

Contrast gain control in the central auditory system

Neil Rabinowitz

Merton College

University of Oxford

A thesis submitted for the degree of Doctor of Philosophy

Michaelmas 2011

Abstract

The auditory system must represent sounds with a wide range of statistical properties. One important property is the spectrotemporal contrast in the acoustic environment. The level of some sounds varies only a little over frequency and time, while the level of other sounds can vary a lot. This raises a *contrast problem* for neural coding: auditory neurons, with a limited dynamic range of firing rates, must be able to efficiently encode the sound level fluctuations in both low and high contrast sounds. In this thesis, I show how the contrast problem is solved in the primary areas of the ferret auditory cortex (A1/AAF).

One hypothesis is that different neurons specialise for representing sounds of different contrasts. I find little evidence for such specialisation in the auditory cortex. Rather, the system adapts its coding to operate under different contrast conditions. I demonstrate that neurons in A1/AAF rescale their gain to partially compensate for the spectrotemporal contrast of recent stimulation. When contrast is low, neurons increase their gain, becoming more sensitive to small changes in the stimulus, without changing their tuning. I quantify these gain changes and find that they resemble divisive normalisation, a phenomenon observed in many other neural systems. In a given stimulus context, auditory cortical neurons determine their gain predominantly on the basis of spectrotemporally local statistics. By comparing neural responses in A1/AAF with those in the inferior colliculus (IC), I show that this adaptive strategy of the auditory cortex is not simply inherited from the IC. When stimulus contrast changes, IC neurons undergo a variety of different changes in coding, which are on average weaker than in cortex.

Together, these results suggest that the auditory cortex attempts to divide out contrast in its representation of an ongoing stimulus. This appears to be a novel property of the higher auditory pathway.

Contrast gain control in the central auditory system



Neil Rabinowitz

Merton College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2011

To ignorance:

boo unto you

Acknowledgements

There's a certain irony in beginning a thesis with the statement that a D.Phil. is so much more than a thesis. I appreciate irony. It reminds me that a D.Phil. is so much more than a thesis.

Several years ago, I made the rather abrupt and curious decision to start sticking my hands in brains. For someone who kinda knew how to wield some numbers, but had never done a real experiment before, this required a major shift in thinking. It would not have been so easy if it were not for the support and encouragement of people around me.

Professors Andy and Jan: thank you for making my initial decision to join the lab the right one. The environment you have fostered here has allowed me to flourish. My deepest thanks for all your advice, inspiration, and enthusiasm.

To Ben Willmore, who has occupied a role for me somewhere halfway between older brother and mentor. Our circuits in the park have injected clarity when it was needed most. Thank you for calling me out on ideas that didn't make sense. Thank you for encouraging me on the ones that did. In many ways, your name deserves to be on my official supervisor list. You should nevertheless thank me for saving you the administrative hassle.

The folk here in the Auditory Neuroscience Group have made every day joyous. A special thanks to Sandra Tolnai, not only for your moral support and friendship, but also for transcending my coding-induced hypnotic state with precisely timed cups of tea.

I am grateful to Merton College for keeping me above the poverty line while I pursued this endeavour.

And to my wonderful Jo, who travelled around the world and back again for me.

Statement of Originality

In vivo experiments were conducted in collaboration with Ben Willmore and Sandra Tolnai, who performed the surgery. The remainder of this thesis is entirely my own work. A major portion of the results of Chapter 4 and of Sections 5.2 and 5.4 were written up for publication in *Neuron* together with Ben Willmore, Jan Schnupp and Andrew King.

Abstract

The auditory system must represent sounds with a wide range of statistical properties. One important property is the spectrotemporal contrast in the acoustic environment. The level of some sounds varies only a little over frequency and time, while the level of other sounds can vary a lot. This raises a *contrast problem* for neural coding: auditory neurons, with a limited dynamic range of firing rates, must be able to efficiently encode the sound level fluctuations in both low and high contrast sounds. In this thesis, I show how the contrast problem is solved in the primary areas of the ferret auditory cortex (A1/AAF).

One hypothesis is that different neurons specialise for representing sounds of different contrasts. I find little evidence for such specialisation in the auditory cortex. Rather, the system adapts its coding to operate under different contrast conditions. I demonstrate that neurons in A1/AAF rescale their gain to partially compensate for the spectrotemporal contrast of recent stimulation. When contrast is low, neurons increase their gain, becoming more sensitive to small changes in the stimulus, without changing their tuning. I quantify these gain changes and find that they resemble divisive normalisation, a phenomenon observed in many other neural systems. In a given stimulus context, auditory cortical neurons determine their gain predominantly on the basis of spectrotemporally local statistics. By comparing neural responses in A1/AAF with those in the inferior colliculus (IC), I show that this adaptive strategy of the auditory cortex is not simply inherited from the IC. When stimulus contrast changes, IC neurons undergo a variety of different changes in coding, which are on average weaker than in cortex.

Together, these results suggest that the auditory cortex attempts to divide out contrast in its representation of an ongoing stimulus. This appears to be a novel property of the higher auditory pathway.

List of Abbreviations

A1	Primary Auditory Cortex
AAF	Anterior Auditory Field
AIC	Akaike Information Criterion
AN	Auditory Nerve
ATS	Anti-Test Sound
BIC	Bayesian Information Criterion
BF	Best Frequency
CN	Cochlear Nucleus
CoM	Centre of Mass
CRF	Context Receptive Field
DRC	Dynamic Random Chord
EM	Expectation Maximisation
GLM	Generalised Linear Model
HPR	High Probability Region
IC	Inferior Colliculus
LGN	Lateral Geniculate Nucleus
LN	Linear-Nonlinear
LSO	Lateral Superior Olive
MAP	Maximum A Posteriori
MCMC	Markov Chain Monte Carlo
MD	Monotonically Decreasing
MGB	Medial Geniculate Body
MI	Monotonically Increasing
MSO	Medial Superior Olive
MT	Middle Temporal
MTF	Modulation Transfer Function
NM	Non-Monotonic
NP	Noise Power
NR	Noise Ratio
PCA	Principal Component Analysis
PSTH	Peri-Stimulus Time Histogram
RC-DRC	Random Contrast Dynamic Random Chord
RGC	Retinal Ganglion Cell
RMS	Root Mean Square
RSS	Random Spectral Stimuli
SAM	Sinusoidally Amplitude-Modulated
SP	Signal Power
SPE	Signal Power Explained
SPL	Sound Pressure Level
SRF	Spectral Receptive Field
STRF	Spectrotemporal Receptive Field
TS	Test Sound
UK	The United Kingdom of Great Britain and Northern Ireland
VE	Variance Explained

Contents

List of Figures	x

List of Tables	xv

1 Introduction	1

1.1 Two problems for hearing	1

1.2 The intensity problem	4

1.3 Contrast gain control in the visual system	6

1.4 The contrast problem in the auditory system	12

1.4.1 Evidence for the division of labour strategy	13

1.4.2 Evidence for the adaptation strategy	15

1.5 The auditory cortex	20

1.6 Open questions	21

1.7 Structure of the thesis	23

2 Theory and methods	25

2.1 Stimuli for model fitting	25

2.1.1 DRCs	25

2.1.2 Previous use of DRCs	28

2.1.3 DRCs with variable contrast	30

2.1.3.1	Controlling the statistics of DRCs	30
2.1.3.2	Relationship between pressure contrast and level standard deviation	32
2.1.3.3	Spectral and temporal contrast	34
2.2	Models of cortical neurons	37
2.2.1	Linear models: STRFs	37
2.2.1.1	Separable STRFs	39
2.2.2	Linear-nonlinear (LN) models	41
2.2.3	Other extensions to STRF models	42
2.3	Experimental methods	44
2.3.1	Introduction	44
2.3.2	Surgery and recording	45
2.3.2.1	Cortex (anaesthetised)	45
2.3.2.2	Cortex (awake)	46
2.3.2.3	IC (anaesthetised)	46
2.3.3	Stimulus presentation	47
2.3.3.1	Anaesthetised	47
2.3.3.2	Awake	47
2.3.4	Spike sorting	48
2.3.5	Unit selection criteria	50
2.3.5.1	Acoustically-responsive activity	50
2.3.5.2	Noise ratio criterion	50

3	Contrast tuning	52
3.1	Introduction	52
3.2	Methods	55
3.2.1	Stimuli	55
3.2.1.1	DRC stimuli	55
3.2.1.2	Test sound stimuli	56
3.2.2	Response metrics	58
3.2.3	Unit selection criteria	58
3.2.4	Models	59
3.2.5	Model comparisons	61
3.3	Results	62
3.3.1	Mean firing rate and stimulus contrast	63
3.3.2	Signal power and stimulus contrast	64
3.3.3	Transient stimuli	67
3.4	Discussion	69
3.4.1	Comparison with previous studies	70
3.4.2	The effect of total stimulus level	74
4	Contrast gain control in auditory cortex	75
4.1	Introduction	75
4.2	Methods	78
4.2.1	Animals and recordings	78
4.2.2	DRC Stimuli	78
4.2.3	Unit selection criteria	79
4.2.4	Linear models of unit responses: STRF estimation	81

4.2.5	Nonlinear models of unit responses: LN model estimation	82
4.2.6	Curve transformations	83
4.2.7	Fisher information	85
4.2.8	Normalisation models	86
4.3	Results	86
4.3.1	Initial observations of gain changes	88
4.3.1.1	An elementary measure of gain	88
4.3.1.2	Caveats	90
4.3.2	Spectrotemporal receptive fields	91
4.3.3	Stimulus contrast has no effect on tuning	92
4.3.4	Quantifying gain changes: linear-nonlinear models	95
4.3.4.1	Coding accuracy	99
4.3.4.2	Control experiment: total sound level	100
4.3.5	Contrast gain control	102
4.3.6	Dependence on mean tone level	103
4.4	Discussion	108
4.4.1	The role of contrast gain control	109
4.4.2	Gain control vs. division of labour	110
4.4.3	Predictive models for auditory neurons	111
5	Spectral integration in contrast gain control	113
5.1	Introduction	113
5.2	Spectral Experiment 1: Local and global contributions to gain	116
5.2.1	Methods	116
5.2.1.1	Animals and recordings	116

5.2.1.2	Stimuli	116
5.2.1.3	STRF and LN models	119
5.2.1.4	Overlap between the test band and the SRF . .	120
5.2.1.5	Unit selection criteria	121
5.2.1.6	Normalisation models	121
5.2.2	Results	122
5.2.2.1	Contributions to gain from within and outside the tuning curve	123
5.2.2.2	Measuring the relative contributions	125
5.2.3	Preliminary Discussion	129
5.3	Contrast kernel models	130
5.3.1	Introduction	130
5.3.2	Contrast kernels	132
5.3.2.1	Parameters of output nonlinearities	132
5.3.2.2	Contrast kernels	133
5.3.2.3	Consistency with previous models	136
5.3.2.4	Simplifications of contrast kernel models	137
5.3.2.5	Extensions of contrast kernel models	139
5.3.3	Random Contrast DRCs	141
5.4	Spectral Experiment 2: Contrast frequency kernels	143
5.4.1	Methods	143
5.4.1.1	Animals	143
5.4.1.2	Stimuli	144
5.4.1.3	Unit selection criteria	145

5.4.1.4	Model structure: contrast kernels	146
5.4.1.5	Model structure: cross-frequency interactions	147
5.4.1.6	Model fitting	149
5.4.1.7	Bayesian methods	151
5.4.1.8	Priors on nonlinearity parameters	153
5.4.1.9	Priors on contrast kernels	154
5.4.1.10	Model success	155
5.4.2	Results	158
5.4.2.1	STRF and LN models	158
5.4.2.2	Contrast produces stereotyped changes to out- put nonlinearities	160
5.4.2.3	Gain contrast kernels	160
5.4.2.4	Model success	164
5.4.2.5	Negative coefficients of gain contrast kernels	165
5.4.2.6	Multiple contrast kernels	170
5.4.2.7	Cross-frequency interactions	171
5.5	Discussion	172
5.5.1	Local vs. remote spectral contributions to the gain	174
5.5.2	Mechanisms for gain control	175
5.5.3	Implications for modelling	179
5.5.4	Computational principles	181

6	Temporal integration in contrast gain control	182
6.1	Introduction	182
6.2	Temporal Experiment 1: Fixed stimulus in different contexts . .	184
6.2.1	Methods	184
6.2.1.1	Animals and recordings	184
6.2.1.2	Stimuli	184
6.2.1.3	Unit selection criteria	185
6.2.1.4	Responses to the test sound	186
6.2.2	Results	186
6.2.2.1	Time course of gain changes	187
6.2.3	Preliminary Discussion	189
6.3	Contrast kernel models	190
6.3.1	Contrast time kernels	190
6.3.2	Simplifications	193
6.4	Temporal Experiment 2: Contrast time kernels	193
6.4.1	Methods	193
6.4.1.1	Animals and stimuli	193
6.4.1.2	Contrast time kernels	194
6.4.1.3	Model success	195
6.4.2	Results	195
6.4.2.1	Contrast time kernels	196
6.4.2.2	Contrast time kernels: simplifications	198
6.5	Discussion	201
6.5.1	Mechanisms for gain control	201

6.5.2	Implications for modelling	203
6.5.3	Computational principles	204
7	Contrast gain control in the inferior colliculus?	207
7.1	Introduction	207
7.2	Methods	209
7.2.1	Animals and stimuli	209
7.2.2	Models	210
7.2.3	Fisher information	210
7.3	Results	211
7.3.1	How contrast affects IC nonlinearities	213
7.3.2	Gain contrast kernels	214
7.3.3	Multiple contrast kernels	217
7.3.4	Coding accuracy	221
7.3.5	Cross-frequency interactions	222
7.3.6	Gain time kernels	223
7.4	Discussion	225
7.4.1	Stimulus representations in the IC and A1/AAF	227
8	Conclusions	229
8.1	The contrast problem	229
8.2	Comparing the auditory with the visual	230
8.3	Adaptation and gain control	231
8.4	From synthetic to natural sounds	232
8.5	Further experiments	235

Appendices

A Contrast kernel models: derivatives	238
A.1 Model parameters	238
A.2 Log likelihood	240
A.3 Log posterior	241
B Cross-frequency kernel models: derivatives	244
B.1 Model parameters	244
B.2 Global effect model	246
B.3 Local effect model	247
B.3.1 Derivatives	247
B.4 Regional effect model	249
References	252

List of Figures

1.1	The intensity problem and the contrast problem	3
1.2	The contrast problem for neural coding	4
1.3	The division of labour solution to the intensity problem in the auditory nerve	7
1.4	The division of labour solution to the contrast problem	8
1.5	Contrast gain control in the visual system	10
1.6	The adaptation solution to the contrast problem	12
2.1	Example of a DRC stimulus	26
2.2	Spectral and temporal cross-sections of a DRC	27
2.3	Varying the widths of DRC level distributions	31
2.4	Relationship between level statistics and sound pressure statistics	35
2.5	Tensor representation of a DRC	39

2.6	Schematic of the LN model	42
3.1	Schematic of test sound construction	57
3.2	Idealised contrast-response relationships	60
3.3	Mean firing rate as a function of contrast	65
3.4	Signal power as a function of contrast	66
3.5	Response to test sound as a function of contrast	68
4.1	Output nonlinearities for two example units	83
4.2	Evidence of gain control in the relationship between contrast and signal power	89
4.3	Most units show gain control	90
4.4	Effect of stimulus contrast on tuning: example	92
4.5	Effect of stimulus contrast on tuning: statistics	94
4.6	Schematic of the contrast-dependent LN model	95
4.7	Effects of stimulus contrast on output nonlinearities: examples .	97
4.8	Effects of stimulus contrast on output nonlinearities: statistics .	98
4.9	Effects of stimulus contrast on output nonlinearities: additional statistics	98
4.10	Effects of stimulus contrast on coding: Fisher information	100

4.11	Gain as a function of contrast: population	103
4.12	Gain as a function of σ_L : individual units	104
4.13	Contrast-dependent changes in threshold and saturation	105
4.14	Contrast gain control is modulated by μ_L only at low mean levels	106
4.15	Output nonlinearities for two example units across different μ_L , σ_L conditions	107
4.16	Schematic of gain control model	108
4.17	Contrast gain control expressed as a function of c	108
5.1	Schematic of test/mask DRC stimuli	118
5.2	Changes in gain from varying σ_{test} and σ_{mask}	124
5.3	Summary of changes in gain from varying σ_{test} and σ_{mask}	125
5.4	Parameters of the sigmoidal static nonlinearity	133
5.5	Illustration of the contrast kernel model	135
5.6	Illustration of cross-contrast kernels	140
5.7	Schematic of Random Contrast DRC stimuli	143
5.8	Uncertainty in estimates of static nonlinearity parameters	150
5.9	Measuring model success	157
5.10	Model predictive power: STRF and LN models	159

5.11	Output nonlinearities in low and high contrast: A1/AAF	161
5.12	Gain contrast kernels: examples	163
5.13	Correlation between linear (STRF) kernel and gain (contrast) kernel coefficients	164
5.14	Model predictive power: Gain model	165
5.15	Gain model: statistics on contrast-dependent gain changes . . .	166
5.16	Contrast gain kernels: examples when $\kappa > 0$	168
5.17	Model predictive power: Gain model, constrained coefficients . .	169
5.18	Effective scale of the gain effect	172
6.1	Schematic of test sound stimuli	185
6.2	Response to the test sound when presented in different contrast contexts	188
6.3	Time course of gain changes, measured via the test sound	189
6.4	Illustration of the contrast time-frequency kernel model	192
6.5	Contrast time kernels, for four example units	197
6.6	Mean contrast time kernel across the 77 cortical units	197
6.7	Model predictive power: Contrast time kernels	198
6.8	Comparison of assumptions for contrast time kernels	200

7.1	Electrophysiological determination of the central nucleus of the IC	212
7.2	Output nonlinearities for three example units in IC	213
7.3	Output nonlinearities in low and high contrast, IC	215
7.4	Gain model: statistics on contrast-dependent gain changes . . .	216
7.5	Gain contrast kernels: examples in IC	218
7.6	Correlation between linear (STRF) kernel and gain (contrast) kernel coefficients, IC	219
7.7	Model predictive power: Gain model with constrained coefficients, IC	220
7.8	Model predictive power: Contrast kernels for multiple nonlinearity parameters, IC	221
7.9	Contrast-dependent changes in Fisher information: IC vs. A1/AAF	223
7.10	Effective scale of the gain effect in IC	224
7.11	Contrast time kernels, for 4 example units in IC	225
7.12	Mean contrast time kernel: IC vs A1/AAF	225
7.13	Comparison of assumptions for contrast time kernels, IC	226

List of Tables

4.1	Summary of stimulus statistics	80
5.1	Summary of stimulus statistics for test/mask DRCs	119
5.2	Success of test/mask models	128

Chapter 1

Introduction

1.1 Two problems for hearing

Somehow, somewhere along a tortuous evolutionary trajectory, some early ancestors of ours figured out that the air or water or rock that surrounded them would, from time to time, push on them, very, very gently. No doubt their own ancestors had occasionally been affected in some small way by these little nudges — but the pressure applied was pretty minuscule, so it had been a pretty fair bet that these tiny taps were benign in and of themselves. These spectacularly weak events, it turned out, were not simply random jostles, but actually carried useful information about what was changing in the local environment. Those who learned how to detect these signals and interpret them ended up doing much better than their neighbours, who were probably eaten, or crushed, or starved, or destined to die lonely.

This is but a romanticised imagining of the origins of hearing. Unfortunately, not only is it vastly oversimplified, it is also probably as far as we will go into a Devonian underwater fantasy land over the course of the next few hundred pages. But it serves a purpose to remind us that the pressure waves that reach our ears only carry information if we are able to extract it.

When we survey the field of sounds which are, in one way or another, relevant to us, we quickly notice that it's a rather diverse collection. Sounds are produced by a huge variety of sources, in many different conditions. They are emitted by our fellow apes, and by lions and by cars and by wind. They are bellowed by apes who are close at hand or mumbled by those who are far away; they are growled by lions to our left or who pace underneath our tree; they are hummed by cars which are accelerating towards us or slowing to a stop. The general success of our auditory system in each of these conditions comes from its capacity to pull out content from every one of these signals, in spite of their variety.

Dealing with such a range of potential signals raises many difficulties. Here, we shall restrict our view to two particular challenges, relating to the overall statistics of sounds. The first is the *intensity problem*. Some sounds are quiet, meaning that the energy in the sound pressure waves is, on average, rather small; other sounds are loud, meaning that the energy in their sound pressure waves is, on average, rather large. This is illustrated in Figure 1.1A.

A second, related challenge is the *contrast problem*. Independent of their *average* sound intensity (i.e., whether they are generally quiet or loud), the energy of some sounds varies a lot over time, while the energy of other sounds varies only a little (Figure 1.1B). This is also true of different environments: while the average sound intensity may be high or low in a given situation, the important sensory information within that context may lie either in small or large deviations from the average sound. For example, detecting a subtle increase in the loudness of an approaching car's engine in a mostly constant background of traffic noise can be just as crucial as hearing a pronounced honk. Thus, while the intensity problem is about the *mean* of the sound energy, the contrast problem is about its *variance*.

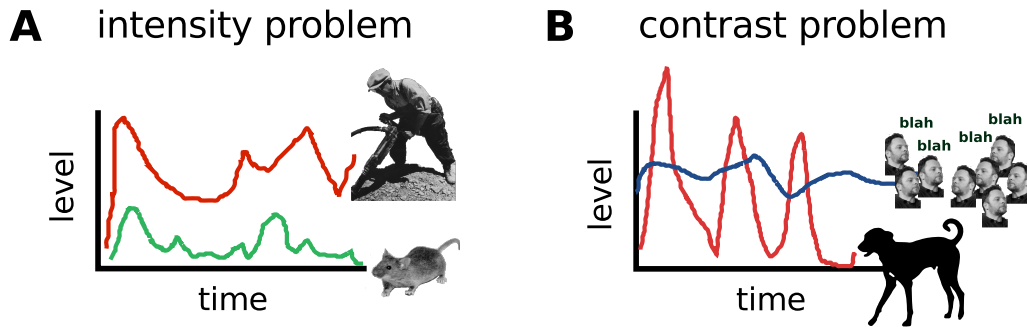


Figure 1.1: The intensity problem and the contrast problem. (A) The intensity problem. The auditory system has to be able to represent sounds which are loud (e.g. the grinding of a jackhammer), and sounds which are soft (e.g. the squeaking of a mouse). These sounds differ in their mean sound level. (B) The contrast problem. The auditory system has to be able to represent sounds with levels which vary a lot over time (e.g. the barking of a dog), and sounds with levels which vary only a little over time (e.g. the hubbub of a crowded room). These sounds differ in their contrast.

The fundamental question of this thesis is to ask how the central auditory system — in particular, the primary auditory cortex — manages the contrast problem. This, as we shall see, represents a major challenge for neurons in the auditory system. The firing rates of these neurons, to first approximation, follow the fluctuations in sound level (within their receptive field) of an ongoing sound, albeit in a nonlinear manner. They have only a limited dynamic range in their firing rate to encode how these sound levels change over time. The contrast problem is, therefore, a problem of neural coding (Figure 1.2).

Before laying out how I intend to address this question, it is necessary to review some existing literature, and to cover some of the core theoretical ideas. I begin here by considering two related stories: how the peripheral auditory system solves its intensity problem; and how the visual system solves both its intensity and its contrast problems. Together with a review on previous studies on auditory contrast, these motivate a set of experiments designed to directly probe how the mammalian auditory system is able to extract information from sounds that have very different contrasts.

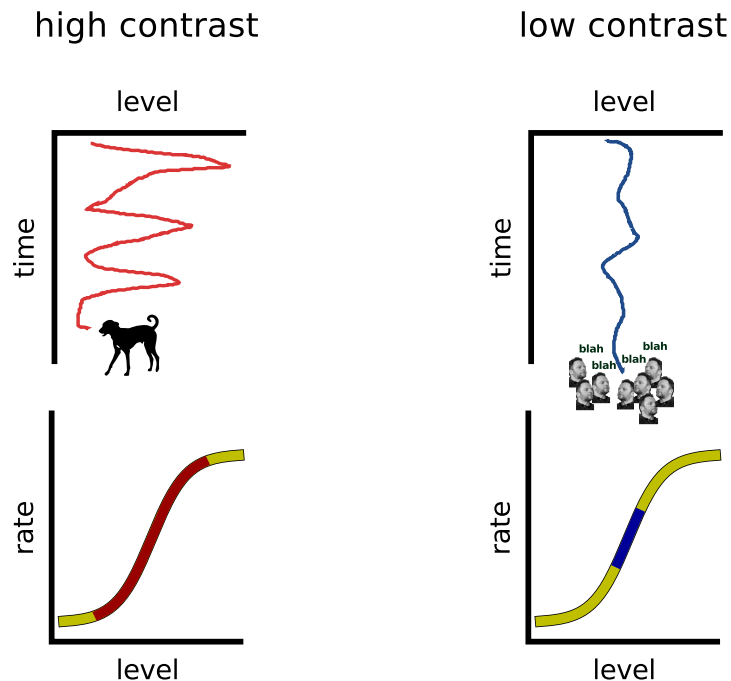


Figure 1.2: The contrast problem for neural coding. Auditory neurons which encode the intensity of a stimulus through their firing rate have only a limited dynamic range. This is illustrated via a hypothetical stimulus-response relationship, which has a lower threshold and an upper saturating firing rate. In the absence of other effects, if neurons use their whole dynamic range under high contrast stimulation (left), then they will only use a fraction of it under low contrast stimulation (right).

1.2 The intensity problem

The range of sound energies we encounter is enormous. The difference between the quietness of a night in the forest and the hooting drama of crossing the street may seem trivial given the ease with which we handle both, until we actually measure the size of the difference between the sound energies involved. An extremely quiet sound at the threshold of hearing, such as gentle breathing, carries sound energy on the order of tens of micropascals, while very loud sounds, such as jet engines, have energies in the range of tens of pascals. Thus the human ear is able to detect and provide function over a range of approximately 6 orders of magnitude of sound pressure, or 12 orders of magnitude of sound power. Given this huge range, it is typical to describe sound intensities

using a logarithmic scale of sound pressure level (decibels; dB), rather than the linear scale of sound pressure (Pa).

Broadly speaking, there are two potential strategies which our auditory system could employ in order to manage this massive range. One, which I shall call the *division of labour strategy*, is to maintain separate, independent circuits that are each specialised for dealing with sounds in a particular range of intensities. The other, which I shall call the *adaptation strategy* is to change the way that a set of circuits operate, in order to provide for the demands of a specific context.

The mammalian peripheral auditory system makes use of both strategies to address the intensity problem. The cochlea itself invokes adaptive processes in order to amplify soft sounds and compress loud sounds, mediated through the action of outer hair cells on the basilar membrane. This increases the overall dynamic range of the peripheral sensory organ by boosting its sensitivity to low sound levels [Ruggero and Rich, 1991; Yates, 1995]. I class this as *adaptive*, as the cochlea shifts its operating point for different contexts so that the same inner hair cells can code both soft and loud sounds, reducing the need for specialisation. Further adaptation is provided by the olivocochlear reflex, which scales down the amplification of the outer hair cells during sustained bursts of sound energy [Dallos, 1992], as well as by the middle ear muscle reflex, which acts on the conducting apparatus of the middle ear [Liberman and Guinan, 1998]. The inner hair cell synapse also appears susceptible to a depletion of pre-synaptic stores of neurotransmitters [Furukawa et al., 1978; Moser and Beutner, 2000; Spassova et al., 2004] and a desensitisation of post-synaptic receptors [Goutman and Glowatzki, 2007], causing short-term firing rate adaptation in the auditory nerve. Yet despite these positive and negative feedback mechanisms, no single element of the auditory pathway remains able to encode the full range of sound intensities [Barbour, 2011]. Thus, in order to carry information from the cochlea to the brainstem, the auditory nerve

(AN) adopts a division of labour strategy, and maintains a range of separate, independent circuits for different sound intensities. Amongst AN fibres, there is a broad distribution of threshold and saturation levels, which allow for the processing of low and high intensity sounds [Sachs and Abbas, 1974]. This arrangement is illustrated schematically in Figure 1.3. These strategies, combined, are effective: across the full intensity range of hearing, we maintain the capacity to discriminate level differences of between 1–2 dB [Rabinowitz et al., 1976; Viemeister, 1983; Florentine, 1987] (although performance is not uniform throughout [Pienkowski and Hagerman, 2009]).

Each of these two strategies alone, or combined, might provide a solution to the contrast problem. In the division of labour approach, the auditory system would use separate, specialised circuits to process stimuli with different contrasts. This would require some neurons to perform optimally — or at least near optimally — at low contrasts, while others must do so at high contrasts. One way that this may manifest is illustrated in Figure 1.4. Alternatively, neurons may change the way they encode stimuli under different contrast conditions, so that they remain informative about the stimulus. To get an idea as to how this might happen, we turn to the visual system.

1.3 Contrast gain control in the visual system

The intensity and contrast problems have clear counterparts in vision. Visual scenes vary in their overall luminance (intensity), which can range over approximately 10 orders of magnitude. This problem is resolved at the retina by fast-acting light adaptation, which alters the sensitivity of the peripheral organ based on the local mean intensity [Shapley and Enroth-Cugell, 1984; Dunn et al., 2007]. The net result of this is that retinal ganglion cells' luminance-

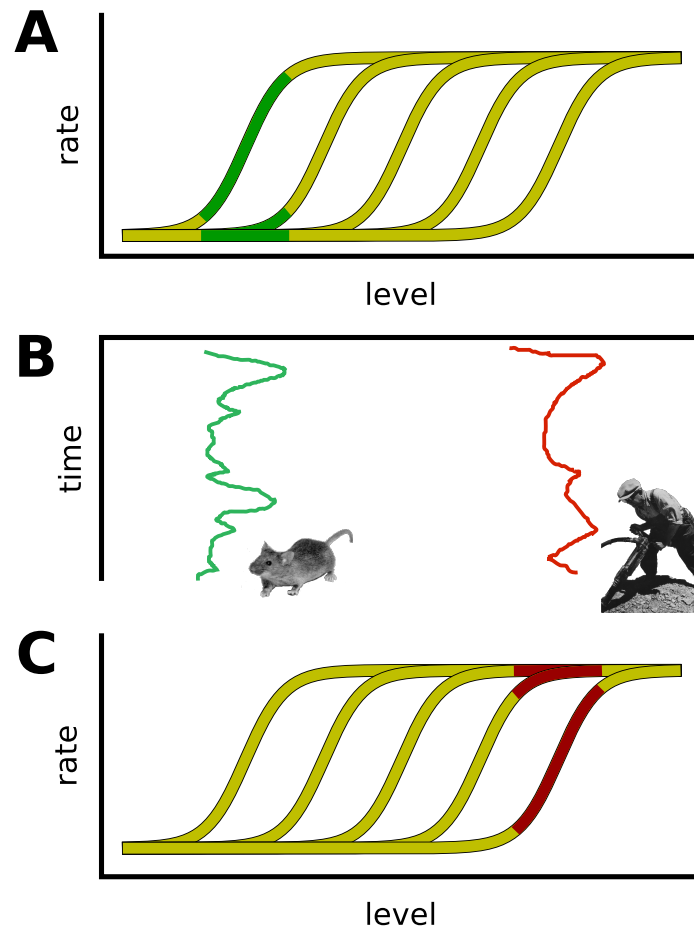


Figure 1.3: The division of labour solution to the intensity problem in the auditory nerve. This diagram shows a simplified schematic of how auditory nerve fibres encode sound intensity through their firing rate. (A, C): Each fibre has only a limited dynamic range, which it uses to represent a finite region of sound intensity (level) space. Different fibres have different threshold and saturation points, so that the full range of naturally-encountered sound levels is encoded through specialised circuits. Each yellow curve shows the stimulus-response relationship for a single fibre. The quiet sound shown in (B) uses the dynamic range of the low-threshold/low-saturation fibres, whose activation is shown in green in (A). The loud sound shown in (B) uses the dynamic range of the high-threshold/high-saturation fibres, whose activation is shown in red in (C). This is only a simplified schematic, as the spontaneous rate and dynamic range of auditory nerve fibres vary with threshold.

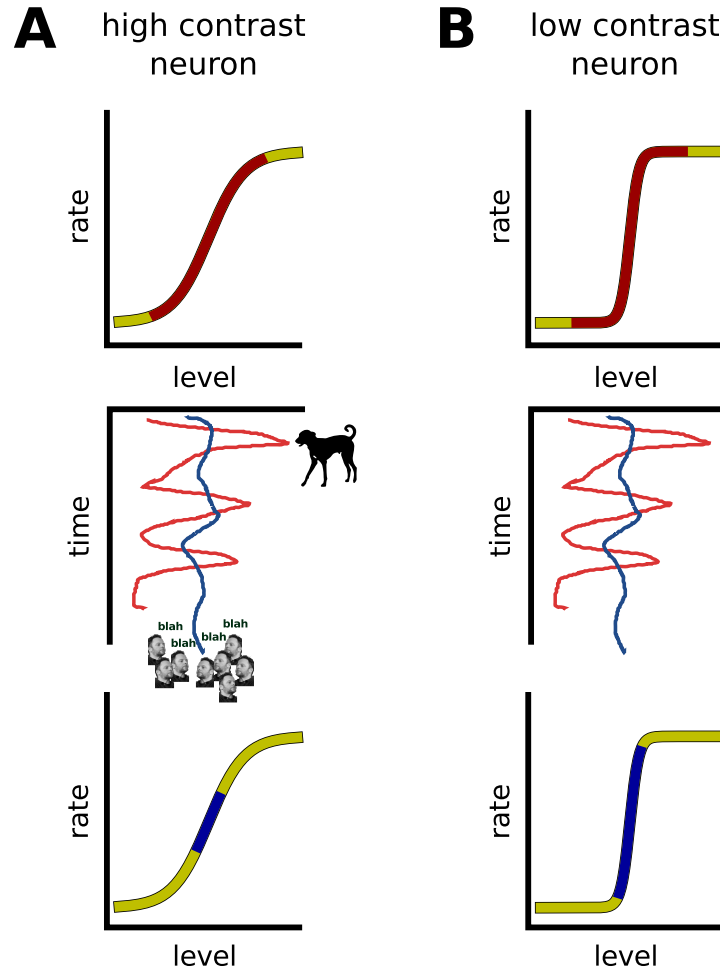


Figure 1.4: One way which the division of labour solution to the auditory contrast problem might take form. Here, we consider two neurons, (A) and (B), with fixed stimulus-response relationships. The neuron in (A) is specialised for high contrast sounds (red), for which the neuron uses its full dynamic range to represent. Under low contrast stimulation (blue), this cell only uses a fraction of its dynamic range. The neuron in (B) is specialised for low contrast sounds; under high contrast stimulation, it is not able to capture level variations beyond a certain range.

response relationship is shifted horizontally to compensate for the mean stimulus intensity.

Just as the mean intensity changes across scenes, so does its variance. This gives rise to a contrast problem (Figure 1.5). The visual system solves this via contrast gain control, a process that begins in the retina [Shapley and Victor, 1978, 1981; Benardete et al., 1992; Chander and Chichilnisky, 2001; Baccus

and Meister, 2002; Zghloul et al., 2005], and continues in the lateral geniculate nucleus (LGN) and possibly in the primary visual cortex (V1) [Kaplan et al., 1987; Sclar, 1987; Sclar et al., 1990; Cheng et al., 1995]. When contrast is reduced, neurons' stimulus-response relationships undergo a rapid change in slope, so that the neurons use more of their dynamic range to encode the observed distribution of stimulus values. This rescaling occurs as a function of root-mean-square contrast (the standard deviation of light intensities relative to the mean) [Sclar et al., 1990; Bonin et al., 2006]. In visual neurons from retina to cortex, this process is also accompanied by a slow adaptation to stimulus contrast, thought to be due to a prolonged membrane afterhyperpolarisation effect [Smirnakis et al., 1997; Brown and Masland, 2001; Kim and Rieke, 2001; Baccus and Meister, 2002; Solomon et al., 2004; Manookin and Demb, 2006]. These two separate processes have historically been termed (fast) contrast gain control and (slow) contrast adaptation respectively; they are nevertheless both realisations of the *adaptation strategy* I defined above.

The visual system thus presents us with a striking example of how an adaptation strategy can be used to solve the contrast problem. One particularly noteworthy feature of this solution is that it embodies the principles of efficient coding [Barlow, 1961; Atick, 1992]. By scaling gain up in low contrast, the visual system increases the signal-to-noise ratio in its representations, and makes use of coding capacity that would otherwise be wasted. By scaling gain down in high contrast, the system ensures that information is not lost due to thresholding and response saturation. Unlike the division of labour strategy in the auditory nerve, employing gain control also has the advantage of avoiding a potentially costly duplication of resources.

One form in which gain control is often cast is that of *divisive normalisation*. In the abstract, this is a gain standardisation process by which an initial set of responses — usually the result of information fed forward from earlier brain

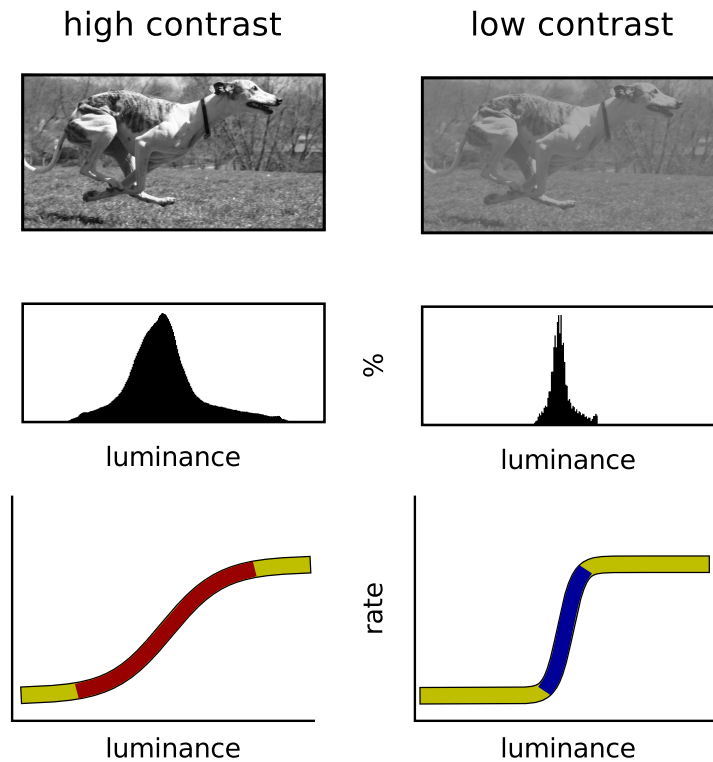


Figure 1.5: Contrast gain control in the visual system. The contrast problem is solved in the visual system by contrast gain control and contrast adaptation. This diagram shows a simplified schematic of the effect of this process on the coding of a retinal ganglion cell (RGC). Under high contrast stimulation (left), the distribution of luminance values within the cell’s receptive field has a high variance over space and time (middle). As a result, as the cell codes the visual stimulus over time, it uses its full dynamic range of firing rates. Under low contrast stimulation (right), the distribution of luminance values is narrow. The system adopts an adaptation strategy, and rescales the gain of the RGC to more efficiently encode the observed distribution.

areas — is rescaled. The scaling factor takes the form of a local response normaliser: the activity of each neuron is divided by the pooled activity over other neurons in a local neighbourhood [Heeger, 1992; Carandini et al., 1997].

This is not merely a computational abstraction: there is considerable evidence for normalisation in a large number of systems beyond the retina, including V1 [Heeger, 1992; Carandini et al., 1997; Rust et al., 2005], extrastriate visual cortex [Miller et al., 1993; Recanzone et al., 1997; Missal et al., 1997; Simoncelli and Heeger, 1998; Britten and Heuer, 1999; Heuer and Britten, 2002; Zoccolan

et al., 2005], superior colliculus [Basso and Wurtz, 1997], and the *Drosophila* antennal lobe, which mediates olfaction [Olsen et al., 2010].

There remains considerable debate as to what combination of cellular and circuit mechanisms actually underscores divisive normalisation in the visual system [Kohn, 2007]. Nevertheless, divisive normalisation has proved to be a powerful idea for advancing our understanding of the computations actually being performed by a given system. Normalisation promotes efficient coding, not only by shifting stimulus representations to use more of neurons' dynamic range, but also by encouraging decorrelated, higher-entropy representations of natural signals [Ruderman and Bialek, 1994; Olshausen and Field, 1996; Brady and Field, 2000; Fairhall et al., 2001; Schwartz and Simoncelli, 2001]. It might also subserve other computations, such as decoding [Deneve et al., 1999; Ringach, 2010] and marginalisation [Beck et al., 2011].

One final aspect of gain control in the visual system is worth mentioning at this stage. If we return to the two images of Figure 1.5, there is no ambiguity to a normal human observer that these are exactly the same image, depicting the same content under different conditions. Both pictures are pictures of a dog. Contrast, in this instance, is a *nuisance parameter*. The goal of gain control and divisive normalisation is to integrate out nuisance parameters [Beck et al., 2011]. By doing just that, many systems appear to construct representations that are invariant to the normalised statistic, including visual representations in V1 that are contrast-invariant [Albrecht and Hamilton, 1982; Heeger, 1992; Busse et al., 2009; Ringach, 2010], velocity representations in MT that are spatial-pattern-invariant [Heeger et al., 1996; Simoncelli and Heeger, 1998], and odour representations in the antennal lobe that are concentration-invariant [Luo et al., 2010; Olsen et al., 2010]. Links such as these between gain control and normalisation will prove to be important later in this thesis.

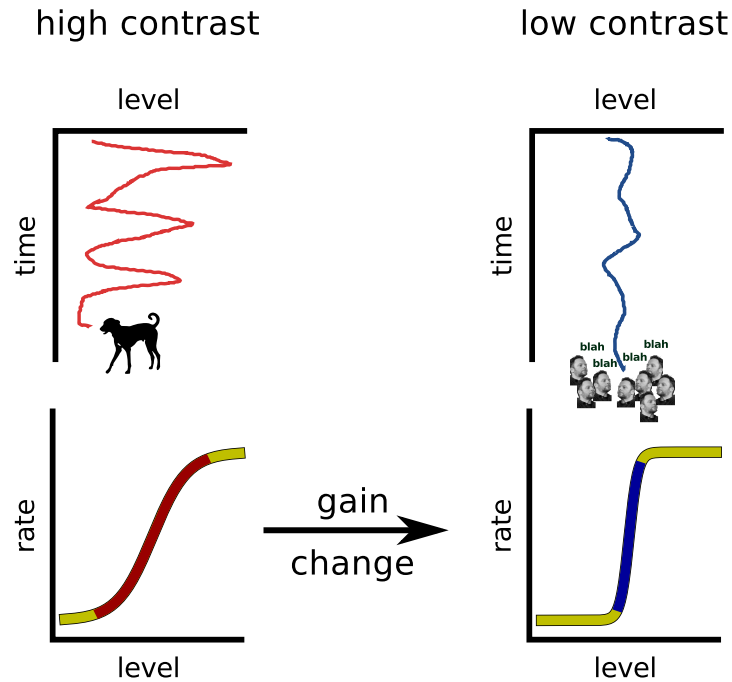


Figure 1.6: One way which the adaptation solution to the auditory contrast problem might take form. Unlike the division of labour strategy, which requires different neurons for different contrasts (Figure 1.4), this solution involves a single neuron for both high (left) and low (right) contrast. Like the visual case (Figure 1.5), this neuron rescales its gain to operate efficiently in both contexts.

1.4 The contrast problem in the auditory system

There are two competing possibilities for how the auditory contrast problem might be resolved in the auditory cortex. Either different neurons specialise for different contrasts, as in Figure 1.4, or neurons adapt the way they operate under different contrast conditions, as in Figure 1.6. Here I shall review the existing evidence for whether these strategies are used in the auditory pathway at large.

1.4.1 Evidence for the division of labour strategy

Two particular studies seem to suggest that the central auditory system employs a division of labour strategy in order to cope with sounds of both high and low contrast. [Escabi et al. \[2003\]](#) recorded from the inferior colliculus (IC) of the anaesthetised cat while presenting a spectrotemporally complex ripple stimulus (the details of which are described in the next chapter, in Section 2.1.2). In turn, they varied the size of the fluctuations in sound level present in these ripples, thereby controlling stimulus contrast (see Section 2.1.3.3).

The authors found that the stimulus selectivities of IC neurons, as measured via the neurons' spectrotemporal receptive fields (STRFs; see Section 2.2.1), were unaffected by contrast. However, they also found that statistics of neurons' responses — characterised by their mean firing rate and the extent to which their firing rate was modulated by fluctuations in the stimulus — could depend on stimulus contrast in one of two ways. One subset of neurons increased their responses monotonically with contrast, while, for another set, responses were maximal for an intermediate contrast. The coding of the contrast-monotonic class of neurons appears specialised for high contrast sounds, while the coding of the contrast-nonmonotonic class of neurons appears specialised for sounds of intermediate contrast.

Incidentally, while the authors pursued an analysis of the population data in terms of mutual information rate — which would directly couch the coding properties of contrast-monotonic and contrast-nonmonotonic neurons in terms of stimulus informativeness — their focus was on a different question, and they therefore did not distinguish between the two classes when reporting their data. They also did not report whether there were any gain changes across conditions, so this data cannot rule out the adaptation hypothesis in the IC. Finally, the resolution of their study was limited, as they tested only four

contrast conditions (a fact they obfuscate by smooth interpolation on their graphs). Notwithstanding these limitations, their data suggests that the IC may address the contrast problem by dividing its labour between neurons which prefer different contrasts.

A study by [Barbour and Wang \[2003b\]](#) in awake marmoset A1 argues more directly that auditory cortical neurons specialise for sounds with particular contrast statistics. Unlike the work of [Escabi et al. \[2003\]](#), the authors here did not consider the statistics of sound over both frequency and time, but just over frequency. They presented a class of “static” sounds known as Random Spectral Stimuli (RSS). This stimulus paradigm attempts to provide an auditory analogue for static images: each RSS sound has a power spectrum that varies as a function of frequency, but does not change with time (see Section 2.1.1). Each sound was nevertheless presented for only a short time. The contrast of a sound could be controlled by changing the variance (across frequency) in its power spectrum. In order to distinguish this experimental paradigm from the control of *spectrotemporal* contrast by [Escabi et al. \[2003\]](#), I refer to it as control of the *spectral contrast* (see Section 2.1.3.3)

When the authors looked at how A1 neurons responded to sounds of different spectral contrasts, they found that some neurons tended to *increase* their firing rate with increasing spectral contrast, while other neurons tended to *decrease* their firing rate with increasing spectral contrast. They therefore argue that A1 contains some cells which prefer high contrast (merely by acting as linear filters) as well as other cells which are “tuned” for lower contrasts.

The promise of the results of [Barbour and Wang \[2003b\]](#) as evidence for the division of labour strategy is unfortunately diminished by a number of methodological and statistical issues in their experiment and analysis. These are addressed in more detail later, in Section 3.4.1. More generally, however, the

extent to which their stimulus paradigm provides an appropriate model for the contrast problem is limited. This is because sounds, by their very nature, are time-varying. It is not possible to extrapolate from how neurons differentially respond to sounds of different *spectral contrasts* to predict how they would respond under different *spectrotemporal contrast* conditions.

Nevertheless, given the data of [Escabi et al. \[2003\]](#) and [Barbour and Wang \[2003b\]](#), we would not be surprised to find evidence that the primary auditory cortex has specialised circuits for processing sounds of high contrast, and specialised circuits for processing sounds of low contrast. This would need to be weighed alongside evidence for the alternative.

1.4.2 Evidence for the adaptation strategy

Several studies have investigated how auditory neurons from the auditory nerve to the auditory cortex respond to sounds of different *temporal contrasts*. This involves manipulating the statistics of how the intensity of sound changes over time, while keeping the spectral content fixed (described in more detail in Section 2.1.3.3). These studies argue that the adaptation strategy is widely used throughout the auditory pathway.

One simple way of controlling the temporal contrast is to change the modulation depth of sinusoidally amplitude-modulated (SAM) tones. In general, when the modulation depth of the stimulus is increased, so the modulation depth of auditory neurons' firing rates also increases; this is true for neurons in the auditory nerve [[Joris and Yin, 1992](#)], inferior colliculus [[Rees and Møller, 1983](#); [Nelson and Carney, 2007](#)], and primary auditory cortex [[Eggermont, 1994](#); [Liang et al., 2002](#); [Malone et al., 2007](#)]. However, there is not a simple linear relationship between these two variables. Rather, when modulation depths are low, many

neurons are hypersensitive, and their responses to the small stimulus modulations are still considerable. This is often described as *modulation gain*. In turn, at large modulation depths, the size of their response modulations saturates. Thus, these neurons change their gain as a function of modulation depth. This has been demonstrated to occur for neurons in the auditory nerve of the anaesthetised cat [Joris and Yin, 1992], and for neurons in the inferior colliculus of the anaesthetised rat [Rees and Møller, 1983] and awake rabbit [Nelson and Carney, 2007]. Malone et al. [2010] observed the same phenomenon in the firing rates of A1 neurons in the awake macaque. When SAM tones of low modulation depth are presented, A1 neurons are as hypersensitive as those in the auditory nerve, if not more so. Finally, Krishna and Semple [2000] observed a similar effect in the IC of the anaesthetised mongolian gerbil, but found that at some modulation frequencies, increasing modulation depth led to a monotonic increase of the firing rate, while at other modulation frequencies, increasing modulation depth led to a monotonic decrease of the firing rate. In auditory cortex, however, neurons' preferred modulation frequency remains unchanged with depth [Eggermont, 1994; Liang et al., 2002].

One issue with the SAM tone paradigm is that it is highly simplistic. In a drive to understand how the contrast problem might be resolved under more complex stimulation, a number of studies have explored how neural coding changes when the distribution of sound levels presented over time is varied in more sophisticated ways. These studies have generally taken a spectrally simple carrier — such as a pure tone or noise — and modulated its sound level over time. By controlling the statistics of the sound level distributions, the authors have been able to construct sounds with envelopes much like in Figure 1.1, and thus directly investigate both the intensity problem and the contrast problem.

Kvale and Schreiner [2004] looked at how the higher-order statistics of sound level affect the gross response properties of neurons in the central nucleus of the

inferior colliculus in the anaesthetised cat. They modulated the sound level of pure tones over time, while controlling the moments of the sound level distributions: the mean was fixed, but the variance (i.e. contrast) and the kurtosis were subject to experimental control. While the authors found that decreasing contrast led to a decrease in the average firing rate, there was also an increase in response gain. This was measured via the neurons' temporal filters, which capture the shape of the amplitude modulations that would elicit spikes: under low contrast stimulation, the amplitude of these filters was greater. When the authors switched from the low contrast condition to the high contrast condition, response properties changed with a time constant of 140 ms; when they switched from the high contrast condition to the low contrast condition, responses changed with a slower time constant of 300 ms. Changes in the kurtosis of the level distribution also affected average firing rates, with a slower time course of adaptation than for variance.

A similar study was performed by [Nagel and Doupe \[2006\]](#) in field L of the awake zebra finch, an avian analogue of the primary auditory cortex. They altered the mean and variance of the sound level distribution of a similar stimulus. The shapes of these neurons' temporal filters were sensitive to mean level, but insensitive to the variance. The gain, however, changed with the stimulus variance, much in the same way as observed by [Kvale and Schreiner \[2004\]](#). This effect manifested within 100 ms of a change in stimulus variance.

These studies together show that when the auditory system is faced with sounds whose levels vary over a narrow range — whether they are SAM tones of low modulation depth, or sound level distributions of low variance — many of its neurons rescale their gain to become more sensitive to the smaller fluctuations in sound level. This is reminiscent of the gain changes seen in the visual system.

A final set of related work was performed by Dean and various colleagues [[Dean et al., 2005, 2008; Wen et al., 2009](#)]. As with the work of [Kvale and Schreiner](#)

[2004] and Nagel and Doupe [2006], they modulated the sound level of a carrier (here, noise), while controlling the shape of the envelope's level distribution. These distributions were defined to have a high probability region (HPR), such that some sound levels were presented more frequently than others.

In the first study [Dean et al., 2005], the authors recorded from the inferior colliculus of the anaesthetised guinea pig, and measured the stimulus-response functions of these neurons (i.e. the relationship between sound level and firing rate for these neurons, akin to what is depicted in Figures 1.2 – 1.6). When they varied the mean sound level — by shifting the HPR to higher or lower levels — IC neurons typically shifted their stimulus-response functions horizontally (i.e. changed their thresholds), towards the new HPR. When the authors measured coding accuracy in terms of Fisher information, they found that neurons were, on average, adjusting to improve the accuracy of their coding over the high probability region. It is important to note that there was a lot of variation seen in this effect across IC neurons, such that this trend in coding accuracy could only be observed in the population average.

The authors then looked at changing the width of the HPR, i.e. changing the contrast. While some neurons changed their stimulus-response functions as a result of this manipulation, many neurons did not. Nevertheless, the net sum of these small changes was such that when the HPR was wide (i.e. the contrast was high), the Fisher information of the population responses was spread over a slightly wider region of stimulus space than when the HPR was narrow (i.e. the contrast was low). A similar effect was seen when the level distribution was bimodal, with two HPRs: here, the population Fisher information spread its weight over the two peaks of the stimulus distribution.

In two follow-up studies, the authors focused on the coding shifts induced by changes in mean level. Dean et al. [2008] found that these changes occur

with an adaptation time constant of 160 ms. Similar to the findings of [Kvale and Schreiner \[2004\]](#) (who manipulated variance, rather than mean), the time constants for adaptation were found to be slower for decreases in mean level compared with increases in mean level. Finally, [Wen et al. \[2009\]](#) presented the same stimuli to anaesthetised cats and recorded the responses of auditory nerve fibres. They found that changing the mean level of the HPR also causes a similar horizontal shift in the rate-level functions of AN fibres. However, the adaptive effects at the auditory periphery are too small to account for the observations in IC.

The three experiments of Dean and colleagues demonstrate that when the level statistics of sound stimulation change, the population of IC neurons adjusts its coding to be more informative about the statistics of recent stimulation. Some, but not all of these coding changes are inherited from earlier in the pathway. It is necessary to emphasise that much of the focus of these three studies was on the intensity problem, for which it appears that an adaptive strategy is being employed. However, regarding the data of [Dean et al. \[2005\]](#) on the contrast problem — wherein they changed the width of the HPR — their results do not unequivocally show that adaptation is used to solve the contrast problem. Changing the width of the level distributions led to only small, or barely noticeable changes in the stimulus-response functions of individual neurons. While, at the population level, changes in the Fisher information profile appear to signal an adaptation, these results are still consistent with a division of labour strategy: under low contrast stimulation, the IC might selectively amplify the firing rate of specialist low contrast neurons and suppress the firing rate of specialist high contrast neurons.

1.5 The auditory cortex

A brief pause is necessary here to place the auditory cortex in the context of the auditory pathway. While much is known about the subcortical processing of sound, the mammalian auditory cortex has remained somewhat elusive. This is partly due to the fact that by the time sensory information reaches the auditory cortex, it has already undergone considerably pre-processing. Information from the cochlea is carried, via the auditory nerve, to the cochlear nucleus (CN) in the brainstem. The CN contains a large number of different cell types, each with different firing properties [Pfeiffer, 1966; Brawer et al., 1974; Young and Brownell, 1976]. These, in turn, project to a number of monaural and binaural nuclei, scattered through the brainstem. Information from all these streams then converges in the inferior colliculus [Casseday et al., 2002]. In turn, the IC provides the principal input to the medial geniculate body (MGB) of the thalamus [Calford and Aitkin, 1983; Winer, 1992], and thence to cortex. This is by no means a simple, feed-forward pathway: the cortex sends vast downwards projections to the MGB and IC [Colwell, 1975; Andersen et al., 1980; Huffman and Henson Jr., 1990; Winer et al., 1998].

Neurons in the primary auditory cortex show many nonlinear properties, including a hypersensitivity to stimulus onsets [Heil and Irvine, 1996; Fishbach et al., 2001; Heil and Neubauer, 2003], context dependence [Blake and Merzenich, 2002; Valentine and Eggermont, 2004; Gourévitch et al., 2009; Ahrens et al., 2008] and stimulus-specific adaptation [Ulanovsky et al., 2003, 2004]. As a result, although these neurons are often frequency tuned like subcortical neurons, attempts to capture their responses using linear or near-linear models such as the spectrotemporal receptive field (covered in detail in Section 2.2.1) often fail to generalise across different stimulus conditions [Rotman et al., 2001; Machens et al., 2004; David et al., 2009]. Thus, there is consid-

erable scope for extending our understanding of how auditory cortex processes complex sound.

In this thesis, I focus on neural coding within the primary fields of the auditory cortex of ferrets. The ferret, initially used as an animal model for sensory development, has in recent years become an established model for coding in the auditory cortex (e.g. [Depireux et al., 2001; Schnupp et al., 2001; Fritz et al., 2005b; Garcia-Lazaro et al., 2006; Fritz et al., 2007; David et al., 2009; Bizley et al., 2010; Campbell et al., 2010]). The ferret has two primary fields in its auditory cortex: the primary auditory cortex (A1), and the anterior auditory field (AAF) [Kelly et al., 1986; Phillips et al., 1988; Kowalski et al., 1995; Bizley et al., 2005], as well as several higher-level fields [Bizley et al., 2005]. Like in many of the subcortical stations, the tonotopic map of the cochlea is preserved in A1/AAF.

The choice to study the contrast problem in the primary fields of the auditory cortex was initially motivated by the power of divisive normalisation models in the primary visual cortex, V1. The study of Barbour and Wang [2003b], though not without its flaws, provides a challenging counter-hypothesis to the possibility of contrast-dependent gain changes in the auditory cortex. As choices such as these are always somewhat arbitrary, later in this thesis I extend the experiments pursued in auditory cortex to the inferior colliculus in order to address whether sensitivity to contrast changes within the auditory pathway.

1.6 Open questions

Studies to date do not resolve the question as to how the auditory cortex manages the contrast problem. There are several reasons for this. First, of the two studies advocating the division of labour hypothesis, one was poorly constructed, while the other's results may or may not extend to cortex. Second,

the studies advocating the adaptation hypothesis all used simple stimuli, especially those studying gain changes within the auditory cortex. It is unclear to what extent these results would persist under stimulation with sounds with *spectrotemporal* dynamics. Third, those studies which have directly explored how the statistics of sound level distributions affect gain (in the IC and in avian field L) have only tested two contrast conditions. If the contrast problem is solved via gain changes, we should be able to quantify a functional relationship between stimulus statistics and neural gain. We should also be able to determine how stimulus information is integrated to inform the appropriate operating point for the system. Fourth, an understanding of how the contrast problem is managed at the level of cortex should not merely be descriptive, but also predictive: it should allow us to extend existing quantitative models of neurons' firing behaviour to cover stimulus-response relationships under different stimulus statistics. Finally, we may not simply be able to extrapolate from the way the contrast problem is managed at lower stations to how it is managed in cortex. The observation that the IC has a greater adaptive capacity to solve the intensity problem than the auditory nerve raises the possibility that the cortex has a greater capacity to solve the contrast problem than its subcortical sources. This can only be answered with a direct comparison between auditory cortex and a subcortical station. We need only look at the known richness of contrast gain control, adaptation, and normalisation in the visual cortex to see how little we currently understand of the resolution of the contrast problem in the auditory cortex.

