



Run Run Shaw Library

香港城市大學
City University of Hong Kong

Copyright Warning

Use of this thesis/dissertation/project is for the purpose of private study or scholarly research only. ***Users must comply with the Copyright Ordinance.***

Anyone who consults this thesis/dissertation/project is understood to recognise that its copyright rests with its author and that no part of it may be reproduced without the author's prior written consent.

CITY UNIVERSITY OF HONG KONG
香港城市大學

Natural Sound Statistics in Auditory Scene
Analysis
聽覺場景分析中的自然聲音統計

Submitted to
Department of Biomedical Sciences
生物醫學系
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
哲學博士學位

by

MISHRA Ambika Prasad

September 2020
二零二零年九月

Abstract

We come across a wide variety of sounds every day. Most of the time we receive sounds from a multitude of sources rather than a single source. In a complex auditory environment, the brain's ability to segregate the flux of incoming sounds into separate auditory sources or streams plays a crucial role in auditory perception. In auditory scene analysis, how the brain represents different sound objects still remains an open area of research. Among natural sounds, "sound textures" have recently been recognized as an important class of sounds. Textures are stochastic streams of sounds which have temporal homogeneity, i.e. the statistical properties of these sounds do not vary significantly over a period of time. Typical examples might include the noises made by waves on a beach or the buzzing of a swarm of insects. Such sound textures are easily identified, and segregated into foreground or background sounds in the course of scene analysis, suggesting that the auditory system must be sensitive to the statistical features of sounds that make sound textures identifiable and discriminable.

In a recent psychoacoustic study, [[McDermott and Simoncelli, 2011](#)] have described methods that make it possible to synthesize naturalistic sounds from white noise by systematic imposition of statistical features, such as mean, variance, skew, kurtosis

of the amplitudes in sound envelopes in cochlear frequency channels, correlations between frequencies, and modulation power. However, how neurons at mid and higher level auditory stations encode or represent these statistical features is not known in detail. Moreover, the space of all theoretically possible sound textures is huge, making the exploration of this sound space in a systematic or representative way a challenging task. My thesis therefore has two objectives:

1. To compile and survey a sufficiently large corpus of natural sound textures to estimate the distributions of statistical features that are typically found in our environment, given that knowledge of these distributions will enable us to explore the sensitivity of the auditory system in a systematic manner.
2. To characterize the sensitivity of neurons in the auditory pathway to statistical features, using synthetic stimuli selected to form a representative sample of the “natural sound texture space” characterised in objective 1.

To address the first objective, I collected a corpus of 200 natural sounds and established a statistical framework based on principal component analysis to explore the natural sound texture space. I found that the large dimensionality of the statistical parameters of the natural sound texture space are mostly redundant and with only a few statistical parameters the natural sound texture space can be explored efficiently. To address the second objective, I selected a set of sounds from the sound corpus which I call as representative textures. I resynthesized these sounds to generate a set of synthetic stimuli or morphed textures from white noise by systematically morphing and imposing

these statistics in a hierarchical fashion to explore the impact of different statistics. I have used these morphed textures for electrophysiological recordings from the inferior colliculus (IC) and auditory cortex (AC) of young adult female Wistar rats. Subsequent analysis revealed that above 70% of the neurons in the inferior colliculus during onset and around ~30% of auditory cortical neurons are sensitive to only power to variance the statistical transition present in the natural sound textures. For other transitions auditory cortical neurons remain insensitive. On the other hand ~2-30% IC neurons are sensitive to other statistical transitions during onset response. For ongoing response around ~10-90% of IC neurons are sensitive whereas only ~2% cortical neurons are sensitive to modulation power only.

CITY UNIVERSITY OF HONG KONG

Qualifying Panel and Examination Panel

Surname: MISHRA
First Name: Ambika Prasad
Degree: PhD
College/Department: Department of Biomedical Sciences

The Qualifying Panel of the above student is composed of:

Supervisor(s)

Prof. Jan SCHNUPP Department of Neuroscience
City University of Hong Kong

Qualifying Panel members(s)

Dr. LAU Chun Yue Geoffrey Department of Neuroscience
City University of Hong Kong
Dr. YANG Sungchil Department of Neuroscience
City University of Hong Kong

This thesis has been examined and approved by the following examiners:

Dr. CHAN Ho Man Department of Electrical Engineering
City University of Hong Kong
Prof. HE Jufang Department of Neuroscience
City University of Hong Kong
Prof. Jan SCHNUPP Department of Neuroscience
City University of Hong Kong
Dr. SUMNER Christian School of Social Sciences
Nottingham Trent University

DECLARATION

I, MISHRA Ambika Prasad, declare that this thesis entitled, "**Natural Sound Statistics in Auditory Scene Analysis**" represents my original work and the contents of this thesis has never been submitted to this University and other institutions for a degree or any other qualifications in the form of thesis or other report.

MISHRA Ambika Prasad

To my family

Nana, Maa, Putli and Baya

Acknowledgements

If a doctorate student is very lucky, they will happen to pick an excellent supervisor. I was fortunate enough to choose Prof. Jan W. H. Schnupp. Generous with his time and able to provide both high-level inspiration and also detailed technical help, he has allowed me great freedom to work on a range of problems of my choice. I am extremely grateful for his wisdom and his kindness. “I want you to be successful”, and “you are so close to your results” are his two most encouraging statements that I am going to embrace in my heart throughout my life.

I am grateful to my late parents whose teachings have guided me through every journey of life. Life’s journey is beyond imagination without a good soulmate. This journey of Ph.D. would not have materialized itself without my wife’s encouragement. She has walked in parallel with me in this difficult journey. She has taken all the pain alone to take care of my lovely son Krutartha (Baya) in my absence despite she went through two major surgeries. My son has also cooperated silently in all possible ways despite he has missed me always.

I am grateful to Dr. Nicol S. Harper, Department of Physiology, Anatomy, and Genetics, University of Oxford, for his help in stimuli design and significant contribution in data analysis. My heartfelt thanks to Dr. Fei Peng for her help in data analysis, stimuli design and cortical electrophysiological recordings. I am thankful to Dr. Nicole Rosskothén-Kuhl, for training me the basic skills.

I am grateful to Prof. Christian Sumner, School of Social Sciences, Nottingham Trent University, UK, Prof. Jufang He, Department of Neuroscience, City University of HongKong and Dr. Rosa H.M.Chan, Department of Electrical Engineering, City University of HongKong for their valuable time in reading my thesis, giving a lot of constructive comments and for a scientific discussion during the defense.

I consider myself fortunate to have studied for my PhD. at the same time as Cecilia. I am greatly thankful for her help in all electrophysiological recordings. My sincere

thanks to Jacinta, Vani, Ryszard, Hijee, Drew, Ahn, Alexa and Chloe for their analytical thoughts and support.

My sincere thanks to Dr. Sungchil Yang and Dr. Geoffrey Lau to assess my progress for three years and providing valuable comments to improve my study. I would also thank to academicians, administration and technical staff of Department of Biomedical Sciences for their continuous support.

I would also like to thank Dr. Pallavi Asthana and Dr. Gajendra Kumar who have always stood by me in every aspect of my life here in Hong Kong. My heartfelt thanks to Dr. Abhimanyu Thakur, who literally stood with me during my entire stay in Hong Kong.

Finally, I would also like to thank my family - Nani, Bhaina, Sanbhaina and Jwain for their constant motivation and support. My sincere gratitude to my (Late.)father-in-law and mother-in-law for being there at home whenever I have needed. Last but not the least, my sincere and heartfelt gratitude to my brother-in-law Mr.Saroj Kumar Panda and Mrs.Chanchala Panda. Without their support, I would never have come to Hong Kong at the first place. He took care of the most important responsibilities that I should have. They took the best care of my family throughout my absence. My heartfelt thanks to Lori who was always there with my son to fill my absence. While I am writing this thesis, the world is going through COVID-19 pandemic. Millions have lost their lives and millions are affected. My heartfelt condolences to those families who lost their near and dear once. My sincere gratitude to all those who are working day and night to save people all around the globe. My sincere gratitude to all those who are working hard to keep the environment clean.

Lots of Love!

List of Abbreviations

A1	Primary auditory cortex
ABR	Auditory brainstem responses
AC	Auditory Cortex
AVCN	AnteroVentral Cochlear Nucleus
BIC	Brachium of the IC
BM	Basilar Membrane
C	Cochlear cross-band correlations
C1	Cross band modulation correlation
C2	Within band modulation correlation
CASA	Computational Auditory Scene Analysis
CF	Characteristic Frequency
DCN	Dorsal Cochlear Nucleus
ERB	Equivalent Rectangular Bandwidth
IC	Inferior Colliculus
ICc	Central nucleus of IC
ICd	Dorsal nucleus of IC
ICx	External nucleus of IC
ILD	Interaural Level Difference
ITD	Interaural Time Difference
K/k	Kurtosis

LSO	Lateral Superior Olive
M/mod.pow	Modulation power
MFCC	Mel-frequency cepstral coefficients
Mgv	Ventral medial geniculate body
Mgd	Dorsal medial geniculate body
MNTB	Medial Nucleus of the Trapezoid Body
MSO	Medial Superior Olive
m	Mean
PCA	Principal Component Analysis
PC1	First principal Component
PC2	Second principal Component
PVCN	PosteroVentral Cochlear Nucleus
RMS	Root Mean Square
S	Skew
SPL	Sound Pressure Level
v	Variance

Table of contents

List of figures	xvi
List of tables	xviii
1 Introduction to sound textures	1
1.1 Importance of natural sound stimuli: Brief literature review	3
1.2 The Big Questions	5
1.3 Auditory model	7
1.4 Definition of some statistical parameters:	11
1.5 Statistical parameters computed in the model	13
1.6 Why did we consider these statistics in our study?	18
1.7 Subcortical and cortical processing of sounds	21
1.7.1 Sound transduction in the ear	22
1.7.2 The auditory nerve	23
1.7.3 The cochlear nucleus	23
1.7.4 The superior olive	25

1.7.5	The inferior colliculus and thalamus	25
1.7.6	Role of auditory cortex	27
1.7.7	Receptive fields and organization of A1	27
1.8	Thesis overview	29
2 Exploring the Distribution of Statistical Feature Parameters for Natural		
	Sound Textures	31
2.1	Introduction	32
2.2	Methods	36
2.2.1	Sound collection	37
2.2.2	Statistical parameter extraction	38
2.3	Results	42
2.3.1	Principal Components of the Marginal Statistics of Sound Textures	44
2.3.2	Principal Components of the Cochlear Correlations of Sound Textures	51
2.3.3	Principal Components of the Modulation Power Statistics of Sound Textures	56
2.4	Discussion	60
3 Sensitivity of Auditory Midbrain Neurons to Statistical Features of Sound		
	Textures	65
3.1	Introduction	65
3.2	Materials and method	69

Table of contents

3.2.1	Animal Preparation	69
3.2.2	Stimuli selection from Principal component space	69
3.2.2.1	Sound corpus	69
3.2.2.2	Representative sound textures selection:	69
3.2.2.3	Synthesized stimuli construction	71
3.2.3	Electrophysiological recordings	77
3.2.4	Data acquisition:	79
3.2.5	Data analysis	79
3.2.5.1	Neural data quantification	79
3.2.5.2	Measuring neuronal response to the transitions in statistics	80
3.3	Results	84
3.4	Discussion	87
4	Sensitivity of Auditory Cortical Neurons to Statistical Features of Sound Textures	93
4.1	Introduction	94
4.2	Materials and Methods	98
4.2.1	Animal preparation	98
4.2.2	Stimulus construction	99
4.2.3	Electrophysiological recordings	100
4.2.4	Data acquisition	102

Table of contents

4.3	Data analysis	103
4.3.1	Quantifying the neural responses	103
4.3.2	Measuring the neuronal responses to the statistical transitions	103
4.4	Results	104
4.5	Discussion	108
5	General Discussion and Conclusion	111
5.1	General Discussion	111
	References	115
	Appendix A Publications	127
	Appendix B List of Sound Textures	130
	Appendix C Figure Permission	143

List of figures

1.1	Auditory model	7
1.2	Figures illustrating the moments of different distributions	12
1.3	Cochlear envelope marginal moments	18
1.4	Modulation powers for selected textures	19
1.5	Identification improves with more statistics	21
2.1	Workflow for exploring the statistical parameter space of a corpus of natural sounds.	39
2.2	Distribution of statistical parameter values for the entire corpus	43
2.3	Principal Components of the Marginal Parameters.	46
2.4	Comparison between the envelope statistics of “sea at night” from one end of PC1 dimension and “clock ticks” from the other end.	48
2.5	Marginals statistics of selected textures across PC2 space	50
2.6	Distribution of cochlear correlation parameter values for the entire corpus:	52
2.7	Principal Components of Cochlear Correlation Parameters.	54
2.8	Distribution of modulation power parameter values for the entire corpus.	56

2.9	Principal Components of the Modulation Parameters.	59
3.1	Selection of the natural sound textures from the PCA space	71
3.2	Revisiting the auditory model	73
3.3	Sound synthesis steps	73
3.4	Spectrograms of an example of synthesized texture	76
3.5	Step-wise morphed stimuli and IC multiunit activity for " <i>Cackling Geese</i> "	80
3.6	Response of one IC multiunit in response to 6 samples of the texture (Cackling Geese)	84
3.7	Percentage of significant IC multiunits showing transient response . .	85
3.8	Percentage of significant IC multiunits showing sustained response .	86
4.1	Examples of one AC multiunit in response to 6 samples of the texture " <i>Cackling Geese</i> "	104
4.2	Percentage of AC multiunits showing significant transient responses .	106
4.3	Percentage of AC multiunits showing significant sustained responses .	107

List of tables

3.1	List of sounds selected from PCA spaces of marginals, correlations and modulation power statistics	72
-----	--	----

Chapter 1

Introduction to sound textures

In the natural environment we rarely listen sound from a single sound source. It is usually an aggregation of different sound streams that forms a highly “*textured sound-space*” environment around us, e.g. buzzing bees, insect swarms, fire, sound of running air conditioner, keyboard typing, coughing, sneezing, speech, coffee-sipping, walking foot-steps, closing sounds of lift, whispering etc. Some of these sounds are aggregation of many similar acoustic events and have “*temporal homogeneity*” *i.e.* statistical properties of such sounds remain consistent over a period of time. This category of sounds are referred as “*sound textures*”.

A psychoacoustic study on “*sound texture*” perception by [[McDermott and Simoncelli, 2011](#)] reported that “*sound textures*” that may have seemingly limitless complexity can nevertheless be described by a finite set of stationary statistical parameters and highly realistic exemplars of such sounds can be synthesized from scratch by morphing

random noise samples to assume the spectral, modulation power and cross-frequency correlation structure characteristic of that type of texture.

Such a stochastic definition of sound textures excludes highly deterministic sounds like regular rhythms or rule-based sounds like spoken sentences and pieces of music, even if such sounds form a significant proportion of sounds in natural and man-made environment.

Many other parametric descriptions of sounds have been described, including, for example, *Mel-frequency cepstral coefficients* (MFCC), *band energy ratio*, *spectral flux* and the *wavelet subspace cepstrum* [Chachada and Kuo, 2014], and these have found application in machine learning based sound event classification and computational auditory scene analysis (CASA), but these schemes are very general, and unlike the statistical parameters developed by [McDermott and Simoncelli, 2011], these schemes do not take advantage of the fact that the statistical parameters of textures are stationary over the entire duration of the sound.

[McDermott and Simoncelli, 2011] were able to show that natural sounding textures can be synthesized de novo by “*shaping-noise*” to impose the statistical features of the desired sound on random noise samples. Synthesized sounds from “*white noise*” by [McDermott and Simoncelli, 2011] were often easily identifiable as exemplars of a particular type of natural sound, and in many cases indistinguishable from a natural recording. The statistical parameters they adopted in their study were inspired by knowledge of the peripheral filtering of sounds performed by the peripheral and central

1.1 Importance of natural sound stimuli: Brief literature review

auditory system. The details of their model, different statistics computed and their importance are elaborated in section 1.3-1.6.

In the section 1.1, the importance of natural stimuli is discussed in the light of previous literature. Section 1.2 outlines the important questions that are explored in this thesis. Sections 1.5 and 1.6 deal with the auditory model [McDermott and Simoncelli, 2011], and the importance of the different types of statistical parameters in the model respectively. Section 1.7 gives an overview of auditory system for interested novice readers. Finally, section 1.8 provides a chapter-wise outline of the thesis.

1.1 Importance of natural sound stimuli: Brief literature review

Hubel and Wiesel [1977] have reported the significance of the “*right*” kind of stimuli to study the feature processing mechanism of the cortical neurons in primary visual cortex. Since then, one common understanding has developed that neurons or neural populations that respond to some “*adequate*” stimuli should have two properties. Firstly, the neurons or neural populations in question must exhibit stronger response to the “*adequate*” stimuli in comparison to other stimuli. Secondly, the parameters of “*adequate*” stimuli are mapped in some orderly way on the cortical surface [Nelken et al., 1994].

Many studies in literature have reported that auditory neurons are tuned for a number of independent feature parameters of simple stimuli including *frequency*, *intensity*,

1.1 Importance of natural sound stimuli: Brief literature review

amplitude modulation, frequency modulation, and binaural structure.

However, beginning with spectral analysis at the level of cochlea up to cortex, auditory stimuli goes through various transformations in the auditory pathway, and most of these transformations are not well understood [Aitkin, 1990; Ehret and Merzenich, 1988; Sachs and Blackburn, 1991; Yeshurun et al., 1985]. Two types of organizations are reported at the level of auditory cortex, the **tonotopic gradient** and the **binaural interaction bands**. Tonotopic order at the level of auditory nerve [Aitkin, 1990; Schreiner and Merzenich, 1988] is also preserved at the level of auditory cortex [Goldstein and Abeles, 1975; Goldstein Jr et al., 1970; Merzenich et al., 1975]. Reports on binaural interaction map suggest that ipsilateral stimulation either facilitates or suppresses the contralateral stimulation [Middlebrooks and Pettigrew, 1981]. Binaural interaction is also reported at the level of *superior olive* (SO). These organizations indicate that computations performed at the lower auditory stations are preserved at other higher auditory stations. Study on these two organizations have mostly used single tones or some simple modifications to simple tones.

Some other studies on auditory cortical units using more complex stimuli have been reported (e.g. FM sweeps: [Heil et al., 1992; Shamma et al., 1993; Whitfield and Evans, 1965]; AM sounds: [Schreiner and Urbas, 1986, 1988]; Harmonic complexes:[Schreiner et al., 1938; Schwarz and Tomlinson, 1990]; species-specific natural calls:[Newman and Wollberg, 1973]; speech sounds:[Steinschneider et al., 1990]. Similar studies by [Newman and Wollberg, 1973; Whitfield and Evans, 1965] have reported that there

are auditory units which are unresponsive to simple tones but respond to wide range of complex stimuli.

Tonotopic gradient simply cannot explain the response behaviour of such auditory units because these are built upon single tone maps or some simple modifications of these tone maps. Recent reverse correlation study by [Theunissen et al., 2000] have reported the superiority of spectrotemporal receptive field (STRFs) using natural sound over white noise [DeCharms et al., 1998] in modelling the response properties of higher auditory neurons.

1.2 The Big Questions

1. The natural sound texture space may at first glance seem limitless, it is nevertheless useful to ask whether all this variety of environmental sound can be captured in a more or less bounded, finite parameter space with knowable parameter distributions. If so, estimating the parameter distributions that characterize perhaps not all but at least a very large portion of the diversity of environmental sounds could be very useful, as it would allow us to ask how sound stimuli used in psychoacoustic or physiological studies of the auditory system relate to the types of sounds the auditory system actually encounters, and may have adapted to. Here I describe my attempt to characterize these distributions by collecting and statistically analysing a large corpus of natural sound textures.

2. McDermott and Simoncelli [2011] hypothesized that the sensitivity to each of these types of statistical features may already be present at the level of the auditory midbrain, but the extent to which neurons in the *inferior colliculus* (IC) are already sensitive to each of these statistical features have not yet been examined experimentally.
3. It is also unknown whether the neurons in the *auditory cortex* (AC) are sensitive to each of these statistical features.

The second and third objectives aim to explore how pervasive sensitivity to each of these statistical feature types is at the level of the IC and AC. If IC or AC neurons are sensitive to a particular statistical sound texture feature, then it is expected to see a change in neural responses when the parameters for the statistical feature under investigation suddenly changes even if all other properties of the sound remain essentially unaltered. In contrast, if the neuron is deaf to that feature, its response should remain unchanged.

This thesis adopts the auditory model and the statistical features elaborated by [McDermott and Simoncelli, 2011], and a description of the model and the statistics provide essential background material. It is also necessary to realize why these statistics are important and hence discussed in 1.3 and 1.5-1.6.

1.3 Auditory model

The auditory model by [McDermott and Simoncelli, 2011] which is used in this study is shown in figure 1.1. This model is influenced by their understanding of “visual texture synthesis” from some studies by [Heeger and Bergen, 1995; Portilla and Simoncelli, 2000; Zhu et al., 1997]. This model implements some significant properties of auditory pathway beginning from cochlea up to thalamus as discussed below:

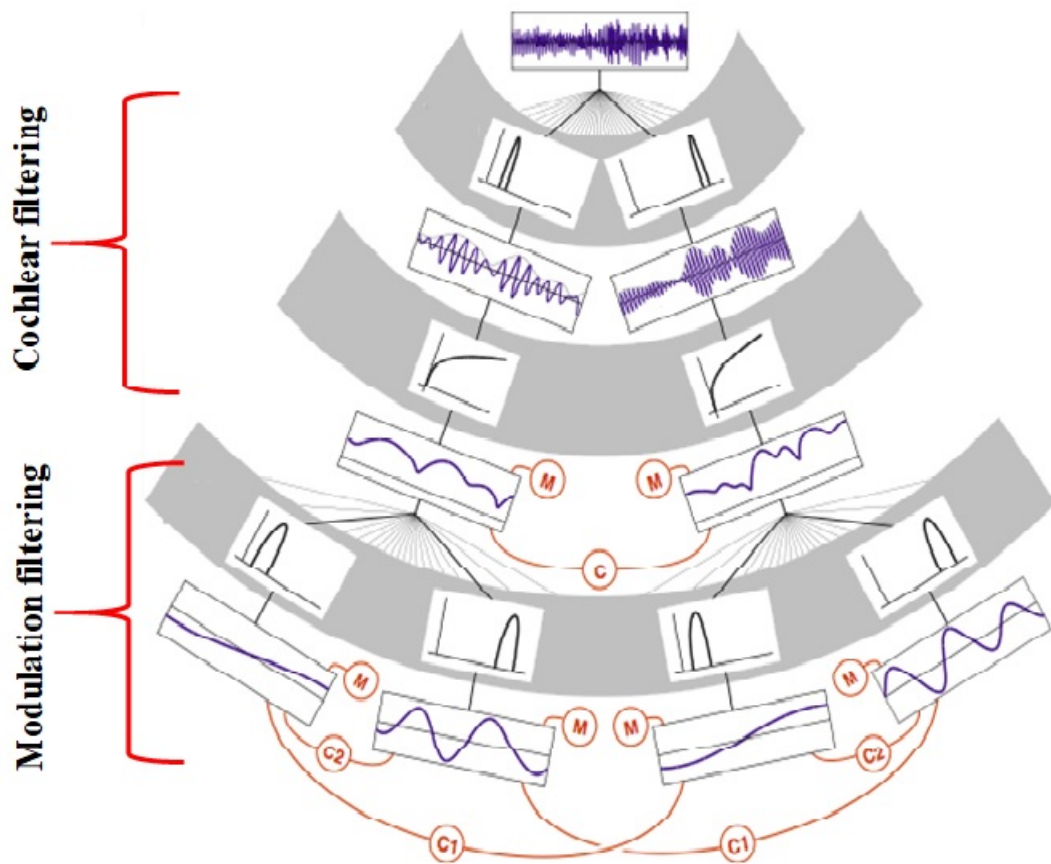


Fig. 1.1 Auditory model from [McDermott and Simoncelli, 2011] Raw sound waveform is filtered by a bank of *log* spaced band pass filters. The cochlear filter responses are subjected to power compression by a factor of 0.3 to mimic *cochlear transduction* process. From the compressed cochlear envelopes marginal moments and cross band correlations are computed. The compressed envelopes again goes through the second stage of *log* spaced bandpass modulation filters. From the modulation filter response, modulation power and modulation correlations are computed.

1. Cochlea as a frequency analyser:

Environmental sounds stimulate the cochlea, via vibrations of the stapes, the innermost of the middle ear ossicles. Sounds produce displacement waves and travel on the elongated and spiral *basilar membrane* (BM). The location of maximum BM motion is a function of stimulus frequency, with high-frequency waves being localized to the “*base*” of the cochlea and low-frequency waves approaching the “*apex*” of the cochlea. Cochlea behaves like a frequency analyser [Robles and Ruggero, 2001] as specific cochlear site respond maximally to specific frequency also known as the “**characteristic frequency**” (CF).

Properties of the filters used in the model which mimic cochlear filtering effects:

1. A bank of 30 bandpass zero-phase filters whose *Fourier* amplitudes are shaped as positive portion of raised *cosine* function.
2. The center frequencies of these filters are equally spaced on an *Equivalent Rectangular Bandwidth* ERB_N scale [Glasberg and Moore, 1990] spanning 52-8844 Hz.
3. All the filter banks are invertible.

Many studies related to natural sounds [Gygi et al., 2004; Shannon et al., 1995; Smith et al., 2002], show that information in envelopes are useful for signal reconstruction which are perceptually indistinguishable from their original counterpart. Hence amplitude envelopes are extracted from the cochlear frequency bands

after passing the cochlear responses through a low-pass filter. All the cochlear envelopes are downsampled to 400 Hz for computational efficiency.

The filter banks are used to mimic the mechanical filtering property of the cochlea. Modulation tuning is also reported at the level of auditory midbrain. The modulation filter bank is consistent with previous auditory models [Bacon and Grantham, 1989; Dau et al., 1997]. The second set of filters are also used to mimic the midbrain neurons. The log spaced filter banks cover the frequency spectrum of the auditory range of the animal under study and are in line with the previously reported studies [Joris et al., 2004]. Both the cochlear and modulation filters in the model had bandwidths that increased with their center frequency (such that they were approximately constant on a logarithmic scale), as is observed in biological auditory systems. The filters did not replicate all aspects of biological auditory filters, but perfectly tiled the frequency spectrum. [McDermott and Simoncelli, 2011] have shown that increasing the number of filters by four times the current number of filters does not increase the quality of synthesis of sounds significantly.

2. Cochlear transduction:

The extracted envelopes after cochlear filtering may be subjected to *power* or *log* compression to imitate the cochlear transduction mechanism. The cochlea shows a **”compressive nonlinearity”** *i.e.*: its response to high intensity sounds is proportionally smaller than that to low intensity sound due to non-linear, level-

dependent amplification. In this model, power compression by a factor of 0.3 has been applied to envelopes. Statistics such as the mean (m), variance (V), skew (S), kurtosis (K) and cochlear correlation (C) between bands are computed from these compressed cochlear envelopes. The Mean and variance are known as first and second order moments respectively whereas the skew and kurtosis are known as the third and fourth order moments respectively.

3. Modulation property of mid brain and thalamic neurons:

Some studies [[Baumann et al., 2011](#); [Joris et al., 2004](#); [Miller et al., 2002](#); [Rodríguez et al., 2010](#)] have reported that modulation tuning is observed in midbrain and thalamic neurons. [Nelken et al. \[1999\]](#) have also reported cross band correlation as a major source of variation in natural sounds.

Properties of the filters used in the model for midbrain and thalamic neurons:

1. A bank of 20 bandpass zero-phase filters whose *Fourier* amplitudes are shaped as positive portion of raised *cosine* function.
2. The centre frequencies are in the range of (0.5 - 200 Hz) and are equally spaced on the *log* scale.

The modulation filters that are used in this model are consistent with the previously reported human auditory model [[Dau et al., 1997](#)] and are also broadly consistent with animal models [[Miller et al., 2002](#); [Rodríguez et al., 2010](#)].

