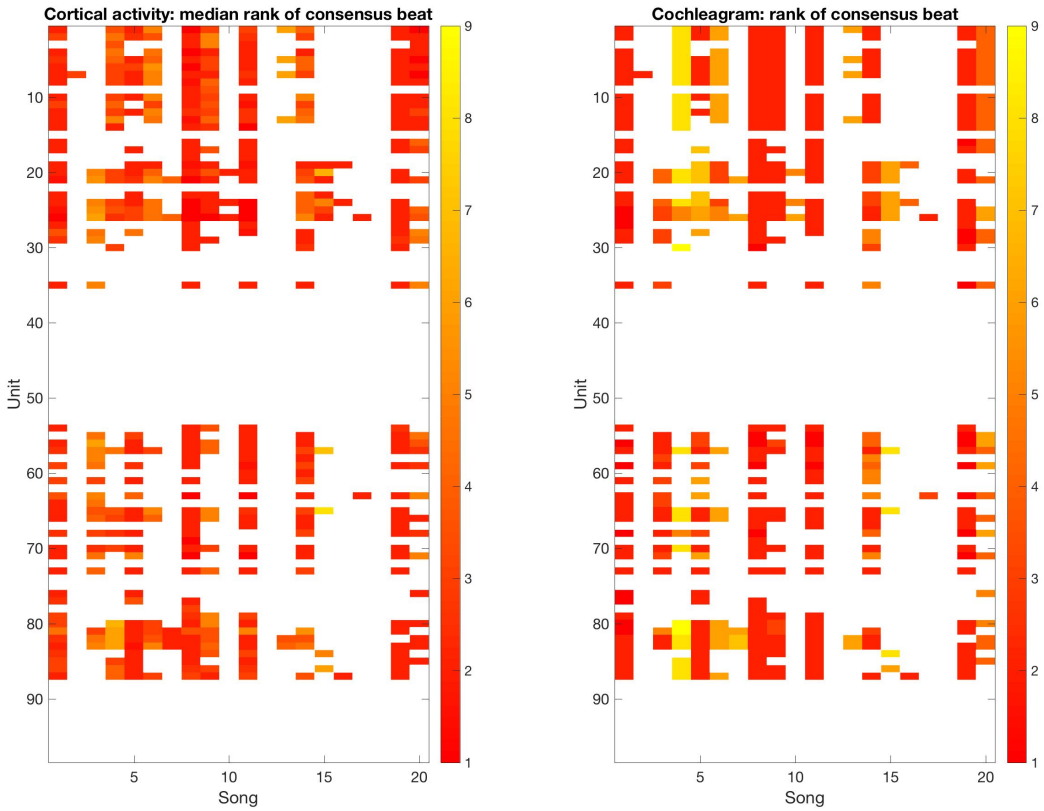


Figure 4.8: **Comparison of how the consensus beat ranks against other plausible beat structures based on cortical activity and cochleagram activation.** (Left) Median rank across trials for the consensus beat structure based on on:off-beat values for all plausible beat structures. Ranks were calculated for all unit/song combinations where cortical activity showed significant amplification of on:off-beat values over that of the cochleagram. (Right) Consensus beat rank based on cochleagram activation in the frequency band corresponding to that unit's BF.

might be a feature that distinguishes the consensus beat from other plausible beat structures, and that a consequence of auditory processing is to further restrict the possible beat structures that are likely to be heard.

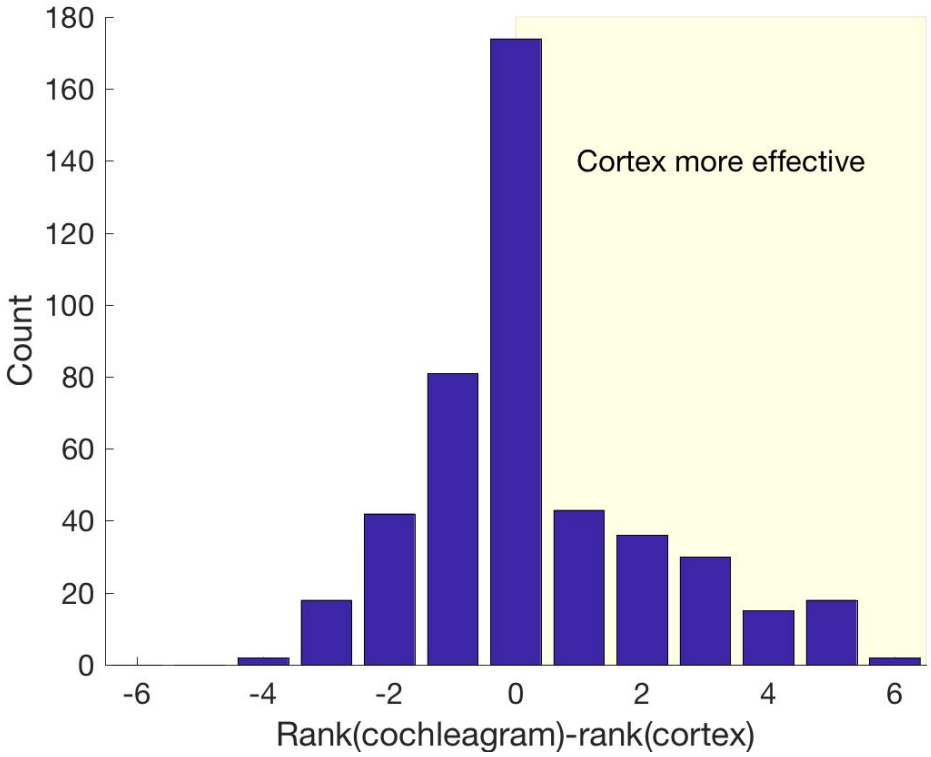


Figure 4.9: **Difference in consensus beat ranks based on cortical activity and based on cochleagram activation.** Values shown are the consensus beat's rank based on cortical activity subtracted from the median rank across trials based on cochleagram activation. Positive values mean that cortical activity would have been more effective than cochleagram activation at distinguishing the consensus beat out of a pool of plausible beat structures based on a higher on:off-beat value.

4.2.5 What about the songs that showed little to no on-beat neural emphasis?

The results so far are in agreement with the idea that large neural emphases that result from auditory processing may be a relevant precursor to the induction of the beat percept. However, the handful of songs where there was little to no significant on-beat neural emphasis cast doubt on this conclusion. While a more thorough feature search could uncover a metric that explains the location of the consensus beat across all songs, such a feature search would be extremely exploratory, and there is a risk of overfitting to this particular dataset. Therefore, a better way might be to simply assess the relevance of the presence (or absence) of an on-beat neural emphasis by asking whether it has any bearing on the consistency of tapping behaviour. The hypothesis here is that if an on-beat neural emphasis is relevant to beat perception, then a song that has a substantial number of units that show significant on-beat cortical gain should have a perceptually clearer beat, which should result in tapping behaviour that is more precise. Figure 4.10 tests this hypothesis by plotting the variance around consensus tap times for each song as a function of the number of cortical units in this dataset that showed significant on-beat gain relative to cochleagram activation. There is a strong negative correlation between tap variance and the number of units that show a significant on-beat gain ($R = -0.775$, $p < 0.0001$). It is also worth noting that the four songs that had the highest tap variance were Song 10, 17, 2, and 18, songs which across all analyses have shown very little evidence for an on-beat neural emphasis. This suggests that a song that evokes a stronger on-beat emphasis in cortical activity than in the auditory nerve might be perceived as having a clearer beat, and conversely, that a song that fails to reliably evoke stronger on-beat neural activity in the auditory cortex might be perceived as having a less clear beat. However, given that a consensus beat is found anyway for songs showing little to no

on-beat cortical gain, this suggests that beat perception may not rely exclusively on the presence of an on-beat neural emphasis, or at least not exclusively on the one present in primary auditory cortex.

4.3 Discussion

The aim of this this study was to extend the methods and findings in the previous chapter to understand whether an on-beat neural emphasis also accompanies, and possibly defines, where the beat is felt in real music. Like in the gerbil IC, population firing rates in the rat auditory cortex were on average higher on the beat than off the beat reported by human listeners. Many units in the recorded population also showed a significant on-beat neural emphasis, as well as significant on-beat "gain" over any on-beat emphasis present in the cochleagram, for single trials on individual songs. Furthermore, a large on-beat emphasis in cortical activity was better able to distinguish the beat interpretation overwhelmingly chosen by listeners out of a wider set of plausible beat structures. The relevance of a prominent on-beat neural emphasis was further supported by a final analysis that showed that songs that were more "difficult" for listeners also had fewer units in the auditory cortex showing a significant on-beat neural emphasis. All together, these findings add evidence to the hypothesis that low-level auditory processing creates points of emphasis that may shape where the beat is felt, not only in musical rhythms, but also in real music.

It has been shown previously in the human auditory cortex that strong beat positions in music are accompanied by larger source currents than weak or non-beat positions, and that these differences might be pre-attentive [Todd and Lee, 2015][Tierney and Kraus, 20 While the human auditory cortex results are likely to involve higher-level processing, the early response differences are suggestive of something pre-attentive, and the higher firing rates observed in the gerbil IC under anesthesia at on-beat positions in musical

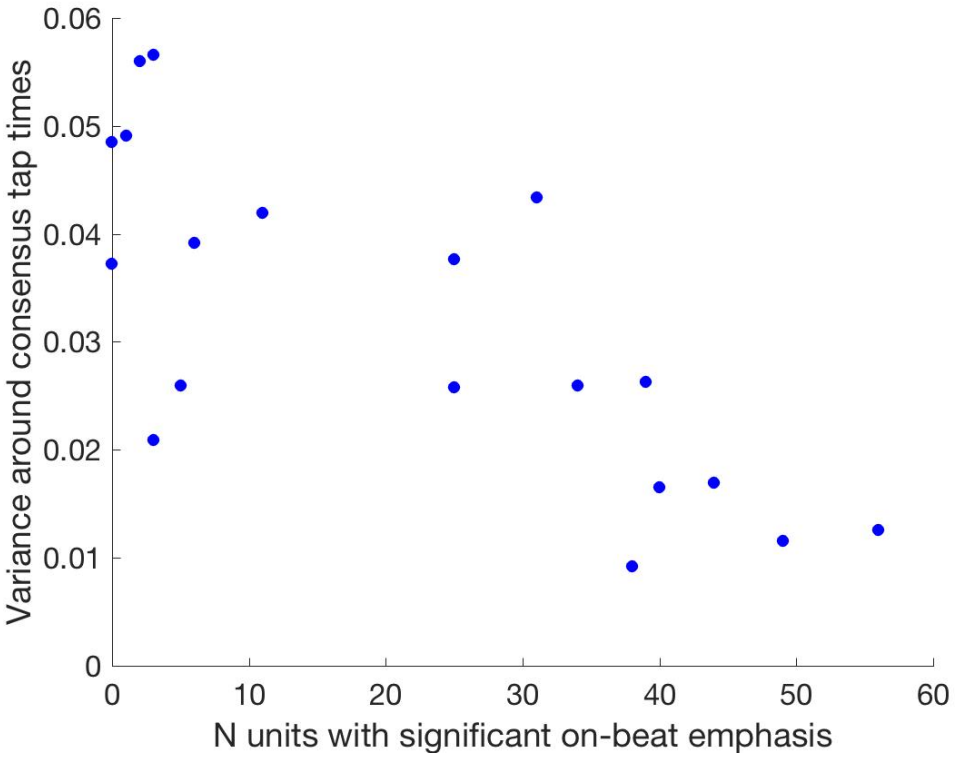


Figure 4.10: **Tapping precision as a function of on-beat neural emphasis.** The x-axis is a simple count of the number of units (out of the 98 recorded from in this dataset) that showed significant on-beat cortical gain relative to cochleagram activation on a trial-by-trial (see Figure 4.7). The y-axis is the variance around consensus tap times as a fraction of the beat period. This was calculated by first pooling all taps across all annotators for a given song, then computing the distance between every tap and its nearest consensus tap time, and then normalising by the consensus beat period for that song so that distances were expressed as a fraction of the beat period. Finally, the variance was taken across these normalised distances to arrive at the tap variance for each song. Each data point in the figure represents one song.

rhythms are furthermore suggestive of relevant bottom-up precursors to beat perception [Rajendran et al., 2017a]. The present findings in the rat auditory cortex are a first demonstration that auditory processing creates a pre-attentive neural emphasis on the beat in real music, corroborating those previous studies and adding evidence to the idea that the perceptual emphasis felt as the beat is accompanied by a bottom-up neural emphasis.

These findings also add support to the hypothesis that a large on-beat emphasis is a distinguishing feature of the beat interpretations actually perceived by listeners. Out of a set of plausible beat structures, the consensus beat structure had a relatively higher on-beat emphasis in the auditory cortex than in the cochleagram. Since adaptation time constants generally become larger with each higher anatomical stage in the auditory pathway [Langner, 1992], a possible explanation for this result could be that adaptation at each stage in auditory processing extracts information about temporal context across longer and longer timescales. The on-beat neural emphasis quantified here could be considered a measure "global contrast" since it describes the relative strength of neural activity at on-beat positions against background neural activity at all other times. These results would suggest that the global contrast that may be relevant to beat perception is not yet prominent at the auditory nerve and may therefore emerge at later stages in auditory processing when adaptation is occurring across a sufficiently long timescale.

The diversity of neural responses and songs is also noteworthy. Not all neurons showed an on-beat neural emphasis to any song, and not all songs had any neurons showing an emphasis on the beat. This is a valuable demonstration that the story developed here is not entirely simple and straightforward, and that real insight into beat perception will come from not just looking at aggregate responses averaged across stimuli. An advantage provided by this diversity was that it enabled an additional test on the possible relevance of on-beat cortical gain on beat perception. Specifically, the

finding that listeners tapped more precisely to songs that had more neurons showing significant on-beat gain suggests that on-beat neural gain might affect how clearly the beat is perceived, and/or how precisely the motor system is able to synchronise to it. There is a tendency for sensorimotor synchronisation errors to scale with longer beat periods [Repp and Su, 2013], but since the variances quantified here are relative to each song's beat period, song tempo would not be an explanation for the variability in tapping behaviour found here. These observations provide some big-picture insights into the consistencies - and complexities - of auditory processing as it relates to musical beat. At the same time, it could be interesting to drill deeper into the different neural responses through a clustering analysis to understand how any why responses differ. Similarly, it would also be interesting to better understand what about the spectrotemporal dynamics of different songs results in vastly different levels of beat-related engagement of auditory cortical neurons.

Another open question is whether or not rat cortical neurons entrain to musical beat. Human studies have consistently shown evidence for the entrainment of beta, gamma, and delta-theta oscillations to beat and meter [Iversen et al., 2009, Fujioka et al., 2015, Snyder and Large, 2005, Zanto et al., 2006, Nozaradan et al., 2011, Lehmann et al., 2016]. While there are a few cases of nonhuman species that show a natural ability to synchronise movements to a musical beat [Patel et al., 2009][Cook et al., 2013][Ro most electrophysiological studies of beat perception have been done in nonhuman primates who do not show evidence of being able to extract a beat from rhythms more complex than isochronous [Honing et al., 2012]. When isochronous auditory or visual rhythms are attended, phase-locked entrainment of delta oscillations has been observed in the macaque auditory cortex [Lakatos et al., 2008]. Isochronous auditory stimulation also evokes anticipatory "echo responses" in the auditory of mice even under anaesthesia, although these mice had been trained to associate each of a sequence of isochronous broadband noise bursts with a water reward for which they

had to lick [Li et al., 2017]. Human studies also suggest that low-frequency entrainment to rhythmic auditory stimuli requires task engagement [Lakatos et al., 2013, Doelling and Poeppel, 2015]. Together, these studies would suggest that it might be possible to observe phase-locked entrainment-type responses to isochronous sounds in our preparation, but that entrainment specifically to the musical beat is unlikely because there is currently no evidence that rodents can extract a beat in rhythmic stimuli more complex than isochronous, nor were the rats exposed to the musical excerpts in the context of a task prior to electrophysiological recordings. Future work is required to determine whether rats can perceive musical beat, and whether the ability to do so is accompanied by an emergence of neural entrainment at the beat frequency. However, a follow-on analysis of local field potentials in this dataset could provide first a glimpse into whether any isochronous spectral features in the tested musical excerpts drive entrainment.

Related to this, there is also a lack of clear understanding about how entrainment manifests through spiking activity. At present, the most compelling context that bottom-up processing should struggle to explain is the "missing-pulse" phenomenon [Tal et al., 2017], where an on-beat interval can evoke a clear neural response despite a lack of acoustic energy there. However, as described earlier, the anaesthetised auditory cortex is perfectly capable of responding in an anticipatory manner to sounds that are expected but not played for isochronous stimuli presented at rates as low as 1/3 Hz [Li et al., 2017]. Is this entrainment? Furthermore, Tal and colleagues (2017) use rhythmic stimuli specifically designed to show no energy at the pulse frequency because of phase cancellation, despite the fundamental note rate, as defined by inter-onset intervals, being at the pulse rate. Is it therefore entrainment if the brain finds the inter-onset intervals and shows energy at the pulse frequency? In light of these doubts, it would appear that a key study that is missing is one that performs a side-by-side comparison of firing rates recorded from the human auditory cortex and

from a rodent auditory cortex under identical conditions and in response to musical excerpts and isochronous stimuli. Such a study would shed light on the similarities and differences in the dynamics of entrainment-like neural activity in response to isochronous stimuli, which are likely to be similar, and to music, which are likely to be a point of divergence. By using methods such as those described here, in addition to classic frequency domain analyses, a clearer picture may be obtained about the extent to which signals can be explained by sensory processing, and whether what remains is really due to top-down entrainment.