The goal of this thesis is to address these shortcomings, by characterising how the primary fields of the ferret auditory cortex deal with the contrast problem under spectrotemporally complex stimulation.

1.7 Structure of the thesis

In Chapter 2, I review the theoretical and methodological background behind the core tools used in this thesis. In particular, I introduce the Dynamic Random Chord stimulus paradigm (Section 2.1.1), define metrics for contrast (Section 2.1.3), and cover a set of standard models for the stimulus-response relationships of auditory cortical neurons: the spectrotemporal receptive field (STRF; Section 2.2.1), and the linear-nonlinear model (LN; Section 2.2.2). This chapter also provides a description of the experimental methods used for in vivo extracellular electrophysiology (Section 2.3).

In Chapter 3, I test the division of labour hypothesis: that neurons in the primary auditory cortex (A1/AAF) of ferrets are specialised, or “tuned” for sounds of particular contrasts. To do this, I measure the contrast-response functions of A1/AAF neurons using both ongoing and transient sounds.

In Chapter 4, I test the adaptation hypothesis: that neurons in A1/AAF change their gain to compensate for changes in stimulus contrast. To do so, I measure how cortical neurons’ stimulus selectivity (via their STRFs) and their stimulus-response functions (via their output nonlinearities) change with contrast. I also quantify the observed gain changes, to establish a functional relationship between stimulus statistics and gain changes.

As Chapters 4 and 5 provide strong evidence that neurons in the auditory cortex use the adaptation strategy rather than the division of labour strategy, in the remaining chapters I explore gain changes in much greater detail.

In Chapter 5, I characterise how neuronal gain depends on the spectral profile of stimulus contrast. This involves two experiments. In Section 5.2, I determine whether gain changes depend on spectrally-local or spectrally-global contrast statistics. In Section 5.3, I develop a new class of *contrast kernel* models for

auditory neurons. I use these in Section 5.4 to measure the spectral integration which informs gain control on a neuron-by-neuron basis.

In Chapter 6, I characterise the time course over which gain changes manifest. As in Chapter 5, this involves two experiments. I also extend the contrast kernel models of Section 5.3, in order to measure the temporal integration which informs gain control for each neuron.

Finally, in Chapter 7, I ask whether the gain changes observed in cortex are inherited from earlier in the auditory pathway. This involves a direct comparison between how cortical and IC neurons respond to the same contrast problem, via the same set of stimuli.

Chapter 2

Theory and methods

2.1 Stimuli for model fitting

2.1.1 DRCs

The main stimulus used throughout this work was a dynamic random chord sequence (abbreviated as a DRC, or DRCs; see Figure 2.1). These have a distinguished history in auditory research, as they allow for efficient estimation of the spectrotemporal sensitivities of auditory neurons [deCharms et al., 1998; Schnupp et al., 2001; Rutkowski et al., 2002; Linden et al., 2003; Sahani and Linden, 2003a; Ahrens et al., 2008; Christianson et al., 2008; Bitterman et al., 2008] (see below, Section 2.1.2). DRC are comprised of a sequence of chords, each of which is a superposition of a fixed set of pure tones. The sound levels of each tone during each chord are independently drawn from a particular distribution. A central methodology advanced in this thesis is to control the statistics of these distributions. This is expounded in detail below in Section 2.1.3.1.

In different experiments, different sets of pure tones were used to build the DRCs. In experiments designed to characterise contrast-tuning curves (Chapter 3), to explore the effect of contrast on tuning and gain (Chapter 4), and to

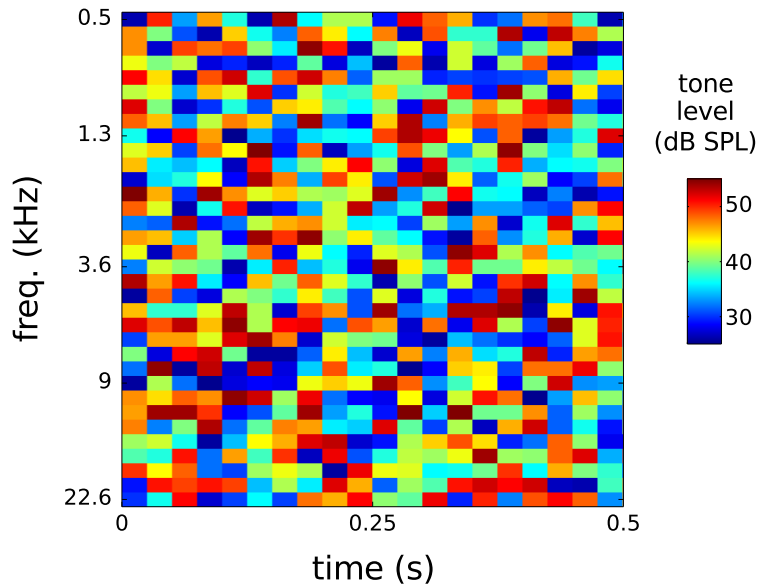


Figure 2.1: Example of a DRC stimulus. The elements of the sequence are chords of pure tones, the levels of which are randomly chosen from a particular distribution. The colour grid shows the sound level at a particular time (abscissa) of a particular tone frequency (ordinate).

investigate the elementary spectral and temporal properties of the gain effect (Sections 5.2 and 6.2), chords were built from $N_F = 34$ pure tones, with frequencies log-spaced between $f_{low} = 500$ Hz and $f_{high} = 22.6$ kHz at $1/6$ octave intervals. In experiments designed to explore gain kernels (Sections 5.4 and 6.4), and to compare IC and cortical gain functions (Chapter 7), a sparser set of pure tones covering the same frequency range was used, being built from $N_f = 23$ pure tones, with frequencies log-spaced at $1/4$ octave intervals.

It is useful to consider two views of a DRC stimulus. First, we can take a *spectral cross-section* of the DRC: at a given point in time t , the stimulus is a single chord, with different sound energies assigned to different frequency components, as described above (Figure 2.2A). From this perspective, a DRC resembles a Random Spectral Stimulus set (RSS), as used by Yu and Young [2000], Barbour and Wang [2003a,b], and Reiss et al. [2007]. However, the dynamic component of the DRC stimulus is essential. This is visible by taking a *temporal cross-section* of the DRC at a given tone frequency, f . Here, each

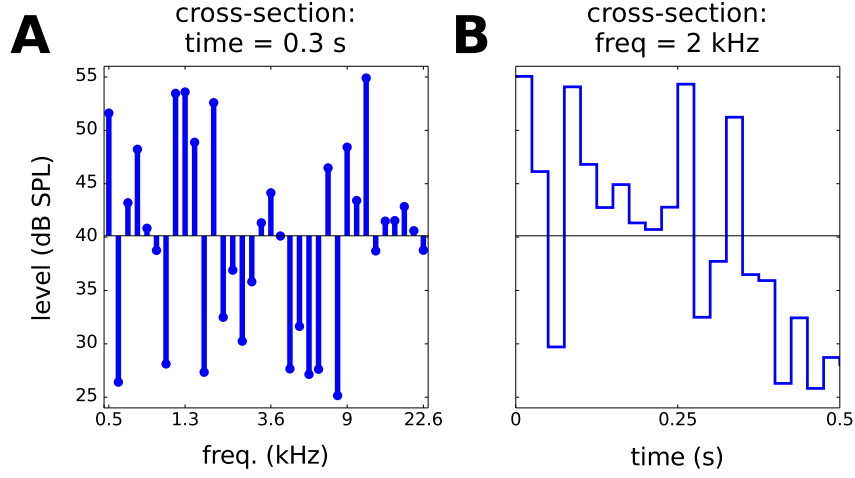


Figure 2.2: Spectral and temporal cross-sections of a DRC. (A): The spectral profile of a chord at a fixed time within the DRC shown in Figure 2.1. (B): The temporal profile, i.e. level envelope, of a single tone at a fixed frequency within the same DRC.

carrier tone is modulated by a piecewise-linear envelope (Figure 2.2B).

In the DRCs used throughout this thesis, individual chords were always 25 ms in duration, and presented in long, continuous sequences totalling 15-40 s in duration. To build such sequences, random levels were first generated for each tone in each chord. A new random seed was used for each electrode penetration and stimulus condition. Each carrier tone was synthesised in turn, and multiplied by modulation envelopes based on the random levels (described below in Section 2.1.3.1). These envelopes included 5 ms linear ramps between chords. These modulated signals were then superimposed. This process ensured that there were no amplitude or phase discontinuities in the signal, which would manifest as audible clicks.

I represent DRCs in the following way. Each DRC stimulus grid can be uniquely identified by a matrix, X_{tf} , where each component of the matrix describes the sound pressure level (in dB SPL) of a tone with frequency f at time t . For simplicity, f is considered an index over component tone frequencies $[1, \dots, N_f]$, and t an index over (25 ms) time bins $[1, \dots, N_t]$.

2.1.2 Previous use of DRCs

DRC are particularly useful in that they provide an efficient means to estimate the spectrotemporal receptive fields (STRFs) of auditory neurons, using reverse-correlation or spike-triggered averaging [deCharms et al., 1998; Schnupp et al., 2001; Rutkowski et al., 2002; Bitterman et al., 2008]. Other techniques for STRF estimation are described below in Section 2.2.1.1. The efficiency of STRF estimation derives from the fact that the levels of any two spectrotemporal bins (i.e. the elements of X_{tf}) are independent. For stimulus ensembles where this independence property does not hold (for example, if natural sounds are used to estimate the stimulus sensitivities of auditory neurons), bias correction procedures such as normalised reverse correlation [Theunissen et al., 2000] or sparse estimation protocols such as boosting [Zhang, 2005; David et al., 2007, 2009] are necessary to mitigate stimulus correlations.

It is not within the purpose of this thesis to evaluate the myriad stimulus ensembles that could be used to estimate the spectrotemporal receptive field properties of auditory neurons. Nevertheless, a commonly used alternative to pure tones and DRCs known as ripple stimuli are worth mentioning. These comprise a basis set for the space of sound spectrograms, defined in the temporal and spectral modulation domain [Kowalski et al., 1996; Calhoun and Schreiner, 1998; Klein et al., 2000; Miller et al., 2002; Escabi and Schreiner, 2002; Fritz et al., 2005b, 2007]. This modulation domain is equivalent to the 2D Fourier transform of the spectrotemporal domain. As a result, ripple stimuli can be used to estimate cells' modulation transfer function (MTF) directly, and their spectrotemporal receptive field only indirectly, via the inverse Fourier transform. As these basis stimuli do not have the spectrotemporal independence properties of random chords, estimation of STRFs from ripples can yield biased estimates of neurons' spectrotemporal sensitivities in the presence of response nonlinearities [Christianson et al., 2008]. Nevertheless, ripples allow

for a characterisation of neurons' MTFs where DRCs cannot. As shall be seen in Chapter 5, DRCs provide the opportunity to control the statistics of stimulation within and outside the spectrotemporal receptive field of a given unit, which is a fundamental reason for their choice as the stimuli used in this work.

In previous studies using random chord sequences, the DRCs have typically been *spectrotemporally sparse*, insofar as only a subset of all tones have non-zero energy (i.e. root mean square (RMS) power) at any given time. In this work, I shall depart from this form by using dense DRCs, in which all tones have non-zero energy at all times. This is necessary for constructing the tone level distributions described below in Section 2.1.3.1.

The STRFs for cortical neurons derived from dense and sparse DRCs are likely to show some differences. For neurons in A1 of owl monkeys, [Blake and Merzenich \[2002\]](#) showed that increasing the sound density of a similarly-constructed dynamic tone pip ensemble reduced sensitivity to changes in sound level, sharpened excitatory tuning, and revealed lateral inhibitory sidebands. The narrowing of cortical STRFs with density was also observed by [Valentine and Eggermont \[2004\]](#) in cat A1. Recent reports from the auditory thalamus [[Williamson et al., 2011](#)] suggest that increasing tone density makes these neurons more nonlinear in their stimulus sensitivity.

Since I use high density DRCs throughout this thesis, I am therefore studying neurons under conditions where their excitatory tuning is sharpened, lateral inhibition is present, and their stimulus-response functions are nonlinear. Since the high density of DRCs is kept constant while contrast is changed, there should be few changes to such density-dependent nonlinearities across experimental conditions, which might otherwise confound the study of contrast-dependent effects. The extent to which contrast and density interact is beyond the scope of this work. Therefore, compared with some of the simple contrast

studies previously attempted (see Section 2.1.3.3), the data presented here provide more appropriate models for the kinds of response nonlinearities that arise within acoustically challenging environments.

2.1.3 DRCs with variable contrast

2.1.3.1 Controlling the statistics of DRCs

A primary question of this thesis is whether — and how — the statistics of recent stimulation affect the coding of neurons in the central auditory system. To address this, I parametrically controlled the marginal (i.e. within-band) statistics of a set of DRCs.

The sound levels of each tone in each time bin were independently drawn from a uniform distribution, with mean level μ_L (dB SPL). The distribution was chosen to be uniform across (logarithmic) level, rather than (linear) RMS pressure, for two reasons. Firstly, this better matches the range of sound intensities and modulations present in natural signals [Escabi et al., 2003; Gill et al., 2006]. Secondly, the statistics of this distribution in the level and the pressure domains are simply related, as described below in Section 2.1.3.2.

By varying the halfwidth of the tone level distribution for a given channel, w_L , the standard deviation changes via:

$$\sigma_L = 2 w_L / \sqrt{12} \quad (2.1)$$

The range of halfwidths used differs from experiment to experiment, and is described in the respective methods of each chapter. The effect of changing the halfwidth is illustrated in Figure 2.3.

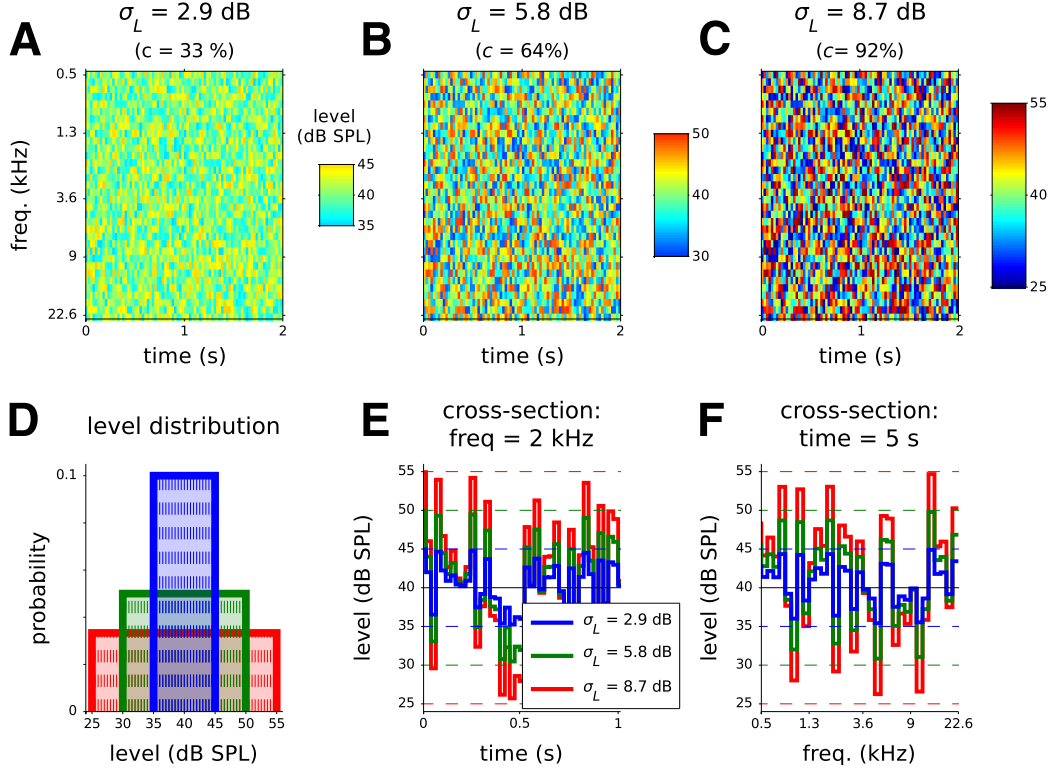


Figure 2.3: Varying the widths of DRC level distributions. (A-C): Dynamic random chord sequences, with different distribution halfwidths, w_L . The elements of each sequence are chords of pure tones, whose levels are randomly chosen from the distributions in (D). (D): Tone level distributions for the DRC sequences in (A-C). These all have the same mean level μ_L , but different widths w_L , and hence different level standard deviations σ_L . As a result of this, the DRCs have different contrasts, c . The blue region corresponds to the low contrast DRC (A); the green region is the medium contrast DRC (B); the red is the high-contrast DRC (C). (E)-(F): The widths of the level distribution determine both the spectral and the temporal contrast of the individual sequences, as shown by the temporal (E) and spectral (F) cross-sections of the DRC, as per Figure 2.2.

Several factors influenced the choice of what range of halfwidths to use. Firstly, I aimed to roughly cover the same range of statistics used in previous studies where contrast was varied. Secondly, the range of statistics tested should be roughly consistent with the range of level variations undergone in natural scenes. Although such statistics have not been extensively characterised, some relevant values are provided by Escabi et al. [2003], who measured the *long-term* statistics of sound level of some natural sounds. They measured values of $\sigma_L \approx 13$ dB for animal vocalisations, and $\sigma_L \approx 9$ dB for environmental

sounds. Given that these were long-term measurements, shorter time windows may have smaller σ_L values associated with them. More challenging acoustic environments, in turn, are likely to present even narrower distributions of sound levels. Since we know that humans are able to detect level differences on the order of 1 dB [Rabinowitz et al., 1976; Viemeister, 1983; Florentine, 1987], I therefore considered the vocalisation statistic ($\sigma_L \approx 13$ dB) as being a very high contrast sound, and $\sigma_L = 1 - 2$ dB as being a very low contrast sound. A final factor for motivating the range of σ_L values used in this work was ensuring that level fluctuations of the stimuli did not grow so loud as to deafen the animal.

2.1.3.2 Relationship between (pressure) contrast and (level) standard deviation

In order to discuss auditory contrast, a formal definition of the contrast of a sound is required. In the visual system, the contrast of a simple stimulus is defined as the ratio of the intensity difference to the mean intensity:

$$c = \Delta I / I \quad (2.2)$$

This definition can be generalised to complex stimuli as the ratio of the standard deviation to the mean:

$$c = \sigma_I / \mu_I \quad (2.3)$$

In principle, the same definitions can be applied directly in the auditory system as a function of the (RMS) pressure statistics of a sound:

$$c = \sigma_P / \mu_P \quad (2.4)$$

However, it is normal to describe sounds using sound pressure level (SPL), rather than pressure:

$$L = 20 \log_{10}(p/p_0) \quad (2.5)$$

with $p_0 = 20 \text{ }\mu\text{Pa}$, the standard reference RMS sound pressure.

The effect of this log transform is that the standard deviation of the SPL of a sound (σ_L) can provide an excellent approximation of the contrast of the sound pressure (c). This can be derived as follows. The distributions of tone levels in this study were defined as uniform distributions in the level domain, i.e.:

$$P(L) = \begin{cases} 1/(2w) & \mu_L - w_L \leq L \leq \mu_L + w_L \\ 0 & \text{otherwise} \end{cases} \quad (2.6)$$

where P denotes probability.

In the pressure (p) domain, these distributions become:

$$P(p) = \begin{cases} 1/(2\xi p) & \nu e^{-\xi} \leq p \leq \nu e^{\xi} \\ 0 & \text{otherwise} \end{cases} \quad (2.7)$$

with the transformed variables

$$\xi = \frac{\log(10)}{20} w_L = \frac{\log(10)\sqrt{12}}{40} \sigma_L \quad (2.8)$$

$$\nu = p_0 \cdot 10^{\mu/20} \quad (2.9)$$

Computing expectations gives:

$$E[p] = \frac{\nu}{\xi} \sinh(\xi) \quad (2.10)$$

$$E[p^2] = \frac{\nu^2}{2\xi} \sinh(2\xi) \quad (2.11)$$

The square of the contrast can then be computed by:

$$c^2 = \left(\frac{\sigma_P}{\mu_P} \right)^2 = \frac{E[p^2]}{E[p]^2} - 1 \quad (2.12)$$

This gives:

$$c = \sqrt{\xi \cdot \coth(\xi) - 1} \quad (2.13)$$

The relationship between the pressure contrast, c , and the level standard deviation, σ_L is plotted in Figure 2.4B. For the set of distributions used in this study, the effect of the log transform is that σ_L can provide an excellent linear approximation of c :

$$c \equiv \frac{\sigma_P}{\mu_P} \approx a \cdot \sigma_L + b \quad (2.14)$$

where a and b are constants.

For this reason, we can consider a change in the width of the tone level distributions as an effective manipulation of the contrast of the DRC stimuli.

It is worth noting that this approximation does not necessarily hold under all sets of stimulus statistics, and appears to break down as the kurtosis of the sound level distribution is increased.

2.1.3.3 Spectral and temporal contrast

Similar metrics to auditory contrast have been defined in previous work. Several studies have investigated the effects of *temporal contrast* on neural responses. In these experiments, the overall level of a stimulus was designed to change over time, while maintaining the same spectral structure. Control was then exercised over the statistics of these temporal level variations.

One simple way of controlling temporal contrast is to vary the modulation depth, m , of sinusoidally amplitude-modulated tones [Joris and Yin, 1992; Krishna and Semple, 2000; Liang et al., 2002; Nelson and Carney, 2007; Malone et al., 2007, 2010]. The size of envelope fluctuations is quantified as:

$$m = \frac{A_{max} - A_{min}}{A_{max} + A_{min}} \quad (2.15)$$

Where A_{min} and A_{max} are the minimum and maximum amplitudes respectively [Joris and Yin, 1992]. This can be measured in either the pressure domain or the level domain.

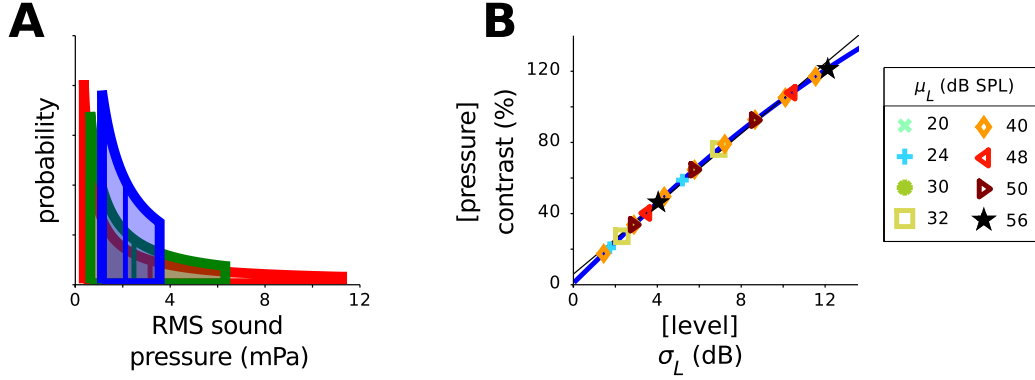


Figure 2.4: Relationship between level statistics and sound pressure statistics. (A): Distributions of tone sound pressures (mPa) corresponding to the level distributions shown in Figure 2.3. In the (logarithmic) level domain, the blue (low σ_L) distribution is uniform, with a mean $\mu_L = 40$ dB SPL and a halfwidth of $w_L = 5$ dB; the green (medium σ_L) and red (high σ_L) distributions have the same level mean but level halfwidths of 10 and 15 dB, respectively. In the (linear) sound pressure domain, as shown here, these are no longer uniform distributions, and their means are no longer identical. (B) The contrast ($c = \sigma_P/\mu_P$) of the sound pressure distribution is highly correlated with the standard deviation (σ_L) of the sound level distribution. Coloured symbols shown correspond to the range of stimulus statistics presented during this study. Blue thick line shows the analytic solution of the relationship between these variables, given by Equations (2.8), (2.9), and (2.13). The black thin line shows a linear fit between the sound pressure contrasts and level standard deviations of the presented distributions. This provides an excellent fit ($r^2 = 0.997$), indicating that these measures are interchangeable, up to a scale factor and constant offset, i.e. Equation (2.14).

More sophisticated experiments have probed the auditory system with sounds composed of pure tone or noise carriers that have been modulated by envelopes with parametrically-defined statistics. Much like the individual tone levels in the DRCs, these envelopes can be piecewise-linear in time, with values drawn from arbitrary distributions. Such envelopes thus follow dynamics similar to those depicted in Figure 2.2B. In this way, [Kvale and Schreiner \[2004\]](#) manipulated the variance and kurtosis of Gaussian and M-sequence distributions of sound levels; while [Dean et al. \[2005\]](#) controlled the first and second moments, as well as the number of modes, of a piecewise-linearly-defined level probability distribution. As a different approach, [Nagel and Doupe \[2006\]](#) convolved a noise carrier with a continuous autoregressive process of configurable mean

and variance. However, in all these studies, temporal contrast was investigated in a categorical manner, with only two levels explored. They therefore do not provide comparable metrics for contrast.

As an alternative to investigating temporal contrast, several studies have varied the *spectral contrast* of auditory stimuli, while simplifying the temporal dynamics as much as possible. In this paradigm, static chords are generated with control over the spectral statistics, i.e. the distributions over which the levels of the constituent tones can be drawn. Importantly, chords have no time-varying component: they are each presented for a set duration of time, preceded and postceded by silence. An early study by [Calhoun and Schreiner \[1998\]](#) used tone levels that were sinusoidally-modulated in (logarithmic) frequency space; their measure of contrast consisted of the spectral modulation depth. In work by [Barbour and Wang \[2003a,b\]](#) and [Reiss et al. \[2007\]](#), the spectral energy profile was randomised, with tone levels drawn from normal distributions of variable σ_L (which served as the corresponding metric for spectral contrast). The spectral profiles of these Random Spectral Stimuli (RSS) resemble the spectral cross-sections of DRCs shown in Figure 2.2, though, as in the study by [Calhoun and Schreiner \[1998\]](#), they were not presented in a rapid, sequential, dynamical fashion.

Both temporal contrast and spectral contrast have their limitations. Most importantly, all real sounds have spectrotemporal dynamics, of which the respective stimulus paradigms each miss a vital dimension. If temporal statistics alone are varied, then one does not have experimental access to the spectral sensitivities of auditory neurons. If RSS stimuli of different spectral contrasts are presented in silence, it is not clear whether or how any changes in neurons' response properties could be extrapolated to time-varying stimuli.

The DRC manipulations explored in this thesis involve control over the *spectrotemporal contrast* of sound. One previous study has used this concept. [Es-](#)

[Escabi et al. \[2003\]](#) generated spectrotemporal ripple stimuli, whose spectrograms were scaled to control the contrast. Although these stimuli had spectrotemporal correlations, unlike the DRCs used here, the marginal distributions of sound levels within any spectrotemporal bins were also uniform, with controlled σ_L . Their metric for spectrotemporal contrast thus coincides with that used in this thesis.

2.2 Models of cortical neurons

2.2.1 Linear models: STRFs

The *spectrotemporal receptive field* (STRF) is one of the most widely used models for the stimulus selectivity of auditory neurons in the midbrain and the cortex [[Aertsen et al., 1980](#); [Aertsen and Johannesma, 1981](#); [deCharms et al., 1998](#); [Theunissen et al., 2000](#); [Klein et al., 2000](#); [Miller et al., 2002](#); [Escabi and Schreiner, 2002](#); [Linden et al., 2003](#); [Fritz et al., 2003](#); [Gill et al., 2006](#); [Christianson et al., 2008](#); [Gourévitch et al., 2009](#); [David et al., 2009](#)]. Neurons in the higher auditory stations of the ferret, as for many other mammalian species, are tuned to particular stimulus frequencies, and modulate their firing rate in response to the shape of the modulation envelope, rather than the phase of the carrier [[Shamma et al., 1993](#); [Kowalski et al., 1996](#); [Bizley et al., 2005](#)]. The STRF captures these properties by modelling the time-varying firing rate of a neuron, $y(t)$ as a linear function of the spectrogram of the sound, $X(f, t)$:

$$\hat{y}(t) = k_0 + \sum_{f, h} [k(f, h) \cdot X(f, t - h)] \quad (2.16)$$

Here, and throughout this thesis, I use the hat operator, $\hat{\cdot}$, to denote an estimate; thus the STRF tries to fit $y(t)$ with $\hat{y}(t)$.

The spectrotemporal kernel, or STRF, k , describes the joint spectrotemporal features of sound that linearly drive the neuron in an optimal way. The kernel weights the sound energy in frequency bin f , at a time lag of h , to yield the predicted firing rate.

The notation is much simpler when this is expressed in tensor form. Tensors can be considered a generalisation of vectors and matrices. A vector, $\mathbf{v}$, is a 1-dimensional array that can be indexed by a single variable, such as v_i . A matrix, M is a 2-dimensional array that is indexed by two variables, M_{ij} . A tensor is an n -dimensional array, that requires n indices. Vectors and matrices are therefore treated from here on as tensors.

Equation (2.16) can be conveniently expressed in tensor form, if we first define the tensor:

$$X_{tfh} = X(f, t - h) \quad (2.17)$$

For a DRC, this relationship is illustrated in Figure 2.5. The STRF model can then be expressed via:

$$\hat{y}_t = k_0 + \sum_{f, h} X_{tfh} \cdot k_{fh} \quad (2.18)$$

The STRF model can be interpreted as defining a preferred dimension of stimulus space, k_{fh} . The firing rate is a measure of how much the history of recent stimulation matches this preferred dimension, via the linear projection of X_{tfh} onto k_{fh} .

For simplicity, all models in this thesis have been fitted using trial-averaged data, assuming a homoscedastic (i.e. constant variance) Gaussian error term $\epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2)$:

$$y_t = \hat{y}_t + \epsilon_t \quad (2.19)$$

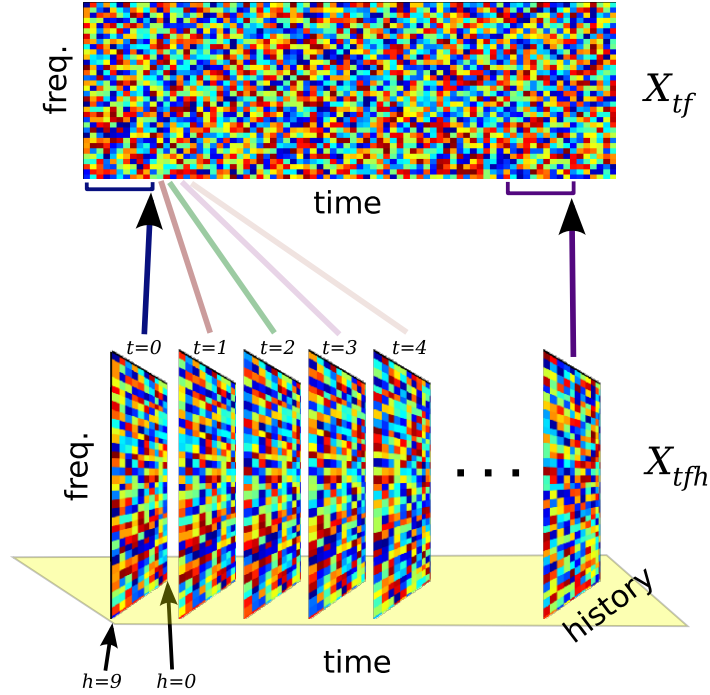


Figure 2.5: Tensor representation of a DRC. The levels of the tones in a DRC define a parameterisation of its spectrogram, X_{tf} (top). According to the STRF model, at any moment in time, a neuron’s firing rate is a linear combination of the recent history of the spectrogram. This can be visualised by considering, at any moment in time t , a “snapshot” of the last N_h time bins of the spectrogram. This forms a 2-tensor (i.e. a matrix), indexed by frequency f and history h . An example snapshot is shown at the bottom left for $t = 0$. There is a snapshot of history for each moment of time t , shown as the sequence on the bottom of the plot: as time progresses, the snapshot shifts by one time bin. For the purposes of the model, it is therefore convenient to parameterise the stimulus with an extra history index, X_{tfh} , as given by Equation (2.17).

2.2.1.1 Separable STRFs

The STRF k_{fh} , as in Equation (2.18), can be fitted via standard maximum-likelihood (ML) techniques, i.e. using least-squares linear regression to minimise an error function, such as $E = \sum_t (y_t - \hat{y}_t)^2$. However, this typically results in considerable overfitting [Sahani and Linden, 2003a; Ahrens et al., 2008]. This is particularly problematic where cortical neurons are concerned, as firing rates in response to auditory stimuli are typically noisy and unreliable [Sahani

and Linden, 2003b]. Furthermore, for the STRFs fitted in Chapter 4, the STRF model of Equation (2.18) has 341 parameters; this large parameter load is exacerbated by the experimental constraints which limit sampling of the stimulus/response space.

There are several ways of reducing the extent of overfitting. One option is to include an explicit regularisation term in the model fit, such as via an L1- or L2-norm penalty term on the individual coefficients of k_{fh} , or penalties for smoothness across neighbouring coefficients [Machens et al., 2004]. These depend on hyperparameters — i.e. parameters which determine the strength of regularisation — that must be manually chosen. Bayesian approaches, such as Automatic Relevance Determination and Automatic Smoothness Determination [Sahani and Linden, 2003a; Linden et al., 2003; Ahrens et al., 2008] circumvent this manual requirement by applying sparseness and smoothness priors to the coefficients of k_{fh} , using hyperparameters that are tuned to the data by maximising model evidence.

A third possibility is to assume that the receptive field is separable in frequency and time (i.e. history). This amounts to an approximation of the STRF kernel as the outer product of a frequency kernel k_f and a time kernel k_h :

$$k_{fh} = k_f \otimes k_h \quad (2.20)$$

For the majority of cortical neurons, separable STRFs provide as good, if not a better predictions of neural responses than full, inseparable STRFs [Linden et al., 2003; Simon et al., 2007; Ahrens et al., 2008]. Despite this model’s inability to capture inseparable frequency-time stimulus dependencies, the extent of overfitting using a separable STRF is considerably less. For the STRFs fitted in Chapter 4, there are $N_F = 34$ frequency bins and $N_H = 10$ time bins, such that the separable approximation has 45 parameters (including the offset term, k_0), rather than 341 for the inseparable STRF. The majority of these

additional parameters generally fit to the noise in the training dataset, rather than capture actual structure in data, making the model generalise poorly to new data.

Separable STRFs can be fit by constructing a low rank approximation to k_{fh} via singular value decomposition [Golub and Reinsch, 1970]. However, a more direct approach is to minimise an objective error function in an iterative manner [Ahrens et al., 2008]:

$$E = \sum_t \left(y_t - \sum_{f,h} [k_0 + X_{tfh} \cdot (k_f \otimes k_h)] \right)^2 \quad (2.21)$$

In this process, k_f is first fixed; finding the solution for k_h (and k_0) reduces to simple linear regression. Then, with k_h fixed, the solution for k_f is found in the same manner. This process is repeated until convergence.

2.2.2 Linear-nonlinear (LN) models

A fundamental limitation of an STRF model is its assumption of a linear relationship between the stimulus spectrogram and a neuron’s firing rate. While this assumption may hold in a narrow range of stimulus space, it inevitably breaks down. At minimum, rectification of a linear model is necessary to prevent the prediction of negative spike rates. Many neurons exhibit spike thresholds and firing rate saturation, which cannot be captured in simple linear models.

While it is in principle possible to extend a STRF model to include nonlinear relationships via higher-order interaction terms, this requires a very large data set [Sakai, 1992], or careful assumptions about the structure of such interactions [Ahrens et al., 2008].

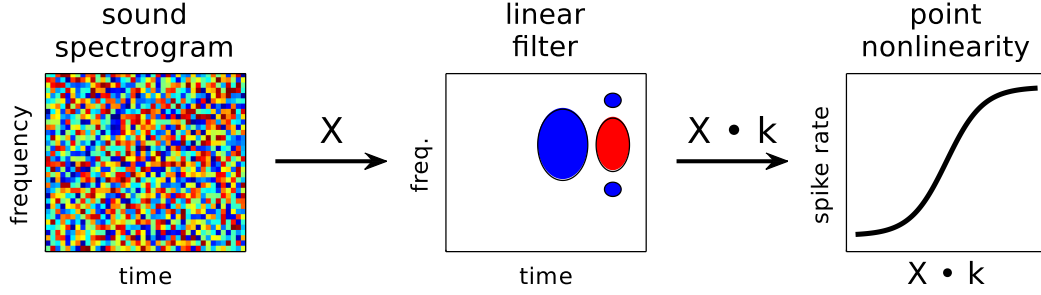


Figure 2.6: Schematic of the LN model. The relationship between stimulus and neuronal response is characterised here using a linear-nonlinear (LN) model. The sound input is treated as a spectrogram, X_{tf} ; the STRF k_{fh} acts as a linear filter on X , extracting the relevant features of the sound via the dot product. The output of the linear filter is (optionally) matched to the output spike rate through a nonlinearity that captures features such as thresholding. These nonlinearities were parameterised as sigmoids (see Section 4.2.5).

A useful method for extending linear firing rates is to pass the output of the linear model, z_t , through a static (i.e. memory-less), nonlinear function F [Chichilnisky, 2001; Simoncelli et al., 2004]:

$$\hat{y}_t = F[z_t] \quad (2.22)$$

$$z_t = \sum_{f,h} X_{tfh} \cdot k_{fh} \quad (2.23)$$

This process is illustrated in Figure 2.6. Specific methods for fitting LN models are described in Sections 4.2.5 and 5.4.1.10.

2.2.3 Other extensions to STRF models

Several extensions to the STRF have been developed in other work that do not figure in this thesis. These are briefly reviewed here.

The *context-dependent STRF* [Ahrens et al., 2008; Williamson et al., 2011] extends the linear STRF in two ways. Firstly, it employs an input nonlinearity,

rather than an output nonlinearity, such that Equation (2.18) becomes:

$$\hat{y}_t = k_0 + \sum_{f,h} F[X_{tfh}] \cdot k_{fh} \quad (2.24)$$

Secondly, it adds a (simplified) higher-order term, that allows for interactions between spectrotemporal components of the stimulus at different times and frequencies. This second-order kernel, or *context receptive field* (CRF) captures known physiological effects such as two-tone and forward suppression, and improves the predictive power of STRF models. Its successful implementation depends on choosing dimensional separabilities and introducing priors over parameters.

A reduced version of the context-dependent STRF has also been used by [David et al. \[2009\]](#) to simulate synaptic depression. Here, the context receptive field functions to dampen the magnitude of X_{tfh} , based on the previous history of stimulation in the same frequency bins. This simplifies the CRF to a two-parameter system (with parameters measuring the strength and time constant of depression), governed by a simple dynamical equation simulating a depressing synapse. This model can partly explain why using different stimulus ensembles produce certain differences in STRF shape.

From a different direction, linear-nonlinear models can be extended to Generalised Linear Models (GLMs) [[Paninski, 2004](#)]. These allow for a dynamic modification of the nonlinear output function F via a post-spike filter, which captures the dependence of a neuron's firing rate on its own recent spiking behaviour. This has been implemented most successfully in the visual system for modelling spatio-receptive fields in the retina, where it improves the predictions of single cell [[Pillow et al., 2005](#)] and population [[Pillow et al., 2008](#)] receptive field models. Recently, [Calabrese et al. \[2011\]](#) applied GLMs to auditory midbrain neurons of the zebra finch, to capture stimulus-selectivity in

response to natural sounds. While their model performed better than linear STRFs fitted using normalised reverse-correlation, this was not the result of either the static nonlinearity nor the post-spike filter. Rather, the improved predictive power of the GLM model lay in its different regularisation approach for compensating for spectrotemporal correlations in the ensemble of natural stimuli. As stimulus correlations are not an issue when using DRCs, I shall not focus on post-spike filters for the remainder of this thesis.

As a final note, there have been no systematic studies of whether input or output nonlinearities provide better extensions of STRF models for auditory neurons. Given that both of these model classes approximate a complex cascade of nonlinear filters into more simple, tractable models, each can potentially provide different insights into the dynamics of stimulus processing in the auditory pathway.

2.3 Experimental methods

2.3.1 Introduction

Chapters 3, 4, 5, 6, and 7 describe experiments that were performed using a common set of surgical and electrophysiological procedures. These methods are documented here.

All animal procedures were approved by the local ethical review committee, and performed under license from the UK Home Office in accordance with the Animal (Scientific Procedures) Act 1986. Experiments were carried out on adult pigmented ferrets (*Mustela putorius*), which were chosen for electrophysiological recordings based on normal health and hearing.

2.3.2 Surgery and recording

2.3.2.1 Cortex (anaesthetised)

Animals received initial intramuscular injection of 0.1 mL medetomidine hydrochloride (Pfizer Ltd, Sandwich UK) and 0.15 mL atropine sulfate (C-Vet Veterinary Products, Leyland UK). Their radial vein was cannulated, and a continuous infusion (4-8 mL/hr) of a mixture of medetomidine hydrochloride (0.022 mg/kg/hr), ketamine (5 mg/kg/hr), and dexamethasone (0.5 mg/kg/hr; Dexadreson, Intervet UK Ltd, Milton Keynes UK) in physiological saline containing 5% glucose was maintained over the course of the experiment. Ventilation was supplied to the animal using an Ugo Basile 7025 ventilator (Comerio, Italy), with careful, regular measurement of body temperature, inspired and expired CO₂, and electrocardiogram (ECG) signals to ensure that an adequate and stable anaesthesia was maintained throughout the experiment.

Animals were placed in a stereotaxic frame, and the temporal muscles retracted to expose the dorsal and lateral skull. A metal headbar was attached near the midline using screws and dental cement. Craniotomies were performed on both sides to expose the region over the auditory cortices (AC) in each hemisphere, demarcated by the suprasylvian sulcus and pseudosylvian sulcus [Kelly et al., 1986; Versnel et al., 2002]. The dura was removed, and the cortex covered with silicone oil to prevent dehydration. Recordings were made from the region of the middle ectosylvian gyrus where A1 and AAF are located, in both hemispheres. These areas were identified on the basis of surrounding sulcal landmarks and the frequency tuning and latencies of the units recorded there [Bizley et al., 2005].

Extracellular recordings were made using silicon probe electrodes (Neuronexus Technologies, Ann Arbor, MI), with 16 sites on a single shank, vertically spaced at 50 μm or 150 μm .

2.3.2.2 Cortex (awake)

Recordings were also performed in one awake ferret, which was passively listening to the stimuli. One month prior to the first recording session, a cranial mount was surgically implanted. This was performed as follows. Anaesthesia was induced with an intramuscular injection of medetomidine and ketamine prior to surgery. The animal was intubated, and anaesthesia was maintained with 1-2% isoflurane in oxygen-enriched air. The animal was placed in a stereotaxic frame, and the temporal muscles retracted and partially removed as for the anaesthetised animal. A craniotomy was performed to expose the auditory cortex of both hemispheres, and the cranial mount was fixed to the skull above the craniotomy. The cranial mount was made of bone cement, with a resealable metal well, as well as a metal fitting that enabled the head to be fixed to a solid recording frame. The animal was trained to accept the head restraint over a month after implant surgery, via positive food reinforcement.

Recordings were made from the awake animal using up to five quartz/platinum-tungsten electrodes (Thomas Recording, Giessen, Germany). The animal's head was restrained while the electrodes were lowered through the dura using a Mini Matrix System microdrive. Only data from recording sites located in A1 and AAF were included for further analysis.

2.3.2.3 IC (anaesthetised)

Animals chosen for IC recordings underwent the same anaesthesia protocols as for the anaesthetised cortex preparation. The skull was exposed and a stainless steel bar was attached near the midline with screws and dental cement. Craniotomies were performed over both the left and the right lateral gyri, the dura was removed, and silicone oil was applied to protect the cortical surface.

Recordings were made via a silicon probe electrode (Neuronexus Technologies) with 16 recording sites on a single shank, vertically spaced over a length of 1.5 mm. The probe was lowered through the cortex into the central nucleus of the IC. The position of the probe was confirmed by inspecting the units' frequency response areas and by the existence of the characteristic dorso-ventral tonotopic gradient.

2.3.3 Stimulus presentation

2.3.3.1 Anaesthetised

Stimuli were presented to anaesthetised animals via Panasonic RPHV27 earphones (Bracknell, UK), coupled to otoscope speculae. The speculae were inserted into each ear canal, and driven by Tucker-Davis Technologies (Alachua, FL) System III hardware. All stimuli were presented at a sample rate of 48 kHz. Earphones were calibrated using a Brüel and Kjær (Nærum, Denmark) Type 4138 1/8-inch pressure-field microphone. The headphone calibrations were used to create inverse filters, ensuring that a flat frequency output from 200 Hz to 24 kHz (correct to within ± 3 dB).

2.3.3.2 Awake

Recordings in the awake, passively listening ferret were made while presenting stimuli in the free field. This was done in an anechoic room via an Audax TWO26M0 speaker (Audax Industries, Château du Loir, France) located approximately 80 cm from the animal's head. The speaker was similarly calibrated with a Brüel and Kjær open-field Type 4191 1/2-inch microphone, placed near the position of the ferret's head within the recording setup.

2.3.4 Spike sorting

Neural signals were amplified and digitised at 25 kHz using TDT System III digital signal processors. Spike detection and sorting were performed using a custom software package which I designed and implemented in Matlab. Signals from each electrode channel were band-pass filtered between 300 Hz and 3 kHz, using an acausal, 4th order elliptic filter, with 40 dB attenuation and 1 dB pass-band ripple. Candidate events were identified when the signal on a channel i , x_{it} , crossed a (negative) threshold, T_i , defined as:

$$T_i = -4 \hat{\sigma}_i \quad (2.25)$$

where $\hat{\sigma}_i$ is an estimate of the standard deviation of the background noise on that channel [Donoho and Johnstone, 1994; Quiroga et al., 2004]:

$$\hat{\sigma}_i = \text{median} \left(\frac{|x_{it}|}{0.6745} \right) \quad (2.26)$$

The time t_j of a candidate event j was set as the sample when the signal on that channel reached an absolute minimum within the 1 ms following the threshold crossing. As single spike events were typically visible across many channels, especially for the 50 μm -spaced electrode configuration, spike sorting was performed using information about the spike shape across all channels. Each event j was then represented as a matrix E_{itj} , with i as the channel index, ranging across the 16 channels of the probe, t the time index spanning a 2 ms duration from 0.5 ms prior to 1.5 ms after t_j , and j the event index.

Next, a low-dimensional representation of each candidate event was constructed. For each channel, principal component analysis (PCA) was performed on the set of spike shapes. The first three principal components for each channel (v_{itk} for $k \in [1, 2, 3]$) were collected as a basis set, and candidate events were projected onto the set of 3×16 basis vectors. This yielded a 48-dimensional feature-space representation of each candidate event, denoted here as F_{itk} .

Clusters in feature space were fitted with a mixtures-of-Gaussian model. This was performed by KlustaKwik (K.D. Harris, Rutgers, NJ), a fast implementation of an Expectation Maximisation (EM) algorithm, with an Akaike Information Criterion (AIC) penalty term to constrain the number of clusters.

It was found that this process tended to group event collisions (i.e. when two or more spike-like events occurred simultaneously on well-separated channels) in separate clusters. To circumvent this problem, the feature-space representation of each event was limited to channels local to the channel j on which the spike amplitude was greatest. This process required specifying a set of replacement values for the spatially-distant coefficients, as well as a threshold distance δ , beyond which the coefficients of F_{itk} for that event would be ignored or replaced.

Zeroing the spatially-distant coefficients of F_{itk} (i.e. those for which $|k - j| \geq \delta$) biased the EM algorithm towards producing exactly 16 clusters, whereby cluster j contained all and only those spikes whose spike amplitude was greatest on channel j . This was a result of the EM algorithm converging to a mixture model whose components had the infinitely precise covariance structure that was imposed by the zeroing process. A better solution was to draw replacement values for the spatially-distant coefficients of F_{itk} from a multivariate Gaussian with a diagonal covariance matrix, whose marginal distributions were moment-matched to the empirical statistics of the observed ensemble of F_{itk} values. For the 50 μm electrode spacing configuration, a heuristic balance between the spatially-distributed nature of observed events and collision-avoidance was found at $\delta = 3$. Thus I approximated the 16-channel electrodes as a set of local pentatodes.

Finally, cluster separation was visually inspected to ensure that the automated clustering yielded well-separated clusters. Clusters with similar event shapes and event autocorrelations were merged, guided by information on cluster separability from the feature-space representation. As many electrode penetrations

needed to be maintained for long durations of time for several of the experiments presented in this thesis, particular care was taken to ensure that spiking statistics were stationary. Over the course of 1-2 hours, it was often the case that average spike rates would drift due to mechanical or physiological instability. Where such drift was considerable, only the responses during the largest stationary period were considered in the analysis of that unit.

Clusters were deemed to be single units based on spike shape and the presence of a refractory period in the autocorrelation history, and were otherwise labelled as multi-units.

2.3.5 Unit selection criteria

2.3.5.1 Acoustically-responsive activity

Only units showing acoustically-responsive activity were included. This was assessed at the beginning of each electrode penetration, by presenting a stimulus consisting of repeated short bursts of white noise at 70 dB SPL. The noise was presented in alternating fashion from the left earphone, the right earphone, then bilaterally. If units at this location did not respond to the noise bursts with a well-defined peri-stimulus time histogram (PSTH), the electrode was retracted and a new site chosen for further recording.

2.3.5.2 Noise ratio criterion

The analyses carried out in this work can only be applied to acoustically driven units that produce reasonably reliable, repeatable responses. The reliability of a unit can be measured by estimating the *signal power*: the part of the total response variance (the *total power*: the average variance in a single-trial PSTH) that is reliably stimulus-locked [Sahani and Linden, 2003b]. The signal power

captures how much of the variance in responses across stimuli (or time) can be explained, given the variance in responses across multiple repeats of the same stimulus. In turn, the *noise power* is an estimate of the residual variance:

$$\text{total power} = \text{signal power} + \text{noise power} \quad (2.27)$$

The reliability of response can then be measured by the *noise ratio*:

$$\begin{aligned} \text{noise ratio} &= \frac{\text{noise power}}{\text{signal power}} \\ &= \frac{\text{total variance} - \text{explainable variance}}{\text{explainable variance}} \end{aligned} \quad (2.28)$$

When signal power is equal to the total response power, the noise ratio is 0, indicating that responses were identical for repeated stimulus presentations. Higher noise ratio (equivalently, lower signal power) indicates that responses were less reliable. In each experiment, an exclusion criterion was asserted, so that analyses were performed only those units with sufficiently reliable responses.

Chapter 3

Contrast tuning

3.1 Introduction

One of the main reasons why we are able to have sensory systems at all is that sensory space is inherently structured. For auditory signals (as for visual and somatosensory signals, for example), much of this organisation occurs along continuous dimensions: sounds vary from loud to soft, high to low frequency, short to long periodicity, and from left to front to right to back, to name a few.

A key strategy for managing a high-dimensional stimulus space is neuronal specialisation. The most notable specialisation in the auditory pathway is frequency tuning, whereby neurons from the cochlea through to the primary auditory cortex respond preferentially to sounds in a particular frequency range [Merzenich and Reid, 1974; Merzenich et al., 1975; Kelly et al., 1986; Bizley et al., 2005]. Another example is found with sound localisation: the mammalian brain decomposes the spatial dimension of incoming sounds through highly specialised neurons in the lateral (LSO) and medial (MSO) nuclei of the superior olive. The firing of LSO neurons provide a rate code for interaural level differences [Tollin and Yin, 2002; Park et al., 2004], and MSO neurons a rate code for interaural time differences [Brand et al., 2002; McAlpine, 2005].

Sound intensity is also a target of specialisation: auditory nerve fibres exhibit range fractionation, whereby neurons distribute their limited dynamic ranges across a huge intensity spectrum, allowing for specialised representation within each of these intensity domains [Nomoto et al., 1964; Sachs and Abbas, 1974; Winter et al., 1990; Winter, 1991]. It has also been variously suggested that neurons in the central auditory pathway are specialised for periodicity [Patterson et al., 2002; Bendor and Wang, 2005; Bizley et al., 2009], or for spectral or temporal modulation frequency [Bacon and Grantham, 1989; Dau, 1997; Miller et al., 2002; Joris et al., 2004; Woolley et al., 2005, 2006; Rodriguez et al., 2010; Baumann et al., 2011].

Broadly, this strategy is useful as there may be unique computational demands for processing stimuli from different regions of a stimulus dimension. For example, extracting the spatial source of low frequency sounds from binaural information requires a fundamentally different strategy (interaural time differences) from estimating the source of high frequency sounds from binaural information (interaural level differences); together with the monaural processing of the dorsal cochlear nucleus [May, 2000; Grothe et al., 2010] the mammalian auditory system thus has three independent pathways for sound localisation.

To provide a full account of how the auditory system approaches sounds of different contrasts, it is instructive to consider whether cortical neurons are specialised for representing particular regions of contrast space.

As introduced at the start of this thesis (Section 1.4.1), the question of contrast specialisation has been raised in a few other systems. Escabi et al. [2003] presented a ripple stimulus to anaesthetised cats while recording from the central nucleus of their inferior colliculus (IC), and varied the spectrotemporal contrast of the ripples. They found that approximately half of IC neurons increased their average firing rate monotonically with stimulus contrast, while the

remainder had a lower average firing rate at the highest contrast tested. Thus IC contains a mixture of neurons with monotonic and non-monotonic *contrast tuning curves*. Nonmonotonic contrast tuning curves were not, however, found by [Krishna and Semple \[2000\]](#), who merely controlled the temporal contrast of sinusoidally amplitude modulated (SAM) tones, while recording from the central nucleus of the IC of the mongolian gerbil. They reported that at any given modulation frequency, a neuron’s firing rate either monotonically increased, or monotonically decreased, as a function of the contrast (i.e. modulation depth) of the SAM tone. A number of other studies have similarly reported that response modulation monotonically increases with modulation depth for neurons in the auditory nerve [[Joris and Yin, 1992](#)], inferior colliculus [[Rees and Møller, 1983](#); [Nelson and Carney, 2007](#)], and primary auditory cortex [[Eggermont, 1994](#); [Liang et al., 2002](#); [Malone et al., 2007, 2010](#)]. In a final study, [Barbour and Wang \[2003b\]](#) presented random spectral stimuli (RSS) to the awake marmoset, while recording from primary auditory cortex. They investigated how the firing rate of these neurons varied as a function of the spectral contrast of the chord stimuli. They reported a heterogeneity of contrast-response functions, including low-contrast-preferring neurons (i.e. with monotonically decreasing contrast-response functions), and high-contrast-preferring neurons (i.e. with monotonically increasing contrast-response functions).

Given the lack of agreement amongst this prior work, I sought to characterise whether neurons in ferret primary cortex respond preferentially to chords and chord-sequences of particular contrasts. The contrast values were more finely spaced than used in previous studies, allowing me to fit parametric contrast-response curves to the measured data. I thereby amassed evidence as to whether different types of contrast-response functions are exhibited by these neurons.

3.2 Methods

3.2.1 Stimuli

Two sets of stimuli were presented in this study. Firstly, dynamic random chords (DRCs), as described in Section 2.1.1, were constructed to explore how neural response properties depend on the (spectrotemporal) contrast of *ongoing* stimuli. Secondly, single short chords, or “test sounds”, were specially designed for units at each recording location, and presented in isolation, in order to explore how neural response properties depend on the (spectral) contrast of *transient* stimuli. These are now described in more detail.

3.2.1.1 DRC stimuli

Each chord in the DRCs was a superposition of 34 pure tones, with frequencies log-spaced between 500 Hz and 22.5 kHz at 1/6 octave intervals. The levels of each tone during each chord were independently drawn from a uniform distribution, with a mean level of $\mu_L = 40$ dB SPL. The halfwidth of the distribution was varied from $w_L = 2.5$ dB to $w_L = 20$ dB in 2.5 dB steps (i.e. σ_L from 1.4 dB to 11.5 dB, and contrast, c , from 17% to 116%; see Equations (2.1), (2.8), (2.9) and (2.13)). Chords were each 25 ms in duration, and presented in sequences of 15 s duration. The overall RMS level of the stimuli ranged from 71 to 77 dB SPL. Each DRC sequence was presented 10 times, randomly interleaved, with 15-20 s silence between each sequence.

Units’ responses to the DRCs, aside from being analysed in terms of contrast tuning (as explained below in Section 3.2.2), were also used to estimate spectrotemporal receptive fields (STRFs). Separable STRFs were fitted for each neuron by minimising the objective function given in Equation (2.21) (see Section 2.2.1).

3.2.1.2 Test sound stimuli

Test sounds were constructed for each electrode penetration. A representative unit was chosen for that location, and its STRF estimated online. The frequency component of the STRF, k_f , approximates the spectral profile of a chord that would drive the neuron best, if presented within the context of an ongoing DRC. k_f was therefore scaled, and noise added, to create an effective single-chord stimulus: a test sound template that would be close to the optimal chord, while having extra variation in sound level across neighbouring frequency bands. This was done as follows. k_f was normalised to have zero mean and unitary standard deviation. Uniformly distributed random noise (with zero mean and a halfwidth of 0.3) was added to the coefficients of k_f . The tone levels represented by this vector were then offset to have a mean of 40 dB SPL, and scaled to 12 (spectral) contrasts: w_L from 2.5 dB to 30 dB in 2.5 dB steps (i.e. σ_L from 1.4 dB to 17.3 dB, and contrast c from 17% to 157%). From these level vectors, 12 test sounds were generated, using the same set of component tones as used in the DRCs.

As a control, I also generated a set of chords whose spectral profile was inverted relative to this template above. The spectral profile of this “anti-test sound” was scaled to yield 12 further stimuli, in the same way as the test sound. Finally, a chord of zero spectral contrast (i.e. all tones presented at 40 dB SPL) was generated, yielding a total of 25 chords for each electrode penetration. The set of transient stimuli for an example electrode penetration is illustrated in Figure 3.1.