1.4 Definition of some statistical parameters:

Let us consider a random dataset X , consisting of values, $[x_1, x_2, \dots, x_n]$. The weighted mean, variance, skew and kurtosis of the dataset X are computed as given in the following equations.

a. **weighted mean:**

$$\mu = \sum_{i=1}^n w_i x_i, x_i \in X \text{ and } \sum_{i=1}^n w_i = 1 \quad (1.1)$$

b. **weighted variance:**

$$v = \sum_{i=1}^n w_i \cdot (x_i - \mu)^2 \quad (1.2)$$

c. **weighted skew:**

$$S = \frac{\sum_{i=1}^n w_i \cdot (x_i - \mu)^3}{\sigma^3} \quad (1.3)$$

d. **weighted kurtosis:**

$$K = \frac{\sum_{i=1}^n w_i \cdot (x_i - \mu)^4}{\sigma^4} \quad (1.4)$$

σ : standard deviation of the dataset X .

The mean and variance are the first two statistical moments and give information about the central tendency and about the spread of a distribution. The skew and kurtosis

1.4 Definition of some statistical parameters:

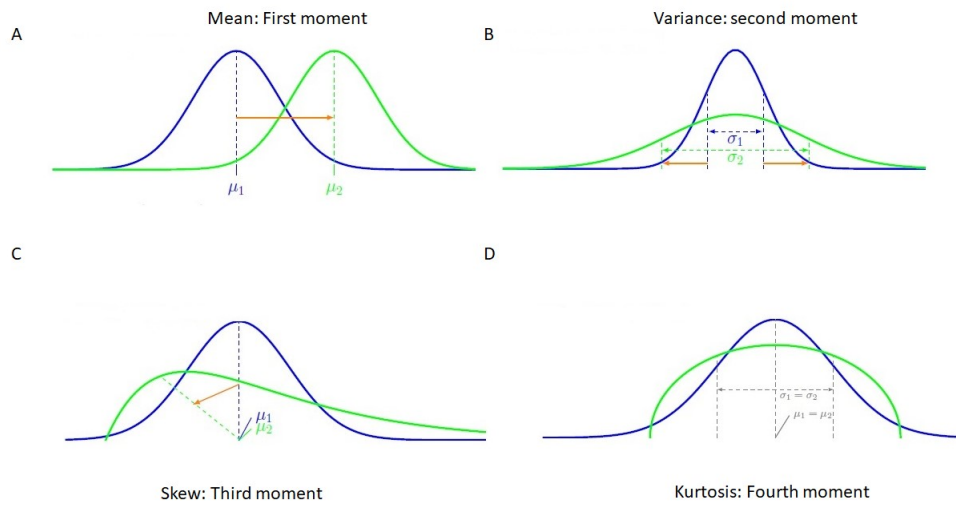


Fig. 1.2 Illustration of moments of distribution. A. Two distributions differing in their mean. B. Two distributions differing in variance. C. Positively Skewed data D. Kurtosis of the dataset. Skewness and kurtosis are considered as shape statistics. Skewness measures the symmetry about the mean whereas kurtosis measures the "flatness" or "peakedness". If data has a long tail towards positive x-axis, it is said to be "positively skewed" and if a long tail is towards negative x-axis then data is said to be "negatively skewed". If a distribution looks more flat then it is said to be "platykurtic" and if is more peaked then it is said to be "leptokurtic".

are the third and fourth moments respectively and are also known as the "shape statistics" as they provide information about the shape of the distribution. Skewness is a measure of the "symmetry" of the shape of a distribution. If a distribution is symmetric, the skewness will be zero. If there is a long tail in the positive direction, skewness will be positive, while if there is a long tail in the negative direction, skewness will be negative. The kurtosis on the other hand is a measure of the "flatness" or "peakedness" of a distribution. The flat-looking distributions are referred to as *platykurtic*, while the peaked distributions are referred to as *leptokurtic*. In their model, [McDermott and Simoncelli, 2011] they found that the envelopes of the sound textures present in their sound corpus have marginal distribution. Therefore, the marginal parameters (mean, variance, skew and kurtosis) have been used in the model to explain the marginal

distribution of the envelopes.

1.5 Statistical parameters computed in the model

1. Parameters computed from compressed cochlear envelopes

In their model, both the cochlear and modulation filtering operations are performed in the discrete frequency domain, and circular boundary conditions have been assumed.

As per [McDermott and Simoncelli, 2011] to avoid boundary artifacts, the statistics measured in original recordings were computed as weighted time-averages. The weighting window fell from one to zero (half cycle of a raised cosine) over the 1 s intervals at the beginning and end of the signal (typically a 7 s segment), minimizing artificial interactions. From figure 1.1 consider k_{th} . cochlear subband as C_k .

Consider half cycle of raised *cosine* as a *window function*, and $\sum_t w(t) = 1$.

$b_{k,n}$: n^{th} . modulation band of cochlear envelope $C_k(t)$ computed via convolution with filter f_n .

For the subband envelope C_k , weighted mean, variance, skew and kurtosis are computed as given below:

1.5 Statistical parameters computed in the model

a. **weighted mean**

$$\mu_k = \sum_t w(t) C_k(t) \quad (1.5)$$

b. **weighted average of variance:**

$$v = \frac{\sum_t w(t) (C_k(t) - \mu_k)^2}{\mu_k^2} \quad (1.6)$$

The variance was normalized by the squared mean, so as to make it dimensionless like the skew and kurtosis.

c. **weighted average of skew:**

$$S = \frac{\sum_t w(t) (C_k(t) - \mu_k)^3}{\sigma_k^3} \quad (1.7)$$

d. **weighted average of kurtosis:**

$$K = \frac{\sum_t w(t) (C_k(t) - \mu_k)^4}{\sigma_k^4} \quad (1.8)$$

e. **Cochlear cross band correlation statistics:**

The weighted correlation between subband envelopes $C_k(t)$ and $C_j(t)$ is measured as follows:

$$Cr_{j,k} = \frac{w(t) (C_j(t) - \mu_j) (C_k(t) - \mu_k)}{\sigma_j \sigma_k} \quad (1.9)$$

1.5 Statistical parameters computed in the model

μ_k and μ_j : means of the subbands C_k and C_j respectively.

σ_k and σ_j : standard deviation for the subbands C_k and C_j respectively.

They mention that the mathematical form of cochlear correlation that has been used in their model is not the unique way of specifying neural instantiation.

Cochlear correlations can also be computed as squared sums and difference which are common in functional models of neural computation [[Adelson and Bergen, 1985](#)]

2. Parameters computed from the envelopes after modulation filtering.

a. Modulation power:

For n^{th} the modulation band of cochlear envelope S_k modulation power is measured as:

$$Modulation\ power_{k,n} = \frac{\sum_t w(t) b_{k,n}(t)^2}{\sigma_k^2} \quad (1.10)$$

Modulation power has been normalized by the variance of the corresponding cochlear envelope in order to make measured statistics independent of cochlear statistics. Thus modulation power here represents the proportion of total envelope power captured by each modulation band.

1.5 Statistical parameters computed in the model

b. Modulation correlation statistics (C1):

This is computed between bands centred on the same modulation frequency but different acoustic frequency. Between modulation band $b_{k,n}$ and $b_{j,n}$, the weighted modulation correlation CI is measured as follows:

$$CI_{jk,n} = \frac{w(t) b_{k,n}(t) b_{j,n}(t)}{\sigma_{j,n} \sigma_{k,n}}, j \in [1..32], (k-j) \in [1..2], n \in [1..7] \quad (1.11)$$

$$\sigma_{k,n} = \sqrt{\sum_t w(t) b_{k,n}(t)^2} \quad (1.12)$$

$$\sigma_{j,n} = \sqrt{\sum_t w(t) b_{j,n}(t)^2} \quad (1.13)$$

c. Modulation correlation (C2):

This is measured between bands of different modulation frequencies but derived from the same acoustic frequency which measures phase relationship between the modulation frequencies.

Conventional measurement of temporal asymmetry do not preserve phase information. Hence a complex-valued correlation measure has been used in the model [Portilla and Simoncelli, 2000]. In signal processing, analytic signal is a complex-valued function that has no negative frequency components. Both the real and imaginary parts of an analytic signal are real-valued functions related to each other by *Hilbert transform*. Due to *Hermitian*

1.5 Statistical parameters computed in the model

symmetry of the Fourier spectrum the negative frequency components are redundant. These negative frequency components can be removed without any loss of information. [Smith, 2007].

The analytic extension of modulation band $b_{k,n}$ is represented as:

$$a_{k,n}(t) \equiv b_{k,n}(t) + iH(b_{k,n}(t))$$

where, $H = \text{Hilbert transform}$

$$i = \sqrt{-1}$$

$C2$ is measured as the correlation between analytic modulation bands tuned to modulation frequencies an octave apart, with the frequency of the lower band doubled. Frequency doubling is done by squaring the complex-valued analytic signal:

$$d_{k,n}(t) = \frac{a_{k,n}^2(t)}{\|a_{k,n}\|} \quad (1.14)$$

$$C2_{k,mn} = \frac{\sum_t w(t) d_{k,m}^*(t) a_{k,n}(t)}{\sigma_{k,m} \sigma_{k,n}}, k \in [1..32], m \in [1..6] \quad (1.15)$$

In equation 1.14-1.15, $*$ and $\| \|$ represent complex conjugate and modulus operator respectively.

1.6 Why did we consider these statistics in our study?

Variability of these statistics across the natural sound textures has been reported by [McDermott and Simoncelli, 2011]. In their study they have explored the statistics of 168 natural sound textures. They show that both marginal and correlation statistics vary substantially across natural sound textures as shown in figures 1.3 and 1.4. Cochlear marginal moments for entire sound corpus are shown in Figure 1.3. Cochlear envelope correlations for sounds “Fire”, “applause” and “Stream” has been shown in figure 1.4B. Modulation power statistics for “insects”, “Waves” and “Stream” are shown in figure 1.4A. Modulation correlations $C1$ and $C2$ are shown in figure 1.4C and 1.4D respectively.

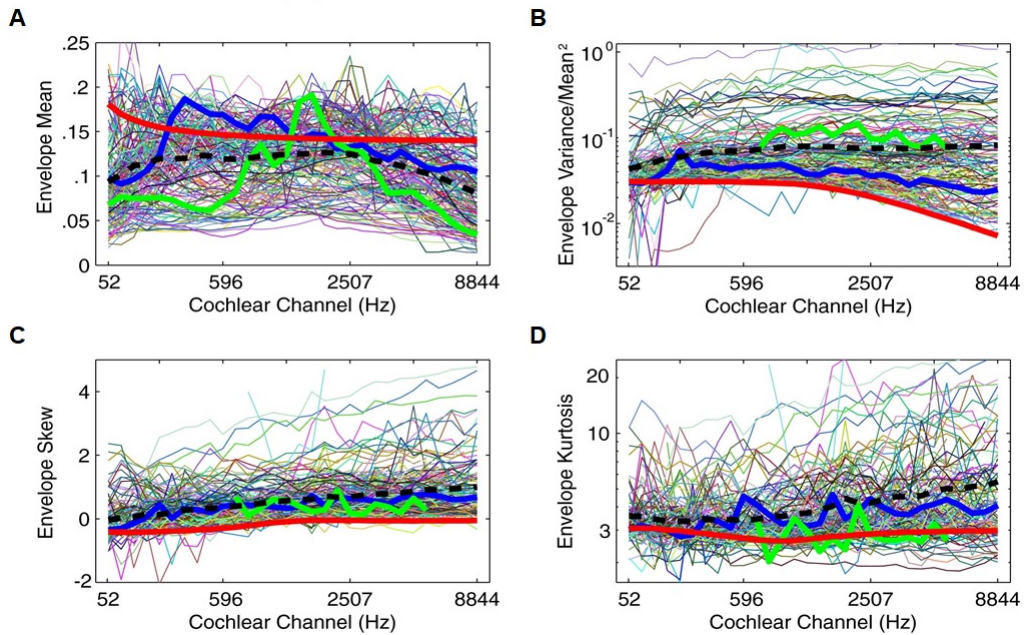


Fig. 1.3 Reproduced from [McDermott and Simoncelli, 2011] Cochlear envelope marginal moments of 168 natural sound textures (A) Envelope mean (B) Envelope variance normalized by square of the envelope mean (C) Envelope skewness (D) Envelope kurtosis. **Thick lines:** Blue: “Stream”, Green: “Geese calls”, Red: “Noise”. Black dotted: Mean value of each statistics across all sounds. Thin lines: All 168 natural sound textures considered in the study.

1.6 Why did we consider these statistics in our study?

As per [McDermott and Simoncelli, 2011] all these statistics have distinct contribution in identifying different sounds in “*natural sound texture space*”. Sparse sounds (e.g. “*geese calls*”, and “*hedgehog*”), are usually with a burst of energy and are efficiently captured by marginal moments. Envelope correlations are measured both after cochlear filtering and modulation filtering. Cochlear envelope correlations (C) can distinguish sound textures like “*applause*”, “*fire*” which are widely correlated across many frequency bands to sound textures like “*stream*”, “*water tape*” which are poorly correlated across frequency bands.

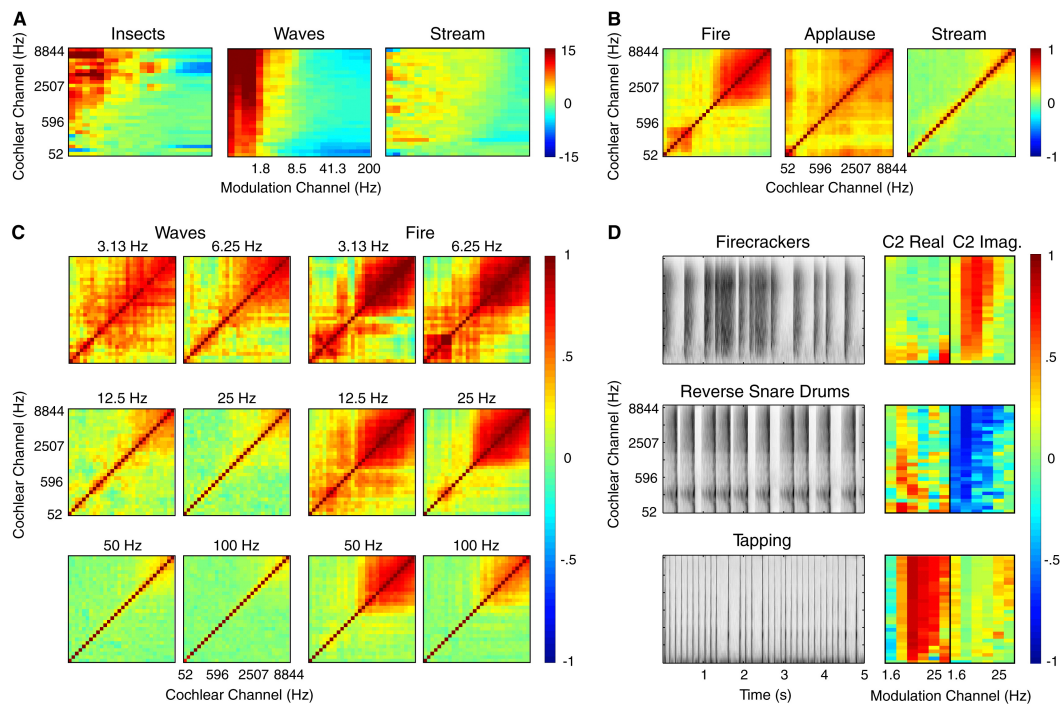


Fig. 1.4 Reproduced from [McDermott and Simoncelli, 2011] (A) Modulation powers for “*Insects*”, “*Waves*” and “*Stream*”. Modulation has been normalized by their corresponding cochlear envelope variance. (B) Cochlear envelope correlations (C) for “*Fire*”, “*Applause*” and “*Stream*”. (C) Modulation Correlations (C1) for “*Waves*” and “*Fire*”. (D) Modulation Correlations (C2) for “*Firecrackers*”, “*Reverse Snare Drums*” and “*Tapping*”.

1.6 Why did we consider these statistics in our study?

Sound textures like “*waves*” and “*wind*” have highly correlated C1 for lower modulation frequencies, whereas for sound like “*fire*”, have widely correlated C1 across all modulation frequencies. Sounds like “*fire crackers*”, “*bomb explosions*”, “*tapping*” have sudden onsets-offsets have higher C2 across all acoustic frequencies. Modulation powers are measured over the envelopes after modulation filtering. This measures the amount of amplitude modulation of the cochlear envelopes in given modulation frequency bands. Insect sounds like “*buzzing bee*”, “*mosquito whine*” have a lot of modulation power at high modulation rates, while sounds like “*waves*” have more of their modulation power at slower rates. These different types of statistics can be thought of as forming a “**hierarchy**”, given that the auditory system might be able to measure the marginal from observing the activity of auditory nerve fibres individually, whereas cochlear correlations require information to be combined across processing channels along with the tonotopic array. Similarly, modulation power requires that envelopes are extracted first, making it in a sense a “higher order” statistic than the marginals of the envelopes.

1.7 Subcortical and cortical processing of sounds

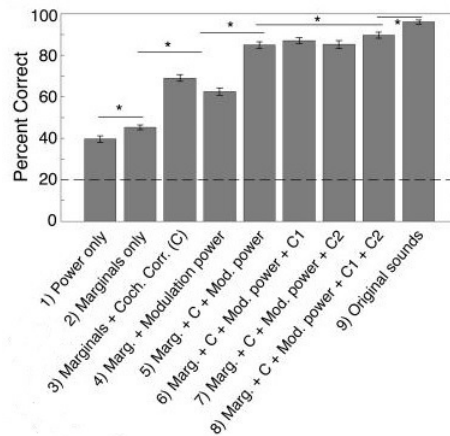


Fig. 1.5 Reproduced from [McDermott and Simoncelli, 2011]. Identification of sound textures improves as more statistics are included in the synthesis process. Asterisks (*) denote significant differences between conditions ($p < 0.01$ paired t-tests corrected for multiple comparison). Error bars denote standard error bars and dashed line denotes the chance level of performance.

In their study they have also demonstrated that by systematic imposition of these statistics, “natural-like” sound textures can be synthesized from “noise”. As shown in Figure 1.5, perceptual recognition for different sound textures improves as more and more statistics are added to the synthesis process.

1.7 Subcortical and cortical processing of sounds

Sound traverses through the peripheral and subcortical auditory pathway before it arrives in the auditory cortex. In the auditory pathway sound is subjected to many auditory processing in different auditory station. Since there are multiple pathways of afferent and efferent projections among sub cortical and cortical areas [JA, 2007], auditory processing at different auditory stations simply cannot be linear. More precisely function of these multiple auditory pathways are poorly understood and are beyond the scope

1.7 Subcortical and cortical processing of sounds

of this thesis. This thesis also assumes linear processing of information in different auditory stations which is an oversimplification of auditory system. To orient the reader, the following section provides a brief overview of processing of sound in important auditory stations of mammalian auditory system from ear to cortex as studied in many literature.

1.7.1 Sound transduction in the ear

Objects make sounds and different objects make different sounds. Sound waves created by different objects carry valuable clues about the physical properties about the objects. Our auditory brain can extract information from these pressure waves so effortlessly to recognize these objects. Sound pressure waves passes through the ear canal and then reach the tympanic membrane or ear drum. The purpose of the ear drum is to separate the middle ear from the outer. Middle ear which has three small bones ("*malleus*", "*incus*" and "*stapes*") are also called as "*ossicles*". As sound travels from air-filled space of middle ear to fluid filled (watery) space of cochlea it faces greater resistance because acoustic impedance of water is much higher than air.

The mechanical properties of middle ear ossicles ensure that sound waves are efficiently transmitted to the fluid filled interior of the cochlea. Mammalian cochlea is a coiled structure. Outer cochlear wall consists of solid bone with membrane lining. The only openings to the bony shell of the cochlea are the "*oval window*" and "*round window*". The "*stapes*" pushes against the oval window as it vibrates to and fro to the rhythm of the sound and thereby increases the pressure of the fluid filled cochlear space.

1.7 Subcortical and cortical processing of sounds

The entire fluid filled space of cochlea is separated into two separate compartment by the basilar membrane. Near the oval and round windows of the cochlea basilar membrane is narrow, thick and stiff while at the apical end it is wide, thin and floppy. Due to differential mechanical property of basilar membrane different part of basilar membrane resonate with different frequency making it an efficient mechanical frequency analyser. The basilar membrane essentially performs a mechanical frequency analysis on the incoming sound, thereby setting up a topographic map of sound frequency (i.e. a “tonotopic” map). The frequency acuity of this map is refined by active mechanisms in the outer hair cells of the cochlea.

1.7.2 The auditory nerve

Upon leaving the cochlea, the auditory nerve fibres join the *vestibulocochlear nerve* (VIII cranial nerve). Auditory nerves have been reported to adopt different coding strategies to preserve different information in the sound. The “*tonotopic gradient*” observed in the auditory nerve fibres is preserved across the auditory pathway up to the level of cortex. As auditory nerves use “*place code*” to preserve frequency information they also use “*rate code*” to represent the intensity information. “*Temporal coding*” has also been reported to preserve phase information of sound up to $1.5kHz$.

1.7.3 The cochlear nucleus

Auditory nerve fibers after entering into cochlear nucleus in the brain stem immediately bifurcate. The ascending branch enters into the *anteroventral cochlear nucleus* (AVCN)

1.7 Subcortical and cortical processing of sounds

and the descending branch runs through the *posteroventral* (PVCN) to the *dorsal cochlear nucleus* (DCN). There, AN fibers contact different populations of second order cell types, which are distinguished by their anatomical location, morphologies, cellular physiologies and firing properties [Brawer et al., 1974; Pfeiffer, 1966; Young and Brownell, 1976]. The *AVCN* is populated with spherical and globular *bushy cells*. The firing pattern of *bushy cells* is very much similar to the auditory nerve fibres to which they are connected. *Bushy cells* therefore preserve information in the temporal firing pattern of the auditory nerve fibres. Star-like *stellate cells* in *AVCN* and *PVCN* respond to pure-tone stimuli with rhythmic bursts. These neurons do not preserve the timing of their input spikes but they have narrow frequency tuning. They appear to be sensitive to more complex spectral profiles than auditory nerve fibres. The onset cells in *PVCN* which are either *stellate or octopus* shaped often respond to pure-tone bursts with just single action potential at the stimuli onset. These cells are broadly tuned and receive convergent inputs from many nerve fibres. These cells represent the fundamental frequency of a complex sound as the reciprocal of their interspike intervals, thereby converting the all-order interspike interval code of stimulus periodicity in auditory nerve fibres into a first-order interspike interval pitch code [Winter et al., 2001]. These cells can also provide more details about the time structure of a complex tone. Cells in *DCN* with pyramidal morphology are “*excited*” by some frequencies and “*inhibited*” by some other frequencies. *VCN* cells are more specialized to process temporal structure of the sound whereas *DCN* cells play important role in detecting spectral contrast.

1.7.4 The superior olive

Though almost all of the Cochlear nucleus output reaches the first major midbrain auditory processing station *inferior colliculus*, the output of the bushy cells of *AVCN* project to superior olivary complex. Auditory signals from the two ears are first combined at the level of the superior olivary nucleus of the brainstem. Spatial cues such as *inter aural level difference* (ILD) and *inter aural time difference* (ITD) are important for sound localization. Processing of these spatial cues are carried out by separate area at the level of brainstem. The neurons of *lateral superior olive* (LSO) are biased toward high-frequencies. *ILD* being a high-frequency sound localization cue are mostly handled by *LSO* neurons. *LSO* neurons are excited by sounds from ipsilateral ear but are inhibited by sounds from contralateral side. The excitatory inputs are directly received from the bushy cells in the *AVCN* whereas the inhibitory inputs are received from the neurons in the *medial nucleus of the trapezoid body* (MNTB), which in turn receive their input from the bushy cells in *AVCN* of the contra-lateral side [Phillips and Irvine, 1981; Sanes, 1990]. *ITD* cues are processed in the *medial superior olive* (MSO) using the relative timing of inhibitory inputs from cochlear nuclei on both sides of the brainstem [Brand et al., 2002].

1.7.5 The inferior colliculus and thalamus

Axons emerging from cochlear and olivary nuclei travel along lateral lemniscus to connect to the inferior colliculus. Neurons in the midbrain and cortex are more excited

1.7 Subcortical and cortical processing of sounds

to sounds presented contralaterally, as the paths between *CN* and *IC* are predominantly crossed. The left and right *IC* have commissural connection between them. The central nucleus of *IC* (*ICc*) receives most of the brainstem inputs. The *ICc* is surrounded by dorsal nucleus (*ICd*), external nucleus (*ICx*) and the nucleus of the brachium of the *IC* (*BIC*). Axons leaving from *IC* nuclei are mostly connected to the *median geniculate body* (*MGB*), which is a significant relay centre of thalamus. Tonotopic gradient is clearly observed in *ICc* and *ventral medial geniculate body* (*Mgv*), hence said to be “*lemniscal*” structure of midbrain. *ICx*, *BIC* and *dorsal medial geniculate body* (*Mgd*) do not show clear tonotopic organization, hence termed as “*paralemniscal*”. Signals arriving to the *IC* are already considerably pre-processed. *ILD* and *ITD* representations established in the *superior olive* also undergo further processing in the *IC*. These binaural cues remain to be encoded in largely segregated neural populations, although recent studies have shown that some *IC* neurons are sensitive to both *ILD* and *ITD* cues [Chase and Young, 2005; Langner et al., 2002; Langner and Schreiner, 1988] have suggested that a “*periodotopic*” map of best modulation frequency exists in *IC*, running orthogonal to the direction of the tonotopic map. *MGB* acts as a “*relay station*” between midbrain and cortex. Conditioning of emotional reactions to acoustic stimuli are reported to be regulated by the amygdala which receives thalamic fibres from *MGB* [Clugnet et al., 1990; Farb and Ledoux, 1997].

1.7.6 Role of auditory cortex

Primary auditory cortex is reported to act as an auditory object aggregator or novelty detector [Nelken and Bar-Yosef, 2008; Nelken et al., 2003]. Research data about auditory cortex is still at its nascent stage and hence very much inconclusive. Studies from [Schulze et al., 2002; Schulze and Langner, 1997] have reported that *A1* stores multiple maps of complex stimulus attributes such as pitch. Many recent studies report that *A1* is more suitably tuned to sound dynamics of natural sounds and it is more species-specific [Garcia-Lazaro et al., 2006]. It has also been reported that plasticity of *A1* neurons are responsible for perceptual learning [Dahmen and King, 2007] and directing attention [Fritz et al., 2007].

1.7.7 Receptive fields and organization of A1

Tonotopic organization beginning at peripheral auditory system is preserved throughout auditory pathway and is also preserved at the level of *A1* [Kelly et al., 1986; Merzenich et al., 1975]. Many studies report that *A1* as aggregator of auditory objects integrates acoustic information over relatively wide temporal windows [Joris et al., 2004; Liu et al., 2006; Miller et al., 2002; Wallace et al., 2002]. An imaging study [Boemio et al., 2005] on human auditory cortex has reported that the temporal window for auditory object formation widens from 25-50ms to 200-300ms progressing from primary auditory cortex to higher auditory cortical regions. Studies from [Depireux et al., 2001; Linden et al., 2003; Theunissen et al., 2000] have reported that *A1* neurons have preferential

1.7 Subcortical and cortical processing of sounds

reference to specific spectral and temporal features of acoustic signals. Receptive field properties of neurons across the *AI* surface have been reported in many studies. Many studies found these to be non-random on *AI* surface. [Heil et al., 1992; Recanzone et al., 1999; Schreiner and Mendelson, 1990; Shamma et al., 1993] have reported gradual change of spectral bandwidth along the axis orthogonal to *CF* gradient. [Recanzone et al., 1999; Shamma et al., 1993] have studied the symmetry of spectral response filter properties of neurons around their *CF* along the isofrequency bands. Most of these neurons have symmetric spectral response filters, i.e. neurons which are at one end of the isofrequency band are more inhibited by frequencies below the *CF*, whereas neurons at the other end of the isofrequency bands are more inhibited by frequencies above the *CF*. Similarly, [Cheung et al., 2001; Heil et al., 1994, 1992; Recanzone et al., 1999] have reported a non-random organization of stimulus intensity and first-spike latency within the *AI* isofrequency bands. Distribution of binaural sensitivities across the *AI* surface has also been reported to be non-random [Imig et al., 1977; Kelly and Judge, 1994; Middlebrooks and Pettigrew, 1981]. [Hall and Goldstein Jr, 1968; Middlebrooks and Pettigrew, 1981; Mrsic-Flogel et al., 2001; Schnupp et al., 2001] reported that *AI* neurons are not biased towards specific spatial cue. Rather they are tuned to specific direction along the azimuthal plane and take input from a number of integrated binaural and monaural directional cues.