In the previous chapter, I showed that adaptation could explain the beat-related preprocessing observed in the gerbil IC. Such an analysis was not possible to perform and interpret conclusively here. This is because spectrotemporal receptive fields (STRFs) in the auditory cortex consist of not one, but several multiple independent components, which is is in contrast to IC responses, which can be well-described by a single linear filter [Atencio et al., 2012]. Linear-nonlinear (LN) models, such as the one used here to arrive at a rough estimation of a unit's best frequency (BF), typically underestimate abrupt peaks in firing rate [Harper et al., 2016]. The accuracy of firing rate predictions in the auditory cortex based on an LN model can be improved upon substantially if an additional input nonlinearity designed to capture the adaptive properties of the IC is introduced [Willmore et al., 2016], or when the feedforward properties of the nonlinear network in which auditory cortical neurons are embedded is introduced [Harper et al., 2016]. The problem is that a simple LN model is known to produce inaccurate predictions of firing rate, but more accurate models incorporate complexity that goes beyond the scope of simple adaptation. It would be an interesting follow-on to this study to explore whether predicted firing rates based on relatively accurate physiologically-inspired computational models would adequately capture the findings here, and whether any discrepancy between predicted and measured firing rates is evidence of any beat-specific processing.

To conclude, I demonstrate in this exploratory work that firing rates in the rat auditory cortex are on average higher on the beat than off the beat in real music. This asymmetry might explain why certain interpretations of beat are overwhelmingly chosen over others, and why listeners are able to tap to some songs more precisely than others. This further underscores the importance of low-level auditory processing in creating a representation of sound where certain features are emphasised based on temporal context, a representation on which other high-level processes such as perception and sensorimotor synchronisation ultimately depend.

4.4 Methods

4.4.1 Stimuli

The 20 songs tested were taken from the MIREX 2006 dataset available here: $http://www.music-ir.org/mirex/wiki/2006:Audio_Beat_Tracking$. The dataset consists of 20 song excerpts 40 s in duration and beat annotations from 40 different human volunteers. Only the first 10 s of songs and beat annotations were used in this study.

4.4.2 Surgical protocol

All procedures were approved and licensed by the University College of London, London, UK (UCL) Animal Welfare and Ethical Review Body (AWERB) as well as the UK home office in accordance with governing legislation (ASPA 1986). Three female Lister Hooded rats weighing approximately 250 grams were anesthetised with an intraperitoneal injection of 0.05 ml domitor and 0.1 ml ketamine. To maintain anesthesia, a saline solution containing 16 ug/kg/h domitor, 4 mg/kg/h ketamine, and 0.5 mg/kg/h torbugesic were infused continuously during recording at a rate of

1 ml/h. A craniotomy was performed 4.7 mm caudal to bregma and extending 3.5 mm lateral from the midline on the right hand side.

Recordings were made using a 64 channel silicon probe (Neuronexus Technologies, Ann Arbor, MI, USA) with 175 um^2 recording sites arranged in a square grid pattern at 0.2 mm intervals along eight shanks with eight channels per shank. The probe was inserted into the auditory cortex in a medio-lateral orientation wherever possible.

The 20 songs were played in randomised order for a total of 12 repeats, with 3 seconds of silence separating each song from the next. Stimuli were presented binaurally through headphones at 60 dB SPL. Sounds were presented with a sampling rate of 48828 Hz, and data were acquired at a sampling rate of 24414 Hz using a TDT system 3 recording setup (Tucker Davis Technologies).

4.4.3 Data analysis

4.4.3.1 Data preprocessing

Offline spike sorting and clustering was done on the raw data using an automated expectation-maximization algorithm (Spikedetekt/Klustakwik) [Kadir et al., 2013], and clusters were manually sorted using Klustaviewa (Cortical Processing Lab, University College London). Firing rates for multi-units were calculated by binning spike times into 5 ms bins, which resulted in firing rate traces at an effective sampling rate of 200 Hz.

To determine whether spikes were reliably stimulus-driven, a noise power to signal power cutoff of 40 was chosen [Linden, 2003]. Song 1 was arbitrarily chosen to the be the stimulus for which the repeatability of responses was measured. Units that failed to show a noise power to signal power ratio less than 40 based on the 12 repeats were excluded from further analysis, leaving a total of 98 multi-units. All subsequent analyses were performed using custom-written Matlab code.

4.4.3.2 Calculating the STRF using an LN model

First, the sound stimuli were transformed into a simple approximation of the activity pattern received by the auditory pathway by calculating the log-scaled spectrogram ('cochleagram'). For each sound, the power spectrogram was taken using 10 ms Hamming windows, overlapping by 5 ms. The power across neighboring Fourier frequency components was then aggregated using overlapping triangular windows comprising 27 frequency channels with center frequencies ranging from 50 Hz to 20,319 Hz (1/3 octave spacing). Next, the log was taken of the power in each time-frequency bin, and finally any values below a low threshold were set to that threshold. These calculations were performed using code adapted from melbank.m (http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html). The LN model was trained to predict the firing rate at time t from a snippet of the cochleagram extending 100 ms (20 time bins) back in time from time t. The linear weights describing the firing rate of each neuron were estimated by regressing each neuron's firing rate at each time point against the 100 ms cochleagram snippet directly preceding it. A sigmoidal nonlinearity was then fitted to map from the linear activation to the predicted PSTH such that it minimises the error between the predicted PSTH and the observed PSTH. The STRF was calculated 20 times, each time setting a different song aside as the "test" set. Initially this was done so that firing rate predictions could be made based on model fits, though this was not done for the reasons described in the discussion. A unit's best frequency (BF) was taken to be the frequency channel containing the maximum weight based on that unit's average STRF.

Chapter 5

General Discussion

Most of the content in this chapter has been published in:

Rajendran, V. G., Teki, S., & Schnupp, J. W. (2017). Temporal Processing in Audition: Insights from Music. Neuroscience.

ST contributed to the final section on the neural mechanisms of timing, and both ST and JWHS helped revise the manuscript. The conceptualisation, research, and writing are my own.

Abstract

In the work contained in this thesis, I showed that rhythmic sound patterns are easier to detect, that adaptation in the auditory system may be a mechanism by which information about temporal structure at the rhythm timescale is encoded, and that low-level auditory processing may play a substantial part in shaping complex percepts such as where and how clearly we can feel the beat in music. These findings occupy an interesting space at the intersection of several fields of research ranging from auditory neuroscience to music psychology. It is perhaps not a coincidence that the rhythm timescale is at the fringe of both of these fields; for auditory neuroscientists, this temporal scale is arguably "high-level" in that many aspects of neural coding and perception of sound at timescales shorter than this one are still not fully understood. On the other hand, perception of rhythmic timing is a fairly "low-level" question in the field of music psychology, which is also concerned with understanding larger themes of emotion, cognition, and culture. The feature that really unites these fields in the context of this work is the central importance of timing, temporal sequencing, and temporal anticipation. The ability to hear a beat in music relies on an ability to form temporal predictions, which is a general feature of auditory temporal processing that is equally relevant to everyday tasks such as auditory scene analysis, pattern detection, and speech perception. Therefore, this Discussion will highlight the cross-disciplinary implications of the work contained in the previous chapters by

bringing together findings from the music and auditory literature as they relate to temporal sequence processing at the rhythm timescale. The five broad areas touched upon in this chapter include music psychology, sensorimotor synchronisation, neurophysiology of beat perception, temporal predictability in auditory perception, and neural mechanisms for timing perception.

5.1 Music psychology and the natural bounds of human temporal processing

Rhythm is an aspect of music that occurs on a medium temporal scale (hundreds of milliseconds to one or two seconds), longer than that of pitch (up to tens of milliseconds), but shorter than that of global musical form and structure (several seconds to minutes, e.g. phrases, sections, movements). Crucially, it is at the temporal scale of rhythm that a number of overt motor processes in humans tend to occur, such as the swing of the arms and legs during walking or the inhaling and exhaling of air during breathing. Dance, for example, is movement to the rhythm of music. In the Western music tradition, movements such as dance are typically synchronised to a periodic pulse, or beat. It is important to highlight that pulse and beat are not physical properties of the music itself, but are perceptual phenomena that arise from music through beat induction (BI). BI refers to our ability to extract a periodic pulse from music and is widely considered a cognitive skill, though its species-specificity and domain-specificity are topics of current debate [Honing, 2012].

5.1.1 Timescales

Studies into sensorimotor synchronisation to beat tend to employ simple movements such as tapping a finger. These studies find that beat is generally perceived between 0.5-4 Hz, corresponding to time intervals of 250 milliseconds to 2 seconds, a range beyond which precise coordination of motor movements becomes difficult [Repp, 2005, McAuley et al., 2006, Repp and Su, 2013]. Even within this range, perception of time differs between shorter and longer time intervals. When asked to judge the duration of time intervals, there is a systematic tendency for human listeners to overestimate shorter time intervals (roughly 250-400 ms) and underestimate

long ones (600 ms to 2 seconds). The transition point in between, measured by various researchers to lie between 400-600 ms, is termed the indifference interval and also corresponds to the rate at which people spontaneously tap [Fraisse et al., 1958, Fraisse, 1963, Fraisse, 1978, Fraisse, 1982, Clarke, 1999].

In the context of rhythm perception, it is also the boundary between temps courts and temps longs [Clarke, 1999]. When human subjects are asked to tap rhythmically, almost invariably they employ a 1:1 or 1:2 ratios to the time intervals between successive taps. A ratio of 1:2 refers to a tapping pattern of long and short intervals where the short intervals are precisely half the duration of the longer ones. This alludes to the theory that temps longs are intervals during which a listener is aware of the passage of time, whereas temps courts do not evoke a sense of time passage by themselves, but listeners are aware that a certain number of them grouped together make up a longer interval. Tapping ratios of 1:2 observed almost always span the indifference interval, with the longer interval belonging to temps longs and the shorter one to temps courts (see Figure 5.1). A preference for time intervals with integer ratios also shapes the way rhythmic patterns are perceived [Jacoby and McDermott, 2017]. Compared withinh intervals with noninteger ratios, intervals with integer ratios are more accurately reproduced by listeners [Essens and Povel, 1985] and show a distinct pattern of neural activity [Sakai et al., 1999]; see also Section 3). Interestingly, while a preference for integer ratios spans different cultures, the specific ratios preferred by listeners is primarily determined by their music listening experience and is not strongly affected by musical expertise [Jacoby and McDermott, 2017].

5.1.2 Beat - a perceptual accent

The timescales are one aspect of what determines where a musical beat might be felt, but not all sound events in music are equally likely to induce a beat percept. Certain events in music have been described as giving rise to perceptual accents, which

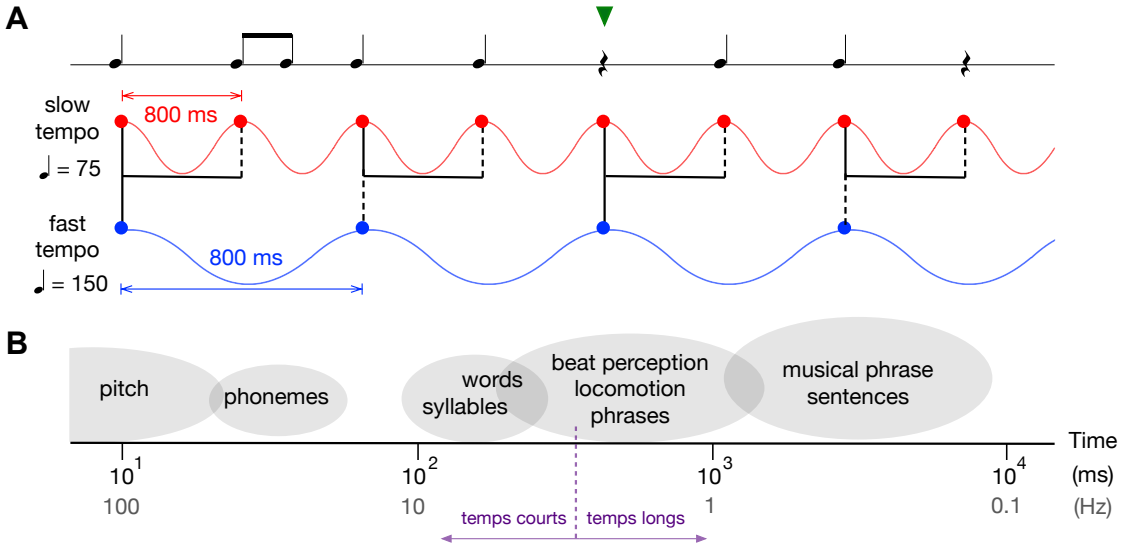


Figure 5.1: **Timescales in auditory perception.** A) The beat perceived depends on the tempo at which a musical rhythm is played. In this simple, recognizable example rhythm, notes represent sound events and those with a single stem are quarter notes, notes with the attached stem are eighth notes, and the remaining symbol is a quarter rest (silence). The basic unit of time here is the quarter note; a quarter rest is the same duration as a quarter note, and each eighth note is half the duration of the quarter note. Tempo is conventionally specified in beats per minute, so for the slow tempo (in red), there would be 75 quarter notes per minute, and each quarter note is therefore 800 ms in duration. The fast tempo (in blue) is twice the speed of the slow tempo. In both cases, the beat may be comfortably perceived at 800 ms intervals or 1.25 Hz (filled circles), but depending on the tempo this may coincide with different events in the music. The alternation of strong (solid lines) and weak beats (dotted lines) are illustrated for each tempo. Syncopation (green triangle), or when a beat is felt where there is silence, is very common in music. B) This schematic illustrates the timescales over which common auditory events unfold. Time is on a log scale from small intervals (fast rates) to large intervals (slow rates), with values shown in milliseconds and in Hz. The indifference interval is marked in purple; shorter intervals are temps courts, longer intervals are temps longs.

together with the temporal constraints described earlier, form the basis of where the beat is felt.

Perceptual accents may be felt at points that differ in loudness or in frequency relative to surrounding events. However, perceptual accents can also arise purely through temporal context. Essens and Povel [Povel and Essens, 1985] proposed a theoretical framework for metrical complexity based on empirical observations. They posit that 1) an isolated acoustic event will be perceived as accented, 2) the second of a set of two similar or identical acoustic events played in sequence will be perceived as accented, and 3) the first and last of three or greater similar events in a sequence will be perceptually accented. Based on the location of perceptual accents within a rhythm (which themselves may not be periodic), the period and phase of a periodic pulse can be determined.

Neural adaptation such as that described in Chapter 3 could provide an explanation for why isolated acoustic events and the first in a rapid succession of acoustic events are perceptually accented. However, the empirical finding that the second of two sounds played in sequence or the last of a train of successive sounds are perceptually accented would not necessarily be explained by adaptation. It is currently unclear if there is a neural signature is that matches these empirical findings, but knowing whether such a signature exists, as well as where and what its dynamics look like, could unlock key insights into how the perception of beat arises.

5.1.3 Not all beats are created equal, nor is there always an accent: subjective rhythmisation

The basic temporal structure of a piece of music can be described by its meter, or its recurring pattern of strong beats and weak beats. Again, 'strong' and 'weak' in this context are perceptual notions, much akin to identical ticks of a clock being instinctively perceived as tick-tock-tick-tock [Bolton, 1894, Brochard et al., 2003,

van Noorden and Moelants, 1999, Bååth, 2015]. This tick-tock of a clock could be described as having a binary meter, or a beat pattern based on the number two (most commonly two or four beats in a bar) and have the beat pattern of strong-weak-strong-weak. Ternary meters, or bars based on the number three, have a pattern of strong-weak-weak-strong-weak-weak, the most common example being a waltz. Other more complex meters, for example based on 5 or 7 beats in a bar are also common in Western music, though binary and ternary meters are more often studied because they are generally more effective in inducing a clear beat percept. The preference or natural acceptance of binary meter could be due to a likeness of such meters to common rhythmic motor patterns such as breathing or walking.

5.1.4 To summarise

Within the range of frequencies that a periodic pulse can typically be perceived, there is a further distinction between longer timescales across which the passage of time is noticeable, and shorter timescales of which several together are perceived to fit into a longer timescale. The boundary between the two is the indifference interval, which lies somewhere between 400-600 ms. This is where temporal perception is most accurate in humans [Fraisse, 1978], and incidentally also corresponds to a comfortable walking pace [Styns et al., 2007]. Beats themselves are thought to arise as a result of the combination of perceptual accents and the constraint of a periodic pulse within the range of perceivable beat frequencies. A repeating pattern of strong and weak beats group together to form the musical meter of a piece, and some meters (binary and ternary) are more easily interpreted generally, perhaps due to their semblance to binary motor patterns or to the harmonic series on a fundamental frequency.