In total, this stimulus set constitutes a subset of a Random Spectral Stimulus (RSS) corpus as used by [Yu and Young \[2000\]](#), [Barbour and Wang \[2003a,b\]](#), and [Reiss et al. \[2007\]](#) (also described in Section 2.1.1). Here, the RSS sounds

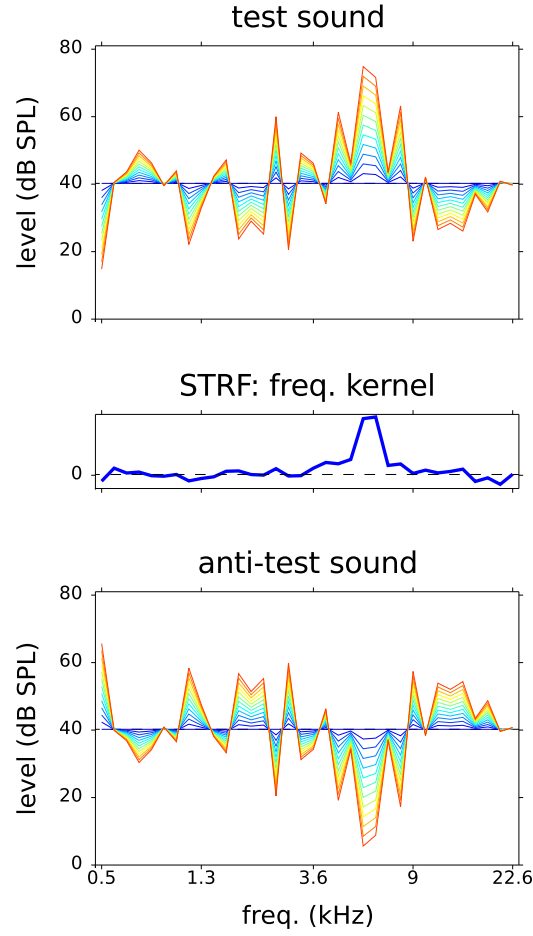


Figure 3.1: Schematic of test sound construction. A test sound (TS) was generated for each electrode penetration, comprising a single chord of 50 ms duration, with a spectral profile similar to the frequency component of a representative unit’s STRF, k_f (middle panel). The levels of the tones in the chord were scaled to have a fixed mean level, but varying contrasts (upper panel; from low contrast in blue to high contrast in red). A similar set of anti-test sounds (ATS) was also generated, with an inverted spectral profile relative to the test sounds (bottom panel).

were chosen to be particularly matched to the frequency specificity of neurons within each electrode penetration.

Envelopes for the test sounds were 50 ms in duration, with 5 ms cosine on-set/offset ramps. Each TS was presented 35-40 times in random order, embedded in silence. Test sounds were presented irregularly, with an inter-stimulus interval randomly chosen from 250, 275, 300, 325, or 350 ms, and the responses to each TS aggregated. The same sounds were also presented embedded within

a DRC of intermediate contrast ($w_L = 10$ dB; $\sigma_L = 5.8$ dB; $c = 64\%$).

3.2.2 Response metrics

Two simple metrics were used to measure units' responses to the DRC stimulus: the mean firing rate, and the signal power (i.e. stimulus-locked variance; see Section 2.3.5.2 and [Sahani and Linden, 2003b]). The first 2 s of data from each stimulus presentation were discarded to ensure that a constant adaptation state had been reached.

Responses to the transient test sound were measured as a unit's mean firing rate over an appropriate response window. This window was determined as follows. The unit's responses to all the test sounds were aligned, and a post-stimulus time histogram (PSTH) with 5 ms bins was calculated over the pooled responses. I measured the mean (μ_r) and standard deviation (σ_r) of the baseline firing rate over the 100 ms prior to each stimulus presentation, and estimated the response interval as beginning from when the average PSTH first exceeded $\mu_r + 2\sigma_r$ after stimulus onset, until when it first dropped below $\mu_r + \sigma_r$. As for the DRCs, responses to the first 2 s of TS stimulation in each block were discarded.

3.2.3 Unit selection criteria

For the DRC stimuli, I only included units which modulated their firing rate in response to the DRC. This was asserted by excluding units whose noise ratio (see Section 2.3.5.2) was greater than 40 for any one condition.

For the transient stimuli, only those units whose response to the test sound differed from the unit's spontaneous rate were included. This was assessed by comparing the number of spikes evoked during the response interval defined

above (pooled over presentations of test sounds of all contrasts) against the number of spikes evoked during an interval of the same duration immediately prior to each test sound presentation. Two models for these spike counts were then compared: firstly, that both pre- and post-stimulus spike counts were generated from a Poisson process of the same mean; secondly, that these two sets of counts were generated from two Poisson processes of different means. Model comparisons were performed using the Bayesian Information Criterion (see Section 3.2.5). Only those units for whom the latter model was selected were retained for further analysis.

While many units satisfied the above criterion for the transient stimuli (i.e. they responded, on average, to the test sounds), their responses were either insufficiently modulated by contrast, or there was too much trial-to-trial variability in responses to fit putative contrast-response relationships. This was assessed by calculating the noise ratio (NR) for the set of responses to trial stimuli, with a cutoff for inclusion of $\text{NR} = 40$.

3.2.4 Models

The primary question in this study was whether neural responses showed evidence of preference (i.e. “tuning”) for particular stimulus contrasts. To assess this, I considered the relationship between stimulus contrast and each of the above response metrics. In each case, I considered four different candidate functional relationships between stimulus contrast and neural response:

Null model (N):

$$\hat{y} = \mu_y \tag{3.1}$$

Monotonically Increasing model (MI):

$$\hat{y} = a + \frac{b}{1 + e^{-(x-c)/d}} \quad b, d > 0 \tag{3.2}$$

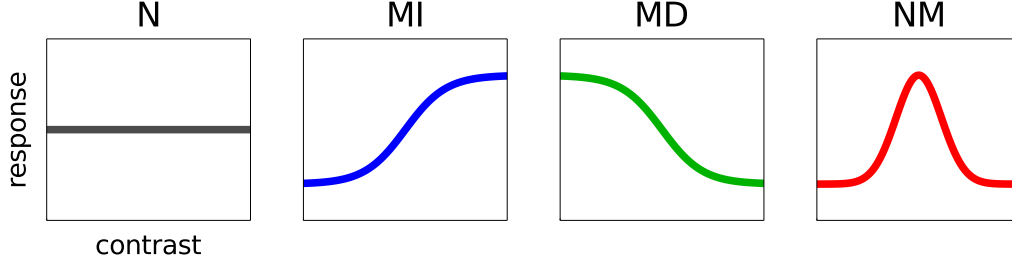


Figure 3.2: Idealised contrast-response relationships, as defined by Equations (3.1) – (3.4). N (Null) units have responses which are independent of contrast. MI (Monotonic Increasing) and MD (Monotonic Decreasing) units have sigmoidal contrast-response relationships. NM (Non-Monotonic) units have a Gaussian contrast-response relationship.

Monotonically Decreasing model (MD):

$$\hat{y} = a + \frac{b}{1 + e^{(x-c)/d}} \quad b, d > 0 \quad (3.3)$$

Non-Monotonic (Gaussian) model (NM):

$$\hat{y} = a + b \cdot \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right) \quad b, \sigma > 0 \quad (3.4)$$

These models are illustrated in Figure 3.2.

To weight against the chance of the fitted models merely attempting to capture single outliers (especially at the edges), I placed constraints on several parameters. For Equations (3.2) and (3.3), a lower bound on d such that $\frac{d}{\Delta x} > 0.2$ was enforced, where Δx is the minimum difference between consecutive x values. This prevented these models from behaving as step functions. Similarly, for Equation (3.4), the constraint $\sigma > \Delta x$ was enforced, preventing this model from behaving as a delta function to fit a single outlier. For Equation (3.4), μ and σ were constrained so that at least one standard deviation on either side of the mean lay within the range of x values. This was necessary to ensure that the Gaussian functional form would only be suitable when there was clear

evidence for the existence of regions of both increasing and decreasing values of the response metric with respect to changing contrast.

Given these constraints, maximum likelihood values of the parameters in each of the models (save for N) were fitted using gradient descent, assuming normally-distributed errors. This procedure was repeated 20 times, with random initial conditions.

3.2.5 Model comparisons

Model comparison was performed using the Bayesian Information Criterion (BIC; [Schwarz, 1978]). The BIC for model M_i , assuming normal residuals, is given by:

$$\begin{aligned} BIC_i &= -2 \log L_i + k_i \log n \\ &= n \log(\sigma_{e_i}^2) + k_i \log n + \text{const}(n) \end{aligned} \tag{3.5}$$

where $L_i = P(y|M_i; \theta)$ is the (maximum) likelihood under model M_i , k_i is the number of parameters in model M_i , n is the number of data points, and $\sigma_{e_i}^2$ is the error variance. The BIC balances model fit (the likelihood term) with a penalty for model complexity. A good model has a high likelihood (so $-2 \log L_i$ is small), using few parameters (so $k_i \log n$ is also small); hence the best model is the one that minimises the BIC. In the figures of this chapter, I plot the negative of the BIC (which aids interpretation): higher values of $-\text{BIC}$ denote better fits.

It is possible to fit relationships between contrast and neural response using fewer parameters (e.g. a linear model for a monotonic relationship; quadratic or piecewise-linear for a non-monotonic relationship). However, I chose the form of the MI, MD, and NM models to all have four parameters, so that their model complexities (and hence the penalty terms in Equation (3.5)) were identical. Selection between these three models thus reduced to a comparison of their log

likelihood terms, i.e. their error variances. However, when assessing whether there was evidence for any of these models above the Null model, the penalty term was relevant.

The BIC is a rough approximation to the (logarithm of the) Bayes factor, a general measure of the evidence in favour of a model [Kass and Raftery, 1995]. I use Jeffrey’s guide [Jeffreys, 1961] of a Bayes factor of 10 as strong evidence of one model in favour of another. I also assume equal prior probabilities for each of the models M_i . Finally, I favour parsimony. In my analysis, I considered two nested null hypotheses: first, that the Null model holds; second, that the MI model holds (for which there is considerable prior evidence). Comparing two models amounts to computing the difference in their BICs, with a criterion for selecting a more sophisticated model of:

$$\Delta BIC > 2 \log(10) \quad (3.6)$$

Ideally, the Bayes factor would be computed analytically. However, this involves evaluating an integral which is intractable for the models considered here. As the BIC is only one simple (though easily calculated) approximation to the Bayes factor, I cross-checked the results here using an alternative approximation, Laplace’s method [Kass and Raftery, 1995]. This produced similar results to the BIC, albeit more conservative: the Laplace approximation favoured the MI model over the NM model to a greater degree than when using the BIC comparison.

3.3 Results

I first recorded from 48 electrode penetrations in the primary auditory cortex (A1) and anterior auditory field (AAF) of five anaesthetised ferrets, while diotically presenting dynamic random chord (DRC) sequences of different contrasts.

Sequences comprised chords which changed every 25 ms, with the levels of the constituent tones drawn from uniform distributions in SPL space. The spectrotemporal contrast of the sequences was manipulated by changing the (dB) width of these distributions. The tone level distributions had identical means ($\mu_L = 40$ dB SPL), but halfwidths ranging from $w_L = 2.5$ dB ($\sigma_L = 1.4$ dB; $c = 17\%$) to $w_L = 20$ dB ($\sigma_L = 11.5$ dB; $c = 116\%$) in 8 steps (see Figure 2.3).

I identified 75 units which responded reliably to the DRCs, as measured via a maximum noise level criterion (see Section 3.2.2). These had best frequencies (BFs) that were roughly evenly spread between 3 kHz and 22 kHz.

3.3.1 Mean firing rate and stimulus contrast

The question motivating this study is whether neurons exhibit a preference for sounds of particular contrasts. The first assessment I made was of the nature of the relationship between the contrast of an ongoing stimulus and mean firing rate. This relationship could take on four forms: there could be no observable dependence of mean rate on contrast (Null model (N), Equation (3.1)); mean rate could increase monotonically as a function of contrast (Monotonically Increasing model (MI), Equation (3.2)); mean rate could decrease monotonically as a function of contrast (Monotonically Decreasing model (MD), Equation (3.3)); or there could be a non-monotonic dependence of rate on contrast (Non-Monotonic model (NM), Equation (3.4); see Figure 3.1). I determined the most explanatory model via the Bayesian Information Criterion (BIC): the model with the lowest BIC best described the relationship between mean rate and contrast.

Of the 75 units, the vast majority (68/75) increased their firing rate monotonically with contrast (Figure 3.3A). Of the remainder, 3/75 showed evidence of non-monotonicity, while 4/75 did not show a consistent relationship between

mean rate and contrast (and were therefore best fitted by N). No units were best described by the monotonically decreasing model. Amongst all these units, only a single one showed strong evidence against the MI model, in favour of MD or NM (Figure 3.3B-D). In terms of firing rate, most DRC-responsive neurons thus prefer high contrast stimulation.

3.3.2 Signal power and stimulus contrast

Collapsing neural responses over long time periods to mean firing rate provides only a weak measure of a neuron's preference for different temporally-varying stimuli. In particular, mean firing rate does not relate the extent to which a neuron modulates its firing rate to changes in the stimulus. I thus considered a measure of the stimulus-locked variability of neural responses over time under different contrast conditions, the signal power (SP; see Section 2.3.5.2).

As for the mean firing rate, signal power increased monotonically with contrast for almost all units (72/75; Figure 3.4A). Only one unit showed strong evidence in favour of non-monotonicity (Figure 3.4B-D). The similarities between the dependence of mean rate and signal power on contrast can be attributed to these two measures being correlated, both across units (Figure 3.4E), and within units (Figure 3.4F).

While neurons' signal power generally increased with contrast, their noise power, as a whole, did not change with contrast.

In total, these data suggest that neurons in ferret primary auditory cortex generally operate more effectively under higher contrast stimulation.

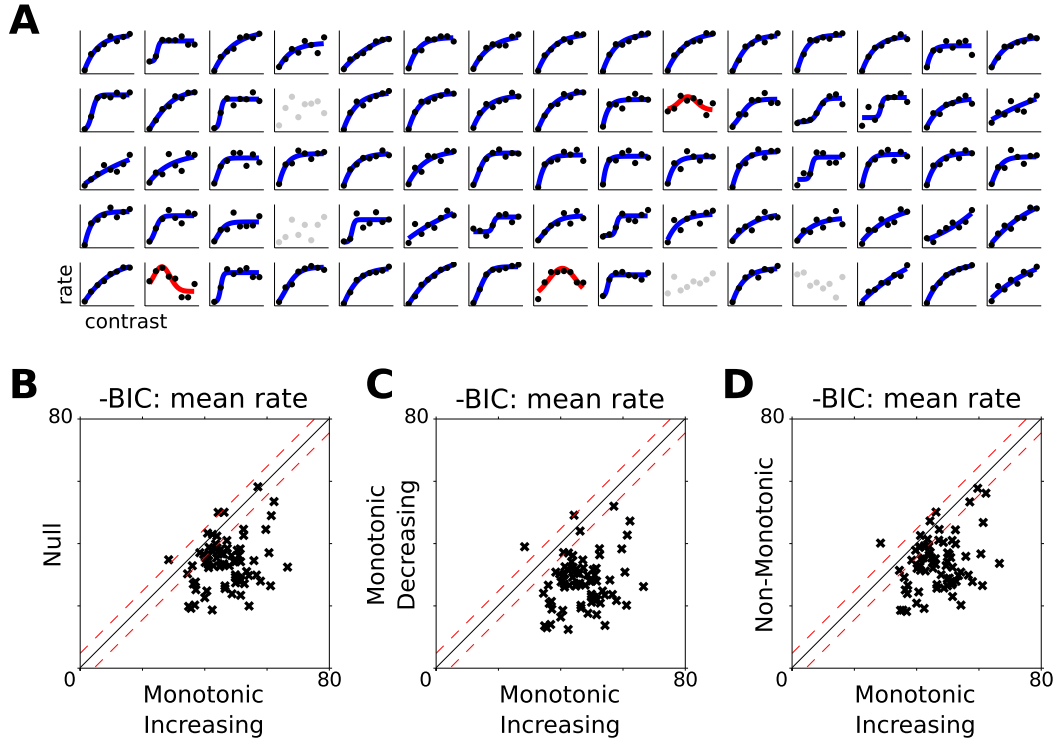


Figure 3.3: (A): Mean firing rate as a function of contrast for 75 units. As the spatiotemporal contrast of the DRCs was increased, the majority of units increased their firing rate monotonically. Black dots denote measured values; model of best fit indicated by solid line of different colours. Blue: Monotonically Increasing model (MI; Equation (3.2)). Green (none visible): Monotonically Decreasing model (MD; Equation (3.3)). Red: Non-monotonic (tuned) model (NM; Equation (3.4)). Where the null model provided the best fit, no curve is drawn and the measured response rates are shown in grey. Axes have been scaled and positioned to emphasise the shape and trend of the contrast-response relationship, rather than the absolute values of the measured response. (B-D): Comparison of how well several models capture the contrast-response relationship of each neuron, as measured by the Bayesian Information Criterion (BIC). The negative of the BIC has been plotted, so that higher values of $-BIC$ denote better fits. Black diagonal line shows equal BIC (i.e. equal evidence for the two models). For example, in (B), points below the diagonal indicate neurons where there is more evidence for the MI model than for the N model. Red dashed lines mark a difference of BICs of $2\log(10)$, indicating strong evidence in favour of a particular model. For simplicity of visualisation, the $-BIC$ of the MI model has been plotted as a reference for all comparisons. For the majority of units, there is strong evidence in favour of a monotonically increasing relationship between contrast and mean firing rate. There is only one unit for which there is strong evidence for non-monotonicity in its firing rate. Incidentally, this unit appears above the diagonal line in both (C) and (D), but is better fit by the NM model than the MD model.

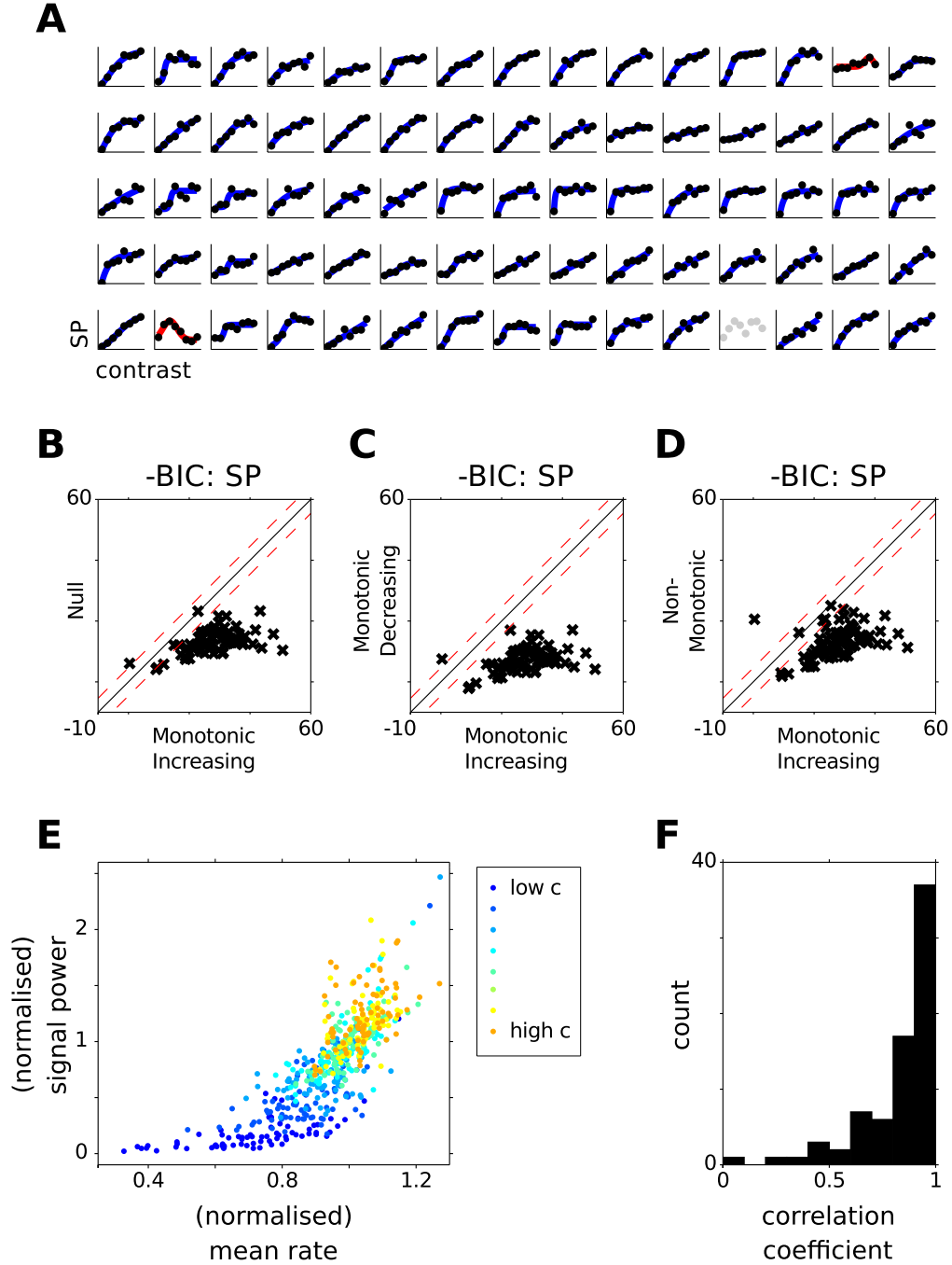


Figure 3.4: As the spectrotemporal contrast of the DRCs was increased, the majority of units increased their stimulus-locked variance monotonically. (A): Relationship between signal power (SP) and contrast for the 75 neurons shown in Figure 3.3; colours and lines as in Figure 3.3A. (B-D): Comparison of models, as in Figure 3.3B. There is only strong evidence for 1/75 units preferring a lower contrast; this unit was best described by the non-monotonic model. (E): Signal power and mean firing rate were correlated with each other ($r = 0.81$). Data are shown for all units and contrast conditions. (F): Histogram of correlation coefficients between SP and mean rate across stimulus conditions for each individual neuron.

3.3.3 Transient stimuli

I next asked whether neurons showed a preference for the contrast of transient stimuli. To a further three ferrets, I presented an additional set of stimuli. In each electrode penetration, I designed a special “test sound” (TS) that would drive the neurons in that location well. The TS was a chord of 50 ms duration, comprising tones identical to those used in the DRCs, and with spectral structure that resembled the frequency component of the spectrotemporal receptive fields (STRFs) of a representative unit from the penetration. In turn, the chord was scaled to have a range of spectral contrasts, with halfwidths from $w_L = 0$ dB ($\sigma_L = c = 0$), to $w_L = 30$ dB ($\sigma_L = 17.3$ dB, $c = 157\%$; Figure 3.5A). Test sounds were presented at jittered intervals of ~ 300 ms, in silence.

Of 248 units, 224 responded to the test sound. Amongst these, only 118 varied their responses sufficiently — and sufficiently reliably — with the contrast of the TS to assess the contrast-dependence of their response (see Section 3.2.3).

As was observed with ongoing stimuli, the vast majority of units (108/118) responded more strongly as the contrast of the stimulus was increased, in a monotonically increasing fashion (Figure 3.5A). None of the units showed a monotonically decreasing response as a function of contrast. 7/118 had non-monotonic relationships between their evoked spike count and contrast, of which 5 showed strong evidence against the MI model (Figure 3.5D, black crosses). The remaining 3 units were best described by the Null model.

I also presented these same test sounds, while embedded in ongoing DRCs, to a subset of 205 units. However, amongst these, only 34 responded differently to the TS compared with the ongoing sound. In turn, only 14 of these sufficiently varied their responses to the TS in a contrast-dependent fashion to be

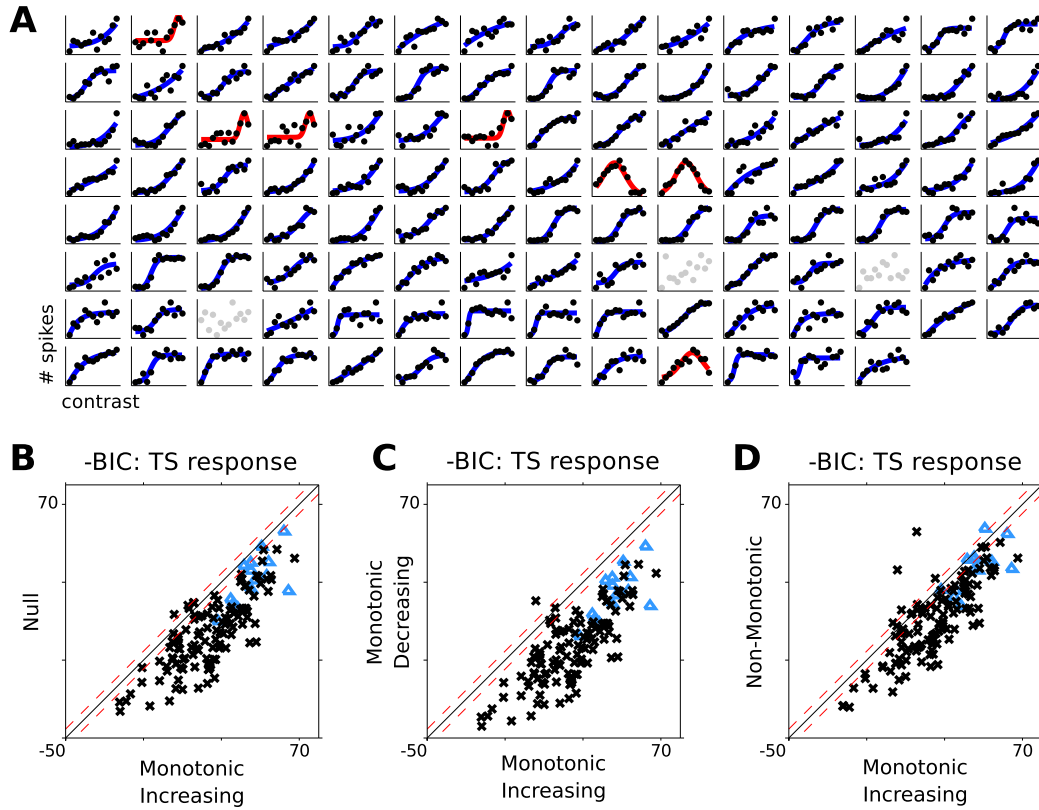


Figure 3.5: As the spectral contrast of a transient stimulus was increased, the majority of units increased their firing rate monotonically. (A): Response firing rate elicited by the test sound for 118 units. Colours and lines as in Figure 3.3A. (B-D): Comparison of model fits. Black crosses show model fits to responses when the test sound was embedded in silence (as shown in (A)); blue triangles show the fits when the test sound was embedded in an ongoing DRC. Only $\sim 5\%$ of units show strong evidence for a non-monotonic contrast-response relationship. No units were best fit by the monotonically decreasing contrast-response relationship.

reliably modelled. Nevertheless, amongst these, the same pattern was observed as with stimuli presented in silence: the majority (12/14) increased their responses monotonically with contrast; the remaining two were non-monotonic (Figure 3.5B-D, blue triangles). Thus auditory cortical neurons generally increase the strength of their response to transient stimuli as the contrast of these stimuli increase, regardless of whether they are presented in isolation, or as part of ongoing stimulation.

Finally, I considered whether this monotonic dependence on stimulus con-

trast extends to sounds which are not spectrally matched to the preferences of the neurons being recorded. I therefore presented to these neurons a set of “anti-test sounds” (ATS), whose levels were inverted relative to the test sounds. As the contrast of these sounds was increased, so the amount of energy near the best frequency of the neurons typically decreased with increasing contrast (though energy at non-preferred frequencies still within the tuning of the neuron may have either increased or decreased). Responses to the ATS were considerably weaker on average than to the TS, and the responses of only 85 units reliably varied as contrast was changed. Amongst these, most units (52/85) showed no smooth relationship between the ATS contrast and response. Amongst the remainder, the nature of the relationship between response strength and contrast was mixed: about half (16/33) monotonically increased with contrast; 6/33 monotonically decreased; and 16 responded most to intermediate contrasts.

3.4 Discussion

I measured whether neurons in the primary fields (A1/AAF) of the auditory cortex of ferrets are specialised for encoding auditory stimuli of low, high, or intermediate contrasts. While not all neurons exhibit contrast-sensitive response properties, amongst those that do, almost all share the same contrast-response relationship: responses become stronger as stimulus contrast is increased. When the spectrotemporal contrast of ongoing sounds is increased, more than 90% of contrast-dependent neurons monotonically increase their mean firing rate and stimulus-locked variance. When probed with short sounds whose spectral content is near to neurons’ optimal linear stimulus, more than 90% of contrast-dependent neurons respond more vigorously as the spectral contrast of the short sound is increased.

Once it is acknowledged that the majority of neurons have monotonic contrast-response relationships, we can also consider the variation across the population of auditory cortical neurons. While most neurons' firing rate and signal power continue to increase up to the highest contrast level tested, some saturate their response at intermediate contrasts. This latter behaviour may be a signature of specialisation of a different sort than considered previously. As illustrated in Figure 1.4B, a specialised low-contrast neuron with a strongly saturating *level*-response curve may produce *contrast*-response curves that saturate at intermediate contrasts. However, as we shall see with LN-models presented in subsequent chapters, the level-response curves of auditory cortical neurons do not typically saturate (see Figures 4.1, 4.7, 4.15, and 5.11; c.f. neurons in the inferior colliculus in Figures 7.2 and 7.3). A more plausible explanation for these saturating contrast-response curves will emerge from the results of the next chapter: for intermediate and greater contrasts, these neurons undergo sufficiently strong gain control to compensate for changes in stimulus contrast.

In total, these data suggest that ferret A1/AAF neurons do not in general specialise for particular contrasts, but rather improve the strength and quality of their representations as contrast increases.

3.4.1 Comparison with previous studies

These results differ from those of other studies of contrast preferences in the auditory system. Most notably, Barbour and Wang [2003b] presented similar transient stimuli to awake, passively listening marmosets, while varying the spectral contrasts of these sounds. They reported that approximately half the neurons in auditory cortex increased their response with increasing contrast, while the other half decreased. The number of neurons I observed of the latter type is negligible.

There are several potential reasons to explain this discrepancy. There may be species-specific reasons for marmosets (and perhaps higher primates) — but not ferrets — to have evolved contrast-preferring neurons and circuits. This relates to a similar discrepancy between the observed prevalence of *intensity*-non-monotonic neurons in the primary auditory cortex of different species. These have been reported to comprise 60-80% of monkey A1 neurons [Pfungst and O'Connor, 1981; Sadagopan and Wang, 2008; Watkins and Barbour, 2008], but only 20-30% of cat A1 neurons [Phillips and Irvine, 1981; Sutter and Schreiner, 1995; Sutter, 2000], and approximately 15% of ferret A1/AAF [Bizley et al., 2005] and rat A1 neurons [Polley et al., 2004, 2006]. Although the nature of the stimuli used to characterise intensity-response relationships (pure tones) and contrast-response relationships (chords, and chord-sequences) are different, the rather small proportion of contrast-“tuned” neurons observed here is consistent with the proportion of level-tuned neurons in ferret A1/AAF, but not with those of the monkey. As has been argued for the prevalence of level-selective neurons in these species, contrast-tuned neurons may turn out to be useful for the processing of ethologically relevant stimuli, such as vowels or speech [Davis et al., 2003; DiMattina and Wang, 2006], which may not be of value to the ferret. This idea receives some support from the observation that the proportion of neurons in rat A1 that are intensity-non-monotonic increases if rats are trained on level discrimination tasks [Polley et al., 2004, 2006].

A second point of difference between this study and what I report here is the use of anaesthesia. Although very few studies have compared awake and anaesthetised responses directly in a controlled manner, it is generally recognised that anaesthetics such as barbiturates and ketamine reduce the overall activity of auditory midbrain and cortical neurons [Zurita et al., 1994; Gaese and Ostwald, 2001]. Neurons with sustained, rather than phasic, firing patterns also appear to be more common in awake than anaesthetised animals [Qin et al., 2003; Wang et al., 2005]. However, several studies have identified a wealth of tonic

response properties in the ferret auditory cortex under different anaesthetic regimes [Bizley et al., 2005; Mrsic-Flogel et al., 2005; Campbell et al., 2010]. Furthermore, Walker et al. [2008] found that complex firing patterns of cortical neurons were similar in both the awake and anaesthetised preparation, while Dahmen et al. [2008] observed stimulus-timing-dependent plasticity to be independent of anaesthesia. Finally, in the results of the next chapter, I show that contrast gain control is present in both the anaesthetised and awake animal. Thus anaesthesia is unlikely to explain the discrepancies between the results of this study and those of Barbour and Wang [2003b]. Incidentally, a far greater modulator of auditory cortical activity appears to be an animal’s behavioural state: cortical neurons’ response properties can change considerably when an animal actively engages in an auditory task [Fritz et al., 2005a, 2007; Otazu et al., 2009; Lee and Middlebrooks, 2011].

A final, methodological consideration is the set of statistics used to support the conclusions. A major limitation of the study of Barbour and Wang [2003b] is that they fail to statistically justify their assignment of neurons into monotonically increasing (MI) and monotonically decreasing (MD) classes. They presented only 4 or 5 stimulus conditions to each unit, an undeclared number of times. In turn, their units yielded very noisy contrast-response curves, to which they fitted linear regression lines (Figure 4 of that paper). The sign of the slope of those lines was the basis for their assignment of MI/MD. A visual inspection of their unit-by-unit results suggest that, for most units, one should not reject the null hypothesis that the unit’s response was completely unrelated to contrast. To correct these mistakes, I repeated their experiment by presenting a finer-grained set of contrast conditions over the same approximate range, and using a stricter set of evidence criteria for asserting a model fit. Such measures lower the likelihood of assigning spurious contrast-response relationships. In turn, I found very few units whose responses were monotonically decreasing, or non-monotonically related to contrast. Even in spite of these efforts, it

was often difficult to distinguish between monotonically increasing and tuned (non-monotonic) response curves (e.g. Figure 3.5, third row). This also means that although many neurons appeared insensitive to contrast by the metrics tested here, this could be the result of limited data. However, it is difficult to see how a weak non-monotonicity that is only revealed through hours of careful stimulation might be functionally relevant.

The results on ongoing sounds also diverge from those of [Escabi et al. \[2003\]](#). They presented dynamic ripple stimuli to anaesthetised cats and recorded the firing rate of neurons in the central nucleus of the inferior colliculus (IC) as a function of the spectrotemporal contrast of the stimulus. They observed that the majority of neurons increase their firing rate as contrast increased, in agreement with what I find here. However, they also observed that approximately half of these neurons decreased their rate at the highest contrast value.

A direct comparison between these results is difficult as I only measured responses to transient stimuli, and not to ongoing stimuli, at the highest contrast value presented in that study ($w_L = 30$ dB). It is tempting to extrapolate from the results on transient stimuli — where the response magnitudes generally continued to increase up to this maximum contrast — to expect a similarly heightened firing rate in ferret A1/AAF neurons in response to this maximum contrast. However, this prediction does not take into account factors such as contrast gain control (described in later chapters), or the multitudinous forms of cortical adaptation [[Ulanovsky et al., 2004](#)], which may suppress responses in this range. A further complication in extrapolating from responses to short stimuli to responses to ongoing stimuli is that contrast-response relationships can have different shapes for spectrally preferred and spectrally non-preferred transient chords. The DRCs I presented contain mixtures of both. Finally, the balance between excitation and inhibition is likely to be different when chords are presented in silence compared with when they are presented in quick succes-

sion. Despite all these caveats, non-monotonic contrast dependence in some IC neurons (as observed by [Escabi et al. \[2003\]](#)) does not necessarily lead to non-monotonicity in cortical neurons: the connections from the inferior colliculus via the thalamus to auditory cortex are highly recurrent, with the descending pathway greatly influencing the properties of neurons in the IC [[Ma and Suga, 2001b,a](#); [Yan and Ehret, 2002](#); [Yan et al., 2005](#); [Zhou and Jen, 2005](#); [Nakamoto et al., 2008](#); [Bajo et al., 2010](#)].

3.4.2 The effect of total stimulus level

When building synthetic stimuli, it is not always possible to control all stimulus statistics independently. For the DRCs and transient sounds presented in this study, increasing the stimulus contrast also resulted in a small increase in the overall sound level of the whole DRC (see Table 4.1 for details for the ongoing DRCs). As a consequence, the highest contrast DRC was an average of ~ 6 dB louder than the lowest contrast DRC, with similar changes for the transient sounds.

The net result of this is that the increase in spike rates and signal power with increasing contrast are also partly the result of the increase in stimulus level. This latter contribution is likely to be small, given the magnitude of the increase in sound level. It should be noted that neither [Escabi et al. \[2003\]](#) nor [Barbour and Wang \[2003b\]](#) controlled for this. More thorough controls for changing total stimulus sound level are presented in the following chapter in Section 4.3.4.2 in relation to the gain control phenomena documented there.

Chapter 4

Contrast gain control in auditory cortex

4.1 Introduction

The brain must be able to detect and represent both small and large changes in sound level. Not only do we experience a wide range of sound levels, but within a given signal or context, the important sensory information may lie either in small or large deviations from the average sound. This highlights a fundamental challenge for the auditory system: using neurons with limited dynamic range, the system has to be able to represent large changes in sounds that are highly variable (high contrast), without losing the ability to represent subtle changes in sounds whose level is relatively constant (low contrast).

In Chapter 1, I presented two ways in which the auditory system could manage a range of contrasts. One way is to use separate, specialised circuits to process stimuli with different statistics. This requires some neurons to perform optimally (or at least near optimally) at low contrasts, while others must do so at high contrasts. Evidence for such a “division of labour” solution to the problem of varied sensory contexts has come from two studies. [Barbour and Wang \[2003b\]](#) varied the spectral contrast of random spectral stimuli, and reported

that some marmoset A1 neurons responded preferentially to sounds of low contrast, while others responded more to sounds of high contrast. However, as demonstrated in Chapter 3, these results do not hold up for the ferret primary auditory cortex (A1/AAF), and may be an artefact of limited sampling. Escabi et al. [2003] found that mean firing rates of IC neurons can have non-monotonic dependencies on spectrotemporal contrast, while retaining their spectrotemporal preference. Again, as shown in Chapter 3, this also does not appear to be the case in the ferret A1/AAF. Here, the vast majority of neurons increase the strength and quality of their representations as contrast is increased. Only a minority ($< 10\%$) show evidence of a tuning of their response properties for low contrast.

A second, more efficient solution is contrast gain control — where the responsiveness of neurons is dynamically adjusted according to the contrast of recent stimulation. This approach is efficient in two ways: it avoids the potentially costly duplication of resources necessary to uphold the division of labour solution; and it optimises neural coding to be informative over the current distribution of stimuli. As described in Section 1.3, considerable evidence suggests that the mammalian visual system uses contrast gain control [Shapley and Victor, 1978] so that it can operate in both high and low contrast environments. This mechanism is well described by *divisive normalisation*, whereby the range of visual input is adjusted according to the contrast of recent visual stimulation [Heeger, 1992; Carandini et al., 1997; Schwartz and Simoncelli, 2001; Bonin et al., 2005].

In the auditory system, several studies have investigated the effects of temporal (i.e. within-band) contrast on neural responses, and have provided evidence for gain control. The effect has been observed in studies where temporal contrast was controlled: when the modulation depth of sinusoidally amplitude-modulated tones is reduced, neurons from the auditory nerve [Joris and Yin,

1992] to the auditory cortex [Malone et al., 2010] can rescale their gain to partially compensate for the lower contrast. Similar effects have been found in the inferior colliculus (IC; [Kvale and Schreiner, 2004; Dean et al., 2005]) and in the songbird forebrain [Nagel and Doupe, 2006] when the temporal contrast of more complex stimuli is altered. Such gain changes improve the efficiency with which neurons encode frequently-presented sound levels [Dean et al., 2005].

Division of labour and gain control solutions are not mutually exclusive. The observations of Barbour and Wang [2003b] and Escabi et al. [2003] are compatible with contrast gain control, so long as gain changes are slow (compared to spike generation), or do not completely compensate for changes in contrast.

In this study, I ask whether gain control is implemented by the mammalian auditory cortex, and whether it adjusts neural gain according to the spectrotemporal contrast of recent stimulation. There are several possible experimental outcomes to this question. One possibility is that neurons' responses do not adjust at all to the statistics of recent stimulation. This would suggest that the problem of variable environmental statistics is ignored. As a second possibility, neurons may be informative only about stimuli with a particular contrast. For example, their receptive fields may be more well-defined, or the sensitivity of their firing rate may be maximal for a limited range of environmental statistics. This would suggest a division of labour strategy. Thirdly, cortical neurons may undergo complex changes in their spectrotemporal tuning as contrast varies, suggesting that changing the environmental statistics induces a reallocation of resources in the auditory system. This would not be entirely without precedent: tuning of auditory neurons has been shown to depend on stimulus context, such as tone density [Blake and Merzenich, 2002; Valentine and Eggermont, 2004; Williamson et al., 2011], stimulus bandwidth [Gourévitch et al., 2009], and the history of recent stimulation [Ahrens et al., 2008]. Finally, gain control could be observed in some or all cortical neurons. To distinguish between these hypothe-

ses, I probed cortical neurons with DRC stimuli, while controlling the statistics of level variations within individual frequency bands. From their responses, I estimated spectrotemporal receptive fields (STRFs), using both linear (Section 2.2.1) and linear-nonlinear (LN; Section 2.2.2) models. A comparison between neurons' stimulus-response functions under different conditions reveals how the ferret auditory cortex responds to the challenges imposed by changing the contrast of its inputs.

4.2 Methods

4.2.1 Animals and recordings

Eight adult pigmented ferrets (6 male, 2 female) were chosen for electrophysiological recordings under anaesthesia. Further recordings were made in an awake, passively listening female ferret. A full description of the anaesthesia, surgical, stimulation and recording procedures are provided in Section 2.3.

4.2.2 DRC Stimuli

Dynamic random chords (DRCs), as described in Section 2.1.1, were presented during this study. Each chord in the DRCs was a superposition of 34 pure tones, with frequencies log-spaced between 500 Hz and 22.5 kHz at 1/6 octave intervals. The levels of each tone during each chord were independently drawn from a uniform distribution. The properties of this distribution were varied to produce three separate stimulus corpora.

In the first corpus, the mean level was fixed at $\mu_L = 40$ dB SPL, and the halfwidth of the distribution was varied to give three stimulus contrasts ($w_L = 5, 10, 15$ dB; $\sigma_L = 2.9, 5.8, 8.7$ dB; $c = 33, 64, 92\%$). For the duration of this

chapter, we shall refer to these as simply the low, intermediate, and high contrast conditions. The overall RMS level of the stimuli was 71.0 ± 0.5 dB SPL in low contrast, 72.4 ± 1.0 dB SPL in intermediate contrast, and 74.5 ± 1.5 dB SPL in high contrast. A control experiment was performed to show that these small differences in the overall level did not account for gain control (see Section 4.3.4.2). This set of DRCs is illustrated in Figure 2.3.

In the second corpus, the mean level was again fixed at $\mu_L = 40$ dB SPL, and a broader range of halfwidths was presented ($w_L = 2.5$ dB to $w_L = 20$ dB in 2.5 dB steps; σ_L from 1.4 dB to 11.5 dB; c from 17% to 116%). This was used to quantify the functional relationship between σ_L and gain, as in Section 4.3.5.

In the third corpus, the mean level of the tones was varied from $\mu_L = 20$ dB SPL to $\mu_L = 56$ dB SPL, and a range of halfwidths was presented. The conditions are summarised in Table 4.1. This was used to explore the extent to which gain control depended on the mean tone level, in Section 4.3.6.

Each DRC sequence was presented 5-20 times (10 times for the awake animal), randomly interleaved, with 15-20 s of silence between each sequence. The first 2 s of data from each presentation were discarded to ensure that a constant adaptation state had been reached.

4.2.3 Unit selection criteria

Only units which modulated their firing rate in response to the DRCs were included for analysis. This was measured via the noise ratio (NR; see Section 2.3.5.2): units with $\text{NR} > 10$ in any one stimulus condition, i.e. whose explainable variance was less than 9.1% of the total variance, were excluded.

Lower NR is associated with better predictions and less overfitting in STRF data [Sahani and Linden, 2003b]. As was demonstrated in Section 3.3.2, the

tone level distribution			tone pressure distribution			(total) DRC level		
μ_L [dB SPL]	w [dB]	σ_L [dB]	μ_P [mPa]	σ_P [mPa]	c [%]	min [dB SPL]	RMS [dB SPL]	max [dB SPL]
Constant mean level								
40	2.5	1.4	2	0.3	17%	69.9	70.8	71.5
40	5	2.9	2.1	0.7	33%	69.4	71.1	72.6
40	7.5	4.3	2.3	1.1	49%	69.1	71.7	73.8
40	10	5.8	2.5	1.6	64%	69	72.5	75.2
40	12.5	7.2	2.8	2.2	78%	69.1	73.4	76.7
40	15	8.7	3.1	2.9	92%	69.4	74.6	78.3
40	17.5	10.1	3.7	3.8	104%	69.8	75.8	80.1
40	20	11.5	4.3	5	116%	70.4	77.2	82
Varied mean level								
20	5	2.9	0.2	0.1	33%	49.4	51.1	52.6
20	10	5.8	0.2	0.2	64%	49	52.5	55.2
20	15	8.7	0.3	0.3	91%	49.4	54.6	58.3
24	3	1.7	0.3	0.1	20%	53.8	54.8	55.7
24	9	5.2	0.4	0.2	58%	53	56.1	58.6
30	5	2.9	0.7	0.2	33%	59.4	61.1	62.6
30	10	5.8	0.8	0.5	64%	59	62.5	65.2
30	15	8.7	1	0.9	91%	59.4	64.6	68.3
32	4	2.3	0.8	0.2	26%	61.6	62.9	64.2
32	12	6.9	1.1	0.8	75%	61.1	65.2	68.4
48	6	3.5	5.4	2.1	39%	77.3	79.3	81.1
48	18	10.4	9.5	10.1	107%	77.9	84.1	88.5
50	5	2.9	6.7	2.2	33%	79.4	81.1	82.6
50	10	5.8	7.8	5	64%	79	82.5	85.2
50	15	8.7	10	9.1	91%	79.4	84.6	88.3
56	7	4	14	6.4	45%	85.2	87.6	89.6
56	21	12.1	29.1	35	120%	86.6	93.8	98.8

Table 4.1: Summary of stimulus statistics. Tone level and tone pressure statistics have been calculated from the respective analytic distributions of these variables given in Section 2.1.3.1. Statistics of the total level of the DRC (last three columns of the table) have been estimated from sample DRCs of 30 s duration.

SP of cortical neurons generally increases with contrast, while the noise power (NP) remains constant; thus NR declines with increasing contrast. For this reason, the high contrast condition ($\mu_L = 40$ dB SPL, $w_L = 15$ dB), which was presented to all units, has been used as the reference condition for comparisons across stimulus conditions.

4.2.4 Linear models of unit responses: STRF estimation

STRFs were estimated by correlating the stimulus history with the spike peri-stimulus time histogram (PSTH). The PSTH was binned at 25 ms; bins were offset by between 0 and 25 ms to allow for response latency. The offset was chosen on a unit-by-unit basis to minimise the NR. I estimated separable STRFs for each unit, as given in Section 2.2.1.1. Separable STRFs gave more accurate predictions than fully inseparable STRFs (which had more parameters). STRFs were trained on 9/10 of the available data for each unit, and were used to predict a PSTH for the remaining 1/10. The prediction score is defined as the proportional reduction in the signal power of the response in the prediction dataset; if this was positive, the STRF was deemed predictive.

STRFs were estimated separately for each stimulus condition and for the pooled dataset. The separate set of STRFs was used for the linear analysis (Figures 4.4 and 4.5; Section 4.3.3); the pooled STRFs were used thereafter. In each case, units whose STRFs or LN models (see below) were not predictive were excluded from analysis.

Statistics on the STRF were calculated as follows. The best frequency (BF) was calculated as the centre of mass (CoM) of the (element-wise) squared frequency kernel, k_f^2 . Given that a finite range of frequencies was used, noise in the estimates of k_f^2 exerted a bias in the CoM values towards the centre of this range. To compensate for this bias, I calculated the CoM on a circular basis. I

manually checked that there were no wrapping artefacts when the true BF was at either the lower or upper extreme of the frequency range. Bandwidth was estimated as the (circular) standard deviation of k_f^2 . Circular statistics were performed using CircStat in Matlab [Berens, 2009]. I confirmed that all results held when BFs and bandwidths were estimated on non-circular (i.e. standard) bases.

Noise in the STRF had a considerable impact on the estimate of bandwidth. To compensate for this, I calculated a second bandwidth measure after excluding all coefficients of k_f that were not significantly non-zero. This was established via a bootstrap procedure, as follows. First I constructed confidence intervals for the components of k_f . I drew (with replacement) 1000 subsamples of 90% of the stimulus/response data from the pooled dataset, and estimated an STRF from each. Those components of k_f not significantly non-zero across the ensemble (via studentisation, $P < 0.01$) were set to zero.

4.2.5 Nonlinear models of unit responses: LN model estimation

The linear STRFs were refined by fitting a Linear-Nonlinear (LN) model to units' time-varying responses, y_t (Section 2.2.2). If the time-varying output of the linear STRF model is denoted as z_t , as per Equation (2.23), then the LN model fits a function F such that $\hat{y}_t = F[z_t]$. To fit F , the set of pairs of (z_t, y_t) were divided into 40 bins along the z axis, and the values of z_t and y_t averaged within each bin, as per [Chichilnisky, 2001; Simoncelli et al., 2004]. The edges of the bins were spaced to contain equal number of points within each bin. Examples are provided in Figure 4.1.

When viewed in this way, it was clear that a sigmoid (logistic) parameterisation of the nonlinearity captured the observed stimulus/response relationship for

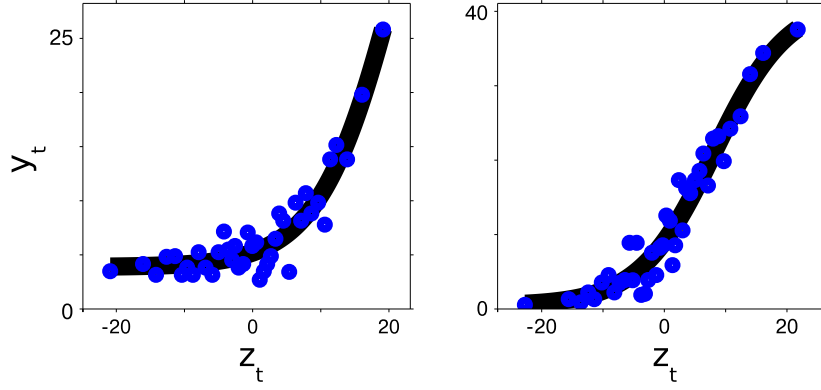


Figure 4.1: Output nonlinearities for two example units in the high contrast condition. On the abscissa is the output of the linear model, z_t , i.e. the projection of the stimulus onto the STRF. On the ordinate is the actual spike rate, y_t . Blue dots indicate binned data for the respective units (see Section 4.2.5). Black lines show fitted sigmoid functions.

cortical units:

$$F[z_t] = a + \frac{b}{1 + e^{-(z_t - c)/d}} \quad (4.1)$$

The four parameters of the sigmoid were fitted to the binned data for each neuron in each condition, by minimising the squared loss function $E = \sum_t (y_t - \hat{y}_t)^2$ via gradient descent. As described in the results (Section 4.3.4), the saturation points of these nonlinear functions F were never reached; in many cases, an exponential function sufficed to fit these data. However, for the purposes of consistency, sigmoid forms for F were fitted throughout for all units.

4.2.6 Curve transformations

Independent sigmoids were fitted to the responses from each contrast condition. To describe the differences between the sigmoids, I chose the nonlinearity for the high contrast condition ($w_L = 15$ dB, $\sigma_L = 8.7$ dB, $c = 92\%$; see Section 4.2.2) for every unit as a reference. I then found the linear transformations required to map the sigmoid nonlinearity for the reference condition onto the sigmoids obtained under the other conditions (see Section 4.3.4).

A full set of linear transformations allows for scaling (expansion/dilation) and translation in each of the horizontal (z) and vertical (y) directions of the output nonlinearity. These transformations map directly onto changes into the respective parameters in Equation (4.1). Changes in a amount to vertical translations of the curve; changes in b amount to vertical scaling; changes in c amount to horizontal translations; and changes in d amount to horizontal scaling.

Because none of the units measured operated near their saturation points, the extent of vertical scaling (changes in b) could not be accurately estimated alongside horizontal scaling (changes in d). In practice, only allowing for horizontal scaling provided considerably better fits than only allowing for vertical scaling. Thus changes in the parameters $\{a, c, d\}$ are considered.

Fitting the linear transformations is equivalent to solving, for each contrast condition σ_i :

$$\hat{F}_{\sigma_i}[z_t] = F_{\sigma_0}[G \cdot z_t + \Delta z] + \Delta y \quad (4.2)$$

Here, σ_0 is the reference condition; F_{σ_0} denotes the sigmoid nonlinearity in the reference condition, G is the horizontal scaling factor (i.e. gain change), Δz is the z -offset, and Δy is the y -offset. For a given unit, Δz is expressed as a percentage of the size of the domain of z in the reference condition for that unit, while Δy is expressed as a percentage of $F_{\sigma_0}[0]$.

Equation (4.2) was fitted by minimising a quadratic loss function. This measured the squared difference between the target curve, F_{σ_i} , and the transformed reference curve $\hat{F}_{\sigma_i}$, over the domain of the target curve only. I used gradient descent to find the optimal parameters of the gain, G , the z -offset, Δz , and the y -offset, Δy , using initial conditions of $G = 1$ and $\Delta z = \Delta y = 0$. Gradient descent was repeated a further ten times from random initial conditions (with $G \sim \text{Exp}(1)$, and $\Delta z, \Delta y \sim \mathcal{N}(0, 10)$), and the best of all these fits chosen. Results were cross-checked by minimising the mean squared error between the

transformed curve and the original data by which the original sigmoid was fitted.

4.2.7 Fisher information

In order to compare the abilities of cortical neurons to encode stimulus distributions, I computed the Fisher information $I_F[z]$ of the output nonlinearities in each condition. The Fisher information, for a single neuron, provides a measure of its coding accuracy: the higher $I_F[z]$, the greater the ability of an ideal observer to decode the stimulus presented at that time. More precisely, since this was calculated as a function of z , $I_F[z]$ quantifies the ability of an ideal observer to decode the (current) value of the projection of the current stimulus onto that neuron's STRF, based on that unit's current firing rate.

Assuming that the firing rate of a unit at time t is Poisson-distributed with mean $F[z_t]$, then the Fisher information is given by:

$$\begin{aligned} I_F[z] &\equiv \left\langle -\frac{d^2}{dz^2} \log P[y|z] \right\rangle_y \\ &= \frac{(F'[z])^2}{F[z]} \end{aligned} \tag{4.3}$$

where $\langle \cdot \rangle_y$ denotes the expectation over firing rates y .

The stimulus of maximum Fisher information, z^* , can then be determined by solving $\partial I_F[z]/\partial z = 0$. Using the parameterisation of the sigmoidal output nonlinearity of Equation (4.1), this yields a unique solution:

$$z^* = c - d \cdot \log \left[-b + \frac{1}{4a} \sqrt{b^2 + 16a(a+b)} \right] \tag{4.4}$$

4.2.8 Normalisation models

To model the effects of stimulus statistics on neural gain, I extended a well-known class of gain normalisation equations used in the visual system [Heeger, 1992; Carandini et al., 1997], which take on the general form given in Equation (4.9). As all gain values were computed relative to a reference curve (Equation (4.2)), I fixed $A = 1 + b \sigma_L^n$ to constrain $G(\sigma_0) = 1$.

To model the effects of varying both the width of the tone level distribution (via σ_L), and its mean (μ_L), I fitted separate values for B (and therefore for A) for each mean level, as given in Equation (4.11). Here, $A(\mu_L)$ was constrained by $A(\mu_L) = 1 + B(\mu_L) \cdot \sigma_L^n$, so that $G(\sigma_0, \mu_L) = 1$ for all μ_L (as observed in the data). n was set as constant with respect to μ_L . The fit obtained was slightly better than if n was allowed to vary as a function of μ_L and B was kept constant with respect to μ_L .

Following the empirical fitting of $B(\mu_L)$ values, B was parameterised using an exponential form (Equation (4.12)), to capture the saturation of $B(\mu_L)$ at high μ_L .

4.3 Results

I recorded from 1840 sites in the primary auditory cortex (A1) and anterior auditory field (AAF) of eight anaesthetised ferrets, while diotically presenting dynamic random chord (DRC) sequences. The chords were changed within each sequence every 25 ms, with the levels of their constituent tones (1/6 octave spaced) drawn from uniform distributions in SPL space. The contrast of the sequences was manipulated by changing the (SPL) standard deviation (σ_L) of these distributions.

In what follows, three sets of stimulus statistics were explored. In the first set, the tone level distributions had identical mean ($\mu_L = 40$ dB SPL), but three different halfwidths of low, medium, and high contrast ($w_L = 5, 10, 15$ dB; $\sigma_L = 2.9, 5.8, 8.7$ dB; $c = 33, 64, 92\%$) were used. In the second set, DRCs with a greater range of contrast values was presented. Finally, μ_L was varied as well as σ_L . The range of tone level statistics is documented in Table 4.1.

As these distributions are primarily defined in (logarithmic) SPL space, and as I performed analyses on units' stimulus-response relationships using stimulus representations in SPL space, I present the data and models here in terms of σ_L rather than $c = \sigma_P/\mu_P$, so as not to mix together the sound pressure and level domains.

The RMS sound level of the total stimulus ranged from 70-80 dB SPL. Although the overall sound level of the total stimulus did vary slightly with σ_L , a control experiment was performed to ensure that this did not materially contribute to the results presented here (see Section 4.3.4.2).

I identified 1001 units which responded reliably to the DRCs, as measured via a maximum noise level criterion (see Section 4.2.3).

Although the anaesthetised preparation allowed for precise control of stimulation and eliminated the possibility of attentional modulation, to confirm that the observations made under anaesthesia apply in awake animals, I also presented the same stimuli through a free-field speaker to an awake, passively listening ferret, and recorded spiking activity from 62 sites in A1 and AAF. I identified 19 units which responded reliably to the DRCs. I found no differences in the response characteristics of neurons in the two preparations, and therefore have combined these data for the analyses below.

4.3.1 Initial observations of gain changes

The first insight into the existence of gain changes in cortical neurons can be obtained from the results in Chapter 3.

Consider a sample of noiseless neurons, with linear stimulus-response relationships. If the neurons do not change their gain with the statistics of the stimulus distribution, then we should expect the standard deviation of neural responses, $\sigma^{(R)}$, to be directly proportional to the standard deviation of the stimulus, $\sigma^{(S)}$. On the other hand, if neurons adjust their gain to completely compensate for $\sigma^{(S)}$, then any reduction in stimulus contrast should be compensated for by an increase in gain, so that $\sigma^{(R)}$ is unaffected by contrast.

Real neurons are, however, not noiseless, and so a more useful estimate of the *stimulus-driven* response variability is the signal power (SP), as defined in Section 2.3.5.2. The focus in Chapter 3 was on the *shape* of the relationship between SP and $\sigma^{(S)}$: in Figure 3.4, SP was demonstrated to monotonically increase with $\sigma^{(S)}$ for the majority of neurons. Here it is instructive to consider the relationship between SP and $\sigma^{(S)}$ in absolute terms, as shown in Figure 4.2.