1.8 Thesis overview

This thesis is built upon my understanding and findings on sensitivity of midbrain and cortical neurons to statistical features of sound textures and also explores the distribution of statistical features of natural sound texture space. In particular, neural responses to specially designed stimuli that transits through different statistics are recorded from inferior colliculus and auditory cortex in anaesthetised animals to assess what percentage of neural population are sensitive to any of the statistical features presented in the stimuli. Experimental Chapters 2-4 progress chronologically through investigations of these research questions that I have undertaken throughout my Ph.D. Studies.

Chapter 2: Exploring the Distribution of Statistical Feature Parameters for Natural Sound Textures

In this chapter, I have analyzed the marginal statistics (mean, variance, skew, and kurtosis), as well as cochlear envelope correlations and modulation power statistics of a sound corpus of 200 sound textures. Using principal component analysis, I explored the distributions of these statistical parameters. This study suggests that large “acoustic variability” of natural sound texture space can be compensated by significantly small “statistical variability”.

Chapter 3: Sensitivity of Auditory Midbrain Neurons to Statistical Features of Sound Textures

The objective of this chapter is to explore the pervasive sensitivity to different statistical features observed at the level of the IC. To determine

this, I recorded extracellular responses of IC multiunits with silicon array electrodes implanted into the IC of anaesthetized young adult female wistar rats to synthesized stimuli. I found that subcortical processing of auditory textures may already be sufficient to encode all the types of statistical features identified by [McDermott and Simoncelli, 2011] as being important in identifying and discriminating natural sound textures. I found that ~80% of the IC multiunits were sensitive to the statistical features.

Chapter 4: Sensitivity of Auditory Cortical Neurons to Statistical Features of Sound Textures

This chapter describes the sensitivity of auditory cortical neurons to different statistical parameters of natural sound textures. For this I recorded extracellular responses of auditory cortical multiunits with silicon array electrodes implanted into the AC of anaesthetized young adult female wistar rats. I found that around ~5-20% of cortical neural population are sensitive to the statistical features present in natural sound texture stimuli.

Chapter 2

Exploring the Distribution of Statistical Feature Parameters for Natural Sound Textures

Abstract

Sounds like “running water” and “buzzing bees” are classes of sounds which are a collective result of many similar acoustic events and are known as “sound textures”. Recent psychoacoustic study using sound textures by [[McDermott and Simoncelli, 2011](#)] reported that natural sounding textures can be synthesized from white noise by imposing statistical features such as marginals and correlations computed from the outputs of cochlear models responding to the textures. The outputs being the envelopes of bandpass filter responses, the “cochlear envelope”. This suggests that the perceptual

qualities of many natural sounds derive directly from such statistical features, and raises the question of how these statistical features are distributed in the acoustic environment.

To address this question, we collected a corpus of 200 sound textures from public online sources and analyzed the distributions of the textures' marginal statistics (mean, variance, skew, and kurtosis), cross-frequency correlations and modulation power statistics. A principal component analysis of these parameters revealed a great deal of redundancy in the texture parameters. For example, just two marginal principal components, which can be thought of as measuring the sparseness or burstiness of a texture, capture as much as 66% of the variance of the 128 dimensional marginal parameter space, while the first two principal components of cochlear correlations capture as much as 90% of the variance in over 1000 correlation parameters. Knowledge of the statistical distributions documented here may help guide the choice of acoustic stimuli with high ecological validity in future research.

2.1 Introduction

Be it buzzing bees, a flowing river, flocks of squawking birds or howling wind, the natural world is filled with a huge diversity of different sound textures, and humans have added further to that variety with all manner of traffic and machine noises. While this variety may at first glance seem limitless, it is nevertheless useful to ask whether all this variety of environmental sound can be captured in a more or less bounded, finite parameter space with knowable parameter distributions. If so, estimating the

parameter distributions that characterize a large portion of the perceptual diversity of environmental sounds could be very useful, as it would allow us to ask how sound stimuli used in psychoacoustic or physiological studies of the auditory system relate to the types of sounds the auditory system actually encounters, and may have adapted to. Here I describe my attempt to characterize these distributions by collecting and statistically analysing a large corpus of a class of natural sounds known as sound textures. I understand sound textures in the sense popularized by [McDermott and Simoncelli, 2011], as sounds that may have a lot of complexity, like for example the sound of waves breaking on a pebble beach, but which are nonetheless fully described by a finite set of stationary statistical parameters, so that highly realistic exemplars of such sounds can be synthesized from scratch by morphing random noise samples to assume the spectral, modulation power and cross-frequency correlation structure characteristic of that type of texture. While textures defined in this way are fundamentally stochastic, and thereby exclude some important classes of sounds which are highly deterministic (such as highly regular rhythms) or rule based (such as a spoken sentence or a piece of music) they nevertheless cover a large proportion of the sounds encountered in natural and man-made environments, and the fact that they appear to be well characterized by a potentially large but finite number of stationarity statistical parameters makes the research questions I am pursuing here tractable.

Previous studies have identified a variety of parameters that are in principle suitable for characterizing sounds, including, for example, features like Mel-frequency cepstral coefficients (MFCC), band energy ratio, spectral flux and the wavelet subspace cepstrum.

These have been nicely reviewed by [Chachada and Kuo, 2014], and are often used in applications such as sound event classification and computational auditory scene analysis (CASA)[Rosenthal and Okuno, 1998]. However, cepstral coefficients tend to look at relatively short time windows, so here I chose to use the auditory texture statistics by [McDermott and Simoncelli, 2011], which were inspired by the previous characterization of features used in visual texture discrimination research, [Julesz, 1962; Julesz et al., 1978; Portilla and Simoncelli, 2000]. [McDermott and Simoncelli, 2011] were able to show that natural sounding textures can be synthesized de novo by “shaping-noise” to impose the statistical features of the desired sound on random noise samples. Synthesized sounds from “white noise” by [McDermott and Simoncelli, 2011] were often easily identifiable as exemplars of a particular type of natural sound, and in many cases indistinguishable from a natural recording. The statistical parameters they adopted in their study was inspired by knowledge of the filtering of sounds known to be performed by the peripheral and central auditory systems. In their model, the input signal is band pass filtered into a range of frequency bands which mimics cochlear filtering. The amplitude envelope of the signal in each frequency band is extracted and cochlear transduction of sound is simulated by applying compressive nonlinearity (raising the envelope by a power 0.3) to the amplitude envelopes. From the compressed cochlear envelopes, statistics such as mean, variance, skew, kurtosis and correlation between bands are computed for the amplitude distributions of these “cochlear envelopes”. The mean, variance, skew and kurtosis (also referred to as the first, second, third and fourth moment respectively) of the envelope amplitudes, will collectively be referred to

as “marginal moments”, or “marginals” of the sound texture. In addition, the pairwise correlations between cochlear envelope amplitudes (“cochlear correlations”, for short) are computed. Previous studies by [Heeger and Bergen, 1995; Portilla and Simoncelli, 2000] reported that both marginal moments and correlations are important features of visual textures, and the same is clearly true for auditory textures. Furthermore, the compressed cochlear envelopes are also passed through a second bank of band pass filters to measure the distribution of amplitude modulations (the “modulation power” statistics) and to compute correlations between modulation channels.

To appreciate how the types of statistical parameters extracted by the [McDermott and Simoncelli, 2011] model, distinguish types of sound textures, consider that some textures are “sparse”, exhibiting periods of relative silence with burst of energy of widely varying amplitude distribution (e.g. grains of hail bouncing on a tin roof), while others have a much more constant stream of sound (e.g. a high pressure jet of water rushing out of a faucet). Marginal moments will distinguish sparse from less sparse sounds easily, with sparse sounds having relatively greater variance, skew and kurtosis. Indeed, the usefulness of marginal moments in distinguishing natural sounds and images has been appreciated for a while [Attias and Schreiner, 1997; Field, 1987]. Similarly, modulation power statistics may be useful to distinguish “buzzing insects” sound textures, which have low modulation power at low frequencies but higher modulation power at higher frequency bands, from “waves on a beach” for which modulation power is relatively uniform across all modulation frequency bands.

In addition to cochlear marginals and modulation power distributions, [McDermott and Simoncelli, 2011] also examined the role of correlations, either between cochlear envelopes, or between modulation filters. Cochlear correlations (C) turned out to be a perceptually very powerful feature, discriminating, for example, the sound of applause, in which the ebb and flow of acoustic energy is highly correlated across many cochlear frequency channels, from "running water" type sounds, in which correlations between cochlear envelopes are small. Unlike cochlear correlations, modulation correlations are computed between outputs of modulation filter banks, and they come in two "flavors", cross-modulation-frequency-band (C1) and within-modulation-frequency-band (C2). But unlike cochlear correlations, modulation correlations do not appear to play an important role in auditory perception [McDermott and Simoncelli, 2011], and I shall not consider modulation correlations further in this study.

The rest of the chapter has been organized as follows. Section 2.2 describes the methodology used for statistical parameters extraction from a corpus of 200 natural sound textures. Section 2.3 describes the results. Section 2.4 concludes with a scholarly discussion of my observations.

2.2 Methods

To examine the statistical parameter space spanned by natural sound textures, I collected a corpus of natural sound recordings, computed their statistical parameters using the

[[McDermott and Simoncelli, 2011](#)] framework, and subjected the resulting database of statistical parameters to dimensionality reduction by principal component analysis (PCA). This allowed us to identify parsimonious "principal feature axes" which explain a substantial portion of the variability among sound textures typically found in the environment, and to determine the ranges of parameter values that environmental sound textures typically occupy.

2.2.1 Sound collection

I collected 450 high quality raw sound samples from freesound, a freely available web resource [[Font et al., 2013](#)]. After a preliminary inspection, I selected 200 sound samples which were deemed to be "texture like". Sound clips with long duration of silence were excluded. All the sounds are of 48 kHz sample rate, and each clip is of 15 s duration. All the sounds in the sound corpus are normalized to RMS power.

The criterion for inclusion of a sound in this study is that the sound should be approximately stationary (as judged by me) and should not be predictable like music sounds which are mostly rule-based. At least they should not be rule based like music. The duration of 15s is enough because the property of "temporal homogeneity" for sound textures can be justified over a time window of 5-7s [[McDermott and Simoncelli, 2011](#)]. Even the sounds from the same source will not satisfy "temporal homogeneity" over a large window like 15s. Only a few of the sounds in the sound corpus have unwanted duration (>30s) of silence. These silence intervals are eliminated because they are not useful for statistics calculation. For the final study, all the sounds in

my sound corpus are of minimum 15s length. The sound synthesis toolbox has also mechanisms to set many parameters on the fly that allows to decide the length of the input sound texture from which the different statistical parameters are computed.

2.2.2 Statistical parameter extraction

Hence in this study, I explore the distribution of marginal moments, cochlear correlation, and modulation power over my corpus of natural sound textures. There are thus three aspects to this study of the statistical parameters of natural textures:

- a. Exploring the marginal moments.
- b. Exploring the correlation statistics.
- c. Exploring the modulation power statistics.

The workflow of our exploration process for each statistical feature type is shown in Figure 2.1.

The *Sound Synthesis Toolbox V1.7* by [McDermott and Simoncelli, 2011] segregates each input sound into a number of “cochlear” frequency sub bands and computes the marginal statistics for each sub band. Here I have taken 32 cochlear filters with center frequencies equally spaced on an *Equivalent Rectangular Bandwidth* (ERB) scale [Glasberg and Moore, 1990], spanning 80-20000 Hz. which is similar to the model by [McDermott and Simoncelli, 2011]. The output of each of the 32 cochlear filters undergoes envelope extraction and compression, and four “marginal moments”, mean, variance, skew and kurtosis, of the envelope values are computed, yielding

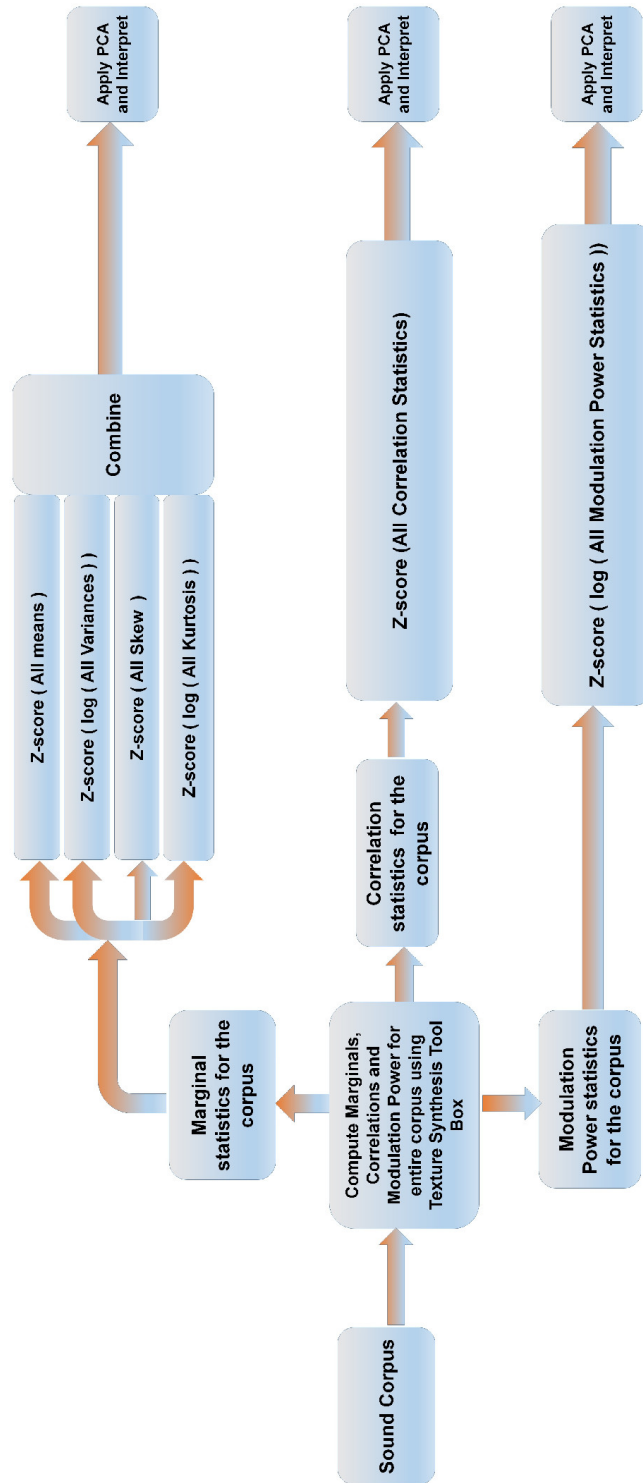


Fig. 2.1 Workflow for exploring the statistical parameter space of a corpus of natural sounds. Using the sound synthesis tool box, marginal, correlation and modulation power statistics were computed for the entire corpus. Envelope variance and kurtosis, as well as modulation power parameters, are log transformed to make their distributions more symmetric around the mean. Each of these parameter sets is normalized and centered by z-scoring, and the z-scored parameter sets are subjected to PCA and interpreted.

$32 \times 4 = 128$ marginal parameters for each sound.

Why statistical moments beyond kurtosis are not considered in this study?

[Julesz, 1962] in their visual study had reported that statistical moments up to 2nd/ 3rd order are useful. Adding higher order moments are also not useful in generating more realistic sounds as justified by [McDermott and Simoncelli, 2011]. It's important to realize that the objective of [McDermott and Simoncelli, 2011] was not to synthesize more and more realistic sounds. In order to do that many other algorithms and sound features are used in the sound community. Their objective was to identify those statistics which are biologically explainable or can be expected that the brain might be computing. Neurons cannot be expected to carry out complex mathematics other than some linear operations.

The *Sound Synthesis Toolbox V1.7* also computes the pair-wise correlation between the cochlear envelopes, yielding $32 \times 32 = 1024$ correlation parameters (although these are somewhat redundant given that the correlation matrix is symmetric around the main diagonal). To compute the modulation power parameters, the output of each cochlear envelope is passed through another set of 20 "modulation" bandpass filters. The center frequencies of these modulation filters are equally spaced on a log scale from 0.5 to 200Hz (same parameters to those used by [McDermott and Simoncelli, 2011]). Modulation power is then measured as the variance (mean sum of squares) of the output of each modulation filter, normalized by the variance of the respective

cochlear envelope. For each sound, a total of 32 (cochlear channels) \times 20 (modulation channels) = 640 modulation power parameters are computed.

Thus, each sound in my corpus is described by a parameter set of 128 marginal values, 1024 correlation values and 640 modulation values, a very high-dimensional parameter space, but also one that is expected to be highly redundant, given that, for example, the marginal moments in adjacent frequency bands are bound to be highly correlated. To examine this redundancy, and to arrive at a low-dimensional parametrization of my sound corpus which would make it feasible to examine the ranges and distributions of statistical features that are common among environmental sounds, I subjected each of the parameter sets (marginals, correlations, modulations) to PCA. Prior to PCA, the raw parameter values underwent the following two pre-processing steps: Firstly, the distributions of the envelope variance and kurtosis parameters, as well as the modulation power parameters, were strongly positively skewed, as might be expected. They were therefore log-transformed to yield more symmetric and compact parameter distributions. Secondly, the distributions of means, log(variances), skew and log(kurtosis), as well as those of the correlations and the log(modulation power) values were normalized and centred by z-scoring, respectively. After these preprocessing steps, the matrices of 200(*soundexamples*) \times 128 marginal parameters, 200 \times 1024 correlations, and 200 \times 640 modulation power values were independently analysed with PCA.

Why these statistics are separately subjected to PCA ? All these statistics provide different perceptual information about the sounds and they vary across the natural sounds as shown by [McDermott and Simoncelli, 2011]. As the nature of marginals

and correlations are very different and there is no reason apriori to expect them to be linked, so in a first step, analysing them separately is sensible.

2.3 Results

The distributions of the original and transformed (pre-processed) statistical parameters computed for the corpus are shown in Figure 2.2A-H. Figure 2.2A shows the distribution of mean of envelope amplitudes across frequency bands for the entire corpus. The distribution of variances of envelope amplitudes is shown in Figure 2.2B, and the distributions of skew and kurtosis are depicted in Figures 2.2E and 2.2G respectively. The distributions of the raw variance and kurtosis parameters in particular are quite asymmetric, with a noticeable positive skew. This asymmetry is reduced after log transformation and z-scoring, as can be seen in Figures 2.2D and 2.2H.

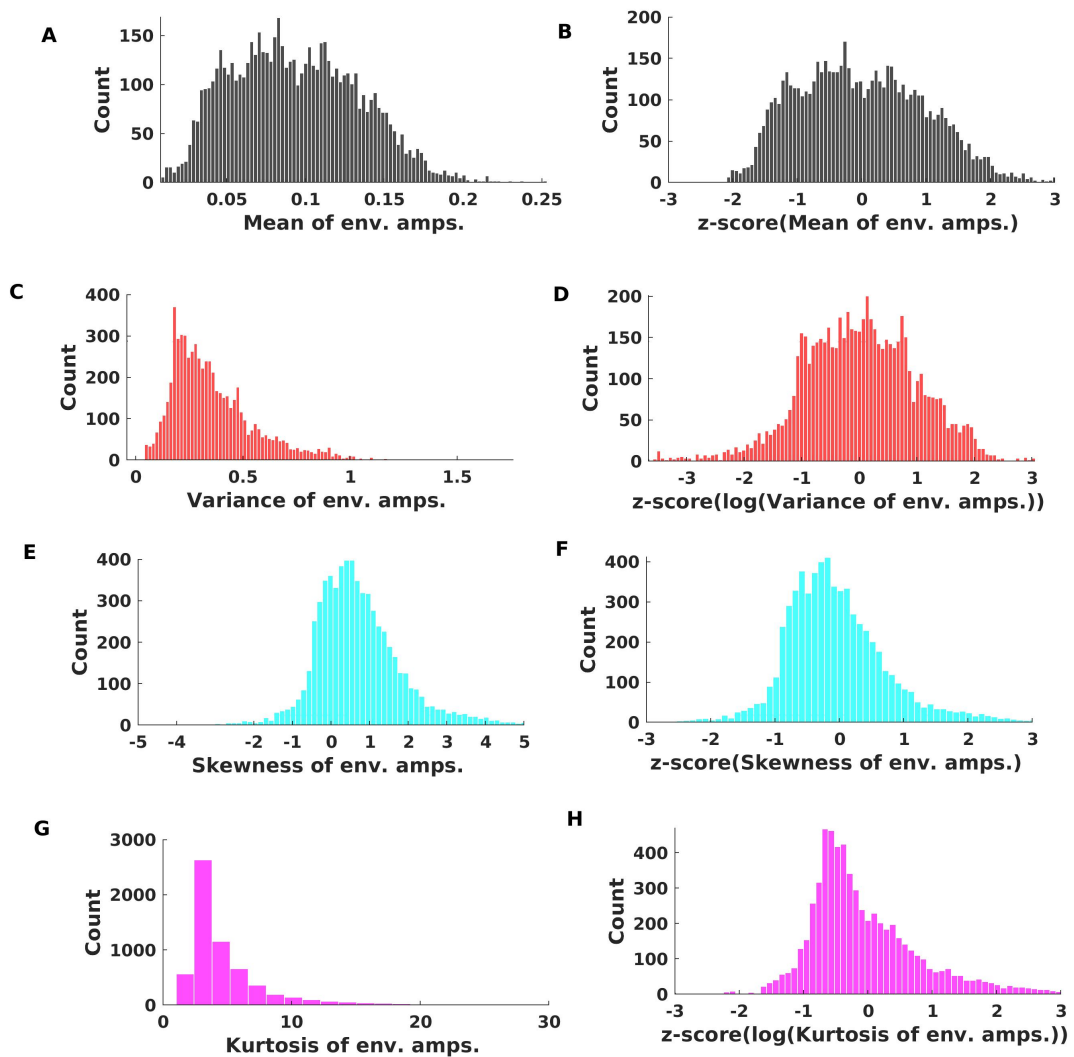


Fig. 2.2 (A, B) Distribution of mean of envelope amplitudes. The median, 5th and 95th centile of the distribution were 0.09, 0.034 and 0.16 respectively. (B) Mean of envelope amplitudes after z-scoring. (C) Distribution of variances of envelope amplitudes. The median, 5th and 95th centile were 0.31, 0.13 and 0.77 respectively. (D) Variances of envelope amplitudes after log transformation and z-scoring (E) Distribution of skew of envelope amplitudes. The median, 5th and 95th centile were 0.56, -0.77 and 3.0 respectively. (F) Skew of envelope amplitudes, after z-scoring. (G) Distribution of kurtosis of envelope amplitudes. The median, 5th and 95th centile were 3.79, 2.13 and 17.4 respectively. (H) kurtosis of envelope amplitudes after log transformation and z-scoring.

2.3.1 Principal Components of the Marginal Statistics of Sound

Textures

The results of the PCA on the marginal statistics are shown in figure 2.3. A key indicator of whether PCA is a useful and appropriate tool to identify major underlying trends and patterns in the data is whether the first few principal components capture (“explain”) a large proportion of the variability between samples. The proportion of variance explained by the first few principal components (PCs) is shown in Figure 2.3A. Perhaps surprisingly, the first two principal components are sufficient to capture about 65% of the variance of the 128 marginal parameters for the corpus of 200 acoustically very diverse samples of environmental sounds. Figure 2.3B shows the distribution of our sound corpus over the “*marginal space*” spanned by the first two principal components, and Figures 2.3C and 2.3D show the “shapes” of the first and second PCs for the marginals. When inspecting the heatmap plots of these PCs, it is worth remembering that, because the parameters were z-scored prior to PCA, the units of the color scale are standard deviations above or below the mean parameter values for the entire corpus. The first PC (Figure 2.3C) is characterized by low means but large variances and skews, with perhaps slightly above average kurtosis, and these trends apply more or less uniformly across all frequency channels. Consequently, PC1 will discriminate sounds that are sparse, with silent periods punctuated by occasional bursts of sound which drive the large variance and skew in the envelopes. The low mean envelope values compensate for the large variance and positive skew: as all sound

samples were normalized for RMS power, sounds that are characterized by bursts with positive skew in their envelope amplitudes must have a relatively lower mean envelope "baseline").

Why did I consider the first two PCs only? I have considered only the first two PCs because the variances explained by the 3rd and higher order PCs are substantially less. Apparently we can visualize things more clearly in two-dimensional space than higher dimensional space. My other objective was to get an approximate range of these statistics through which the natural sounds vary where the ranges need not be some solid boundary. It's impractical to find a solid range of values because the natural sound space is infinite.

Adding more statistics will only add complexity to the model, which will no longer be explainable biologically. It will also be meaningless to explore sounds which are statistically similar. In fact it's hard to say that any two sounds will ever be same statistically. By taking similar sounds in the study we cannot explore the acoustic variability of natural sound space.

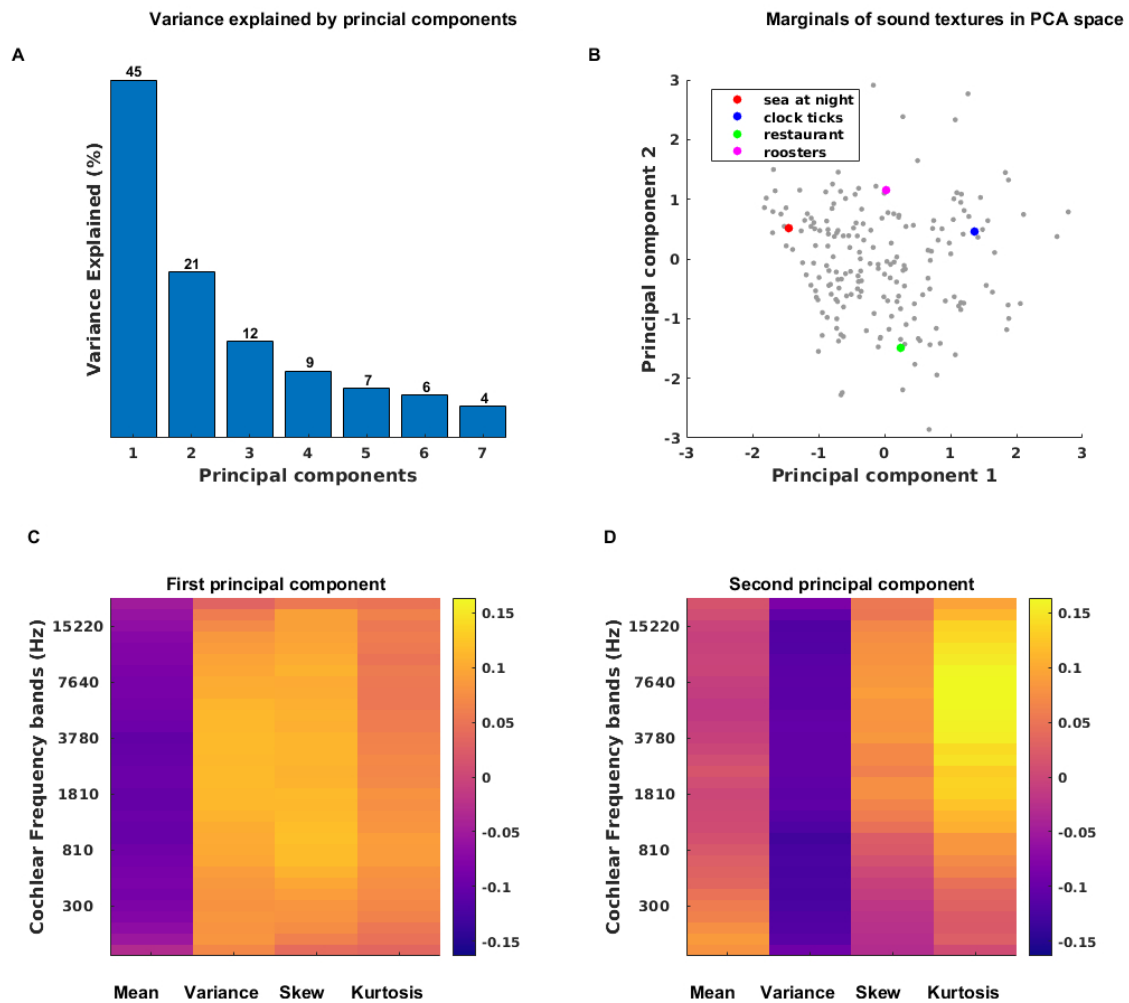


Fig. 2.3 Principal Components of the Marginal Parameters. (A) Percent variance explained by the first 7 PCs of the marginal parameters. The first two PCs capture 65% of the variance explained. (B) Distribution of the sounds in our corpus along the first two PCs of the marginals. Four example sound textures examined further in Fig 4 are highlighted in color. (C, D) Shape of the first and second PCs of the marginals, respectively. The 1st PC distinguishes textures of relatively low mean and high variance, skew and kurtosis from textures for which the reverse is true. The 2nd PC has mean and skew values that are near zero, and thus mostly distinguishes textures with low variance and but high kurtosis, particularly for frequency bands above 800 Hz, from sounds with the opposite feature combination.