5.2 The psychoacoustics of beat perception: sensorimotor synchronisation

With beat and meter defined, we are now equipped to explore how we synchronise to beat. When we hear a beat in music, we almost instinctively want to move with it, and it has been shown that listeners often cannot reliably maintain movements that are out of sync [Repp, 2002b]. The synchronisation of our movements to an external rhythm is known as sensorimotor synchronisation (SMS). SMS has been studied extensively (see [Repp, 2005] and [Repp and Su, 2013] for reviews), but here I highlight a few observations from SMS studies that may be of particular relevance to the neurophysiological processes that underlie rhythm perception.

5.2.1 When tapping along, we are usually early

Negative mean asynchrony (NMA) is a testament to the predictive nature of synchronising a motor action with an expected stimulus. NMA refers to the observation that listeners, when asked to tap along with an isochronous pacing stimulus such as a metronome, tend to anticipate stimulus onsets with their taps by tens of milliseconds, rather than tapping with a distribution that is symmetric around sound onsets (sometimes early, sometimes late). Interestingly, listeners are often unaware of their own NMA, suggesting a general incongruence between objective and subjective synchrony. Musicians tend to show less NMA than nonmusicians, and the neurophysiological differences between the two groups may therefore shed some insight into the interaction between the sensory, motor, and cognitive processes involved. A final observation is that NMA decreases as the tempo of the pacing stimulus increases, which may allude to the tendency to overestimate short time intervals and underestimate longer ones described in the previous section. For a more comprehensive review of NMA, see

[Aschersleben, 2002].

5.2.2 Beat period and phase may have distinct underlying representations

A number of intriguing insights into SMS have also been uncovered through studies that systematically perturb the pacing stimulus, for example by introducing a phase offset, tempo change, or a sequence of distractors. Overwhelmingly, the evidence points to an interesting behavioral dissociation between phase correction and period correction [Repp, 2005]. Phase correction in an isochronous sound sequence refers to a subtle adjustment of tapping so that it returns to synchrony following an unexpected inter-onset interval (IOI) that is abnormally short or long, which would result in an abrupt phase shift in the sequence that is either temporary or persistent (Anomaly and Phase shift in Figure 5.2, respectively). Period correction refers to the adjustment of taps to a sudden tempo change, or an abruptly shorter or longer IOI (Tempo change in Figure 5.2).

The phase correction mechanism appears to be automatic; the timing of the tap subsequent to the perturbation shifts according to whether the preceding IOI was shortened or lengthened, even when the phase offset is imperceptible to listeners [Repp, 2002a]. Similarly, shifting a single tone in an isochronous stimulus such that it results in a shorter IOI on one side of it and a longer IOI on the other induces an involuntary shift in tap times after the perturbation, even when participants were told to ignore the perturbation. If a distractor sequence of isochronous tones is introduced, taps shift towards it, and interestingly this effect appears to be insensitive to the pitch difference between the tones of the pacing and distractor sequence. In this case, temporal coherence seems to be key: if a target and distractor tone are within 120 ms of each other, tapping behavior would suggest that they are treated as a joint referent. In contrast, period correction to a step change in tempo appears to require the change

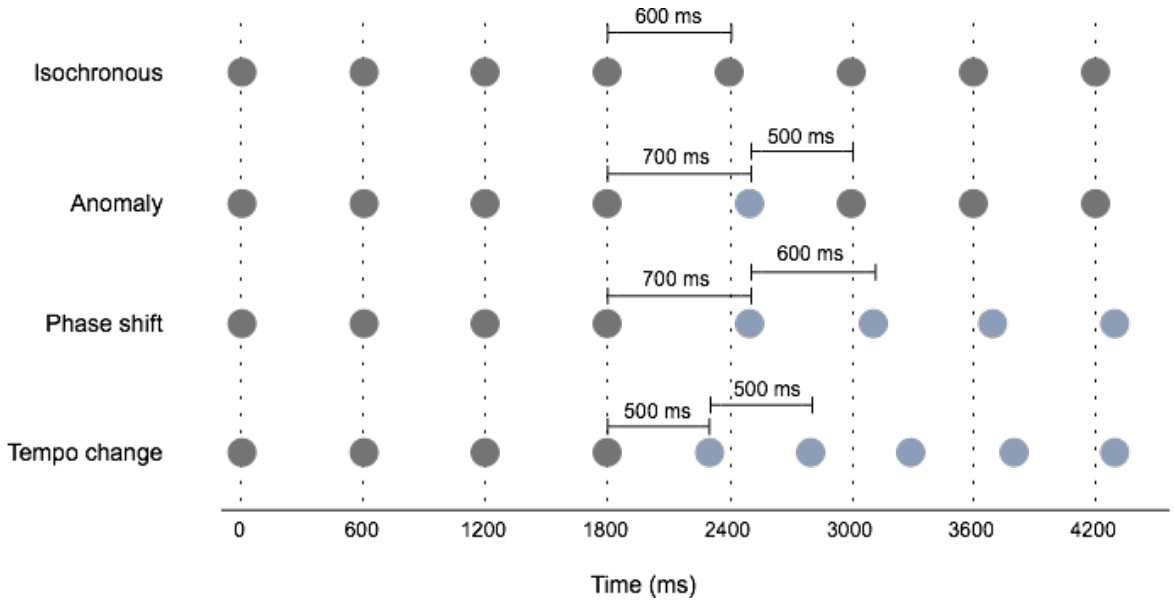


Figure 5.2: **Illustration of a selection of perturbations used to study period and phase correction in sensorimotor synchronisation.** The x-axis represents time, and here a temporal grid representing 600 ms intervals is marked by the vertical dotted lines. Circles represent clicks to which a listener would align their taps, and blue circles mark sounds whose timing would be a departure from the isochronous condition where there is no perturbation. In the Isochronous condition (top), a click is played every 600 ms. In Anomaly, a single click in the sequence is manipulated such that the IOIs that flank it are too long and too short by 100 ms, allowing the remainder of the sequence to remain unchanged. In Phase Shift, a single IOI is lengthened by 100 ms but this time is not gained back, resulting in a phase shift of 100 ms that persists for all remaining clicks, even though their IOI remains 600 ms. In Tempo Change, the IOI changes abruptly from 600 ms to 500 ms. This would be perceived as a faster tempo and would require and adjustment in the period of taps.

in tempo to be perceptible [Repp and Keller, 2004]. Listeners' ability to completely ignore a tempo change and continue tapping at the original tempo without showing any period correction is further evidence that period correction requires cognitive control, in contrast to phase correction, which in the same task proved impossible for participants to suppress. Under the looser constraint of self-paced movements, there does appear to be a natural tendency to synchronise movements to the period and phase of a musical beat [Peckel et al., 2014].

5.2.3 How fast we can tap along depends on what we are tapping to

Depending on the nature of the task, different studies reports somewhat different ranges within which beat perception and sensorimotor synchronisation can occur. In truth, the context-dependent nature of SMS is in itself a reflection of the different sensory and biomechanical constraints. At the slow extreme, IOIs longer than 1.8 s result in inaccurate synchronisation where taps begin to lag the pacing stimulus. At the fast extreme, finger tapping with an isochronous pacing stimulus can be done at a rate of up to 5-7 taps per second. However, if the task is 1:n synchronisation, the IOIs in the pacing signal can be as short as 100-120 ms for trained musicians, which suggests that audiomotor processing can cope with these fast rates. This so-called subdivision benefit too depends on the exact subdivision required. 1:2, 1:3, 1:4, and 1:8 tapping can be done at lower IOIs than 1:5 or 1:7, with 1:6 and 1:9 tapping falling somewhere in between. This suggests a certain level of automaticity to subdivision by 2, 3, 4, and 8, while the cognitive demands of counting groups of 5 interferes with sensory processing. A similar effect is observed when listeners tap an isochronous beat in non-isochronous rhythmic patterns. Rhythmic patterns differ in their complexity, and while very complex rhythms are difficult to synchronise to [Povel and Essens, 1985], rhythmic patterns of medium complexity are what

elicit the greatest desire from listeners to move [Witek et al., 2014]. This may relate to beat salience, which has been shown to correlate with listeners' desire to move [Madison et al., 2011]. Rhythmic complexity and the strength of the beat percept also influence the precision of temporal judgments [Grube and Griffiths, 2009] and may be due to differences in neural representation of metrically simple, complex, and non-metrical sound patterns [Sakai et al., 1999, Vuust and Witek, 2014].

5.2.4 To summarise

Perception appears to be based on a judgment of intervals, whereas synchronisation of movements appears to be the result of a joint computation based on stimulus onsets and ongoing taps. Beat has both a period and a phase, and perturbation studies suggest that dissociable processes underlie adjustment of each. Specifically, phase correction of motor actions appears to be automatic and involuntary, while period correction requires cognizance of a tempo change and can be suppressed at will. The temporal limits of synchronisation ability are also context-dependent and are a result of biomechanical, sensory, and cognitive constraints. These factors are also at play in the context of more complicated rhythmic patterns and real music, where the temporal structure of the sound affects listeners' ability and desire to synchronise to a steady pulse.

5.3 Neurophysiology of beat perception

A number of electrophysiological studies in humans have attempted to identify the neural correlate of the beat percept. Neural signatures of beat perception have been identified through direct and indirect means and involve distributed cortical and subcortical networks [Teki et al., 2012]. Comparisons with studies in newborn humans [Winkler et al., 2009b] and nonhuman species would suggest that some aspects of

rhythm perception may be innate to humans and to some nonhuman species, whereas other aspects may be unique to humans.

5.3.1 Strong beats differ physiologically from weak beats

As described in earlier, subjective rhythmisation can generate a metrical percept of alternating strong and weak beats even if all sounds are identical. This paradigm arguably would allow for the dissociation between the cognitive and the sensory aspects of beat perception in the context of identical isochronous sounds. Electroencephalography (EEG) studies that investigate the neural correlates of subjective accenting do so either directly or indirectly. Indirect methods involve the measurement of event-related potentials (ERPs) resulting from rare deviant (e.g. an omission or change in loudness) in a series of standard or expected sounds. Differences in the ERP to perturbations coinciding with strong and weak beats may therefore signify neurophysiological differences in processing that result from subjective accenting. Early components of the ERP are known as the mismatch negativity (MMN) and are considered to be pre-attentive in contrast to later components (300-600 ms post-stimulus onset), which are presumed to reflect cognitive mechanisms. Though subjective accenting is cognitive by definition, the setting of temporal expectations may influence the processing of forthcoming sounds in a predictive manner, and indeed both early and late ERP differences have been found between deviants at strong and weak beat positions [Brochard et al., 2003, Abecasis, 2005, Schaefer et al., 2010, Geiser et al., 2010, Bouwer et al., 2014, Honing et al., 2014].

The more direct approach compares sound-evoked responses at strong and weak positions in rhythmic sequences. Here too, strong beats evoke higher source current activity than weak beats in temporal and frontal areas, despite sounds being acoustically identical [Todd and Lee, 2015]. Similarly, a target sound played over a background of pop music evokes stronger cortical and brainstem responses if it was

presented on the beat, rather than shifted off the beat by of the inter-beat interval [Tierney and Kraus, 2013]. All together, these event-based studies suggest that metrically strong positions are accompanied by larger source currents than metrically weak positions, and that these differences may be pre-attentive. It is worth noting, however, that by design these studies look at differences in predictions of not only when an auditory event is expected, but also what that auditory event should be [Teki and Kononowicz, 2016]. Behavioral evidence suggests that these two types of predictions may have distinct neural substrates [Morillon et al., 2016, Rajendran and Teki, 2016], and it is therefore not yet possible to say whether pre-attentive responses are a result of temporal expectation alone or a combination of expectations of what and when [Arnal, 2012, Arnal and Giraud, 2012, Schwartze et al., 2013].

5.3.2 Entrainment of oscillatory activity to musical beat

In addition to event-based descriptions of the beat percept, cortical oscillations have also been shown to reflect metrical structure. This is noteworthy because it suggests that neural oscillations, in addition to entraining to the rate of individual events in a rhythmic sequence [Doelling and Poeppel, 2015], are also able to entrain to higher-level temporal regularities, but the precise mechanism behind this is still unknown. Modulation of auditory cortical activity in the beta band has been shown to track the clicks of a metronome, while gamma oscillations appear to encode anticipated stimulus timing as evidenced by a peak in gamma activity even in the absence of a click [Fujioka et al., 2009]. Beta oscillations have also been demonstrated to encode beat and meter imagery [Iversen et al., 2009, Fujioka et al., 2015], and the dynamics of induced beta oscillatory activity both in humans [Teki, 2014] and in nonhuman primates [Bartolo et al., 2014, Bartolo and Merchant, 2015], have been shown to vary according to the temporal regularity of sound sequences. In addition to beta, gamma band oscillations also appear to encode beat and meter

[Snyder and Large, 2005, Zanto et al., 2006], and entrainment in the low-frequency delta-theta band (¡8 Hz) has also been shown to correlate with years of musical training [Doelling and Poeppel, 2015]. Low-frequency entrainment to the beat has also been observed in the bulk electroencephalogram signal [Nozaradan et al., 2011, Henry et al., 2014] (see [Zhou et al., 2016] for a guide on the interpretation of low-frequency components in the Fourier spectrum). A hierarchical organisation of oscillatory activity in the auditory cortex is thought to facilitate temporal processing of auditory stimuli and coordinate activity between sensory and other brain areas [Lakatos, 2005]. Cortical oscillations have furthermore been hypothesised to provide a mechanism for attentional selection and may be entrained by rhythmic auditory stimuli [Lakatos et al., 2008, Schroeder and Lakatos, 2009, Gomez-Ramirez et al., 2011, Lakatos et al., 2013].

5.3.3 Brain areas involved in beat perception

In addition to the auditory cortex, musical rhythms have been shown to engage a number of distributed brain areas, including several that would traditionally be considered part of the brain's motor system, and hence might not immediately be thought of as playing a key role in beat perception. These include the basal ganglia, supplementary motor area, striatum, cerebellum, sensorimotor cortex, and premotor cortex [Parsons, 2001, Grahn and Brett, 2007, Zatorre et al., 2007, Chen et al., 2008, Grahn, 2009, Teki et al., 2011]. Engagement of motor-related areas appears to be automatic since it is observed consistently even when listeners are instructed not to make overt movements [Chen et al., 2008]. Activation in auditory and motor areas furthermore correlates with individual differences in beat perception [Grahn and McAuley, 2009].

The activation of brain areas during beat perception depends on several factors including the duration of intervals [Lewis and Miall, 2003], temporal context [Teki et al., 2011], and task demands [Merchant et al., 2013]. The core timing areas

of the brain, specifically the striatum and the cerebellum [Ivry and Schlerf, 2008] are activated in perceptual timing depending on the temporal regularity of the sequences. For isochronous sequences, where a clear beat can be perceived, timing relies more on a network involving the striatum, while for jittered sequences, where the percept of a beat is negligible and intervals are encoded in an absolute manner, timing relies more on an olivocerebellar network [Teki et al., 2011, Teki et al., 2012]. Examination of individuals who exhibit beat deafness [Phillips-Silver et al., 2011], a rare condition that is associated with poor motor synchronisation and/or impoverished beat perception [Sowiński and Bella, 2013], provides further evidence that beat perception may recruit distinct circuits depending on the implicit/explicit timing aspect of the task [Bégel et al., 2017]. The dissociation of striatal and cerebellar responses for beat-based versus duration-based sequences has recently been observed to hold not only for the perception but also for working memory for single time intervals in sequences with different rhythmic structures [Teki and Griffiths, 2014, Teki and Griffiths, 2016].

Beat perception itself may be subcategorised into the processes of finding, continuing, and adjusting the beat, and the evidence points strongly towards the basal ganglia being involved in the continued representation of beat rather than its detection or adjustment [Chapin et al., 2010, Grahn and Rowe, 2013]. In one fMRI study [Chapin et al., 2010], participants were played six cycles of each of a set of complex rhythm and were tasked with attending to the rhythm, holding it in memory over 12 s, then reproducing it by tapping. During the attending phase, the basal ganglia showed significant activation only if the auditory stimulus was attended to, and if sufficient cycles of the rhythm had passed for listeners to perceive the beat. The basal ganglia also remained active during the rehearsal period. Similarly, in another fMRI study [Grahn and Rowe, 2013] where beat and nonbeat rhythms were played consecutively, the preceding rhythm determined whether the beat in the subsequent rhythm, if any, was a continuation from the previous rhythm (beat continuation),

was sped up or slowed down (beat adjustment), or needed to be found afresh (beat finding). Here, the basal ganglia were most active in beat continuation conditions and less active for beat adjustment conditions, with no apparent difference between the beat finding and the nonbeat (where no beat was present) conditions.