The red diagonal lines in Figure 4.2 show, for each unit, the expected signal power under each condition if two assumptions hold: that neurons respond linearly to the stimulus, and that no gain control processes operate. For two thirds of these units, as the contrast of the stimulus was reduced, the signal power on average exceeded this prediction.

4.3.1.1 An elementary measure of gain

These results can be used to motivate an elementary measure of the strength of gain control. Given the assumptions above, we can compare the ratio



Figure 4.2: Evidence of gain control in the relationship between contrast and signal power (SP). Figure 3.4A has been adjusted to show the absolute scale of the SP as a function of σ_L . Black triangles are the observed data as in Figure 3.4A. The red diagonal line connects the SP at the highest contrast stimulation to $SP = 0$ at $\sigma_L = 0$, showing the expected relationship for linear neurons with no gain control. The purple dashed line shows the expected relationship for units with complete gain control.

between the SP of the responses to two stimuli of high and low contrast (SP_{low}/SP_{high}) with the corresponding ratio between the stimulus standard deviations ($\sigma_{low}^{(S)}/\sigma_{high}^{(S)}$):

$$G_{SP} = \frac{SP_{low}/SP_{high}}{\sigma_{low}^{(S)}/\sigma_{high}^{(S)}} \quad (4.5)$$

If there is no gain control, then SP is proportional to $\sigma^{(S)}$, and so $G_{SP} = 1$. If there is complete compensation for changes in stimulus statistics, then SP is independent of $\sigma^{(S)}$, and $G_{SP} = \sigma_{high}^{(S)}/\sigma_{low}^{(S)}$.

For the 1001 units recorded, I calculated G_{SP} for the comparison between high ($\sigma_L = 8.7$ dB) and low ($\sigma_L = 2.9$ dB) contrast conditions. The population statistics are shown in Figure 4.3. The population median was $G_{SP} = 1.78$, indicating that cortical neurons generally show some gain control. Only a

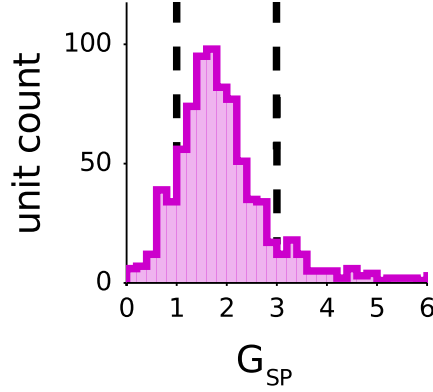


Figure 4.3: Most units show gain control. Histogram of the relative scaling of signal power, G_{SP} , as σ_L was reduced by a factor of 3 (Equation (4.5)). A value of 1 indicates no gain control, while a value of 3 indicates complete compensation for the change in σ_L .

minority reached or exceeded $G_{SP} = 3$, the predicted value for complete gain control. Finally, less than 10% (98/1001) of the population of units had $G_{SP} \leq 1$, indicating that compensatory gain changes are exhibited by the majority of units in A1/AAF.

4.3.1.2 Caveats

This analysis suffers from two shortcomings. Firstly, the signal power is both an imprecise and a poor measure of how auditory neurons respond to DRCs. Error bars on the estimates of signal power were calculated (as per [Sahani and Linden, 2003b]), but these were often large, obscuring illustration of the mean estimate and making quantitative comparisons across conditions difficult. Furthermore, far better measures of contrast-dependent changes in units' responsiveness are available if we characterise responses along the dimensions of stimulus space to which neurons are actually sensitive. These dimensions can be determined by units' STRFs.

The second faulty assumption is that neurons' stimulus-response relationships are linear. Even in the complete absence of any gain control process, the thresh-

olding and saturation properties of neurons — properties which necessitate extending linear STRF models to linear-nonlinear (LN) models, as described in Section 2.2.2 — break down any expected linear scaling of signal power with stimulus contrast. The resultant skew can bias changes in SP either upwards or downwards, depending on the local convexity of the response nonlinearities. As a result, any potential gain changes are better evaluated through changes to the output nonlinearities in LN models.

These considerations motivate a thorough evaluation of the effect of contrast on the responses of auditory cortical neurons. This constitutes the remainder of this chapter.

4.3.2 Spectrotemporal receptive fields

I first asked whether the tuning of cortical neurons is affected by changes in stimulus contrast. If this were the case, it would not be appropriate to describe such a response as gain control. I characterised the tuning of each unit by estimating one STRF for each contrast condition, $k_{fh}^{(\sigma_L)}$ (e.g. Figure 4.4):

$$\hat{y}_t = k_0 + \sum_{f, h} X_{tfh} \cdot k_{fh}^{(\sigma_L)} \quad (4.6)$$

Only STRFs which had predictive power (see Section 4.2.4) were included in further analysis.

As contrast was reduced, fewer units yielded predictive STRFs ($P \ll 0.001$; Kruskal-Wallis test). In general, units were less reliably driven by low contrast stimuli. Signal power under low contrast stimulation was a median 53% less than that measured under high contrast. Similarly, the intermediate contrast stimulus yielded signal powers that were a median 17% less than that measured

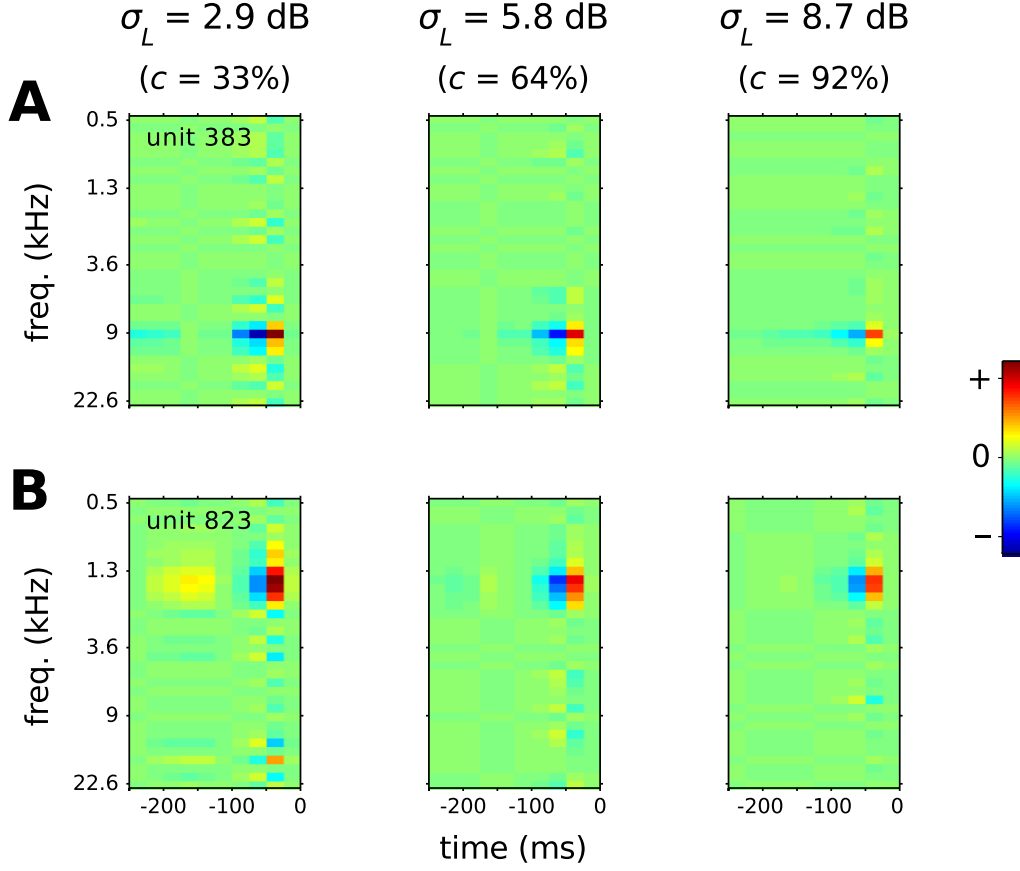


Figure 4.4: Effect of stimulus contrast on tuning. Example units (A) and (B) show the same spectrotemporal selectivity in the receptive fields (STRFs) estimated from the three contrast conditions. Red denotes components of the stimuli that excite the unit, blue denotes components that inhibit. As the colour scale is uniform across the plots, it is clear that the dominant variation across the conditions lies in the magnitude of the drive to the unit: this is stronger (the cortex appears to listen harder) under low-contrast than under high-contrast stimulus conditions.

under high contrast. Noise power, however, remained constant across conditions. Thus, units were more reliably driven under high contrast stimulation, and hence their responses were more predictable.

4.3.3 Stimulus contrast has no effect on tuning

Changing stimulus contrast produced only small changes in STRF shape (Figures 4.4 and 4.5A-B). Of 261 units with predictive STRFs, 223 maintained

the same best frequency (BF) across conditions (within 1/6 of an octave; Figure 4.5A). 26 units had STRFs that were too diffuse to give clear BF estimates. Only 12 units showed evidence of changes ($\leq 1/3$ octave) in BF across conditions. Tuning bandwidths were slightly broader under low contrast stimulation (sign-rank test; $P \ll 0.001$); however, this may reflect the noisier estimates of STRF coefficients at low contrast. Tuning bandwidth did not change systematically between medium and high contrast regimes ($P > 0.5$; Figure 4.5B). I also observed no systematic changes in the temporal structure of STRFs, though this was limited by the 25 ms time resolution of the analysis.

The most notable change across the STRFs for any unit was in the gain: as the contrast was reduced, the magnitude of the largest coefficient increased (Figure 4.5C). These are discussed in more detail below in Section 4.3.4.

To assess the importance of any unmeasured STRF shape changes, I modelled each neuron using a single linear STRF for all contrast conditions, multiplied by a contrast-dependent gain factor G_{σ_L} :

$$\hat{y}_t = G_{\sigma_L} \left[k_0 + \sum_{f,h} X_{tfh} \cdot k_{fh} \right] \quad (4.7)$$

In this way, STRFs from one stimulus condition could be used to predict responses in the other conditions. Cross-conditions prediction scores (using Equation (4.7)) were as good as the within-condition STRFs (using Equation (4.6); median difference in prediction scores of 3.5%, $P > 0.4$; see Figure 4.5D), indicating that any shape changes in the STRFs were negligible. Thus auditory cortex neurons exhibit similar spectrotemporal preferences regardless of contrast. This is similar to previous observations in the IC [Escabi et al., 2003], but different from the visual system, where contrast has a considerable effect on the temporal dynamics of neural responses [Mante et al., 2005].

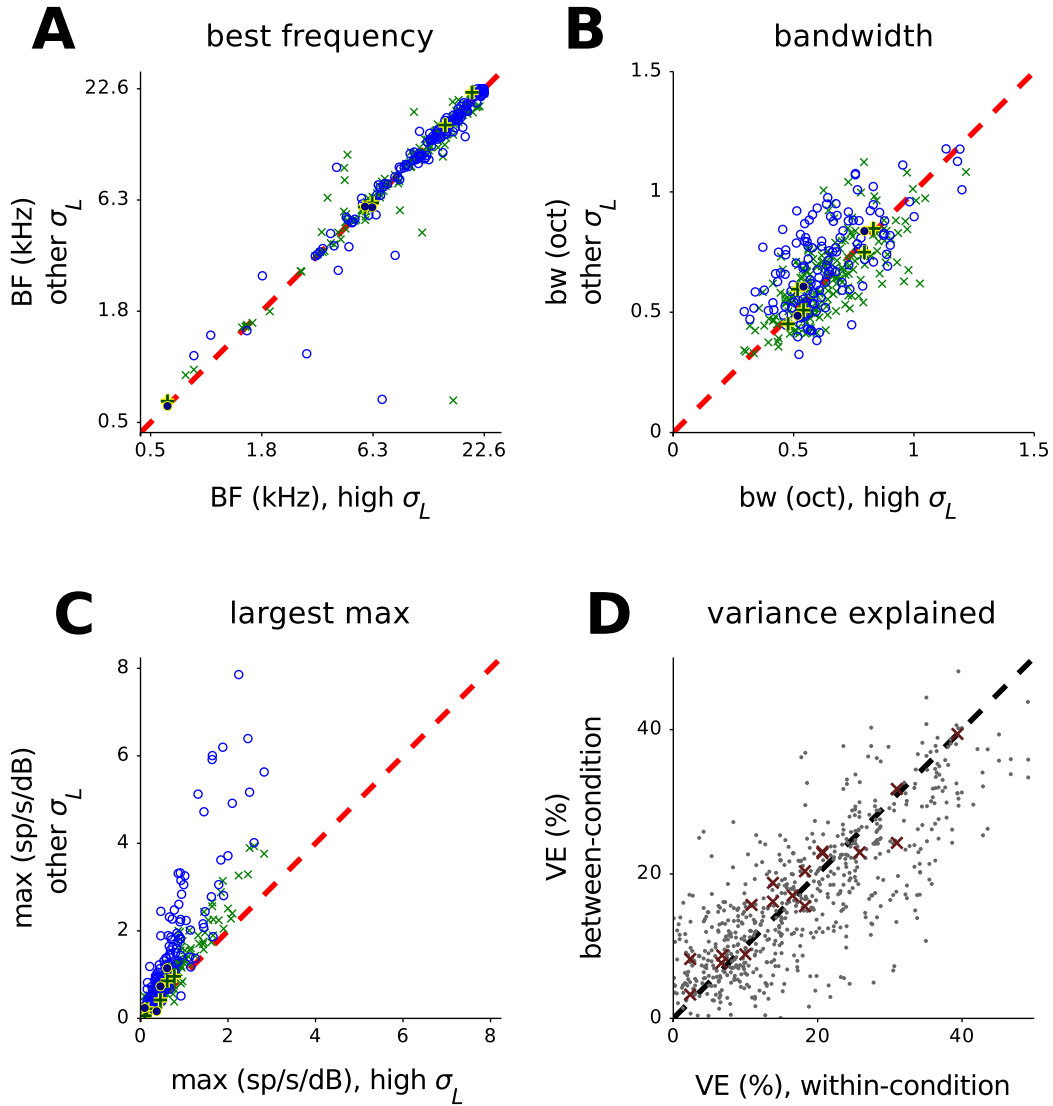


Figure 4.5: Statistics of the effect of stimulus contrast on tuning. (A-C) Best frequency (A) does not vary systematically with stimulus contrast across units. Tuning bandwidth (B) shows a small, significant broadening at low contrast. In almost all units, the gain of the STRF (C), measured by the maximum magnitude of any coefficient, increases as contrast decreases. STRF properties under low contrast stimulation are shown as blue circles, under medium contrast as green crosses. Filled dark blue circles and dark green pluses indicate data in these two same conditions from the awake recordings. (D) The linear STRFs fitted from DRCs with different contrasts are sufficiently similar in tuning properties that, once adjusted for differences in STRF gain, they predict responses across stimulus conditions on average 96.5% as well as they do within their own conditions. Thus, most of the contrast dependence of STRFs is captured by a change in gain. Red crosses indicate data from the awake recordings.

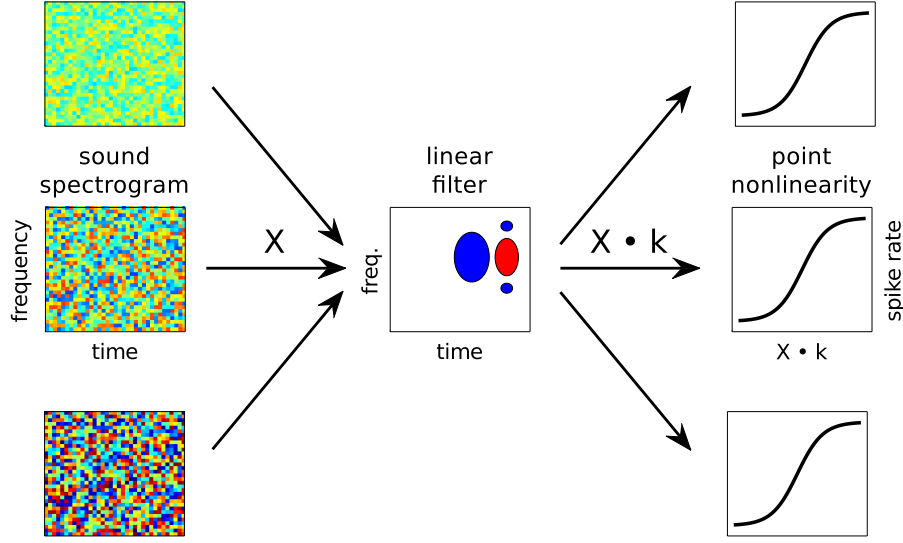


Figure 4.6: Schematic of the contrast-dependent LN model. Given that units' STRFs did not change shape as a function of contrast (Section 4.3.3), a linear-nonlinear model was fitted to the responses of the unit to each contrast condition, with a shared STRF across all conditions.

4.3.4 Quantifying gain changes: linear-nonlinear models

The above results show that the primary effect of changing contrast is that neurons increase their gain. To characterise gain changes more accurately, I extended the simple linear model to a linear-nonlinear (LN) one (see Sections 2.2.2 and 4.2.5, and Figure 2.6). This comprised a single linear STRF for each unit, estimated from its responses across all conditions, followed by a sigmoidal output nonlinearity. Separate nonlinearities were fitted for each contrast condition. This model is illustrated in Figure 4.6.

The LN model far outperformed the linear models: prediction scores were a median 38.5% higher than the within-condition linear models ($P \ll 0.001$; sign-rank). I found 315 units where LN models were predictive in all three contrast conditions.

In an LN model, differences in gain manifest through changes in the shape of the output nonlinearity. To quantify these changes, I calculated the set of

linear transformations required to map the output nonlinearity for high contrast stimulation ($\sigma_L = 8.7$ dB, $c = 92\%$) onto those for other stimulus conditions. In principle, this mapping could combine a scaling of the curve along the horizontal (z) and vertical (y) axes, and a translation of the curve along these axes (z - and y -offset, respectively). However, none of the units under investigation operated near their saturation point (see Figure 4.1 for examples), making an estimate of vertical scaling difficult. Thus, I measured changes in the remaining three degrees of freedom (see Equation (4.2)). Horizontal scaling corresponds to a change in gain, z -offset to a threshold shift, and y -offset to a change in minimum firing rate. Examples of output nonlinearities for the different contrast conditions, and the corresponding curve transformations, are provided in Figure 4.7.

I observed a robust relationship between stimulus contrast and gain across the population of units. An approximately 3-fold decrease in σ_L from 8.7 dB ($c = 92\%$) to 2.9 dB ($c = 33\%$) increased gain by a median factor of 2.01; for a 1.5-fold decrease in σ_L from 8.7 dB ($c = 92\%$) to 5.8 dB ($c = 64\%$), gain increased by 1.34x (Figure 4.8A).

The gain effect was also strongest — or at least most clearly observed — among units with the most robust, repeatable spike trains (Figure 4.9A). The lower the noise power (see Section 2.3.5.2), the greater the observed G . There is no established *theoretical* reason for this to be the case: it may be the case that the estimator for G is biased towards 1 for noisy data. Nevertheless, by taking the approach of extrapolating from the population data to a hypothetical zero-noise neuron [Sahani and Linden, 2003b; Ahrens et al., 2008], relative gain changes of 2.21 ± 0.16 and 1.40 ± 0.06 are obtained for the comparisons made above (median $\pm$ 99% confidence interval). The gain of auditory cortical neurons therefore changes in the appropriate direction to compensate for changes in stimulus contrast, but this compensation is not complete.

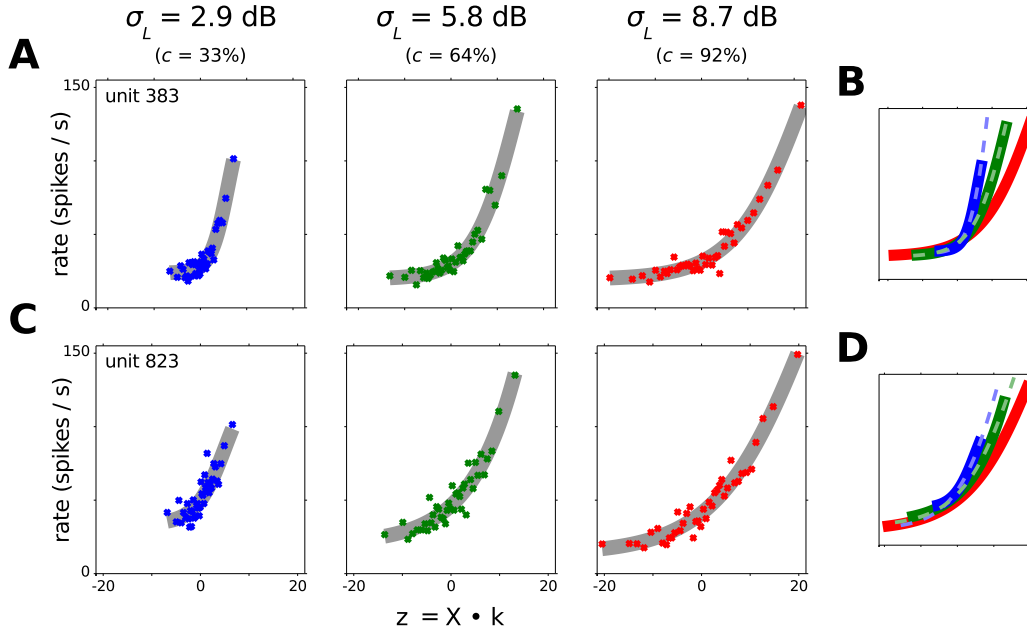


Figure 4.7: Effects of stimulus contrast on output nonlinearities: examples. Output nonlinearities for two example units show gain rescaling as a function of stimulus contrast. (A) For each unit, I fitted a single linear STRF, then calculated a separate output nonlinearity for each contrast condition. The abscissa denotes the output of the linear STRF, z ; the ordinate is the spike rate, y . (B) The differences between the nonlinearities were quantified by the linear transform required to convert the high-contrast nonlinearity into each of the medium- and low-contrast nonlinearities. Solid curves show the original sigmoids for the unit shown in (A); dashed lines show the result of the transformation of the high-contrast (red) curve into the low contrast (blue) and medium contrast (green) curves. The parameters for the high-to-low transform for this unit were $G = 2.9$, $\Delta z = 17.6\%$, $\Delta y = 8.9\%$; the high-to-medium transform parameters were $G = 1.5$, $\Delta z = 2.0\%$, $\Delta y = -0.6\%$. (C,D) Nonlinearities for a second example unit; panels equivalent to (A,B). Parameters for these transforms were $G = 1.5$, $\Delta z = -27.6\%$, $\Delta y = 2.4\%$ (high-to-low) and $G = 1.2$, $\Delta z = -3.8\%$, $\Delta y = 8.5\%$ (high-to-medium).

Decreasing stimulus contrast also caused nonlinearities to shift by a small amount to the right (median z -offset of 5.5% and 1.4% for low and medium contrast; $P < 0.001$ and $P < 0.05$ respectively, sign-rank test; Figure 4.8B), but there was no corresponding vertical translation of these curves (Figure 4.8C). Although the change in z -offset is nominally indicative of a small increase in threshold, the gain and z -offset measures were correlated with each other across units ($r^2 = 0.195$ in high-to-low- and $r^2 = 0.11$ in high-to-medium-contrast curve transformations; Figure 4.9B), suggesting that the rightwards shift in

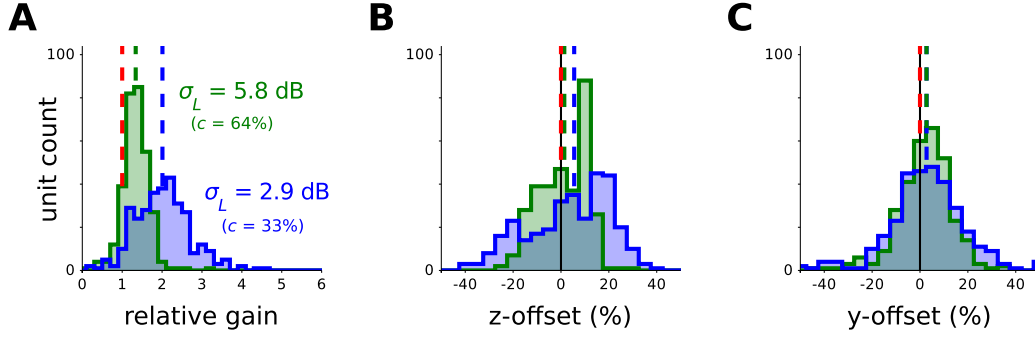


Figure 4.8: Effects of stimulus contrast on output nonlinearities: statistics. (A) Histogram of gain (relative to the reference condition) under low (blue) and medium contrast (green) conditions, for $n = 315$ units. Coloured dashed lines are population medians. Red dashed line indicates $G = 1$, i.e. no scaling of the nonlinearity. (B) As in (A), for changes in z-offset. Positive values denote rightward shifts in nonlinearities. (C) As in (A), for changes in y-offset. Positive values denote upward shifts in nonlinearities.

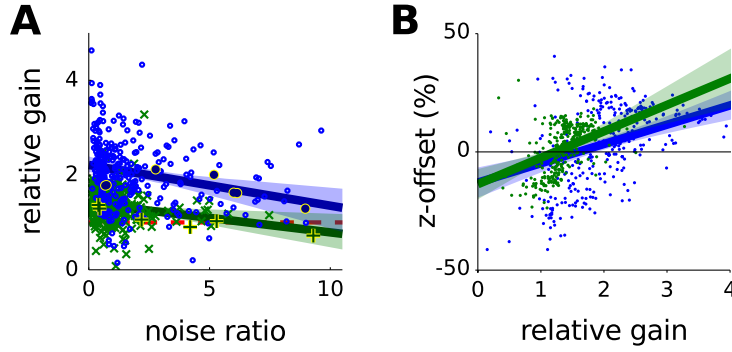


Figure 4.9: Effects of stimulus contrast on output nonlinearities: additional statistics. (A) The gain effect was strongest among units with the most robust, repeatable spike trains. Units with low noise ratios (NR) showed a greater modulation of their gain as a function of contrast. Colours as in Figures 4.5 and 4.8. Shaded regions show the (bootstrapped) 99% confidence intervals about the mean regression line (solid). Red dashed line indicates $G = 1$ across conditions, i.e. no gain modulation. (B) For both the high-to-low-contrast (blue), and the high-to-medium-contrast (green) transformations, changes in gain (abscissa) and z-offset (ordinate) were positively correlated across the population of units. Lines as in (A).

curves partly acts to compensate for gain changes (see Figure 4.13A). The lack of systematic y-offset changes indicated that minimum firing rate did not change across conditions. Unlike for gain, there was no dependence of z-offset or y-offset on noise level. Therefore, the primary consequence of decreasing stimulus contrast is that cortical cells increase their gain.

By transforming output nonlinearities across conditions, neural responses to each contrast stimulus could be predicted as successfully as by using separate nonlinearities for each condition as described above (median difference in prediction scores of 0.7%; sign-rank, $P > 0.5$).

4.3.4.1 Coding accuracy

These effects are similar to the changes in coding accuracy previously observed in the IC. [Dean et al. \[2005\]](#) presented white noise stimuli to guinea pigs, where the amplitude of the stimulus was varied every 50 ms. The sound level of the envelope in any block was drawn from a particular distribution, much like the individual tones used here in the DRCs (see Section 2.1.3.3). They found that IC neurons adapt to the shape of this level distribution, by repositioning their stimulus-response functions to maximise the Fisher information over the presented distribution.

To compare the results observed here with those of [Dean et al. \[2005\]](#), I calculated the Fisher information for each unit along its (parameterised) output nonlinearity (see Section 4.2.7). This quantifies the extent to which an ideal observer can decode the current position of the stimulus along the dimension spanned by the unit's STRF — i.e. the current value of z — based on the firing rate y of the neuron.

The position of maximum $I_F[z]$, z^* , shifted with the contrast of the stimulus (Figure 4.10). An approximately 3-fold reduction of σ_L from 8.7 dB to 2.9 dB reduced z^* by a median factor of 2.45; an approximately 1.5-fold reduction of σ_L to 5.8 dB reduced z^* by a median factor of 1.39. Neurons in auditory cortex thus adapt their coding strategy to be most informative about stimuli within the current stimulus distribution.

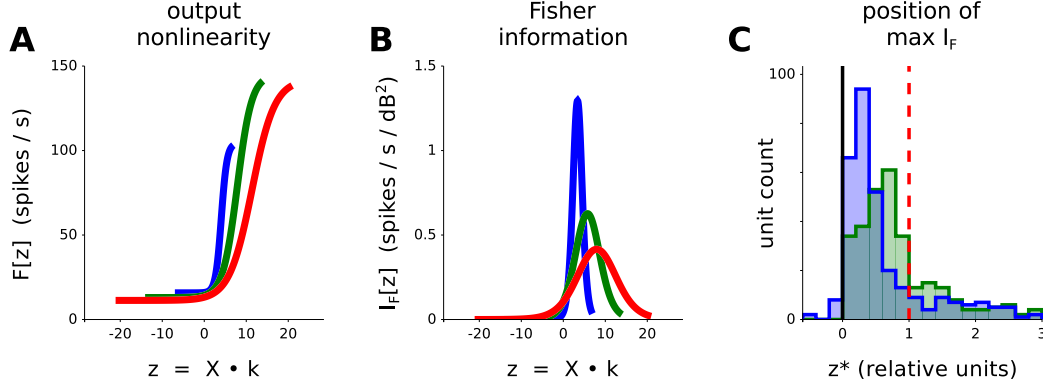


Figure 4.10: Neurons adapt their stimulus-response relationship to maximise Fisher information for the presented distribution. (A-B) The output nonlinearities for an example unit (A) not only tell us about the overall gain of that unit under different stimulus conditions, but also to which region of stimulus space that unit is informative. The Fisher information, $I_F[z]$ (B) quantifies this by providing a bound on the accuracy of a stimulus decoder based on this neuron’s firing rate. The neuron is most informative about stimuli X whose projection, z , onto the (normalised) STRF, k , lies at the peak of this curve. This occurs at a different value of z under low, intermediate, and high contrast stimulus conditions. (C) The stimulus of maximum $I_F[z]$ across the population of neurons shifts as contrast changes. For each unit, the value $z^* = \operatorname{argmax}_z (I_F[z])$ was determined for each contrast condition, and normalised relative to the z^* value for the high contrast condition.

4.3.4.2 Control experiment: total sound level

I have here presented data which show that neural coding properties change as a result of changing the contrast (via σ_L) of the DRC sequences. However, as noted in Section 4.2.2, and documented in Table 4.1, although the mean tone level, μ_L , was kept constant, increasing σ_L of the tone distributions also increased the (average) total sound level of the overall DRC stimulus, L_{total} . This is because L_{total} (in dB) is not a linear sum of the tone levels (also in dB), but rather:

$$L_{total,t} = 20 \cdot \log_{10} \left[\sum_f 10^{X_{tf}/20} \right] \quad (4.8)$$

Due to this nonlinear transform, the high contrast condition was an average 3.5 dB louder than the low contrast condition. This difference is small, but audible — humans can discriminate sound level differences of approximately 1 dB

[Green, 1993]. Though there is no data on intensity discrimination thresholds in ferrets, the difference in the overall level of DRCs lies above the (pure tone) intensity discrimination thresholds of rats [Hack, 1971], cats [Rosenzweig, 1946; Elliott and McGee, 1965], and macaques [Sinnott et al., 1985]. This could, in theory, account for any differences in STRFs or output nonlinearities across conditions.

To explicitly test such a possibility, I presented a control set of DRCs whose σ_L was varied, but with L_{total} kept constant. This required raising μ_L (the mean tone level) for low σ_L DRCs by 3.5 dB. For human listeners, this was sufficient to remove the percept of an audible total level difference between the different stimuli.

I presented these L_{total} -adjusted DRCs to a subset of neurons, and performed the same analyses. The results were qualitatively identical to that reported above (data not shown). Thus changes in neuronal gain cannot be attributed to changes in the overall level of the stimuli.

A final possibility is that neurons are affected by the overall stimulus level within some frequency range, such as that defined by their receptive field, or by the width of a single peripheral critical band. However, any differences in the overall sound level across contrast conditions within such a band is bounded above by differences in L_{total} (in the limit of large band sizes), and bounded below by zero (in the limit of small band sizes). Thus the control experiment of this section argues that the changes in neural coding are due to manipulations of σ_L .

4.3.5 Contrast gain control

To fully quantify the relationship between stimulus contrast and gain, I presented to a subset of these cells a larger set of DRCs, with 8 different σ_L values ranging from 1.4 dB to 11.5 dB ($c = 17\%$ to 116%). I obtained 80 units for which the above analysis could be performed over the whole contrast range. On average, these showed a clear, monotonic increase in gain as the contrast of the stimulus was reduced (Figure 4.11). The relationship between relative gain and contrast was extremely well described by a standard normalisation equation ([Heeger, 1992; Carandini et al., 1997]; see Section 4.2.8):

$$G(\sigma_L) = \frac{A}{1 + B \cdot \sigma_L^n} \quad (4.9)$$

where G denotes the gain, and A , B and n are constants. Maximum likelihood fits for the parameters were $n = 1.31$ and $B = 0.15$, with A constrained so that $G(\sigma_0) = 1$ (although values of n between 1.2 to 1.8 gave almost equally good fits). This model explained 99.9% of the variance in the population average of relative gain values.

This model also provided a good description of the relative gain values for individual units (Figure 4.12). However, in some units, the model failed at the lowest contrasts. For these units, gain increased as contrast was reduced down to a threshold, below which gain either levelled off, or decreased. For 46/80 units, this threshold was $\sigma_L = 1.4$ dB ($c = 17\%$); for a further 26 units, this threshold was $\sigma_L = 2.9$ dB ($c = 33\%$); and a further four units had a threshold of $\sigma_L = 4.3$ dB ($c = 49\%$). At these thresholds and above, gain was fitted well on a cell-by-cell basis by Equation (4.9) for 76/80 units. The model produced marginally better predictions of neural responses than fitting individual nonlinearities to each contrast condition (median improvement in

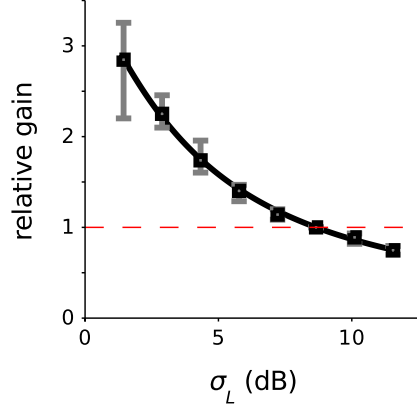


Figure 4.11: Gain as a function of contrast. The population median gain over a wide range of contrasts was well modelled by a standard gain normalisation function (Equation (4.9)). Gray bars indicate 99% confidence intervals on the median, via bootstrapping.

prediction scores of 0.3%, $P < 0.01$ by sign-rank test). Thus, across a wide range of contrasts, gain normalisation is a robust phenomenon for individual units.

Finally, the effect of contrast on the threshold and minimum firing rate, as per Figure 4.8 were characterised in more detail, given the broader range of σ_L values (Figure 4.13). The same pattern was observed: z-offsets increased slightly as contrast was reduced, and y-offsets remained constant. Z-offset changes were correlated with gain changes. Thus the contrast-dependent changes in nonlinearities can be summarised as:

$$\hat{y}_t = F[G(\sigma_L) \cdot (z_t - H)] \quad (4.10)$$

with $G(\sigma_L)$ given by Equation (4.9), and H constant.

4.3.6 Dependence on mean tone level

In the experiments presented so far, the mean SPL of each tone in the DRC, μ_L , was kept fixed. To explore the effect of mean tone level, I presented a

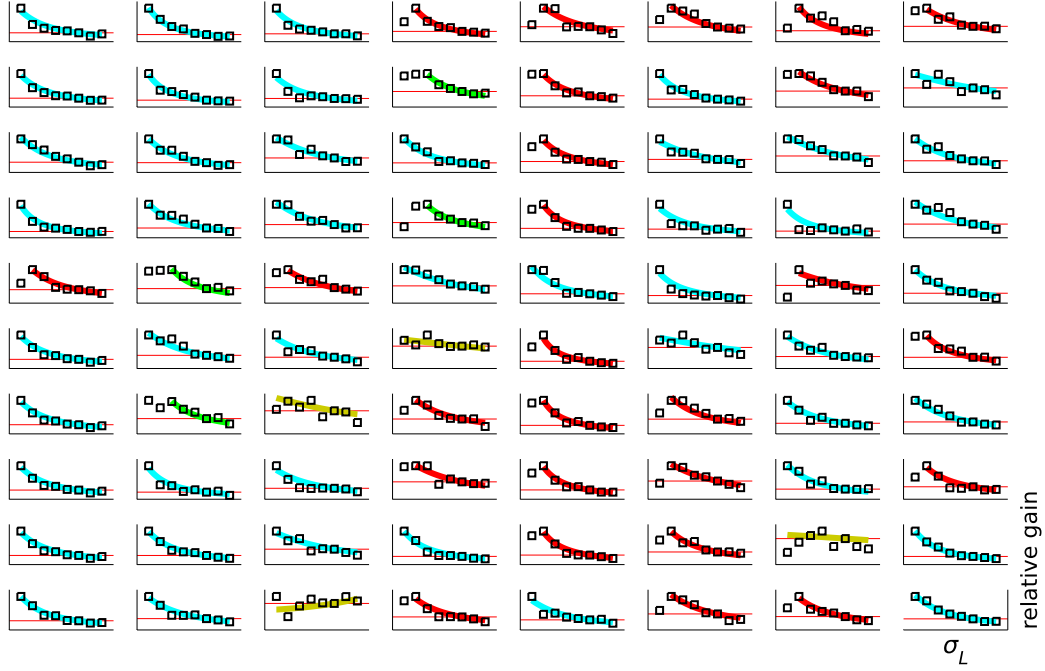


Figure 4.12: Gain as a function of σ_L : individual units. Relationship between contrast (σ_L ; abscissa) and gain (ordinate) is shown for 80 units. σ_L values as in Figure 4.11, thin red horizontal line denotes $G = 1$. Individual units show the same relationship between contrast and gain as the population average, although many failed to maintain gain control down to the lowest contrast conditions. The gain for individual units was fitted well by Equation (4.9) down to their respective maximum-gain contrast values (cyan: relationship held down to $\sigma_L = 1.4$ dB; red: relationship held down to $\sigma_L = 2.8$ dB; green: relationship held down to $\sigma_L = 4.3$ dB; yellow: remaining units). This demonstrates that gain normalisation is evident on a unit-by-unit basis, although individual units may have limits on how much they can compensate for arbitrarily low stimulus contrasts.

further set of stimuli in which both the mean of the level distributions (μ_L) and the contrast (σ_L) were manipulated independently. I estimated LN models from responses to a range of mean/contrast conditions, together with curve transformations from each stimulus condition relative to the $\mu_L = 40$ dB SPL, $\sigma_L = 8.7$ dB ($c = 92\%$) nonlinearity. Of the 1001 units above, 56 units yielded predictive LN models across the whole range of conditions. Only data from these 56 units are analysed below, in order to maintain the same sample set across stimulus conditions. Nevertheless, data from all units where LN models were predictive in only a subset of conditions ($n = 217$) yielded similar results.

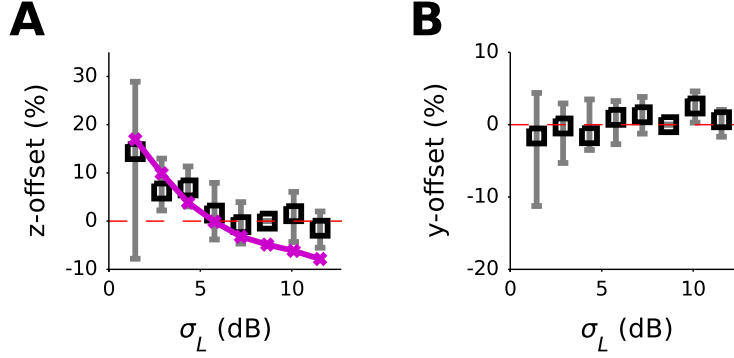


Figure 4.13: Contrast-dependent changes in threshold and saturation. Dependence of z-offset and y-offset values on a broad range of stimulus contrasts. These show the relationship between stimulus contrast (σ_L) and z-/y-offset in an equivalent manner to Figure 4.8B-C. Magenta line shows the predicted z-offset calculated from the (within-contrast) relationship between gain and z-offset in Figure 4.9B, and the dependence of gain on σ_L in Figure 4.11. The increase in z-offset as contrast is reduced can therefore be explained primarily by the pronounced increase in gain at low contrast. There is no systematic effect of stimulus contrast on y-offset.

At all mean levels tested, decreasing contrast caused gain to increase across the population of cells. However, the degree of gain normalisation depended on the mean level: at low μ_L , reducing contrast did not yield as much compensatory gain change compared with reducing contrast at high μ_L . While increasing μ_L therefore increased the dependence of gain on contrast, this trend saturated above $\mu_L \approx 35$ dB SPL (Figure 4.14A). At higher mean levels, gain was decoupled from the mean sound level, and varied with contrast alone. Interestingly, though changing mean level had no systematic effect on z-offset in this data (Figure 4.14B), it did on the y-offset: reducing the mean level typically increased the minimum firing rate (Figure 4.14C; examples in Figure 4.15).

Given the success of Equation (4.9) in modelling the relationship between σ_L and gain, I extended this model to include mean level, μ_L . The most explanatory model was a simple extension of the contrast-dependent model where B could vary with μ_L . This allows μ_L to directly modulate the dependence of gain on contrast:

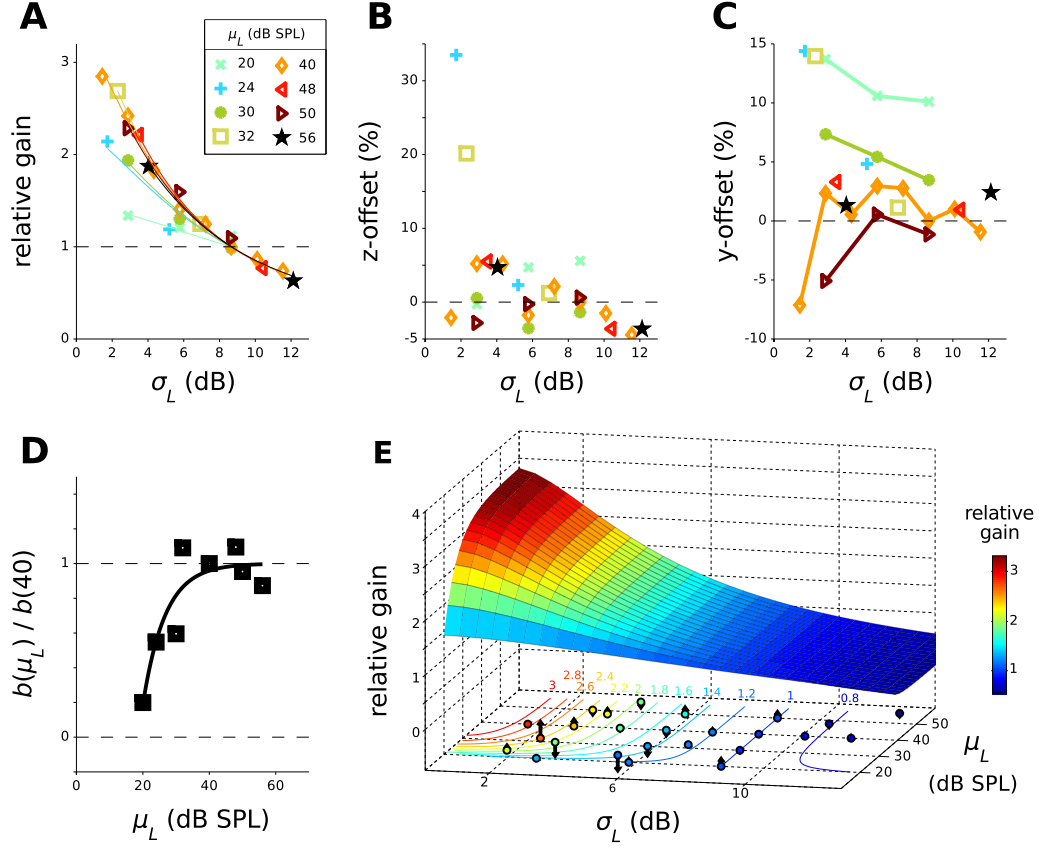


Figure 4.14: Contrast gain control is modulated by μ_L only at low mean levels. (A) Sigmoid nonlinearities were fitted to units' responses to a range of DRCs with different contrast and μ_L statistics, and compared with a reference curve. Population median gain factors are plotted against contrast, with different symbols/colours for each mean level. Coloured lines show independent fits of the model in Figure 4.11. (B) As in (A), for z-offset. (C) As in (A), for y-offset. (D) Values of B fitted to the model given in Equation (4.9), as shown in (A). These measure the relative sensitivity of neural gain to the stimulus contrast, as a function of mean level. At mean levels $\mu_L \geq 35$ dB SPL, B is relatively independent of μ_L , while it becomes sensitive to μ_L at lower mean levels. Solid line denotes an exponential fit to $B(\mu_L)$. (E) Illustration of the full model $G(\sigma_L, \mu_L)$ for gain normalisation (coloured surface). A contour plot of this surface is projected below. Coloured dots on the contour plot show the population median data used to constrain the model, from (A). Their position is a function of both σ_L and μ_L ; their colour denotes the median measured gain in these stimulus conditions; vertical arrows denote the residual between the measured gain and that described by the model.

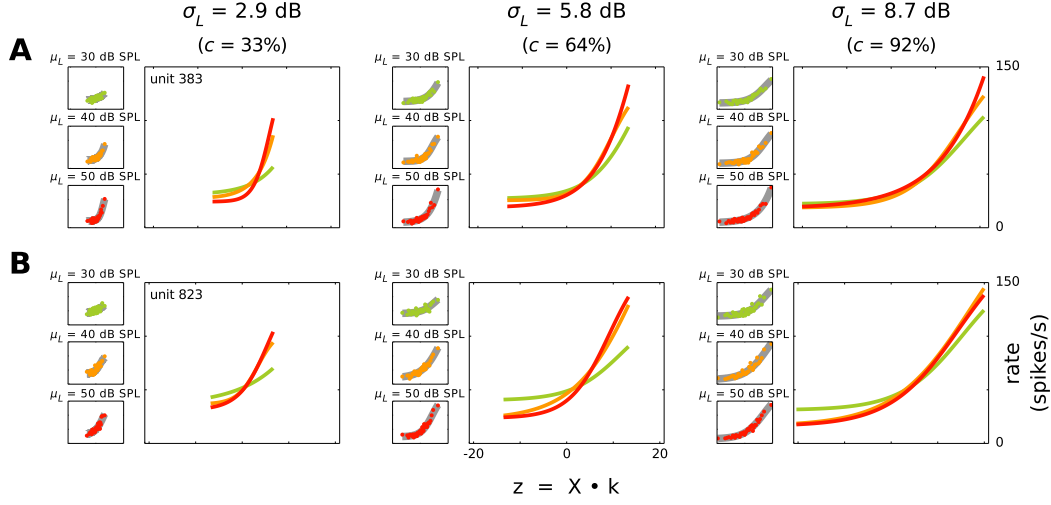


Figure 4.15: Output nonlinearities for two example units across different μ_L , σ_L conditions. (A) Output nonlinearities for a given unit for three contrast values (left, middle and right blocks), and three stimulus mean levels (small sub-blocks; green $\mu_L = 30$ dB, orange $\mu_L = 40$ dB, red $\mu_L = 50$ dB SPL). Larger panels show all fitted sigmoids for a given contrast but different μ_L values. For this unit, gain rescaling is strong at $\mu_L = 50$ dB and 40 dB, but relatively weak at $\mu_L = 30$ dB SPL. (B) As in (A), for a second example unit.

$$G(\sigma_L, \mu_L) = \frac{A(\mu_L)}{1 + B(\mu_L) \cdot \sigma_L^n} \quad (4.11)$$

Fitted values for $B(\mu_L)$ are presented in Figure 4.14D, showing that at low μ_L , B is modulated by μ_L , whereas B saturates with high μ_L . For simplicity, I modelled this with an exponential function:

$$B(\mu_L) = B_{max} (1 - e^{-c(\mu_L + k)}) \quad (4.12)$$

Fitted values of the parameters were $B_{max} = 0.17$, $c = 0.11$ and $k = -18.6$ dB. This model explained 97% of the total variance in the median population gains (Figure 4.14E). A schematic of the full stimulus-response model for auditory neurons is shown in Figure 4.16.

All of the above results remained unchanged when gain was expressed as a

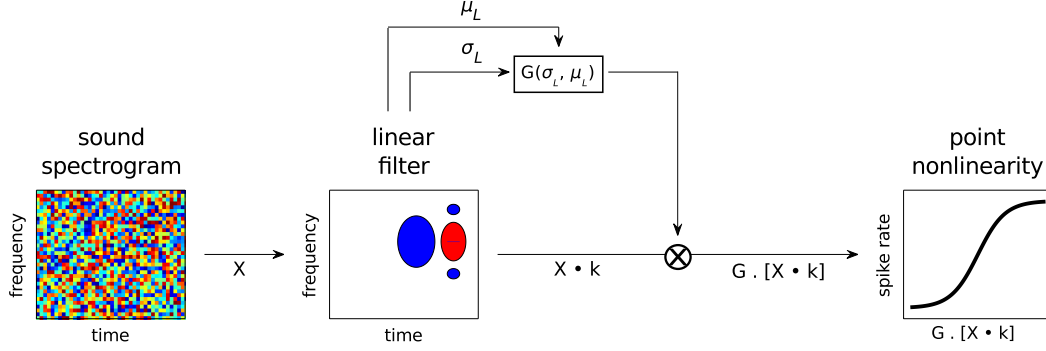


Figure 4.16: Schematic of gain control model.

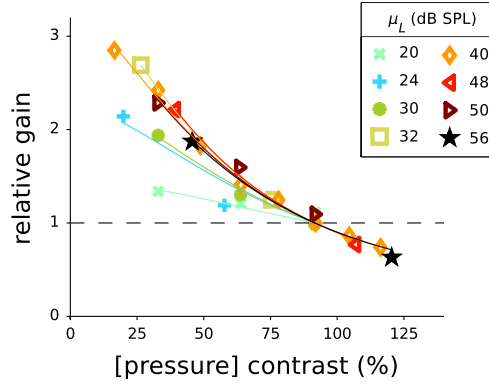


Figure 4.17: Contrast gain control expressed as a function of c . As explained in Section 2.1.3.2, I used the standard deviation of the distribution of tone levels (σ_L) as a measure of contrast, as this variable was directly controlled by varying the width of the tone level distribution, and it correlates very strongly with the definition of contrast in sound pressure space $c = \sigma_P / \mu_P$. For completeness, I provide here a version of Figure 4.14A where the abscissa is expressed directly in units of c , rather than in units of σ_L . All analyses and model fitting produced almost identical results when contrast was expressed as a percentage (via c) rather than in dB (via σ_L).

function of $c = \sigma_P / \mu_P$ rather than σ_L (Figure 4.17), confirming the validity of using σ_L as a measure of contrast (Section 2.1.3.2).

4.4 Discussion

The data presented in this chapter show that the gain of neurons in auditory cortex is dynamically modulated according to the spectrotemporal statistics of

recently-heard sounds. The primary determinant of gain is stimulus contrast, which is approximated well by the standard deviation of the sound pressure level (SPL; σ_L). Gain decreases as stimulus contrast increases, thereby partially compensating for changes in contrast. Mean stimulus level also influences gain: when the mean level is low, the effectiveness of contrast gain control is reduced.

4.4.1 The role of contrast gain control

Gain control is primarily useful for adapting the limited dynamic range of a neuron to the statistics of the stimulus. When spectrotemporal contrast is low, firing rates are sensitive to smaller changes within their receptive field than under higher contrast conditions. Thus the representation of stimulus space is effectively expanded under low contrast stimulation, and compressed under high contrast stimulation. Consequently, gain control should improve the ability of individuals to detect small changes in low-contrast sounds. Indeed, a related phenomenon has been demonstrated in the adaptation to reverberation, whereby listeners are better able to discriminate (low-contrast) reverberant words when embedded within a reverberant context sentence, than within a (high-contrast) anechoic context [Watkins, 2005], an effect which is also frequency-band specific [Watkins and Makin, 2007]. Perceptual adaptation is not, however, complete, as a general increase in the spectrotemporal contrast of speech still leads to demonstrable gains in intelligibility [Steeneken and Houtgast, 1980; van Veen and Houtgast, 1985; Miller et al., 1999]. The data presented here predict that perceptual adaptation to stimulus contrast should be observable with non-speech stimuli as well.

Neurons in the visual system are subject to contrast gain control, which is thought to be desirable for efficient coding of natural images [Schwartz and Simoncelli, 2001]. Since the contrast of natural images is correlated across space and time, normalisation by stimulus contrast reduces the redundancy

of the neural code [Barlow, 1961; Vinje and Gallant, 2002]. The contrast of a complex visual stimulus can be defined as σ_I/μ_I , which is strongly related to the two contrast measures which I have shown to determine auditory gain control (σ_L , Figure 4.11A; $c = \sigma_P/\mu_P$, Figure 4.17). Auditory gain control may therefore have a similar redundancy-reducing effect. Although the ensemble (i.e. long time scale) distributions of natural sounds have been explored [Attias and Schreiner, 1997; Escabi et al., 2003; Singh and Theunissen, 2003], a deeper understanding of the relationship between contrast gain control and the statistics of natural sounds will require a characterisation of natural sound level distributions at the spectral and temporal scales over which gain control operates.

4.4.2 Gain control vs. division of labour

In Chapter 1, the contrast problem was raised: in some environments, sound levels vary considerably over time, while in others, sound levels vary by only a small amount. Two potential solutions were offered to the contrast problem: the division of labour strategy, whereby neurons *specialise* for particular statistics; and gain control, whereby neurons *adapt* to the environmental statistics. This experiment argues that the latter solution is implemented by the auditory cortex. However, there are two important caveats.

Firstly, although the gain changes observed here are strong, they do not completely compensate for changes in stimulus contrast: even at high mean stimulus levels (where contrast gain control is most effective and independent of sound level), an approximately 3-fold reduction in spectrotemporal contrast only yields an approximately 2-fold increase in gain. Thus, gain control does not result in a complete normalisation (i.e. a total invariance) to stimulus contrast. This incompleteness of gain control also suggests that there is a preferred range of stimulus contrasts for which neural coding is optimal; outside this

range, gain control will fail to adjust gain enough to bring the stimuli into the neurons' dynamic range. It is possible that this preferred distribution is defined by the ensemble of natural sounds [Attneave, 1954; Barlow, 1961; Schwartz and Simoncelli, 2001; Lewicki, 2002].

Secondly, it does not appear that gain normalisation operates with equal measure from neuron to neuron. Not only does the strength of the effect differ across neurons, only a subset continues to increase their gain as stimulus contrast is reduced to ever smaller levels (Figure 4.12). This implies that different cortical neurons will be optimal encoders of different spectrotemporal level distributions. Similar diversity in adaptive properties has also been found in awake marmoset cortex, where subclasses of cells either adapt to the mean sound level of a stimulus, or maintain a fixed preference for a particular intensity range [Watkins and Barbour, 2008]. Just as such cells retain the ability to detect soft sounds in a loud environment, a variation in the degree of gain normalisation between neurons may help retain the ability to detect small changes in high contrast environments. These are particularly important tasks in audition, where superimposed sound sources need to be detected and dissected.

In total, while gain control is a feature of cortical neurons, the diversity in its expression may itself be a kind of specialisation for the contrast statistics in the ensemble of natural auditory scenes. If this is the case, then the gain behaviours of auditory cortical neurons may covary with other functional properties that would be particularly suited for processing sounds of particular statistics.

4.4.3 Predictive models for auditory neurons

In this chapter, I have advanced an extension of STRF/LN models for A1 neurons that explicitly includes a gain normalisation term. Given the magnitude of gain changes observed in this study, the inclusion of gain control should

prove to be a generally important factor in improving the predictive power of STRF models for cortical neurons. However, the implementation details may prove crucial, and depend on a characterisation of how spectral and temporal information are integrated to determine neural gain. This is the subject of the next two chapters.

Chapter 5

Spectral integration in contrast gain control

5.1 Introduction

A fundamental property of the central auditory system is that neurons integrate information over a particular spectral range. As explained in Section 2.2.1, this is a core principle that motivates the development of the STRF model: neurons determine their firing rate as a weighted combination of the sound level in different frequencies, and at different times. Level variations in only a subset of all sound frequencies are (linearly) responsible for firing rate variations. I shall refer to this region as the *(level) spectral receptive field*, or simply the *SRF*.

In the previous chapter, I demonstrated that auditory cortical neurons are subject to gain control, which depends on the spectrotemporal contrast of the stimulus. In these stimuli, contrast was only ever changed over the entire spectral range. Given the facts above, we might ask the same question for contrast as we do for sound level: what is the *contrast spectral receptive field* for an auditory cortical neuron, i.e. the range of frequencies whose contrast variations are responsible for gain variations?

This question has significance on several levels. From a computational perspective, if the contrast-SRF which determines gain matches the level-SRF which determines firing rate, then these neurons operate at a local level only. As a result, all the (gain-related) computations that the system performs as a whole must be comprised of interactions between such units. If this is not the case, i.e. if gain is determined from the contrast in a broader region of frequency space than the level-SRF, then the statistics of the wider spectral context influence the coding of spectrally local level variations.

From an implementation perspective, experimental results on this would constrain any possible mechanisms for gain control. For example, gain control could be the result of a local or a network mechanism. If a neuron's gain depends only on the statistics of the stimuli presented within its level-SRF, then gain control could be implemented by local mechanisms within or across individual neurons, for example by synaptic depression [Carandini et al., 2002]. However, synaptic depression is unlikely to account for gain effects that result from the statistics of stimuli outside the level-SRF, in which case gain control is more likely to arise from network mechanisms, such as the leveraging of balanced excitation and inhibition (e.g. [Abbott et al., 2005]).

From a modelling perspective, incorporating gain control is likely to improve characterisations and predictions of the firing patterns of auditory neurons. While the results of the previous chapter go some way towards estimating the gain changes that will be experienced by a neuron under different conditions, knowing how this operates in more spectrally complex situations is necessary for extending these results to more naturalistic stimulus paradigms. This is especially important as natural sounds do not have uniform statistics across all frequency bands [McDermott and Simoncelli, 2011].

To address these questions, I manipulated the contrast of DRCs in limited frequency ranges. I did this in two separate experiments, which are described

below. The first experiment (Section 5.2) provides a coarse estimate of whether gain changes are due to local contrast, i.e. the statistics within a neuron's level-SRF, or whether they are due to the global contrast statistics of a stimulus. To achieve this, I divided up the frequency bands of the DRC into two spectral regions — one positioned over a given neuron's level-SRF, and the other constituting the remaining frequency bands — and then I varied the contrast independently in these two regions. This allowed for a measurement of how much the local and remote contrasts contribute to that neuron's gain.