To verify and illustrate that the first PC of marginals distinguishes sound textures along a “sparseness” dimension, I examine the marginal statistics of two example sounds from my corpus, “sea at night” and “clock ticks”, in figure 2.4. These sounds

in the PCA space are highlighted by red and green dots respectively in Fig 2.3B, and they were chosen to be approximately at opposite ends of the distribution along PC1 but with nearly identical PC2 values. “Sea at night” has lower PC1 values in contrast to “clock ticks”. From Fig 2.3C, we expect that “clock ticks” should have higher on average lower envelope means, but higher envelope variance, skew and kurtosis, than “sea at night”. The panels in figure 2.4 confirm this. “Clock ticks”, a texture of the sound of multiple clockworks - and brief silences between the ticks - is also a much “sparser” sound than “sea at night”, which features rolling wave and wind sounds that sustain constantly elevated sound pressures. Others before us have remarked that marginal moments can capture the sparseness of natural sounds [Attias and Schreiner, 1998; Field, 1987; McDermott and Simoncelli, 2011], but note from Fig 2.3A that the first PC of the marginal distributions which captures sparseness in the manner just described accounts for almost half of the variability observed in our corpus of natural sound textures, suggesting that “sparseness” is indeed a major discriminating feature of environmental sounds.

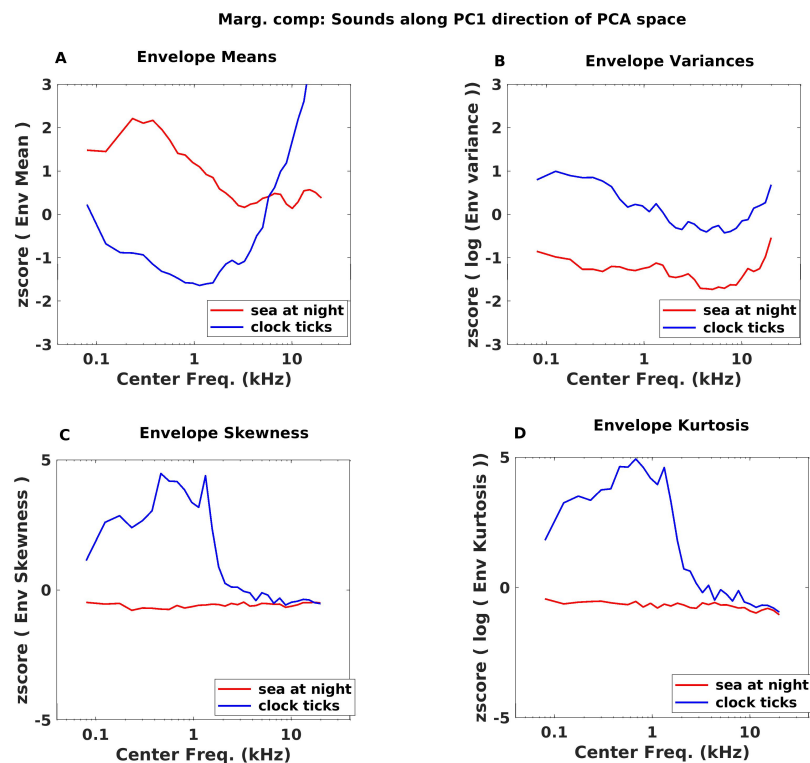


Fig. 2.4 Comparison between the envelope statistics of “sea at night” from one end of PC1 dimension and “clock ticks” from the other end. (A) “Sea at night” has higher envelope mean than “clock ticks” as it is in the lower end of PC1 dimension (B, C, D) Envelope of “clock ticks” with high variance, skewness and kurtosis than “sea at night” for frequencies above 800Hz.

The first PC in figure 2.3C can reasonably be interpreted as capturing the sparseness of sounds, but does the second PC shown in figure 2.3D also lend itself to an intuitive interpretation? The 2nd PC is characterized by envelope mean z-scores near zero, relatively small (negative) values for variance, but large values for skew and particularly for kurtosis, the latter with some high-frequency bias. To interpret this result, consider that variance, skew and kurtosis all measure excursions from the mean, but skew and kurtosis, as higher order moments, are “more sensitive” to such excursions, growing with the third and fourth powers of the deviation from the mean respectively, rather than just the square. Thus, an envelope distribution with a large kurtosis but a small variance

will have a particularly long, thin “tail”, meaning that sound amplitudes can shoot up to very large values relatively frequently, but will not spend much time at “middling” amplitude levels, while for a texture with relatively larger variance and smaller kurtosis, the converse is true. We would therefore expect sounds with large marginal PC2 scores to be not just sparse, but “bursty”, exhibiting intermittent bouts of very high sound energy and fluctuating quite wildly between loud and quiet, but relatively little in between, unlike textures with low PC2 scores which would exhibit comparatively “less extreme” amplitude fluctuations. In PC2, large kurtosis goes hand in hand with positive skew. This is likely attributable to the fact that sound envelope amplitudes cannot be negative, and the large amplitude excursions of “bursty” sounds with high kurtosis are therefore bound to be positively skewed. Thus, PC2 appears to rank sound textures on how “bursty” they are. In figure 2.5 we illustrate the marginal statistics for two sounds chosen to vary systematically along PC2, but have approximately the same values for PC1: “restaurants”, and “roosters”. Figure 2.3C highlights the coordinates of these two sounds in marginal PC space with a green and a magenta spot respectively, and shows that “roosters” has a much larger PC2 value than “restaurants”. As can be seen in figure 2.5, the two sounds exhibit the expected trends, with “roosters” having on average smaller variance but greater kurtosis than “restaurants”. Both sounds are of average “sparseness”, but while in “restaurants” there is a variety of background sound events of differing levels (voices, cutlery sounds, footsteps, etc), “roosters” jumps wildly between periods of relative quiet and moments of loud and forceful crowing, making it the substantially more “bursty” sound of the two.

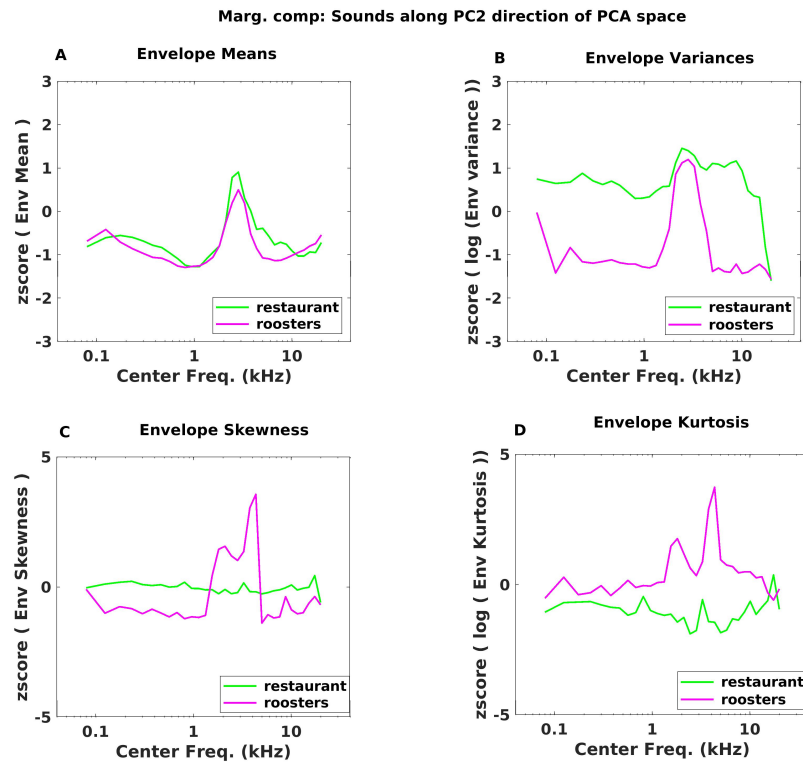


Fig. 2.5 Comparison of “restaurant ambience” and “roosters” across PC2 dimension of Marginal statistics. (A) “roosters” has a lower mean for cochlear envelope than “restaurant”(B) “roosters” has a higher variance cochlear envelope than “restaurant”. PC2 in Fig 2.3D indicates that as we move along the PC2, sounds should have opposing trends in mean and variance values of their cochlear envelopes. (C, D) As we move in the PC2 direction “skew” and “kurtosis” should be higher. “roosters” has higher skew and kurtosis than “restaurant”.

In summary, the first two PCs of the marginals of our corpus of sound textures can be interpreted as capturing features of “sparseness” and “burstiness”, which between them account for two thirds of the variance (45% and 21% respectively, see Figure 2.3A) in envelope marginals across the corpus.

2.3.2 Principal Components of the Cochlear Correlations of Sound

Textures

The distributions of the cochlear correlations between the envelope amplitudes for different cochlear frequency bands computed for the corpus and pooled over all frequency bands, are shown in Figure 2.6A. The distribution shows a number of interesting features. Firstly, anticorrelations (that is, negative correlation coefficients) are extremely rare. That is perhaps unsurprising given that positive correlations between frequency bands arise easily whenever a broadband source modulates activity simultaneously in multiple adjacent frequency channels, but physical mechanisms that would lead to anticorrelated sound envelopes in different frequency bands are hard to envisage. Secondly relatively large correlations ($R > 0.5$) are somewhat more common than smaller ones ($R < 0.5$), although the full positive range of correlation coefficients is very well represented. The median R value was 0.5536, and the 5th and 95th centile values were 0.0052 and 0.9163 respectively. Figure 2.6B shows the distribution of correlations after pre-processing for the PCA via z-scoring to achieve a more symmetric distribution.

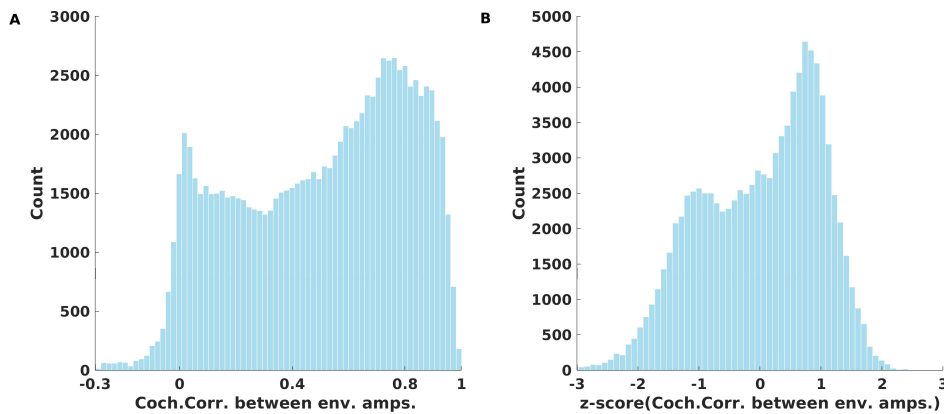


Fig. 2.6 (A) Distribution of cochlear correlations between the envelope amplitudes of different cochlear frequency bands. The median, 5th and 95th centile of the distribution were 0.55, 0.0052 and 0.91 respectively. (B) Distribution of cochlear correlations after z-scoring.

The results of PCA on the correlation statistics are depicted in figure 2.7. In figure 2.7A we can see that the first PC accounts for a remarkably high proportion of the variance, with 79%. The second PC, in comparison, captures a comparatively modest 11% and the percent variance explained by the remaining components is in the single digits. Despite the great diversity of the corpus and the large number of correlation parameters, 90% of the variability in correlation parameters can be accounted for by the first two principal components only. Figure 2.7C and 2.7D show the “shapes” of the first and second PCs for the correlation statistics. The units of the color scale in the heatmaps (figure 2.7C and 2.7D) are once again standard deviations of the correlation values for the entire corpus. Only the upper triangular matrix of the PCs is shown, as these are symmetric correlation matrices.

The first PC (figure 2.7C) is essentially completely “flat”, and it will therefore distinguish textures for which envelope amplitudes are highly correlated between frequency

bands from those that are poorly correlated, irrespective of frequency. High correlations among frequencies in sound textures typically arise if many broad-band clicks or noise-bursts contribute to the texture, as these will create synchronized, abrupt changes in sound level across many frequency bands. Thus, an applauding crowd, or pouring gravel onto a hard surface would generate highly correlated sound textures. Examples of less correlated textures are generated from sources that are more narrow band and which become active independently. The sound of running water is a typical example. In running water, much of the sound is created from the excitation of small air bubbles. Each bubble is a resonator with a more or less narrow band resonance that depends on the bubble's size, and different sized bubbles may burst or become otherwise excited at different times, creating sound patterns that are poorly correlated across frequency.

Indeed, the first PC is very good at discriminating “applause”-like sounds from “water”-like sounds, as can be appreciated from figure 2.7B, 2.7E and 2.7F. Figure 2.7E and 2.7F show the cochleagrams for a sample of dripping water sounds and the sound of an applauding crowd respectively. These cochleagrams have been normalized or sound level in each band by z-scoring the envelope amplitudes in each band independently. Correlations across cochlear frequency bands are visible as vertical bands in these cochleagrams, and the normalization ensures that such bands are not obscured by overall sound level differences in different sound frequency bands. The “dripping water” sound shown in figure 2.7E is relatively weakly correlated, as can be seen from its low PC1 coordinate in figure 2.7B (red dot), and while there are clear horizontal stripes in the high frequency part of its cochleagram, there are also many prominent narrow band

features, particularly in the lower frequencies. In contrast, the “applauding crowds” sound in figure 2.7F has a high PC1 correlation coordinate (figure 2.7B, blue dot), and a lot of prominent vertical striping throughout its cochleagram.

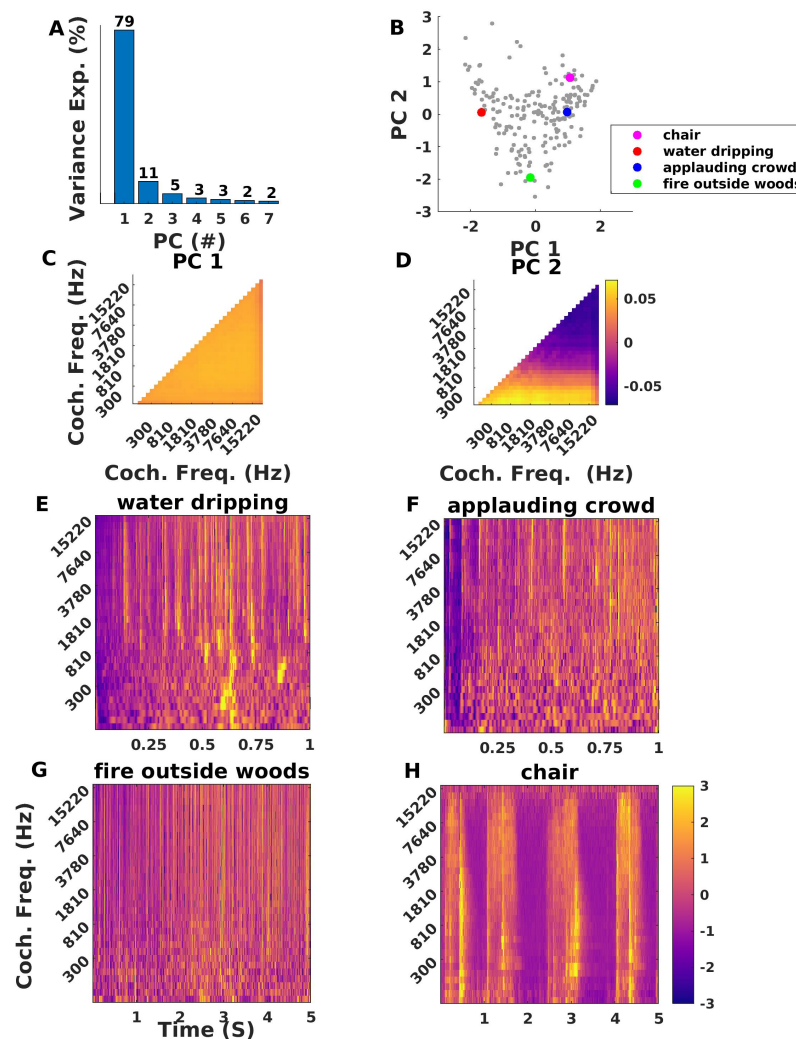


Fig. 2.7 Principal Components of Cochlear Correlation Parameters. (A) The first PC alone captures 79% of variance, whereas the second PC captures merely 11% and higher PCs capture only very small proportions of the variance. (B) Coordinates of all sound textures in our corpus. Coloured dots represent four example sounds examined further in panels E-H. (C) PC1 shows elevated correlation across all pairs of cochlear frequencies, and thus distinguishes “highly correlated” from “poorly correlated” sounds. (D) PC2 captures whether correlations are more pronounced among low or high frequencies. (E-H) normalized cochleagrams for the four sound texture examples highlighted in (B) by colored dots

The second PC of the correlation parameters captures whether correlations are more prominent in low or high frequency bands (see figure 2.7D). Normalized cochleagrams of sound textures with very different PC2 coordinates are shown in figure 2.7G and 2.7H. The sound texture “fire outside woods” (figure 2.7G, green dot in figure 2.7B) has a strongly negative PC2 coordinate, and indeed, the vertical stripes that are characteristic of envelope correlations are prominent only in the higher frequency bands. These features likely originate from broad band but somewhat high-pitched crackling sounds which come about when small twigs in a fire burst, and which contribute to the characteristic “fire” sound. In contrast, the “dragging chair” sound texture (Fig 7H, cyan dot in figure 2.7B) has a PC2 coordinate near zero, and correlations are more or less evenly distributed throughout the frequency bands.

In summary, while the number of possible pairwise correlations between cochlear frequency bands is very large (1024 parameters per sound texture in our study), almost 80% of the variance in these correlation statistics is captured by a PC that simply measures the extent to which amplitude envelopes are correlated regardless of frequency band. A relatively modest additional 11% of variance is explained by a PC that distinguishes sounds with correlations in high frequencies from sounds with correlations in lower frequencies. Like marginals, correlation statistics are therefore very highly redundant.

2.3.3 Principal Components of the Modulation Power Statistics of Sound Textures

The distributions of the original and transformed (pre-processed) modulation power parameters pooled over the entire corpus are shown in Figure 2.8. Fig 2.8A shows that the original modulation power distribution for the sound corpus is highly asymmetric and positively skewed. Its median was 0.0278, and its 5th and 95th centile were 0.0037 and 0.1141 respectively. Log transformation and z-scoring for PCA preprocessing made the distribution more much symmetric (Fig 2.8B).

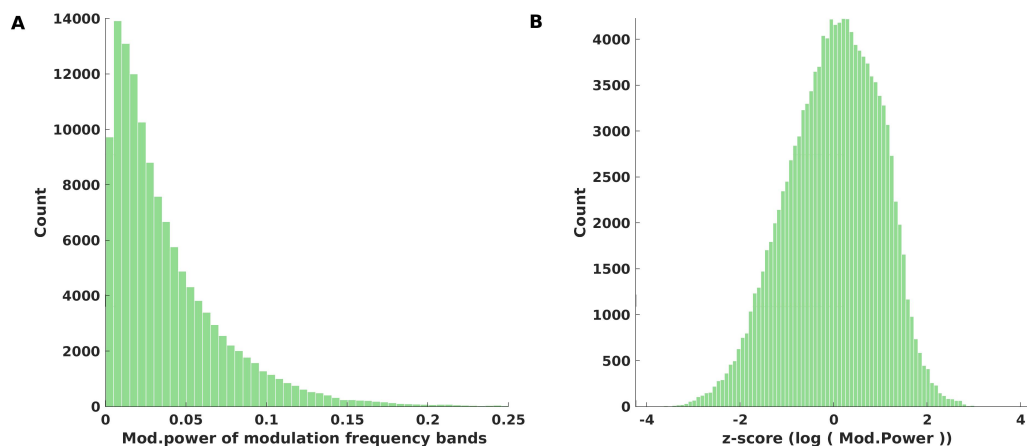


Fig. 2.8 (A) Distribution of modulation power parameters for the entire corpus which are computed over the cochlear envelope amplitudes after modulation filtering. The median, 5th and 95th centile of the distribution were 0.028, 0.0037 and 0.11 respectively. (B) Distribution modulation power parameters after log transformation and z-scoring.

The results of PCA on the modulation power parameters of our sound corpus is shown figure 2.9. Analysis of the 640 dimensional modulation power statistics indicates that 73% of variability of our sound corpus are explainable by the first two principal components as shown in figure 2.9A. The distribution of our sound corpus along the

first two PC dimensions is shown figure 2.9B, while figure 2.9C and 2.9D depict the “shapes” of first and second PCs respectively. PC1, which captures 48% of the variance in modulation parameters, discriminates sounds which are modulated at fast modulation frequencies (greater than 60 Hz) from those that are slowly modulated, and this again holds in a very similar manner across all cochlear frequency bands (Fig 7C). Meanwhile, PC2 (shown in figure 2.9D) accounts for 25% of the variance and is sensitive to the extent to which sound textures exhibit amplitude modulations at “middling” modulation frequencies of around 30-100 Hz. We again illustrate these dimensions with examples chosen from the corpus which span either the first or the second PC axes, and which are highlighted in figure 2.9B with colored dots. Thus “gunshots” (red dot in figure 2.9B, z-scored modulation spectra shown in figure 2.9E) lies at the low end of PC1, and the texture is dominated by modulation frequencies of typically less than 10 Hz, while “bees” (blue dot in figure 2.9B, z-scored modulation spectra shown in figure 2.9F) is dominated by high modulation frequencies, typically above 100 Hz. The causes of the different amplitude modulation rates for these two examples are intuitive: bees beat their wings at much faster rates than users of firearms typically pull triggers. Meanwhile the texture sample “applauding crowd” (figure 2.9G, green dot in figure 2.9B) has a PC2 coordinate of approximately -1.8, and its modulation spectrum exhibits the expected dearth of modulations near 60 Hz, while the texture “vacuum cleaner” (figure 2.9H, cyan dot in figure 2.9B) has a PC2 coordinate of +1.5 and prominent 60 Hz modulations.

In summary, just like marginal and correlation parameters, modulation parameters too exhibit a high degree of redundancy, such that almost three quarters (73%) of the variance across the 640 parameters could be captured with just two PC coordinates. And again, the PCs obtained lend themselves to simple interpretations, in this case fast vs slow (for PC1), and with or without much modulation in a mid, 60 Hz range (for PC2).

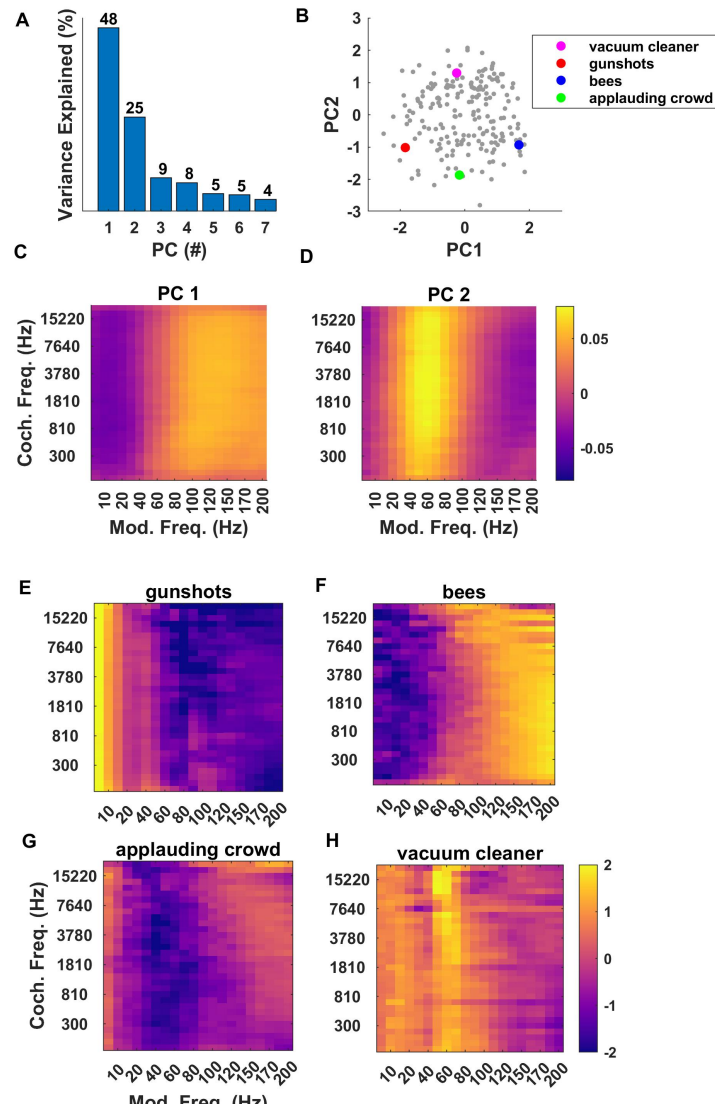


Fig. 2.9 Principal Components of the Modulation Parameters. (A) Proportion variance explained by the first seven PCs of modulation power parameters. (B) Distribution of sound textures from our corpus along the first two PC coordinates for modulation power. The first two PCs capture 73% of the variance between them. (C, D) Shape of first and second PC respectively. PC1 discriminates sounds “slowly” from “rapidly” modulating sounds, with a boundary near 60 Hz for all cochlear frequencies. PC2 discriminates sounds with prominent modulations in a “mid range” (near 60 Hz) from sounds lacking such modulations. (E) Modulation spectrum of sound texture sample “gunshots”, showing prominent modulation at low rates. (F) Modulation spectrum for “bees”. High modulation frequencies (> 80 Hz) dominate. (G, H) The modulation spectrum for “applauding crowd” shows a relative dearth of modulations near 60 Hz, while that for “vacuum cleaner” shows prominent 60 Hz modulations.

2.4 Discussion

The idea that statistical regularities may govern the types of sensory stimuli we encounter in our environment has a long history, as does the idea that the sensory systems may be adapted to some of these statistical features or regularities [Attneave, 1954; Barlow et al., 1961]. This idea has arguably been much more influential in vision research than in hearing research. For example, an attempt by [Olshausen and Field, 1997] to explain the centre-surround structure of primary cortex visual receptive fields as nature's solution to the problem of having to encode the structure of visual scenes in a sparse, and hence energy-efficient, manner, has become enormously influential. (Note, however, that more recent work by Singer et al. [2018] proposes an intriguing alternative explanation, namely that cortical receptive field structure not just of visual but also auditory cortical neurons may be optimized to facilitate prediction of future inputs, rather than energy efficiency.)

An early example of work looking for statistical regularities specifically in the auditory modality comes from [Voss and Clarke, 1975], who already reported over 40 years ago that pitch and amplitude fluctuations over long segments of music and speech streams recorded from the radio exhibited a so-called $1/f$ distribution. Garcia-Lazaro and colleagues [Garcia-Lazaro et al., 2006] later built on that observation and showed that auditory cortex neurons appear to be tuned to these statistics, in that they respond more strongly and reproducibly to artificial sound streams that follow $1/f$ distributions than to sounds which fluctuate according to slower ($1/f^{0.5}$), or faster

($1/f^2$) distributions. This was later shown to be an emergent property of the ascending auditory pathway, as inferior colliculus neurons generally prefer more rapidly fluctuating sounds, and neurons in the medial geniculate exhibited no particular preference for fluctuations that were either faster or slower than $1/f$ [Garcia-Lazaro et al., 2011]. These studies are conceptually similar to earlier work by Also highly relevant are studies by [Attias and Schreiner, 1997, 1998] which had described power-law statistics in amplitude distributions of natural sounds, and reported evidence that midbrain neurons can encode synthetic sounds with higher accuracy (as quantified by mutual information), when these stimuli match the statistical parameters typical of natural sounds. Other noteworthy examples of studies concerned with the distributions of environmental sounds and their relevance to auditory processing include a well-known study by [Lewicki, 2002], which presented an efficient coding argument alongside an analysis of natural sounds to explain the cochlear frequency tuning characteristics, or a study by [Singh and Theunissen, 2003] which described the low-pass nature of spectral and temporal modulations in natural sounds, in a manner corroborating and extending the findings by [Voss and Clarke, 1975].