The superior temporal gyrus, premotor cortex, and venterolateral prefrontal cortex show activity during beat detection and synchronisation through tapping [Kung et al., 2013]. When tapping to rhythmic sequences that contain syncopation (the absence of sound on a perceived beat, see Figure 5.1), differences in activation of the premotor cortex, supplemental motor area, basal ganglia, and lateral cerebellum were observed, and these differences were present even when motor actions were not executed and the beat was simply imagined [Oullier et al., 2005]. Syncopation is among the factors that determine how engaging listeners find a piece of music, and pleasant music appears to more effectively entrain neural responses in the caudate nucleus of the basal ganglia [Trost et al., 2014]. Premotor and cerebellar areas are also more heavily recruited in response to subjectively more beautiful rhythms, and activity in the ventral premotor cortex (PMv) is enhanced by rhythms that are at a preferred tempo [Kornysheva et al., 2010b]. Repetitive transcranial magnetic stimulation (TMS) over the PMv changed people's preferred tempo, suggesting that the PMv may be involved in beat rate preference [Kornysheva et al., 2010a].

Findings from a number of functional imaging studies begin to allude back to some of the observations from early studies on temporal processing in the context of music. For example, beat induction is poorer for a slow (1500 ms) tempo compared to a faster one (600 ms), and activity in the basal ganglia, premotor and supplementary motor regions, and thalamus is correspondingly reduced [McAuley et al., 2012]. This is consistent with accounts that the motor system is preferentially engaged in the measurement of sub-second time intervals [Lewis and Miall, 2003]. Basal ganglia activity peaks for intervals around 500-600 ms [Riecker et al., 2003], which is comparable to the indif-

ference interval and the rate of spontaneous tapping in humans [Repp and Su, 2013]. The upper tempo limit to beat perception (approx. 200 ms) may be determined by the time constant for temporal integration [Loveless et al., 1996], which is comparable to the duration of auditory short term sensory memory, or short auditory store [Cowan, 1984]. Recent work, however, suggests that temporal memory resources may not be fixed for a discrete number of items but flexibly distributed according to the number of intervals to be encoded in a sequence [Teki and Griffiths, 2014, Joseph et al., 2015].

5.3.4 Model-based accounts of beat perception

A number of theoretical models have been proposed that capture neural and behavioral aspects of beat perception. Neural resonance theory is an influential computational model that consists of two sets of dynamic nonlinear oscillators, one that receives sensory input (an auditory layer) and one that receives input and projects back to the auditory layer (motor layer). The interaction between these layers can be modelled as a dynamical system, and the results resemble both neurophysiological and behavioral aspects of beat perception [Large et al., 2015]. Neural resonance theory is compatible with the dynamic attending theory, which postulates oscillatory fluctuations in attention [Large and Jones, 1999]. The active sensing hypothesis [Schroeder et al., 2010] postulates similar interactions between the auditory and motor system (see [Henry and Herrmann, 2014] for a comparison of the two hypotheses). The action simulation for auditory prediction (ASAP) hypothesis goes a step further by suggesting that auditory perception is sharpened by the explicit simulation of periodic movement in motor planning regions of the brain [Patel and Iversen, 2014]. The precise mechanism for beat induction remains unknown, though the entrainment of neural oscillations is a common thread between these competing hypotheses. Modulation filters, or neural populations with

a bandpass characteristic for the detection of amplitude modulation, which were initially theorised to computationally account for the brain's ability to detect periodic temporal structure [Dau et al., 1997a, Dau et al., 1997b], appear to be present at various stages in the auditory system [Joris et al., 2004, Schreiner and Urbas, 1986, Schreiner and Urbas, 1988, Kilgard and Merzenich, 1999] and have also been supported by some physiological evidence in the human brain [Xiang et al., 2013, Malone et al., 2015]. Neural modulation filters also provided a computational account for why rhythmic sound patterns were easier to detect in Chapter 2, and could therefore be a mechanism by which periodicity is extracted from temporally structured transients that are present in the adapted representation of rhythmic sounds. However, as current studies have only described modulation filters in the context of sinusoidally amplitude modulated (in other words, periodic) signals, the question of whether putative modulation filters would be able to detect and encode a periodicity present in signals that are not strictly periodic remains to be tested experimentally.

5.3.5 Beat processing in nonhuman species

So far, the discussion has focused on findings from human studies. Beat perception studies in nonhuman species are numerous, but apart from notable exceptions such as a cockatoo [Patel et al., 2009] and a sea lion [Cook et al., 2013, Rouse et al., 2016], nonhuman species have shown little compelling evidence of being able to perceive and synchronise to the beat as precisely as humans [Geiser et al., 2014]. Though chimpanzees appear to show some synchronisation ability [Hattori et al., 2013], it appears to be weak and quite limited in tempo range compared to that of the human. This may be somewhat surprising given that humans are not the only species to rely on rhythmic sounds such as vocalisations and produce rhythmic movements. Indeed, some signatures of rhythm perception in humans have also been observed in macaques, such as interval duration-selective modulation of beta oscillations [Bartolo et al., 2014].

In this and other related studies, the macaques were given a serial continuation task where they tapped along to a metronome that produced an isochronous series of clicks and continued tapping at the same rate after the metronome was stopped. Though tap times tended to lag metronome clicks by 100-250 ms, these lags were shorter than the macaque's reaction times, suggesting that there was a predictive element, though not strong enough to mimic the near-zero or negative lags in humans. Like in humans, beta oscillations in the basal ganglia (putamen) show preference for the continuation of a beat, and overall, similar timing circuits have been identified in both human and nonhuman primates, though macaques show better performance when synchronising their movements to a visual rather than auditory metronome [Merchant et al., 2015]. This is in contrast to a clear auditory bias in humans [Honing and Merchant, 2014]. Larger responses in primary auditory cortex to tones at strong beat positions in a rhythmic sequence than to the same tones in a rhythmically irregular sequence have also been observed in macaques, in addition to enhanced deviance detection ability [Selezneva et al., 2013]. However, this may be due to sensitivity to rhythmic grouping rather than to beat perception itself, since certain aspects of beat-specific neural activity observed in human adults and newborns are not observed in macaques [Honing et al., 2012].

5.3.6 To summarise

Human imaging studies have provided glimpses into the complex and highly distributed neural dynamics that are set into motion by musical rhythms. A key conceptual advance is the finding that rhythmic sequences engage auditory and motor areas more strongly than arrhythmic sequences, even during passive listening and in the absence of movement. Another is that perceptually strong beats evoke stronger neural activity than weak beats, which suggests a close link between neural activity and perception. Underlying both are oscillatory processes that are capable of en-

training to the beat, but are also coordinated across sensory, frontal, parietal, and motor-related areas both cortically and subcortically. Some of these neural dynamics have been observed in nonhuman primates, and it therefore remains an open question to what extent humans are unique in their ability to perceive musical beat, and what differences in connectivity and neural response dynamics give rise to humans' seemingly superior ability to spontaneously synchronise movements to music. These have proven difficult questions to answer since there is a wide methodological and philosophical gap between human and animal work related to beat perception. Approaches such as the ones developed in Chapters 2-4, as well as an awareness of parallels with work in other fields described in this chapter, may prove useful in overcoming some of these challenges.

5.4 Predictable timing in auditory perception

As alluded to at the beginning of this chapter, an appreciation of music is only one of many consequences of our ability to perceive rhythmic patterns. I will now begin to shift the focus away from music to explore rhythm perception in a more general context. Intrinsic to the perception of a musical beat is the prediction of when the next beat will occur, and the perceptual advantages afforded by our general ability to form temporal predictions will be the subject of this section.

5.4.1 Temporal predictability in pattern detection

Humans show a remarkable ability to detect repeating patterns that are quite complex in their acoustic content [Agus et al., 2010, McDermott and Simoncelli, 2011, Kumar et al., 2014, Barascud et al., 2016]. To do so is an impressive feat; the brain must be able to hold arbitrary sounds of arbitrary length and complexity in memory over timescales that can range from milliseconds up to tens of seconds [Kaernbach, 2004].

It is therefore relevant that a feature of repeating sounds in nature is that they tend to be rhythmic and indicative of animate sound sources. Rhythm detection may therefore be an advantageous sensory capability, and in addition to decreasing detection thresholds [Lawrance et al., 2014], I showed in Chapter 2 that rhythmic presentation of repeating sounds facilitates detection of complex acoustic patterns [Rajendran et al., 2016].

The entrainment of oscillatory activity in the brain, mentioned earlier in the context of beat perception, could be a likely explanation for these results too. Rhythmic input is widely thought to entrain attentional resources [Lakatos et al., 2008, Bolger et al., 2013, Calderone et al., 2014] such that neuronal excitability is highest when the next stimulus is predicted to occur [Lakatos et al., 2009, Besle et al., 2011]. Low frequency entrainment of oscillations may therefore serve as a mechanism for sensory selection [Schroeder and Lakatos, 2009] and improve the quality of sensory information received [Rohenkohl et al., 2014]. It is worth noting that the rhythmic form of temporal expectation is just one of several forms of temporal expectation, each resulting in subtle differences in perception that may arise from differences in the underlying neural substrates [Nobre et al., 2007, Breska and Deouell, 2017]. For example, an enhancement of perceptual sensitivity has been demonstrated in both periodic and non periodic sequences that are temporally predictable, but motor facilitation through faster response latencies were only observed in the periodic condition [Morillon et al., 2016, Rajendran and Teki, 2016].

However, it is also important to note that there is an as yet unresolved tension, or apparent conflict, in the physiological literature regarding the nature of the neural responses involved in the processing of periodic or rhythmic stimuli. The aforementioned studies posit that entrainment due to temporal expectation and attention would result in periods of heightened sensitivity in phase with the rhythm, which would be expected to lead to enhanced, stronger responses. This is in contrast to well

documented phenomena such as repetition suppression in auditory evoked responses measured through EEG [Costa-Faidella et al., 2011] and stimulus specific adaptation (SSA) observed in neural responses recorded extracellularly in auditory cortex and non-lemniscal parts of the inferior colliculus and thalamus [Malmierca, 2014, Nieto-Diego and Malmierca, 2016, Khouri and Nelken, 2015], which find that responses to simple periodic stimuli are reduced or suppressed, rather than enhanced. How can isochronous stimuli on the one hand produce entrainment that is suggestive of periodically heightened sensitivity but at the same time elicit reduced response amplitudes as evidenced by SSA or repetition suppression? The simple answer is that we do not yet know. The methodologies of studies of entrainment versus SSA are too different to allow direct comparisons. Entrainment studies typically use EEG or LFP measures, the amplitude of which depends at least as much on the degree of synchronisation of neural activity as on net response amplitudes of individual neurons. Additionally, they are often carried out on awake human volunteers or animal subjects who may be attending to the rhythmic sounds, while SSA studies typically use anesthetised preparations to measure extracellular response amplitudes that are essentially independent of neural synchrony. Consequently, while the take-home messages from studies of entrainment and of SSA at present appear somewhat contradictory, how they may be reconciled will need to be addressed in future studies using unified methodologies.

5.4.2 Temporal predictability in auditory scene analysis

Another practical advantage of forming predictions based on temporal patterns is that it allows us to parse a complex auditory scene into distinct perceptual objects [Winkler et al., 2009a]. In addition to temporal coherence of sound features [Turgeon et al., 2002, Turgeon et al., 2005, Shamma et al., 2011], the predictability of features such as location, pitch, loudness, and timbre play a pivotal role in auditory scene analysis [Bendixen, 2014]. The segregation of a set of sounds from another set of

sounds is known as auditory stream segregation and has often been probed experimentally using an alternating tone paradigm of A-B-A, where A and B tones are typically different frequencies of a certain frequency separation [Bregman and Campbell, 1971, van Noorden, 1975]. Temporal regularity within these paradigms influences whether these sequences are perceived as integrated (A-B-A-B-A) or whether they segregate into two perceptually distinct streams (A—A—A and –B—B) [Bendixen et al., 2010, Andreou et al., 2011, Rajendran et al., 2013]. Together with attentional effects, predictive coding based on temporal and other feature regularities may account for the stability of auditory objects [Pressnitzer and Hupe, 2006, Denham and Winkler, 2006, Chait et al., 2007, Winkler et al., 2012].

Current theories suggest that the formation of auditory objects may rely on both basic sensory neural mechanisms [Fishman et al., 2012] and attention-driven oscillatory mechanisms [Lakatos et al., 2008, Schroeder and Lakatos, 2009]. Though the question of how a perceptual object is formed from potentially noisy and conflicting information is still an open one, the final representation of an auditory object is remarkably distinct and robust, even if it overlaps spectrotemporally with the unattended background [Ding and Simon, 2012, Ding and Simon, 2013]. A key question here, which may also be relevant to how beat and meter emerge from music, is whether and how different oscillatory populations of neurons entrain to different time-varying sound features, and how their relative contributions are weighted and integrated to form a coherent percept of, for example, a single person speaking in a noisy room.

5.4.3 Rhythms in speech perception

Speech is perhaps the most pervasive and critical context in which we rely on our ability to derive meaning from complex temporal patterns. The intelligibility of speech has been shown to correlate with the entrainment of oscillatory neural responses to the speech envelope [Ahissar et al., 2001, Peelle and Davis, 2012], partic-

ularly in the 4-8 Hz range [Luo and Poeppel, 2007]. This range corresponds to the syllable rate of speech production [Greenberg et al., 1999] and dominates the temporal modulations in the speech envelope [Chi et al., 1999, Chandrasekaran et al., 2009, Elliott and Theunissen, 2009]. The syllabic rate is nearly an order of magnitude slower than fine structure elements such as formants (30-50 Hz), and a few-fold faster than intonation contours that are typical of phrasal units (1-2 Hz). Content at all of these timescales are parsed concurrently to extract meaning.

Speech, like music, is built hierarchically from elements that span short to long timescales. A recent survey of temporal modulations in speech and music reveals a consistent peak in temporal modulations around 5 Hz for speech across nine languages, and around 2 Hz for music across several (Western) musical genres [Ding et al., 2017]. It is worth emphasising, however, that the temporal structure present in speech is not periodic like it is in music [Nolan and Jeon, 2014]. The temporal modulations in speech are constrained by the motor system though, specifically by the biomechanics of the articulators, and this results in clear temporal structure in both the auditory and visual components to speech [Chandrasekaran et al., 2009]. There is strong evidence that speech contains sufficient temporal structure to robustly entrain oscillatory neural activity [Giraud and Poeppel, 2012], and that this entrainment serves to maximise processing efficiency of future inputs by ensuring that intervals of high neuronal excitability coincide with when critical information is expected to arrive [Peelle and Davis, 2012, Ding et al., 2017].

Interestingly, temporal manipulations to speech more drastically impair intelligibility [Adank and Janse, 2009] than extreme spectral manipulations do [Shannon et al., 1995]. Model-based accounts [Ghitza, 2011, Giraud and Poeppel, 2012] suggest that phase-locking and nested theta-gamma oscillations could explain why an extremely impoverished speech signal can be understood if the syllabic rhythm is preserved [Ghitza and Greenberg, 20
The asymmetric sampling in time (AST) hypothesis suggests that the two cerebral

hemispheres sample an auditory signal at different rates; the left auditory areas extract information from 20-40 ms temporal integration windows, while auditory areas in the right hemisphere sample using 150-250 ms temporal integration windows [Poeppel, 2003]. A related hypothesis suggests that the left hemisphere has better temporal resolution and the right hemisphere has better spectral resolution, and that this functional organisation reflects an optimisation of processing for speech and music, respectively [Zatorre et al., 2002]. Both of these ideas are consistent with the observation that the left hemisphere dominates during speech processing while the right hemisphere dominates during music processing [Tervaniemi and Hugdahl, 2003]. The parallels drawn here between music and speech deal strictly with timing and do not suggest that music has any meaning that is analogous to the semantic meaning of speech [Lerdahl and Jackendoff, 1983]. However, given these parallels, it is possible that music and speech co-evolved [Fitch, 2000, Hauser and McDermott, 2003, Fitch, 2006, Patel, 2007] and are built on overlapping circuit mechanisms for auditory working memory [Hickok et al., 2003, Joseph et al., 2015] and timing [Patel, 2011].

5.4.4 To summarise

The temporal predictability that results from rhythmic stimulation helps us detect basic patterns, which ultimately helps us parse an auditory scene into distinct auditory objects and understand speech. Entrainment of neural oscillations, which by virtue of aligning to temporal modulations in a rhythmic acoustic signal generates predictions about future input, is thought to underlie all of these abilities. The acoustic stimuli used in these studies range from extremely simple alternating tone paradigms, to the parsing of two people speaking simultaneously. Much remains to be understood about which periodic or quasi-periodic features in a spectrotemporally complex sound entrain oscillations, and how such oscillations are ultimately integrated to form distinct auditory objects or extract meaning. Answering these questions would also likely

shed light on the mechanisms and functional roles of oscillatory entrainment in the context of musical rhythm perception.

5.5 Neural mechanisms of timing

Music, speech, and the parsing of complex auditory scenes all rely on an ability to detect temporal regularities in order to form the temporal predictions that drive how sounds in the future are perceived. This requires some form of timekeeping in the brain. The timing field is vast and is a likely reflection of the complexity of the neural circuits that are able to track time [Teki, 2016]. Those findings most relevant to this discussion are reviewed here.