The second experiment (Section 5.4) refines this view considerably, by measuring how the contrast in different frequency bands contributes to a neuron's gain. This required a type of DRC stimulus where the widths of the tone level distributions were different for every frequency band. To analyse neurons' responses to these, I designed a class of models for characterising stimulus-response relationships, which extends the STRF and LN framework presented in Section 2.2. These new models, which are introduced in Section 5.3, define *contrast kernels* for neurons, which capture how their major response properties change with the contrast statistics of a sound. Contrast kernels are analogous to STRFs, in that they have a frequency component (explored in Section 5.4), and a temporal component (explored in Section 6.4). In this chapter, I focus on the former components, as the contrast frequency kernels of auditory cortical neurons provide an estimate of how these neurons integrate the spectral profile of contrast to determine their gain.

5.2 Spectral Experiment 1: Local and global contributions to gain

5.2.1 Methods

5.2.1.1 Animals and recordings

Data in this experiment were obtained from three of the ferrets used in Chapter 4. A full description of the anaesthesia, surgical, stimulation and recording procedures is provided in Section 2.3. As described below in Section 5.2.1.2, the stimuli were tailored to the response properties of units being recorded. Thus the units whose responses form the basis for this experiment form a subset of those described in the previous chapter, with the stimuli of this experiment being presented immediately after the standard contrast DRCs of Chapter 4.

5.2.1.2 Stimuli

The goal of these experiments was to determine the spectral region whose contrast statistics influence a given neuron's gain. As a first pass to answering this question, I divided up the 34 frequency bands (Φ) used for the DRCs (as constructed in Chapter 4) into two groups. One group, denoted here as the *test band*, Φ_{test} , comprised a contiguous set of 7 of the 34 tone frequencies. The remainder I denote as the *mask band*, Φ_{mask} . As is described below, varying the contrast independently in these two bands provided the means for estimating how contrast in different regions of frequency space influences a neuron's gain.

The key to this approach lay in situating the test band around a given unit's best frequency (BF). During recording, units' STRFs and best frequencies (BFs) were estimated on-line. A small number of representative units were chosen from a given electrode penetration based on the criteria of a low noise

ratio (NR; see Section 2.3.5.2) and a clearly defined STRF. As simultaneous recordings were made from both hemispheres, and as recording electrodes were often not inserted perfectly perpendicular to the cortical surface, there was often variation across the BFs of the units being recorded at any one time. Given that stationary recordings could only be maintained for a limited time, it was generally not possible to present test/mask DRCs tailored to each individual unit. Thus this dataset contains units where the test band precisely overlaps the SRF, and others where they do not.

For a representative unit, the test and mask bands were defined. The test band usually included 7 frequencies, spanning 1.2 octaves centred around the unit's BF. The goal was to contain the unit's level-SRF within the test band. In particular, the test and mask were chosen so that all the tone frequencies whose level variations additively determined a unit's firing rate via its STRF would lie in the test band, while the mask band would contain only tone frequencies whose STRF components were zero (see below, Section 5.2.1.4). Some units' BFs lay in the 2-3 highest frequency bands of the DRCs; for these units, the test band was reduced in size to cover only 3 or 4 tone frequencies. Results from these units were similar to those for which the test included 7 frequencies, and so results from all three cases were pooled.

In turn, for each test/mask definition, a set of 9 *test/mask DRCs* was constructed. In this set, the levels of the tones in the test frequency bands were drawn from a different distribution from the tones in the mask frequency bands. As for the first corpus of DRCs in Chapter 4, the tone level distributions were uniform in SPL space, with mean tone levels of $\mu_L = 40$ dB SPL, and halfwidths of $w_L = 5$ dB ($\sigma_L = 2.9$ dB; $c = \mu_P/\sigma_P = 33\%$; low contrast), $w_L = 10$ dB ($\sigma_L = 5.8$ dB; $c = 64\%$; intermediate contrast) or $w_L = 15$ dB ($\sigma_L = 8.7$ dB; $c = 92\%$; high contrast). A full set of statistics for these conditions is provided in Table 5.1. The stimulus set for an example unit is illustrated in Figure 5.1.

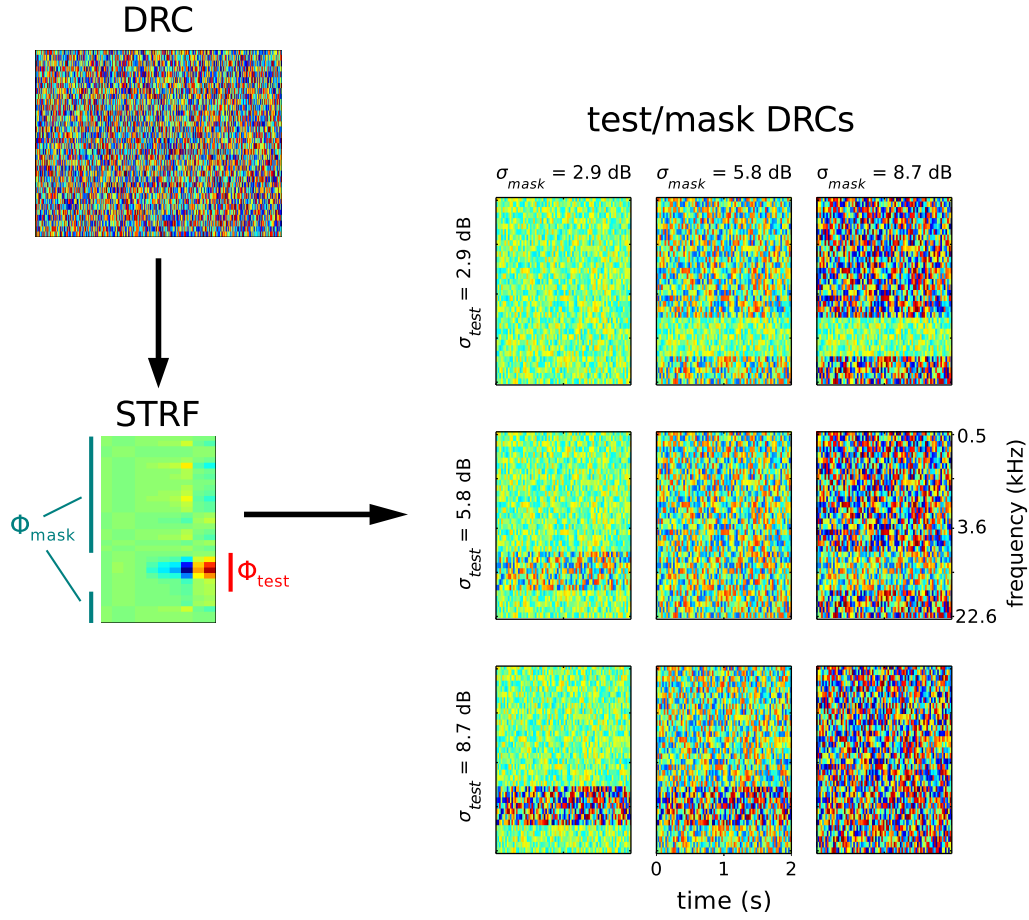


Figure 5.1: Schematic of test/mask DRC stimuli. During each electrode penetration, the STRF of a representative unit was mapped from responses to the DRCs presented in Chapter 4. A 0.5 - 1.2 octave band centred around the best frequency (BF) of that unit was designated the test band (Φ_{test}), while the set of remaining frequency bands was designated the mask (Φ_{mask}). A new set of *test/mask DRCs* was constructed, where the contrasts in the test and mask bands were varied independently. The unit pictured here had a BF of 9 kHz.

The 9 test/mask DRCs were presented five times each, randomly interleaved, with 15-20 s silence between each sequence. The first 2 s of data from each presentation were discarded to ensure that a constant adaptation state had been reached.

	tone level distribution			tone pressure distribution			(total) DRC level		
	μ_L [dB SPL]	w [dB]	σ_L [dB]	μ_P [mPa]	σ_P [mPa]	c [%]	min [dB SPL]	RMS [dB SPL]	max [dB SPL]
Test:	40	5	2.9	2.1	0.7	33%	69.1	71.1	72.6
Mask:		5	2.9	2.1	0.7	33%			
Test:	40	5	2.9	2.1	0.7	33%	68.3	72.2	74.7
Mask:		10	5.8	2.5	1.6	64%			
Test:	40	5	2.9	2.1	0.7	33%	68.1	73.9	77.4
Mask:		15	8.7	3.1	2.9	92%			
Test:	40	10	5.8	2.5	1.6	64%	69.3	71.4	73.4
Mask:		5	2.9	2.1	0.7	33%			
Test:	40	10	5.8	2.5	1.6	64%	68.5	72.5	75.3
Mask:		10	5.8	2.5	1.6	64%			
Test:	40	10	5.8	2.5	1.6	64%	68.3	74.2	77.9
Mask:		15	8.7	3.1	2.9	92%			
Test:	40	15	8.7	3.1	2.9	92%	69.2	71.9	74.7
Mask:		5	2.9	2.1	0.7	33%			
Test:	40	15	8.7	3.1	2.9	92%	69	72.9	76.4
Mask:		10	5.8	2.5	1.6	64%			
Test:	40	15	8.7	3.1	2.9	92%	68.9	74.6	78.6
Mask:		15	8.7	3.1	2.9	92%			

Table 5.1: Summary of stimulus statistics for test/mask DRCs. Tone level and tone pressure statistics have been calculated from the respective analytic distributions of these variables given in Section 2.1.3.1. Statistics of the total level of the DRC (last three columns of the table) have been estimated from sample DRCs of 30 s duration.

5.2.1.3 STRF and LN models

For all units, a linear STRF was calculated from the pooled set of responses to the test/mask DRCs. Although STRFs had been calculated for the same electrode penetrations using the standard contrast DRCs described in the previous chapter, non-stationarities in the recordings (for example, due to electrode or physiological drift) often led to small divergences between the STRFs estimated from these two sets of stimuli.

Next, separate linear-nonlinear (LN) models were fitted to units' responses in

each condition, as per Sections 2.2.2 and 4.2.5. All LN models for a given unit used the same shared STRF, in the manner illustrated in Figure 4.6.

Comparisons between output nonlinearities were performed in the same manner as in Chapter 4. The reference condition was chosen as the $\sigma_{test} = \sigma_{mask} = 8.7$ dB condition, and the curve transformations required to map the parameterised sigmoid nonlinearity from the reference condition to those obtained in other conditions were estimated as per Section 4.2.6.

5.2.1.4 Overlap between the test band and the SRF

To estimate the extent to which the test band covered the full extent of units' receptive fields, I formally defined the *spectral receptive field*, Φ_{RF} , of a unit to be those coefficients of the frequency component of the STRF (k_f) that were significantly non-zero, via a bootstrap procedure. This involved constructing confidence intervals for the components of k_f : I drew 1000 subsamples (with replacement) of 90% of the stimulus/response data from the pooled dataset, and estimated a STRF from each. Those frequencies whose k_f coefficients were significantly non-zero across the ensemble of STRFs (via studentisation, $P < 0.01$) were included in Φ_{RF} . The remainder were set to zero for the purposes of the measures below.

The overlap between Φ_{RF} and Φ_{test} was then defined to be:

$$\text{overlap} = \frac{\sum_{f \in \Phi_{test}} |k_f|}{\sum_{f \in \Phi} |k_f|} = \frac{\sum_{f \in \Phi_{test}} |k_f|}{\sum_{f \in \Phi_{RF}} |k_f|} \quad (5.1)$$

This measure of overlap is illustrated in Figure 5.2A.

5.2.1.5 Unit selection criteria

Units were selected for analysis in this study based on three criteria. Firstly, only units which modulated their firing rate in response to the DRC were included for analysis; thus, as in Chapter 4, units with a noise ratio $NR > 10$ in any one stimulus condition, i.e. whose explainable variance was less than 9.1% of the total variance, were excluded.

Secondly, in order for the gain in each $(\sigma_{test}, \sigma_{mask})$ condition to be accurately estimated, the LN models that were fitted for each of the nine stimulus conditions were required to be predictive (as defined in Section 4.2.4).

Finally, since the aim of this experiment was to estimate the relative contributions of the contrast in local and remote frequency bands to the gain of a unit, it was necessary to exclude any units which did not show contrast-dependent gain changes at all. This was enforced by requiring that the curve transformation between the all-high-contrast ($\sigma_{test} = \sigma_{mask} = 8.7$ dB) and all-low-contrast ($\sigma_{test} = \sigma_{mask} = 2.9$ dB) conditions involved a horizontal compression factor of $G > 1.1$. Amongst 61 units which satisfied the first two criteria, only three ($< 5\%$) were excluded on this basis.

5.2.1.6 Normalisation models

The goal of this work was to estimate the extent to which a neuron's gain is determined by stimulus contrast within its SRF, and stimulus contrast outside its SRF. For brevity, I shall refer to these respectively as the *local contrast* (i.e. the contrast in frequency bands *local* to the unit's BF; σ_{local}) and the *remote contrast* (i.e. the contrast in frequency bands remote from the unit's BF; σ_{remote}).

As is explained below in Section 5.2.2.2, I modelled how contrasts in different regions of frequency space contribute to a neuron’s gain. I extended the gain normalisation equation used in Chapter 4 (Equation (5.2)), to include sensitivity to two separate contrast measures (Equation (5.3)). As in Equation (4.9), A was constrained so that $G(\sigma_1 = \sigma_0, \sigma_2 = \sigma_0) = 1$, where $\sigma_0 = 8.7$ dB; and $n = 1.31$ was fixed at the maximum likelihood value fitted in Section 4.3.5. The coefficients B_1 and B_2 measure the respective contributions of the two contrasts to the gain.

5.2.2 Results

I recorded from 448 sites in the primary auditory cortex (A1/AAF) of three anaesthetised ferrets. The goal was to determine to what extent the gain of these units depended on the contrast in frequency bands local to the unit’s best frequency (BF), compared with the contribution from the contrast in more distant frequency bands.

To address this, I varied the contrast of DRC stimuli within two separate bands. One region was denoted the *test band* (Φ_{test}), centred around a chosen unit’s BF, and spanning 0.5, 0.6, or 1.2 octaves. The remaining tone frequencies were denoted the *mask band* (Φ_{mask}). I aimed to situate the test band over the *spectral receptive field* (Φ_{RF} ; see Section 5.2.1.2). The spectral receptive field was defined to cover the significantly non-zero coefficients of the frequency component of the chosen unit’s STRF; the goal of this construction was therefore to capture within the test band all the frequencies whose level variations additively influenced the chosen unit’s firing rate. I presented nine separate DRCs, where the contrasts in the test (σ_{test}) and mask (σ_{mask}) were independently chosen from $\sigma_L = 2.9$ dB, 5.8 dB, or 8.7 dB ($c = 33\%$, 64%, or 92%). The stimulus set is illustrated for an example unit in Figure 5.1.

I identified 97 units which responded reliably to the test/mask DRCs. Amongst these, 61 were well described by predictive LN models. In turn 58 of these showed evidence of contrast-dependent gain changes (see Section 5.2.1.5).

Given that multiple units were recorded simultaneously (usually bilaterally), these units were sampled from a range of conditions where the test band covered a unit's spectral receptive field, overlapped it, or lay entirely outside it (Figure 5.2A). I took advantage of this to measure how the gain depended on the amount of overlap between Φ_{test} and Φ_{RF} .

5.2.2.1 Contributions to gain from within and outside the tuning curve

The contrast in frequency bands nearby to the unit's BF proved to be strong determinants of a unit's gain. This was evident from two observations. Firstly, when the test band completely, or almost completely ($> 80\%$) covered a unit's Φ_{RF} , the test contrast greatly modulated the unit's gain. This can be seen on the right side of Figure 5.2B: as σ_{test} was reduced, those units with large overlap between Φ_{RF} and Φ_{test} increased their gain. Secondly, amongst those units for whom the test did not overlap Φ_{RF} , it was the *mask* contrast that most strongly determined response gain. This can be seen on the left side of Figure 5.2C: as σ_{mask} was reduced, those units with little overlap between Φ_{RF} and Φ_{test} increased their gain.

However, contrast in frequencies remote to units' BFs also had an impact on gain. This manifested in several ways. Firstly, even when the test band completely covered Φ_{RF} , decreasing σ_{mask} still resulted in an increase in gain (Figure 5.2C, right hand side). Secondly, there were interactions between local and remote contrasts. For example, amongst those units whose Φ_{RF} lay entirely inside the test band, the extent to which σ_{test} modulated was reduced by high

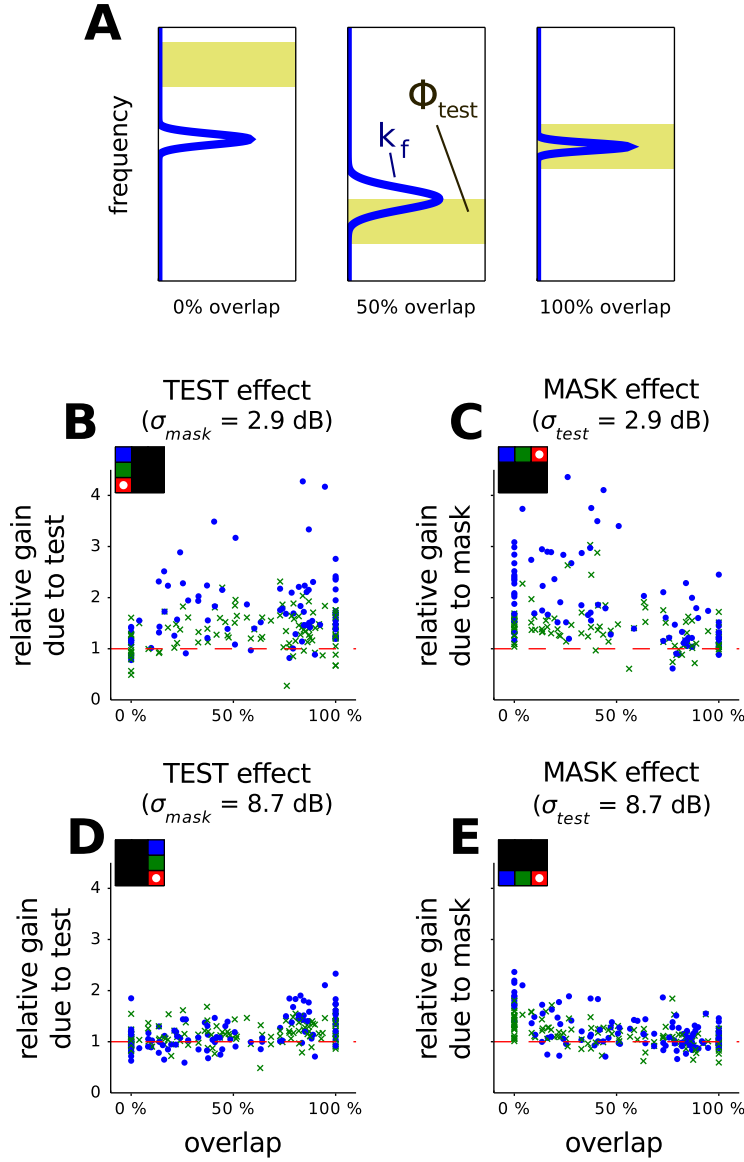


Figure 5.2: Changes in gain from varying σ_{test} and σ_{mask} . (A) Illustration of the measure of overlap between Φ_{RF} (the frequencies for which $|k_f|$, blue, is significantly non-zero) and Φ_{test} (yellow shaded region). (B-E) Relative gain changes from varying either σ_{test} or σ_{mask} , with the other kept constant. Colour boxes in the top left hand corner of each plot illustrate which stimulus conditions from Figure 5.1 are being compared; the box with the white dot is the reference curve for which the relative gain is $G = 1$. Each unit is associated with two points, corresponding to the relative gain in the green (intermediate contrast in either σ_{test} or σ_{mask}) and blue (low contrast in either σ_{test} or σ_{mask}) stimulus conditions. The abscissa measures the overlap between Φ_{RF} and Φ_{test} for that unit, as illustrated in (A).

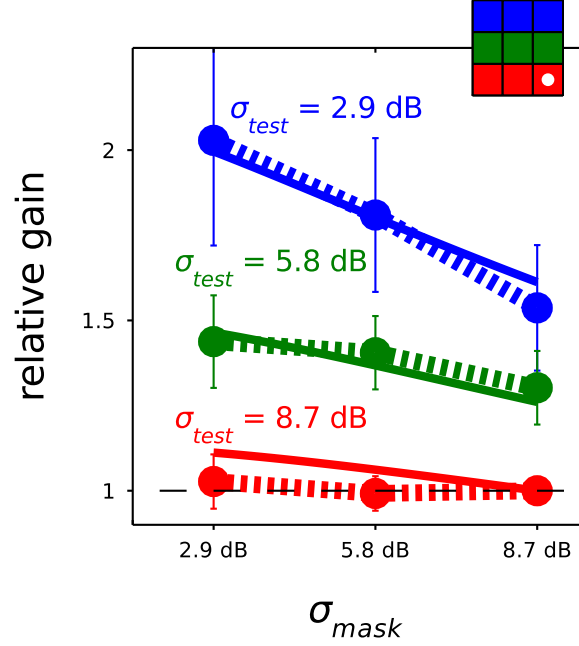


Figure 5.3: Summary of changes in gain from varying σ_{test} and σ_{mask} . Gain is affected by the contrast of stimuli lying both within and outside the tuning of a unit. Population median gain for 24 units with their spectral receptive field lying within the test (dots with dashed lines). Contrast within the test is a stronger determinant of the gain than that in the mask. Solid lines show the fit of Equation (5.3) to these data, with $\sigma_1 = \sigma_{local}$ and $\sigma_2 = \sigma_{global}$.

σ_{mask} (compare Figure 5.2B and D, right hand sides). The converse was also true amongst those units whose Φ_{RF} lay entirely inside the mask: here, although σ_{mask} was the primary determinant of gain, the strength of its effect was reduced at high σ_{test} (compare Figure 5.2C and E, left hand sides).

This dataset is summarised in Figure 5.3 for 24 units where the test band completely covered Φ_{RF} .

5.2.2.2 Measuring the relative contributions

To quantify these effects, I extended the (μ_L -independent) contrast gain control model (Equation (4.9)) to include contributions to gain normalisation from stimulus statistics both within and outside the spectral receptive field.

I fitted two classes of model to a training dataset. The training dataset consisted of 9 data points: the relative gain values for the 9 test/mask DRC conditions, averaged (median) across the 24 units where the test band completely covered Φ_{RF} (i.e. the data shown in Figure 5.3).

Models in the first class asserted that units' gain was a function of only one measure of stimulus contrast:

$$G(\sigma) = \frac{A}{1 + B \cdot \sigma^n} \quad (5.2)$$

Models in the second class asserted that units' gain was a function of two measures of stimulus contrast:

$$G(\sigma_1, \sigma_2) = \frac{A}{1 + B_1 \cdot \sigma_1^n + B_2 \cdot \sigma_2^n} \quad (5.3)$$

The measures of stimulus contrast used in these models were sourced from the following list:

- the contrast of the test band, σ_{test}
- the contrast of the mask band, σ_{mask}
- a measure of *local contrast*, defined via the average variance in tone levels within Φ_{RF} :

$$\sigma_{local}^2 = \frac{1}{|\Phi_{RF}|} \sum_{f \in \Phi_{RF}} \sigma_L^2(f) \quad (5.4)$$

- a measure of *remote contrast*, defined via the average variance in tone levels outside Φ_{RF} :

$$\sigma_{remote}^2 = \frac{1}{|\Phi| - |\Phi_{RF}|} \sum_{f \notin \Phi_{RF}} \sigma_L^2(f) \quad (5.5)$$

- a measure of the *global contrast*, defined via the average level variance across all frequency bands, Φ :

$$\sigma_{global}^2 = \frac{1}{|\Phi|} \sum_{f \in \Phi} \sigma_L^2(f) \quad (5.6)$$

where $\sigma_L(f)$ is the standard deviation of the levels in frequency band f . The choice of the exponent of $\sigma_L(f)$ in the sums above was arbitrary: the range of conditions I explored was insufficient to constrain this. Similar results were obtained when weighted sums of the form $\sigma_j = \frac{1}{|\Phi_i|} \sum_{f \in \Phi_i} \sigma_L(f)$ were used.

To simplify the training data, it was assumed that $\Phi_{RF} = \Phi_{test}$ for all 24 units, i.e. that the test band covered exactly the spectral receptive field. This constrained $\sigma_{local} = \sigma_{test}$ and $\sigma_{remote} = \sigma_{mask}$. The parameter n was fixed with a value of $n = 1.31$, given from the results of Section 4.3.5; A was constrained as per Section 4.2.8. Thus Equation (5.2) has only one free parameter, and Equation (5.3) has two.

Each of the fitted models were tested via a validation and a prediction. Firstly, the model, whose parameters were fitted to the population median data, was used to calculate the expected relative gain of the 24 individual units where the test band completely covered Φ_{RF} . This used the unit-by-unit measurement of the five contrast measures as listed above. Secondly, the model was used to predict the relative gain for 34 individual units where the overlap between Φ_{test} and Φ_{RF} was incomplete or zero. Again, this used the unit-by-unit measurement of the five contrast measures as listed above.

The results are reported in Table 5.2. When gain was modelled as a function of only one measure of contrast, a dependence on σ_{local} provided the best validation and prediction scores. This indicates that neural gain is most powerfully modulated by stimulus variance within the frequency bands closest to its BF.

Model	%VE		
	Training	Validation	Prediction
single: $G(\sigma) = \frac{A}{1+B \cdot \sigma^n}$			
$\sigma = \sigma_{test}$	64%	60%	-30%
$\sigma = \sigma_{local}$	64%	62%	76%
$\sigma = \sigma_{mask}$	-18%	-16%	62%
$\sigma = \sigma_{remote}$	-18%	4%	52%
$\sigma = \sigma_{global}$	15%	7%	72%
weighted: $G(\sigma_1, \sigma_2) = \frac{A}{1+B_1 \cdot \sigma_1^n + B_2 \cdot \sigma_2^n}$			
$\sigma_1 = \sigma_{test}, \sigma_2 = \sigma_{mask}$	75%	73%	11%
$\sigma_1 = \sigma_{local}, \sigma_2 = \sigma_{remote}$	77%	65%	83%
$\sigma_1 = \sigma_{local}, \sigma_2 = \sigma_{global}$	77%	71%	83%

Table 5.2: Success of test/mask models. Second to fourth columns show the percentage of variance explained (%VE) in the respective datasets. Each model was fitted using the median relative gain values across 24 units for whom Φ_{RF} was completely covered by Φ_{test} (training data). Models were validated using the gain values for the 24 individual units (validation data). Finally, the models were used to predict the gain values for 34 units for which there was some, or no overlap between Φ_{RF} and Φ_{test} . The best fitting models are highlighted in bold.

Improvements to the model fit to all data sets were obtained by allowing gain to be modelled as a weighted function of both a local measure of stimulus contrast, and a remote (or global) measure. In particular, including σ_{remote} or σ_{global} alongside σ_{local} produced better validation and prediction scores than using the σ_{local} model alone.

When σ_{local} was paired with σ_{remote} , the fit estimated a 5.4x stronger weighting of local stimulus contrast over the contrast in remote frequency bands, with $B_{local} = 0.066$ and $B_{remote} = 0.012$. When σ_{local} was paired with σ_{global} , the fit estimated a 2.4x stronger weighting of local over global stimulus, with $B_{local} = 0.079$ and $B_{remote} = 0.033$.

Results from either model indicate that the statistics in frequency bands near to a unit's BF have a greater influence on that unit's gain than the statistics in more distant frequency bands. The fits of the local/global model are also shown in Figure 5.3. The predictive value of these two models point to either

the existence of a gain control mechanism that strongly weights local over remote stimulus statistics, or else to the presence of two stages of gain control: one local, one global.

5.2.3 Preliminary Discussion

In this experiment, I estimated how the gain of an auditory cortical neuron depends on the contrast within its preferred frequency range (the *local contrast*), and the contrast in more distant frequency bands (the *remote* or *global contrast*). This was done by designing special DRCs stimuli, where a *test band* of frequencies, covering roughly the same spectral range as a unit's spectral receptive field (SRF), had contrast that was manipulated independently of the remaining *mask band* of frequencies. The results show that neuron's gain is more strongly modulated by the local contrast. There is, however, a smaller, but non-significant contribution to gain from the contrast in more distant frequency bands.

This experiment reveals that, at some point in the auditory pathway, the appropriate level of gain for neurons in auditory cortex is determined by a non-uniform integration of stimulus statistics across the spectral range. In particular, when the contrast is large in a band-limited region of a stimulus, neurons representing stimulus dynamics near to this region will experience a greater gain compression than neurons representing distant frequency bands.

To provide a more precise characterisation of this spectral integration, a more sophisticated stimulus paradigm is required. As the test/mask construction effectively subdivides the spectrum into two regions, it can provide only a coarse estimate along these boundaries of how spectral information is integrated. It is not, for instance, possible to determine how the statistics within individual frequency bands are weighted by a neuron to determine its gain. These

limitations inspired a second experiment, which is the subject of the next two sections.

5.3 Contrast kernel models

5.3.1 Introduction

STRF and LN models share a common set of goals. Most directly, they act as *forward models*: given a stimulus, they try to predict the pattern of firing that would be produced by an auditory neuron. This requires finding the properties of the stimulus which change the neuron’s firing rate. The better the predictions, the better the model.

When trying to understand the auditory brain in terms of *computation*, these models take on an additional function as *coding models*: they attempt to capture what a spike from that neuron “means” for the auditory brain. This is often viewed from an information perspective. If, for example, a fitted STRF indicates that a neuron modulates its firing rate as a function of changes in the level of a sound at a frequency of 8 kHz, then, given a spike train from that neuron, we can invert the forward model to extract information about the stimulus. In particular, we can retrieve information about the changes in sound level of the stimulus in the 8 kHz range. Some of this information is likely to be relevant to that neuron’s downstream targets.

It is therefore a common, and highly useful paradigm, to conceive of these models as probing the *representation* of sounds at a level of the auditory brain. If we consider a highly simplified feed-forward view of the auditory pathway, then the firing patterns in the auditory nerve carry the entirety of the information about an incoming sound that the cochlear nucleus gets to work with. The

cochlear nucleus transforms this representation of the stimulus, producing another pattern of spikes. In turn, the inferior colliculus does its computations on the representations of the stimulus it receives from the cochlear nucleus, as well as from other brainstem nuclei. With this in mind, when we build a successful forward model of the firing patterns of an auditory neuron, we obtain a view into the information it provides about the stimulus to its post-synaptic targets. When we build the models for a whole field of neurons, we obtain a view into the representation of the sound which that station, as a whole, provides. This representation, presumably, is the substrate for further computation.

In this section, I shall describe an extension of LN models. These explicitly allow the response properties of neurons, via the parameters of the static output nonlinearities, to depend on the *contrast* properties of a stimulus. I particularly focus on the spectral profile of stimulus contrast (explored in Section 5.4) and the temporal profile of stimulus contrast (explored in Section 6.4). Details of the models aside, they share the same set of goals as for STRF and LN models. First, they aim to be more sophisticated *forward* models, such that when we present a stimulus with a non-trivial contrast profile, we can make better predictions of what the firing rate of that neuron will be. Second, they aim to reveal something about *neural coding*. By seeing how the stimulus-selectivity of cortical neurons changes with contrast, we can understand how the representation of a sound in auditory cortex — *as a whole* — integrates, and is affected by stimulus contrast.

The structure of this section is as follows. In Section 5.3.2.1, I motivate how output nonlinearities can be rendered contrast-dependent. In Section 5.3.2.2, I construct the general form of the contrast kernel model. Section 5.3.2.3 is a small digression, wherein I compare the contrast kernel model to those used in previous sections of this thesis. In Section 5.3.2.4, I consider ways of simplifying contrast kernel models. These simplified forms are used in the experiment which

follows. In particular, I introduce the gain contrast kernel. In Section 5.3.2.5, I consider a number of extensions of the contrast kernel models which are also tested later in this chapter. Finally, in Section 5.3.3, I describe the Random Contrast DRC stimuli that are used to fit the models presented here.

5.3.2 Contrast kernels

5.3.2.1 Parameters of output nonlinearities

We begin by considering the LN model, depicted in Figure 2.6. In Section 4.3.3, I demonstrated that when stimulus contrast is changed, the linear filter (i.e. the STRF kernel) of a cortical neuron is unchanged. However, the shape of the static output nonlinearity — most importantly, its gain — changes with contrast.

If we denote the time-varying output of the linear STRF as z_t , then a sigmoidal output nonlinearity applies a function to z_t of the form:

$$\hat{y}_t = F[z_t] = a + \frac{b}{1 + \exp(-(z_t - c)/d)} \quad (5.7)$$

The four nonlinearity parameters each have a specific interpretation with respect to the shape of the sigmoid (Figure 5.4): a denotes the minimum (or threshold) firing rate; b denotes the dynamic range of the neuron, with $a + b$ being the maximum (saturating) firing rate; c denotes the inflection point; and d denotes the dilation (i.e. inverse gain), such that a larger d means a more horizontally dilated curve (lower gain), and a smaller d means a more horizontally compressed curve (higher gain).

A simple way to impart contrast-dependency on the output nonlinearity is to render each one of these four parameters dependent on the contrast of the

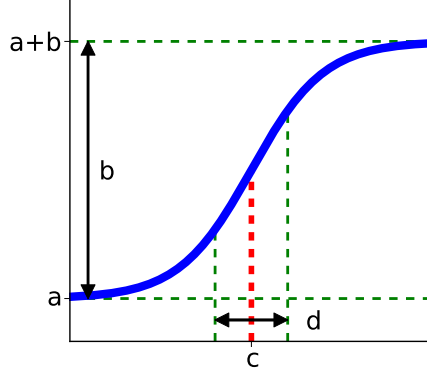


Figure 5.4: Parameters of the sigmoidal static nonlinearity.

stimulus, σ :

$$\hat{y}_t = F[z_t, \sigma] = a[\sigma] + \frac{b[\sigma]}{1 + \exp(-(z_t - c[\sigma])/d[\sigma])} \quad (5.8)$$

where each of the nonlinearity parameters a through d is a function of σ .

In the experimental stimuli considered here, the contrast, σ , changes with frequency and time. I will therefore represent contrast as a 2-tensor, σ_{tf} . Thus Equation (5.8) becomes:

$$\hat{y}_t = F[z_t, \sigma_{tf}] = a[\sigma_{tf}] + \frac{b[\sigma_{tf}]}{1 + \exp(-(z_t - c[\sigma_{tf}])/d[\sigma_{tf}])} \quad (5.9)$$

5.3.2.2 Contrast kernels

There is considerable freedom in Equation (5.9) to specify the form of the functions $a[\sigma_{tf}]$ through $d[\sigma_{tf}]$. The simplest assumption, which I consider here, is that these are linear functions of σ_{tf} . A motivation for this assumption, in relation to the results presented so far, is provided below in Section 5.3.2.3. Here, I shall present the form of the complete model, so that its parts can be dissected and understood.

The model takes the form:

$$\hat{y}_t = a_t + \frac{b_t}{1 + e^{-(z_t - c_t)/d_t}} \quad (5.10)$$

where, as previously, $\hat{y}_t$ is the predicted firing rate, and z_t the output of the linear STRF.

The subscript t on the values $\{a_t, b_t, c_t, d_t\}$ indicates that these values vary with time. These are calculated via:

$$a_t = a_0 + a_1 \cdot \zeta_t^{(a)} \quad \zeta_t^{(a)} = \sum_f \kappa_f^{(a)} \cdot \sigma_{tf} \quad (5.11)$$

$$b_t = b_0 + b_1 \cdot \zeta_t^{(b)} \quad \zeta_t^{(b)} = \sum_f \kappa_f^{(b)} \cdot \sigma_{tf} \quad (5.12)$$

$$c_t = c_0 + c_1 \cdot \zeta_t^{(c)} \quad \zeta_t^{(c)} = \sum_f \kappa_f^{(c)} \cdot \sigma_{tf} \quad (5.13)$$

$$d_t = d_0 + d_1 \cdot \zeta_t^{(d)} \quad \zeta_t^{(d)} = \sum_f \kappa_f^{(d)} \cdot \sigma_{tf} \quad (5.14)$$

Here, all of the parameters $\theta \in \{a, b, c, d\}$ of the output nonlinearity change linearly with the contrast of the stimulus. Since the contrast in each frequency band, σ_{tf} , varies with time, so each parameter θ varies with time. These changes are mediated via a weighted sum of the contrasts in different frequency bands, $\zeta_t^{(\theta)}$. The weighting is provided by the term $\kappa_f^{(\theta)}$, which I shall refer to as the *contrast (frequency) kernel* for the parameter θ . The parameter θ is sensitive to the contrast in those frequency bands where the coefficient of $\kappa_f^{(\theta)}$ is large in magnitude, and is insensitive to the contrast in those frequency bands where $\kappa_f^{(\theta)}$ is near zero. This is illustrated in Figure 5.5.

The notation has been chosen to draw a parallel between contrast kernels and STRF kernels. The STRF acts as a linear filter on the sound spectrogram, X_{tfh} , by projecting it onto a preferred direction, k_{fh} , to give z_t . The contrast kernel for a parameter θ , likewise acts as a linear filter on the ongoing contrast profile of a sound, σ_{tf} , by projecting onto a preferred direction, $\kappa_f^{(\theta)}$, to give $\zeta_t^{(\theta)}$.

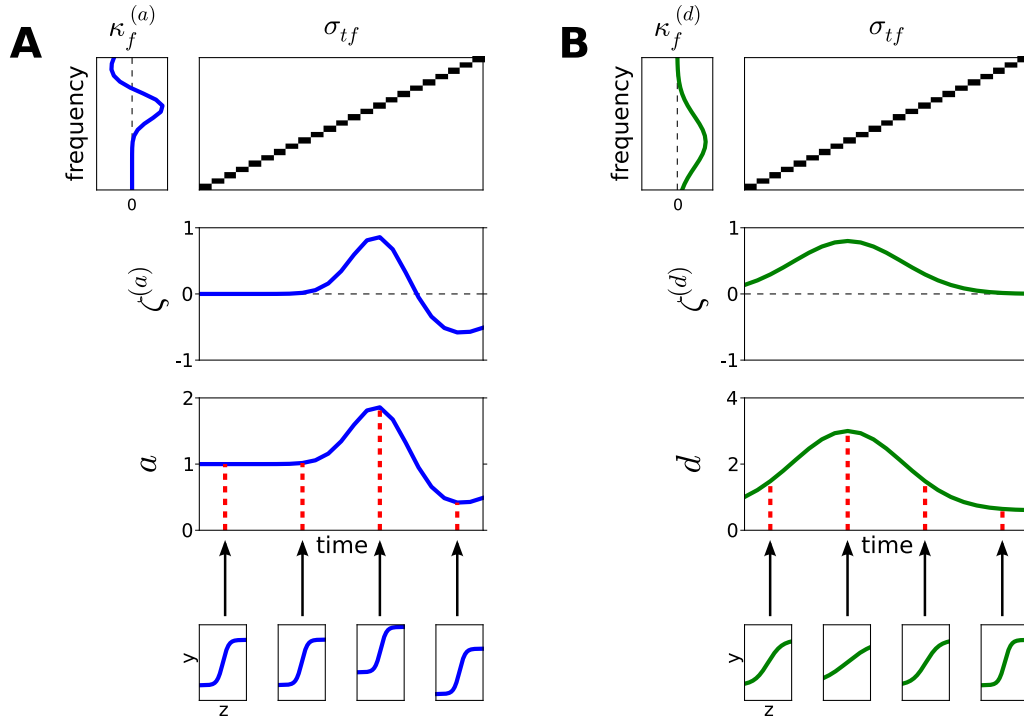


Figure 5.5: Illustration of the contrast kernel model. The diagram shows two examples of how a neuron's output nonlinearity might change over time, due to changing contrast. (A) Effect of a contrast kernel for the nonlinearity parameter a , which sets the minimum (threshold) firing rate. Top left panel: the contrast kernel, $\kappa_f^{(a)}$ is shown. This comprises a set of weights assigned to each frequency band, measuring the effect of the contrast in that band on a . Top panel: the contrast profile of an example stimulus. The abscissa denotes time, the ordinate denotes frequency. Dark areas indicate that the stimulus was high contrast in that band and at that time, white areas indicate that the stimulus was low contrast. Note that this is not the spectrogram, but is a measure of its statistics. At any time, this stimulus is low contrast within all frequency bands, except for within a single narrow band of high contrast; over time, the frequency of this high contrast band changes. Second panel from top: the weighted contrast signal, $\zeta_t^{(a)}$. This is the dot product of the contrast kernel with the contrast, from Equation (5.11), and shows how much the ongoing contrast profile “overlaps” with the contrast kernel. Third panel: as a result of changing contrast, the minimum firing rate, a changes with time. For this neuron, $a_0 = 1$ and $a_1 = 1$. Bottom: the output nonlinearities, as in Figure 5.4, change with time, via their minimum firing rate. These shifts would be combined with the contrast-dependent changes to the other nonlinearity parameters, b , c , and d . (B) Effect of a contrast kernel for the nonlinearity parameter d , which sets the (inverse) gain of the output nonlinearity. This neuron decreases its gain when there is high contrast anywhere within a relatively broad region demarcated by $\kappa_f^{(d)}$.

5.3.2.3 Consistency with previous models

The contrast kernel model advanced in Equations (5.10) – (5.14) is motivated from the results presented in this thesis so far. This section outlines the relationship between the contrast kernel model and those models fitted to those data.

In Chapter 4, I focused on contrast-dependent gain changes, and found that these could be excellently described by transforming the firing rate $F[z_t]$ into $F[G(\sigma) \cdot (z_t - H)]$, where G was given by:

$$G(\sigma) = \frac{A}{1 + B \cdot \sigma^n} \quad (5.15)$$

If we compare this with Equation (5.8), we see that this form for gain changes is equivalent to parameterising d to be linear in σ^n .

Earlier in this chapter, the gain measure of Equation (5.15) was extended to weight a local measure of contrast, σ_1 , with a remote or global measure of contrast, σ_2 (Equation (5.3)):

$$G(\sigma_1, \sigma_2) = \frac{A}{1 + B_1 \cdot \sigma_1^n + B_2 \cdot \sigma_2^n} \quad (5.16)$$

Again, this form of gain changes is equivalent to parameterising d to be a linear function of σ_1^n and σ_2^n .

Here, we are interested in how gain changes as a function of the spectral profile of contrast. Since this changes over time, this profile is denoted σ_{tf} . It is therefore natural to extend Equations (5.15) and (5.16) and allow the gain to vary with time, via:

$$G_t = G(\sigma_{tf}) = \frac{A}{1 + \sum_f B_f \cdot \sigma_{tf}^n} \quad (5.17)$$

Here, the gain of a unit at time t , G_t , depends on a weighted function of the contrast in different frequency bins at that time, σ_{tf} , via the weights B_f . The gain changes can thus be written as changes to d , via:

$$d(\sigma_{tf}) = d_0 + d_1 \cdot \sum_f \kappa_f^{(d)} \cdot \sigma_{tf}^n \quad (5.18)$$

where $\kappa_f^{(d)} = B_f$. For consistency and simplicity, we can generalise from this, and assume that all four parameters change with contrast in a similar manner.

Equations (5.10) – (5.14) differ from Equation (5.18) insofar as they do not include an exponent n for σ_{tf} . Thus they consider the weighting of contrast to be linear in σ_{tf} , rather than linear in σ_{tf}^n . As shall be explained below in Section 5.3.3, in the stimuli used here, only two values of σ_{tf} were presented. As a result, the exponent is superfluous for this data set.

Nevertheless, it is worth noting that both the value of the exponent, n , and the exponentiation itself, were chosen empirically in Section 4.3.5 to fit the gain changes produced by spectrally coarse changes in contrast. It would therefore be an empirical matter to determine to what extent these models would be improved if gain kernels for each parameter θ incorporated input nonlinear transforms (F_{in}) and/or output nonlinear transforms (F_{out}) on the contrast, such as:

$$\theta(\sigma_{tf}) = \theta_0 + \theta_1 \cdot F_{out}^{(\theta)} \left[\sum_f \kappa_f^{(d)} \cdot F_{in}^{(\theta)}[\sigma_{tf}] \right] \quad (5.19)$$

This lies beyond the scope of the work presented here.

5.3.2.4 Simplifications of contrast kernel models

The full model of Equations (5.10) – (5.14) has a sizeable number of parameters. Denoting N_F as the number of frequency bands, this model has $4 \times (N_F + 1)$ parameters (there being redundancy between $\kappa_f^{(\theta)}$ and θ_1), in addition to the

$N_F + N_H$ parameters of the separable STRF. We can make a number of assumptions to simplify this. First, not every one of the nonlinearity parameters $\{a, b, c, d\}$ needs to be rendered contrast-dependent. For such parameters θ , we set $\theta_1 = 0$, such that $\theta_t = \theta_0$. Next, it is possible that changes to some of these parameters are the result of the same physiological process. This would allow us to assume a shared contrast kernel between parameters θ and θ' , i.e. $\kappa_f^{(\theta)} = \kappa_f^{(\theta')}$, with a corresponding $\zeta_t^{(\theta)} = \zeta_t^{(\theta')}$.

For brevity, I shall assign the following notation to individual models. The full model, wherein all nonlinearity parameters have separate contrast kernels, I denote the $a/b/c/d$ -model. When a nonlinearity parameter is assumed contrast-independent, I shall omit the corresponding letter from the name. Thus b does not change with contrast in the $a/c/d$ -model. Finally, I shall concatenate letters when they share the same contrast kernel. Thus, in the a/cd -model, $\kappa_f^{(a)} \neq \kappa_f^{(c)} = \kappa_f^{(d)}$.

One special case is worth particular mention. The results of Chapter 4 demonstrate that the primary effects of changing contrast lie in changes to the gain, via the parameter d , with some correlated changes in threshold, via the parameter c . These effects can be most simply captured by the cd -model, wherein the nonlinearity parameters a and b are contrast-independent, while c and d share a single contrast kernel. For the cd -model, I shall denote the kernel as κ_f , the *gain contrast kernel*. The cd -model thus takes the form:

$$\begin{aligned}
 \hat{y}_t &= a + \frac{b}{1 + e^{-z_t}} \\
 z'_t &= (z_t - c_t)/d_t \\
 c_t &= c_0 + c_1 \cdot \sum_f [\kappa_f \cdot \sigma_{tf}] \\
 d_t &= d_0 + d_1 \cdot \sum_f [\kappa_f \cdot \sigma_{tf}]
 \end{aligned} \tag{5.20}$$

5.3.2.5 Extensions of contrast kernel models

Contrast kernels estimate how the contrast in different frequency bands affects the parameters of the output nonlinearities. If we focus on the contrast kernel for d , we see that $\kappa_f^{(d)}$ specifically measures how a neuron's *overall* gain is determined by the spectral profile of contrast. Any gain changes are applied *equally*, such that the neurons' sensitivities to all sound frequencies are uniformly scaled. In other words, if there is high contrast in a single frequency band, g , and the coefficient κ_g is non-zero, the gain of that neuron's response to level changes in band f is changed, independent of f .

This assumption can be relaxed by extending the contrast kernel for d to allow for non-uniform gain effects. This allows us to consider, for example, what would happen if the gain effects from a high-contrast frequency band were localised to a particular region of frequency space.

We start with the cd -model of Equation (5.20), and apply two major changes. First, rather than the gain changes being applied uniformly as d_t :

$$z'_t = \frac{z_t - c_t}{d_t} = \left[\frac{(\sum_{fh} X_{tfh} \cdot k_{fh}) - c_t}{d_t} \right] \quad (5.21)$$

instead, they can be applied on a frequency-by-frequency basis, as d_{tf} :

$$z'_t = \sum_f \frac{z_{tf} - c_t}{d_{tf}} = \sum_f \left[\frac{(\sum_h X_{tfh} \cdot k_{fh}) - c_t}{d_{tf}} \right] \quad (5.22)$$

with $\hat{y}_t = a + b/(1 + e^{-z'_t})$.

Second, rather than the gain for all frequency bands, d_t , being calculated by a weighted sum of the contrast profile of a sound, via the contrast frequency kernel, κ_g :

$$d_t = d_0 + d_1 \cdot \sum_g \kappa_g \cdot \sigma_{gt}$$

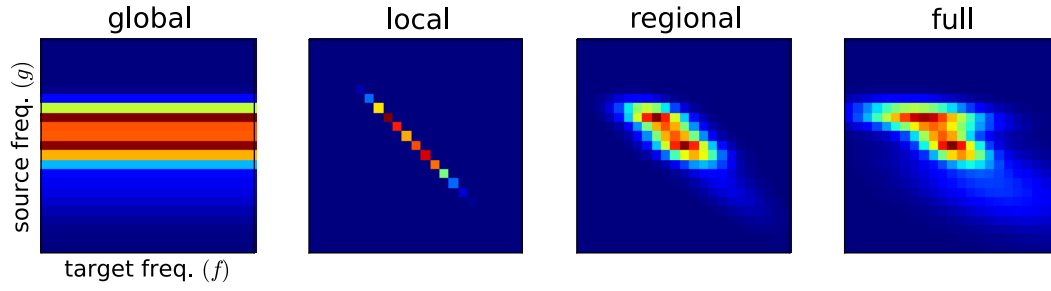


Figure 5.6: Illustration of cross-contrast kernels, C_{fg} . Colour plots show images of C_{fg} for the global effect model, the local effect model, the regional effect model, and a full-rank C_{fg} for an idealised unit. The colour ranges from blue (coefficients near zero; i.e. the contrast of band g does not affect the gain of band f) to red (large coefficients; i.e. the contrast of band g strongly modulates the gain of band f). 1st panel: In the global effect model, high contrast in a source band affects all target bands equally. 2nd panel: In the local effect model, high contrast in a source band affects only that band. 3rd panel: In the regional effect model, high contrast in a source band has gain effects in a local region. Here, a small effective scale, σ_κ , tends to diagonalise C_{fg} towards the local model, while large σ_κ stretches C_{fg} horizontally, towards the global model. 4th panel: A full rank C_{fg} can fit the gain effects induced by the contrast in any one source band on any target band, but is impractical to fit.

instead, the gain for *each* frequency band, f , can be calculated by such a sum, via the *contrast cross-frequency kernel*, C_{fg} :

$$d_{tf} = d_0 + d_1 \cdot \sum_g C_{fg} \cdot \sigma_{gt} \quad (5.23)$$

The coefficients of C_{fg} measure the divisive effect of the contrast in a source band g on the coding of level fluctuations in a target band f .

Calculating a full C_{fg} matrix requires N_F^2 terms, and is therefore impractical to fit. We shall therefore consider three simplifications of the cross-frequency kernel. These are illustrated in Figure 5.6.

First, in the *global effect model*, the contrast in band g affects the gain of all bands equally. Thus $C_{fg} = \kappa_g$. This is identical to the standard contrast kernel model presented above.

Second, in the *local effect model*, the contrast in band g affects only the gain

of that band. Thus $C_{fg} = \delta_{fg} \cdot \kappa_g$.

Third, in the *regional effect model*, the contrast in band g affects the gain of nearby bands, but not distant bands. This is parameterised as follows:

$$C_{fg} \propto \kappa_g \cdot \exp\left(-\frac{1}{2\sigma_\kappa^2}(f - g)^2\right) \quad (5.24)$$

with a constant of proportionality standardised in Appendix B to prevent a redundancy with other parameters. This form depends on two parameters. The vector κ_g is a standard contrast frequency kernel, measuring the total gain effect due to changes in the contrast of band g . The scalar σ_κ is the *effective scale* for the spread of gain effects. Small σ_κ indicates that high contrast in band g only affects the sensitivity of the neuron to level changes in frequencies close to g , while large σ_κ indicates that this would affect the sensitivity to a broad range of frequencies.

The local and global effect models embody two extrema of σ_κ . If we take the limit $\sigma_\kappa \rightarrow 0$, then $C_{fg} \rightarrow \kappa_g \cdot \delta_{fg}$, and we recover the local effect model, wherein σ_{gt} affects only band g . If we take the limit $\sigma_\kappa \rightarrow \infty$, then $C_{fg} \rightarrow \kappa_g$, and we recover the global effect model, wherein σ_{gt} affects all bands equally.

5.3.3 Random Contrast DRCs

Using the models presented above, and a set of experiments which are described in Section 5.4, I aimed to estimate how the contrasts in different frequency bands affect the gain of a neuron (as well as how they affect other response parameters). This task bears a structural similarity to the problem of estimating how the sound levels in different frequency bands linearly affect the firing rate of a neuron, i.e. the problem of estimating a neuron's STRF. As explained in Section 2.1.2, DRCs provide an efficient solution to this latter problem. When the levels of the tones in different regions of the spectrogram are independent,

STRF estimation reduces to simple reverse correlation. In a similar manner, one can efficiently estimate how gain depends on the contrast in band-limited regions when these contrasts are independent of one another.

I therefore introduce a set of stimuli I shall call *Random Contrast DRCs* (RC-DRCs). These provide an ensemble of contrast conditions in an analogous way to how an ordinary DRC provides an ensemble of level conditions. Standard DRCs have tone levels that are spectrotemporally uncorrelated; RC-DRCs additionally have band contrasts that are spectrally uncorrelated.

As with ordinary DRCs, RC-DRCs are constructed as a sequence of chords. Their major distinguishing feature is the organisation of these chords into segments of several seconds' duration. In each segment, the tone level distributions for each of the N_F bands have different parameters.

In the RC-DRCs I construct here, I make a particular simplification. In each segment, a random subset of the tones have their levels drawn from a high contrast level distribution, while the remaining tones have their levels drawn from a low contrast level distribution. Thus generating each new segment requires two sets of random draws: first, a random partitioning of tones into the high and low contrast conditions for that segment; then, conditioned on that partitioning, a sampling of random levels for each time/frequency bin from the respective tone distributions.

The notation I shall use for the RC-DRC stimuli is as follows. As in previous chapters, the sound level of frequency f at time t is denoted X_{tf} . The contrast in frequency band f at time t is denoted σ_{tf} . Given that only two levels of contrast are used, I shall employ a binary representation, such that $\sigma_{tf} \in \{0, 1\}$. The relevance and validity of this linear transformation is discussed below in Section 5.4.1.4. Due to the segmented nature of the RC-DRCs, this 2-tensor is highly redundant: $\sigma_{tf} = \sigma_{(t-1)f}$, except at segment boundaries. A compressed

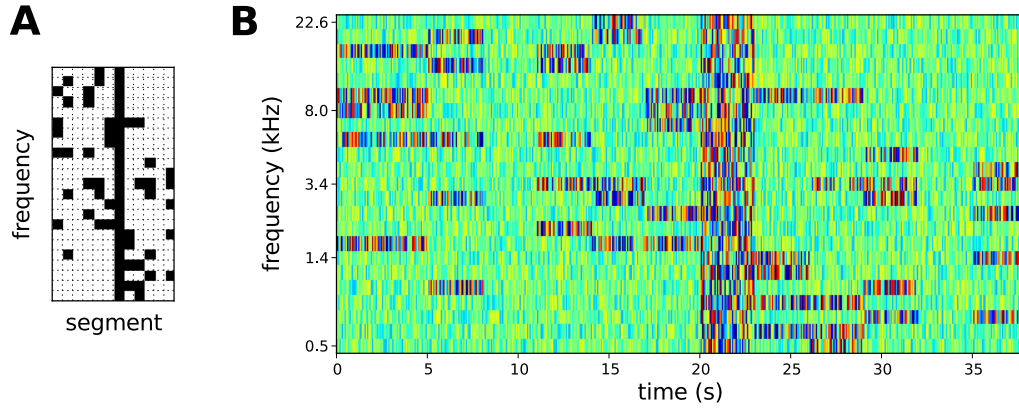


Figure 5.7: Schematic of Random Contrast DRC (RC-DRC) stimuli. This 38 s stimulus is comprised of 12 segments, in which the contrast within different frequency bins is either high or low. (A) Image of the matrix s_{kf} for a RC-DRC. White blocks denote $s_{kf} = 0$, i.e. low contrast in frequency band f during segment k ; black blocks denote $s_{kf} = 1$, i.e. high contrast in that band/segment. (B) Image of the matrix X_{tf} , i.e. the tone levels of the RC-DRC shown in (A). Colours denote tone levels, as in Figure 2.3.

version of this matrix is given as s_{kf} , denoting the (binary) contrast in frequency band f , during segment k .

An example RC-DRC stimulus used in the experiments which follow is illustrated in Figure 5.7.

5.4 Spectral Experiment 2: Contrast frequency kernels

5.4.1 Methods

5.4.1.1 Animals

Data in this experiment were obtained from three female ferrets, which were not used in previous experiments. A full description of the anaesthesia, surgical, stimulation and recording procedures is provided in Section 2.3.

5.4.1.2 Stimuli

I constructed Random Contrast DRCs (RC-DRCs), as defined in Section 5.3.3. To reduce the total number of stimulus conditions, I used a slightly sparser set of tone frequencies than in previous experiments: $N_F = 23$ pure tones, with frequencies log-spaced between $f_{low} = 500$ Hz and $f_{high} = 22.6$ kHz at 1/4 octave intervals (c.f. the 1/6 octave spacing in Chapters 3 and 4, and in Experiment 1 above in Section 5.2). The levels of the tones were changed every 25 ms, with 5 ms linear ramps between chords.

In each segment, a random subset of N_{high} of the N_F tones had their levels drawn from a high contrast ($w_L = 15$ dB; $\sigma_L = 8.7$ dB; $c = 92\%$) level distribution, while the remaining $N_{low} = N_F - N_{high}$ tones had their levels drawn from a low contrast ($w_L = 5$ dB; $\sigma_L = 2.9$ dB; $c = 33\%$) level distribution. All tone distributions had mean level $\mu_L = 40$ dB SPL. By virtue of the 3 s duration, each segment consisted of a sequence of 120 chords, sufficient for a rough approximation of the output nonlinearity during that contrast condition (see Section 5.4.1.4).

In order to explore as large a region of contrast space as possible, between $N_S = 80$ and $N_S = 120$ segments were presented at each electrode penetration. Two types of segment were necessary to establish baselines for gain measurements: one where all tone distributions were low contrast, one where all were high. Given the importance of these two baseline conditions, 9 of the N_S segments were reserved for each. The remaining segments all had a randomised partition of tones into $N_{high} = 5$ high contrast bands and $N_{low} = 18$ low contrast bands, as described above.

While the choice of $N_{high} = 5$ was somewhat arbitrary, there were several motivating factors for this decision. If all possible assignments of high and

low contrast were admitted, then the stimulus space would be extremely large ($2^{23} \approx 8 \times 10^6$ conditions). Decimating the stimulus space in a structured way (to $\binom{23}{5} \approx 3 \times 10^5$ conditions) can reduce the severity of small sampling biases. Next, choosing approximately 20% of the tones (5/23) to be high contrast within any segment is consistent with the ratio of test/mask sizes used above in Section 5.2 (7/34 tones in the test band). Finally, from the shape of the contrast-dependent gain curve (Equation (4.9), Figure 4.11), the gradient $\frac{\partial G}{\partial \sigma_L}$ is greater at low σ_L ; small changes in gain are likely to be easier to measure in this domain.