Despite this relatively long history, the literature on natural sound statistics and their relevance to auditory processing and perception has remained relatively thin, perhaps because it is still unclear which of the many statistical parameters that could in theory be devised or applied to the study of natural sounds is most likely to provide highly useful and practical descriptors of natural sounds. In this context, the study by [McDermott and Simoncelli, 2011] provided a fresh perspective. By being able to

generate recognizable, and often highly realistic, morphs of natural sound textures by imposing the statistical parameters they had identified onto white noise, [McDermott and Simoncelli, 2011] demonstrated that their chosen statistical parameter set comes close to being a set of “sufficient statistics”. The fact that these sets of parameters fully describe many types of natural sound textures also raises the intriguing hypothesis that neurons in the central auditory system may be tuned to statistical parameters similar to those identified in their study. Such tuning could easily explain our perceptual ability to distinguish different sound textures with ease, even though these textures are stochastic signals, and two recordings of the same type of sound texture are essentially guaranteed to be very different sound waves. Identifying a set of statistical parameters that come close to fully characterizing sound textures is therefore a very significant conceptual advance. However, there are issues which make it difficult in practice to build on their work with follow-on psychoacoustic and neurophysiological studies. One such issue is the fact that the number of statistical parameter values used by [McDermott and Simoncelli, 2011] to characterize and synthesize textures is very large, in the order of 1500 parameters in total for each texture. This parameter explosion arises largely because marginals, correlations and envelope modulation spectra are computed independently for every frequency band. In addition, the range of parameter values that one is likely to encounter in natural or environmental soundscapes has not been described.

Ideally one would wish to build on [McDermott and Simoncelli, 2011] approach to devise a characterization of the statistical features of natural sound textures which uses

a far smaller number of numerical parameters and identify their distributions across the ecological acoustic environment. We do not claim that our analysis presented here has achieved this, but it has nevertheless shown that this may be possible in principle, given the enormous redundancy of the statistical parameters we have identified through our PCA of a corpus of environmental texture recordings which we have analysed. Indeed, just two PC coefficients for marginals captured 66% of the variance of a 128 dimensional parameter space.

Similarly, the first two PC coefficients of cochlear correlations cover 90% of 1024 dimensional parameter space, and the first two PC coefficients of modulation power explain 73% of variance of a 649 dimensional modulation parameter space. Lower-dimensional descriptions of natural sound statistics which nevertheless capture much of the richness of the auditory environment should therefore be possible.

Another noteworthy finding of our PCA analysis is that it illustrates the high degree to which many statistical features tend to co-vary greatly across frequency bands. Thus, the first PC across the marginals showed very little variation in Mean, Variance of Kurtosis as a function of cochlear frequency (Fig 3C), the first PC of cochlear correlations is effectively constant across all pairs of frequency bands (Fig 6C), and the first and second PCs of the modulation parameters too exhibit very little variation as a function of cochlear frequency. This observation is not entirely novel, as Attias and Schreiner (1997) had already conducted a filter bank analysis on an ensemble of natural sounds, and reported that temporal lower order statistics for a given sound sample tend to be highly similar across frequency bands. Nevertheless, in combination with the many

additional values which we report here, this confirmatory finding is potentially quite useful. Thus, if someone presented us with a “mystery texture sound”, reproduced at a unit RMS amplitude, and asked us to guess what its statistical parameters are likely to be in some particular frequency band, then we would be able to declare with some confidence, firstly, that the particular frequency band probably does not matter, secondly, that its mean envelope amplitude has a 90% chance of falling between ~ 0.0338 and 0.1618 with a maximum likelihood value of ~ 0.0905 (Fig 2A), the variance of the envelope amplitude has a 90% chance of falling between ~ 0.1292 and 0.7763 with a maximum likelihood of ~ 0.315 (Fig 2C), its skewness has a 90% chance of falling between ~ -1.8 and $+3$ with a maximum likelihood of ~ 0.6 (Fig 2E), and its kurtosis is 90% likely to fall between ~ 1 and 18 (Fig 2G) with a maximum likelihood of ~ 5 . Similarly, envelopes in any two cochlear frequency channels are a priori more likely than not to be substantially correlated, with an $R > 0.55$ (Figure 6).

The data presented here can therefore facilitate informed guesses about as yet unknown natural sounds that we may be presented with in the future, and we hope that a better characterization of the statistical features of natural sounds will enable us to start asking better questions about the extent to which expectations derived from these distributions may be “built into” the functional anatomy of our central auditory nervous system.

Chapter 3

Sensitivity of Auditory Midbrain

Neurons to Statistical Features of

Sound Textures

3.1 Introduction

As we have seen in the previous chapters, sound textures are the collective result of many similar acoustic events, and recent psychoacoustic work indicates that many natural sound textures are largely characterized by key statistical features [[McDermott and Simoncelli, 2011](#)]. Thus, textures with one particular set of statistical features will sound like a crackling fire, and textures with another set may sound like a rushing stream or a swarm of insects. I have also described an auditory model proposed by [[McDermott and Simoncelli, 2011](#)] to extract these key statistics, and mentioned that

they hypothesized that these statistics are measured by the successive stages of neural processing in the auditory pathway.

The significant statistical features of their model were described in greater detail in chapter 1 section 1.5-1.6.

In McDermott & Simoncelli's two-stage bandpass filter model, marginal moments and cochlear correlations are computed from the output of the first stage filters whereas modulation power and modulation correlations are calculated over the results of second stage filters. Filter banks those used in the two stages of the model are similar except for their bandwidths and are in accordance with the frequency response properties seen at the level of cochlea and midbrain neurons. All these parameters are "unique" in the natural sound texture space and that is in the sense that they vary across the domain of natural sound space. More precisely, these parameters can explain the acoustic variability of natural sound space in a limited framework of statistical variability. As we have seen in the last chapter, marginal moments mostly distinguish "sparse" and "bursty" sound which are primarily characterized by intermittent burst in energy of the sound envelopes, from more continuous textures. In contrast, cochlear correlation distinguishes "highly correlated" (e.g. applause) to "poorly correlated" ones (water-like sound textures). Modulation power can differentiate "rapidly modulating sounds" (e.g. fast flapping wings of bee) to "slowly modulating sounds" like ocean waves. Modulation correlations on the other hand can differentiate sound textures which have sudden "phase-changes" or onset-offset like mechanism (e.g. bomb explosion, fire crackers).

These different types of statistics can also in a sense be thought of as forming a "hierarchy", given that the auditory system might be able to measure the marginals from observing the activity of auditory nerve fibers individually, whereas cochlear correlations require information to be combined across processing channels along the tonotopic array. This notion of a "hierarchy" and the types of statistical features chosen by [McDermott and Simoncelli, 2011] were motivated at least in part by known physiological properties of neurons in the auditory pathway, including modulation tuning [Joris et al., 2004], and the sensitivity to temporal coherence [Elhilali et al., 2009; Krishnan et al., 2014].

Furthermore, [McDermott and Simoncelli, 2011] hypothesized that the sensitivity to each of these types of statistical features may already be present at the level of the auditory midbrain, but the extent to which neurons in the *inferior colliculus* (IC) are sensitive to each of these statistical features has not yet been examined experimentally.

The objective of this study is to explore how pervasive sensitivity to each of these statistical feature types is at the level of the IC. If IC neurons are sensitive to a particular statistical sound texture feature, then changes in neural responses should be observed whenever that particular feature of a sound texture changes abruptly, but all other characteristics are held constant. In contrast, if a neuron is deaf to that particular statistical feature, then its response should remain unchanged.

To determine how common sensitivity to each of the types of statistical features is among IC neurons, I therefore recorded extracellular responses of IC multiunits with silicon array electrodes implanted into the IC of *ketamine/xylazine* anesthetized female

wistar rats to sets of texture stimuli, which were synthesized to incorporate, at specific time points, abrupt changes in just one type of statistical feature while leaving all other stimulus parameters unchanged. The recordings were examined for either transient or sustained changes in neural activity evoked by changes in a each type of statistics. My results show that sensitivity to all types of texture statistics can already be observed at the level of the IC, although to a varying extent. For example, sensitivity to changes in variance is more common than sensitivity to changes in modulation correlations. Thus, my results indicate that subcortical processing of auditory textures may already be sufficient to encode all the types of statistical features identified by [McDermott and Simoncelli, 2011] as being important in identifying and discriminating natural sound textures.

The rest of the chapter has been organized as follows. Section 3.2 describes the methodology used to select "representative" sound texture parameters from a corpus of 200 natural sounds which was described in the previous chapter, and for preparing stimuli from these parameters. The methods used for animal preparation, stimulus delivery and electrophysiological recording are also described. Section 3.3 summarizes and analyzes the data. Section 3.4 concludes with a scholarly discussion.

3.2 Materials and method

3.2.1 Animal Preparation

Five young adults (eight weeks old) female Wistar rats weighing approximately 250 – 280gm were used for IC recordings. All rats were purchased from the Chinese University of Hong Kong. The experiment procedures in the study were approved by the Ethics Sub-Committee on the Use and Care of Animals at the City University of Hong Kong and under license by the Department of Health of Hong Kong [Ref. No. (18-167) in DH/HA and P/8/2/5 Pt.5].

3.2.2 Stimuli selection from Principal component space

3.2.2.1 Sound corpus

The sound corpus of 200 sound textures as described in chapter-2 and provided in Appendix- B was used here for representative sound texture selection and stimulus design.

3.2.2.2 Representative sound textures selection:

A set of sound textures were chosen to be "representative" of a wide range naturally occurring textures by selecting textures from my corpus so that these cover a substantial portion of the "PC space" of texture statistics described in the previous chapter. For the entire corpus, marginals , cochlear correlations and modulation power statistics had been measured separately. Modulation correlations statistics have not been considered

in the representative sound texture selection process partly because: though $C1$ and $C2$ statistics are significant in capturing sounds with differential modulation frequencies and sounds with sudden onsets-offsets respectively, results from [McDermott and Simoncelli, 2011] indicate that exclusion of $C1$ and $C2$ statistics from synthesis process does not produce significantly different synthesized sound texture.

We have infinite sounds in nature and sampling is always an arbitrary process. Considering the sound corpus to be a random sample of natural sound textures, I checked post hoc that the random sample collected should not be too biased to ensure that there were not large parts of the distribution which remain un-sampled. Moreover it's unrealistic to take all sounds present in the corpus for study. I need a suitable set of sounds that could fairly span over the entire sound corpus. Choosing another set of sounds from the corpus that can evenly capture the corpus to some reasonable extent will not alter the results.

I handpicked sound textures from the two-dimensional PCA spaces of marginals, cochlear correlations, and modulation power statistics. Visualizing the coordinates of the selected textures within a dimensionality reduced representation of the distribution of the parameters of my diverse corpus allowed me to verify that the textures selected for the current study are widely distributed through, and cover a substantial part of the range of, the parameter space covered by the corpus, and the 13 sample textures can therefore be considered as "representative" samples.

The selected thirteen textures (red dots) are shown in figure 3.1 and listed in table

3.1.

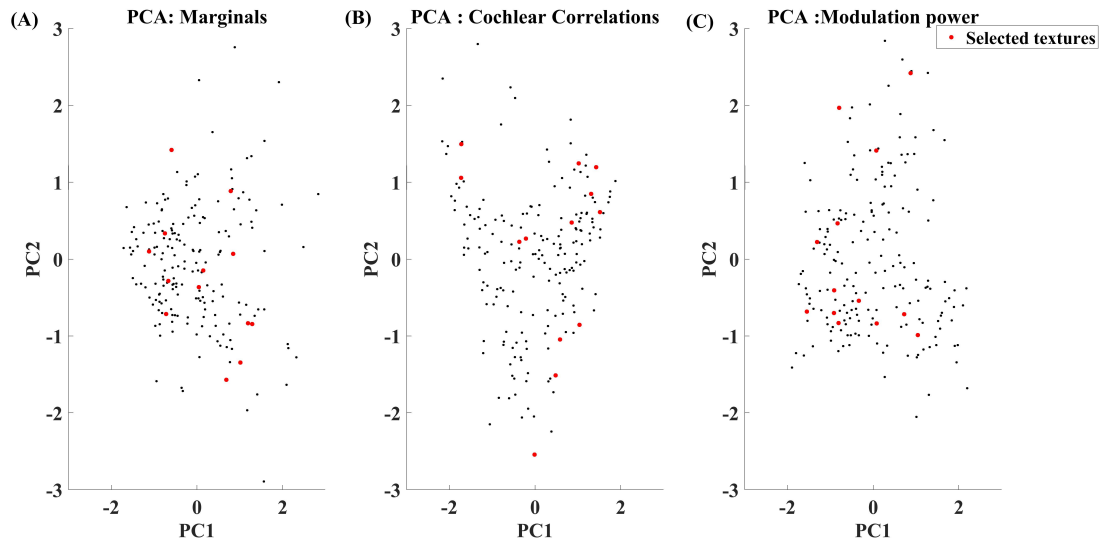


Fig. 3.1 Sound textures in the PCA space of (A) Marginal statistics. (B) Cochlear correlation statistics. (C) Modulation power statistics. x- and y-axis represent the first and second principal components respectively. Grey dots represent sounds in the corpus. Red colored dots represent the selected texture sounds.

3.2.2.3 Synthesized stimuli construction

The *Sound Synthesis ToolBox V1.7* [McDermott and Simoncelli, 2011] was used for stimulus synthesis. It is based on their model as shown in figure 3.2 and described in detail in chapter 1. The synthesis procedure has been summarized in figure 3.3. The toolbox requires a structure detailing the following information : (a) The original sound texture from which the target statistics that are to be imposed are measured, (b) sample rate of the sound texture, (c) random seed for white noise generation, (d) statistics that are to be imposed, (e) no of cochlear and modulation bandpass filters to be used, (f) filter parameters, (g) sound envelope compression factor, (h) boundary handling

Table 3.1 List of sounds selected from PCA spaces of marginals, correlations and modulation power statistics

Sound textures handpicked to represent the PCA space of marginal statistics

- 1 Lawn mower
- 2 Applause
- 3 Cackling geese
- 4 Stirring liquid in glass
- 5 xylophone

Sound textures handpicked to represent the PCA space of correlation statistics

- 1 Tin can
- 2 Barn swallow calls
- 3 Foot-steps walking in water
- 4 Horse galloping

Sound textures handpicked to represent the PCA space of Mod.power statistics

- 1 Church Bell
- 2 Frogs at night
- 3 Fireworks
- 4 Fire outside wood sticks

parameters required for merging the synthesized subbands. With the necessary information, the toolbox breaks sound snippets into *log*-spaced frequency bands, and, for each sub-band, it calculates statistical parameters from the original sound texture. To synthesize textures with desired parameters, the synthesis can be initiated with Gaussian noise, and pre-determined sub-band statistics can then be imposed on each sub-bands in an iterative process which uses conjugate gradient approach until the statistical parameters of the iteratively morphed noise snippet converge toward the desired parameters to a specified level of accuracy.

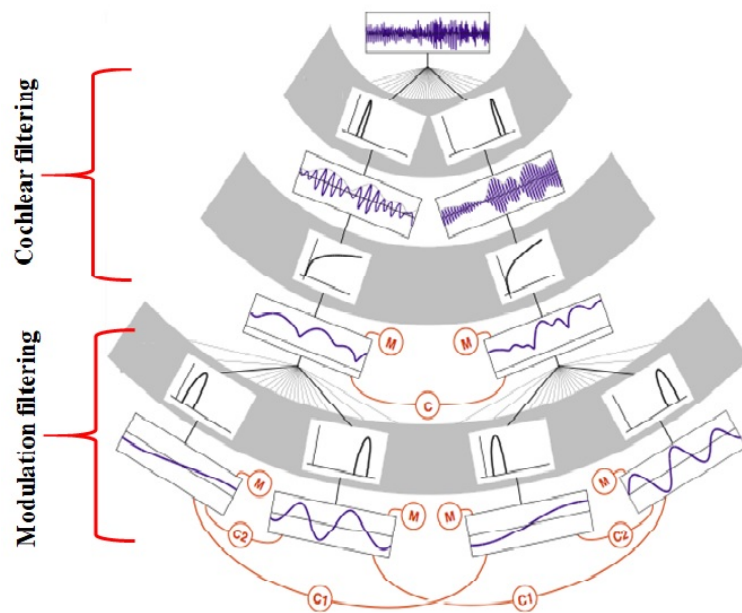


Fig. 3.2 The biological model used in the sound synthesis process [McDermott and Simoncelli, 2011].

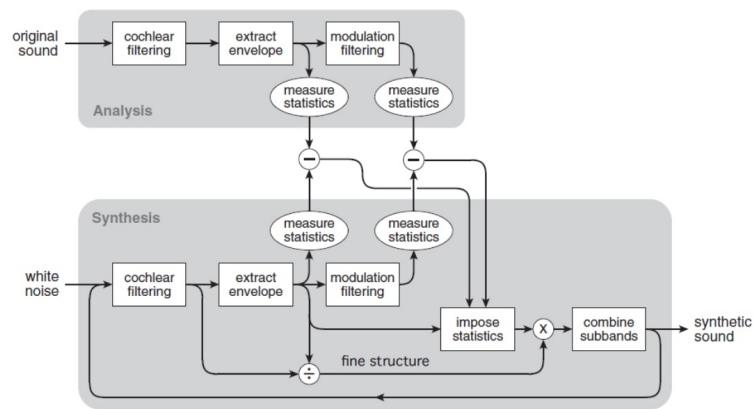


Fig. 3.3 Sound synthesis steps from [McDermott and Simoncelli, 2011].

From each of the 13 chosen textures, I computed their statistical parameters and then synthesized sound samples which morphed white noise into full-fledged textures in a step-wise process, each step representing a sudden transition where just one set

of parameters changes from that of white noise to that of the appropriate texture. The steps were as follows:

(A) For every sound (x):

- a. Measure the *power* (p) statistics of the original texture. Modify only the *power* (p) statistics of the original white noise (WN) sample to match to the measured (p). This will generate a sound ($sound_{x1}$) with matched (spectral) power statistics only. Henceforth, $sound_{x1}$ will be referred to as *+power* matched stimulus.
- b. Measure both *power* (p) and *variance* (v) statistics. Modify both the *power* (p) and *variance* (v) statistics of *WN* until it matches to the measured (p) and (v). This will generate a sound ($sound_{x2}$) with matched (p) and (v) statistics. ($Sound_{x2}$) is referred as a *+Var* stimulus.
- c. Measure (p), (v), *skew* (s) and *kurtosis* (k).
Modify (p, v, s, k) statistics of *WN* to synthesize a sound ($sound_{x3}$) with matched (p, v, s, k) and will be referred to as a *+SK* stimulus.
- d. Measure (p), (v), (s), (k) and *cochlear correlation* (C).
Modify (p, v, s, k, C) statistics of *WN* to synthesize a sound ($sound_{x4}$) with matched (p, v, s, k, C). This will be referred to as a *+Coch.Corr* stimulus.

e. Measure (p, v, s, k, C) and *modulation power* (M) of *sound_x*.

Modify (p, v, s, k, C) and *modulation power* (M) of *WN* to synthesize *sound_x5*. This will be referred to as a *+Mod.power* stimulus.

f. Measure (p, v, s, k, C, M) .

Also measure *cross-band* ($c1$) and *within-band* ($c2$) correlations of *sound_x*.

Modify $(p, v, s, k, C, M, c1, c2)$ of *white noise* to generate *sound_x6* with matched $(p, v, s, k, C, M, c1, c2)$. This will be referred to as a *+C1 + C2* stimulus.

(B) Combining the sound subsets generated in **step (A)**: Including 1 s silence at the beginning and end of sounds, 1.5 s each for synthetic sounds (power, +Var, +SK, +Coch.Corr, +Mod.power, +C1+C2) and 1.5 s of the original sound (Ori. sound) were concatenated to produce a continuous sound that morphs step-wise from noise to the final texture. Each segment was crossfaded with the next segment using 10 ms *cosine* ramps. For each of the 13 chosen textures, six samples were synthesized by using different random seeds for the Gaussian white noise in MATLAB. In total, I have $13 \times 6 = 78$, "morphed textures". All the synthesized signals have the same RMS power across statistical transitions.

The order of the statistical segments in the stimuli cannot be altered i.e. correlation based statistics (cochlear correlation, modulation correlations) should not come before moment based statistics (mean, variance, skew, kurtosis and modulation power) because to calculate the correlations we need the mean and variance statistics. If we play the

3.2 Materials and method

correlations based statistical segments prior to moments based segments (especially mean and variance segments) it will be difficult to quantify the response of neurons to any changes in mean and variance stimuli as they are already embedded in correlations based statistics.

I have chosen the duration of each segment to be 1.5s because I have to measure both “onset” and “ongoing” responses of neurons to statistical transitions. If I choose a very short duration for each segment, then it will be difficult measuring the “ongoing responses” over different time windows. On the other hand considering a large time segment will only consume more electrophysiology recording time.

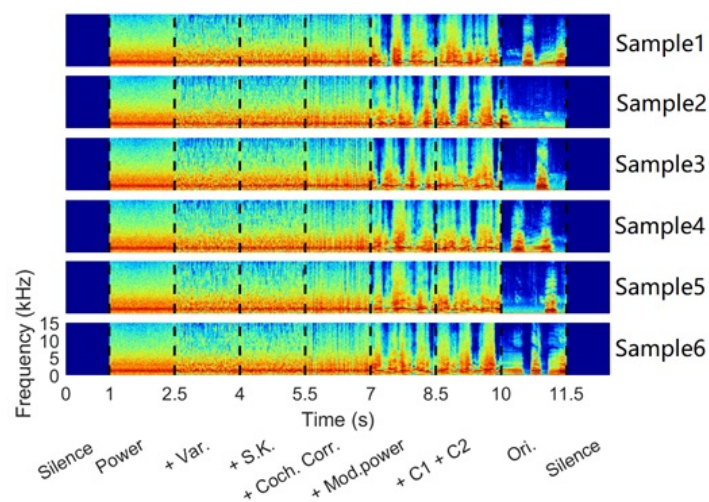


Fig. 3.4 Spectrograms for six different exemplars of synthesized stimuli using six random seeds for *Gaussian noise*. The synthesized stimuli shown in the figure are for sound texture “*Cackling Geese*”. The dashed lines represent the transitions when the statistical features of the sound textures change to incorporate the next “level” of features.

From figure 3.4 it is apparent that the synthesized sounds should not resemble completely, but their envelope statistical properties will remain consistent. For each

sound texture six exemplars are synthesized from different random seeds. This makes the experimental design more robust.

The total duration of each of these morphing probe stimuli is $7 * 1.5 + 2 = 12.5$ s, including 1 s silence at the beginning and end of each sound, and 1.5 s for each of the synthetic sound segments incorporating the next of the 6 levels of statistical features, as well as one randomly selected 1.5 s segment from the original sound texture. Each segment was crossfaded with the next segment using 10 ms *cosine* ramps. Figure 3.4 shows the spectrograms of the six different exemplars of synthesized stimuli created for "Cackling Geese". The sampling rate was 48.828kHz. The orig. to silence (last 1 s) transition is not included in analysis.

3.2.3 Electrophysiological recordings

Acoustic brainstem responses (ABRs) were measured, Preyer's reflex and physical examination were performed to ensure the ears, especially the tympanic membrane, and hearing of the rats had no abnormalities. Five young adult (eight weeks) female Wistar rats weighing approximately 250 – 280gm were used for this study. Healthy rats were anesthetized with an initial induction dose by i.p. injection of a mixture of ketamine (80mg/kg) and xylazine (12mg/kg,). For maintenance dose of anesthesia during electrophysiological recordings, a pump delivered an i.p. infusion of 0.9% saline solution of ketamine (17.8mg/kg/h) and xylazine (2.7mg/kg/h) at a rate of 2.1ml/h. Body temperature was measured rectally and maintained with a heating pad (RWD Life Science, Shenzhen, China) and blanket at 38°C both during surgery

3.2 Materials and method

and recording. The state of the animal was monitored (temperature, and toe-pinch withdrawal reflexes) throughout the experiment, and the anaesthetic infusion rate was adjusted if necessary. The animal was placed inside a sound-attenuating chamber, and head fixed using hollow ear bars in a stereotactic frame (RWD Life Sciences). Auditory brainstem responses (ABRs) were recorded to evaluate the hearing sensitivity of animals before surgery. ABRs were evoked by the clicks ($500\mu s$ white noise pulses) at a rate of $23Hz$, and 400 click presentations were played at each intensity level ($30dB$ SPL to $80 dB$ SPL in $5 dB$ steps) for each rat. The clicks were played through the hollow ear bars using custom-made headphone drivers based on AS02204MR-N50-R (PUI audio, Dayton, USA). Stainless steel needle electrodes placed at the mastoids, vertex, nose, and back, and the ABR corresponded to the averaging of scalp potentials between mastoid and vertex. Normal hearing sensitivity was verified when the threshold of ABR was at or lower than $30 dB$ SPL. For the IC recording, the right temporal muscle and cranium were removed just anterior to lambda.

Extracellular multichannel neural recordings were recorded from IC by using single shank 32-channel ($50\mu m$ spacing between recording sites, **ATLAS Neuroengineering, E32-50-S1-L6**) silicon electrodes. Each morphed texture in the six exemplars were repeated for 10 trials. The neural signals were amplified by a PZ5 preamplifier and recorded at a sampling rate of $24.414kHz$ with an RZ2 system (Tucker-Davis Technologies). In total 480 multiunits were recored.(fifteen recording sites from five animals.No of recording sites/animal=3). All the multiunits were used for subsequent data analysis.

3.2.4 Data acquisition:

Stimuli presentation: The stimuli described above were presented via **AS02204MR-N50-R (PUI audio, Dayton, USA)** earphones, coupled to external metallic ear bars that were inserted into each ear canal, and driven by Tucker-Davis Technologies System III digital signal processor hardware. (48,828.125 Hz sample rate) together with systems running BrainWare software and custom-written MATLAB scripts.

3.2.5 Data analysis

3.2.5.1 Neural data quantification

The neural activity is analyzed offline using an “analog measure of multiunit activity” (*aMUA*), which measures the voltage signal power in the frequency band corresponding to the extracellularly recorded action potentials. The raw signal is bandpass filtered between 300 and 6000 Hz by a zero-phase shifting *Butterworth filter*, and took the absolute value of the filtered signal, and then downsampled it to 2 kHz. This method for quantifying neural activity is essentially identical to that previously used by [Choi et al., 2010; Chung et al., 1987; Kayser et al., 2007; King and Carlile, 1994; Schnupp et al., 2015].

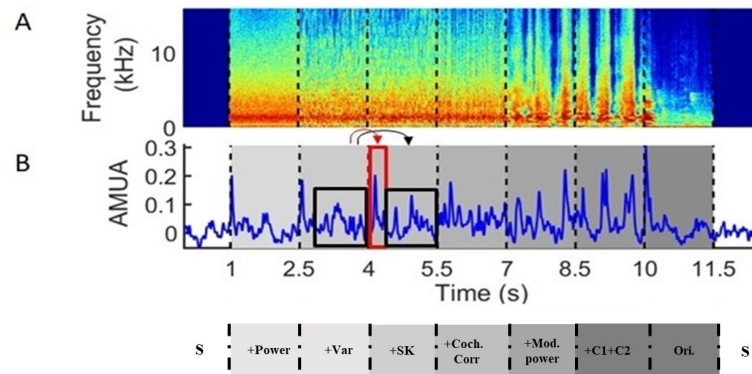


Fig. 3.5 (A) Synthesized stimuli for "Cackling Geese". (B) Neural responses from a single channel quantified as an analog multiunit activity. Red window after the transition point at time 4 s shows an "onset" transition response window, while the black window shows an "ongoing" or "sustained" transition.