5.5.1 Dedicated timekeeping circuits?

The neuronal mechanisms underlying timing have been a subject of investigation for several decades. Braitenberg (1967) [Braitenberg, 1967] proposed the cerebellum as a biological clock in the millisecond range. Since then, the concept of a central clock or internal timekeeper has dominated timing research. Early work highlighted the unique synaptic circuitry of the cerebellum and the inferior olive as being capable of generating precise timing signals. Specifically, inferior olive neurons, which provide climbing fiber input to the Purkinje cells in the cerebellum, possess unique voltage-gated conductances that exhibit rhythmic sub-threshold membrane potential oscillations (5-15 Hz) as well as electrical gap-junctions that synchronise membrane potential oscillations across cells into distinct neuronal clusters that show temporally coherent activity [Llinas et al., 1974, Llinas and Yarom, 1981]. The deep cerebellar nuclei like the dentate nucleus modulate the electrical activity of olivary neurons and decouple them into dynamic cell assemblies. Furthermore, these deep cerebellar nuclei are inhibited by the Purkinje cells, completing a feed-forward inhibitory loop. These neurophysi-

ological properties provide the olivocerebellar network with the capacity to generate accurate absolute timing signals for motor and perceptual timing [Welsh et al., 1995, Yarom and Cohen, 2002, Jacobson et al., 2008, Mathy et al., 2009]. The use of timing signals from the olivocerebellar network has been demonstrated across several timing paradigms in human studies as well [Xu, 2006, Teki et al., 2011, Wu et al., 2011, Lusk et al., 2016].

Motivated by neuropsychological evidence from Parkinson's patients who showed perceptual timing deficits, parallel work focused on the basal ganglia as a core timing network in the brain [Artieda et al., 1992]. Matell and Meck (2004) [Matell and Meck, 2004] proposed an oscillatory timing model: medium spiny neurons in the dorsal striatum act as coincidence detectors of oscillatory cortical activity (5-15 Hz; [Miall, 1989]). The cortical oscillations are proposed to be synchronised at interval onset by phasic dopamine release from the ventral tegmental area, while dopaminergic input from the substantia nigra modulates the activity of the dorsal striatum [Buhusi and Meck, 2005]. Cortico-striatal synapses are strengthened or weakened over experience through long-term potentiation and depression, and after repeated stimulus presentation, medium spiny neurons learn to encode the duration of reinforced time intervals [Gu et al., 2011]. Several studies point to the importance of the dopaminergic basal ganglia network in mediating accurate timing signals [Jin et al., 2009, Bartolo et al., 2014, Gershman et al., 2014, Chiba et al., 2015, Mello et al., 2015, Gouvêa et al., 2015, Soares et al., 2016].

Cortical networks such as primary visual, auditory, parietal and frontal cortices have also been implicated in sensory timing functions (e.g. [Bueti et al., 2008, Bueti and Macaluso, Bueti, 2011, Leon and Shadlen, 2003, Schneider and Ghose, 2012, Hayashi et al., 2015, Jazayeri and Shadlen, 2015, Namboodiri et al., 2015, Bakhurin et al., 2016, Shuler, 2016]). However, it is not completely understood what aspects of timing are respectively mediated by each of these networks, nor are the dynamics of temporal processing across sensory and higher-order cortical networks completely clear (see [Rao et al., 2001]).

The likely hypothesis is that early stage sensory cortices process the stimulus related features while parietal and frontal cortices are engaged by task demands like memory and attention [Finnerty et al., 2015]. More recently, the hippocampus (CA1) has been shown to have time cells' that display increased firing rates in relation to elapsing durations, independent of space and distance [MacDonald et al., 2011]. The prevailing view suggests the existence of time cells' in the striatum, cerebellum and the hippocampus whose output is integrated to obtain a common percept of time [Lusk et al., 2016].

5.5.2 Is the passage of time implicit in neural responses?

An alternative hypothesis to the one positing that time is kept by dedicated sensorimotor circuits is one that suggests that timing is an intrinsic computation that emerges from network-wide neural dynamics [Goel and Buonomano, 2014]. Hardy and Buonomano (2016) [Hardy and Buonomano, 2016] have recently reviewed a number of plausible neurocomputational models of timing. One of the simplest network models of timing is based on synfire chains' where groups of neurons are connected in a feed-forward fashion such that each neuronal population is activated at a different instant in time [Haß et al., 2008]. Synfire chains represent a neurobiologically plausible mechanism for interval timing but are limited because of their pure feed-forward architecture and absence of recurrent connections. Positive feedback models, on the other hand, use recurrent excitatory connections and are compatible with experimental findings on cortical representation of time (e.g. [Namboodiri et al., 2015]). The limitation of these models, however, is that it is not known whether these can be generalised to sequences of intervals. Finally, state-dependent networks of timing and temporal processing are based on the hypothesis that sensory events interact with current states of recurrent networks to form a sequence of network states that encode each event in the context of recent stimulus history [Karmarkar and Buonomano, 2007].

Several studies have demonstrated that cortical networks can be trained to represent time intervals in the hundreds of millisecond range where timing is proposed to emerge from network-wide and pathway-specific changes in evoked neural dynamics (e.g. [Goel and Buonomano, 2016]). It is also well known that neurons in sensory cortices show multiplexed responses to different stimulus features [Walker et al., 2011] as well as temporal scales [Panzeri et al., 2010]. The number of stations information passes through as it ascends the auditory system, combined with the tendency for adaptation to operate over longer and longer timescales the higher up in the auditory system we look, could mean that neurons in higher auditory areas may already have a reach into the past stretching far enough to span the rhythm timescale [Schönwiesner and Zatorre, 2009].

5.5.3 To summarise

Although the notion of population clocks is gaining traction [Hardy and Buonomano, 2016], there is no compelling biologically plausible model that generalises these results from studies based on computation of single time intervals to complex sequences such as those observed in music. Natural motor commands as well as sound sequences like speech and music consist of dynamically varying time intervals with different temporal structures. Several of the circuits reviewed above are specialised for processing time on the scale of tens of milliseconds to a few seconds, but it is not yet clear which of these mechanisms apply in the context of beat perception as this has not directly been tested. Integrating basic mechanisms of sound processing observed along the auditory pathways with models of timing may provide some novel insights into the examination of pattern timing. It may also be worth exploring in more detail any parallels between network models of timing and computational models of beat perception such as neural resonance theory [Large et al., 2015].

Timing functions are distributed across the brain and are expressed in subcorti-

cal motor structures like the basal ganglia and the cerebellum, sensory and motor cortices, as well as higher-order areas like the parietal and frontal cortical networks. These areas overlap to a large extent with the brain areas implicated in beat perception and in general auditory temporal processing. While the timing computations performed by of each these individual brain regions is not fully understood, it is evident that particular areas are specialised for mediating specific timing functions. Future research may benefit from dissecting the precise role within each brain area and as a part of a distributed timing network.

5.6 Conclusions and outlook

A lot of ground has been covered in this discussion because the work contained in this thesis touches on general topics that are relevant to several different fields of research, even though these fields have historically shown relatively little overlap. This is despite these fields sharing common themes that unite them. For example, the timescales that are relevant in music are also relevant in other contexts such as in the production and perception of movement (walking, running, breathing) and speech, and in the parsing of complex acoustic scenes (see Figure 5.1B). Furthermore, the entrainment of neural oscillations through sensorimotor loops may be a central mechanism governing perception and action in all of these contexts. By presenting an overview of these diverse topics that likely rely on similar temporal encoding mechanisms, I hope to provide a thorough contextualisation of the results contained in this thesis as well as an insightful point of departure for future investigations into auditory temporal sequence processing.

An important direction for future research that I believe will be pivotal to advancing our understanding of auditory temporal processing is arriving at a mechanistic understanding of the entrainment of neural oscillations. While a large body of work

points to the importance of neural oscillations (the studies mentioned in this chapter only scratch the surface), this topic is nevertheless not without controversy with many questions that remain unresolved, starting with the functional role that oscillations in different frequency bands play in information coding and retrieval. A number of theories have been proposed that describe functional aspects of oscillatory dynamics, including communication between neuronal groups through coherence of oscillations [Fries, 2015], the prioritisation of sensory streams through pulsed inhibition via alpha oscillations [Haegens et al., 2011, Jensen et al., 2014, Strauss et al., 2014, Mathewson et al., 2011] the retrieval of memories through spiking that is phase-locked to theta oscillations [Hsieh and Ranganath, 2014], and the active sensing of sound through rhythmic temporal priors provided by the motor system [Morillon et al., 2015]. Of particular relevance are the behavioural [Morillon et al., 2016] and neuronal dissociations [Breska and Deouell, 2017] that have been observed between auditory sequences that are periodic versus temporally predictable but not periodic, suggesting that the underlying neural dynamics manifest differently according to the nature of the temporal predictions being maintained. Further work is required to develop a unified understanding of the function served by the entrainment of neural oscillations.

A second question relates to the dynamics of entrainment itself, specifically how entrainment arises and unfolds, how it extracts higher-order temporal regularities in a rhythmic sequence, how it behaves in response to new sensory input, and how possibly different and simultaneous processes interact to guide what we perceive. Much of our current knowledge about the role of neural oscillations and entrainment in the perception of temporally structured stimuli is based on the interpretation of data obtained with non-invasive techniques (EEG, MEG, fMRI), which lack the fine resolution required to provide insights into these phenomena at the level of individual neurons and neural networks. Attaining deeper insights will require data obtained at higher spatial and temporal resolution, as is typically obtained from invasive record-

ings in animal models, but the use of richly structured auditory stimuli such as music in animal electrophysiology experiments remains a highly unusual thing to do. The results contained in this thesis attempt to break ground in this direction. Employing music in addition to traditional stimulus paradigms may be especially fruitful because, as described in this chapter, much is already known about our perception of music and because it is a finely-controllable stimulus paradigm within which nested periodicities across different sound features (frequency, loudness, duration) can be simultaneously present and tuned. Furthermore, since nonhuman model organisms do show some capacity to perceive and discriminate rhythms, and since recognising rhythmic patterns in environmental sounds such as footsteps or vocalisations is of great importance to a wide range of organisms, complementary studies in nonhuman species such as those in Chapter 3 and 4 should begin fill in the gaps in our knowledge that non-invasive psychoacoustic and physiological studies on humans (e.g. Chapter 2) can inform, but alone cannot answer.

Bibliography

[Abecasis, 2005] Abecasis, D. (2005). Differential Brain Response to Metrical Accents in Isochronous Auditory Sequences. *Music Perception: An Interdisciplinary Journal*, 22(3):549–562.

[Adank and Janse, 2009] Adank, P. and Janse, E. (2009). Perceptual learning of time-compressed and natural fast speech. *The Journal of the Acoustical Society of America*, 126(5):2649–2659.

[Agus et al., 2010] Agus, T. R., Thorpe, S. J., and Pressnitzer, D. (2010). Rapid Formation of Robust Auditory Memories: Insights from Noise. *Neuron*, 66(4):610–618.

[Ahissar et al., 2001] Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., and Merzenich, M. M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proceedings of the National Academy of Sciences*, 98(23):13367–13372.

[Andreou et al., 2011] Andreou, L.-V., Kashino, M., and Chait, M. (2011). The role of temporal regularity in auditory segregation. *Hearing Research*, 280(1-2):8–8.

[Arnal, 2012] Arnal, L. H. (2012). Predicting "When" Using the Motor System's Beta-Band Oscillations. *Frontiers in human neuroscience*, 6.

[Arnal and Giraud, 2012] Arnal, L. H. and Giraud, A.-L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences*, 16(7):390–398.

[Artieda et al., 1992] Artieda, J., Pastor, M. A., Lacruz, F., and Obeso, J. A. (1992). Temporal Discrimination is Abnormal in Parkinson's Disease. *Brain*, 115(1):199–210.

[Aschersleben, 2002] Aschersleben, G. (2002). Temporal Control of Movements in Sensorimotor Synchronization. *Brain and cognition*, 48(1):66–79.

[Atencio et al., 2012] Atencio, C. A., Sharpee, T. O., and Schreiner, C. E. (2012). Receptive field dimensionality increases from the auditory midbrain to cortex. *Journal of Neurophysiology*, 107(10):2594–2603.

[Bååth, 2015] Bååth, R. (2015). Subjective Rhythmization. *Music Perception: An Interdisciplinary Journal*, 33(2):244–254.

[Bakhurin et al., 2016] Bakhurin, K. I., Goudar, V., Shobe, J. L., Claar, L. D., Buonomano, D. V., and Masmanidis, S. C. (2016). Differential encoding of time by prefrontal and striatal network dynamics. *The Journal of Neuroscience*, pages 1789–16.

[Barascud et al., 2016] Barascud, N., Pearce, M. T., Griffiths, T. D., Friston, K. J., and Chait, M. (2016). Brain responses in humans reveal ideal observer-like sensitivity to complex acoustic patterns. *Proceedings of the National Academy of Sciences*, 113(5):E616–E625.

[Bartolo and Merchant, 2015] Bartolo, R. and Merchant, H. (2015). beta oscillations are linked to the initiation of sensory-cued movement sequences and the internal guidance of regular tapping in the monkey. *The Journal of Neuroscience*, 35(11):4635–4640.

[Bartolo et al., 2014] Bartolo, R., Prado, L., and Merchant, H. (2014). Information Processing in the Primate Basal Ganglia during Sensory-Guided and Internally Driven Rhythmic Tapping. *The Journal of Neuroscience*, 34(11):3910–3923.

[Bégel et al., 2017] Bégel, V., Benoit, C.-E., Correa, Á., Cutanda, D., Kotz, S. A., and Bella, S. D. (2017). "Lost in time" but still moving to the beat. *Neuropsychologia*, 94(C):129–138.

[Bendixen, 2014] Bendixen, A. (2014). Predictability effects in auditory scene analysis: a review. *Frontiers in neuroscience*, 8(143):60.

[Bendixen et al., 2010] Bendixen, A., Denham, S. L., Gyimesi, K., and Winkler, I. (2010). Regular patterns stabilize auditory streams. *The Journal of the Acoustical Society of America*, 128(6):3658.

[Besle et al., 2011] Besle, J., Schevon, C. A., Mehta, A. D., Lakatos, P., Goodman, R. R., McKhann, G. M., Emerson, R. G., and Schroeder, C. E. (2011). Tuning of the Human Neocortex to the Temporal Dynamics of Attended Events. *The Journal of Neuroscience*, 31(9):3176–3185.

[Böck, 2016] Böck, S. (2016). Event Detection in Musical Audio. *Doctoral Thesis*.

[Bolger et al., 2013] Bolger, D., Trost, W., and Schön, D. (2013). Rhythm implicitly affects temporal orienting of attention across modalities. *Acta Psychologica*, 142(2):238–244.

[Bolton, 1894] Bolton, T. L. (1894). Rhythm. *The American Journal of Psychology*, 6(2):145.

[Bouwer et al., 2014] Bouwer, F. L., van Zuijen, T. L., and Honing, H. (2014). Beat Processing Is Pre-Attentive for Metrically Simple Rhythms with Clear Accents: An ERP Study. *PLoS ONE*, 9(5):e97467.

[Braitenberg, 1967] Braitenberg, V. (1967). Is the Cerebellar Cortex a Biological Clock in the Millisecond Range? In *The Cerebellum*, pages 334–346. Elsevier.

[Bregman and Campbell, 1971] Bregman, A. S. and Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of experimental psychology*, 89(2):244–249.

[Breska and Deouell, 2017] Breska, A. and Deouell, L. Y. (2017). Neural mechanisms of rhythm-based temporal prediction: Delta phase-locking reflects temporal predictability but not rhythmic entrainment. *PLOS Biology*, 15(2):e2001665–30.

[Brochard et al., 2003] Brochard, R., Abecasis, D., Potter, D., Ragot, R., and Drake, C. (2003). The "Ticktock" of Our Internal Clock: Direct Brain Evidence of Subjective Accents in Isochronous Sequences. *Psychological science*, 14(4):362–366.

[Bueti, 2011] Bueti, D. (2011). The Sensory Representation of Time. *Frontiers in Integrative Neuroscience*, 5.

[Bueti et al., 2008] Bueti, D., Bahrami, B., and Walsh, V. (2008). Sensory and Association Cortex in Time Perception. *Journal of cognitive neuroscience*, 20(6):1054–1062.

[Bueti and Macaluso, 2010] Bueti, D. and Macaluso, E. (2010). Auditory temporal expectations modulate activity in visual cortex. *NeuroImage*, 51(3):1168–1183.

[Buhusi and Meck, 2005] Buhusi, C. V. and Meck, W. H. (2005). What makes us tick? Functional and neural mechanisms of interval timing. *Nature Reviews Neuroscience*, 6(10):755–765.

[Bulkin and Groh, 2011] Bulkin, D. A. and Groh, J. M. (2011). Systematic mapping of the monkey inferior colliculus reveals enhanced low frequency sound representation. *Journal of Neurophysiology*, 105(4):1785–1797.

[Calderone et al., 2014] Calderone, D. J., Lakatos, P., Butler, P. D., and Castellanos, F. X. (2014). Entrainment of neural oscillations as a modifiable substrate of attention. *Trends in Cognitive Sciences*, 18(6):300–309.

[Chait et al., 2007] Chait, M., Poeppel, D., de Cheveigné, A., and Simon, J. Z. (2007). Processing asymmetry of transitions between order and disorder in human auditory cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 27(19):5207–5214.

[Chandrasekaran et al., 2009] Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., and Ghazanfar, A. A. (2009). The Natural Statistics of Audiovisual Speech. *PLoS Computational Biology*, 5(7):e1000436.