The segments were packaged into individual RC-DRC sequences, each consisting of 12 segments. The first segment of each sequence was 5 s in duration, so that the first 2 s of each stimulus presentation could be discarded (as in Chapters 3 and 4). From each subsequent segment, the first 0.5 s of data was also discarded to ensure that a constant adaptation state had been reached. An analysis of responses during this period is presented in the next chapter (Section 6.4). The 38 s sequences were presented 10 times each, randomly interleaved, with a 1-2 s period of silence between each sequence.

5.4.1.3 Unit selection criteria

Only units which modulated their firing rate in response to the RC-DRCs were included for analysis. This was measured via the noise ratio (NR; see Section 2.3.5.2). As in previous experiments, the PSTH was binned at 25 ms, with bins offset by between 0 and 25 ms to allow for response latency. The offset was chosen on a unit-by-unit basis to minimise the NR. However, unlike previous experiments (Chapters 3 and 4, and Section 5.2 above), the maximum admitted NR was 40 (averaged across the whole ensemble); units with $NR > 40$, i.e. whose explainable variance was less than $\sim 2.5\%$ of the total variance, were excluded from analysis. In this study, models are evaluated while taking NR

into consideration (see below, Section 5.4.1.10), allowing for a less stringent criterion for unit inclusion.

5.4.1.4 Model structure: contrast kernels

The modelling in this experiment builds extensively on the work of Chapter 4, and Section 5.2 within this chapter. A major departure is that rather than estimating separate output nonlinearities in each contrast condition, from which gain changes would then be estimated via curve transformations, single models that incorporate both these stages are fit directly to the data.

As described above in Section 5.3.2, the models explored here take the general form given by Equations (5.10) – (5.14). These models constitute an extension of the standard sigmoid nonlinearity function (Equation (4.1)) wherein some or all of the parameters $\theta \in \{a, b, c, d\}$ are contrast-dependent. The contrast dependence of parameter θ is mediated via a contrast kernel $\kappa_f^{(\theta)}$, which weights the time/frequency-varying contrast σ_{tf} , via:

$$\theta_t = \theta_0 + \theta_1 \cdot \left[\sum_f \kappa_f^{(\theta)} \cdot \sigma_{tf} \right] \quad (5.25)$$

This contains a redundancy between θ_1 and $\kappa_f^{(\theta)}$. I therefore defined:

$$\kappa_f^{(\theta)} = \frac{\lambda_f^{(\theta)}}{\sum_{f'} \lambda_{f'}^{(\theta)}} \quad (5.26)$$

Here, I shall refer to $\lambda_f^{(\theta)}$ as the *unnormalised contrast kernel* for θ . This ensured that:

$$\sum_f \kappa_f^{(\theta)} = 1 \quad (5.27)$$

As explained above in Section 5.3.2.3, while previous models advanced in this thesis have included an exponent n to σ_{tf} , such an exponent is superfluous

for the stimulus set used here. This is because the contrast in any frequency band in any segment was chosen from only one of two values. Including an exponent in Equation (5.25) for the two-valued $\{\sigma_{tf}\}$ set is equivalent to a linear transformation on this set, which can in turn be mopped up by changes to θ_0 and θ_1 . This justifies the binary $\{0, 1\}$ representation for σ_{tf} used here. It may still be the case that such an exponent may be relevant for more flexible stimulus sets; nevertheless, a greater priority for this experiment was to determine the nature of the contrast kernels, $\kappa_f^{(\theta)}$.

In order to speed up the fitting of these models, I took advantage of the segmented nature of the RC-DRCs. Firstly, as explained in Section 5.4.1.2 above, σ_{tf} does not change with time within a segment k , and thus can be summarised via s_{kf} . Secondly, rather than fitting the parameters directly to the entire, trial-averaged training dataset (with $N_t \approx 8000$), a similar approach was used as in Chapter 4 whereby the set of (STRF-weighted-)stimulus/response pairs (z_t, y_t) within each segment was divided equally into 20 bins along the z axis. For bin j in segment k , I shall denote the bin centre as $\bar{z}_{jk}$, and the average firing rate $\bar{y}_{jk}$. In total, the model could thus be summarised as:

$$\hat{y}_{jk} = a_k + \frac{b_k}{1 + e^{-(\bar{z}_{jk} - c_k)/d_k}} \quad (5.28)$$

$$\theta_k = \theta_0 + \theta_1 \cdot \left[\sum_f \kappa_f^{(\theta)} \cdot s_{kf} \right] \quad (5.29)$$

This greatly reduced the time taken to fit each model.

5.4.1.5 Model structure: cross-frequency interactions

While the contrast kernel models estimate which set of frequencies' contrasts determine a neuron's gain, they assume that any gain changes have an equal effect on that neuron's sensitivity to all sound frequencies. Testing this assumption requires estimating how contrast in a source frequency band g affects a neuron's gain for sound level changes in a target frequency band f .

As explained in Section 5.4.2.7, this was done by extending the contrast frequency kernel, κ_g , to a cross-frequency kernel, C_{fg} . Rather than solving for a full rank estimate of C_{fg} , I considered the three simplifications developed in Section 5.3.2.5: the global effect model, the local effect model, and the regional effect model.

The contrast cross-frequency kernel models were considerably more computationally intensive to fit than the standard contrast kernel models. This was for two major reasons. Firstly, the addition of an extra dimension increased the volume of operations. The standard contrast frequency kernel models shift and scale the output of the linear STRF model, $z_t = \sum_{f,h} X_{tfh} \cdot k_{fh}$. Their output can be seen as a time-varying function of this, $F_t^{(\kappa)}[z_t]$ (see Equation (5.21)). The cross-frequency kernel models, however, do not collapse the frequency dimension of the STRF model, and depend instead on $z_{ft} = \sum_h X_{tfh} \cdot k_{fh}$. Their output is then $F_t^{(C)}[z_{ft}]$ (see Equation (5.22)).

Secondly, as explained above in Section 5.4.1.4, fitting the parameters of $F_t^{(\kappa)}$ can be expedited by binning (z_t, y_t) pairs along the z_t axis within each segment. The additional dimension of the cross-frequency kernel models prevents such binning, so that the parameters of $F_t^{(C)}[z_{ft}]$ needed to be fitted to the entire dataset.

Fitting the cross-frequency kernels models was very slow. Thus only three versions were considered, as given in Section 5.4.2.7. The parameters a and b were considered contrast-independent, and the contrast kernel for c was considered the (target) frequency-weighted average of the contrast cross-frequency kernel used for d :

$$\kappa_g^{(c)} = \frac{1}{N_F} \sum_f C_{fg}^{(d)} \quad (5.30)$$

The coefficients of C_{fg} were constrained to be positive, given that convergence was faster, and required fewer repeats than for unconstrained coefficients. Like

the case for the standard contrast frequency kernels, it did not appear that prediction scores were affected by this constraint (c.f. Figures 5.16 and 5.17)

Although the global effect model reduces to the standard contrast kernel model presented in Section 5.4.2.3, the latter was fitted using the segmented/binning approximation. To facilitate a fair comparison with the local and regional effect models, the global effect model was recomputed over the unbinned dataset. Incidentally, similar prediction scores were obtained for the unbinned dataset as for the segmented/binning dataset; this provides support for the validity of the binning approximation.

Finally, the Markov Chain Monte Carlo stage of estimating parameter uncertainties for these models (see below, Section 5.4.1.7) was omitted, as this was prohibitively slow.

5.4.1.6 Model fitting

Until this point (i.e. Chapter 4 and the first experiment in this chapter in Section 5.2), I estimated gain changes using a multi-step modelling approach. This process ran as follows: first, a single STRF was fitted; next, the parameters of the output nonlinearities were fitted for each condition, with the STRF fixed; finally, the parameters of the curve transformations were fitted, with the nonlinearities fixed.

This approach is problematic. In such sequential fitting, any noise in the parameter estimates of one stage become a target of the optimisation in the following stage. The failures of such sequential fitting were especially salient in this experiment. Parameter estimates for the nonlinearities for many segments often carried large amounts of uncertainty. This is illustrated in Figure 5.8.

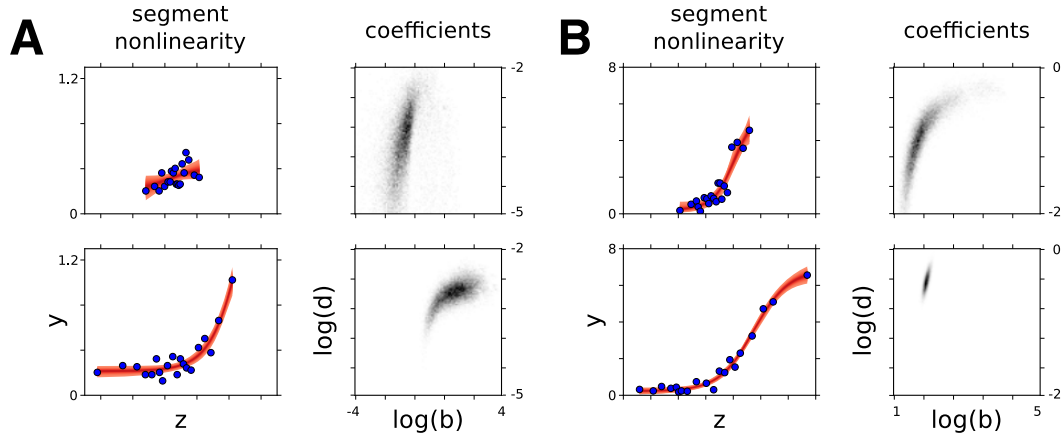


Figure 5.8: Uncertainty in estimates of static nonlinearity parameters. Fitting separate nonlinearities to each segment, then estimating gain changes, is a poorly constructed problem. For many segments, these parameter estimates were very uncertain. (A) Data from two segments (top and bottom) for an example unit. Left panels: output nonlinearities in each of the segments. Blue circles show binned STRF-weighted-stimulus/response pairs (z, y) . Red shaded region shows the 95% credible interval for the predictive distribution over the sigmoid output nonlinearity, estimated using MCMC. Right: MCMC estimates of the marginal posterior distributions for two of the parameters in the respective nonlinearities. (B) As for (A), for two segments' responses from another unit. The precision of the parameter estimates is better for some segments than others.

When nonlinearities were fitted first and fixed, contrast kernels were susceptible to overfitting to the fitted parameters of the segment nonlinearities; for many units, the models generalised poorly from training data to prediction data.

Since the interest is in the final gain parameters, a better approach (if it is tractable to do so) is to construct and optimise a single objective function over all parameters. As a result, all model fitting here was done in two stages only. As previously, separable STRFs were fitted to the pooled dataset (with observed X_{tf} and y_t), and fixed thereafter. Next, Maximum A Posteriori (MAP) estimates of the nonlinearity and contrast kernel parameters were estimated all together, using gradient descent (see below). This assumed that z_t , i.e. the output of the linear STRF model, was observed. As an additional, optional stage, the uncertainty on the model parameters was estimated by approximate

sampling from their joint posterior distribution using Markov Chain Monte Carlo (MCMC; see below).

In principle, one can merge the STRF fitting with the nonlinearity/contrast kernel fitting, and minimise a single objective function. However, optimising all parameters proved too computationally demanding to be rendered practical. Alternatively, one can iterate between optimising the nonlinearity/contrast kernel parameters (with the STRF fixed), and optimising the STRF (with the other parameters fixed). However, I observed that this iterative procedure typically made little to no difference to prediction scores, and STRFs did not noticeably change over successive iterations. This has also been reported for GLMs by [Calabrese et al. \[2011\]](#). Since the focus here is not on the STRF, but on the contrast-dependent changes to output nonlinearities, no successive refinements to STRFs were pursued beyond the initial fit.

5.4.1.7 Bayesian methods

As with models in previous sections, the data set for each unit was subdivided randomly into training and prediction subsets (see below, Section 5.4.1.10). All parameter fitting took place on the training dataset.

Separable STRF models were fitted using maximum likelihood, as in previous sections. All subsequent models were fitted using Bayesian methods. The Bayesian paradigm provided a uniform framework for fitting parameters, applying hard and soft constraints, and estimating parameter uncertainties.

Maximum A Posteriori (MAP) parameter estimates were fitted using gradient descent. For each model, the log posterior probability was calculated, as well as its derivatives with respect to all the parameters. The respective derivatives for each model are documented in Appendix A. Minimisation of the negative

log posterior was performed using the L-BFGS-B algorithm [Zhu et al., 1997], via SciPy [Jones et al., 2001].

Priors were chosen for simplicity of form, rather than analytic tractability. These are listed below in Sections 5.4.1.8 and 5.4.1.9. The log likelihoods (and the log posteriors) were only convex with respect to some parameters; the gradient descent algorithm thus needed to be initialised at a number of different initial conditions to reduce the chances of settling in local minima. Forty different starting locations were chosen by random draws from the prior distributions over the parameters, with an additional initialisation condition at the mean of the priors. Generally, most of these repeats converged to the same fixed point.

A major goal of this work is to estimate the contrast kernels for cortical neurons. This involves estimating not only the best parameter estimates for $\kappa_f^{(\theta)}$, but also their error bounds. In addition to the MAP estimates, which pinpoint the mode of the posterior parameter distributions, I characterised the shape of these posterior distributions by sampling from them using Markov Chain Monte Carlo (MCMC). MCMC models were constructed in Python, using the PyMC package [Patil et al., 2010]. Chains were initialised at the MAP parameter values, and advanced using a Metropolis-Hastings step method. A barrage of diagnostics, including traceplots, Geweke’s diagnostic [Geweke, 1992], and autocorrelation analyses, was used to assess convergence and mixing. From these diagnostics, I found that minimum chain lengths of 120000 samples, with a 20000 sample burn in and 20x thinning were sufficient for a reasonable characterisation of the posteriors. As always for MCMC, longer and parallel chains would improve the representation of the posteriors; nevertheless, the observed results satisfied the above diagnostics, and are therefore used here to provide an approximate measure of the error bounds on the contrast kernels.

5.4.1.8 Priors on nonlinearity parameters

Each parameter $\theta \in \{a, b, c, d\}$ is the sum of a contrast-independent term, θ_0 , and a contrast-dependent term, weighted by θ_1 . Rather than placing priors directly on these terms, it was more convenient to reparameterise the model as follows.

The contrast in segment k and frequency f , s_{kf} , could only take on binary values. In the all-low-contrast segment, when $s_{kf} = 0 \ \forall f$, it follows from Equation (5.29) that $\theta_k = \theta_0$. In the all-high-contrast segment, when $s_{kf} = 1 \ \forall f$, we use the fact that the contrast kernel, $\kappa_f^{(\theta)}$, is normalised (from Equation (5.27)) to find that $\theta_k = \theta_0 + \theta_1$. Defining these two values as θ_{low} and θ_{high} respectively, we can rewrite Equation (5.29) as:

$$\theta_k = \theta_{low} + (\theta_{high} - \theta_{low}) \cdot \left[\sum_f \kappa_f^{(\theta)} \cdot s_{kf} \right] \quad (5.31)$$

From this parameterisation, the parameters $\{a_{low}, b_{low}, c_{low}, d_{low}\}$ describe the output nonlinearity in the all-low-contrast segment, and the parameters $\{a_{high}, b_{high}, c_{high}, d_{high}\}$ describe the output nonlinearity in the all-high-contrast segment. In each segment, the value of θ_k typically lies between θ_{low} and θ_{high} , depending on the projection of s_{kf} onto $\kappa_f^{(\theta)}$ (although θ_k can take on more extreme values when some of the coefficients of $\kappa_f^{(\theta)}$ are negative).

For each θ , identical priors were placed on each of θ_{low} and θ_{high} . The respective priors were primarily chosen to satisfy three purposes: to enforce a set of hard constraints, specifically that a , b and d all be positive; to apply some regularisation, i.e. to ensure that b , c and/or d did not grow excessively large; and to provide a suitable set of initial conditions for MAP fitting. As a result, the priors were relatively broad, with data-driven hyperparameters.

Firstly, I shall define some intermediate statistics on the set of binned stimulus-response data, $(\bar{z}_{jk}, \bar{y}_{jk})$, for a given unit:

$$\begin{aligned} y_{min} &= \min_{j,k} \bar{y}_{jk} & z_{min} &= \min_{j,k} \bar{z}_{jk} \\ y_{max} &= \max_{j,k} \bar{y}_{jk} & z_{max} &= \max_{j,k} \bar{z}_{jk} \\ y_{range} &= y_{max} - y_{min} & z_{range} &= z_{max} - z_{min} \\ & & z_{mean} &= \frac{1}{N_j N_k} \sum_{j,k} \bar{z}_{jk} \end{aligned}$$

where N_j is the number of bins (here, 20), and $N_k = N_s$, the number of segments.

In turn, the priors $P(\theta_{low}) = P(\theta_{high}) = P(\theta)$ were defined via:

$$a \sim \text{Exp}(\beta = (y_{min} + 0.05 \times y_{range})) \quad (5.32)$$

$$b \sim \text{Exp}(\beta = (y_{range} \times 2)) \quad (5.33)$$

$$c \sim \mathcal{N}(z_{mean}, z_{range}^2) \quad (5.34)$$

$$d \sim \text{Exp}(\beta = z_{range}/20) \quad (5.35)$$

where β is the scale parameter of the exponential distribution.

The L-BFGS-B algorithm used to minimise the negative log posterior allows the explicit specification of parameter boundary values; for the exponentially distributed variables, a lower bound of 1×10^{-15} was provided.

Finally, for those models where only a subset of the nonlinearity parameters were contrast-dependent, $\theta_1 = 0$ was enforced, i.e. $\theta_{high} = \theta_{low}$.

5.4.1.9 Priors on contrast kernels

As discussed in the results (see Section 5.4.2.5), three different approaches to the values of contrast kernels, $\kappa_f^{(\theta)}$ were taken. In the first approach, the

kernels were allowed to take any real value; in the second approach, they were constrained to be positive; in the third approach, they were fixed at particular values.

Priors were placed on the coefficients of the unnormalised contrast kernels, $\lambda_f^{(\theta)}$. As these were normalised via Equation (5.26) to give $\kappa_f^{(\theta)}$, the scale of the respective priors was not important.

When real-valued kernels were used, the prior on $\lambda^{(\theta)}$ was chosen to be a spherical Gaussian, with $\lambda_f^{(\theta)} \sim \mathcal{N}(0, 0.1^2)$. When positive-valued kernels were used, the prior on each coefficient was chosen as $\lambda_f^{(\theta)} \sim \text{Exp}(\beta = 0.1)$. In this latter situation, as for the positive nonlinearity parameters, the L-BFGS-B algorithm bounded each coefficient below at a value of 1×10^{-15} .

When real-valued contrast kernels were used, it was possible for the denominator of Equation (5.26) to approach zero, giving untenable values of $\kappa_f^{(\theta)}$. As a result, the minimisation algorithm occasionally yielded errors. When this occurred, the algorithm was reset with a new initial value. This discontinuity also meant that the algorithm was more likely to get stuck in local minima of the negative log posterior, requiring a larger number of repeated fittings from random initial conditions.

5.4.1.10 Model success

A secondary goal in this section is to compare different models for the firing rate behaviour of auditory cortical neurons. To do so, I make use of the strategy developed by [Sahani and Linden \[2003b\]](#).

For each unit, the amount of its total response variance that can be explained is bounded by the signal power (SP). Model success should therefore be measured

as the percentage of signal power explained ($\%SPE$). This is the percentage reduction in the SP from fitting the model, and equivalent to:

$$\%SPE = \frac{Var(y_t) - Var(y_t - \hat{y}_t)}{SP} \times 100\% \quad (5.36)$$

By subdividing the data for each unit into a training and prediction dataset, one can obtain two values for $\%SPE$ for that unit. The $\%SPE$ from the training data is inflated as a result of overfitting to the noise in the training data. As the noise ratio (NR) of the unit grows, the model will continue to fit the non-stimulus-locked portion of the response, so the $\%SPE$ for the training data should grow accordingly. Thus $\%SPE_{training}$ provides an upper bound for the model performance. The $\%SPE$ from the prediction data is expected to be lower, since it tests the generalisability of the model to new data. As overfitting to the training data worsens with increasing NR, so $\%SPE_{prediction}$ declines. Thus $\%SPE_{prediction}$ provides a lower bound for the model performance.

Since these two measures diverge as a function of noise ratio, a suitable method for measuring the model's predictive power from the population data is to extrapolate from the two sets of estimates of model success above to those for a hypothetical zero-noise neuron [Sahani and Linden, 2003b; Ahrens et al., 2008]. The resultant upper and lower estimates bound the *true predictive power* of the model, i.e. that which would be obtained in the limit of zero noise.

This procedure is illustrated in Figure 5.9 for fitting separable and inseparable STRF models to the cortical units used in this study. It is worth noting that while the trend of increasing $\%SPE_{training}$ with larger NR is observed for inseparable STRFs (rightmost panel of Figure 5.9A), it is not the case for separable STRFs (rightmost panel of Figure 5.9B). This shows that, for sufficiently large datasets, the separable STRF models do not overfit to the training data.

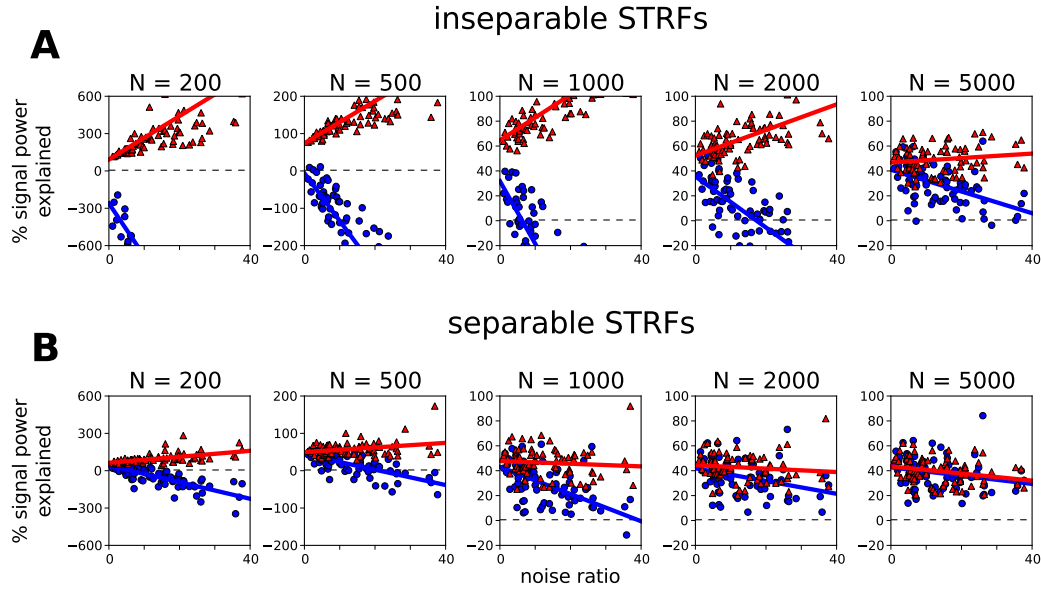


Figure 5.9: Measuring model success. (A): Percentage of signal power explained (%SPE) for an inseparable linear STRF model of the 77 A1/AAF units in this study. Plots are as in [Sahani and Linden, 2003b; Ahrens et al., 2008]. Red triangles show the %SPE for the model fit on the training dataset; blue circles show the %SPE for the model fit on the prediction dataset. The abscissa denotes the noise ratio (NR), i.e. the ratio between noise power and signal power for each unit. Red and blue lines are linear fits to each of these relationships, allowing for an extrapolation from the population data to a hypothetical zero-noise unit. Each column shows the results of fitting the data to a training dataset of increasing size, N (chosen as a random subset of the full training dataset of $N \approx 8000$). Note the difference in scale along the ordinate axis for different plots. When the training dataset is small (leftmost column), there is extreme overfitting, evident as very large %SPE_{training} and poor generalisation performance to %SPE_{prediction}; this gets worse as NR increases. This leads to very wide bounds for the extrapolated model performance as $[-390\%, 90\%]$. For large N , the extent of overfitting is considerably less; extrapolated model performance has tighter bounds of $[40.4\%, 46.2\%]$. (B) Same plots, for the separable linear STRF model. The separable model has considerably fewer parameters than the inseparable model (34 vs. 231), so the extent of overfitting is far less. For $N = 200$, the extrapolated model performance is $[35.0\%, 57.7\%]$; for $N = 5000$, this is $[42.1\%, 42.9\%]$. Importantly, the trend of increasing %SPE_{training} with NR is no longer visible at high N , indicating that overfitting is not a problem for the separable STRFs.

Here, as in Chapter 4, I subdivided the data from each unit randomly so that 90% was used for the training subset, and 10% for the prediction subset. To correct against sampling biases, this cross-validation procedure was repeated for ten different partitions of the data; the %*SPE* values reported here are averages across these ten partitions. The same set of partitions were used for fitting all models to the same unit.

5.4.2 Results

I recorded from 168 units in the primary auditory cortex (A1/AAF) of three anaesthetised ferrets, while presenting Random Contrast DRC (RC-DRC) sequences. These sequences were similar to the DRCs presented in earlier experiments, with two major differences. Firstly, a sparser set of $N_F = 23$ tone frequencies were used, covering the same frequency range, but spaced at 1/4-octave intervals. More importantly, these DRCs were designed so that the contrast could vary independently across frequency bands. Thus, within each 3 s segment of the DRCs, a random subset of the tones had levels drawn from a high contrast distribution, while the remainder had levels drawn from a low contrast distribution (see Figure 5.7).

77 units were identified which responded reliably to the DRCs, as measured via a maximum noise level criterion (see Section 5.4.1.3).

5.4.2.1 STRF and LN models

In this experiment, I extended the STRF and LN models previously used to characterise the relationship between stimuli and the response of auditory cortical neurons. As a baseline for evaluating the success of the models advanced here, I began by fitting simple (separable) STRF and LN models to each unit,

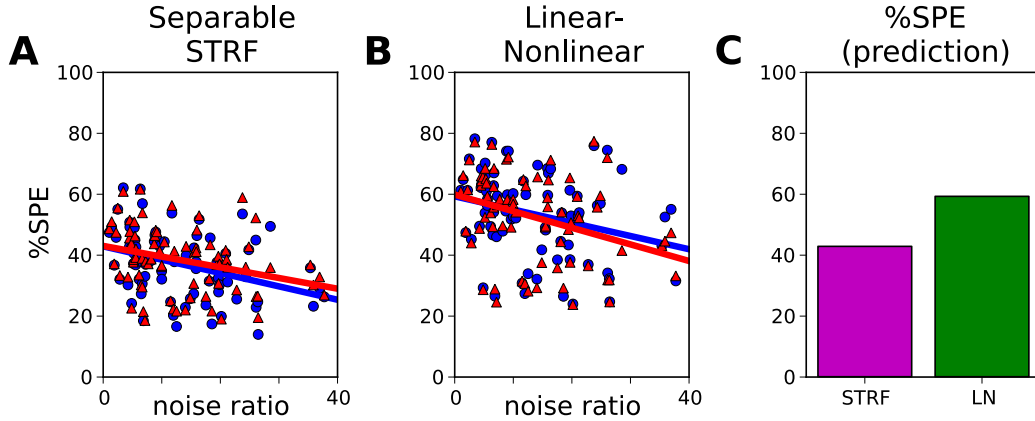


Figure 5.10: Model predictive power: STRF and LN models. For each unit, the models fits to the training data are red triangles, and cross-validation fits to the prediction data are blue circles, measured as the percentage of signal power explained by the model ($\%SPE$). From the population, the solid lines extrapolate to the expected score for a hypothetical zero-noise unit. The linear-nonlinear model performs considerably better than the STRF (see Section 5.4.2.1).

as per Chapter 4. These models were fitted to data that was pooled across all segments of the RC-DRCs, and therefore did not take into account changes in contrast from segment to segment.

For each unit, the performance of each model was measured as the percentage of signal power explained ($\%SPE$) in the training and prediction subsets. To estimate a model's predictive power for the population as a whole, I linearly extrapolated from the pooled set of $\%SPE$ scores to a hypothetical zero-noise unit, providing an upper (from $\%SPE_{training}$) and lower (from $\%SPE_{prediction}$) bound for the model's fit for such a zero-noise unit (see Section 5.4.1.10).

The model fits for the (separable) STRF models, and the LN models are shown in Figure 5.10. The predictive power of the STRF model was 42.8 – 42.9%, while the predictive power of the LN model was 59.2 – 59.8%. This shows that adding an output nonlinearity considerably improves model performance.

5.4.2.2 Contrast produces stereotyped changes to output nonlinearities

In the previous chapter, I found that A1/AAF neurons change their gain in response to changes in contrast. For almost all the neurons recorded from, the same effects were observed. Comparing the output nonlinearity fitted in the all-high contrast segment with that fitted in the all-low contrast segment for each unit, as in Figure 5.11, 72/77 units simply changed their slope (and, in a correlated manner, their inflection point) in response to varying the contrast. This appears to be a stereotyped behaviour amongst cortical units which encode the ongoing stimulus.

5.4.2.3 Gain contrast kernels

The goal of this experiment was to determine how the gain of neurons depends on the contrast within individual frequency bands. This project can be seen as a refinement of the models for gain control advanced to date.

To generalise as directly as possible from the models of the previous chapter (Equation (4.9)), and in the first half of this chapter (Equation (5.3)), I began by fitting the *cd*-model to cortical units, given in Equation (5.20). Thus I fixed the nonlinearity parameters a (minimum firing rate) and b (total dynamic range) to be contrast-independent, by setting $a_1 = b_1 = 0$. Given the previous observation that horizontal shifts of the output nonlinearity were proportional to gain changes (see Figure 4.9), I assumed a single, shared contrast kernel between c and d , the *gain contrast kernel*, κ_f .

Examples of gain contrast kernels fitted to cortical units' responses are shown in Figure 5.12. The most striking aspect of these contrast kernels is their similarity in shape to the STRF frequency kernels. This is particularly noticeable in the

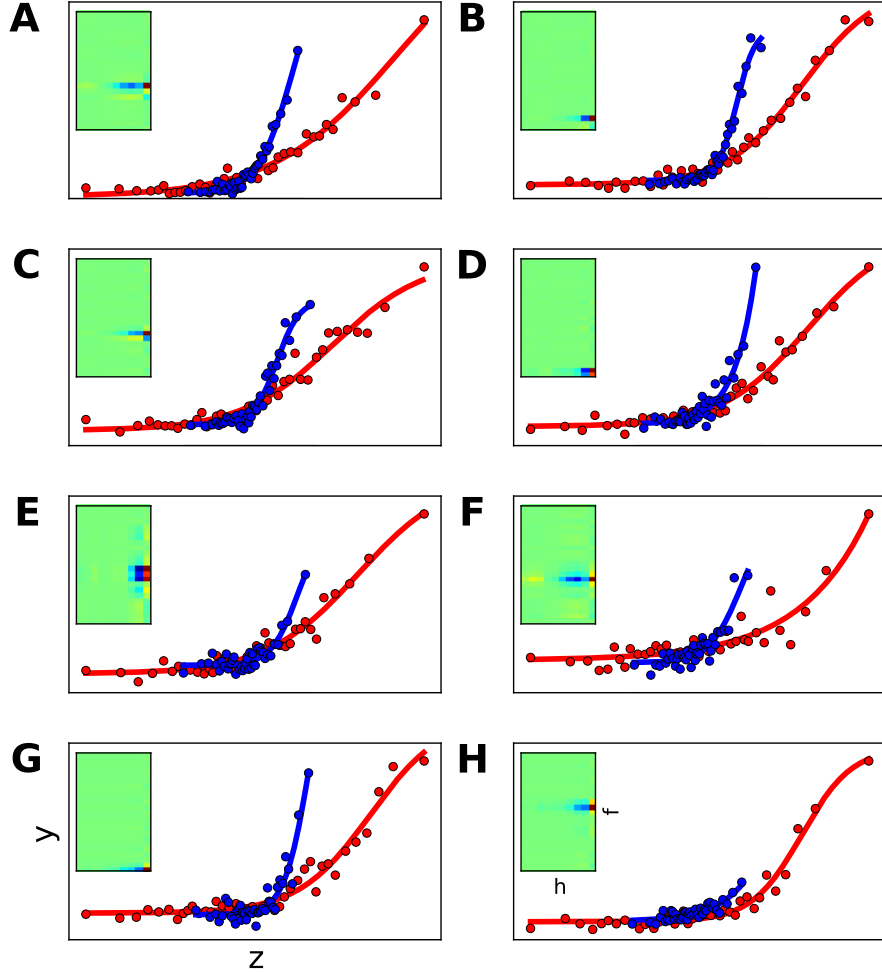


Figure 5.11: Output nonlinearities in low and high contrast: A1/AAF. Data shown for 8 example units in primary auditory cortex. Red circles and line show the output nonlinearity in all-high-contrast segments of RC-DRCs, blue circles and line show the output nonlinearity in all-low contrast segments, as per Figures 4.1 and 4.7. The abscissa denotes the output of the linear STRF (z), the ordinate is the firing rate (y). The inset shows the STRF that was fitted to the entire dataset, as per Figure 4.4, as a function of frequency (f) and time history (h). When the contrast is decreased, the coding properties of cortical units undergo a stereotyped change: the nonlinearities are compressed along the horizontal axis, with a small horizontal shift.

examples from Figure 5.12A-E: for the frequencies in the excitatory component of the receptive field of these units, the weights of κ_f match almost precisely the weights of k_f . As the BF and bandwidth of k_f change across these units, so the BF and bandwidth of κ_f change too.

These kernels reveal an important aspect of contrast gain control: the same frequencies whose level changes additively contribute to the firing rate of a cortical neuron also *divisively* contribute to its gain. In these bands, an increase in tone level *increases* the firing rate of the neuron, while an increase in the contrast of the tone level distribution of these bands *decreases* the gain of the neuron. In turn, the relative size of the gain change produced by varying the contrast in a particular band is roughly proportional to the size of the change in firing rate produced by increasing the level of the band. In the terminology of the previous experiment in this chapter (i.e. Section 5.2), this demonstrates that gain control is predominantly a function of *local contrast*.

I also found that when units had strong inhibitory sidebands in their STRF — i.e. when there were coefficients of k_f , nearby to the BF, which were negative — the gain contrast kernels typically had *positive* coefficients for these same frequencies. Examples of such units are presented in Figure 5.12F-H. In these bands, an increase in tone level *decreases* the firing rate of the neuron; however, an increase in the contrast of the tone level distribution of these bands *also decreases* the gain of the neuron.

For some units, the contrast kernels were slightly wider (e.g. Figure 5.12H), or slightly narrower (e.g. Figure 5.12I) than the excitatory band of the STRF. In addition, not all of the units with inhibitory sidebands produced significantly non-zero κ_f coefficients at the sidebands (e.g. Figure 5.12J). Nevertheless, there was generally a good correlation between units' gain contrast kernels, κ_f , and their linear STRF kernels, k_f ; across units, the median correlation coefficient was $r(\kappa_f, k_f) = 0.69$ (see Figure 5.13A).

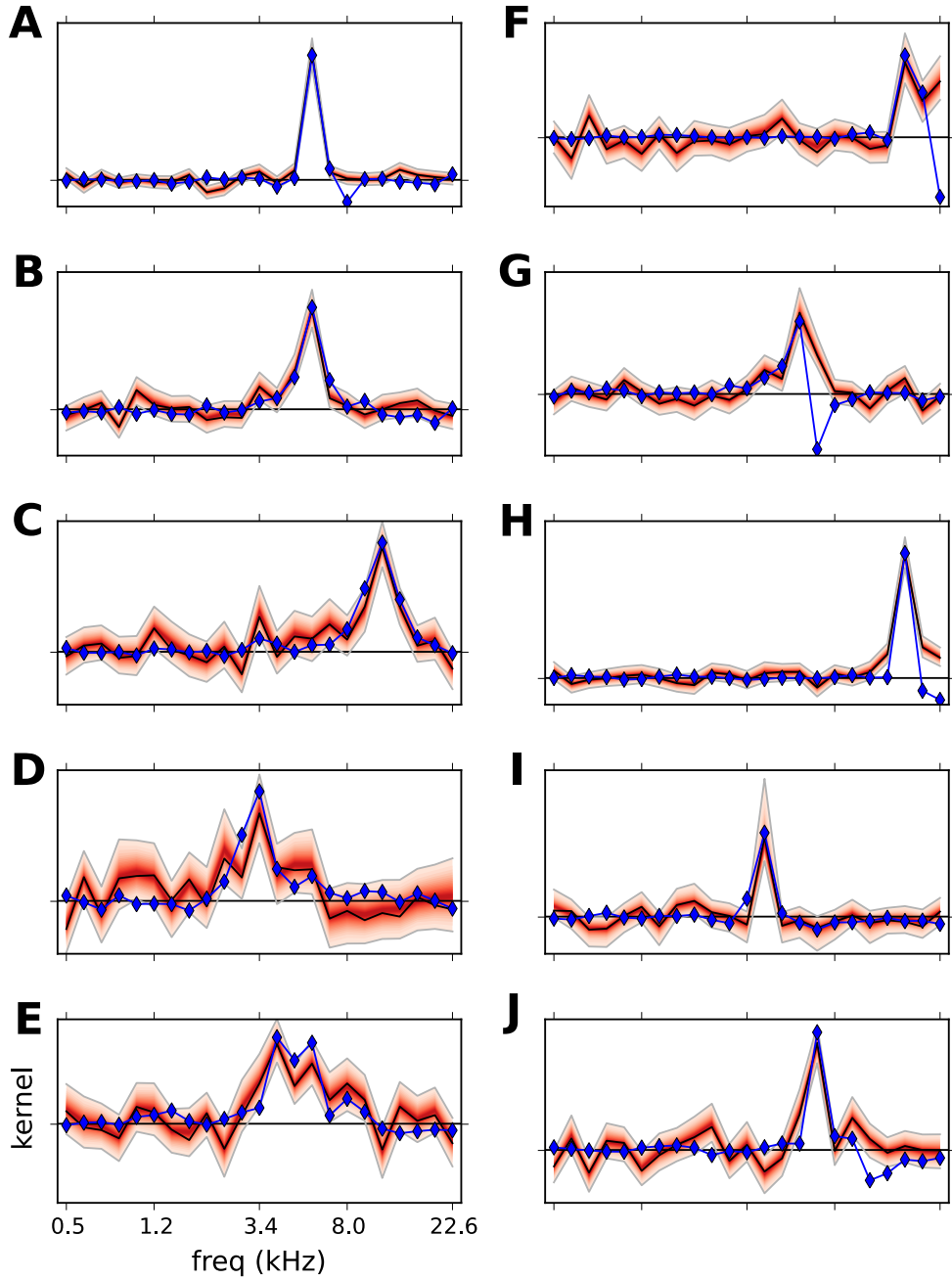


Figure 5.12: Gain contrast kernels, for 10 example units. These are fits of the cd -model (Equation (5.20)), with contrast-independent a and b , and a shared, real-valued contrast kernel, κ_f for c and d . The black line shows the MAP estimate for κ_f ; the red filled region, bounded by the grey lines, shows a 95% credible interval for the posterior distribution over these coefficients. The red shading increases in darkness with probability. The blue line and blue diamonds show the frequency component of the linear, separable STRF, k_f . Both k_f and κ_f have been normalised by their respective standard deviations to facilitate visual comparison.

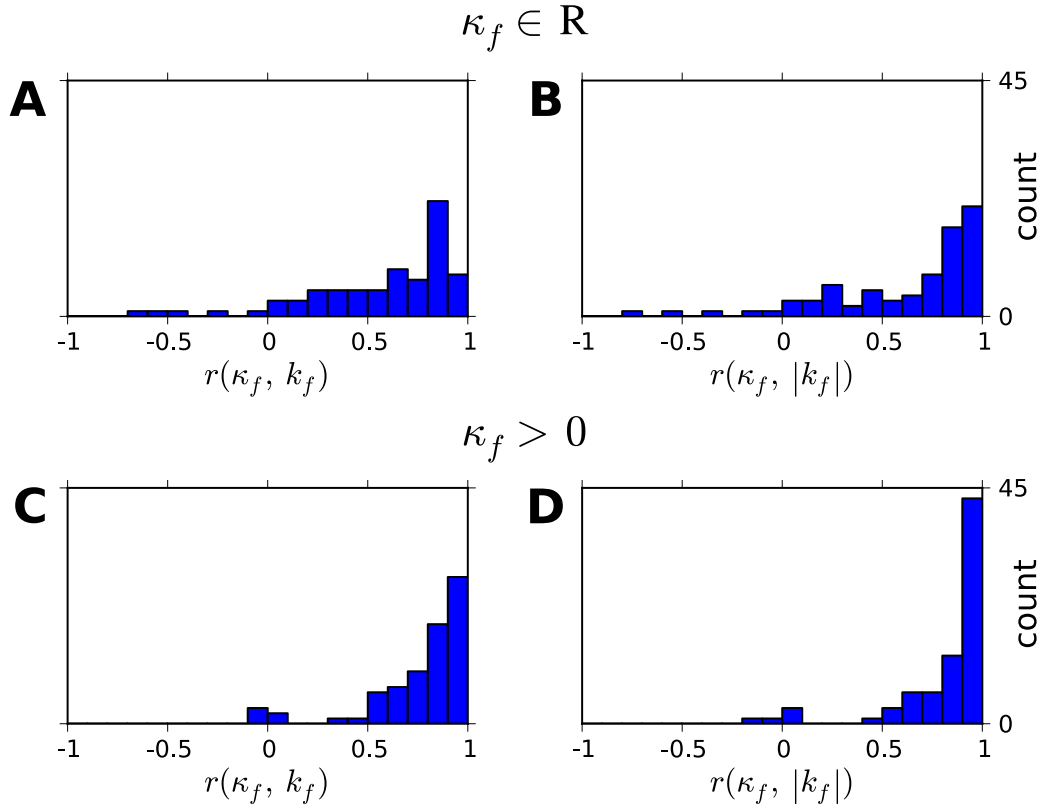


Figure 5.13: Correlation coefficients (r) between the coefficients of the linear STRF kernel (k_f) and the gain contrast kernel (κ_f), on a unit-by-unit basis. (A-B): when coefficients of κ_f were unconstrained (i.e. real valued). (C-D): when these coefficients were constrained to be positive. In general, gain contrast-kernels can be approximated well by $\kappa_f \approx |k_f|$.

Finally, approximately 20% of units (17/77) produced noisy, random-shaped contrast kernels. These models still produced reasonable prediction scores. The results for these units are discussed in more detail below.

5.4.2.4 Model success

Not only did the *cd*-model produce meaningful contrast kernels, it was also very successful at prediction. Using the same model performance measures of Section 5.4.2.1, its predictive power was 68.0 – 68.9% (Figure 5.14). Including a shared contrast kernel for *c* and *d* thus reduces the error variance of the LN model by approximately 23%.

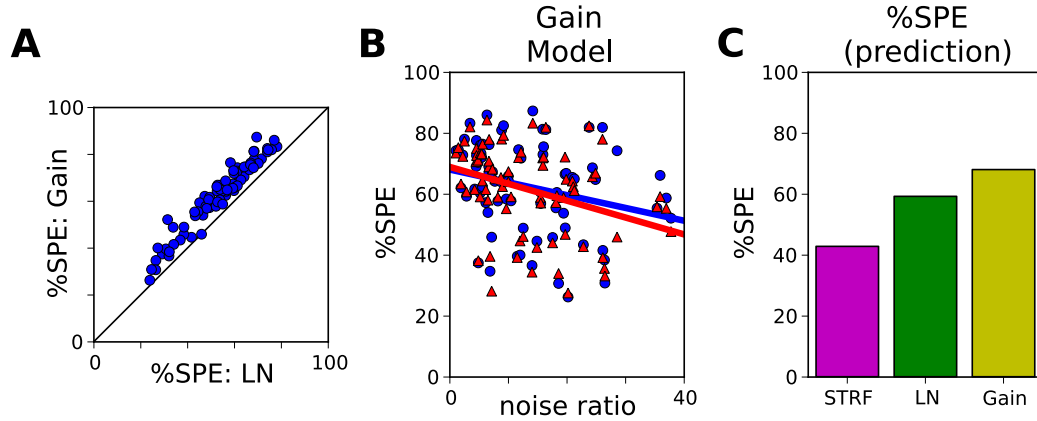


Figure 5.14: Model predictive power: Gain model. (A) On a unit-by-unit basis, the *cd*-model advanced in Equation (5.20) considerably improves upon the LN model. (B) Extrapolation of model success from the population to a hypothetical zero noise neuron. (C) The model predictive power is considerably better than the STRF and LN models. Dots and colours as in Figure 5.10.

According to this model, 72/77 units decreased their gain as contrast increased. The extent of gain changes can be quantified as the ratio $G_d = d_{high}/d_{low}$, which measures the proportional dilation of the output nonlinearity along the z -axis as a result of switching from the all-high-contrast condition to the all-low-contrast condition. A histogram of G_d values for the population of units is shown in Figure 5.15A. The median G_d was 1.92, which is in good agreement with the gain measures of Chapter 4 (Section 4.3.1.1: median $G_{SP} = 1.78$; Section 4.3.4: median $G = 2.01$; both for similar changes in gross stimulus statistics). As expected, units with larger G_d tended to experience the greatest improvements in model prediction by including the gain contrast kernel (Figure 5.15B).

5.4.2.5 Negative coefficients of gain contrast kernels

It is clear from the examples of Figure 5.12 that the most salient features of the gain contrast kernels are their large, positive coefficients in a localised region of frequency space. As mentioned above, these coefficients are often also positive in the inhibitory sidebands. Increasing the contrast of any of these bands thus

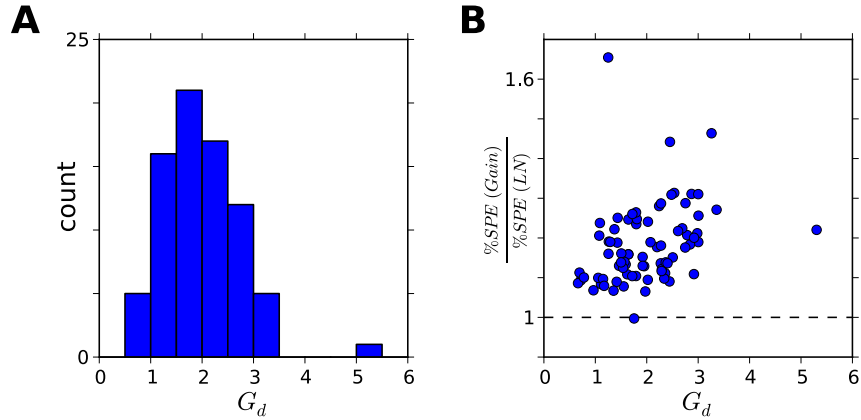


Figure 5.15: Gain model: statistics on contrast-dependent gain changes. (A) The majority of units decreased their gain as contrast was increased, as expected. This is measured here by the ratio $G_d = d_{high}/d_{low}$. (B) The larger a unit's contrast-dependent gain changes, the greater the improvement in model predictive power over the standard linear-nonlinear model. The (non-parametric) Spearman correlation coefficient between G_d and model improvement was 0.44.

yields a decrease in neuronal gain. However, very few κ_f coefficients across the set of models appeared to be genuinely *negative*, such that high contrast in these bands would lead to an *increase* in neuronal gain.

While 45% of all κ_f coefficients were fitted to negative values, these values were typically small in magnitude. They were also generally not significant: the marginal posteriors on these coefficients rarely had all their weight below zero. In total, according to a 95% credible interval criterion, 7% of κ_f coefficients across all units were significantly negative; according to a 99% credible interval criterion, only 3% of κ_f coefficients were significantly negative. These values compare with 19% and 13% for significantly positive κ_f coefficients.

As a result, rather than describing a correlation between the coefficients of units' κ_f and their k_f (as in the previous Section), there was actually a better correlation between κ_f and $|k_f|$ for each unit (see Figure 5.13B), with a median correlation coefficient of $r(\kappa_f, |k_f|) = 0.80$.

There is reason to suspect that none of the coefficients of the gain contrast

kernel should be negative. In principle, negative κ_f values indicate frequency bands for which high contrast would cause an *increase* in neuronal gain. This may not be possible under certain mechanistic implementations of contrast gain control. Negative values may simply arise as a result of overfitting the gain kernels to a small sample of conditions presented.

To test this possibility, I enforced the condition that coefficients of κ_f must be positive. Examples of the resulting kernels are shown in Figure 5.16. This model performed at only a slight disadvantage to using real-valued (i.e. unconstrained) κ_f , but this difference was small: compared with a prediction score for the simple linear-nonlinear model of 59.2% – 59.8%, and for the unconstrained model of 68.0% – 68.9%, the constrained model prediction score was 67.9% – 68.8% (Figure 5.17).

As mentioned above, a number of units' gain contrast kernels were noisy, with little observable structure when κ_f was unconstrained. For all but five of these units, constraining $\kappa_f > 0$ yielded contrast kernels which more closely resembled the units' respective STRF frequency kernels.

A striking feature of the positive-definite kernels is that the correlations between the coefficients of κ_f and $|k_f|$ are even stronger than for the unconstrained models, with a median $r(\kappa_f, |k_f|) = 0.93$ (Figure 5.13D). This suggests that one may be able to approximate the gain contrast kernels simply as $\kappa_f = |k_f|$. I implemented this as a third set of models. This produced promising results: model performance was only slightly impaired, with a prediction score of 67.4% – 68.0%. As a control, I tested a model with $\kappa_f = k_f$, i.e. without the absolute value; this was less successful, with a prediction score of only 65.5% – 66.2%. The main reason for this poorer performance was likely the wrongly-signed κ_f coefficients in the inhibitory sidebands. Finally, ignoring all spectral information and assuming a constant contrast kernel ($\kappa_f = 1/N_F$),

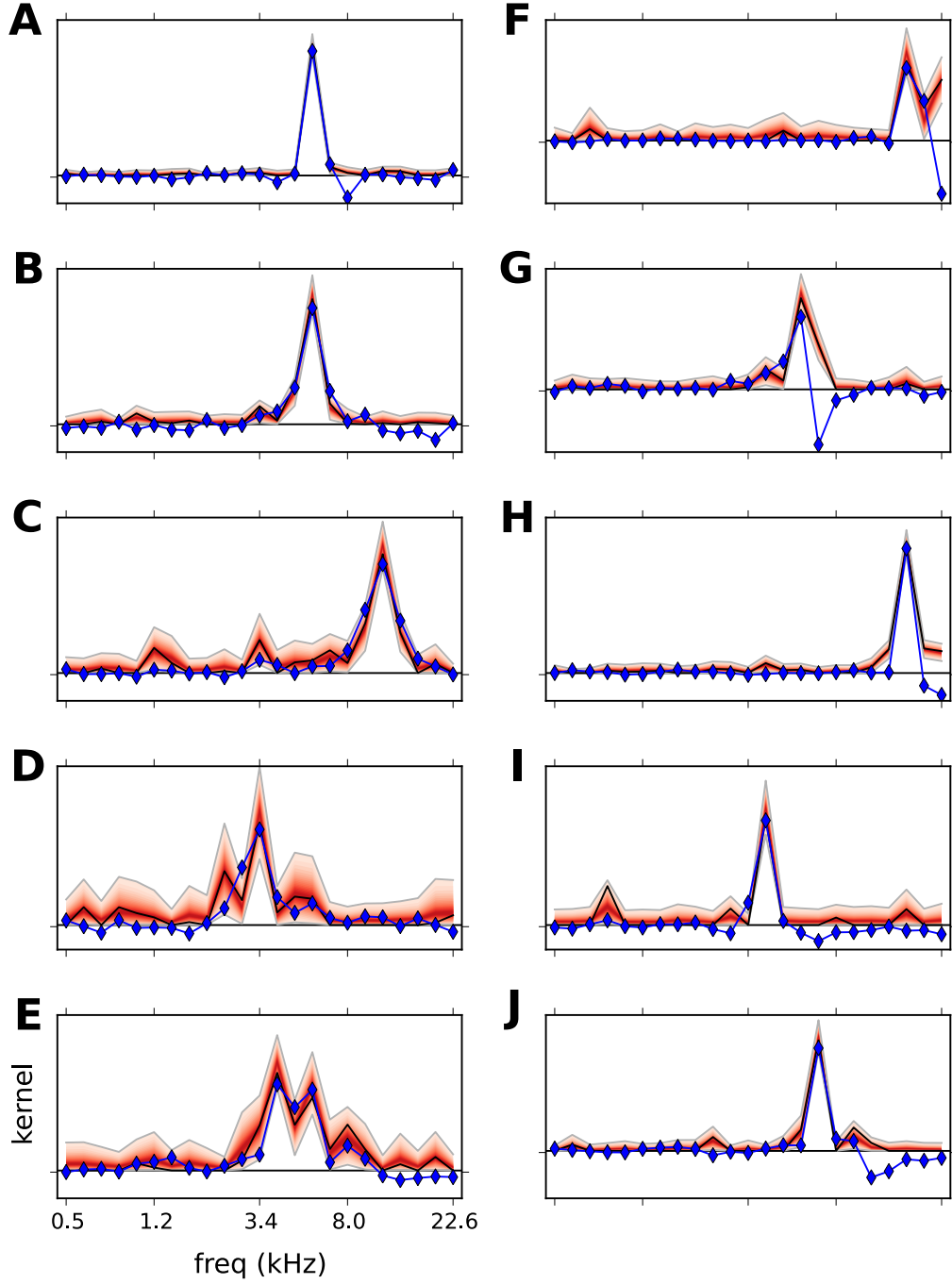


Figure 5.16: Contrast gain kernels, when coefficients were constrained to be positive. This shows the same 10 example units as in Figure 5.12. The same similarities between k_f (blue), and κ_f (black line and red area) can be seen. This model performs almost as well as having unconstrained coefficients.

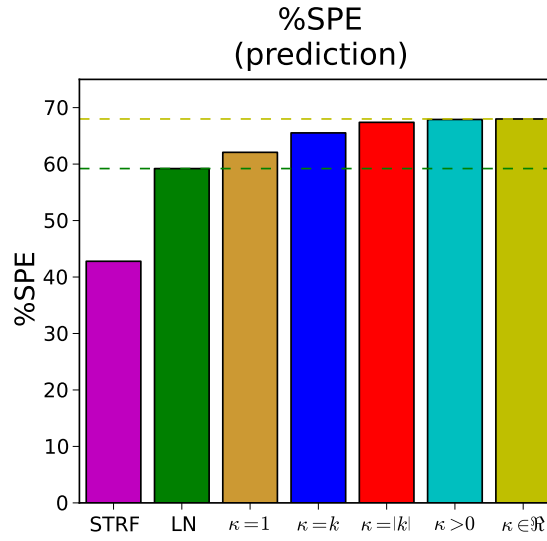


Figure 5.17: Model predictive power: Gain model with constrained coefficients. While the best predictive power is obtained when the contrast kernel coefficients are unconstrained ($\kappa \in \mathbb{R}$; lime green), restricting them to be positive ($\kappa > 0$) yields almost as good prediction scores, as does assuming that they are equal to the absolute value of the STRF frequency kernel ($\kappa = |k|$). Models which do not perform as well include assuming that the contrast kernel is the STRF frequency kernel ($\kappa = k$), or assuming that it is constant with respect to frequency ($\kappa = 1$). These still outperform the simple linear-nonlinear (LN) model. Lime green dashed line shows the model performance for the unconstrained coefficients, green dashed line for the LN model.

the model performed only slightly better than the standard linear-nonlinear model, with a prediction score of 62.1% – 62.7%. These data are summarised in Figure 5.17.

The usefulness of the approximation $\kappa_f \approx |k_f|$ is important. As this experiment demonstrates, the number of contrast conditions needed to estimate gain contrast kernels is large, making it a time-consuming process. On the other hand, including gain changes leads to dramatic increases in model performance. When using the approximation, only two additional variables beyond the standard output nonlinearity need to be estimated (c_1 and d_1), which can be done quickly using only two contrast conditions.

5.4.2.6 Multiple contrast kernels

This experiment began by introducing a general class of models (Equations (5.10) – (5.14)), wherein the four parameters of output nonlinearities could each vary as a function of contrast. Following this, I made a set of simplifications to the model: that parameters a and b were contrast-independent, while c and d shared a common contrast kernel. The prediction score for this cd -model was 68.0% – 68.9%. In this section, I shall now relax these simplifications.

Making the horizontal translation term, c , dependent on contrast was vital to the success of the models. When only d was rendered contrast-dependent, prediction scores were 61.2% – 61.9%, only slightly better than the LN models.

No improvements were made on the models of the previous section by fitting separate contrast kernels for c and d . Rather, prediction scores of the c/d -model declined slightly to 67.4% – 68.6%. There was also little difference between the two contrast kernels fitted for a given unit, with a median correlation coefficient of $r(\kappa_f^{(c)}, \kappa_f^{(d)}) = 0.89$.

Allowing the minimum firing rate, a , to change with contrast, in addition to c and d , also did not improve the model. This was true when I fitted either a shared contrast kernel between all three variables (acd -model: 67.6% – 68.6%), or two separate contrast kernels (a/cd -model; $\kappa_f^{(a)} \neq \kappa_f^{(c)} = \kappa_f^{(d)}$; 67.0% – 68.8%). Finally, allowing all the parameters to change with contrast did not improve predictions either ($abcd$ -model; 67.9% – 68.8%), although a tiny improvement was found with the ab/cd -model (68.1% – 69.8%).

In total, the most parsimonious model for the data is that advanced originally: the cd -model of Equation (5.20). As contrast is varied, auditory cortical neurons' output nonlinearities change their slope and their horizontal shift. These changes can be described as a linear function of the contrast within a spectral

receptive field, which bears strong similarity with the frequencies which make up the neuron's STRF.

5.4.2.7 Cross-frequency interactions

The results presented so far demonstrate that neurons integrate the contrast profile of a stimulus in order to determine their overall gain. The models used to calculate the dependence on stimulus contrast have made a specific assumption: if there is higher contrast in a single frequency band within the contrast kernel, this reduces the gain of that neuron's response to *all* other sounds, regardless of their spectral content.

To test this assumption, I considered what would happen if the gain effects from a high-contrast frequency band were localised to a particular region of frequency space. I built three different models, as described in Section 5.3.2.5: in the *local effect model*, the contrast in a given band, g , could affect the gain of that band alone; in the *global effect model*, the contrast in g could affect the gain of all bands equally; or, in the *regional effect model*, the contrast in g could affect the gain within a range of frequencies centred on g , the width of which was tunable.

The local effect model did not perform as well as the global effect model. The difference was not considerable: the global model scored 65.8% – 69.1%, while the local model scored 64.0% – 67.7%. One reason for this is that many of the units' had SRF bandwidths that were very narrow, such that they were both linearly and divisively sensitive to a single frequency band only; for these units, there is little difference between the local and global models.