3.2.5.2 Measuring neuronal response to the transitions in statistics

Measuring the onset neuronal responses: Testing for the statistical significance of any observed neural response transients to changes in statistical stimulus feature parameters with classical methods is very difficult. Neural responses fluctuations are known not to be normally distributed, but instead are poisson-like, with variances scaling with means. This violates two key assumptions of ANOVA style tests. And even non-parametric type ANOVA tests are difficult to apply to our data, given that the structure of our data is "nested". It would violate the independence assumption of such tests to treat every pair of neural responses around a stimulus transition as an independent sample, because the 6 different exemplars of each of the 13 different texture types constitute nested "random factors" that would need to be taken into account. In the absence of off-the-shelf statistics software that provides an adequate solution for this

3.2 Materials and method

purpose, we devised a nonparametric resampling method to judge the statistical significance of transient changes in neural response amplitudes spanning the stimulus feature transitions.

The objective of this test was to assess whether the absolute difference between the mean response amplitude over the 50 ms preceding the stimulus transition and the 50 ms following the transition was larger than would be expected by chance. Furthermore, to be able to interpret such an apparent, stimulus transition evoked change in neural response as a sensitivity to texture parameters, such significant responses to a particular type of transition should not be an isolated event, but should occur for several of the thirteen different texture types tested here. We carried out our test separately for each of the 13 textures, and then applied the criterion which required that, for a given stimulus transition, a unit would have to exhibit significant responses to a minimum of four of the 13 textures tested. We also conducted a control analysis to verify that this test and the criterion ensure a high degree of specificity.

To judge whether the change in mean neural response amplitude 50 ms on either side of a stimulus transition is larger than expected by chance, we first computed the observed difference (the “true transition response”) by averaging responses over each of the 10 repeats of each exemplar and over each of the 6 exemplars of each texture, and computing the absolute differences in these mean responses, and then we used a bootstrap method to estimate the expected null distribution for such differences by resampling the neural response time series during a steady state response period from 1000 ms to 100 ms prior to the transition. Neural response time series during this steady state response (sampled in 10 ms bins) were averaged over stimulus repeats to yield

3.2 Materials and method

a 6 exemplar by 90 time bin neural response matrix. To generate one “simulated null transition response” we picked, uniformly and independently, one random “simulated transition” time point for the mean response to each exemplar, computed the average responses during the 50 ms before and after that time point and calculated their absolute difference. These absolute differences were averaged over the 6 exemplars to generate one simulated null transition response value. This process was repeated 1000 times to generate a distribution of simulated null transition responses, and the p-value of the true transition response was computed as the percentile of the true transition response value in the distribution of null transition values. To be deemed to exhibit a significant transition response, a multiunit had to yield p-values < 0.05 for at least 4 of the 13 textures.

To verify that this procedure is highly selective and generates very few false positives, we conducted the following control: We simply replaced the true transition response value (which compares 50 ms before the transition against 50 ms after) with a “false” transition value which compares the response 50 ms before the transition against the response observed during the period from 100 to 50 ms prior to the transition. These false transition responses were then compared against the bootstrapped simulated null transition response distribution to compute “sanity check p-values” which would have to be attributable to false alarms. These sanity-check p-values were subjected to the same criterion of requiring at least four values below 0.05 to fulfil our significance criterion. We conducted this test on all 7 stimulus transitions and all 480 multiunits in our sample, and we obtained only a single false positive result for a single multiunit on

a single transition (Mod.power to +C1+C2). This demonstrates that the specificity of our test is very high.

Measuring the ongoing neuronal responses: The method that we described for testing the statistical significance of any observed neural response for transient responses cannot be applied as to ongoing responses due to lack of sufficient time windows. Therefore we developed a separate analysis method for the sustained response analysis.

For estimating the response before any specific statistical transition, we averaged the AMUAs in a time window of 1 s before the transition for each trial. For each texture, we resampled the averaged AMUA over the 6 exemplars and 10 trials with replacement, and then calculated the mean of these 60 numbers. We bootstrapped for 1000 times, and got the distribution of the mean over exemplar and trials before the transition. For the response after the transition we repeated the same procedure in the ongoing time window of 0.5 s to 1.5 s. We generated the distributions of mean responses, one for the pre-transition window and another for the post-transition window. The distributions thus obtained for both of these pre and post transition windows then we estimated the 2.5 and 97.5 percentile values, giving us 95% confidence intervals of the mean response amplitudes. For a specific transition, if the pre-transition and post transition confidence intervals did not overlap then we consider mean responses on either side of the transition to be significant.

3.3 Results

A total of 480 IC multiunits were analyzed in this study. Examples of one multiunit in response to the 6 step-wise morphs generated for the texture "Cackling Geese" are shown in figure 3.6. The multiunit shows, for example, clear onset responses to the +Var and +SK transitions, but the +Mod.power transition shows no obvious onset response. The red and black boxes illustrate the time windows used in statistical tests used to assess whether *transient* (red) or *sustained* (grey) responses to the +Var. stimuli condition.

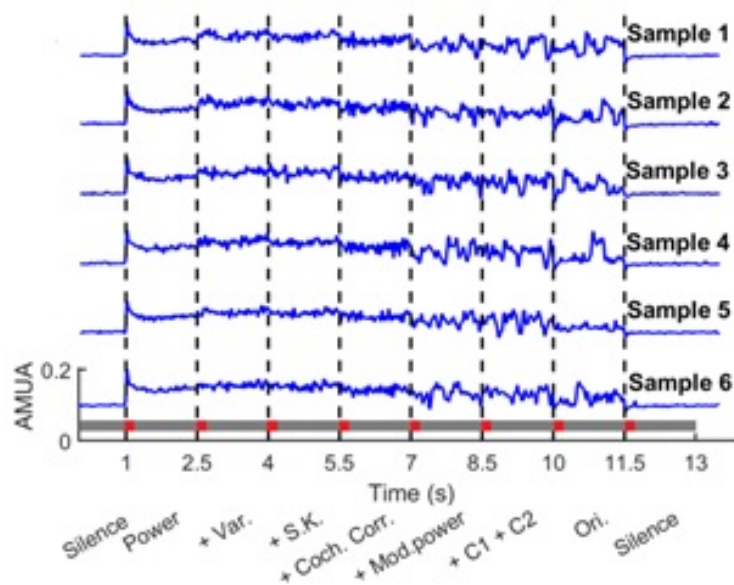


Fig. 3.6 Examples of one multiunit in response to 6 samples of the texture "Cackling Geese". A representative multiunit in the IC. Black dashed lines represent the statistical transitions in the stimulus. The red and gray bars at the bottom of the figure represent the *transient* and *sustained* responses respectively after each statistical transition point.

Figure 3.7 shows the percentage of IC multiunits exhibiting significant *transient* responses to changes in the sounds' statistical parameters.

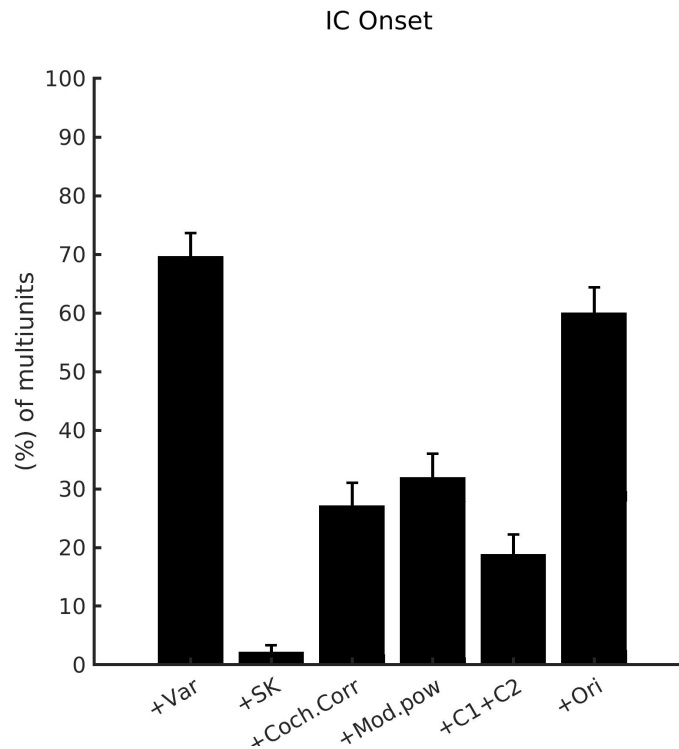


Fig. 3.7 The percentage of multiunits in the IC showed significant changes across statistical transitions for the transient response. The error bars represent the 95% Wilson confidence interval.

Figure 3.8 show the proportion of IC multiunits showing significant *sustained* response change over the statistical transitions.

As shown in figure 3.7 almost 70 % of the IC multiunits were sensitive to the change in envelope variance whereas only around 2% of IC multiunits exhibited significant onset responses to +SK. Approximately 30% multiunits were sensitive to both +Coch.Corr and +Mod.pow transitions whereas only about 15% of the multiunits were sensitive to +C1+C2. But for +Orig condition ~60% of the multiunits were sensitive to onset response.

3.3 Results

For the ongoing responses as shown in figure 3.8 more than 90% of the multiunits showed significant changes to +Var, ~25% of the multiunits were sensitive to +SK, ~60% were sensitive to + Coch.Corr, ~50% were sensitive to +Mod.pow, and only about 10% multiunits were sensitive to +C1+C2. For +Ori condition ~80% of the multiunits were sensitive.

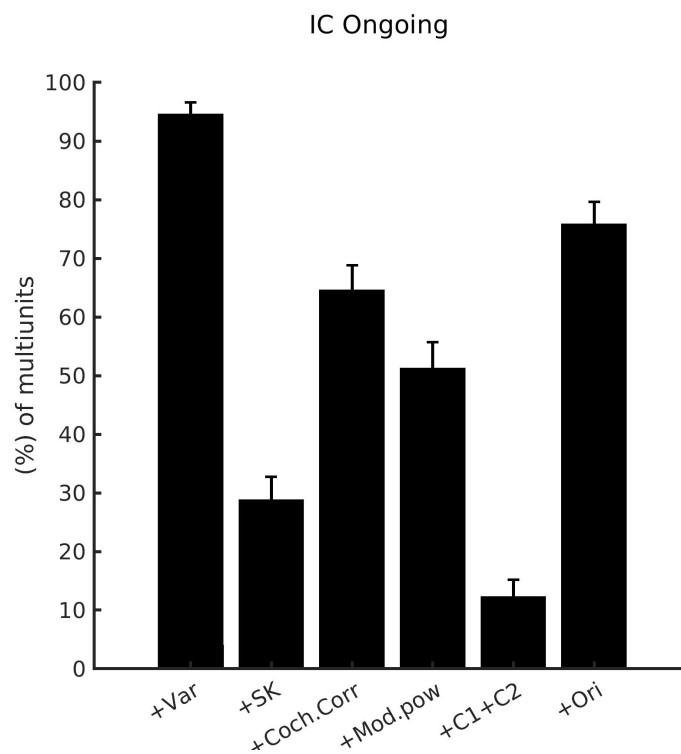


Fig. 3.8 The percentage of multiunits in the IC showed significant changes across statistical transitions for the sustained response. The error bars represent the 95% Wilson confidence interval.

Almost all the multiunits showed significant transient and sustained response change from the synthesized texture to the original sounds and suggesting that the neurons might be sensitive to the other statistics. The results showed that the IC neurons are most commonly sensitive to all the statistics. More multiunits showed significant

sustained responses for +SK, +Coch.corr, and +Mod.power than transient responses, suggesting that the latencies of the transition response are long for these statistical transitions.

3.4 Discussion

In this study, I used a set of synthesized stimuli to assess what percentages of midbrain neural population are sensitive to the statistical parameters known to characterize different types of environmental sound textures. In this study I have focused specifically on how the multiunits respond to the different statistical parameters present in the stimuli. How or even whether these statistical parameters are computed in the brain are beyond the scope of my research question but opens more interesting scientific questions to explore later.

I found that most multiunits in IC are sensitive to all the types of statistical features examined by [McDermott and Simoncelli, 2011]. While some of these results are perhaps unsurprising, given for example that neurons in the IC are known to be sensitive to modulation, other aspects are perhaps less expected. For example, it is not obvious why often narrowly frequency tuned IC neurons should be sensitive to cochlear correlations.

An efficient design of the sensory systems can be guided by the statistically efficient representation of environmental information [Attneave, 1954]. The statistical structure of natural signals are highly conserved across natural sounds [Attias and Schreiner, 1997; Escabi et al., 2003; Nelken et al., 1999; Singh and Theunissen, 2003; Voss and

Clarke, 1975]. Both, the peripheral and central auditory neurons use such statistical regularities to efficiently encode natural sounds [Attias and Schreiner, 1998; Escabi et al., 2003; Holmstrom et al., 2010; Lesica and Grothe, 2008; Nelken et al., 1999; Rieke et al., 1995; Woolley and Casseday, 2005]. In the auditory pathway, the *inferior colliculus* is an obligatory station which receives convergent inputs from the numerous brainstem structures and sends its highly processed outputs to the auditory thalamus, and, subsequently, to the primary auditory cortex.

Various studies have reported that IC neurons are sensitive to the spectral and temporal stimulus attributes.

[Escabi and Schreiner, 2002; Irvine and Gago, 1990; Krishna and Semple, 2000; Kuwada et al., 1997; Langner and Schreiner, 1988; Ramachandran et al., 1999; Rees and Møller, 1983, 1987; Schreiner et al., 1938; Schreiner and Langner, 1988].

However, these studies were mostly confined to pure tones and noise stimuli, which restricts our understanding of auditory encoding of natural sounds. Natural sounds are complex and are difficult to describe quantitatively. How does auditory brain interpret natural sounds still remains an open debate? Primarily, the complexities of natural sounds create a barrier to analyzing the auditory responses to these sounds [Attias and Schreiner, 1998].

“**Natural sound textures**” are suitable candidates for overcoming the limitations posed by less well parameterized natural sounds in studies related to auditory physiology. “**Natural sound textures**” are also a class of “**natural sounds**”, but they have an additional property of “temporal homogeneity” [McDermott and Simoncelli, 2011].

Using the methodology developed by [McDermott and Simoncelli, 2011], here I computed and analysed the marginals, cochlear correlations, modulation power and modulation correlation statistics of 200 natural sound textures and subjected them to *principal component analysis*. The sound textures used in this study are selected from the principal component spaces of the different statistical features of the corpus of 200 natural sound textures (figure 3.1). I used the simple generative model by [McDermott and Simoncelli, 2011], and resynthesized the morphed textures from the selected sound texture stimuli. The morphed textures transit through different types of statistical features present in the natural sound textures. Using these resynthesized morphed textures, I have quantified the percentage of the IC neural population that are sensitive to these statistical features. I found that, IC multiunits are sensitive to all the statistical features of natural sound textures. Further analysis of the *transient* and *sustained* response windows for IC multiunits exhibited that, though the IC multiunits can respond to these statistics during both the transitions, higher percentages of multiunits respond in the sustained windows (figures 3.7 and 3.8). This differential responses by the IC multiunits in transient and sustained windows are also supported by a previous study which has reported that envelopes are encoded differentially by IC neural population for both the transient and sustained response windows [Zheng and Escabí, 2008].

The +Power are like noise and hence drive most of the IC multiunits both during transient and ongoing response window. For +Var transition, 70% and 90% of the multiunits are driven in the transient and sustained response windows respectively. The +SK stimuli are mostly *water-like* [McDermott and Simoncelli, 2011]. The water-like

natural sound textures are mostly poorly correlated across many cochlear frequency bands [McDermott and Simoncelli, 2011] and the +SK morphed textures do not have "noisiness" unlike the +Power morphed textures. For +SK statistical transition ~2% IC multiunits responded during *transient* window and ~25% of the IC multiunits responded during the *sustained* response window.

For the +Coch.corr statistical transition, ~60% of the IC multiunits showed significant response in a *sustained* window whereas ~25% were sensitive in the transient window. The IC, due to its central location in the auditory pathway receives convergent inputs from multiple brainstem structures. The IC neurons have also been reported to perform temporal integration [Voytenko and Galazyuk, 2007]. As more and more statistics are added, the IC neural population may require more time to integrate and analyze.

For a wide variety of natural sounds, the spectrotemporal modulations are important attributes along with frequency components. The IC neurons are selective to both frequency components and spectrotemporal modulations of natural sounds [Escabí et al., 2003; Theunissen et al., 2000; Woolley and Casseday, 2005] and are key information-bearing attributes [Chi et al., 1999; Elliott and Theunissen, 2009; Singh and Theunissen, 2003]. The sensitivity of the IC neural population also varies considerably for natural sound modulations [Krishna and Semple, 2000; Rodríguez et al., 2010; Schreiner and Langner, 1988; Woolley and Casseday, 2005]. I found that around ~30% of the IC multiunits were sensitive to +Mod.power statistics for transient response window whereas ~50% were sensitive for the sustained response window. Increased response to +Mod.power statistical transition during the sustained response window may be attributed to the

differential encoding of the sound envelopes information during the sustained response window [Zheng and Escabi, 2008]. For all the statistics, more multiunits were sensitive in the sustained response window than the transient. Most of the IC multiunits (> 70%) were sensitive to the +Ori., in sustained windows. The +Ori.stimuli are not the morphed textures. They are random snippets (1.5 s long) from the original sound textures. Therefore they have higher information content in contrast to all the morphed textures. The +Ori. stimuli have rich information content in contrast to morphed texture (+C1+C2). Therefore higher percentage(>70%) of the IC multiunits have responded to +Ori. than to the +C1+C2. statistical transition during the *sustained* and more than (> 50%) multiunits are sensitive in the *transient* response window.

Deviance detection is a mechanism that is adopted by the auditory system to purge irrelevant foreseeable stimulation and provide perceptual saliency to those sounds that are unique, unpredictable, and therefore highly informative [Winkler and Schröger, 2015].

Stimulus specific adaptation (SSA) is quantified as the index of change in the firing rate of a neuron in response to a deviant stimulus when compared with its response to that same stimulus played as a standard. SSA was proposed to be a correlate of the deviance detection mechanism at the neuronal level [Ulanovsky et al., 2003], which population activity summation would build up until being detectable on the scalp as mismatch negativity (MMN)[Nelken and Ulanovsky, 2007].

In my experimental stimuli design, there are no frequently repeating sounds (standard stimuli) vs. rare acoustic events (deviant stimuli). Therefore in my experimental setup

it is difficult to establish any link between the sensitivity to the statistical features and stimulus specific adaptation.

Chapter 4

Sensitivity of Auditory Cortical Neurons to Statistical Features of Sound Textures

Abstract

Previous studies by [[McDermott and Simoncelli, 2011](#)] hypothesized that midbrain neural populations are driven by the statistical parameters present in the natural sound texture stimuli.

In chapter 3, I showed that at least around 40% of IC neurons are sensitive to each of the different types of statistical parameters that characterize environmental sound texture stimuli. While that suggests that a great deal of sensitivity to texture features is

already present at the level of the midbrain, it is nevertheless of interest to ask how the representation of these features at the level of cortical neural populations compares to that seen in the midbrain.

In this study, using multi-shank silicone electrodes inserted into the primary *auditory cortex* (AC), I recorded extracellular neural responses from the primary auditory cortices of five anesthetized female Wistar rats to the same set of synthetic "step-wise morphed" stimuli described in chapter 3. Curiously, I observed that overt sensitivity to statistical texture features was noticeably less common among cortical responses than among IC neurons.

4.1 Introduction

The set of all possible sound waves that the auditory system could in theory encounter is infinite, but the physical and statistical structure of the world nevertheless makes some types of sounds a great deal less likely than others [Attneave, 1954; Field, 1987].

A theory of neural representation and neural computation in sensory systems that takes into account the structure of the natural environment, was originally proposed by [Attneave, 1954; Barlow et al., 1961] and led to better understanding of visual system. Singh and Theunissen [2003] suggested that characterization of the statistics of natural sounds is essential for understanding acoustical perception and its underlying neuro-physiological basis. Rieke et al. [1995] have reported that broadband sounds whose power spectrum

matched with the power spectrum of the natural frog call when presented, auditory nerve fibers in the frog transmitted information more efficiently. Recent psychoacoustics study by [McDermott and Simoncelli, 2011] using “*natural sound textures*” suggest the significance of both lower and higher-order statistics in perception. They define “*sound textures*” as the collective result of many similar acoustic events which are distinguished by their “*temporal homogeneity*”.

Many physiological studies on the auditory cortex of mammalian species have reported that auditory neurons are tuned for a number of independent feature parameters such as *frequency*, *intensity*, *amplitude modulation*, *frequency modulation*, and *binaural structure* of simple stimuli. Except a few studies [Margoliash, 1983; Suga et al., 1978] where stimuli are selected based on ethological principles for the species, the underlying feature-processing mechanism of cortical neurons across mammalian species are not well-understood.

Behaviorally relevant stimuli have also been used to probe the physiology of the sensory systems. Specific studies such as pulse-echo tuned neurons in the bat [Suga et al., 1978], song selective neurons in songbirds [Margoliash, 1983] and call selective neurons in the primate [Newman and Wollberg, 1973] indicate that auditory system appears at least to be “*Selective*”. One limitation of classic neuroethological approaches, however, is that there can be a rather narrow focus on an animal’s conspecific vocalizations or similarly restricted set of sounds, neglecting the fact that most land animals are immersed in rich and diverse auditory scenes much of the time.

[[Nelken et al., 1999](#)] analyzed sounds from a range of different environments including both animal-vocalizations and non-animal sounds. They hypothesized that as preferred type of auditory stimulus for cortical neurons are largely unknown a search strategy should be applied on a sound corpus before establishing any relationships between properties of natural soundscapes and neuronal processing mechanisms in the auditory system. Most of these results are from studies of anesthetized animals. In A1, frequency response areas were found to be rather uniform V-shaped under anesthesia in different species [[Sally and Kelly, 1988](#)] whereas more complex patterns have been reported in unanesthetized conditions [[Abeles and Goldstein, 1972](#); [DeCharms et al., 1998](#); [Pelleg-Toiba and Wollberg, 1989](#)]. In many anesthetized preparations (e.g., barbiturate and ketamine), sound-evoked responses are typically transient [[DeWeese et al., 2003](#); [Doron et al., 2002](#); [Heil, 1997](#); [Phillips and Irvine, 1981](#)].

As per the “*efficient coding hypothesis*”, [[Barlow et al., 1961](#)], the sensory processing mechanism should construct an efficient representation of the sensory environment. Sparse encoding strategy in auditory cortex can provide efficient representations for natural scenes [[Olshausen and Field, 1997, 2004](#)].

Sparse representations may also offer energy efficient coding, where fewer spikes are required compared to dense representations [[Attwell and Laughlin, 2001](#); [Laughlin and Sejnowski, 2003](#); [Levy and Baxter, 1996](#)]. It has been reported that a large fraction of cortical neurons remain silent to many stimuli most of the time which may be attributed to high “*stimulus specific selectivity*” of cortical neurons. This can also be attributed to sparse coding strategy adopted by cortical neurons. Sparse coding in unanaesthetized

auditory cortex has been reported by [Hromádka et al., 2008] in rat cortex and also in the visual [Baddeley et al., 1997; Vinje and Gallant, 2000], motor [Brecht et al., 2004], barrel [Margrie et al., 2002]; olfactory systems [Perez-Orive et al., 2002; Rinberg et al., 2006; Szyszka et al., 2005], the zebra finch auditory system [Hahnloser et al., 2002] and cat lateral geniculate nucleus [Dan et al., 1996]. As per [Hromádka et al., 2008], population sparseness in awake rat auditory cortex may be attributed to three factors. First, failure of neural populations to respond to the presented tonal stimuli. Second, brief response duration and third low amplitude response. They found that population response is sparse and less than 5% of neural population respond to stimuli (*tones, frequency-modulated sweeps, white-noise bursts and, natural sound*) at any time. They also argue that response heterogeneity is property of awake auditory cortex.

Both perceptual evidence of recognizing natural sound textures in association with higher order statistics and evidence from neural adaptation to stimuli statistics suggest that auditory cortical neurons should be sensitive to higher-order as well as lower-order statistics. It is anticipated that cortical neurons may adopt heterogeneous and sparse encoding strategy to represent these statistical features.

In this study (as in the previous chapter), 13 natural textures were chosen from a corpus of 200. The marginals, correlations and modulation power statistics of the corpus were computed and projected into their respective principal component spaces. Visualizing the coordinates of the selected textures within a dimensionality reduced representation of the distribution of the parameters of our diverse corpus allowed us to verify that the textures selected for the current study are widely distributed through,

and cover a substantial part of the range of, the parameter space covered by the corpus, and the 13 sample textures can therefore be considered as "representative" samples. The selected stimuli are resynthesized as described in section 4.2 of chapter 3. Acute cortical responses were recorded using multi-shank silicon electrodes. I found that around 1% of cortical neural population were sensitive to cochlear correlations only during onset response. Approximately 30% of the multiunits were sensitive to +Var statistical transition during onset response. During ongoing response about 1% of the multiunits were sensitive to +Mod.powe transition only. For rest of the statistical transitions auditory cortical multiunits showed no response at all.

The rest of the chapter has been organized as follows. Section 4.2 describes materials and methodology. Section 4.4 summarizes the data analysis procedures and results. Section 4.5 deals with discussion and conclusion.

4.2 Materials and Methods

4.2.1 Animal preparation

Five young adult (~eight weeks old) female Wistar rats weighing approximately 250–280gm were used for the AC recordings. All rats were purchased from the Chinese University of Hong Kong. The experiment procedures in the study were approved by the Animal Research Ethics Sub-Committee at the City University of Hong Kong, and performed under license by the Department of Health of Hong Kong [Ref. No. (16-86), (18-167) in DH/HA and P/8/2/5 Pt.5].

4.2.2 Stimulus construction

The original and the "morphed" sound textures that are described in detail in section 3.2 of chapter 3 were also used in this study. A review is given here for the sake of completeness of the chapter. All the "morphed textures" were generated using the *Sound Synthesis ToolBox V1.7* developed by [McDermott and Simoncelli, 2011]. The toolbox requires an input structure for synthesizing the morphed textures from white noise. Most important parameters that are required by the toolbox are given here: (a) the original sound texture from which the target statistics of the subbands are measured. (b) the statistics that are to be modified are predetermined (c) number of cochlear and modulation bandpass filter banks (d) filter properties (cut-off frequencies) (e) seed value for generating white noise (f) envelope compression ratio (g) boundary handling function which is required to merge the synthesized subbands. Initiating with such a structured information the toolbox breakdown the input sound texture into a predefined number of subbands and measures the required statistics for each subband. Then, the subband statistics of the white noise are modified iteratively using conjugate gradient approach until they achieve a predetermined level of the target statistics. Each "morphed texture" has 7 different segments, where each segment has different statistical features. Segment 1 of the "morphed texture" is only "power matched" to the original texture, hereafter called as "+Power". Segment 2 has both power and variance matched, hereafter called as "+Var". Segment 3 have additionally skew and kurtosis matched and will be referred as "+SK". Segment 4 have additionally cochlear

correlations matched, subsequently will be referred as ”+Coch. Corr”. Segment 5 has additionally modulation power matched, hereafter known as ”+Mod.power”. Segment 6 has additionally modulation correlations (C1 and C2) matched, will be referred as ”+C1+C2”. The last segment has a snippet from the original texture and will be referred as ”Ori”. At the beginning and end of each morphed texture 1 s silence is also added. All the pair-wise segments are crossfaded with *10ms cosine* ramps in order to avoid any spectral splatter.

In my data analysis section I have eliminated the last one second of silence as responses to the original to silence transition (last segment in the ”morphed texture”) are not interpretable in terms of the questions and parameters considered in this study. A total of 78 ”morphed texture stimuli” were created (13 original textures · 6 exemplars with different random seeds for each each texture) for this study.

4.2.3 Electrophysiological recordings

I used Preyer’s reflex mechanism and also verified the tympanic membranes to ensure the normal hearing capability of the animals before all recording experiments. For acute recordings, the rats were anesthetized with an initial induction dose by i.p. injection of a mixture of ketamine (*80mg/kg*) and xylazine (*12mg/kg*). For maintenance dose of anesthesia during electrophysiological recordings, a pump delivered an i.p. infusion of 0.9% saline solution of ketamine (*17.8mg/kg/h*) and xylazine (*2.7mg/kg/h*) at a rate of *2.1ml/h*. Eye gel (LubriThal, Dechra Veterinary Product A/S Mekuvej 9 DK-7171 Uldum) was applied to prevent the eyes from drying. Their outer ear canals and

4.2 Materials and Methods

tympanic membranes were inspected under microscope (RWD Life Sciences, China). Body temperature was measured rectally and maintained with heating pad (RWD Life Science, Shenzhen, China) and blanket at $38^{\circ}C$ both during surgery and recording. The state of the animal was monitored (temperature, and toe-pinch withdrawal reflexes) throughout the experiment, and the infusion rate was adjusted accordingly.