[Chapin et al., 2010] Chapin, H. L., Zanto, T., Jantzen, K. J., Kelso, S. J. A., Steinberg, F., and Large, E. W. (2010). Neural Responses to Complex Auditory Rhythms: The Role of Attending. *Frontiers in psychology*, 1.

[Chemin et al., 2014] Chemin, B., Mouraux, A., and Nozaradan, S. (2014). Body Movement Selectively Shapes the Neural Representation of Musical Rhythms. *Psychological science*, 25(12):0956797614551161–2159.

[Chen et al., 2008] Chen, J. L., Penhune, V. B., and Zatorre, R. J. (2008). Moving on Time: Brain Network for Auditory-Motor Synchronization is Modulated by Rhythm Complexity and Musical Training. *dx.doi.org*, 20(2):226–239.

[Chi et al., 1999] Chi, T., Gao, Y., Guyton, M. C., Ru, P., and Shamma, S. (1999). Spectro-temporal modulation transfer functions and speech intelligibility. *The Journal of the Acoustical Society of America*, 106(5):2719–2732.

[Chiba et al., 2015] Chiba, A., Oshio, K. i., and Inase, M. (2015). Neuronal representation of duration discrimination in the monkey striatum. *Physiological Reports*, 3(2):e12283–e12283.

[Clarke, 1999] Clarke, E. F. (1999). Rhythm and Timing in Music. In *The Psychology of Music*, pages 473–500. Elsevier.

[Cook et al., 2013] Cook, P., Rouse, A., Wilson, M., and Reichmuth, C. (2013). A California sea lion (Zalophus californianus) can keep the beat: Motor entrainment to rhythmic auditory stimuli in a non vocal mimic. *Journal of Comparative ...*, 127(4):412–427.

[Correa et al., 2006] Correa, Á., Lupiáñez, J., Madrid, E., and Tudela, P. (2006). Temporal attention enhances early visual processing: A review and new evidence from event-related potentials. *Brain Research*, 1076(1):116–128.

[Costa-Faidella et al., 2011] Costa-Faidella, J., Baldeweg, T., Grimm, S., and Escera, C. (2011). Interactions between "What" and "When" in the Auditory System: Temporal Predictability Enhances Repetition Suppression. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 31(50):18590–18597.

[Cowan, 1984] Cowan, N. (1984). On short and long auditory stores. *Psychological bulletin*, 96(2):341–370.

[Dau et al., 1997a] Dau, T., Kollmeier, B., and Kohlrausch, A. (1997a). Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers. 102(5 Pt 1):2892–2905.

[Dau et al., 1997b] Dau, T., Kollmeier, B., and Kohlrausch, A. (1997b). Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration. *The Journal of the Acoustical Society of America*, 102(5 Pt 1):2906–2919.

[Dean et al., 2005] Dean, I., Harper, N. S., and McAlpine, D. (2005). Neural population coding of sound level adapts to stimulus statistics. *Nature Neuroscience*, 8(12):1684–1689.

[Dean et al., 2008] Dean, I., Robinson, B. L., Harper, N. S., and McAlpine, D. (2008). Rapid Neural Adaptation to Sound Level Statistics. *The Journal of Neuroscience*, 28(25):6430–6438.

[Denham and Winkler, 2006] Denham, S. L. and Winkler, I. (2006). The role of predictive models in the formation of auditory streams. *Journal of physiology, Paris*, 100(1-3):154–170.

[Ding et al., 2017] Ding, N., Patel, A. D., Chen, L., Butler, H., Luo, C., and Poeppel, D. (2017). Temporal modulations in speech and music. *Neuroscience & Biobehavioral Reviews*, pages 1–7.

[Ding and Simon, 2012] Ding, N. and Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proceedings of the National Academy of Sciences*, 109(29):11854–11859.

[Ding and Simon, 2013] Ding, N. and Simon, J. Z. (2013). Adaptive Temporal Encoding Leads to a Background-Insensitive Cortical Representation of Speech. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 33(13):5728–5735.

[Doelling and Poeppel, 2015] Doelling, K. B. and Poeppel, D. (2015). Cortical entrainment to music and its modulation by expertise. *Proceedings of the National Academy of Sciences of the United States of America*, 112(45):E6233–42.

[Downie and Welge, 2017] Downie, J. S. and Welge, M. (2017). MIREX 2006: Audio Beat Tracking.

[Elliott and Theunissen, 2009] Elliott, T. M. and Theunissen, F. E. (2009). The Modulation Transfer Function for Speech Intelligibility. *PLoS Computational Biology*, 5(3):e1000302.

[Essens and Povel, 1985] Essens, P. J. and Povel, D. J. (1985). Metrical and nonmetrical representations of temporal patterns. *Perception & Psychophysics*, 37(1):1–7.

[Finnerty et al., 2015] Finnerty, G. T., Shadlen, M. N., Jazayeri, M., Nobre, A. C., and Buonomano, D. V. (2015). Time in Cortical Circuits. *The Journal of Neuroscience*, 35(41):13912–13916.

[Fishman et al., 2012] Fishman, Y. I., Micheyl, C., and Steinschneider, M. (2012). Neural mechanisms of rhythmic masking release in monkey primary auditory cortex: implications for models of auditory scene analysis. *Journal of Neurophysiology*, 107(9):2366–2382.

[Fitch, 2000] Fitch, W. T. (2000). The evolution of speech: a comparative review. *Trends in Cognitive Sciences*, 4(7):258–267.

[Fitch, 2006] Fitch, W. T. (2006). The biology and evolution of music: A comparative perspective. *Cognition*, 100(1):173–215.

[Fraisse, 1963] Fraisse, P. (1963). *The psychology of time.* Harper & Row.

[Fraisse, 1978] Fraisse, P. (1978). *Time and rhythm perception. in e. carterette & m. friedman (eds.), Handbook of perception VIII (pp. 203-254).* New York: Academic-Press.

[Fraisse, 1982] Fraisse, P. (1982). Rhythm and Tempo. In *The Psychology of Music.* Elsevier.

[Fraisse et al., 1958] Fraisse, P., Oléron, G., and Paillard, J. (1958). Sur les repères sensoriels qui permettent de contrôler les mouvements d'accompagnement de stimuli périodiques. *L'année psychologique*, 58(2):321–338.

[Frenkel-Toledo et al., 2005] Frenkel-Toledo, S., Giladi, N., Peretz, C., Herman, T., Gruendlinger, L., and Hausdorff, J. M. (2005). Treadmill walking as an external pacemaker to improve gait rhythm and stability in Parkinson's disease. *Movement Disorders*, 20(9):1109–1114.

[Fries, 2015] Fries, P. (2015). Rhythms for Cognition: Communication through Coherence. *Neuron*, 88(1):220–235.

[Fujioka et al., 2014] Fujioka, T., Fidali, B. C., and Ross, B. (2014). Neural correlates of intentional switching from ternary to binary meter in a musical hemiola pattern. *Frontiers in psychology*, 5(31):549.

[Fujioka et al., 2015] Fujioka, T., Ross, B., and Trainor, L. J. (2015). Beta-Band Oscillations Represent Auditory Beat and Its Metrical Hierarchy in Perception and Imagery. *The Journal of Neuroscience*, 35(45):15187–15198.

[Fujioka et al., 2009] Fujioka, T., Trainor, L. J., Large, E. W., and Ross, B. (2009). Beta and Gamma Rhythms in Human Auditory Cortex during Musical Beat Processing. *Annals of the New York Academy of Sciences*, 1169(1):89–92.

[Geiser et al., 2012] Geiser, E., Notter, M., and Gabrieli, J. D. E. (2012). A corticostriatal neural system enhances auditory perception through temporal context processing. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 32(18):6177–6182.

[Geiser et al., 2010] Geiser, E., Sandmann, P., Jäncke, L., and Meyer, M. (2010). Refinement of metre perception - training increases hierarchical metre processing. *European Journal of Neuroscience*, 32(11):1979–1985.

[Geiser et al., 2014] Geiser, E., Walker, K. M. M., and Bendor, D. (2014). Global timing: a conceptual framework to investigate the neural basis of rhythm perception in humans and non-human species. *Frontiers in psychology*, 5:159.

[Gershman et al., 2014] Gershman, S. J., Moustafa, A. A., and Ludvig, E. A. (2014). Time representation in reinforcement learning models of the basal ganglia. *Frontiers in Computational Neuroscience*, 7.

[Ghitza, 2011] Ghitza, O. (2011). Linking Speech Perception and Neurophysiology: Speech Decoding Guided by Cascaded Oscillators Locked to the Input Rhythm. *Frontiers in psychology*, 2.

[Ghitza and Greenberg, 2009] Ghitza, O. and Greenberg, S. (2009). On the Possible Role of Brain Rhythms in Speech Perception: Intelligibility of Time-Compressed Speech with Periodic and Aperiodic Insertions of Silence. *Phonetica*, 66(1-2):113–126.

[Giraud and Poeppel, 2012] Giraud, A.-L. and Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience*, 15(4):511–517.

[Goel and Buonomano, 2014] Goel, A. and Buonomano, D. V. (2014). Timing as an intrinsic property of neural networks: evidence from in vivo and in vitro experiments. *Phil. Trans. R. Soc. B*, 369(1637):20120460–20120460.

[Goel and Buonomano, 2016] Goel, A. and Buonomano, D. V. (2016). Temporal Interval Learning in Cortical Cultures Is Encoded in Intrinsic Network Dynamics. *Neuron*, 91(2):320–327.

[Goldberger et al., 2000] Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C. K., and Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation*, 101(23):E215–20.

[Gomez-Ramirez et al., 2011] Gomez-Ramirez, M., Kelly, S. P., Molholm, S., Sehatpour, P., Schwartz, T. H., and Foxe, J. J. (2011). Oscillatory Sensory Selection Mechanisms during Intersensory Attention to Rhythmic Auditory and Visual Inputs: A Human Electrocorticographic Investigation. *The Journal of Neuroscience*, 31(50):18556–18567.

[Gouvêa et al., 2015] Gouvêa, T. S., Monteiro, T., Motiwala, A., Soares, S., Machens, C., and Paton, J. J. (2015). Striatal dynamics explain duration judgments. *eLife*, 4:2473.

[Grahn, 2009] Grahn, J. A. (2009). The Role of the Basal Ganglia in Beat Perception. *Annals of the New York Academy of Sciences*, 1169(1):35–45.

[Grahn and Brett, 2007] Grahn, J. A. and Brett, M. (2007). Rhythm and Beat Perception in Motor Areas of the Brain. *Journal of cognitive neuroscience*, 19(5):893–906.

[Grahn and McAuley, 2009] Grahn, J. A. and McAuley, J. D. (2009). Neural bases of individual differences in beat perception. *NeuroImage*, 47(4):1894–1903.

[Grahn and Rowe, 2013] Grahn, J. A. and Rowe, J. B. (2013). Finding and feeling the musical beat: striatal dissociations between detection and prediction of regularity. *Cerebral cortex (New York, N.Y. : 1991)*, 23(4):913–921.

[Greenberg et al., 1999] Greenberg, S., Arai, T., Kingsbury, B., Morgan, N., Shire, M., Silipo, R., and Wu, S. L. (1999). Syllable-based speech recognition using auditorylike features. *The Journal of the Acoustical Society of America*, 105(2):1157–1158.

[Grossman et al., 2013] Grossman, R. L., Greenway, M., Heath, A. P., Powell, R., Suarez, R. D., Wells, W., White, K., Atkinson, M., Klampanos, I., Alvarez, H. L.,

Harvey, C., and Mambretti, J. J. (2013). The Design of a Community Science Cloud: The Open Science Data Cloud Perspective. In *2012 SC Companion: High-Performance Computing, Networking, Storage and Analysis (SCC)*, pages 1051–1057. IEEE.

[Grube and Griffiths, 2009] Grube, M. and Griffiths, T. D. (2009). Metricality-enhanced temporal encoding and the subjective perception of rhythmic sequences. *Cortex*, 45(1):72–79.

[Gu et al., 2011] Gu, B. M., Cheng, R. K., Yin, B., and Meck, W. H. (2011). Quinpirole-induced sensitization to noisy/sparse periodic input: temporal synchronization as a component of obsessive-compulsive disorder. *Neuroscience*, 179:143–150.

[Haegens et al., 2011] Haegens, S., Nácher, V., and Luna, R. (2011). α-Oscillations in the monkey sensorimotor network influence discrimination performance by rhythmical inhibition of neuronal spiking. In *Proceedings of the ....*

[Hardy and Buonomano, 2016] Hardy, N. F. and Buonomano, D. V. (2016). Neurocomputational models of interval and pattern timing. *Current Opinion in Behavioral Sciences*, 8:250–257.

[Harper et al., 2016] Harper, N. S., Schoppe, O., Willmore, B. D. B., Cui, Z., Schnupp, J. W. H., and King, A. J. (2016). Network Receptive Field Modeling Reveals Extensive Integration and Multi-feature Selectivity in Auditory Cortical Neurons. *PLoS Computational Biology*, 12(11):e1005113–30.

[Haß et al., 2008] Haß, J., Blaschke, S., Rammsayer, T., and Herrmann, J. M. (2008). A neurocomputational model for optimal temporal processing. *Journal of Computational Neuroscience*, 25(3):449–464.

[Hattori et al., 2013] Hattori, Y., Tomonaga, M., and Matsuzawa, T. (2013). Spontaneous synchronized tapping to an auditory rhythm in a chimpanzee. *Scientific Reports*, 3:1566.

[Hausdorff et al., 2007] Hausdorff, J. M., Lowenthal, J., Herman, T., Gruendlinger, L., Peretz, C., and Giladi, N. (2007). Rhythmic auditory stimulation modulates gait variability in Parkinson's disease. *European Journal of Neuroscience*, 26(8):2369–2375.

[Hauser and McDermott, 2003] Hauser, M. D. and McDermott, J. (2003). The evolution of the music faculty: a comparative perspective. *Nature Neuroscience*, 6(7):663–668.

[Hayashi et al., 2015] Hayashi, M. J., Ditye, T., Harada, T., Hashiguchi, M., Sadato, N., Carlson, S., Walsh, V., and Kanai, R. (2015). Time Adaptation Shows Duration Selectivity in the Human Parietal Cortex. *PLOS Biology*, 13(9):e1002262.

[Henry and Herrmann, 2014] Henry, M. J. and Herrmann, B. (2014). Low-frequency neural oscillations support dynamic attending in temporal context. *Timing & Time ...*, 2(1):62–86.

[Henry et al., 2014] Henry, M. J., Herrmann, B., and Obleser, J. (2014). Entrained neural oscillations in multiple frequency bands comodulate behavior. *Proceedings of the National Academy of Sciences*, 111(41):14935–14940.

[Henry and Obleser, 2012] Henry, M. J. and Obleser, J. (2012). Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proceedings of the National Academy of Sciences*, 109(49):20095–20100.

[Hickok et al., 2003] Hickok, G., Buchsbaum, B., Humphries, C., and Muftuler, T. (2003). Auditory–Motor Interaction Revealed by fMRI: Speech, Music, and Working Memory in Area Spt. *Journal of cognitive neuroscience*, 15(5):673–682.

[Honing, 2012] Honing, H. (2012). Without it no music: beat induction as a fundamental musical trait. *Annals of the New York Academy of Sciences*, 1252(1):85–91.

[Honing et al., 2014] Honing, H., Bouwer, F. L., and Háden, G. P. (2014). Perceiving Temporal Regularity in Music: The Role of Auditory Event-Related Potentials (ERPs) in Probing Beat Perception. In *Neurobiology of Interval Timing*, pages 305–323. Springer New York, New York, NY.

[Honing and Merchant, 2014] Honing, H. and Merchant, H. (2014). Differences in auditory timing between human and nonhuman primates. *Behavioral and Brain Sciences*, 37(06):557–558.

[Honing et al., 2012] Honing, H., Merchant, H., Háden, G. P., Prado, L., and Bartolo, R. (2012). Rhesus Monkeys (Macaca mulatta) Detect Rhythmic Groups in Music, but Not the Beat. *PLoS ONE*, 7(12):e51369.

[Hsieh and Ranganath, 2014] Hsieh, L.-T. and Ranganath, C. (2014). Frontal midline theta oscillations during working memory maintenance and episodic encoding and retrieval. *NeuroImage*, 85:721–729.

[Ingham and McAlpine, 2004] Ingham, N. J. and McAlpine, D. (2004). Spike-Frequency Adaptation in the Inferior Colliculus. *Journal of Neurophysiology*, 91(2):632–645.

[Iversen et al., 2009] Iversen, J. R., Repp, B. H., and Patel, A. D. (2009). Top-Down Control of Rhythm Perception Modulates Early Auditory Responses. *Annals of the New York Academy of Sciences*, 1169(1):58–73.

[Ivry and Schlerf, 2008] Ivry, R. B. and Schlerf, J. E. (2008). Dedicated and intrinsic models of time perception. *Trends in Cognitive Sciences*, 12(7):273–280.

[Jacobson et al., 2008] Jacobson, G. A., Rokni, D., and Yarom, Y. (2008). A model of the olivo-cerebellar system as a temporal pattern generator. *Trends in Neurosciences*, 31(12):617–625.