In the regional effect model, the effective scale for the spread of gain effects, σ_κ , was allowed to vary. For approximately 60% of units (45/77), the regional

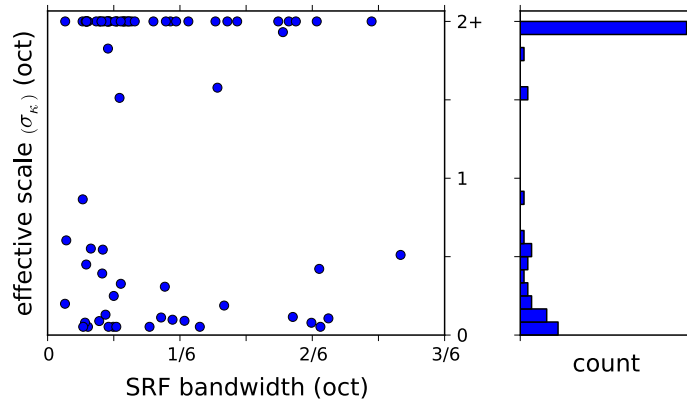


Figure 5.18: Effective scale of the gain effect. To assess whether a single band of high contrast affects a neuron’s response to all frequency bands equally, I estimated a contrast cross-frequency kernel of the form given in Equation (5.24) for each unit. This depends on a parameter, σ_κ , which measures how far the gain effect spreads from the contrast in band g . The fitted values of σ_κ are plotted on the left as a function of the bandwidth of k_f , the spectral receptive field, and on the right as a marginal histogram. Values of $\sigma_\kappa > 10$ have been truncated for the purposes of illustration. There was no systematic relationship between SRF bandwidth and σ_κ .

model converged to the global model, producing a large σ_κ . Amongst the remaining units, the fitted set of σ_κ values was approximately exponentially distributed with a mean value of approximately a third of an octave. However, σ_κ did not appear to be correlated with the bandwidth of units’ SRF (Figure 5.18). Ultimately, the regional effect model did not improve upon the success of the global effect model when considered across the population, producing identical scores of 65.8% – 69.1%. Thus the global effect model, wherein a single band of high contrast induces a non-specific reduction of gain in a unit’s responses to all sound frequencies, remains the most parsimonious description of the contrast gain control observed in ferret primary auditory cortex.

5.5 Discussion

In this chapter, I sought to address whether the gain of auditory cortical neurons depends on the contrast in a limited spectral range, in the same way that

their firing rate depends on the sound level variations within such a range. The data presented in this chapter demonstrate that this is the case.

In the first experiment, the contrast in a test band of frequencies overlapping a unit's spectral receptive field had a greater impact on its gain than the contrast outside of this test band. I estimated that the *local contrast*, i.e. the statistics in bands near to the unit's SRF, had an effect on gain that was approximately five times greater than from the *remote contrast*, i.e. the statistics further from the unit's SRF. The contrast of the remote bands, however, were not irrelevant to gain.

In the second experiment, the contrasts within each frequency band of the DRC were controlled independently. This allowed me to estimate the relative contributions of the contrast in *each* frequency band to the gain of a unit, via a *contrast frequency kernel*, κ_f . These kernels typically placed their weight on the same frequency bands that contributed to a neuron's STRF (Figure 5.12). Thus, when a neuron's firing rate is *linearly* sensitive to the level variations in a particular band, then it is also *divisively* sensitive to changes in the contrast of that band. I also found that $\kappa_f \approx |k_f|$, where k_f is the frequency component of the STRF: the extent to which a band's contrast contributed to a neuron's gain is approximately proportional to the extent to which that band's level contributed to the neuron's firing rate (Figures 5.12 and 5.13). Neurons with narrow tuning curves were sensitive to contrast in a narrow frequency window; the gain of neurons with wide tuning curves could be influenced by contrast over a similarly wide frequency range. Curiously, the spectral region whose statistics determine gain includes a neuron's inhibitory sidebands: high contrast in the sidebands also reduces neural gain.

5.5.1 Local vs. remote spectral contributions to the gain

Both experiments reveal that gain is predominantly determined by the contrast in the preferred frequency range of each neuron. However, there is a potential discrepancy between the two sets of results. In Section 5.2, I observed a significant contribution from the contrast outside the neuron's SRF (Figure 5.3). On the other hand, in Section 5.4, units' gain kernels were found to be highly correlated with the magnitude of the STRF's frequency kernel (Figure 5.13). The results from the first experiment thus suggest that gain control is mediated by a combination of local and remote effects, while the results from the second experiment appear to rule out remote contributions.

This discrepancy can be explained in several ways. In the first experiment, the spectral receptive field was defined by those frequencies whose components of k_f were significantly non-zero, according to a bootstrap procedure. This process could have underestimated the true scope of the SRF, especially if the neuron had a weak dependence on the level of some frequencies more remote from its BF. In this case, the mask band would actually have covered some of the unit's level-SRF. Though the coefficients of k_f would be small over this region, their effect could add over the scope of the mask band.

Another explanation is that gain does depend on the contrast in remote bands, but the gain kernel experiment was not appropriately set up to measure their contributions. As explained below, the contrast kernel model is a linearisation of the relationship between contrast and gain over a constrained subset of contrast space. The linearisation itself is likely to be flawed: like the relationship between sound level and firing rate, the true relationship is undoubtedly highly nonlinear. For instance, high contrast well outside the spectral receptive field of a neuron may only reduce the neuron's gain when it is somewhat broadband, as in the test/mask stimulus paradigm. The subset of contrast space explored

in the second experiment may also have been insufficient to reveal the contributions from remote bands, which may be sufficiently weak that they can only be detected as a compound effect, again, as presented in the test/mask paradigm.

Finally, it is worth pointing out that neurons show different spectral sensitivities under different stimulus conditions. In particular, it has been demonstrated that as the density of DRCs increases, the SRF bandwidths of cortical neurons narrow and lateral sidebands are revealed [Blake and Merzenich, 2002; Valentine and Eggermont, 2004]. The DRCs I used are considerably dense, with 1/4 to 1/6-octave-spaced tones being presented at all times. This means that tone frequencies outside the unit’s SRF — i.e. those for which k_f is of low- or zero-magnitude — may drive the neuron under other stimulus contexts. The gain of cortical neurons may remain sensitive to the statistics of these more remote bands under dense stimulation.

5.5.2 Mechanisms for gain control

These data focus on the effects of gain control, rather than on its specific implementation. Thus, for example, although the models presented here refer to stimulus contrast and level, no assumptions are made about how (or even whether) these parameters are explicitly computed by the brain. Nevertheless, by investigating how the gain signal depends on the spectral integration of stimulus statistics, certain features of the mechanisms underlying gain changes come to light.

Both local and network mechanisms may contribute towards gain control. Here, I observe that STRF frequency kernels and contrast frequency kernels are generally similar, with some small contribution to gain from the contrast in more distant frequency bands.

Synaptic depression is one potential candidate for a contributing mechanism as it has been implicated in gain control across a broad range of neural systems [Stratford et al., 1996; Carandini et al., 2002; Chung et al., 2002; Rothman et al., 2009]. Within the auditory system itself, depression has been observed at the thalamocortical synapse [Rose and Metherate, 2005] and at intracortical synapses [Oswald and Reyes, 2008]. The suppression of synaptic conductances may explain the phenomenon of forward suppression, whereby the response of neurons to a sound is reduced when another sound precedes it [Wehr and Zador, 2005].

The challenge for the synaptic depression hypothesis is to explain why high contrast in one frequency band generally appears to have a uniform effect on the gain of a neuron's representation of *all* frequencies. Synaptic depression is characterised primarily by its selectivity; this, for instance, makes it a favoured candidate for explaining the stimulus-specific adaptation observed in the auditory cortex [Ulanovsky et al., 2004; Percaccio et al., 2005; Wehr and Zador, 2005]. If this process underlies contrast gain control, then we would expect that a single band of high contrast would only reduce a neuron's gain in a limited spectral region, with a bandwidth given by the respective pre-synaptic neuron's spectral selectivity. For many units measured here, this was not the case: a single band of high contrast typically reduced neuronal gain for the entire frequency range, rather than only for neighbouring frequencies. If synaptic depression accounts for the global scope of local contrast changes within the SRF of a single A1/AAF neuron, then at the site of the depressing synapse, the pre-synaptic cell must have at least the same frequency bandwidth as the cortical neuron. Studies on the anatomical and physiological basis for auditory cortical neurons' tuning curves suggest the issue is complex: their spectral receptive fields are produced by a combination of inputs from various subdivisions of the medial geniculate body of the thalamus [Huang and Winer, 2000; Winer et al., 2005] and intracortical inhibition of slightly wider spectral bandwidth

[Wu et al., 2008]; they are potentially also influenced by long-range intracortical connections [Wallace et al., 1991; Budinger et al., 2000; Moeller et al., 2010] and intracolumnar excitatory amplification [Liu et al., 2007].

As an alternative local process, intrinsic cellular mechanisms may contribute towards gain control. Rat A1 neurons in vitro generate hyperpolarising K⁺ currents in response to repeated stimulation [Abolafia et al., 2010]. These provide an input-nonspecific form of adaptation [Nanou et al., 2008]. If this were the case, GLMs (Section 2.2.3) may provide a suitable modelling framework for capturing contrast gain control.

The observation from the test/mask experiment that there is some contribution to gain from the statistics of more remote bands is, however, more difficult to explain via purely within-neuron mechanisms such as synaptic depression or intrinsic hyperpolarising currents. This is because single neurons do not have access to all the information required to measure contrast in these bands and adjust gain accordingly. The influence of distant frequent bands is also unlikely to result from masking in the auditory periphery: although higher spectrotemporal contrast produces larger variation in the level of each frequency band, the total level of the DRCs remains relatively constant both over time and between conditions. It may therefore be the case that there are multiple processes at work, yielding the spectral profile of contrast sensitivity observed here. Finally, it is worth returning to the previous discussion, and noting that the contribution from remote bands was somewhat weak, and did not bear considerable influence over the results of Section 5.4. This may be either a secondary feature, or a secondary mechanism to a more dominant local process.

Other possible mechanisms for contributing towards gain control include the mediation of intracortical networks [Carandini et al., 1997], or of cortico-thalamic feedback, via recurrent excitation and inhibition [Abbott et al., 2005]. In the

primary visual cortex, there is still much debate over the relative contribution of excitation and inhibition to contrast-dependent gain changes (e.g. [Freeman et al., 2002; Webb et al., 2005; Finn et al., 2007; Nauhaus et al., 2009]). Regardless of the underlying mechanism, network-mediated gain changes are compatible with the spectral integration I find here, so long as a given neuron's activity is predominantly modulated by a pool of neurons with similar spectral receptive fields.

It is also likely that gain control in cortex is at least partly inherited from earlier auditory structures. It has been shown, for example, that responses of neurons in the mammalian IC [Kvale and Schreiner, 2004; Dean et al., 2005, 2008] alter their gain to compensate for the temporal contrast of the level of a noise stimulus. If the mechanisms in cortex and midbrain are identical, we would expect gain modulation in the IC to show the same spectral spread as we observe here. This is the subject of Chapter 7.

Finally, there may be a number of independent gain control stages at different levels of the auditory system. These may have different characteristics, reflecting different underlying mechanisms. Such a hierarchy has been observed in the visual system, where at least both the retina and V1 engage separate gain control mechanisms [Carandini et al., 1997; Brown and Masland, 2001; Chander and Chichilnisky, 2001; Baccus and Meister, 2002]. In the extreme, gain control may be performed at every stage along the pathway (for review, see [Kohn, 2007]). If there are multiple, independent stages of gain control, then the local (within-receptive-field) gain effects, and the global (extra-receptive-field) gain effects may be realised by different mechanisms and at different levels of the pathway. Further experiments will be required to distinguish these components by separately measuring their parameters.

These points are explored further in the next two chapters, where the tem-

poral aspects of contrast gain control in cortex (Chapter 6) and the contrast-dependent changes to coding in IC (Chapter 7) are investigated.

5.5.3 Implications for modelling

The contrast kernel model advanced in this chapter provides considerably better predictions of neurons' responses compared with STRF and LN models. For the stimulus sets presented here, the predictions of a linear STRF model could explain 43% of the variance in the responses of an idealised (noise-free) neuron, which the LN model raised to 59%; the gain kernel model outperformed these at 68% (Figure 5.17). Given that the gain kernel can be approximated well from the frequency component of the STRF, this offers a simple and powerful way of extending existing models for the responses of auditory cortical neurons.

The gold standard for models such as these is to be able to predict responses of auditory neurons to natural stimuli [Wu et al., 2006]. Studies which have estimated receptive field models using synthetic stimuli such as these have repeatedly found that the models fail to generalise well to natural sounds [Theunissen et al., 2000; Rotman et al., 2001; Machens et al., 2004; David et al., 2009]. One compelling reason for this is that natural sounds likely engage nonlinear coding mechanisms, which are not activated within the spaces of synthetic stimuli such as DRCs or ripples [Theunissen et al., 2000; Woolley et al., 2006; David et al., 2009]. Furthermore, the linear approximations made during model construction are sensitive to the statistics of the subspace of stimuli explored [Christianson et al., 2008].

Since natural scenes vary in their statistics over time, it is likely that including gain control will improve the predictions of STRF-based models. Interestingly, a model of auditory neurons including divisive normalisation has recently been

tested by [David et al. \[2009\]](#), with little success. However, their model effectively assumed that neurons' gain was determined by the contrast over the whole spectral range, rather than from the contrast in a band-limited region, or as weighted by a contrast kernel. They did, however, implement a simple model of synaptic depression, and found that it was broadly successful in explaining the differences between STRFs estimated under different stimulus regimes. Their synaptic depression model is similar to the local-effect limit of the contrast cross-frequency kernel model advanced here, using a 1/4-octave spectral resolution. As I demonstrate in Section 5.4.2.7, the scope of gain effects is generally wider than this. Nevertheless, this demonstrates the potential for this class of models.

It is worthwhile considering the assumptions made in constructing the models of Section 5.4. The criticisms I raised above of extending synthetic stimulus models to natural domains are especially salient: by design, the DRCs used are spectrotemporally dense, their modulation properties are fixed, and their contrast statistics were varied parametrically, while keeping other statistics constant. These are significant departures from the properties of natural sounds. Furthermore, neuronal gain was only measured after sufficient time had been given for a constant adaptation state to be reached (the time course of gain control is explored in more detail in the next chapter). The models that arise should thus be considered a linearisation about local changes in these parameters. Despite mechanistic interpretations of these models, they are in essence phenomenological, and are constrained by tractability and ease of representation. It is likely that adding more physiological detail will eventually improve their power, when such information is widely available.

5.5.4 Computational principles

In this chapter, I show that when stimulus level statistics are not uniform across the spectrum, gain control is also unevenly applied to neurons, depending on their frequency tuning. A spectrally-limited band of high contrast has the greatest compressive effect on neurons if their tuning curves overlap this band. Conversely, neurons tuned to other frequencies maintain sensitivity to small changes in their input. Because natural sounds do not cover the entire audible frequency range evenly, such an arrangement might make it possible to match contrast gain control to the challenges posed by each particular acoustic environment.

In Section [5.3.1](#), I introduced contrast kernels models as subserving a dual purpose. First, they act as better forward models: the results of this chapter demonstrate that they provide more accurate predictions of neural responses when the stimulus has a non-trivial spectral pattern of contrast statistics. The models also provide the opportunity to understand how the brain's representation of sounds is affected by contrast. Before this can be addressed, it is helpful to know not only how cortical neurons integrate contrast over frequency, but also over time. This is the subject of the next chapter.

Chapter 6

Temporal integration in contrast gain control

6.1 Introduction

Thus far, we have seen that the contrast of an ongoing sound has a direct impact on how neurons in the ferret auditory cortex encode it. In Chapter 4, I demonstrated that decreasing stimulus contrast leads to an increase in neuronal gain, such that smaller changes in sound level are amplified to have a larger effect on neurons' firing rate. In Chapter 5, I investigated the spectral integration that underlies this effect. I showed that neurons are mostly sensitive to the contrast within their spectral receptive field. Thus, when level changes in a frequency band additively (or subtractively) contribute to a neuron's firing rate, then contrast changes in this same frequency band divisively contribute to its gain.

In an analogous manner to the previous chapter, this chapter describes an investigation of the *temporal* integration that underlies contrast gain control. This involves posing two related questions. In the first question, we consider the time course of changes in gain. If the contrast of a stimulus is changed, how long does it take for gain changes to manifest? In the second question,

we consider this from the perspective of a kernel. In Chapter 5, I measured neurons' *contrast spectral receptive fields*, i.e. how they weight the contrast in different frequency bands to determine their gain. Here I shall ask: what are neurons' *contrast temporal receptive fields*, i.e. how do they weight the *recent history* of contrast to determine their gain?

To address these questions, the structure of this chapter is divided into three parts, which parallel those of Chapter 5. I conduct two experiments here: in Section 6.2, I embed a probe stimulus into DRCs of high and low contrast, and assess how the contrast of the context affects the response to this “test sound”. By placing the test sound at different delays after a switch in context, I estimate how long it takes for the change in context to yield a change in the test sound response. This experiment directly addresses the first question. It also uses DRC stimuli similar to those used in Chapter 4 and Section 5.2. In Section 6.3, I extend the models of Section 5.3 to depend not only on the spectral profile of the stimulus contrast, but also its recent history. I then fit these models to the responses of auditory cortical neurons in Section 6.4, using the datasets of Section 5.4. This experiment thus directly addresses the second question.

As for spectral integration, these results provide further insights into the mechanisms underlying contrast gain control, and underpin a set of improved models for the coding of sound features by auditory cortical neurons. They also inform a unified view of how the primary auditory cortex deals with contrast.

6.2 Temporal Experiment 1: Fixed stimulus in different contexts

6.2.1 Methods

6.2.1.1 Animals and recordings

Data in this experiment were obtained from three of the ferrets used in Chapter 4. These were not the same animals used in Experiment 1 of Chapter 5. A full description of the anaesthesia, surgical, stimulation and recording procedures is provided in Section 2.3. As in Section 5.2, the stimuli were tailored to the response properties of units being recorded, and were thus presented immediately after the standard contrast DRCs of Chapter 4.

6.2.1.2 Stimuli

The goal of this experiment was to determine the time course over which contrast-dependent gain effects are established. As a first pass for measuring this, I constructed a test stimulus which could be embedded within DRCs, and used it as a specific probe of the current gain state. As will be explained, the time course of gain changes could be measured by presenting the probe at different delays after a change in stimulus contrast.

For each electrode penetration used, the STRF of a representative unit was estimated on-line, and used to create a test sound. The frequency component of the STRF, k_f , was scaled to create a single chord of 25 ms duration, $X_f^{(T)}$, that roughly fit the statistics of a DRC segment with intermediate contrast ($\sigma_L = 5.8$ dB, $c = 64\%$). A set of new DRCs was generated for that electrode penetration, consisting of 25 alternating 1 s segments of low ($\sigma_L = 2.9$ dB, $c = 33\%$) and high contrast ($\sigma_L = 8.7$ dB, $c = 92\%$). The test sound was inserted

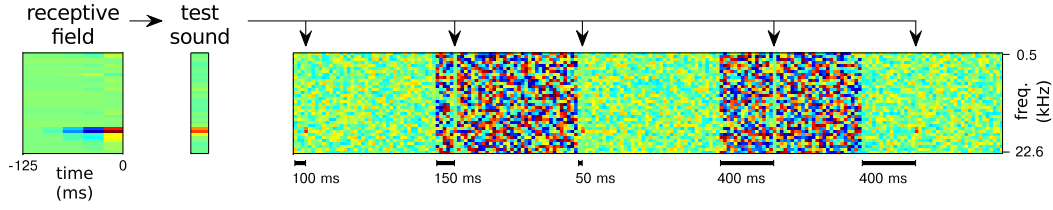


Figure 6.1: Schematic of test sound stimuli. During each electrode penetration, the STRF of a representative unit was used as a basis for a test sound. This was inserted at random times into special DRCs in which stimulus contrast switched every 1 s between low contrast ($\sigma_L = 2.9$ dB, $c = 33\%$) and high contrast ($\sigma_L = 8.7$ dB, $c = 92\%$). The test sound itself was identical within each stimulus regime; only the contrast of its context differed.

once into each segment, at a random delay after each segment transition. The test sound thus replaced the chord of the DRC sequence at that time. This is illustrated in Figure 6.1. Forty sequences, with different random seeds and test sound timing, were presented.

6.2.1.3 Unit selection criteria

Units were selected for this study on the basis of several criteria. These were similar to those in Section 5.2.1.5. Firstly, units were required to modulate their firing rate in response to the DRC; the minimum noise ratio (in both high and low contrast DRC stimulus) that was admitted was $NR = 10$. Secondly, the LN models that were fitted for the low and high contrast conditions were required to be predictive (as defined in Section 4.2.4).

Next, it was necessary to ensure that the test sound actually drove a given unit. Since either one or two test sounds were constructed for each (bilateral) electrode penetration, within which there was some variation in receptive field properties, I applied two criteria. I required that the test sound have energy within the unit's receptive field: only those units for which $\frac{X_f^{(T)} \cdot k_f}{\|k_f\|} > 10$ were retained for analysis. I also required that the firing rate response to the test sound was larger than the average firing rate in response to the DRC stimulus

(as measured below in Section 6.2.1.4). I found 77 units which satisfied these criteria. These were used for Section 6.2.2.

To ensure that reliable estimates of time constants were obtained, I limited my analysis of adaptation time course to units that satisfied a further two criteria. Firstly, it was necessary that the mean peak response (defined below in Section 6.2.1.4) was different in the high and low contrast context conditions. Any fitting of exponentials of the form given in Equation (6.1) to units where there was no difference between peak responses in high and low conditions would only be able to capture noise. Secondly, I calculated the noise ratio of the set of peak responses to the test sound, across all conditions for that unit. As for the peri-stimulus time histograms (PSTHs), NR=0 indicates that the peak responses were identical for each repeated stimulus presentation, while higher NR indicates that responses were less reliable. As for the STRFs, I restricted my analysis to units whose NR for peak responses was less than 10.

6.2.1.4 Responses to the test sound

To measure how a unit responded to the test sound, I first estimated its response latency. I binned the spiking response to the test sound at a 5 ms resolution, and averaged the resultant PSTH over all conditions, i.e. each combination of context (contrast of the DRC segment), and timing (delay after transition). I then defined a 15 ms window about the peak of the PSTH. For each condition, spiking within this fixed window was defined as the peak response, r .

6.2.2 Results

In this experiment, I make use of the DRC stimuli in a somewhat different way from Chapters 4 and 5. In those experiments, I changed the contrast

statistics of the DRCs, and showed that the way that neurons responded to the stimuli depended on those statistics. Thus, the DRCs functioned both as the context (which set the gain state), and the probe itself (from which the gain was measured). Here I make a stronger distinction between these two functions, and consider how the *recent* spectrotemporal statistics encountered by a neuron modulate its responses to a *new* sound. More specifically, I predicted that if a particular sound was presented in a low contrast context, it would generate stronger responses than if it were presented in a high contrast context.

To test this prediction, I embedded a fixed “test sound” into DRC segments of differing contrasts. This sound was designed to drive many units recorded from during an electrode penetration, by having stimulus energy within the receptive fields of these units (Figure 6.1). The different contexts were provided by a DRC sequence which alternated between high ($\sigma_L = 8.7$ dB, $c = 92\%$) and low contrast ($\sigma_L = 2.9$ dB, $c = 33\%$) every 1 s. The same test sound was presented once per 1 s block, at a random time relative to the onset of that block, i.e. the last switch in context.

Among 77 units which responded reliably to the test sound (see Section 6.2.1.3), all but two responded more vigorously when this sound was presented in a low contrast context than in a high contrast context. The firing rate was a median 2.6 times greater in low contrast context ($P \ll 0.001$, sign-rank test; Figure 6.2). This confirmed the prediction.

6.2.2.1 Time course of gain changes

The goal of this experiment was to estimate how long it takes for contrast-dependent gain changes to take effect in cortical neurons. The test sound paradigm provided an opportunity to do so. I therefore not only presented the

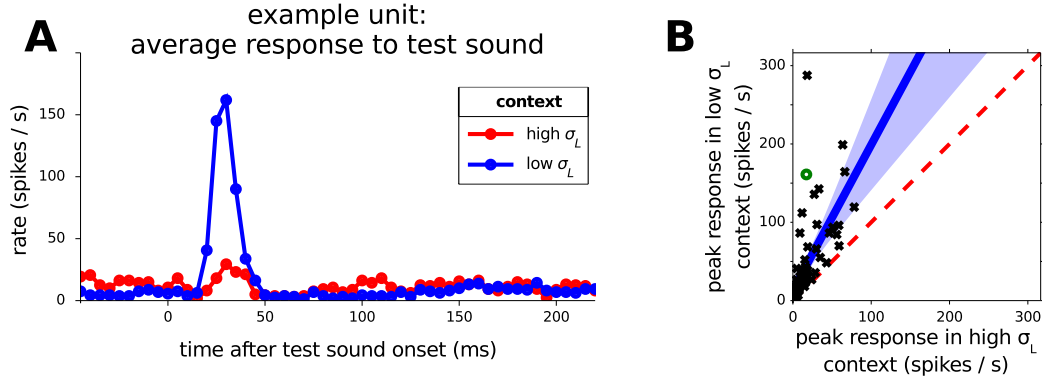


Figure 6.2: Response to the test sound when presented in different contrast contexts. (A) Mean response to the test sound for an example unit, when presented in high contrast context (red), or low contrast context (blue). Responses were averaged within each contrast context over all post-switch delays from 150 – 800 ms. (B) Peak responses to the test sound for $n = 77$ units, during the low contrast and high contrast contexts. Red dashed line shows the expected response relationship if the contrast context was irrelevant. The green circle indicates the unit in (A).

test sound in both high and low contrast contexts, but also at a range of delays after a switch in context had been made (Figure 6.1).

A number of units' responses visibly changed as a function of the post-switch delay. This is exemplified in Figure 6.3A. For the unit shown here, the later the test sound was presented after a switch to the high contrast context, the smaller was the amplitude of the peak response. The reverse was seen in low contrast contexts. For each of these observations, the amplitude of the peak response could be described well by an exponential decay (e.g. Figure 6.3B), given by:

$$r(t) = a + b \cdot e^{-t/\tau} \quad (6.1)$$

For 18 units, I was able to obtain reliable estimates of time constants for both the switch from low-contrast to high-contrast context ($\tau_{L \rightarrow H}$) and the switch from high-contrast to low-contrast context ($\tau_{H \rightarrow L}$). The adaptation to high-contrast context occurred with a median $\tau_{L \rightarrow H}$ of 86 ms, compared with a slower adaptation to low-contrast context with a median $\tau_{H \rightarrow L}$ of 157 ms. This

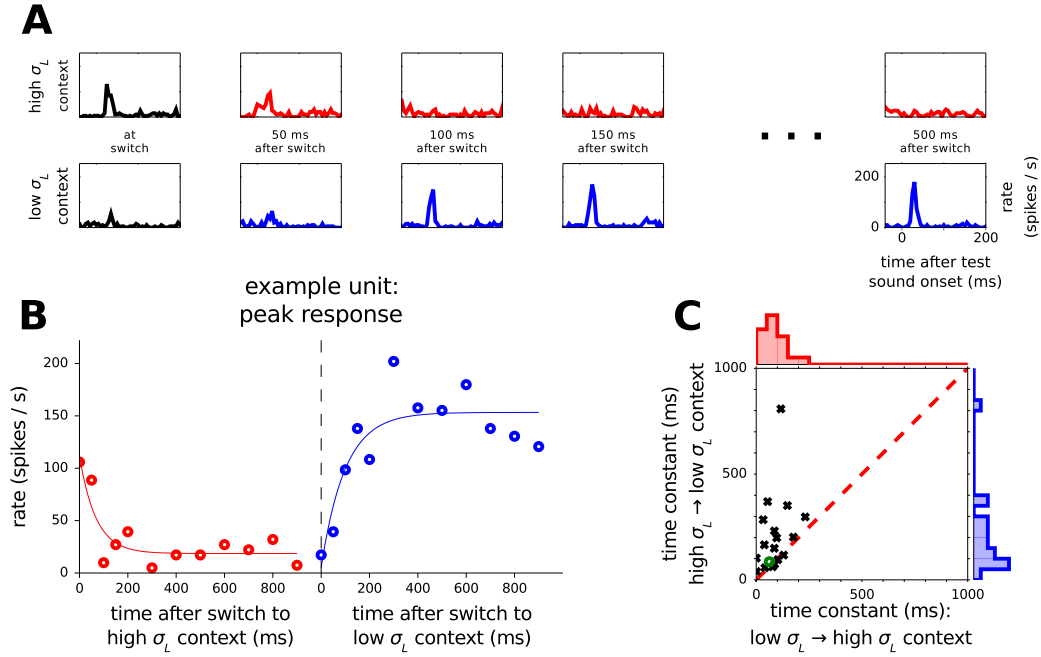


Figure 6.3: Time course of gain changes, measured via the test sound. (A) Mean response to the test for the same example unit as in Figure 6.2A, delineated (by column) by the time since the last switch in context at which the test sounds were presented. (B) Peak response for the unit in (A), as a function of time after context switch at which the test sound was presented. Solid lines show exponential fits to these data, with time constant $\tau_{L \rightarrow H} = 62$ ms after an increase in the contrast of the context, and $\tau_{H \rightarrow L} = 85$ ms after a decrease in the contrast of the context. (C) Time constants for context adaptation, as in (B), for 18 units for which both $\tau_{L \rightarrow H}$ and $\tau_{H \rightarrow L}$ could be reliably estimated. Data are plotted both as a scatter plot, and as marginal histograms of $\tau_{L \rightarrow H}$ (red) and $\tau_{H \rightarrow L}$ (blue). Green circle denotes unit in (A) and (B).

difference was significant ($P < 0.001$, sign-rank), and evident for 14/18 of the individual units (Figure 6.3C). Thus, the time courses for increases and decreases in neural gain are asymmetric.

6.2.3 Preliminary Discussion

This experiment addressed the question of how long it takes for contrast changes to produce gain changes. The results presented here show that these changes occur with an exponential shape, with a time constant of approximately 110 ms.

The effects of increasing contrast manifested slightly faster than the effects of decreasing contrast.

In the next experiment, I ask a related question: what weight do neurons give to the *recent history* of contrast to determine their *current* gain? This involves extending the models developed in Section 5.4 by explicitly introducing a contrast time kernel alongside the contrast frequency kernel. The results above predict that the contrast time kernels of auditory cortical neurons will have an approximately exponential shape.

6.3 Contrast kernel models

6.3.1 Contrast time kernels

The models developed in Section 5.3 aimed to capture how contrast affects the response properties of auditory neurons via changes to the shape of their output nonlinearities. These nonlinearities generally had a sigmoidal shape, governed by four parameters (as in Figure 5.4):

$$\hat{y}_t = a + \frac{b}{1 + e^{-(z_t - c)/d}}$$

In turn, each parameter $\theta \in \{a, b, c, d\}$ was rendered contrast-dependent (and hence time-dependent), by making θ a weighted sum of the spectral profile of contrast at that time t , σ_{tf} . The weightings were captured by a contrast kernel, $\kappa_f^{(\theta)}$:

$$\begin{aligned} \hat{y}_t &= a_t + \frac{b_t}{1 + e^{-(z_t - c_t)/d_t}} \\ \theta_t &= \theta_0 + \theta_1 \cdot \left[\sum_f \kappa_f^{(\theta)} \cdot \sigma_{tf} \right] \end{aligned}$$

These models assume either that neurons have no memory for the recent history contrast, or that there had been no recent changes in σ_{tf} . The latter was, for the large part, true for the data considered in Section 5.4: the first 500 ms of responses from each segment were discarded, so that for the remainder of each segment, σ_{tf} was constant over time. This meant that the values of each parameter θ were constant within each segment.

In the more general case, we want to consider how neurons might integrate the recent history of contrast, and how we might appropriately determine the values of θ . A simple approach to this problem, which I consider here, is to broaden the linear contrast kernels from the frequency domain, $\kappa_f^{(\theta)}$, to the frequency-time domain, $\kappa_{fh}^{(\theta)}$.

To do this, I first extend the contrast 2-tensor, σ_{tf} , into a 3-tensor, σ_{tfh} , in the same way as the sound level 2-tensor, X_{tf} , was extended to X_{tfh} (see Section 2.2.1 and Equation (2.17)). The contrast 3-tensor is indexed by an additional history index, h , such that σ_{tfh} gives the contrast in frequency band f , at time $t - h$. Thus, at time t and in frequency band f , σ_{tfh} reports what the value of the contrast was h time bins in the past.

Next, each parameter θ is calculated via a linear weighting over the recent history of contrast:

$$\hat{y}_t = a_t + \frac{b_t}{1 + e^{-(z_t - c_t)/d_t}} \quad (6.2)$$

$$\theta_t = \theta_0 + \theta_1 \cdot \left[\sum_{fh} \kappa_{fh}^{(\theta)} \cdot \sigma_{tfh} \right] \quad (6.3)$$

Each coefficient of the contrast kernel for θ , $\kappa_{fh}^{(\theta)}$, measures how the contrast in band f , h time bins in the past, linearly affects the parameter θ . This is illustrated in Figure 6.4.

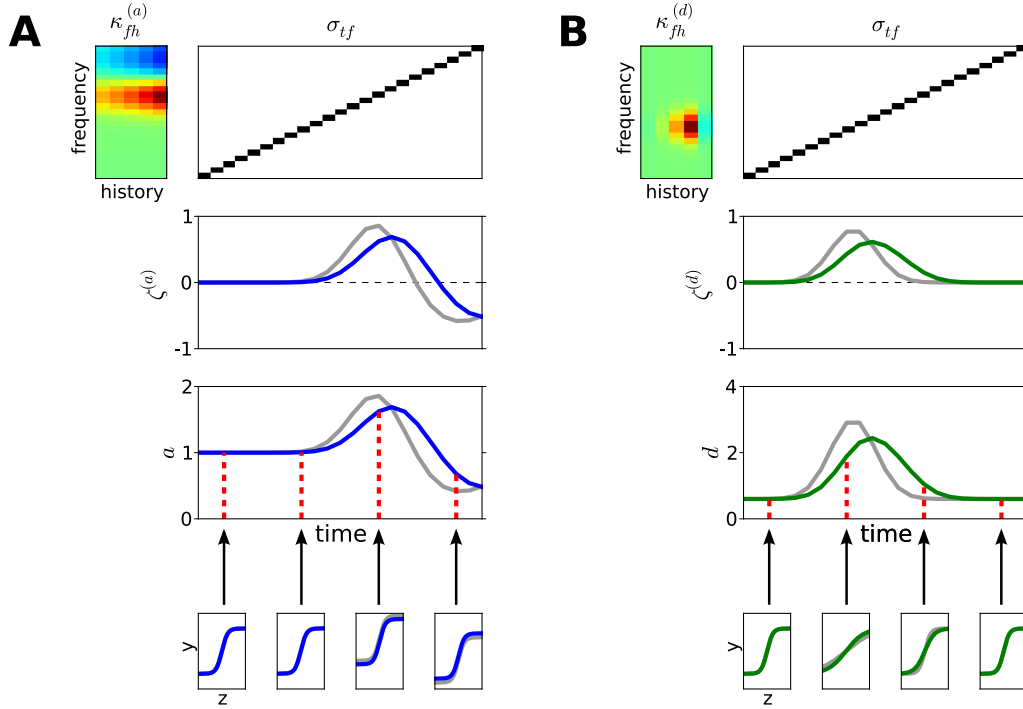


Figure 6.4: Illustration of the contrast time-frequency kernel model. The diagram shows two examples of how introducing a temporal component to the contrast kernels can capture a dependence of the nonlinearity parameters on the recent history of contrast. This figure parallels Figure 5.5. (A) Effect of a contrast time-frequency kernel for the nonlinearity parameter a , which sets the minimum (threshold) firing rate. Top left panel: the contrast kernel, $\kappa_{fh}^{(a)}$ is shown. This comprises a set of weights assigned to each time-frequency bin, measuring the effect of the contrast in that time-frequency bin on a . Red denotes positive coefficients, blue negative. The contrast time kernel has a long memory. Top panel: the contrast profile of an example stimulus. Second panel from top: the weighted contrast signal, $\zeta_t^{(a)}$, in blue. Grey shows the contrast signal if history is ignored, and only the contrast frequency kernel, κ_f is used. The long memory of the contrast time kernel slows changes in $\zeta_t^{(a)}$. Third panel: as a result of changing contrast, the minimum firing rate, a changes with time. Bottom: the output nonlinearities. (B) Effect of a contrast time-frequency kernel for the nonlinearity parameter d , which sets the (inverse) gain of the output nonlinearity. The contrast time kernel includes a delay (and a short memory).

6.3.2 Simplifications

As for STRFs, the contrast time-frequency kernels contain a large number of parameters that need to be fitted. This number can be dramatically reduced if we consider a separable approximation of $\kappa_{fh}^{(\theta)}$ as the outer product of a contrast frequency kernel, $\kappa_f^{(\theta)}$, and a *contrast time kernel*, $\kappa_h^{(\theta)}$, in the same manner as for separable STRFs (Equation (2.20)):

$$\kappa_{fh}^{(\theta)} = \kappa_f^{(\theta)} \otimes \kappa_h^{(\theta)} \quad (6.4)$$

As in Section 5.3.2.4, we can also limit the model so that only a subset of the parameters are contrast-dependent. Thus, for example, in the *cd*-model, only a single, shared contrast time-frequency kernel needs to be fitted.

In the experiment below, I use the separable approximation to the *cd*-model. In particular, this means that the fitted values of κ_f from Section 5.4 can be fixed, and we can focus on the temporal component of the contrast kernels, κ_t .

6.4 Temporal Experiment 2: Contrast time kernels

6.4.1 Methods

6.4.1.1 Animals and stimuli

The data presented here come from the same animals, units, and stimulus presentations as described in Section 5.4. As explained below, this section describes an additional analysis of responses to the RC-DRCs.

In Section 5.4, the focus was on estimating how the contrast in different frequency bands contributed towards a neuron's gain, i.e. the contrast frequency

kernels, κ_f . I therefore excluded data from the first 500 ms following each segment transition, as, presumably, the neurons would require some time to reach a constant gain state. As segments were 3 s in duration, the remaining 2.5 s were available for fitting κ_f . Here, the focus is on these periods of transition between segments of different contrast profiles. I therefore return to the previously discarded data.

6.4.1.2 Contrast time kernels

As explained in Section 6.4.2.1, the modelling of this experiment extends the equations of Section 5.4 by providing a time dimension to the contrast kernels. Together with a separable approximation to the contrast kernels, the full model is given in Equations (6.2) – (6.4). I fitted contrast kernels with the index h running over $N_H = 20$ time bins, starting at $h = 0$ (i.e. the current contrast). This model thus considered the contrast history of the last 500 ms of stimulation.

In the previous chapter, I only considered periods of the RC-DRCs where the recent contrast (for the last 500 ms) was constant. Thus, the model fits and predictions in that work are independent of the values of κ_h .

As with Equation (5.27), I normalised κ_h to sum to unity by defining an unnormalised kernel λ_h , such that:

$$\kappa_h = \frac{\lambda_h}{\sum_{h'} \lambda_{h'}} \quad (6.5)$$

Derivatives of the model with respect to the coefficients of λ_h are similar to those for the contrast frequency kernels, λ_f , as given in Appendix A. They can be obtained by replacing the index f with h on all tensors, and $\sigma_{tf}^{(\theta)}$ with $\sigma_{th}^{(\theta)}$, where $\sigma_{th}^{(\theta)} = \sum_f \kappa_f^{(\theta)} \cdot \sigma_{tfh}$.

6.4.1.3 Model success

As in the previous chapter, model success was measured by the percentage of signal power explained (%SPE) on the training and prediction datasets for each unit. These values were then extrapolated to an idealised zero-noise neuron. The process was repeated for 100 partitions of the data, and the median of the scores taken.

Incidentally, this extrapolation procedure often yielded better scores on the prediction dataset than on the training dataset. Since the prediction scores should provide a lower bound, and the training scores an upper bound on the model's usefulness (Section 5.4.1.10), these results are somewhat paradoxical. However, this can likely be attributed to the small sample sizes: the datasets used to fit and predict models in this section were approximately 15% of the size of those used in Section 5.4. As a result, there was more variance in estimates of signal power and fractional variance explained. To avoid any confusion with these numbers, I therefore generally cite the prediction scores and omit the training scores when presenting data in this section. Given that the models presented here are nested (e.g. exponential κ_h is a special case of positive κ_h , which is a special case of real-valued κ_h), training scores always increased with the degrees of freedom of the model.

6.4.2 Results

The goal of this experiment was to provide a second, independent estimate of the time course of contrast-dependent gain changes, using an analysis similar to that of Section 5.4.2.

I therefore analysed the same 77 units used in Section 5.4.2, but considered a different subset of the RC-DRC data. In Section 5.4.2, the focus was on

estimating how the contrast in different frequency bins contribute to a neuron's gain; it was therefore necessary to exclude a short transition period after a change in the contrast profile of the DRC, to ensure that a constant gain state had been reached. Here, I focus on this transition period alone.

6.4.2.1 Contrast time kernels

In this experiment, I estimated the contrast time kernels of cortical units. I made several simplifying assumptions, motivated, in part, from the results of the previous chapter. I used the cd -model, such that a and b were assumed to be independent of contrast, and the contrast kernels for c and d were shared, such that $\kappa_{fh} = \kappa_{fh}^{(c)} = \kappa_{fh}^{(d)}$. I also assumed that the contrast kernels were separable in frequency and time (Equation (6.4)). I constrained the coefficients of κ_f and of κ_h to be positive. Finally, all parameters except κ_h were fixed for each unit at their maximum a posteriori values, as previously fitted in Section 5.4.2.5. In total, the only parameters to be fitted were the coefficients of κ_h . Specific details of the fitting procedures are given in Section 6.4.1.2.

Examples of contrast time kernels are shown in Figure 6.5. As for the data in the first experiment of this chapter, these appear to be roughly exponential in shape: units were most sensitive to the contrast in the most recent 50 ms of stimulation, and retained a weak dependence on the contrast statistics further back in history. The exponential shape of κ_h is more pronounced in the population average (Figure 6.6), which has a time constant of $\tau = 86$ ms.

Including the time kernel for gain changes led to an overall improvement in the model predictive power. Using the same model assessments metrics as in the previous chapter (Section 5.4.1.10), the prediction scores were 43.1% for the STRF model, 59.7% for the linear-nonlinear model, 64.3% for the

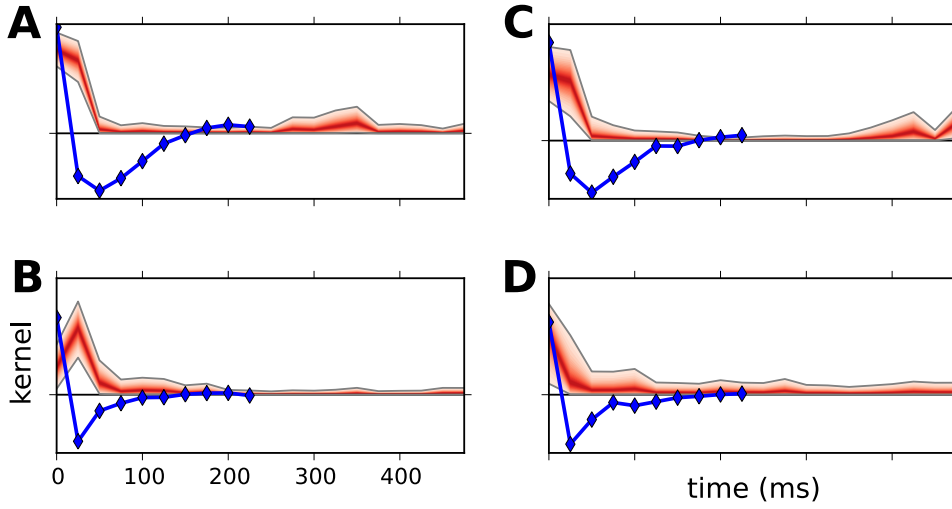


Figure 6.5: Contrast time kernels, for four example units. These are fits of Equations (6.2) – (6.4). As in Figure 5.12, the red filled region, bounded by the grey lines, shows a 95% credible interval for the posterior distribution over these coefficients. The red shading increases in darkness with probability. The blue diamonds and line show the time component of the respective units' STRFs, k_h . These were fit over a shorter history compared with the contrast kernels.

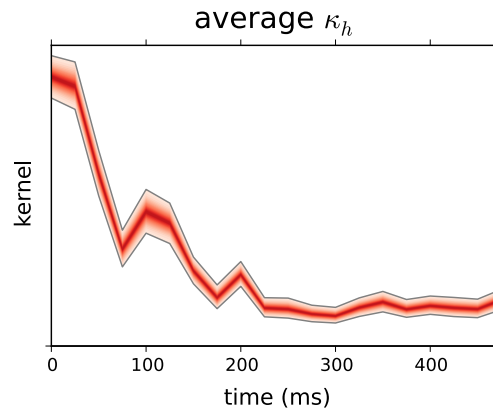


Figure 6.6: Mean of the contrast time kernels from the 77 cortical units, $\bar{\kappa}_h$. This shows the approximately exponential shape of the time kernels, with a fitted time constant of 86 ms. As the coefficients of κ_h were constrained to be positive, and a number of units yielded noisy fits of κ_h kernels, there is an upward bias on the smaller coefficients of $\bar{\kappa}_h$ (and, given the normalisation from Equation (6.5), a downward bias on the larger coefficients). Therefore, the fact that $\bar{\kappa}_h$ does not drop to zero at large h does not imply that history dependence extended beyond 500 ms, or even significantly beyond 250 ms. Testing this was beyond the scope of this experiment.

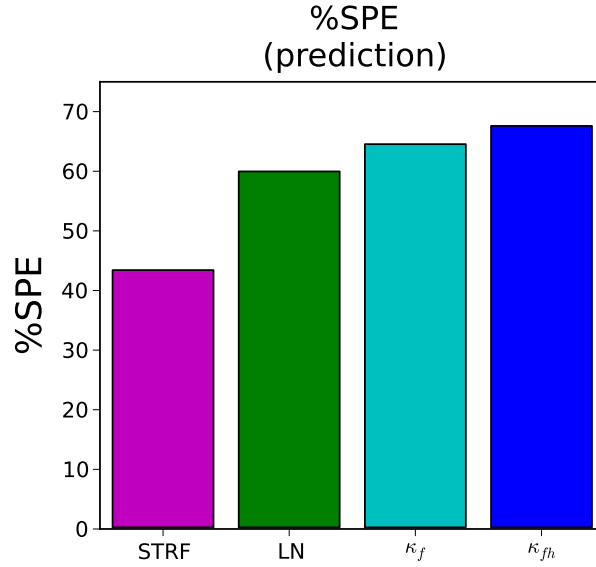


Figure 6.7: Model predictive power: Contrast time kernels. Including a history component to the contrast kernels (κ_{fh}) improves the performance of the model, compared with an assumption that only the current contrast matters (κ_f). Prediction scores for the simple STRF model and the linear-nonlinear (LN) model are shown for comparison.

contrast-history-independent model, and 67.3 – 67.5% for the contrast-history-dependent model of Equations (6.2) – (6.4) (see Figure 6.7).

The constraint that the coefficients of κ_h must be positive was an overwhelmingly necessary one. When this was relaxed, and the coefficients of κ_h allowed to take on any value, then the model performance dropped to 64.2 – 69.4%. This indicates that there was considerable overfitting: the prediction scores were worse than when the history of contrast was ignored (Figure 6.8A, dark green bar).

6.4.2.2 Contrast time kernels: simplifications

As in the previous chapter, a secondary goal of this work is to develop simple approximations to contrast kernels that can be applied without requiring the time-consuming exploration of stimulus space attempted here. I therefore considered a number of simplifications to the contrast time kernel model.

Given the results of Section 6.2, and the approximately exponential shape of the population average of κ_h (Figure 6.6), I modelled the (unnormalised) time kernels by a simple exponential:

$$\lambda_h = e^{-h/\tau_H} \quad (6.6)$$

The median time constant fitted to the 77 units was $\tau_H = 117$ ms (Figure 6.8B). The model performed well compared with fitting a full contrast time kernel, with a score of 67.2% (Figure 6.8A, blue bar). I also considered whether, for the purposes of parsimony, a single time constant could be used for the whole population. By fixing τ_H at different values, I found that the most predictive model had $\tau_H = 85$ ms — in line with the findings in Figure 6.6 — with a prediction score of 67.1%. There was, however, a reasonably broad range of τ_H values between 80 – 120 ms which gave similarly respectable scores (Figure 6.8C).

When investigating the spectral integration underlying contrast gain changes in Section 5.4, I found that the contrast frequency kernels, κ_f , could be approximated, up to a normalisation constant, as the absolute value of the STRF frequency kernel, i.e. $\kappa_f \approx |k_f|$ (see Section 5.4.2.5). Given the usefulness of such an approximation, I considered whether the contrast time kernels, κ_h , could be similarly approximated as the absolute value of the STRF time kernel, i.e. $\kappa_h \approx |k_h|$. As a technical aside, the STRF kernels were estimated using $N_H = 10$ history bins (i.e. with a 250 ms memory), while the contrast time kernels were estimated using $N_H = 20$ (i.e. with a 500 ms memory). For an appropriate control, I fitted κ_h coefficients with $N_H = 10$, yielding a prediction score of 67.2%. When $\kappa_h = |k_h|$ was fixed, the prediction score was 67.1%. Thus the absolute value of the STRF provides an excellent approximation for the contrast kernel of a cortical neuron.

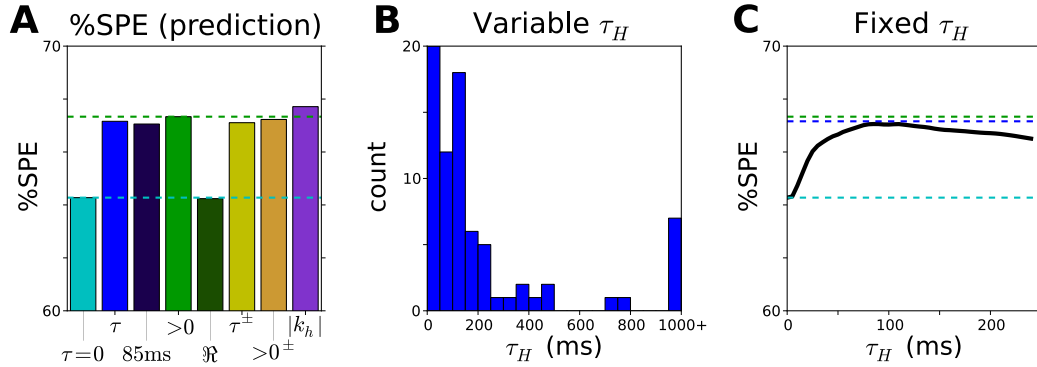


Figure 6.8: Comparison of assumptions for contrast time kernels. (A) Prediction scores for a range of models for the contrast time kernels. In order, from left to right, these models are the following. Cyan ($\tau = 0$): no history dependence, i.e. $\kappa_h = \delta_{h0}$. Blue (τ): Exponential model of Equation (6.6), as shown in (B). Dark blue (85 ms): As for the blue, except τ_H was fixed at 85 ms (see (C)). Light green (> 0): κ_h constrained to be positive. Dark green ($\Re$): κ_h allowed to take on any real value. Lime green ($\tau^\pm$): separate exponential kernels fitted for increases and decreases in contrast. Orange ($> 0^\pm$): separate positive-definite kernels fitted for increases and decreases in contrast. Purple ($|k_h|$): κ_h approximated as the absolute value of the STRF time kernel. Cyan and green dashed lines show the worst (no history dependence) and best (positive κ_h) performing models that were fitted. Note that the ordinate axis only starts at 60% to make the differences between the model scores more visible. (B) Fits of the time constant τ_H for the exponential model (Equation (6.6)) for all 77 units. The median time constant was 117 ms. (C) Prediction scores for the exponential model (Equation (6.6)), when τ_H was fixed, rather than fitted. Abscissa denotes the fixed value of τ_H , ordinate as in (A). The cyan and green horizontal dashed lines are as in (A), denoting, respectively, the model success when $\tau_H = 0$ (i.e. no history dependence), and the model success when a full (positive) contrast time kernel was fitted. Blue dashed line shows the model success when separate τ_H values were fitted for each unit.

Finally, I considered whether fitting separate time kernels for increases and decreases in contrast — as measured in the first experiment in this chapter — would improve the model. However, this just resulted in overfitting, and prediction scores declined. While using a single (positive) time kernel for each unit yielded a prediction score of 67.3%, having separate kernels for increases and decreases in contrast did not improve predictions, with a corresponding score of 67.2%. This trend was the same when a pair of exponential kernels was used, as in Equation (6.6): the score for using a single exponential kernel was 67.2%, compared with 67.1% for using two. Single contrast kernels are therefore

sufficient for capturing the time course of gain changes in these neurons.

6.5 Discussion

In this chapter, I characterised the time course by which changes in stimulus contrast affect the gain of auditory cortical neurons. The data from the experiments presented here demonstrate that gain changes are applied with a time constant on the order of 100 ms.

In the first experiment, gain changes were estimated by embedding a fixed test sound within a DRC of variable contrast. The strength of the neural response to the test sound was dependent on the time since the last switch in contrast. This estimated that changes in gain have an exponential shape, with a time constant of approximately 110 ms. The time course for increases and decreases in gain were found to be asymmetric, with time constants of 157 ms and 86 ms respectively (Figure 6.3).

In the second experiment, non-instantaneous gain changes were integrated into the models of Section 5.3. Such history-dependent contrast kernels significantly improved model performance. These, in turn, could be approximated reasonably well by an exponential time kernel, with a median (across units) time constant of 117 ms. In a similar manner to the contrast frequency kernels of Section 5.4, a unit's contrast time kernels could also be approximated well as the absolute value of the time component of its STRF (Figure 6.8).

6.5.1 Mechanisms for gain control

In the previous chapter (Section 5.5.2), I raised a number of potential implementations of gain control which could contribute towards the predominantly

local spectral integration. These included within-neuron (or across-neuron) mechanisms such as synaptic depression or intrinsic hyperpolarising currents, network mechanisms such as the balancing of excitation and inhibition, an inheritance of the effect from a gain control process earlier in the ascending auditory pathway, and the accumulation of gain changes from a cascade of processes operating across several successive stations.

The results of this chapter and the previous chapter demonstrate a close similarity between units' (linear) STRF kernels and their (divisive) contrast kernels. The similarity of these integration windows in both the spectral and temporal domains provide some further support for the hypothesis that local processes — such as synaptic or intracellular mechanisms — contribute towards gain effects. Synaptic depression can, in principle, operate on a range of different time scales [Varela et al., 1997; Carandini et al., 2002]. In vitro evidence suggests that synaptic depression in auditory cortex can last longer than 100 ms [Wehr and Zador, 2005], and that it shows a similar temporal asymmetry to that observed here (Figure 6.3C) [Dobrunz et al., 1997; Chung et al., 2002; Hosoya et al., 2005]. The time course (though not the asymmetry) is also consistent with the activation of intrinsic suppressive currents [Abolafia et al., 2010]. Nevertheless, the data presented in the past two chapters do not categorically rule out network phenomena, nor do they eliminate the possibility that a combination of local and network mechanisms contribute towards the observed effects. A thorough dissection of the origin of gain control is likely to require experimental paradigms with access to intracellular or synaptic dynamics, and/or activity across larger-scale ensembles of cortical neurons.

There is also insufficient evidence in this data to localise these effects to the cortex. Several studies have identified compensatory gain changes in IC [Kvale and Schreiner, 2004; Dean et al., 2005, 2008]. These effects have similar time

constants to those I observe in cortex, and show a similar asymmetry for increases and decreases in gain.

Finally, it is important to consider that the time course of gain changes is likely to be dependent on the temporal structure of the stimulus. From a computational perspective, if an ideal observer wished to estimate the temporal contrast of any discretely-updated stimulus, they need to integrate over a number of successive samples. In the DRCs used throughout this thesis, the chords changed in unison every 25 ms, so that the stimulus had a fixed temporal modulation frequency of 40 Hz. We might therefore expect that if stimuli with a slower temporal modulation frequency were used, gain changes would take longer to manifest. This is likely to be particularly relevant for natural sounds, which have a wide range of modulations [Singh and Theunissen, 2003; McDermott and Simoncelli, 2011]; here, the time scales of gain control may turn out to be dependent on an individual sound’s modulation spectrum. A similar inference argument explains that an ideal observer would take longer to detect a decrease in contrast than an increase [DeWeese and Zador, 1998], which provides some justification for the asymmetries observed here.

6.5.2 Implications for modelling

In the last several chapters, I have developed a more sophisticated model for the responses of auditory cortical neurons under stimulus conditions of varying contrasts. For the stimulus sets presented here, the predictions of a linear STRF model could explain 44% of the variance in the response of an idealised neuron. In turn, a linear-nonlinear model could explain 60%. Incorporating a contrast frequency kernel raised this value to 65%, while a contrast time-frequency kernel performed at 69% (Figure 6.7). The availability of simple approximations to the contrast kernels is likely to bolster the predictive power of existing models for auditory cortical neurons.

One difficulty in extending these models to other domains is knowing how to measure stimulus contrast. For the synthetic stimuli I used here, the contrast was specified by design; I therefore used the set of stimulus parameters as the input into the models. However, the contrast tensor, σ_{tfh} is not a well-defined quantity for an arbitrary sound. While an algorithm can be specified for estimating σ_{tfh} , the parameters of any neural model that follows would likely depend on the fine details of this algorithm. It is also worthwhile considering that the models advanced here are phenomenological in nature: while they posit that neurons are making measurements of the recent history of contrast in the same way that they measure the recent history of sound level, unlike for the latter, there is as yet no direct evidence for an explicit representation of σ_{tfh} in the auditory pathway.

6.5.3 Computational principles

In the last four chapters, we have observed a remarkable set of properties of ferret A1/AAF neurons. When the contrast of a DRC stimulus is decreased, neurons simply change their gain to partially compensate for the reduced contrast. Indeed, they do this predominantly by weighting the contrast within their own spectrotemporal integration windows. As a result, a cortical neuron's firing rate does not encode the *absolute* changes to sound level that occur within its STRF, but rather the extent of these changes along a *relative* scale, which is partially normalised to the sound statistics within that window.

This has strong implications for decoding the output of neurons' spike trains. It makes it difficult to infer, from the firing rates of auditory cortical neurons, the spectrogram of the stimulus which evoked these responses. This can be seen, for example, in Figure 6.2: when a single, fixed chord is presented in different contrast contexts, these neurons can produce responses of very different magnitudes. Without conditioning on the context, these firing rates might

therefore yield little information by which we could determine, for example, what the magnitude (or spectral contrast) of the probe chord was. The same problem would arise if we were to try to decode the spectrogram of a DRC from the ongoing firing rate of one, or even many, primary auditory cortical neurons. If the firing rate of some neuron at some time is large, then we have evidence that the stimulus recently had energy along the dimension specified by that neuron's STRF. But if we consider the output nonlinearities in Figure 5.11, inverting a given firing rate, y , to find the magnitude of the projection of the stimulus onto that dimension, z , gives different answers in different contexts.