The animal was placed inside a sound-attenuating chamber, and head fixed using ear bars in a stereotactic frame (RWD Life Sciences). Auditory brainstem responses (ABRs) were recorded to evaluate the hearing sensitivity of animals both before and after craniotomy. ABRs were evoked by the clicks ($500\mu s$ rectangular pulse) at a rate of $23Hz$, and 400 click presentations were played at each intensity level ($30dB SPL$ to $80dB SPL$ in 5 dB steps) for each rat. The clicks were played through hollow ear bars. Stainless steel needle electrodes placed at the mastoids, vertex, nose and back, and the ABR corresponded to the averaging of scalp potentials between mastoid and vertex. Normal hearing sensitivity was verified when the threshold of ABR was at or lower than $30dB SPL$. A deep cut in the midline of the skull was made and surgical field was exposed. Local anesthesia Lignocaine (0.3 mL, 20 mg/mL, Troy Laboratories Pty Ltd, Australia) was applied on top of the surgical area. Craniotomy was made over the right temporal cortex. From a point 2.5 mm posterior to bregma, a line was drawn perpendicular to the sagittal suture to the right temporal ridge, and the cross point of this line and the ridge was marked. The right temporal muscle was removed, and a 4x6 mm opening was drilled in the temporal bone, and the cranium was removed to expose

the right AC [Polley et al., 2007]. The dura was removed, and the recording site was kept moist with 0.9% saline during entire recording period.

Extracellular neural recordings were made with multi-shank 64-channel silicon probe electrodes (100 μm spacing between recording sites, *ATLAS Neuroengineering, E64-100-S4-L6-600*). A total of 576 multiunits were recorded during 9 multielectrode penetrations into the cortices of five female Wistar rats. (Two penetrations for four animals, and one penetration from the fifth). The neural signals were amplified by a PZ5 preamplifier and recorded at a sampling rate of 24.414 kHz with an RZ2 system (Tucker-Davis Technologies).

4.2.4 Data acquisition

Stimuli were presented via AS02204MR-N50-R (PUI audio, Dayton, USA) earphones, coupled to external metallic ear bars that were inserted into each ear canal, and driven by Tucker-Davis Technologies System III digital signal processor hardware, (48,828.125 Hz: sample rate) together with systems running BrainWare software and custom written MATLAB scripts. Pure-tone stimuli were used to obtain frequency response areas, to determine tonotopic gradients which were analysed offline to confirm the cortical fields from which the recordings were made.

The "morphed textures", that are described in chapter 3 were also used for the current study. The recordings included 6 exemplars of morphed textures for each of the 13 natural texture samples chosen from the corpus. Each morphed texture stimulus gradually morphs shaped noise to the full texture in steps with the same for

all electrophysiological recordings.

The "morphed textures" were presented contralaterally to the exposed AC at 80dB SPL in a randomized order (10 repeats/morphed texture) , with 1 s of silence between subsequent "morphed textures".

4.3 Data analysis

4.3.1 Quantifying the neural responses

The neural activity is analyzed offline using an "*analog measure of multiunit activity*" (AMUA) [Choi et al., 2010; Chung et al., 1987; Kayser et al., 2007; King and Carlile, 1994; Schnupp et al., 2015; Schroeder et al., 1998], which is a measurement of the voltage signal power in the frequency bands from the extracellularly recorded action potentials. The raw signals are bandpass filtered between 300-6000 Hz by a zero-phase shifting *Butterworth filter*, and took the absolute value of the filtered signal, and then downsampled it to 2 kHz to enhance computational efficiency.

4.3.2 Measuring the neuronal responses to the statistical transitions

For estimating the number of auditory cortical multiunits that are sensitive to transient and ongoing responses I adopted the same statistical procedures that has been elaborated in chapter 3 section 3.2. The percentage of multiunits that showed the significant transient and sustained responses are summarized in figure 4.2 and 4.3 respectively.

4.4 Results

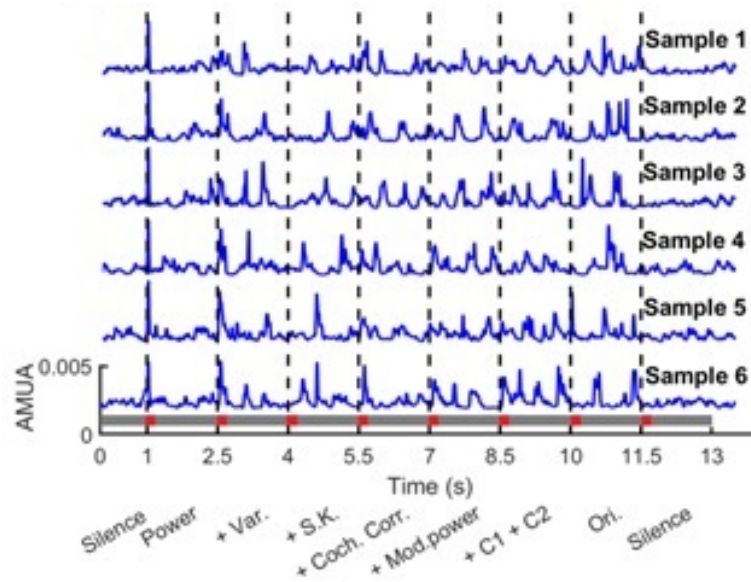


Fig. 4.1 Example of one AC multiunit in response to 6 samples of the texture (“Cackling Geese”). Black dashed lines represent the statistical transitions in the stimulus. The red and gray bars at the bottom of the figure represent the transient and sustained responses after each statistical transition point.

We recorded 576 multiunits and we had 78 (=13 textures*6 exemplars) ”morphed stimuli” in our stimuli set. For every multiunit the mean baseline amplitude (1s before the onset) was calculated over all trials. For every multiunit we also calculated the onset peaks in a time window of [0- 200 ms]. To determine the onset peaks we imposed the following condition [Beckers and Gahr, 2012]:

An onset amplitude was considered to be a *peak amplitude* if and only if

$$Peak_{amplitude} > Baseline_{amplitude} + 3 * std(Baseline_{amplitude}) \quad (4.1)$$

We imposed a very restrictive condition to select multiunits (out of 576) for further statistical analysis. We selected only those multiunits which had more than 70 onset peaks (as we had 78 "morphed stimuli"). Using this method 129 multiunits were selected (out of 576 multiunits) for statistical analysis.

An example of responses of one multiunit in response to 6 samples of the texture is shown in figure 4.1. The neural responses change over some of the statistical transitions. For example, we can clearly see both transient (red bar) and sustained (gray bar) responses following the transition from +Power to +Var.

The response of the AC multiunits was analyzed using custom written MATLAB scripts. The percentage of multiunits that showed the significant transient response and sustained response is summarized in figure 4.2.

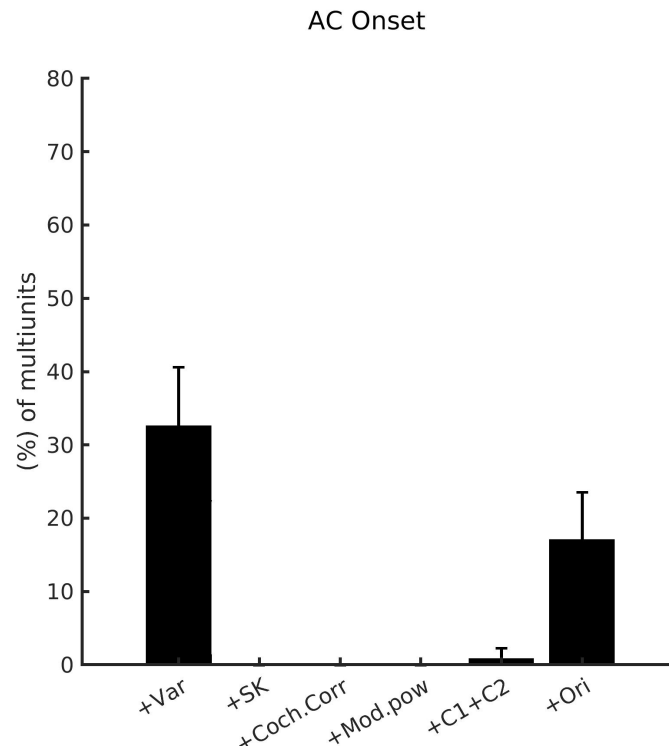


Fig. 4.2 The percentage of AC multiunits that showed significant changes across statistical transitions for onset response. The error bars represent the 95% Wilson confidence interval.

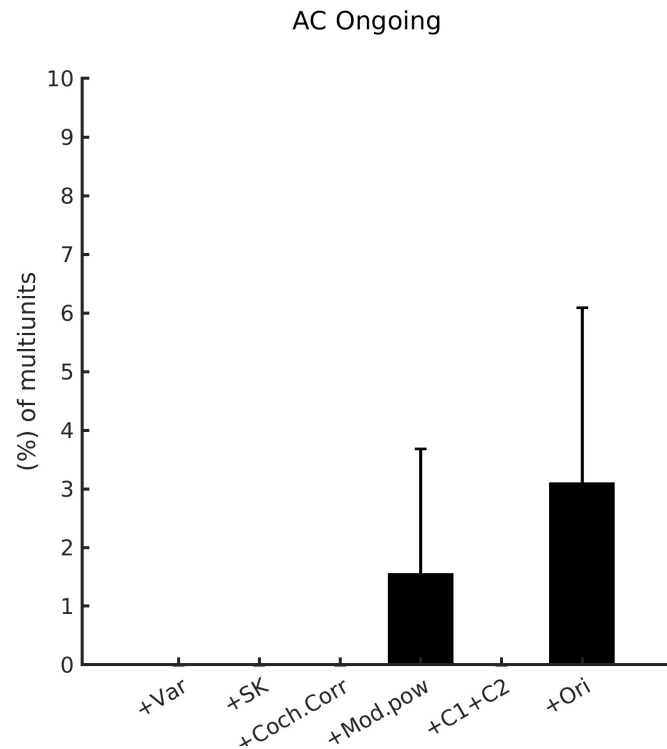


Fig. 4.3 The percentage of AC multiunits that showed significant changes in their sustained firing following changes in the statistical features of texture stimuli. The error bars represent the 95% Wilson confidence interval.

The percentage of multiunits that showed the significant sustained responses to changes in texture features are shown in figure 4.3. For onset response, only 30% of auditory cortical multiunits responded to +Var statistical transition and ~1% only were sensitive to +C1+C2 whereas for ongoing response only about 2% of multiunits were sensitive to only for +Mod.pow transition. This is also surprising to see that auditory cortical units are not sensitive to any statistical transitions (except for a tiny number of significant transition responses for +Mod.pow). It is interesting to see that approximately 20% of the multiunits were sensitive to +Ori during onset response and only about 3% multiunits were sensitive during the sustained response. This indicates

that there are yet unidentified features of the environmental sounds, which the current synthesis model does not include.

4.5 Discussion

The current study has quantified the percentage of neural populations in auditory cortex that are sensitive to different statistical structures present in “*natural sound textures*”. It is important to note here that auditory cortical properties were not modeled in the original model suggested by [McDermott and Simoncelli, 2011]. It is interesting to see that cortical neurons are also sensitive to these statistical features in the first place but significantly less prominent than the IC neurons.

One possible reason may be because in the auditory pathway, the inferior colliculus (ICC) is an obligatory station which receives convergent inputs from numerous brainstem structures and sends its highly processed outputs to the auditory thalamus and, to the primary auditory cortex subsequently.

The sustained firing rates of the cortical neurons have been reported to be much lower for tonal stimuli [Decharms and Merzenich, 1996].

The auditory cortex of awake animals show higher sustained response to sounds [Gaese and Ostwald, 2003]; [Barbour and Wang, 2003; Chimoto et al., 2002; Recanzone, 2000]. Moving one step ahead [Hromádka et al., 2008] have hypothesized that response heterogeneity is a hallmark of awake auditory cortex. However I found that around

30% of neural population respond to +Var only for transient response window 4.2 and only 2% of the multiunits responded to +Mod.pow during ongoing response window.

It can be seen in figures 4.2 and 4.3 that sustained response of cortical neurons to different statistical parameters are different than transient response.

Auditory system progress from a distributed and redundant encoding strategy at the periphery to a more heterogeneous encoding in cortical structures. [Averbeck et al., 2006; Shamir and Sompolinsky, 2004; Sompolinsky et al., 2001].

Auditory cortical neurons by and large do not respond to the change of these statistics (in contrast to IC neurons, as we have seen in the previous chapter). There has been a lot of study that supports the idea of sparse encoding strategy adopted by cortical neurons. Auditory system as a whole, keeps processing a lot of sounds all the time and many of them are redundant. Therefore if it does not adopt redundancy reducing mechanisms like mismatch negativity (MMN) [Carbajal and Malmierca, 2018], contrast gain control [Rabinowitz et al., 2011], stimulus specific adaptation (SSA)[Carbajal and Malmierca, 2018] then it will be really hard for the system to keep processing sounds all the time. Sparse encoding strategy is a kind of energy saving and division of labor mechanism. Specific areas of cortical neurons may be responding to specific sound types.

It has also been reported that at least some auditory cortical neurons may be tuned to conspecific animal vocalizations, and are poorly driven by white noise (primates: [Rauschecker et al., 1995]; bats:[Ohlemiller et al., 1996]; birds:[Margoliash, 1983, 1986; Scheich et al., 1979]). A study by [Theunissen et al., 2000] has used reverse correlation analysis to shown that cortical neurons show higher responses to natural

stimuli than to noise stimuli. It has also been reported that the auditory cortical neurons exhibit non-linear response to animal vocalizations i.e. response to components of vocalizations are poor whereas to the complete vocalization they respond very strongly. In the absence of conspecific animal vocalizations in the current study, only ~15% and ~3% of cortical neurons are driven by original natural stimuli during onset and ongoing windows respectively, than to other statistical transitions.

Chapter 5

General Discussion and Conclusion

5.1 General Discussion

The major findings of my thesis can be summarized as following:

1. **The statistical parameters space of the natural sound texture space are mostly redundant.**

In this study, I have explored the distribution of statistical feature parameters of a corpus of 200 natural sound textures.

Principal component analysis of the statistical feature space of the corpus revealed that, the seemingly large "dimensionality" of natural sound texture space are mostly redundant and can be fairly compensated by only first two principal components. I found it interesting that, the natural sound textures can be broadly classified into: (a) "highly correlated" vs. "poorly correlated" (c) "sparse" vs. "continuous" (d) "fast modulating" vs. "slowly modulating" sounds.

2. The IC and AC multiunits and their sensitivities to the statistical transitions

Auditory system progress from a distributed and redundant encoding strategy at the periphery to a more heterogeneous encoding in the cortical structures [Averbeck et al., 2006; Shamir and Sompolinsky, 2004; Sompolinsky et al., 2001].

In the auditory pathway, the *inferior colliculus* (IC) acts as an obligatory station which receives convergent inputs from the numerous brainstem structures. Then it sends the highly processed outputs to the auditory thalamus, and, subsequently, to the primary auditory cortex. Cortex on the other hand adopts the sparse encoding strategy which has been reported to be an efficient approach for the representations of natural scenes [Olshausen and Field, 1997, 2004]. The IC and AC multiunits and their responses to the statistical transitions present in the morphed textures can be summarized as following:

- (a) In this study, I found that, for the +Var morphed textures, above 90% of the IC multiunits displayed significant response during ongoing response. On the other hand merely 30% of AC multiunits were sensitive only during onset response. Surprisingly, none of the AC multiunits responded to +Var during ongoing response.
- (b) Only about 2% of the IC multiunits exhibited significant responses to +SK during the transient window whereas >20% responded during the ongoing window. Approximately 20-30% of IC multiunits were sensitive +Coch.Corr

and +Mod.power during the onset window. But during the ongoing response 50-60% of the multiunits were sensitive to +Coch.Corr and +Mod.power.

- (c) For the transient response window, around 60% of the IC multiunits and ~15% AC multiunits showed significant responses to +Ori.
- (d) For the sustained response window, around 25% of the IC multiunits showed significant response for +SK whereas none of the AC multiunits showed significant response. For the +Coch.corr statistical transition, ~60% of the IC multiunits and none of the AC multiunits showed significant response in the sustained response window.
- (e) For the +Mod.power statistical transition, ~50% of the IC multiunits and ~1% of the AC multiunits showed significant response in the sustained response window.
- (f) For the +C1+C2 statistical transition, ~10% of the IC multiunits and no cortical multiunits showed significant response in the sustained response window.
- (g) For the Ori. statistical transition, ~75% of the IC multiunits and ~3% of the AC multiunits showed significant response in the sustained response window.

I found that, though the IC multiunits can respond to these statistics during both the *transient* and *sustained* response windows, higher percentage of the multiunits respond during sustained windows. This differential responses in the transient and sustained

5.1 General Discussion

windows are also supported by a previous study which has reported that the sound envelopes are encoded differentially for both the transient and sustained response windows [Zheng and Escabi, 2008]. Due to the central location of the IC, in the auditory pathway it receives convergent inputs from multiple brainstem structures. Therefore, IC multiunits showed higher percentages of sensitivities to the statistical transitions both in the *transient* and *sustained* response windows than the AC multiunits.

References

- Abeles, M. and Goldstein, M. H. (1972). Responses of single units in the primary auditory cortex of the cat to tones and to tone pairs. *Brain research*.
- Adelson, E. H. and Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Josa a*, 2(2):284–299.
- Aitkin, L. (1990). *The auditory cortex: structural and functional bases of auditory perception*. Chapman & Hall.
- Attias, H. and Schreiner, C. E. (1997). Temporal low-order statistics of natural sounds. In *Advances in neural information processing systems*, pages 27–33.
- Attias, H. and Schreiner, C. E. (1998). Coding of naturalistic stimuli by auditory midbrain neurons. In *Advances in neural information processing systems*, pages 103–109.
- Attneave, F. (1954). Some informational aspects of visual perception. *Psychological review*, 61(3):183.
- Attwell, D. and Laughlin, S. B. (2001). An energy budget for signaling in the grey matter of the brain. *Journal of Cerebral Blood Flow & Metabolism*, 21(10):1133–1145.
- Averbeck, B. B., Latham, P. E., and Pouget, A. (2006). Neural correlations, population coding and computation. *Nature reviews neuroscience*, 7(5):358–366.
- Bacon, S. P. and Grantham, D. W. (1989). Modulation masking: Effects of modulation frequency, depth, and phase. *The Journal of the Acoustical Society of America*, 85(6):2575–2580.
- Baddeley, R., Abbott, L. F., Booth, M. C., Sengpiel, F., Freeman, T., Wakeman, E. A., and Rolls, E. T. (1997). Responses of neurons in primary and inferior temporal visual cortices to natural scenes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 264(1389):1775–1783.
- Barbour, D. L. and Wang, X. (2003). Auditory cortical responses elicited in awake primates by random spectrum stimuli. *Journal of Neuroscience*, 23(18):7194–7206.
- Barlow, H. B. et al. (1961). Possible principles underlying the transformation of sensory messages. *Sensory communication*, 1:217–234.

- Baumann, S., Griffiths, T. D., Sun, L., Petkov, C. I., Thiele, A., and Rees, A. (2011). Orthogonal representation of sound dimensions in the primate midbrain. *Nature neuroscience*, 14(4):423–425.
- Beckers, G. J. and Gahr, M. (2012). Large-scale synchronized activity during vocal deviance detection in the zebra finch auditory forebrain. *Journal of Neuroscience*, 32(31):10594–10608.
- Blackwell, J. M., Taillefumier, T. O., Natan, R. G., Carruthers, I. M., Magnasco, M. O., and Geffen, M. N. (2016). Stable encoding of sounds over a broad range of statistical parameters in the auditory cortex. *European Journal of Neuroscience*, 43(6):751–764.
- Boemio, A., Fromm, S., Braun, A., and Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature neuroscience*, 8(3):389–395.
- Brand, A., Behrend, O., Marquardt, T., McAlpine, D., and Grothe, B. (2002). Precise inhibition is essential for microsecond interaural time difference coding. *Nature*, 417(6888):543–547.
- Brawer, J. R., Morest, D. K., and Kane, E. C. (1974). The neuronal architecture of the cochlear nucleus of the cat. *Journal of Comparative Neurology*, 155(3):251–299.
- Brecht, M., Schneider, M., Sakmann, B., and Margrie, T. W. (2004). Whisker movements evoked by stimulation of single pyramidal cells in rat motor cortex. *Nature*, 427(6976):704–710.
- Carbajal, G. V. and Malmierca, M. S. (2018). The neuronal basis of predictive coding along the auditory pathway: from the subcortical roots to cortical deviance detection. *Trends in hearing*, 22:2331216518784822.
- Chachada, S. and Kuo, C.-C. J. (2014). Environmental sound recognition: A survey. *APSIPA Transactions on Signal and Information Processing*, 3.
- Chase, S. M. and Young, E. D. (2005). Limited segregation of different types of sound localization information among classes of units in the inferior colliculus. *Journal of Neuroscience*, 25(33):7575–7585.
- Cheung, S. W., Bedenbaugh, P. H., Nagarajan, S. S., and Schreiner, C. E. (2001). Functional organization of squirrel monkey primary auditory cortex: responses to pure tones. *Journal of neurophysiology*, 85(4):1732–1749.
- 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.
- Chimoto, S., Kitama, T., Qin, L., Sakayori, S., and Sato, Y. (2002). Tonal response patterns of primary auditory cortex neurons in alert cats. *Brain research*, 934(1):34–42.

- Choi, Y.-S., Koenig, M. A., Jia, X., and Thakor, N. V. (2010). Quantifying time-varying multiunit neural activity using entropy-based measures. *IEEE Transactions on Biomedical Engineering*, 57(11):2771–2777.
- Chung, S., Jones, L. C., Hammond, B., King, M., Evans, R., Knott, C., Keating, M., and Anson, M. (1987). Signal processing technique to extract neuronal activity from noise. *Journal of neuroscience methods*, 19(2):125–139.
- Clugnet, M.-C., LeDoux, J. E., and Morrison, S. F. (1990). Unit responses evoked in the amygdala and striatum by electrical stimulation of the medial geniculate body. *Journal of Neuroscience*, 10(4):1055–1061.
- Dahmen, J. C. and King, A. J. (2007). Learning to hear: plasticity of auditory cortical processing. *Current opinion in neurobiology*, 17(4):456–464.
- Dan, Y., Atick, J. J., and Reid, R. C. (1996). Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory. *Journal of Neuroscience*, 16(10):3351–3362.
- Dau, T., Kollmeier, B., and Kohlrausch, A. (1997). Modeling auditory processing of amplitude modulation. i. detection and masking with narrow-band carriers. *The Journal of the Acoustical Society of America*, 102(5):2892–2905.
- DeCharms, R. C., Blake, D. T., and Merzenich, M. M. (1998). Optimizing sound features for cortical neurons. *science*, 280(5368):1439–1444.
- Decharms, R. C. and Merzenich, M. M. (1996). Primary cortical representation of sounds by the coordination of action-potential timing. *Nature*, 381(6583):610–613.
- Depireux, D. A., Simon, J. Z., Klein, D. J., and Shamma, S. A. (2001). Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. *Journal of neurophysiology*, 85(3):1220–1234.
- DeWeese, M. R., Wehr, M., and Zador, A. M. (2003). Binary spiking in auditory cortex. *Journal of Neuroscience*, 23(21):7940–7949.
- Ding, N. and Simon, J. Z. (2013). Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. *Journal of Neuroscience*, 33(13):5728–5735.
- Doron, N. N., Ledoux, J. E., and Semple, M. N. (2002). Redefining the tonotopic core of rat auditory cortex: physiological evidence for a posterior field. *Journal of Comparative Neurology*, 453(4):345–360.
- Ehret, G. and Merzenich, M. M. (1988). Complex sound analysis (frequency resolution, filtering and spectral integration) by single units of the inferior colliculus of the cat. *Brain Research Reviews*, 13(2):139–163.
- Elhilali, M., Ma, L., Micheyl, C., Oxenham, A. J., and Shamma, S. A. (2009). Temporal coherence in the perceptual organization and cortical representation of auditory scenes. *Neuron*, 61(2):317–329.

- Elliott, T. M. and Theunissen, F. E. (2009). The modulation transfer function for speech intelligibility. *PLoS computational biology*, 5(3).
- Escabí, M. A., Miller, L. M., Read, H. L., and Schreiner, C. E. (2003). Naturalistic auditory contrast improves spectrotemporal coding in the cat inferior colliculus. *Journal of Neuroscience*, 23(37):11489–11504.
- Escabí, M. A. and Schreiner, C. E. (2002). Nonlinear spectrotemporal sound analysis by neurons in the auditory midbrain. *Journal of Neuroscience*, 22(10):4114–4131.
- Farb, C. R. and Ledoux, J. E. (1997). Nmda and ampa receptors in the lateral nucleus of the amygdala are postsynaptic to auditory thalamic afferents. *Synapse*, 27(2):106–121.
- Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Josa a*, 4(12):2379–2394.
- Fish, R., Danneman, P. J., Brown, M., and Karas, A. (2011). *Anesthesia and analgesia in laboratory animals*. Academic press.
- Font, F., Roma, G., and Serra, X. (2013). Freesound technical demo. In *Proceedings of the 21st ACM international conference on Multimedia*, pages 411–412.
- Fritz, J. B., Elhilali, M., David, S. V., and Shamma, S. A. (2007). Does attention play a role in dynamic receptive field adaptation to changing acoustic salience in a1? *Hearing research*, 229(1-2):186–203.
- Gaese, B. H. and Ostwald, J. (2003). Complexity and temporal dynamics of frequency coding in the awake rat auditory cortex. *European Journal of Neuroscience*, 18(9):2638–2652.
- Garcia-Lazaro, J., Ahmed, B., and Schnupp, J. (2006). Tuning to natural stimulus dynamics in primary auditory cortex. *Current Biology*, 16(3):264–271.
- Garcia-Lazaro, J. A., Ahmed, B., and Schnupp, J. W. (2011). Emergence of tuning to natural stimulus statistics along the central auditory pathway. *PloS one*, 6(8):e22584.
- Glasberg, B. R. and Moore, B. C. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing research*, 47(1-2):103–138.
- Goldstein, M. H. and Abeles, M. (1975). Single unit activity of the auditory cortex. In *Auditory system*, pages 199–218. Springer.
- Goldstein Jr, M., Abeles, M., Daly, R., and McIntosh, J. (1970). Functional architecture in cat primary auditory cortex: tonotopic organization. *Journal of neurophysiology*, 33(1):188–197.
- Gygi, B., Kidd, G. R., and Watson, C. S. (2004). Spectral-temporal factors in the identification of environmental sounds. *The Journal of the Acoustical Society of America*, 115(3):1252–1265.