[Jacoby and McDermott, 2017] Jacoby, N. and McDermott, J. H. (2017). Integer Ratio Priors on Musical Rhythm Revealed Cross-culturally by Iterated Reproduction. *Current Biology*, 27(3):359–370.

[Jaramillo and Zador, 2011] Jaramillo, S. and Zador, A. M. (2011). The auditory cortex mediates the perceptual effects of acoustic temporal expectation. *Nature Neuroscience*, 14(2):246–251.

[Jazayeri and Shadlen, 2015] Jazayeri, M. and Shadlen, M. N. (2015). A Neural Mechanism for Sensing and Reproducing a Time Interval. *Current Biology*, 25(20):2599–2609.

[Jensen et al., 2014] Jensen, O., Gips, B., Bergmann, T. O., and Bonnefond, M. (2014). Temporal coding organized by coupled alpha and gamma oscillations prioritize visual processing. *Trends in Neurosciences*, pages 1–14.

[Jin et al., 2009] Jin, D. Z., Fujii, N., and Graybiel, A. M. (2009). Neural representation of time in cortico-basal ganglia circuits. *Proceedings of the National Academy of Sciences*, 106(45):19156–19161.

[Jones et al., 2002] Jones, M. R., Moynihan, H., MacKenzie, N., and Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychological science*, 13(4):313–319.

[Joris et al., 2004] Joris, P. X., Schreiner, C. E., and Rees, A. (2004). Neural processing of amplitude-modulated sounds. *Physiological reviews*, 84(2):541–577.

[Joseph et al., 2015] Joseph, S., Kumar, S., Husain, M., and Griffiths, T. D. (2015). Auditory working memory for objects vs. features. *Frontiers in neuroscience*, 9:20738.

[Jorgensen and Dau, 2011] Jorgensen, S. and Dau, T. (2011). Predicting speech intelligibility based on the signal-to-noise envelope power ratio after modulation-frequency selective processing. *The Journal of the Acoustical Society of America*, 130(3):1475–13.

[Kadir et al., 2013] Kadir, S. N., Goodman, D. F., and Harris, K. D. (2013). High-dimensional cluster analysis with the Masked EM Algorithm. *arXiv preprint arXiv:1309.2848*.

[Kaernbach, 2004] Kaernbach, C. (2004). The Memory of Noise. *Experimental Psychology*, 51(4):240–248.

[Karmarkar and Buonomano, 2007] Karmarkar, U. R. and Buonomano, D. V. (2007). Timing in the Absence of Clocks: Encoding Time in Neural Network States. *Neuron*, 53(3):427–438.

[Khouri and Nelken, 2015] Khouri, L. and Nelken, I. (2015). Detecting the unexpected. *Current opinion in neurobiology*, 35:142–147.

[Kilgard and Merzenich, 1999] Kilgard, M. P. and Merzenich, M. M. (1999). Distributed representation of spectral and temporal information in rat primary auditory cortex. *Hearing Research*, 134(1-2):16–28.

[Kornysheva et al., 2010a] Kornysheva, K., von Anshelm-Schiffer, A.-M., and Schubotz, R. I. (2010a). Inhibitory stimulation of the ventral premotor cortex temporarily interferes with musical beat rate preference. *Human Brain Mapping*, 32(8):1300–1310.

[Kornysheva et al., 2010b] Kornysheva, K., von Cramon, D. Y., Jacobsen, T., and Schubotz, R. I. (2010b). Tuning-in to the beat: Aesthetic appreciation of musical rhythms correlates with a premotor activity boost. *Human Brain Mapping*, 31(1):48–64.

[Kumar et al., 2014] Kumar, S., Bonnici, H. M., Teki, S., Agus, T. R., Pressnitzer, D., Maguire, E. A., and Griffiths, T. D. (2014). Representations of specific acoustic patterns in the auditory cortex and hippocampus. *Proceedings. Biological sciences / The Royal Society*, 281(1791):20141000–20141000.

[Kung et al., 2013] Kung, S.-J., Chen, J. L., Zatorre, R. J., and Penhune, V. B. (2013). Interacting cortical and basal ganglia networks underlying finding and tapping to the musical beat. *Journal of cognitive neuroscience*, 25(3):401–420.

[Lakatos, 2005] Lakatos, P. (2005). An Oscillatory Hierarchy Controlling Neuronal Excitability and Stimulus Processing in the Auditory Cortex. *Journal of Neurophysiology*, 94(3):1904–1911.

[Lakatos et al., 2008] Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., and Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science*, 320(5872):110–113.

[Lakatos et al., 2013] Lakatos, P., Musacchia, G., O'Connel, M. N., Falchier, A. Y., Javitt, D. C., and Schroeder, C. E. (2013). The Spectrotemporal Filter Mechanism of Auditory Selective Attention. *Neuron*, 77(4):750–761.

[Lakatos et al., 2009] Lakatos, P., O'Connell, M. N., Barczak, A., Mills, A., Javitt, D. C., and Schroeder, C. E. (2009). The Leading Sense: Supramodal Control of Neurophysiological Context by Attention. *Neuron*, 64(3):419–430.

[Langner, 1992] Langner, G. (1992). Periodicity Coding in the Auditory-System. *Hearing Research*, 60(2):115–142.

[Large, 2008] Large, E. W. (2008). Resonating to musical rhythm: theory and experiment. In *Psychology of Time*.

[Large et al., 2015] Large, E. W., Herrera, J. A., and Velasco, M. J. (2015). Neural Networks for Beat Perception in Musical Rhythm. *Frontiers in Systems Neuroscience*, 9(224):583–14.

[Large and Jones, 1999] Large, E. W. and Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological review*, 106(1):119–159.

[Lawrance et al., 2014] Lawrance, E. L. A., Harper, N. S., Cooke, J. E., and Schnupp, J. W. H. (2014). Temporal predictability enhances auditory detection. *The Journal of the Acoustical Society of America*, 135(6):EL357–EL363.

[Lehmann et al., 2016] Lehmann, A., Arias, D. J., and Schönwiesner, M. (2016). Tracing the neural basis of auditory entrainment. *Neuroscience*, 337:306–314.

[Leon and Shadlen, 2003] Leon, M. I. and Shadlen, M. N. (2003). Representation of Time by Neurons in the Posterior Parietal Cortex of the Macaque. *Neuron*, 38(2):317–327.

[Lerdahl and Jackendoff, 1983] Lerdahl, F. and Jackendoff, R. (1983). *A Generative Theory of Tonal Music. 1983*. MIT Press.

[Lewis and Miall, 2003] Lewis, P. A. and Miall, R. C. (2003). Brain activation patterns during measurement of sub- and supra-second intervals. *Neuropsychologia*, 41(12):1583–1592.

[Li et al., 2017] Li, J., Liao, X., Zhang, J., Wang, M., Yang, N., Zhang, J., Lv, G., Li, H., Lu, J., Ding, R., Li, X., Guang, Y., Yang, Z., Qin, H., Jin, W., Zhang, K., He, C., Jia, H., Zeng, S., Hu, Z., Nelken, I., and Chen, X. (2017). Primary Auditory

Cortex is Required for Anticipatory Motor Response. *Cerebral cortex (New York, N.Y. : 1991)*, pages 1–18.

[Linden, 2003] Linden, M. S. J. F. (2003). How linear are auditory cortical responses? 15:125.

[Llinas et al., 1974] Llinas, R., Baker, R., and Sotelo, C. (1974). Electrotonic coupling between neurons in cat inferior olive. *Journal of Neurophysiology*.

[Llinas and Yarom, 1981] Llinas, R. and Yarom, Y. (1981). Electrophysiology of mammalian inferior olivary neurones in vitro. Different types of voltage-dependent ionic conductances. *The Journal of Physiology*, 315(1):549–567.

[London, 2012a] London, J. (2012a). *Hearing in Time*. Psychological Aspects of Musical Meter. Oxford University Press.

[London, 2012b] London, J. (2012b). Hearing in Time - Justin London - Google Books.

[Loveless et al., 1996] Loveless, N., Levänen, S., Jousmäki, V., Sams, M., and Hari, R. (1996). Temporal integration in auditory sensory memory: neuromagnetic evidence. *Electroencephalography ...*, 100(3):220–228.

[Luo and Poeppel, 2007] Luo, H. and Poeppel, D. (2007). Phase Patterns of Neuronal Responses Reliably Discriminate Speech in Human Auditory Cortex. *Neuron*, 54(6):1001–1010.

[Lusk et al., 2016] Lusk, N. A., Petter, E. A., MacDonald, C. J., and Meck, W. H. (2016). Cerebellar, hippocampal, and striatal time cells. *Current Opinion in Behavioral Sciences*, 8:186–192.

[MacDonald et al., 2011] MacDonald, C. J., Lepage, K. Q., Eden, U. T., and Eichenbaum, H. (2011). Hippocampal "Time Cells" Bridge the Gap in Memory for Discontiguous Events. *Neuron*, 71(4):737–749.

[Madison et al., 2011] Madison, G., Gouyon, F., Ullén, F., and Hörnström, K. (2011). Modeling the tendency for music to induce movement in humans: First correlations with low-level audio descriptors across music genres. *Journal of experimental psychology. Human perception and performance*, 37(5):1578–1594.

[Malmierca, 2014] Malmierca, M. (2014). Neuronal adaptation, novelty detection and regularity encoding in audition. *Frontiers in Systems Neuroscience*, 8.

[Malmierca et al., 2008] Malmierca, M. S., Izquierdo, M. A., Cristaudo, S., Hernandez, O., Perez-Gonzalez, D., Covey, E., and Oliver, D. L. (2008). A Discontinuous Tonotopic Organization in the Inferior Colliculus of the Rat. *The Journal of Neuroscience*, 28(18):4767–4776.

[Malone et al., 2015] Malone, B. J., Beitel, R. E., Vollmer, M., Heiser, M. A., and Schreiner, C. E. (2015). Modulation-frequency-specific adaptation in awake auditory cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 35(15):5904–5916.

[Masquelier et al., 2008] Masquelier, T., Guyonneau, R., and Thorpe, S. J. (2008). Spike Timing Dependent Plasticity Finds the Start of Repeating Patterns in Continuous Spike Trains. *PLoS ONE*, 3(1):e1377.

[Matell and Meck, 2004] Matell, M. S. and Meck, W. H. (2004). Cortico-striatal circuits and interval timing: coincidence detection of oscillatory processes. *Cognitive Brain Research*, 21(2):139–170.

[Mathewson et al., 2011] Mathewson, K. E., Lleras, A., Beck, D. M., Fabiani, M., Ro, T., and Gratton, G. (2011). Pulsed Out of Awareness: EEG Alpha Oscilla-

tions Represent a Pulsed-Inhibition of Ongoing Cortical Processing. *Frontiers in psychology*, 2.

[Mathy et al., 2009] Mathy, A., Ho, S. S. N., Davie, J. T., Duguid, I. C., Clark, B. A., and Häusser, M. (2009). Encoding of Oscillations by Axonal Bursts in Inferior Olive Neurons. *Neuron*, 62(3):388–399.

[McAuley et al., 2012] McAuley, J. D., Henry, M. J., and Tkach, J. (2012). Tempo mediates the involvement of motor areas in beat perception. *Annals of the New York Academy of Sciences*, 1252(1):77–84.

[McAuley et al., 2006] McAuley, J. D., Jones, M. R., Holub, S., Johnston, H. M., and Miller, N. S. (2006). The time of our lives: Life span development of timing and event tracking. *Journal of Experimental Psychology: General*, 135(3):348–367.

[McDermott and Simoncelli, 2011] McDermott, J. H. and Simoncelli, E. P. (2011). Sound texture perception via statistics of the auditory periphery: evidence from sound synthesis. *Neuron*, 71(5):926–940.

[McKinney et al., 2007] McKinney, M. F., Moelants, D., Davies, M. E. P., and Klapuri, A. (2007). Evaluation of Audio Beat Tracking and Music Tempo Extraction Algorithms. *Journal of New Music Research*, 36(1):1–16.

[Meck et al., 2008] Meck, W. H., Penney, T. B., and Pouthas, V. (2008). Cortico-striatal representation of time in animals and humans. *Current opinion in neurobiology*, 18(2):145–152.

[Mello et al., 2015] Mello, G. B. M., Soares, S., and Paton, J. J. (2015). A Scalable Population Code for Time in the Striatum. *Current Biology*, 25(9):1113–1122.

[Merchant et al., 2015] Merchant, H., Grahn, J., Trainor, L., Rohrmeier, M., and Fitch, W. T. (2015). Finding the beat: a neural perspective across humans and non-human primates. *Phil. Trans. R. Soc. B*, 370(1664):20140093–20140093.

[Merchant et al., 2013] Merchant, H., Harrington, D. L., and Meck, W. H. (2013). Neural Basis of the Perception and Estimation of Time. *Annual review of neuroscience*, 36(1):313–336.

[Merchant and Honing, 2014] Merchant, H. and Honing, H. (2014). Are non-human primates capable of rhythmic entrainment? Evidence for the gradual audiomotor evolution hypothesis. *Frontiers in neuroscience*, 7.

[Miall, 1989] Miall, C. (1989). The Storage of Time Intervals Using Oscillating Neurons. *Neural Computation*, 1(3):359–371.

[Morillon et al., 2015] Morillon, B., Hackett, T. A., Kajikawa, Y., and Schroeder, C. E. (2015). ScienceDirect Predictive motor control of sensory dynamics in auditory active sensing. *Current opinion in neurobiology*, 31:230–238.

[Morillon et al., 2016] Morillon, B., Schroeder, C. E., Wyart, V., and Arnal, L. H. (2016). Temporal Prediction in lieu of Periodic Stimulation. *The Journal of Neuroscience*, 36(8):2342–2347.

[Namboodiri et al., 2015] Namboodiri, V. M. K., Huertas, M. A., Monk, K. J., Shouval, H. Z., and Shuler, M. G. (2015). Visually Cued Action Timing in the Primary Visual Cortex. *Neuron*, 86(1):319–330.

[Neuert et al., 2001] Neuert, V., Pressnitzer, D., Patterson, R. D., and Winter, I. M. (2001). The responses of single units in the inferior colliculus of the guinea pig to damped and ramped sinusoids. *Hearing Research*, 159(1-2):36–52.

[Nieto-Diego and Malmierca, 2016] Nieto-Diego, J. and Malmierca, M. S. (2016). Topographic Distribution of Stimulus-Specific Adaptation across Auditory Cortical Fields in the Anesthetized Rat. *PLOS Biology*, 14(3):e1002397.

[Nobre et al., 2007] Nobre, A. C., Correa, A., and Coull, J. T. (2007). The hazards of time. *Current opinion in neurobiology*, 17(4):465–470.

[Nolan and Jeon, 2014] Nolan, F. and Jeon, H. S. (2014). Speech rhythm: a metaphor? *Phil. Trans. R. Soc. B*, 369(1658):20130396–20130396.

[Nozaradan et al., 2016] Nozaradan, S., Peretz, I., and Keller, P. E. (2016). Individual Differences in Rhythmic Cortical Entrainment Correlate with Predictive Behavior in Sensorimotor Synchronization. *Scientific Reports*, 6:20612.

[Nozaradan et al., 2011] Nozaradan, S., Peretz, I., Missal, M., and Mouraux, A. (2011). Tagging the Neuronal Entrainment to Beat and Meter. *The Journal of Neuroscience*, 31(28):10234–10240.

[Nozaradan et al., 2012] Nozaradan, S., Peretz, I., and Mouraux, A. (2012). Selective Neuronal Entrainment to the Beat and Meter Embedded in a Musical Rhythm. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 32(49):17572–17581.

[Oullier et al., 2005] Oullier, O., Jantzen, K. J., Steinberg, F. L., and Kelso, J. A. S. (2005). Neural Substrates of Real and Imagined Sensorimotor Coordination. *Cerebral cortex (New York, N.Y. : 1991)*, 15(7):975–985.

[Panzeri et al., 2010] Panzeri, S., Brunel, N., Logothetis, N. K., and Kayser, C. (2010). Sensory neural codes using multiplexed temporal scales. *Trends in Neurosciences*, 33(3):111–120.

[Parent and Hazrati, 1995] Parent, A. and Hazrati, L.-N. (1995). Functional anatomy of the basal ganglia. I. The cortico-basal ganglia-thalamo-cortical loop. *Brain Research Reviews*, 20(1):91–127.

[Parsons, 2001] Parsons, L. M. (2001). Exploring the Functional Neuroanatomy of Music Performance, Perception, and Comprehension. *Annals of the New York Academy of Sciences*, 930(1):211–231.

[Patel, 2007] Patel, A. D. (2007). *Music, Language, and the Brain*. Oxford University Press.

[Patel, 2011] Patel, A. D. (2011). Why would Musical Training Benefit the Neural Encoding of Speech? The OPERA Hypothesis. *Frontiers in psychology*, 2.

[Patel and Iversen, 2014] Patel, A. D. and Iversen, J. R. (2014). The evolutionary neuroscience of musical beat perception: the Action Simulation for Auditory Prediction (ASAP) hypothesis. *Frontiers in Systems Neuroscience*.

[Patel et al., 2009] Patel, A. D., Iversen, J. R., Bregman, M. R., and Schulz, I. (2009). Experimental Evidence for Synchronization to a Musical Beat in a Nonhuman Animal. *Current Biology*, 19(10):827–830.

[Peckel et al., 2014] Peckel, M., Pozzo, T., and Bigand, E. (2014). The impact of the perception of rhythmic music on self-paced oscillatory movements. *Frontiers in psychology*, 5(183):917.

[Peelle and Davis, 2012] Peelle, J. E. and Davis, M. H. (2012). Neural Oscillations Carry Speech Rhythm through to Comprehension. *Frontiers in psychology*, 3.

[Pérez-González and Malmierca, 2014] Pérez-González, D. and Malmierca, M. S. (2014). Adaptation in the auditory system: an overview. *Frontiers in Integrative Neuroscience*, 8.

[Phillips-Silver et al., 2011] Phillips-Silver, J., Toiviainen, P., Gosselin, N., Piché, O., Nozaradan, S., Palmer, C., and Peretz, I. (2011). Born to dance but beat deaf: A new form of congenital amusia. *Neuropsychologia*, 49(5):961–969.

[Poeppel, 2003] Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech communication*, 41(1):245–255.

[Potter et al., 2009] Potter, D. D., Fenwick, M., Abecasis, D., and Brochard, R. (2009). Perceiving rhythm where none exists: Event-related potential (ERP) correlates of subjective accenting. *Cortex*, 45(1):103–109.

[Povel and Essens, 1985] Povel, D.-J. and Essens, P. (1985). Perception of Temporal Patterns. *Music Perception: An Interdisciplinary Journal*, 2(4):411–440.

[Pressnitzer and Hupe, 2006] Pressnitzer, D. and Hupe, J.-M. (2006). Temporal Dynamics of Auditory and Visual Bistability Reveal Common Principles of Perceptual Organization. *Current Biology*, 16(13):1351–1357.

[Rajendran et al., 2016] Rajendran, V. G., Harper, N. S., Abdel-Latif, K. H. A., and Schnupp, J. W. H. (2016). Rhythm Facilitates the Detection of Repeating Sound Patterns. *Frontiers in neuroscience*, 10(60):464–7.

[Rajendran et al., 2017a] Rajendran, V. G., Harper, N. S., Garcia-Lazaro, J. A., Lesica, N. A., and Schnupp, J. W. H. (2017a). Midbrain adaptation may set the stage for the perception of musical beat. *Proceedings. Biological sciences / The Royal Society*, 284(1866):20171455.

[Rajendran et al., 2013] Rajendran, V. G., Harper, N. S., Willmore, B. D., Hartmann, W. M., and Schnupp, J. W. H. (2013). Temporal predictability as a grouping cue in the perception of auditory streams. *The Journal of the Acoustical Society of America*, 134(1):EL98–EL104.

[Rajendran and Teki, 2016] Rajendran, V. G. and Teki, S. (2016). Periodicity versus Prediction in Sensory Perception. *The Journal of Neuroscience*, 36(28):7343–7345.

[Rajendran et al., 2017b] Rajendran, V. G., Teki, S., and Schnupp, J. W. H. (2017b). Temporal Processing in Audition: Insights from Music. *Neuroscience*.

[Rao et al., 2001] Rao, R. P. N., Eagleman, D. M., and Sejnowski, T. J. (2001). Optimal Smoothing in Visual Motion Perception. *Neural Computation*, 13(6):1243–1253.

[Repp, 2002a] Repp, B. H. (2002a). Automaticity and voluntary control of phase correction following event onset shifts in sensorimotor synchronization. *Journal of experimental psychology. Human perception and performance*, 28(2):410–430.

[Repp, 2002b] Repp, B. H. (2002b). Perception of timing is more context sensitive than sensorimotor synchronization. *Perception & Psychophysics*, 64(5):703–716.

[Repp, 2005] Repp, B. H. (2005). Sensorimotor synchronization: a review of the tapping literature. *Psychonomic Bulletin & Review*, 12(6):969–992.

[Repp and Keller, 2004] Repp, B. H. and Keller, P. E. (2004). Adaptation to tempo changes in sensorimotor synchronization: Effects of intention, attention, and awareness. *Quarterly Journal of Experimental ...*, 57(3):499–521.

[Repp and Su, 2013] Repp, B. H. and Su, Y.-H. (2013). Sensorimotor synchronization: A review of recent research (2006–2012). *Psychonomic Bulletin & Review*, 20(3):403–452.

[Riecker et al., 2003] Riecker, A., Wildgruber, D., Mathiak, K., Grodd, W., and Ackermann, H. (2003). Parametric analysis of rate-dependent hemodynamic response functions of cortical and subcortical brain structures during auditorily cued finger tapping: a fMRI study. *NeuroImage*, 18(3):731–739.

[Rohenkohl et al., 2014] Rohenkohl, G., Gould, I. C., Pessoa, J., and Nobre, A. C. (2014). Combining spatial and temporal expectations to improve visual perception. *Journal of Vision*, 14(4):8–8.

[Rouse et al., 2016] Rouse, A. A., Cook, P. F., Large, E. W., and Reichmuth, C. (2016). Beat Keeping in a Sea Lion As Coupled Oscillation: Implications for Comparative Understanding of Human Rhythm. *Frontiers in neuroscience*, 10(159):403.

[Sakai et al., 1999] Sakai, K., Hikosaka, O., Miyauchi, S., Takino, R., Tamada, T., Iwata, N. K., and Nielsen, M. (1999). Neural representation of a rhythm depends on its interval ratio. *The Journal of Neuroscience*, 19(22):10074–10081.

[Schaefer et al., 2010] Schaefer, R. S., Vlek, R. J., and Desain, P. (2010). Decomposing rhythm processing: electroencephalography of perceived and self-imposed rhythmic patterns. *Psychological Research*, 75(2):95–106.

[Schneider and Ghose, 2012] Schneider, B. A. and Ghose, G. M. (2012). Temporal Production Signals in Parietal Cortex. *PLOS Biology*, 10(10):e1001413.

[Schnupp et al., 2015] Schnupp, J. W. H., Garcia-Lazaro, J. A., and Lesica, N. A. (2015). Periodotopy in the gerbil inferior colliculus: local clustering rather than a gradient map. *Frontiers in Neural Circuits*, 9:423.

[Schönwiesner and Zatorre, 2009] Schönwiesner, M. and Zatorre, R. J. (2009). Spectro-temporal modulation transfer function of single voxels in the human auditory cortex measured with high-resolution fMRI. *Proceedings of the National Academy of Sciences of the United States of America*, 106(34):14611–14616.

[Schreiner and Urbas, 1986] Schreiner, C. E. and Urbas, J. V. (1986). Representation of amplitude modulation in the auditory cortex of the cat. I. The anterior auditory field (AAF). *Hearing Research*, 21(3):227–241.

[Schreiner and Urbas, 1988] Schreiner, C. E. and Urbas, J. V. (1988). Representation of amplitude modulation in the auditory cortex of the cat. II. Comparison between cortical fields. *Hearing Research*, 32(1):49–63.

[Schroeder and Lakatos, 2009] Schroeder, C. E. and Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences*, 32(1):9–18.

[Schroeder et al., 2010] Schroeder, C. E., Wilson, D. A., Radman, T., Scharfman, H., and Lakatos, P. (2010). Dynamics of Active Sensing and perceptual selection. *Current opinion in neurobiology*, 20(2):172–176.

[Schwartze et al., 2013] Schwartze, M., Farrugia, N., and Kotz, S. A. (2013). Dissociation of formal and temporal predictability in early auditory evoked potentials. *Neuropsychologia*, 51(2):320–325.

[Selezneva et al., 2013] Selezneva, E., Deike, S., Knyazeva, S., Scheich, H., Brechmann, A., and Brosch, M. (2013). Rhythm sensitivity in macaque monkeys. *Frontiers in Systems Neuroscience*, 7.

[Shamma et al., 2011] Shamma, S. A., Elhilali, M., and Micheyl, C. (2011). Temporal coherence and attention in auditory scene analysis. *Trends in Neurosciences*, 34(3):114–123.

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

[Shuler, 2016] Shuler, M. G. (2016). Timing in the visual cortex and its investigation. *Current Opinion in Behavioral Sciences*, 8:73–77.

[Snyder and Large, 2005] Snyder, J. S. and Large, E. W. (2005). Gamma-band activity reflects the metric structure of rhythmic tone sequences. *Cognitive Brain Research*, 24(1):117–126.

[Soares et al., 2016] Soares, S., Atallah, B. V., and Paton, J. J. (2016). Midbrain dopamine neurons control judgment of time. *Science*, 354(6317):1273–1277.

[Sowiński and Bella, 2013] Sowiński, J. and Bella, S. D. (2013). Poor synchronization to the beat may result from deficient auditory-motor mapping. *Neuropsychologia*, 51(10):1952–1963.

[Strauss et al., 2014] Strauss, A., WÃ stmann, M., and Obleser, J. (2014). Cortical alpha oscillations as a tool for auditory selective inhibition. *Frontiers in human neuroscience*, 8(154):1926.

[Stupacher et al., 2017] Stupacher, J., Wood, G., and Witte, M. (2017). Neural Entrainment to Polyrhythms: A Comparison of Musicians and Non-musicians. *Frontiers in neuroscience*, 11:1664.

[Styns et al., 2007] Styns, F., van Noorden, L., Moelants, D., and Leman, M. (2007). Walking on music. 26(5):769–785.

[Tal et al., 2017] Tal, I., Large, E. W., Rabinovitch, E., Wei, Y., Schroeder, C. E., Poeppel, D., and Zion Golumbic, E. (2017). Neural Entrainment to the Beat: The "Missing-Pulse" Phenomenon. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 37(26):6331–6341.

[Teki, 2014] Teki, S. (2014). Beta drives brain beats. *Frontiers in Systems Neuroscience*, 8(274):743.

[Teki, 2016] Teki, S. (2016). A Citation-Based Analysis and Review of Significant Papers on Timing and Time Perception. *Frontiers in neuroscience*, 10(1329):656.

[Teki and Griffiths, 2014] Teki, S. and Griffiths, T. D. (2014). Neural Basis of Working Memory for Time Intervals. *Procedia-Social and Behavioral Sciences*, 126:269–270.

[Teki and Griffiths, 2016] Teki, S. and Griffiths, T. D. (2016). Brain Bases of Working Memory for Time Intervals in Rhythmic Sequences. *Frontiers in neuroscience*, 10(13):743.

[Teki et al., 2012] Teki, S., Grube, M., and Griffiths, T. D. (2012). A Unified Model of Time Perception Accounts for Duration-Based and Beat-Based Timing Mechanisms. *Frontiers in Integrative Neuroscience*, 5.

[Teki et al., 2011] Teki, S., Grube, M., Kumar, S., and Griffiths, T. D. (2011). Distinct Neural Substrates of Duration-Based and Beat-Based Auditory Timing. *The Journal of Neuroscience*, 31(10):3805–3812.

[Teki and Kononowicz, 2016] Teki, S. and Kononowicz, T. W. (2016). Commentary: Beta-Band Oscillations Represent Auditory Beat and Its Metrical Hierarchy in Perception and Imagery. *Frontiers in neuroscience*, 10(155):743.

[Tervaniemi and Hugdahl, 2003] Tervaniemi, M. and Hugdahl, K. (2003). Lateralization of auditory-cortex functions. *Brain Research Reviews*, 43(3):231–246.

[Tierney and Kraus, 2013] Tierney, A. and Kraus, N. (2013). Neural responses to sounds presented on and off the beat of ecologically valid music. *Frontiers in Systems Neuroscience*, 7.

[Todd and Lee, 2015] Todd, N. P. M. and Lee, C. S. (2015). Source analysis of electrophysiological correlates of beat induction as sensory-guided action. *Frontiers in psychology*, 6(e97467):611.

[Trost et al., 2014] Trost, W., Frühholz, S., Schön, D., Labbé, C., Pichon, S., Grandjean, D., and Vuilleumier, P. (2014). Getting the beat: Entrainment of brain activity by musical rhythm and pleasantness. *NeuroImage*, 103:55–64.

[Turgeon et al., 2002] Turgeon, M., Bregman, A. S., and Ahad, P. A. (2002). Rhythmic masking release: Contribution of cues for perceptual organization to the cross-spectral fusion of concurrent narrow-band noises. *The Journal of the Acoustical Society of America*, 111(4):1819–1831.

[Turgeon et al., 2005] Turgeon, M., Bregman, A. S., and Roberts, B. (2005). Rhythmic Masking Release: Effects of Asynchrony, Temporal Overlap, Harmonic Relations, and Source Separation on Cross-Spectral Grouping. *Journal of experimental psychology. Human perception and performance*, 31(5):939–953.

[Van Dyck et al., 2015] Van Dyck, E., Moens, B., Buhmann, J., Demey, M., Coorevits, E., Bella, S. D., and Leman, M. (2015). Spontaneous Entrainment of Running Cadence to Music Tempo. *Sports Medicine - Open*, pages 1–14.

[van Noorden, 1975] van Noorden, L. (1975). Temporal coherence in the perception of tone sequences. *Doctoral Thesis*.

[van Noorden and Moelants, 1999] van Noorden, L. and Moelants, D. (1999). Resonance in the Perception of Musical Pulse. *Journal of New Music Research*, 28(1):43–66.

[Vuust and Witek, 2014] Vuust, P. and Witek, M. A. G. (2014). Rhythmic complexity and predictive coding: a novel approach to modeling rhythm and meter perception in music. *Frontiers in psychology*, 5(218):273.

[Walker et al., 2011] Walker, K. M. M., Bizley, J. K., King, A. J., and Schnupp, J. W. H. (2011). Multiplexed and robust representations of sound features in

auditory cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 31(41):14565–14576.

[Welsh et al., 1995] Welsh, J. P., Lang, E. J., Suglhara, I., and Llinás, R. (1995). Dynamic organization of motor control within the olivocerebellar system. *Nature*, 374(6521):453–457.

[Willmore et al., 2016] Willmore, B. D. B., Schoppe, O., King, A. J., Schnupp, J. W. H., and Harper, N. S. (2016). Incorporating Midbrain Adaptation to Mean Sound Level Improves Models of Auditory Cortical Processing. *The Journal of Neuroscience*, 36(2):280–289.

[Winkler et al., 2012] Winkler, I., Denham, S., Mill, R., Bohm, T. M., and Bendixen, A. (2012). Multistability in auditory stream segregation: a predictive coding view. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1591):1001–1012.

[Winkler et al., 2009a] Winkler, I., Denham, S. L., and Nelken, I. (2009a). Modeling the auditory scene: predictive regularity representations and perceptual objects. *Trends in Cognitive Sciences*, 13(12):532–540.

[Winkler et al., 2009b] Winkler, I., Háden, G. P., Ladinig, O., Sziller, I., and Honing, H. (2009b). Newborn infants detect the beat in music. *Proceedings of the National Academy of Sciences of the United States of America*, 106(7):2468–2471.

[Witek et al., 2014] Witek, M. A. G., Clarke, E. F., Wallentin, M., Kringelbach, M. L., and Vuust, P. (2014). Syncopation, Body-Movement and Pleasure in Groove Music. *PLoS ONE*, 9(4):e94446.

[Wu et al., 2011] Wu, X., Ashe, J., and Bushara, K. O. (2011). Role of olivocerebellar system in timing without awareness. *Proceedings of the National Academy of Sciences*, 108(33):13818–13822.

[Xiang et al., 2013] Xiang, J., Poeppel, D., and Simon, J. Z. (2013). Physiological evidence for auditory modulation filterbanks: cortical responses to concurrent modulations. *The Journal of the Acoustical Society of America*, 133(1):EL7–12.

[Xu, 2006] Xu, D. (2006). Role of the Olivo-Cerebellar System in Timing. *The Journal of Neuroscience*, 26(22):5990–5995.

[Yarom and Cohen, 2002] Yarom, Y. and Cohen, D. (2002). The Olivocerebellar System as a Generator of Temporal Patterns. *Annals of the New York Academy of Sciences*, 978(1 THE CEREBELLU):122–134.

[Yogev et al., 2005] Yogev, G., Giladi, N., Peretz, C., Springer, S., Simon, E. S., and Hausdorff, J. M. (2005). Dual tasking, gait rhythmicity, and Parkinson's disease: Which aspects of gait are attention demanding? *European Journal of Neuroscience*, 22(5):1248–1256.

[Zanto et al., 2006] Zanto, T. P., Snyder, J. S., and Large, E. W. (2006). Neural correlates of rhythmic expectancy. *Advances in Cognitive Psychology*, 2(2):221–231.

[Zatorre et al., 2002] Zatorre, R. J., Belin, P., and Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends in Cognitive Sciences*, 6(1):37–46.

[Zatorre et al., 2007] Zatorre, R. J., Chen, J. L., and Penhune, V. B. (2007). When the brain plays music: auditory—[ndash]—motor interactions in music perception and production. *Nature Reviews Neuroscience*, 8(7):547–558.

[Zhou et al., 2016] Zhou, H., Melloni, L., Poeppel, D., and Ding, N. (2016). Interpretations of Frequency Domain Analyses of Neural Entrainment: Periodicity, Fundamental Frequency, and Harmonics. *Frontiers in human neuroscience*, 10:509–8.