- Hahnloser, R. H., Kozhevnikov, A. A., and Fee, M. S. (2002). An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature*, 419(6902):65–70.
- Hall, J. L. and Goldstein Jr, M. H. (1968). Representation of binaural stimuli by single units in primary auditory cortex of unanesthetized cats. *The Journal of the Acoustical Society of America*, 43(3):456–461.
- Heeger, D. J. and Bergen, J. R. (1995). Pyramid-based texture analysis/synthesis. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 229–238.
- Heil, P. (1997). Auditory cortical onset responses revisited. ii. response strength. *Journal of neurophysiology*, 77(5):2642–2660.
- Heil, P., Rajan, R., and Irvine, D. (1994). Topographic representation of tone intensity along the isofrequency axis of cat primary auditory cortex. *Hearing research*, 76(1-2):188–202.
- Heil, P., Rajan, R., and Irvine, D. R. (1992). Sensitivity of neurons in cat primary auditory cortex to tones and frequency-modulated stimuli. ii: Organization of response properties along the ‘isofrequency’ dimension. *Hearing research*, 63(1-2):135–156.
- Higham, N. J. (1988). Computing a nearest symmetric positive semidefinite matrix. *Linear algebra and its applications*, 103:103–118.
- Holmstrom, L. A., Eeuwes, L. B., Roberts, P. D., and Portfors, C. V. (2010). Efficient encoding of vocalizations in the auditory midbrain. *Journal of Neuroscience*, 30(3):802–819.
- Hromádka, T., DeWeese, M. R., and Zador, A. M. (2008). Sparse representation of sounds in the unanesthetized auditory cortex. *PLoS biology*, 6(1).
- Hubel, D. H. and Wiesel, T. N. (1977). Ferrier lecture-functional architecture of macaque monkey visual cortex. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 198(1130):1–59.
- Imig, T., Adria, H., et al. (1977). Binaural columns in the primary field (a1) of cat auditory cortex. *Brain research*, 138(2):241–257.
- Irvine, D. and Gago, G. (1990). Binaural interaction in high-frequency neurons in inferior colliculus of the cat: effects of variations in sound pressure level on sensitivity to interaural intensity differences. *Journal of neurophysiology*, 63(3):570–591.
- JA, W. (2007). Lee cc. the distributed auditory cortex. *Hear Res*, 229:3–13.
- Joris, P., Schreiner, C., and Rees, A. (2004). Neural processing of amplitude-modulated sounds. *Physiological reviews*, 84(2):541–577.
- Julesz, B. (1962). Visual pattern discrimination. *IRE transactions on Information Theory*, 8(2):84–92.

- Julesz, B., Gilbert, E. N., and Victor, J. D. (1978). Visual discrimination of textures with identical third-order statistics. *Biological Cybernetics*, 31(3):137–140.
- Kayser, C., Petkov, C. I., and Logothetis, N. K. (2007). Tuning to sound frequency in auditory field potentials. *Journal of neurophysiology*, 98(3):1806–1809.
- Kelly, J. B. and Judge, P. W. (1994). Binaural organization of primary auditory cortex in the ferret (*mustela putorius*). *Journal of neurophysiology*, 71(3):904–913.
- Kelly, J. B., Judge, P. W., and Phillips, D. P. (1986). Representation of the cochlea in primary auditory cortex of the ferret (*mustela putorius*). *Hearing research*, 24(2):111–115.
- King, A. J. and Carlile, S. (1994). Responses of neurons in the ferret superior colliculus to the spatial location of tonal stimuli. *Hearing research*, 81(1-2):137–149.
- Krishna, B. S. and Semple, M. N. (2000). Auditory temporal processing: responses to sinusoidally amplitude-modulated tones in the inferior colliculus. *Journal of neurophysiology*, 84(1):255–273.
- Krishnan, L., Elhilali, M., and Shamma, S. (2014). Segregating complex sound sources through temporal coherence. *PLoS computational biology*, 10(12).
- Kuwada, S., Batra, R., Yin, T. C., Oliver, D. L., Haberly, L. B., and Stanford, T. R. (1997). Intracellular recordings in response to monaural and binaural stimulation of neurons in the inferior colliculus of the cat. *Journal of Neuroscience*, 17(19):7565–7581.
- Langner, G., Albert, M., and Briede, T. (2002). Temporal and spatial coding of periodicity information in the inferior colliculus of awake chinchilla (*chinchilla laniger*). *Hearing research*, 168(1-2):110–130.
- Langner, G. and Schreiner, C. E. (1988). Periodicity coding in the inferior colliculus of the cat. i. neuronal mechanisms. *Journal of neurophysiology*, 60(6):1799–1822.
- Laughlin, S. B. and Sejnowski, T. J. (2003). Communication in neuronal networks. *Science*, 301(5641):1870–1874.
- Lesica, N. A. and Grothe, B. (2008). Efficient temporal processing of naturalistic sounds. *PLoS One*, 3(2).
- Levy, W. B. and Baxter, R. A. (1996). Energy efficient neural codes. *Neural computation*, 8(3):531–543.
- Lewicki, M. S. (2002). Efficient coding of natural sounds. *Nature neuroscience*, 5(4):356–363.
- Linden, J. F., Liu, R. C., Sahani, M., Schreiner, C. E., and Merzenich, M. M. (2003). Spectrotemporal structure of receptive fields in areas ai and aaf of mouse auditory cortex. *Journal of neurophysiology*, 90(4):2660–2675.

- Liu, L.-F., Palmer, A. R., and Wallace, M. N. (2006). Phase-locked responses to pure tones in the inferior colliculus. *Journal of neurophysiology*, 95(3):1926–1935.
- Margoliash, D. (1983). Acoustic parameters underlying the responses of song-specific neurons in the white-crowned sparrow. *Journal of Neuroscience*, 3(5):1039–1057.
- Margoliash, D. (1986). Preference for autogenous song by auditory neurons in a song system nucleus of the white-crowned sparrow. *Journal of Neuroscience*, 6(6):1643–1661.
- Margrie, T. W., Brecht, M., and Sakmann, B. (2002). In vivo, low-resistance, whole-cell recordings from neurons in the anaesthetized and awake mammalian brain. *Pflügers Archiv*, 444(4):491–498.
- McAlpine, D., Jiang, D., and Palmer, A. R. (2001). A neural code for low-frequency sound localization in mammals. *Nature neuroscience*, 4(4):396–401.
- 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.
- Merzenich, M. M., Knight, P. L., and Roth, G. L. (1975). Representation of cochlea within primary auditory cortex in the cat. *Journal of neurophysiology*, 38(2):231–249.
- Middlebrooks, J. C. and Pettigrew, J. D. (1981). Functional classes of neurons in primary auditory cortex of the cat distinguished by sensitivity to sound location. *Journal of Neuroscience*, 1(1):107–120.
- Miller, L. M., Escabi, M. A., Read, H. L., and Schreiner, C. E. (2002). Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex. *Journal of neurophysiology*, 87(1):516–527.
- Mrsic-Flogel, T. D., King, A. J., Jenison, R. L., and Schnupp, J. W. (2001). Listening through different ears alters spatial response fields in ferret primary auditory cortex. *Journal of Neurophysiology*, 86(2):1043–1046.
- Nelken, I. and Bar-Yosef, O. (2008). Neurons and objects: the case of auditory cortex. *Frontiers in neuroscience*, 2:9.
- Nelken, I., Fishbach, A., Las, L., Ulanovsky, N., and Farkas, D. (2003). Primary auditory cortex of cats: feature detection or something else? *Biological cybernetics*, 89(5):397–406.
- Nelken, I., Prut, Y., Vaadia, E., and Abeles, M. (1994). In search of the best stimulus: an optimization procedure for finding efficient stimuli in the cat auditory cortex. *Hearing research*, 72(1-2):237–253.
- Nelken, I., Rotman, Y., and Yosef, O. B. (1999). Responses of auditory-cortex neurons to structural features of natural sounds. *Nature*, 397(6715):154–157.
- Nelken, I. and Ulanovsky, N. (2007). Mismatch negativity and stimulus-specific adaptation in animal models. *Journal of Psychophysiology*, 21(3-4):214–223.

- Newman, J. D. and Wollberg, Z. (1973). Multiple coding of species-specific vocalizations in the auditory cortex of squirrel monkeys. *Brain research*, 54:287–304.
- Ohlemiller, K. K., Kanwal, J. S., and Suga, N. (1996). Facilitative responses to species-specific calls in cortical fm-fm neurons of the mustached bat. *Neuroreport*, 7(11):1749–1755.
- Olshausen, B. A. and Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583):607–609.
- Olshausen, B. A. and Field, D. J. (1997). Sparse coding with an overcomplete basis set: A strategy employed by v1? *Vision research*, 37(23):3311–3325.
- Olshausen, B. A. and Field, D. J. (2004). Sparse coding of sensory inputs. *Current opinion in neurobiology*, 14(4):481–487.
- Pelleg-Toiba, R. and Wollberg, Z. (1989). Tuning properties of auditory cortex cells in the awake squirrel monkey. *Experimental brain research*, 74(2):353–364.
- Perez-Orive, J., Mazor, O., Turner, G. C., Cassenaer, S., Wilson, R. I., and Laurent, G. (2002). Oscillations and sparsening of odor representations in the mushroom body. *Science*, 297(5580):359–365.
- Pfeiffer, R. R. (1966). Classification of response patterns of spike discharges for units in the cochlear nucleus: tone-burst stimulation. *Experimental Brain Research*, 1(3):220–235.
- Phillips, D. and Irvine, D. (1981). Responses of single neurons in physiologically defined area ai of cat cerebral cortex: sensitivity to interaural intensity differences. *Hearing research*, 4(3-4):299–307.
- Polley, D. B., Read, H. L., Storace, D. A., and Merzenich, M. M. (2007). Multiparametric auditory receptive field organization across five cortical fields in the albino rat. *Journal of neurophysiology*, 97(5):3621–3638.
- Portilla, J. and Simoncelli, E. P. (2000). A parametric texture model based on joint statistics of complex wavelet coefficients. *International journal of computer vision*, 40(1):49–70.
- Rabinowitz, N. C., Willmore, B. D., King, A. J., and Schnupp, J. W. (2013). Constructing noise-invariant representations of sound in the auditory pathway. *PLoS biology*, 11(11):e1001710.
- Rabinowitz, N. C., Willmore, B. D., Schnupp, J. W., and King, A. J. (2011). Contrast gain control in auditory cortex. *Neuron*, 70(6):1178–1191.
- Rabinowitz, N. C., Willmore, B. D., Schnupp, J. W., and King, A. J. (2012). Spectrotemporal contrast kernels for neurons in primary auditory cortex. *Journal of Neuroscience*, 32(33):11271–11284.

- Ramachandran, R., Davis, K. A., and May, B. J. (1999). Single-unit responses in the inferior colliculus of decerebrate cats i. classification based on frequency response maps. *Journal of neurophysiology*, 82(1):152–163.
- Rauschecker, J. P., Tian, B., and Hauser, M. (1995). Processing of complex sounds in the macaque nonprimary auditory cortex. *Science*, 268(5207):111–114.
- Recanzone, G. H. (2000). Response profiles of auditory cortical neurons to tones and noise in behaving macaque monkeys. *Hearing research*, 150(1-2):104–118.
- Recanzone, G. H., Schreiner, C. E., Sutter, M. L., Beitel, R. E., and Merzenich, M. M. (1999). Functional organization of spectral receptive fields in the primary auditory cortex of the owl monkey. *Journal of Comparative Neurology*, 415(4):460–481.
- Rees, A. and Møller, A. R. (1983). Responses of neurons in the inferior colliculus of the rat to am and fm tones. *Hearing research*, 10(3):301–330.
- Rees, A. and Møller, A. R. (1987). Stimulus properties influencing the responses of inferior colliculus neurons to amplitude-modulated sounds. *Hearing research*, 27(2):129–143.
- Rieke, F., Bodnar, D., and Bialek, W. (1995). Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory afferents. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 262(1365):259–265.
- Rinberg, D., Koulakov, A., and Gelperin, A. (2006). Sparse odor coding in awake behaving mice. *Journal of Neuroscience*, 26(34):8857–8865.
- Robles, L. and Ruggero, M. A. (2001). Mechanics of the mammalian cochlea. *Physiological reviews*, 81(3):1305–1352.
- Rodríguez, F. A., Chen, C., Read, H. L., and Escabí, M. A. (2010). Neural modulation tuning characteristics scale to efficiently encode natural sound statistics. *Journal of Neuroscience*, 30(47):15969–15980.
- Rosenthal, D. F. and Okuno, H. G. (1998). *Computational auditory scene analysis*. Lawrence Erlbaum Associates Publishers.
- Sachs, M. and Blackburn, C. (1991). Processing of complex sounds in the cochlear nucleus. *Neurophysiology of Hearing: The Central Auditory System*.
- Sally, S. L. and Kelly, J. B. (1988). Organization of auditory cortex in the albino rat: sound frequency. *Journal of neurophysiology*, 59(5):1627–1638.
- Sanes, D. H. (1990). An in vitro analysis of sound localization mechanisms in the gerbil lateral superior olive. *Journal of Neuroscience*, 10(11):3494–3506.
- Scheich, H., Langner, G., and Bonke, D. (1979). Responsiveness of units in the auditory neostriatum of the guinea fowl (*numida meleagris*) to species-specific calls and synthetic stimuli. *Journal of comparative physiology*, 132(3):257–276.

- Schnupp, J. W., 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*, 9:37.
- Schnupp, J. W., Mrsic-Flogel, T. D., and King, A. J. (2001). Linear processing of spatial cues in primary auditory cortex. *Nature*, 414(6860):200–204.
- Schreiner, C., Urbas, J., and Mehrgardt, S. (1938). Temporal resolution of amplitude modulation and complex signals in the auditory cortex of the cat. In *Hearing—Physiological Bases and Psychophysics*, pages 169–175. Springer.
- Schreiner, C. E. and Langner, G. (1988). Periodicity coding in the inferior colliculus of the cat. ii. topographical organization. *Journal of neurophysiology*, 60(6):1823–1840.
- Schreiner, C. E. and Mendelson, J. R. (1990). Functional topography of cat primary auditory cortex: distribution of integrated excitation. *Journal of neurophysiology*, 64(5):1442–1459.
- Schreiner, C. E. and Merzenich, M. M. (1988). Elements of signal coding in the auditory nervous system. In *Organization of neural networks: structures and models*. VCH Weinheim.
- 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, 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, C. E., Mehta, A. D., and Givre, S. J. (1998). A spatiotemporal profile of visual system activation revealed by current source density analysis in the awake macaque. *Cerebral cortex (New York, NY: 1991)*, 8(7):575–592.
- Schulze, H., Hess, A., Ohl, F. W., and Scheich, H. (2002). Superposition of horseshoe-like periodicity and linear tonotopic maps in auditory cortex of the mongolian gerbil. *European Journal of Neuroscience*, 15(6):1077–1084.
- Schulze, H. and Langner, G. (1997). Periodicity coding in the primary auditory cortex of the mongolian gerbil (*Meriones unguiculatus*): two different coding strategies for pitch and rhythm? *Journal of Comparative Physiology A*, 181(6):651–663.
- Schwarz, D. W. and Tomlinson, R. W. (1990). Spectral response patterns of auditory cortex neurons to harmonic complex tones in alert monkey (*Macaca mulatta*). *Journal of Neurophysiology*, 64(1):282–298.
- Shamir, M. and Sompolinsky, H. (2004). Nonlinear population codes. *Neural computation*, 16(6):1105–1136.

- Shamma, S. A., Fleshman, J. W., Wiser, P. R., and Versnel, H. (1993). Organization of response areas in ferret primary auditory cortex. *Journal of neurophysiology*, 69(2):367–383.
- 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.
- Singer, Y., Teramoto, Y., Willmore, B. D., Schnupp, J. W., King, A. J., and Harper, N. S. (2018). Sensory cortex is optimized for prediction of future input. *Elife*, 7:e31557.
- Singh, N. C. and Theunissen, F. E. (2003). Modulation spectra of natural sounds and ethological theories of auditory processing. *The Journal of the Acoustical Society of America*, 114(6):3394–3411.
- Smith, J. O. (2007). *Introduction to digital filters: with audio applications*, volume 2. Julius Smith.
- Smith, Z. M., Delgutte, B., and Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature*, 416(6876):87–90.
- Sompolinsky, H., Yoon, H., Kang, K., and Shamir, M. (2001). Population coding in neuronal systems with correlated noise. *Physical Review E*, 64(5):051904.
- Steinschneider, M., Arezzo, J. C., and Vaughan Jr, H. G. (1990). Tonotopic features of speech-evoked activity in primate auditory cortex. *Brain research*, 519(1-2):158–168.
- Suga, N., O’Neill, W. E., and Manabe, T. (1978). Cortical neurons sensitive to combinations of information-bearing elements of biosonar signals in the mustache bat. *Science*, 200(4343):778–781.
- Szyszka, P., Ditzen, M., Galkin, A., Galizia, C. G., and Menzel, R. (2005). Sparsening and temporal sharpening of olfactory representations in the honeybee mushroom bodies. *Journal of neurophysiology*, 94(5):3303–3313.
- Theunissen, F. E., Sen, K., and Doupe, A. J. (2000). Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *Journal of Neuroscience*, 20(6):2315–2331.
- Ulanovsky, N., Las, L., and Nelken, I. (2003). Processing of low-probability sounds by cortical neurons. *Nature neuroscience*, 6(4):391–398.
- Vinje, W. E. and Gallant, J. L. (2000). Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*, 287(5456):1273–1276.
- Voss, R. F. and Clarke, J. (1975). ‘1/f noise’ in music and speech. *Nature*, 258(5533):317–318.
- Voytenko, S. and Galazyuk, A. (2007). Intracellular recording reveals temporal integration in inferior colliculus neurons of awake bats. *Journal of neurophysiology*, 97(2):1368–1378.

References

- Wallace, M. N., Shackleton, T. M., and Palmer, A. R. (2002). Phase-locked responses to pure tones in the primary auditory cortex. *Hearing research*, 172(1-2):160–171.
- Whitfield, I. and Evans, E. (1965). Responses of auditory cortical neurons to stimuli of changing frequency. *Journal of Neurophysiology*, 28(4):655–672.
- Willmore, B. D., Schoppe, O., King, A. J., Schnupp, J. W., and Harper, N. S. (2016). Incorporating midbrain adaptation to mean sound level improves models of auditory cortical processing. *Journal of Neuroscience*, 36(2):280–289.
- Winkler, I. and Schröger, E. (2015). Auditory perceptual objects as generative models: Setting the stage for communication by sound. *Brain and language*, 148:1–22.
- Winter, I. M., Wiegrebe, L., and Patterson, R. D. (2001). The temporal representation of the delay of iterated rippled noise in the ventral cochlear nucleus of the guinea-pig. *The Journal of physiology*, 537(Pt 2):553.
- Woolley, S. M. and Casseday, J. H. (2005). Processing of modulated sounds in the zebra finch auditory midbrain: responses to noise, frequency sweeps, and sinusoidal amplitude modulations. *Journal of Neurophysiology*, 94(2):1143–1157.
- Yeshurun, Y., Wollberg, Z., Dyn, N., and Allon, N. (1985). Identification of mgb cells by volterra kernels. *Biological Cybernetics*, 51(6):383–390.
- Young, E. D. and Brownell, W. E. (1976). Responses to tones and noise of single cells in dorsal cochlear nucleus of unanesthetized cats. *Journal of Neurophysiology*, 39(2):282–300.
- Zheng, Y. and Escabí, M. A. (2008). Distinct roles for onset and sustained activity in the neuronal code for temporal periodicity and acoustic envelope shape. *Journal of Neuroscience*, 28(52):14230–14244.
- Zhu, S. C., Wu, Y. N., and Mumford, D. (1997). Minimax entropy principle and its application to texture modeling. *Neural computation*, 9(8):1627–1660.

Appendix A

Publications

(I.) Journals:

- **Mishra, A. P., Harper, N.S., Schnupp, J.W. Exploring the Distribution of Statistical Feature Parameters for Natural Sound Textures.**
(doi: <https://doi.org/10.1101/2020.08.28.271528>).

(II.) Conferences:

- Meng, Q., Zheng, N., **Mishra, A. P.**, Luo, J. D., Schnupp, J. W. (2018, September). **Weighting Pitch Contour and Loudness Contour in Mandarin Tone Perception in Cochlear Implant Listeners.** In Interspeech (pp. 3768-3771).
- Roßkothen-Kuhl, N., Buck, AN, Li, K., **Mishra, A. P.**, Schnupp, JW (2018). **Sensitivity to interaural time differences in a new animal model for**

bilateral cochlear implant users. Laryngo-Rhino-Otologie , 97 (S 02), 10138.

- Li, K., Chloe H. K. Chan, **Mishra, A. P.**, Schnupp, J. W. **Temporal weighting functions for binaural cues in rats.** 50th Annual Society for Neuroscience's Annual Meeting (SFN 2019).
- Roßkothen-Kuhl, N., Buck, AN, Li, K., **Mishra, A. P.**, Schnupp, J. W. **Behavioral sensitivity for interaural time differences in normal hearing rats and rats with cochlear implants.** 41st Annual Association for Research in Otolaryngology MidWinter Meeting.
- Schnupp, J. W., Buck, AN,, Li, K., **Mishra, A. P.**, Roßkothen-Kuhl, N., **Interaural time difference sensitivity in the inferior colliculus of neonatally deafened rats after cochlear implantation .** 41st Annual Association for Research in Otolaryngology MidWinter Meeting.

(III.) Podium Presentation:

- Peng, F ,**Mishra, A. P.**, Harper, N.S., Schnupp, J. W. **Midbrain and Cortical Responses to Natural Sound textures.** Podium presentation, 43rd Annual Midwinter meeting San Jose, California, ARO-2020.

(IV.) **Posters:**

- Peng, F, **Mishra, A. P.**, Harper, N.S., Schnupp, J. W. **Neural Sensitivity to the Statistics of Natural Sound Textures in Rats.** International Symposium on auditory and audiological research (ISAAR-2019), Denmark, 2019.
- **Mishra, A. P.**, Peng, F., Harper, N.S., Schnupp, J. W. Are neurons in the central nervous system sensitive to statistical cues for auditory texture representation? BMS, Research Gala, 2019.
- **Mishra, A. P.**, Harper, N.S., Schnupp, J. W. Low dimensional auditory texture representation and possible impact on auditory scene analysis. Neuroplasticity of Sensory Systems, Gordon Research Conference, 2018.
- **Mishra, A. P.**, Harper, N.S., Schnupp, J. W. **Low dimensional auditory texture representation and possible impact on auditory scene analysis.** Neuroplasticity of Sensory Systems, BMS, Research Gala, 2018.

Appendix B

List of Sound Textures

Sound ID	File name
1	African goosescalls
2	Alarm clock beeping
3	Aluminum foil crumple1
4	Ambience bar restaurant
5	Angry cat roo
6	Chimpanzee heavy fast panting
7	Chimpanzee panting crying
8	applaus1
9	applause_church
10	applause_crowd
11	Arctic fox calls
12	Background shooting1
13	Bald eagle calls

-
- 14 Barking dog
 - 15 Barn swallow calls
 - 16 Bear growl
 - 17 Bees insect tree
 - 18 Belted kingfisher calls
 - 19 Bicycle pump2
 - 20 Bigfly
 - 21 Bike start
 - 22 Blackbird blackforest
 - 23 Blender1
 - 24 Blender2
 - 25 Bongos
 - 26 River
 - 27 Brown headed cowbird
 - 28 Brushing teeth
 - 29 Brushwood fire1
 - 30 Bubbles1
 - 31 Bumble bees blossom

-
- 32 Bumblebee against window
 - 33 Cackling geese1
 - 34 Camel groaning moaning
 - 35 campfire
 - 36 Canary song calls
 - 37 Car difficult start
 - 38 Cardinal song
 - 39 Castanet1
 - 40 Castanet2
 - 41 Castanets3
 - 42 Cathedral1
 - 43 Cat meow
 - 44 Chair
 - 45 Chik chirp
 - 46 Chimps
 - 47 Christmas bells
 - 48 Church bell1
 - 49 Church bell3

-
- 50 Church bell4
 - 51 Church bells2
 - 52 Cloth brush1
 - 53 Clucking chickens crowing rooster
 - 54 Coins pouring2
 - 55 Computer keyboard typing
 - 56 Computer scanner scanning
 - 57 Constant ringing sleigh bells
 - 58 Counting bills in hand
 - 59 Cricket chirping
 - 60 Crickets in woods
 - 61 Crow
 - 62 Crunching food2
 - 63 Crunchy paper1
 - 64 Damaged muffler car
 - 65 Dog barking night
 - 66 Dog whining1
 - 67 Donkey
 - 68 Door creaking
 - 69 Duck quack

-
- 70 Electric sewing machine
 - 71 Elk
 - 72 Excited chickens
 - 73 Bugle music from trumpet
 - 74 Faucet kitchen
 - 75 Faucet leaking water
 - 76 Filling sink with water
 - 77 Fire outside woods sticks
 - 78 Fireplace
 - 79 Fireworks1
 - 80 Fireworks3
 - 81 Footsteps running on road1
 - 82 Footsteps walking in water1
 - 83 Forest fire3
 - 84 Free tailbat calls
 - 85 Frogs at night
 - 86 Toads
 - 87 Frogs and crickets

-
- 88 Funky bongos
 - 89 Gargles1
 - 90 Shaving electronicrazor
 - 91 Glassdebris sweep2
 - 92 Goats
 - 93 Gorilla roars breaths
 - 94 Grunting angry pig
 - 95 gunshoot
 - 96 Hair brushing2
 - 97 Hammer pounding on wood
 - 98 Hand broom1
 - 99 Hand washing dishes1
 - 100 Hand washing dishes2
 - 101 Handling paper
 - 102 Heavy rain1
 - 103 Heavy rain2
 - 104 Hedgehog
 - 105 Hitting metal hammering
 - 106 Horse1

-
- 107 Horse galloping
 - 108 Horse whinnying1
 - 109 Bats nest
 - 110 Jedspear mowing lawn_mower
 - 111 Jigsaw
 - 112 Keys jingling1
 - 113 Keys jingling2
 - 114 clock ticks
 - 115 Knife sharpening
 - 116 Knife stoe sharpen
 - 117 Lake waves
 - 118 Lake waves2
 - 119 Lapping waves
 - 120 Large river rushing
 - 121 Lathe2
 - 122 Lawn mower

-
- 123 Lion growling
 - 124 Lion grunting
 - 125 Madbear1
 - 126 Shaving
 - 127 Metal chimes
 - 128 Mosquito buzzing2
 - 129 Mosquito buzzing1
 - 130 Motorbike idling
 - 131 News paper torn
 - 132 Xylophone
 - 133 Ocean waves crushing
 - 134 Parrots squawking1
 - 135 Peeing toilet1
 - 136 Peeler peeling apple1
 - 137 Pigeons squabs
 - 138 Police3

-
- 139 Pop corn popping1
 - 140 Posh dinner party
 - 141 Radio static3
 - 142 Rail crossing
 - 143 Rain hitting metal
 - 144 Rain thunder
 - 145 Rainy day
 - 146 Raking leaves into a pile
 - 147 Rattle1
 - 148 Rattle2
 - 149 Raven calls
 - 150 Red bellied woodpecker calls
 - 151 Red fox calls
 - 152 Red tail hawk screams
 - 153 Restaurant ambience1
 - 154 Tabla


-
- 155 Ringdove calling
 - 156 River flowing fast
 - 157 Robin calls
 - 158 Rooster crowing
 - 159 Sailing boat bow
 - 160 Scratching skin3
 - 161 Screech owl
 - 162 Sea waves gargl gully
 - 163 sea at night
 - 164 Seashore1
 - 165 Siren from inside ambulance
 - 166 Slow pour of sand in plastic bucket
 - 167 Slow shakes of sleigh bells
 - 168 Small bulldozer engine
 - 169 Snoring man2
 - 170 Squake rtoy

-
- 171 Steam4
 - 172 Stirring liquid in glass
 - 173 Stream1
 - 174 sweeping_up_broken_glass
 - 175 Tapping1
 - 176 Tea kettle whistle
 - 177 Thunderstorm
 - 178 Tibetan chant
 - 179 Tin can
 - 180 Toddler babble3
 - 181 Toilet1
 - 182 Toilet flush
 - 183 Toothbrush
 - 184 Tree frogs
 - 185 Laughing gull_birds calling
 - 186 Typing1
 - 187 Typing2

-
- 188 Vacuum cleaner1
 - 189 Walking through and splashing in water
 - 190 Water boiling strong
 - 191 Water lapping river
 - 192 Water dripping
 - 193 Water running1
 - 194 Waves against shore1
 - 195 Waves at beach cliff
 - 196 Wheel lathe
 - 197 Wild turkey gobbles
 - 198 white board marker
 - 199 Toddler babble2
 - 200 heavyrain

Appendix C

Figure Permission



Thank you for your order!

Dear Mr. Ambika Mishra,

Thank you for placing your order through Copyright Clearance Center's RightsLink® service.

Order Summary

Licensee: Ambika Pr Mishra
Order Date: Apr 10, 2020
Order Number: 4805170377774
Publication: Neuron
Title: Sound Texture Perception via Statistics of the Auditory Periphery:
Evidence from Sound Synthesis
Type of Use: reuse in a thesis/dissertation
Order Total: 0.00 USD

View or print complete [details](#) of your order and the publisher's terms and conditions.

Sincerely,

Copyright Clearance Center

Tel: +1-855-239-3415 / +1-978-646-2777
customerare@copyright.com
<https://myaccount.copyright.com>