It is fair to conclude from this that primary auditory cortex is not representing absolute changes in sound level, and is not trying to recreate the spectrogram. This is unsurprising: given the wealth of adaptive processes that have been described in the peripheral and central auditory pathway, it would be somewhat bizarre to see such sophisticated, dynamic adaptation undone at this stage.

Rather, what we see suggests that primary auditory cortex, as a whole, is attempting to “integrate out” the actual contrast of a sound, and produce a representation of the stimulus that is somewhat invariant to its contrast. Since each neuron merely scales its gain according to the spectrotemporally-local statistics, it may be the case that if we presented a low contrast version of a sound, the global pattern of spiking across A1/AAF would be very similar to what would be produced in response to a high contrast version of the same sound. This may prove to be important in recognising the same auditory “objects” under different stimulus conditions.

In order to explore this idea further, we shall now turn to a comparison between the effects of stimulus contrast on cortical neurons and its effects lower down in the auditory brainstem, at the inferior colliculus. By doing so, we can frame these questions through the ascending auditory pathway and ask: is

gain normalisation, and is this apparent contrast-invariance, a special property of auditory cortex?

Chapter 7

Contrast gain control in the inferior colliculus?

7.1 Introduction

In Chapter 4, we saw that when the contrast of a sound stimulus is reduced, the gain of neurons in the primary auditory cortex of ferrets is increased. As a result of this, their firing rate is more sensitive to smaller changes in sound level than would be the case under high contrast stimulation. In the two chapters that followed, I characterised the spectrotemporal integration that informs these gain changes.

These results, however, are somewhat agnostic as to where the gain changes are actually produced. In particular, they leave open the possibility that the gain changes are not generated within the cortex, but may rather be inherited from earlier stations in the ascending auditory pathway.

As discussed in Chapter 1, there is a large body of work characterising a rich set of adaptive processes in the auditory periphery, brainstem and midbrain. In experiments where the contrast of simple stimuli has been manipulated, neurons in the auditory nerve [Joris and Yin, 1992] and the inferior colliculus [Escabi

[et al., 2003](#); [Kvale and Schreiner, 2004](#); [Dean et al., 2005](#)] have been shown to undergo contrast-dependent changes in coding. However, these studies used different animals, presented different stimuli, and analysed the resulting data in different ways. This make it difficult to quantitatively compare the effects in different stations on the basis of existing literature.

I therefore aimed to explicitly test the hypothesis above — that contrast-dependent changes in neural coding observed in primary auditory cortex are inherited from earlier stations — by recording from neurons in the central nucleus of the inferior colliculus (IC) of ferrets.

The choice of the inferior colliculus was motivated by two considerations. First, the IC is widely considered to be a highly important subcortical integration centre of the auditory pathway. The divergent processing streams of the auditory brainstem converge at the IC, while its efferents provide the principal ascending input to the thalamus. These, in turn, project to the auditory cortex [[Pollak et al., 2003](#); [Xiong et al., 2009](#)]. Thus the majority of auditory information from the cochlea passes through the IC en route to A1/AAF [[Casseday et al., 2002](#)]. Its central role in integrating information has led some authors to dub the IC the “V1 of the auditory pathway” [[Nelken et al., 2003](#)]. The computational strategies enlisted by the mammalian auditory system to cope with sounds of a variety of contrasts are thus likely to involve the inferior colliculus at some stage.

The second consideration for investigating the IC is that other authors have already found that the coding of stimulus features by IC neurons is sensitive to the statistics of recent stimulation [[Escabi et al., 2003](#); [Kvale and Schreiner, 2004](#); [Dean et al., 2005, 2008](#); [Dahmen et al., 2010](#)]. In turn, there is evidence to suggest that at least some of the adaptive shifts in IC coding cannot be explained by the weaker adaptation in the auditory periphery [[Wen et al., 2009](#)].

At the very least, these previous studies guarantee that there will be some contrast-dependent changes in the IC to compare with cortex. At the most, such changes in the IC may explain away the contrast gain control observed in cortex.

To test the hypothesis that contrast-dependent cortical gain changes originate in the IC, I presented the same set of RC-DRC stimuli as in Sections 5.4 and 6.4, and fitted the same set of contrast kernel models developed in Sections 5.3 and 6.3 to the responses of IC neurons. Not only did this facilitate a characterisation of how the response properties of IC neurons depend on contrast, but it also allowed a direct comparison between the effects observed at this level and in cortex. If the gain control observed in cortex is indeed inherited from the IC (or lower stations), we should expect to find the same broad changes in coding, potentially with the same kind of spectral and temporal sensitivity, between the IC and A1/AAF.

7.2 Methods

7.2.1 Animals and stimuli

Data in this experiment were obtained from two ferrets (one male, one female). A full description of the anaesthesia, surgical, stimulation and recording procedures is provided in Section 2.3.

The stimuli presented here consist of the RC-DRC stimuli, as used in Sections 5.4 and 6.4.

7.2.2 Models

The models presented in this section are the same as those in the previous two chapters.

Units were chosen by the same selection criteria described in Section 5.4.1.3. As described below in Section 7.3, a small number of units had non-monotonic output nonlinearities. Since these were a minority, and only a small number were available, such units were excluded from analysis on the basis of insufficient data.

Non-monotonicity was assessed by fitting a difference-of-sigmoids parametric form to the binned data (see Section 4.2.5). This took the form:

$$\hat{y}_t = a + \frac{b_1}{1 + \exp(-(z_t - c_1)/d_1)} - \frac{b_2}{1 + \exp(-(z_t - c_2)/d_2)} \quad (7.1)$$

This was fitted with the constraints that all of $a, b_1, b_2, d_1, d_2 > 0$, and $c_2 > c_1$. The fit to a non-monotonic unit is shown in Figure 7.2C. The first, positive sigmoid term captures the increasing segment of the curve, and the second, negative sigmoid term captures the decreasing segment of the curve. Those units for which the prediction scores for this double-sigmoid nonlinearity were at least 2% better than for the single-sigmoid nonlinearity were labelled non-monotonic.

7.2.3 Fisher information

As a means of summarising the effect that contrast-dependent changes to output nonlinearities have on neural coding, I calculated how the stimulus which maximised Fisher information, z^* (as given in Equation (4.4)) would change with contrast. This involved rendering each of the parameters $\theta \in \{a, b, c, d\}$ in

that equation as functions of contrast, via either the *cd*-model, or the *ab/cd*-model.

If we drop the time index t , and consider the spectral profile of contrast as σ_f , then the optimal stimulus, z^* , is a function of σ_f . Rather than exploring this as a multi-dimensional problem, I considered how z^* would change if the contrast were changed uniformly across frequency, such that $\sigma_f = \sigma$. I therefore calculated the percentage change in z^* in the all-low-contrast condition, from uniformly increasing σ , i.e. $\frac{1}{z^*(\sigma=0)} \times \left[\frac{\partial z^*}{\partial \sigma} \right]_{\sigma=0} \times 100\%$.

7.3 Results

In this experiment, I directly compared the response properties of neurons in the central nucleus of the inferior colliculus (IC) with those of neurons in the primary auditory cortex (A1/AAF). The focus of this comparison was on how the coding properties of these neurons are affected by contrast.

To do this, I presented the identical stimuli as in Sections 5.4 and 6.4 to two anaesthetised ferrets, while recording from neurons in their IC. The central nucleus of the IC was located by the existence of the characteristic dorso-ventral tonotopic gradient, as shown in Figure 7.1. I identified 171 units within the central nucleus, of which 137 responded reliably to the Random Contrast DRC (RC-DRC) stimuli.

Like cortical units, the responses of the IC units could be described well using STRF and LN models. This is in line with previous data from the IC [Escabi et al., 2003; Lesica and Grothe, 2008]. The output nonlinearities for most units could be aptly fitted with a sigmoid function; however, 8 of the 137 units showed strong non-monotonicities (e.g. Figure 7.2C). Since these only

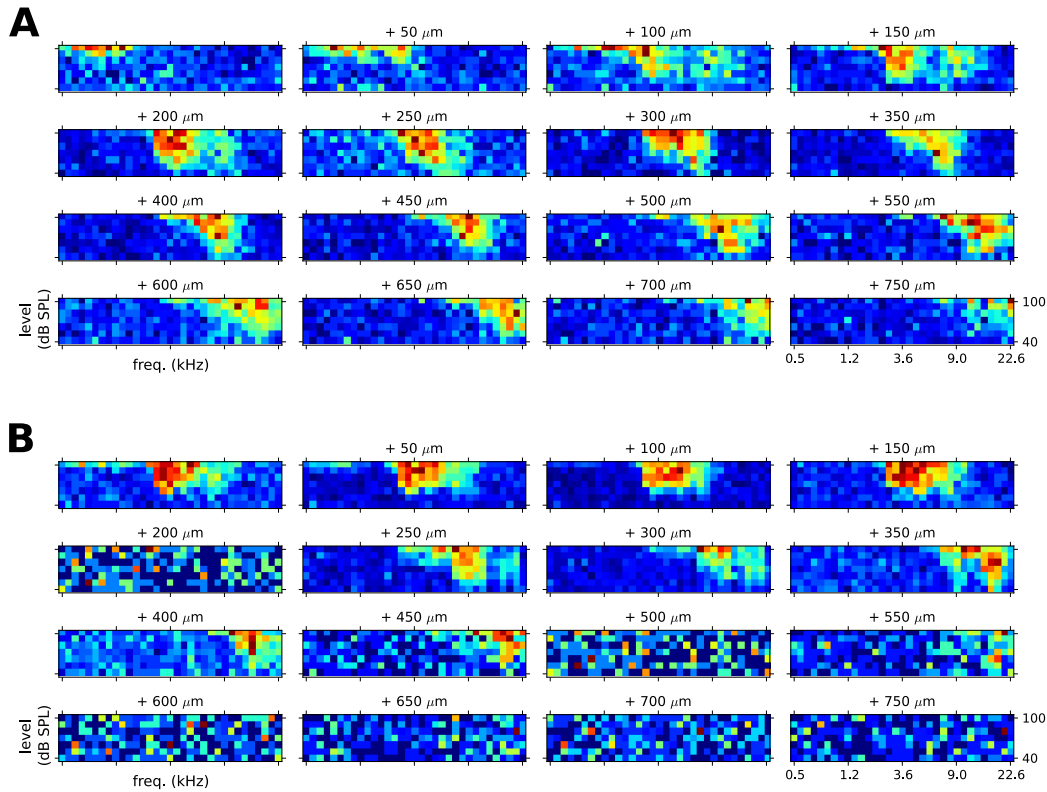


Figure 7.1: The central nucleus of the inferior colliculus was identified by the presence of a characteristic dorso-ventral tonotopic gradient. Each of (A) and (B) show a different electrode penetration. As the probe was lowered in to the IC, I measured the responses of neurons to pure tones of 50 ms duration. These tones had frequencies of same range as used in the DRC, and levels spanning from 40 to 100 dB SPL. Each colour image shows the magnitude of the average spiking response to each tone at a given electrode site, as a function of tone frequency (abscissa) and tone level (ordinate). Red denotes a large firing rate, blue a low. Each electrode site is labelled by the distance from the most dorsally-located electrode; thus the plots show responses along a dorsal (top left) to ventral (bottom right) gradient. In (A), there a clear dorsoventral progression in best frequency, indicating that the electrode was placed within the central nucleus. In (B), this progression only extended as far as the 450 μm position (with a dud site at 200 μm). For this latter electrode penetration, data from sites deeper than the 450 μm position were excluded from analysis as it could not be determined whether these sites were in the central nucleus.

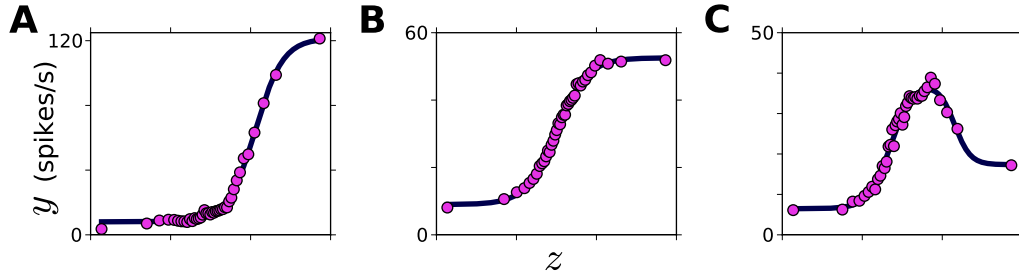


Figure 7.2: Output nonlinearities for three example units in IC. As in Figure 4.1, the abscissa shows the output of the linear model, z , i.e. the projection of the stimulus onto the STRF. On the ordinate is the actual spike rate, y . Magenta circles indicate binned data for the respective units. Dark blue lines show parameterised nonlinearities for the three units: a single sigmoid for (A) and (B), a difference of sigmoids for (C).

constituted a small minority of the units recorded, they were excluded from further analysis.

7.3.1 How contrast affects IC nonlinearities

The models of Section 5.4, which measure how contrast affects the output nonlinearities of cortical units, were motivated by two observations in Chapter 4. First, contrast does not affect the spectrotemporal preferences of A1/AAF units. Similar results have previously been demonstrated in the IC by Escabi et al. [2003]. Second, the main effect of increasing contrast is to reduce the gain of cortical units, by dilating the output nonlinearity. This is accompanied by a small (horizontal) shift in threshold. Shifts in coding have also been reported in the IC as a result of changing contrast [Dean et al., 2005]. This suggests that contrast kernel models may also serve a useful purpose for IC data.

It is useful to return to Figure 5.11 to visualise the cortical results. This figure shows the output nonlinearities for 8 example units in primary auditory cortex, during the all-high-contrast segment (red) and the all-low-contrast segment (blue). There was a highly stereotyped pattern to contrast-dependent shifts in

nonlinearities for cortical units, characterised by a dilation along the horizontal axis. The minimum firing rates did not typically change. Given that the neurons did not often saturate their firing rate, there was no available evidence for changes in maximum firing rate with contrast. The major variation across units lay in the extent of gain changes (Figure 5.15).

When I fitted the same models to IC units, there was no such stereotypy. Figure 7.3 shows the output nonlinearities for 15 example units recorded in IC, in all-high and all-low contrast conditions. Some units, such as Figure 7.3A-C, show the same kind of simple gain changes as observed in A1/AAF units. However, the majority showed more complex patterns. Many units showed dramatic changes in the shape of their output nonlinearity, such as changes in the saturating firing rate (e.g. Figure 7.3D-F), or large shifts in threshold (e.g. Figure 7.3H-L). Some units were entirely insensitive to contrast (e.g. Figure 7.3M-N). Finally, a small number of neurons showed non-monotonic stimulus-response curves in some, but not all conditions (e.g. Figure 7.3O). As mentioned above, there were too few units in this latter group to make quantitative statements about non-monotonicity, so these were excluded from further analysis.

Thus, compared with primary auditory cortex, units in inferior colliculus are relatively heterogeneous in the way their coding depends on contrast.

7.3.2 Gain contrast kernels

The remaining results of this chapter proceed in parallel to the model fitting of Sections 5.4 and 6.4, using the models of Sections 5.3 and 6.3.

In spite of the variety of effects that contrast had on output nonlinearities, the LN models were generally improved by simply including gain changes via the

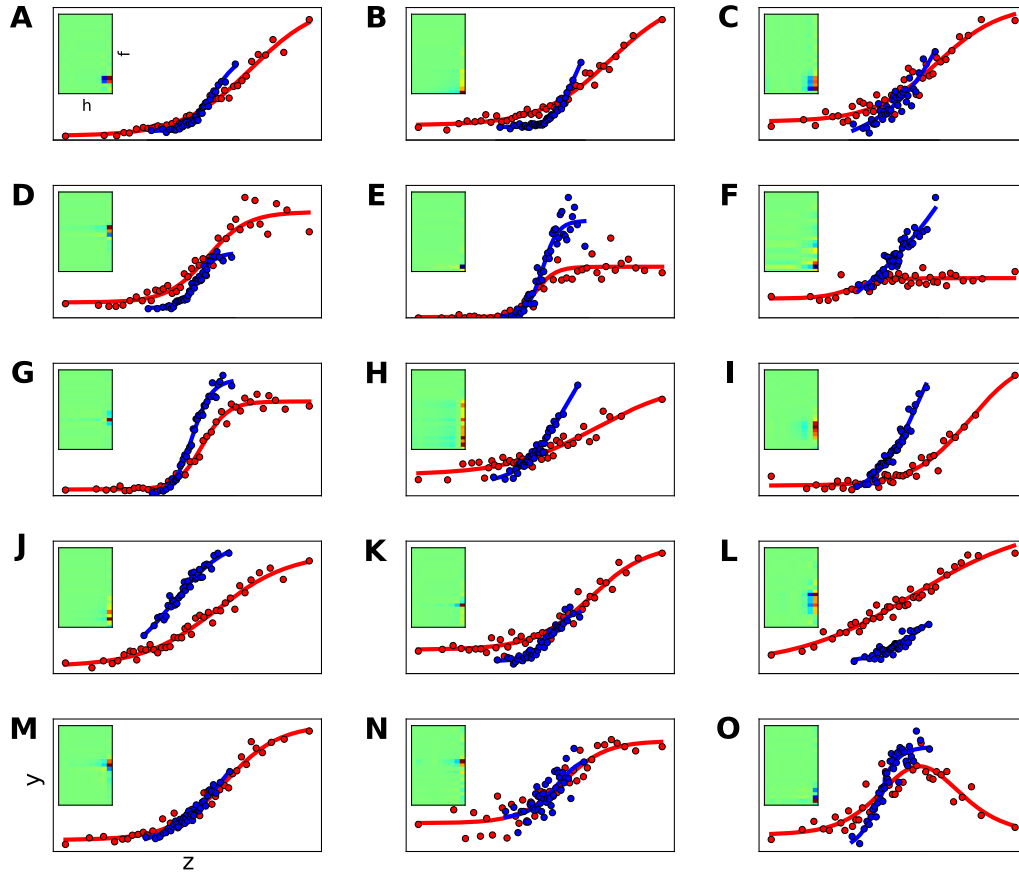


Figure 7.3: Output nonlinearities in low and high contrast: IC. Data shown for 15 example units in the central nucleus of the inferior colliculus, in an identical fashion to Figure 5.11. Unlike in primary auditory cortex, changing contrast does not produce stereotyped changes in the coding properties of IC neurons.

cd-model of Section 5.4.2.3. While the STRF model had scores of 48.8% – 49.3%, and the linear-nonlinear model had scores of 61.6% – 61.9%, the *cd*-model increased these to 65.4% – 65.6%. This improvement was less than what was observed in cortex: while, for the population of A1/AAF units, including a shared contrast for kernel for *c* and *d* reduced the error variance of the LN model by 23%, this value was only 11% for the population of IC units.

IC units did show, on average, contrast-dependent gain changes, but these were weaker than those in cortex. I measured gain changes under this model via $G_d = d_{high}/d_{low}$. The median G_d in IC was 1.59, compared with 1.92 in

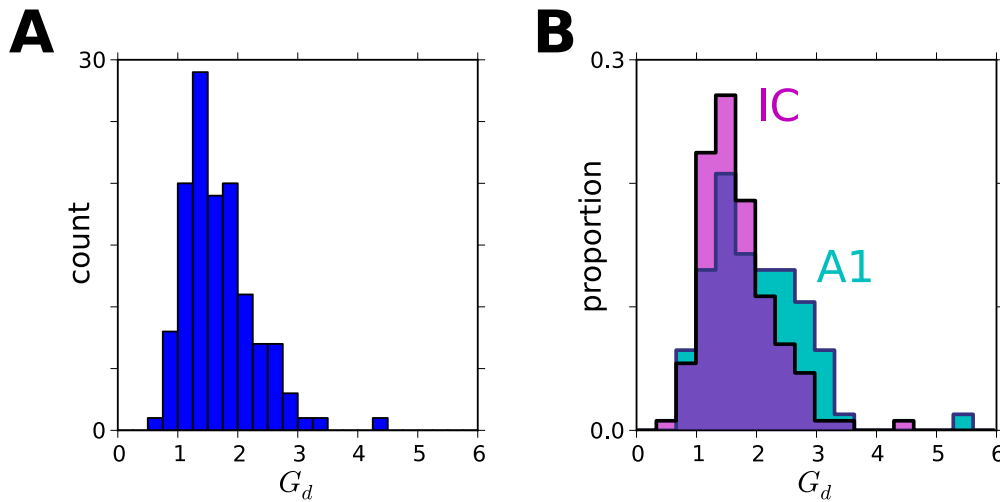


Figure 7.4: Gain model: statistics on contrast-dependent gain changes. (A) The majority of units in the IC decreased their gain as contrast was increased. As in Figure 5.15, this is measured by the ratio $G_d = d_{high}/d_{low}$. (B) Comparison between G_d values for the population of IC units (magenta), and the population of A1/AAF units (cyan) measured in Section 5.4. In this and later figures, the label “A1” refers to combined data from both A1/AAF, as throughout this thesis. Purple shows region of overlap. On average, cortical units experience greater gain control than IC units.

A1/AAF (rank-sum, $p < 0.005$; Figure 7.4).

This raises the possibility that there is additional gain control between the inferior colliculus and A1/AAF. Given the nonlinear nature of the models fitted here, it is necessary to check a simpler, alternative hypothesis: whether summing the output of lower gain neurons yields a higher gain output. This, however, does not appear to be the case. I tested the possibility by taking random subsets of three to five IC units, summing their firing rates, and fitting the *cd*-model to the resultant multi-unit clusters. The population of clusters had a similar distribution of G_d values to the individual source units. In a similar manner, I simulated noiseless IC units, and constructed clusters from their simulated responses. Within each cluster, the simulated units could either have overlapping or non-overlapping STRFs. Again, simulated clusters had the same range of G_d values as the source units. The greater gain in A1/AAF therefore cannot be explained by a non-selective aggregation of responses from the IC.

Like the nonlinearities, there was more variation in the shape of gain kernels in IC than in A1/AAF. Gain contrast kernels are shown for 10 representative units in Figure 7.5. Some units, as in Figure 7.5A-E, showed roughly similar patterns to cortex: gain kernels for these units roughly matched the magnitude of the STRF kernel, covering both the excitatory centre of the receptive field, and the inhibitory sidebands. However, there were coefficients of κ_f that were strongly, and significantly, negative (e.g. Figure 7.5F-H), a property not generally found in the gain kernels of cortical neurons. For these units, presenting high contrast in those bands led to an *increase* in gain. Finally, for some units, there was a pronounced mismatch between the gain and STRF kernels (e.g. Figure 7.5I, J; also C, G).

Because of these latter two properties of IC gain kernels, the approximation $\kappa_f \approx |k_f|$ was not as useful for the IC units as it was for the cortical units. While, for A1/AAF, there was a good correlation between the two variables ($r(\kappa_f, |k_f|) = 0.80$), which was even greater when κ_f was constrained to be positive ($r = 0.93$), these correlations were weaker for the IC data ($r = 0.60$ and $r = 0.66$ respectively; Figure 7.6). Applying the positive-definite constraint on κ_f also reduced the prediction scores of the model (from 65.4% – 65.6% to 65.2% – 65.3%); with the approximation $\kappa_f \approx |k_f|$, this was reduced to 64.0% – 64.3% (Figure 7.7; compare with Figure 5.17).

In summary, gain control was weaker in the IC than in A1/AAF, and often depended on a different balance of spectral integration than seen in the auditory cortex.

7.3.3 Multiple contrast kernels

Figures 5.11 and 7.3 demonstrate that IC units undergo a greater range of changes in shape than cortical units as a result of varying the stimulus contrast.

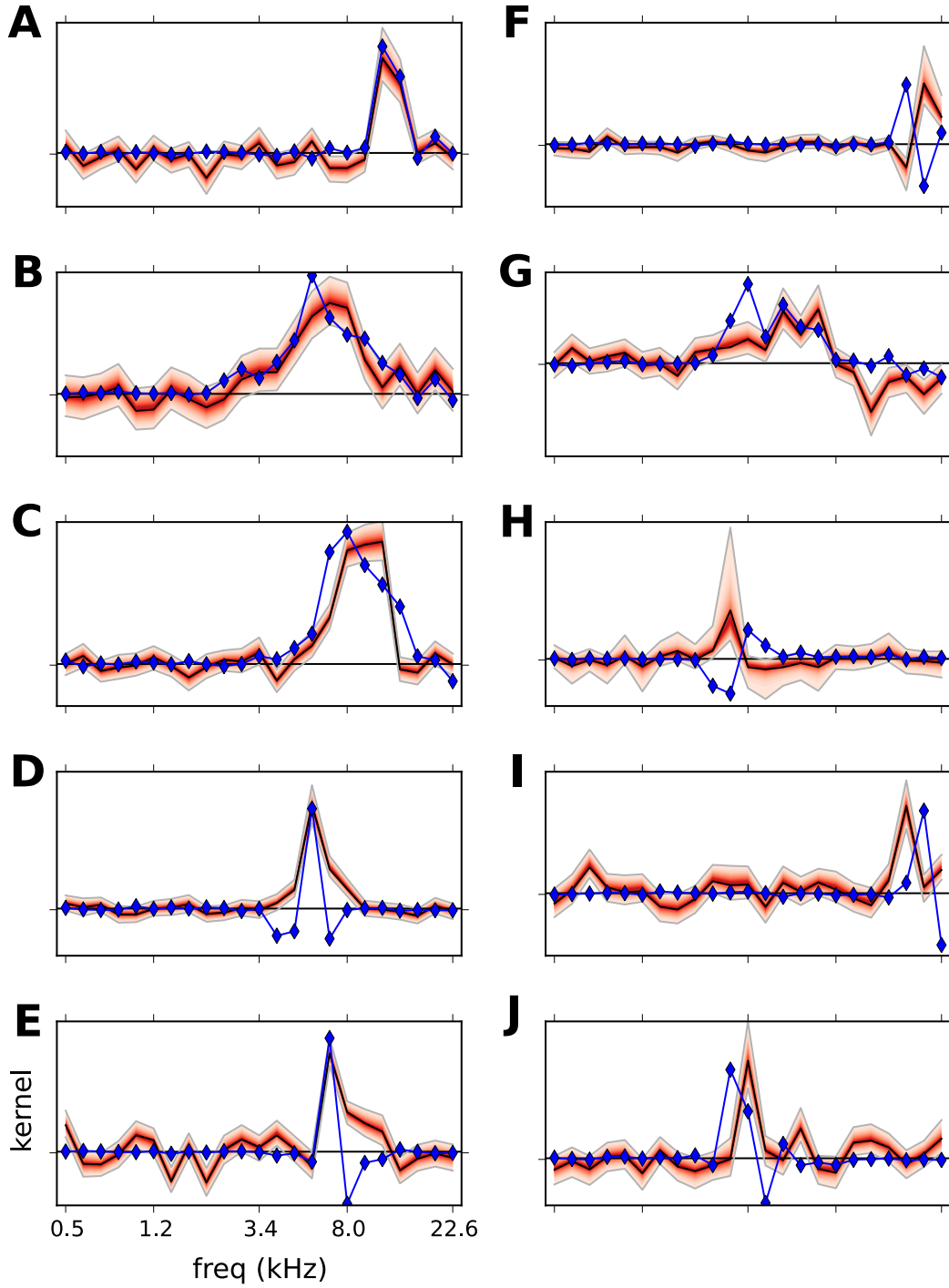


Figure 7.5: Gain contrast kernels, for 10 example units in the IC. This figure parallels the cortical data shown in Figure 5.12. As in that figure, the black line and red area show fits of Equation (5.20), with contrast-independent parameters a and b , and a shared, real-valued contrast kernel, κ_f for parameters c and d . The blue line and diamonds show the frequency component of the linear, separable STRF, k_f .

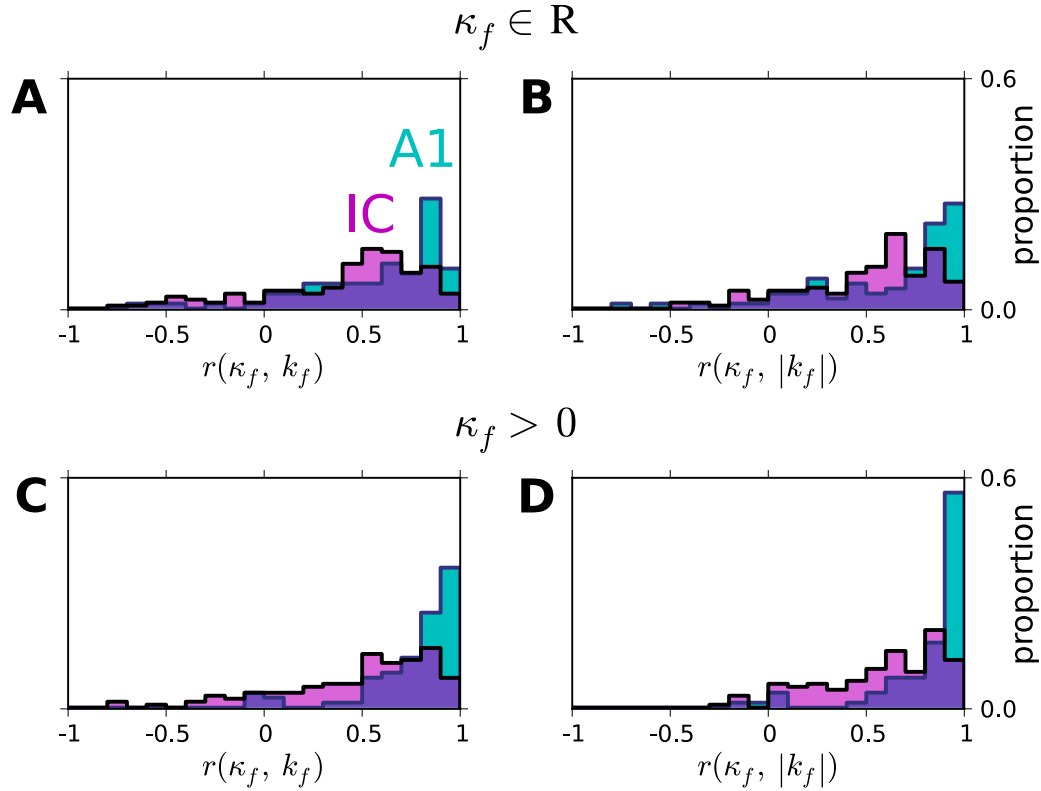


Figure 7.6: Correlation coefficients (r) between the coefficients of the linear STRF kernel (k_f) and those of the gain contrast kernel (κ_f), on a unit-by-unit basis. This parallels Figure 5.13. IC units are shown in magenta, A1/AAF units shown in cyan. (A-B): when coefficients of κ_f were unconstrained (i.e. real valued). (C-D): when these coefficients were constrained to be positive. While the contrast frequency kernels of cortical units were aligned with their STRF kernels, this was not generally the case for IC units.

As a result, unlike in cortex, prediction scores for IC units were generally improved when more general classes of contrast kernel models, beyond the *cd*-model, were used (Figure 7.8).

From this dataset, a couple of trends are worth mentioning. First, the actual predictive power of these models is the result of a trade-off between model power and overfitting. This justifies why, for instance, the *a/cd*-model outperforms the *a/c/d*-model. Second, merely making all the parameters contrast-dependent, with a shared kernel (the *abcd*-model), provided little improvement over the simple gain model (*cd*) that was explored in the section above. The

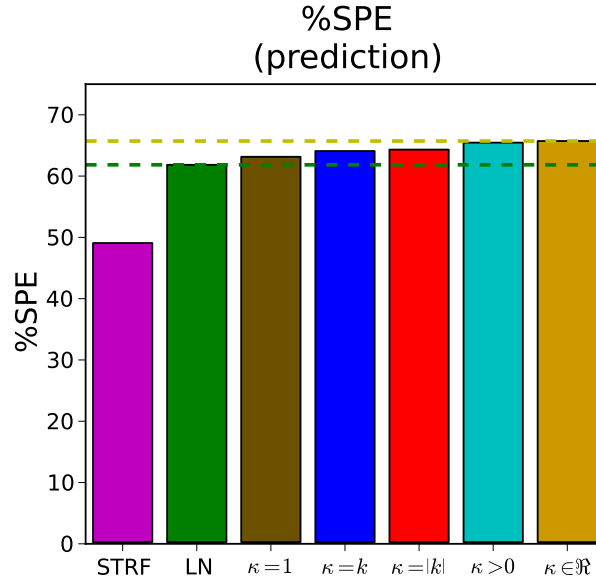


Figure 7.7: Model predictive power: Gain model with constrained coefficients in the IC. This figure parallels the cortical data in Figure 5.17. As for A1/AAF, the best predictive power is obtained when the contrast kernel coefficients are unconstrained ($\kappa \in \mathbb{R}$; yellow). Restricting coefficients to be positive ($\kappa > 0$; cyan) yields almost as good prediction scores. In cortex, the strong correlation between κ_f and $|k_f|$ (Figure 7.6) meant that the latter provided a good approximation to the gain kernel; here, the assumption ($\kappa = |k|$; red) was less useful. Other models which performed even worse assumed that the contrast kernel is the STRF frequency kernel ($\kappa = k$; blue), or assumed that it is constant with respect to frequency ($\kappa = 1$; brown). These still outperform the simple linear-nonlinear (LN) model (green). Yellow dashed line shows the model performance for the unconstrained coefficients, green dashed line for the LN model.

best performing model, ab/cd , had two separate contrast kernels. There are two potential interpretations for its success. One is that there might be different mechanisms subserving changes to the dynamic range of an IC neuron (via the minimum firing rate a , and the maximum firing rate $a + b$) and changes to the gain of an IC neuron (via d , and as a result, c), which, in turn, can have different spectral sensitivities. The other is that the changes to output nonlinearities were, themselves, rather nonlinear. In either case, this diverges from what is seen in cortex.

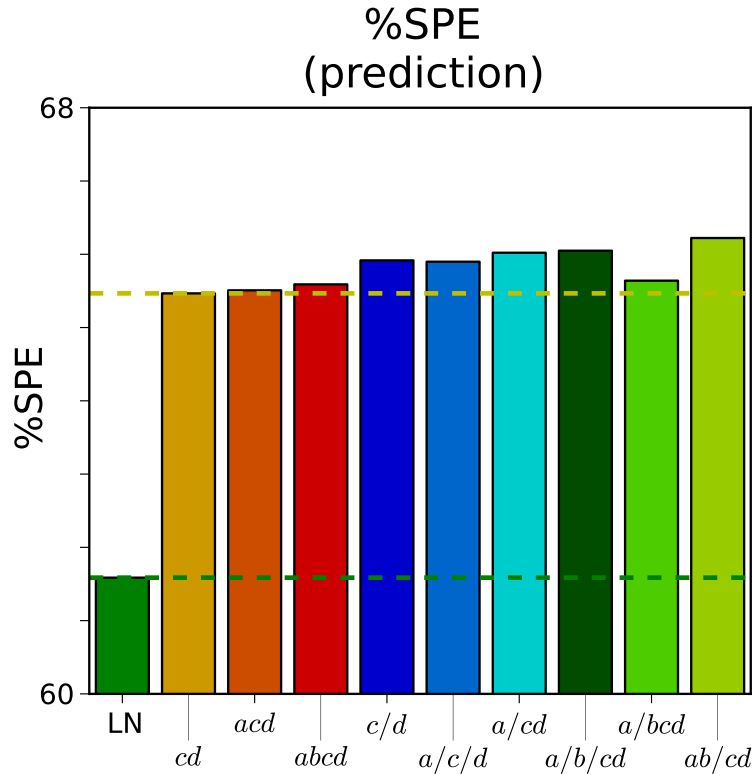


Figure 7.8: Model predictive power: Contrast kernels for multiple nonlinearity parameters in the IC. Prediction scores shown for various permutations of the models of Equations (5.10) – (5.14). Model names as in Section 5.3.2.4: for each model, the listed letters are contrast-dependent, while the omitted letters are contrast-independent; where letters are grouped together, a shared contrast kernel was fitted; where separated by a slash, separate contrast kernels were fitted. Thus, for example, a/cd means that b was contrast-independent, and $\kappa_f^{(a)} \neq \kappa_f^{(c)} = \kappa_f^{(d)}$. Dashed lines as in Figure 7.7.

7.3.4 Coding accuracy

When neurons' output nonlinearities change, they become more informative about some stimuli, and less informative about others. In Section 4.3.4.1, I showed that when contrast is reduced, the gain changes induced in cortical neurons make them more informative about more subtle changes in sound level within their receptive field. Thus, a cortical unit's optimal stimulus, z^* , measured as the sound which maximises the Fisher information in the unit's output nonlinearity, shifts as a function of contrast.

Given the more complex changes in the output nonlinearities of IC neurons, we might ask whether their stimulus-informativeness also shifts with stimulus contrast. Adaptation of IC neurons' coding efficiency to the shape of stimulus distributions has been previously reported by [Dean et al. \[2005\]](#); we should therefore expect to see such an effect here with contrast.

I therefore calculated, on the basis of the contrast kernel models, how the optimal stimulus, z^* , for each unit changes with stimulus contrast. Details of this calculation are provided in Section 7.2.3; the results are shown in Figure 7.9. Like A1/AAF units, IC units also, on average, changed their output nonlinearities to become more informative about smaller-magnitude excursions of the stimulus along the direction of their STRFs. However, there was much more variance in this effect across the population of IC units, and the total effect was much weaker than in cortex: the change for cortical units was a median 3.4–3.7 times greater than that for IC units (rank-sum, $P \ll 0.001$). This difference was seen regardless of which model was used to calculate changes in z^* .

Thus, while neurons in the inferior colliculus change their coding to become more informative about the stimulus distribution, the shift in coding is far more profound in the primary areas of the auditory cortex.

7.3.5 Cross-frequency interactions

In section 5.4.2.7, I measured whether a single band of high contrast would affect the gain of a cortical neuron's coding of *all* sound frequencies *equally*, or whether it would only affect the neuron's gain in a localised region of frequency space nearby to the source band. To do this, I compared three different models: the local effect model (gain changes only for sounds of that high contrast band), the global effect model (gain changes for sounds of all bands), and the regional

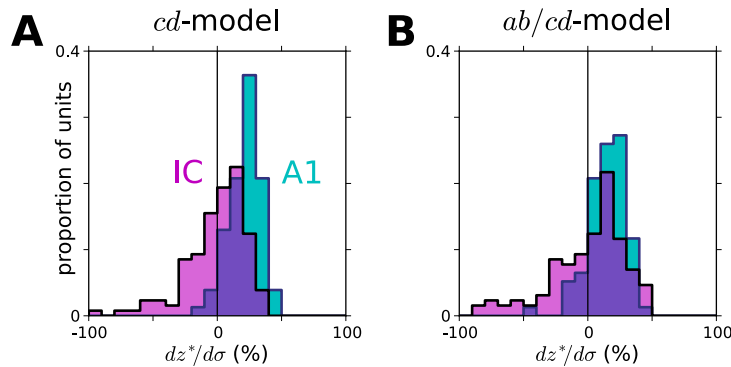


Figure 7.9: Contrast-dependent changes in Fisher information: IC vs. A1/AAF. (A) As shown in Figure 4.10, the stimulus for which an A1/AAF unit is most informative, z^* , changes with contrast. Histogram shows the direction and magnitude of changes to z^* , as the contrast, σ , is increased. Positive values indicate that when contrast increases, the unit becomes more informative about stimuli of higher contrast, i.e. of larger z . These have been calculated using the *cd*-model of Equation (5.20). While A1/AAF units generally increased their z^* with increasing contrast, a wider range of coding changes were exhibited by IC units. The average change in z^* was 3.4 times greater in A1/AAF than in IC. (B) As for (A), except where all nonlinearity parameters were contrast dependent, via the *ab/cd*-model. The average change in z^* was 3.7 times greater in A1/AAF than in IC.

effect model (gain changes over an effective range of frequencies near the high contrast band).

For cortical neurons, the global effect model had the best prediction scores. The effective scale of the gain spread was therefore large. IC neurons, on the other hand, showed the opposite pattern: the local effect model slightly outperformed the global effect model (65.0% – 66.2% vs 64.6% – 65.8%). The regional model, in turn, produced the best scores (65.7% – 66.9%) with a median effective scale of 1/4 octave (Figure 7.10).

7.3.6 Gain time kernels

While the data above show that neurons in the IC do undergo some contrast-dependent gain changes, they suggest that what is observed in A1/AAF is not inherited purely from the IC.

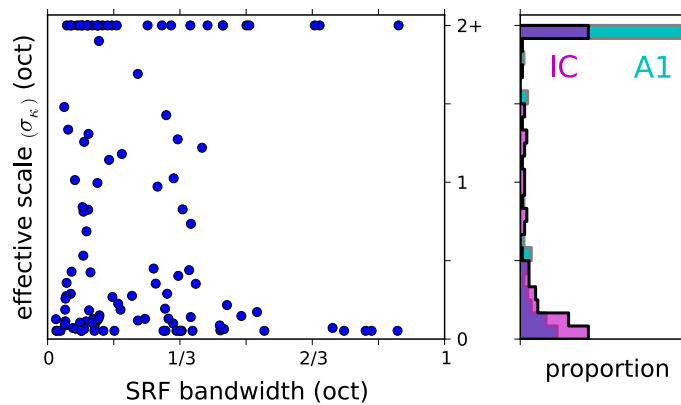


Figure 7.10: Effective scale of the gain effect for IC units. This figure parallels Figure 5.18. Estimates of how the the gain effect spreads from the contrast in band g are measured by the parameter, σ_κ . The fitted values of σ_κ are plotted on the left as a function of the bandwidth of k_f , the spectral receptive field, and on the right as a marginal histogram. There was no systematic relationship between SRF bandwidth and σ_κ . Gain effects were more localised in IC than in A1/AAF.

A crucial test of this is to consider whether the time courses of contrast-dependent changes to coding in the two stations differ. Indeed, I found that the time course of gain changes in the IC was significantly shorter than that observed in cortex. Like cortex, the contrast time kernels for individual units were roughly exponential in shape (e.g. Figure 7.11; c.f. Figure 6.5). When averaged over the respective populations of units, the difference between IC and A1/AAF was striking (Figure 7.12). While the time constant for the population average was 86 ms for the cortex, the corresponding time constant was 28 ms for the IC.

Given that the gain changes typically occurred over a very short period of time, including the time kernel for gain changes did not improve model predictive power as much as it did in cortex. With prediction scores of 49.2% for the STRF model, 62.1% for the linear-nonlinear model, and 65.0% for the contrast-history-independent model, the contrast-history-dependent model gave 65.6%. An exponential kernel with a fixed time constant of 45 ms performed at 65.2%. The best model, as was the case in cortex, was to use the absolute value of the

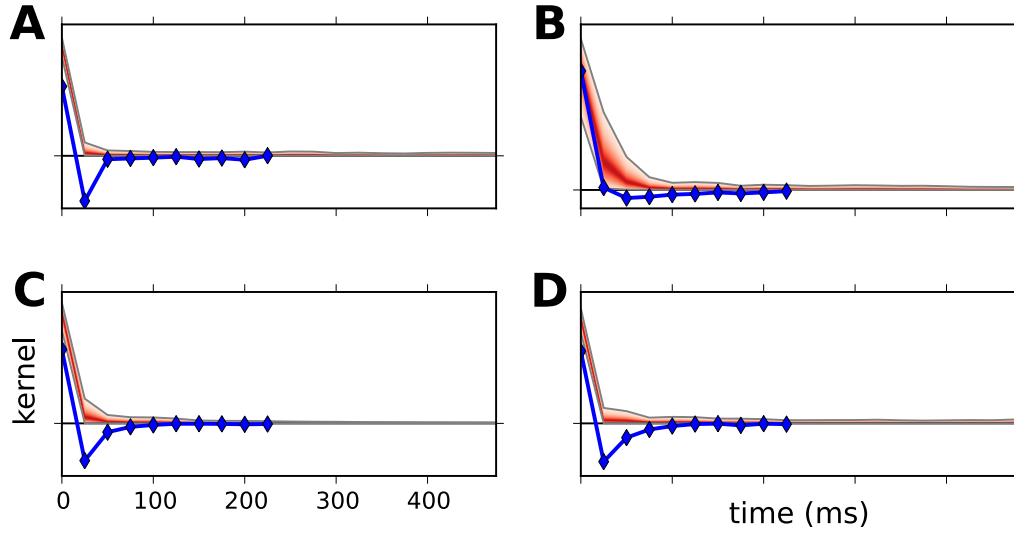


Figure 7.11: Contrast time kernels, for 4 example units in IC. These are fits of Equations (6.2) – (6.4). This figure parallels Figure 6.5.

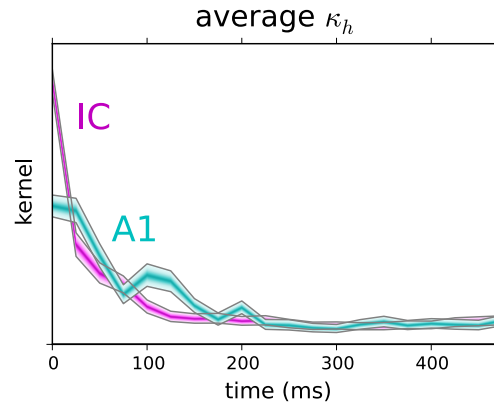


Figure 7.12: Mean of the contrast time kernels: IC vs A1/AAF. A1/AAF data averaged from the 77 cortical units (from Figure 6.6), IC data averaged from 129 units. These have approximately exponential shapes, with fitted time constants of 86 ms (A1/AAF), and 28 ms (IC).

time kernel; the prediction score here was 65.7%, reflecting reduced overfitting.

7.4 Discussion

In this chapter, I asked whether the contrast-dependent changes in neural coding in the ferret primary auditory cortical areas (A1/AAF) are inherited from

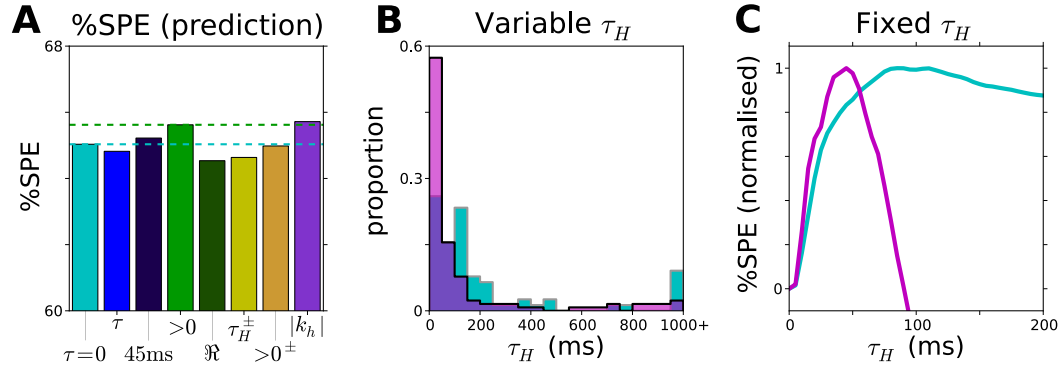


Figure 7.13: Comparison of assumptions for contrast time kernels: IC. This parallels Figure 6.8. (A) Prediction scores for a range of models for the contrast time kernels. In order, from left to right, these models are the following. Cyan ($\tau = 0$): no history dependence. Blue (τ): Exponential model of Equation (6.6), as shown in (B). Dark blue (45 ms): As for the blue, except τ_H was fixed at 45 ms (see (C)). Light green (> 0): κ_h constrained to be positive. Dark green ($\Re$): κ_h allowed to take on any real value. Yellow ($\tau^\pm$): separate exponential kernels fitted for increases and decreases in contrast. Orange ($> 0^\pm$): separate positive-definite kernels fitted for increases and decreases in contrast. Purple ($|k_h|$): κ_h approximated as the absolute value of the STRF time kernel. Cyan and green dashed lines show the worst (no history dependence) and best (positive κ_h) performing models. Note that the ordinate axis only starts at 60% to make the differences between the model scores more visible. (B) Fits of the time constant τ_H for the exponential model (Equation (6.6)), for the 129 IC units (magenta) and 77 cortical units (cyan). The median time constants were 35 ms for IC, and 117 ms for cortex. (C) Prediction scores for the exponential model (Equation (6.6)), when τ_H was fixed, rather than fitted, for IC (magenta) and cortex (cyan). Abscissa denotes the fixed value of τ_H . The ordinate has been normalised so that the prediction score for no history dependence ($\tau_H = 0$) is 0, and the maximum prediction score is 1. The best fixed time constants were 45 ms for IC, and 85 ms for cortex.

the inferior colliculus (IC). I therefore recorded the responses of IC neurons to the same set of RC-DRC stimuli of Sections 5.4 and 6.4. This allowed for a direct comparison with the cortical data. The results show that there are numerous points of difference between the way IC and A1/AAF neurons change their coding for low contrast and high contrast stimuli.

Like cortical neurons, IC neurons also undergo contrast-dependent gain changes. However, this effect is weaker in IC than in A1/AAF (Figure 7.4). More broadly, while varying contrast has a stereotyped effect on cortical units (Figure 5.11), it produces a far greater variety of coding changes in IC neurons

(Figure 7.3). In total, the changes in IC coding are adaptive — insofar as IC neurons shift their stimulus sensitivities to be more informative about the current distribution of stimulus values — but, again, the changes are both more varied, and on average smaller, than the changes in cortical coding (Figure 7.9). The time course of gain changes is shorter in IC (30–40 ms) than in A1/AAF (80–100 ms), consistent with the shorter integration time windows of these neurons (Figures 7.12 and 7.13B-C).

These data, in total, argue against the hypothesis that the cortex inherits its contrast-dependent changes in coding from the IC. Rather, it appears that there is a cascade of at least two (and perhaps more) separate stages of contrast gain control in the ascending auditory pathway.

7.4.1 Stimulus representations in the IC and A1/AAF

In the previous chapter, I raised the question of what the stereotyped gain control of A1/AAF neurons means for the representation of ongoing sounds across the whole of the primary auditory cortex. It appears that the majority of A1/AAF neurons simply rescale their gain based on the local contrast statistics within their receptive fields. This suggests that the population representation of a high contrast sound, and the representation of a low contrast version of the same sound, might be very similar.

It appears, however, that this is very much not the case in the inferior colliculus. As shown in Figure 7.3, the firing patterns of some neurons vary wildly across different contrast conditions. It is therefore likely that were we to present two versions of a sound — one high contrast and the other low contrast — it would be difficult to infer the relationship between these two stimuli from the patterns of spiking in the IC.

This, perhaps, makes the apparent “contrast invariance” of the cortical representation all the more surprising. One typically expects that as we ascend the processing hierarchy, the computations performed by each station should become more complex, with response properties becoming correspondingly more varied. Yet the diversity here appears to be inverted, with the contrast-dependence within A1/AAF simpler than within the IC. We can only speculate as to potential reasons for this. It is possible that integrating out stimulus contrast is necessary for the computations that cortex needs to perform, but either unnecessary (or a hindrance) for the IC. Alternatively, the diversity seen within IC might be necessary in order to build a contrast-invariant representation of sounds in the cortex. In either case, this suggests that sounds are represented in a new way in A1/AAF, in a manner that is likely to be important for computational theories of what the primary auditory cortex actually does.

Chapter 8

Conclusions

8.1 The contrast problem

At the outset of this thesis, I raised a general problem for the auditory system: using neurons with only limited dynamic range, the auditory system must be able to represent and process sounds whose levels vary by only small amounts over time, as well as sounds whose levels vary considerably. The primary goal of this thesis was to determine how this *contrast problem* is solved at the level of the auditory cortex.

I considered two potential solutions. In the division of labour strategy, neurons would specialise for representing sounds of different contrasts. In Chapter 3, I found no evidence for such specialisation amongst neurons in ferret A1/AAF. In the adaptation strategy, neurons would adapt to the statistics of current stimulation. In Chapter 4, I showed that this is indeed the case in ferret A1/AAF. Here, neurons rescale their gain to compensate for changes in stimulus contrast. The gain changes are considerable, but do not completely compensate for changes in stimulus contrast. The gain changes also resemble the divisive normalisation observed in many other neural systems. In Chapters 5 and 6, I demonstrated that in a given stimulus context, auditory cortical

neurons determine their gain predominantly on the basis of spectrotemporally local statistics. In doing so, I developed a class of models that simply and significantly improve the predictions of how auditory cortical neurons respond to ongoing stimuli with complex spectrotemporal statistics. Finally, in Chapter 7, I showed that this adaptive strategy of the auditory cortex is not simply inherited from the inferior colliculus. When stimulus contrast changes, neurons in the IC do change their coding properties, but these changes are varied in form, and on average are weaker than in cortex.

In total, it appears that the auditory cortex deals with the contrast problem by treating it as a nuisance parameter, which it simply (though not completely) divides away. This is achieved using spectrotemporally local calculations in a dynamical process, and appears to be a novel property of the higher auditory pathway.

8.2 Comparing the auditory with the visual

In Section 1.3, I began by characterising what we know about the visual system's solution to its contrast problem. While we have seen that the auditory system also employs gain changes to address its analogous challenge, several differences between these two systems' approaches have come to light. First, in the visual system, the temporal characteristics of neurons' responses is very sensitive to contrast: integration windows typically become shorter at higher contrasts [Chander and Chichilnisky, 2001; Kim and Rieke, 2001; Baccus and Meister, 2002]. This does not appear to occur to a considerable degree in the auditory cortex: at a 25 ms resolution, STRFs do not change. Second, the contrast kernels for V1 neurons are typically broader than the linear receptive fields, in both the spatial and orientation domains [Heeger, 1992; Levitt and Lund, 1997; Angelucci et al., 2002; Webb et al., 2003; Schwabe et al., 2006;

[Angelucci and Bressloff, 2006](#)]. This does not appear to be the case for auditory cortical neurons, whose contrast kernels are mostly co-extensive with their STRFs.

The reasons for these departures might be found on different levels. Mechanistically, the primary auditory cortex is not equivalent to V1. One notable difference is that there is more subcortical processing in the auditory pathway than in the visual pathway, so that the information reaching the cortex has undergone different sets of transformations. This may situate gain control in different sets of underlying processes. From a computational perspective, the statistical structure of visual scenes is considerably different from that of auditory scenes. The relative robustness of the temporal integration windows of cortical neurons may be a consequence of the central role played by temporal modulations in auditory signals, while the size and shape of gain pools in the two systems may reflect processing demands imposed by higher-level correlations within their respective signal domains.

8.3 Adaptation and gain control

Although this thesis has argued in favour of an adaptive — rather than a division-of-labour — solution to the auditory contrast problem, whether the observed effects can truly be considered an “adaptation” has not been addressed in great detail. This label relates predominantly to the duration over which the effect manifests. In the work presented here, I have focused on gain changes which occur over short time scales, on the order of 100 ms in cortex, and 30 ms in inferior colliculus. In the visual system, there is a corresponding process with a similar time scale — contrast gain control — which is distinguished from slow contrast adaptation (which manifests over a course of seconds to minutes). It is therefore more appropriate to describe the observations of this

thesis as contrast gain control rather than adaptation. As the effects occur within the integration windows of these neurons, they could be the result of nonlinear processes within neurons' input-output transformations, rather than more dynamic, slow, adaptive processes.

This does not rule out the possibility that there are slow contrast adaptations (in addition to what has been studied here) operating in the auditory pathway. Given that auditory cortex is known to show many timescales of adaptation, ranging from seconds to tens of minutes [Ulanovsky et al., 2004], it is highly plausible that this is the case. Such adaptations may then have contributed to the changes in neural coding measured during the experiments of this thesis using longer stimulation protocols. Further experiments designed to explore longer-term coding changes would be necessary to characterise the adaptive effects of changing stimulus contrast.

8.4 From synthetic to natural sounds

In order to provide a thorough characterisation of how coding properties changed with stimuli contrast, it was necessary to develop a synthetic stimulus paradigm. Contrast DRCs and RC-DRCs are powerful insofar as they admit parametric control over statistics such as contrast. As with all synthetic stimuli, they diverge in many ways from natural sounds, which have far richer statistical structures [Nelken et al., 1999; Singh and Theunissen, 2003; Escabi et al., 2003; Woolley et al., 2005; Turner, 2010; McDermott and Simoncelli, 2011].

Some of the simplifications that are built into the DRCs are likely to constrain the way that auditory neurons operate. Many neurons' response properties vary as a function of the amplitude modulation frequency of a signal; as the DRCs had fixed modulation frequencies, there may have been many neurons which

responded suboptimally. Next, many natural sounds can be characterised by their sparseness, whereby events occur infrequently in bursts of energy [McDermott and Simoncelli, 2011]. The hypersensitivity of many cortical neurons to onsets [Heil and Irvine, 1996; Fishbach et al., 2001; Heil and Neubauer, 2003], and the diversity of their temporal firing patterns [Eggermont, 1991; Phillips and Sark, 1991; Qin et al., 2003] raises the question of whether the contrast-dependent changes in coding observed under stimulation using dense DRCs would also be present under stimulation using sparse natural sounds. Finally, many natural sounds have strong correlation in stimulus energy across frequency and time. Their considerable comodulation across frequency bands and complex amplitude autocorrelations are not shared by DRCs. It is likely that neural coding is sensitive to such statistics.

Given these issues, it will be necessary to explicitly test whether the auditory cortex also normalises out contrast when presented with natural stimuli. This will require manipulating the spectrotemporal statistics of natural sounds. The results of this thesis predict that if both a high contrast and a low contrast version of the same natural sound were presented, neurons in the auditory cortex would represent these two stimuli in a similar way. This would not, however, be the case for neurons in the inferior colliculus. A positive result would be of considerable significance for understanding how the auditory cortex processes sound in general.

If, however, the results observed here do not extend to natural sounds, then we would need to explain why such contrast invariance is seen at all for DRCs. One potential explanation for such an outcome might be that the dense, uncorrelated, ongoing nature of DRC stimulation drives the auditory cortex to a unique, steady-state mode of operation. If divisive normalisation is only produced in this state, then it may be the product of a canonical cortical circuit operation that is only activated under certain stimulation conditions. This

might draw parallels with the results of Mriganka Sur and colleagues [Sharma et al., 2000; Sur and Rubenstein, 2005], who routed the output of the retina to the auditory pathway, and found visual structures such as orientation tuning in the rewired auditory cortex. An alternative explanation for any discrepancy between DRC results and natural stimuli results might come from the hypothesis that the auditory cortex encodes (or is modulated by) a predictive model for the sound environment [Ulanovsky et al., 2003; Winkler et al., 2003; Gill et al., 2008; Winkler et al., 2009; Taaseh et al., 2011; Todorovic et al., 2011]. Under continual DRC stimulation, the cortex may be building up a simple model for ongoing sound; in this model, contrast could simply be integrated out. These hypothetical explanations, however, would need to be addressed in light of actual data from this experiment.

An important related question is whether the contrast — measured either as the standard deviation of the pressure over the mean pressure, or as the standard deviation of the sound level — is a quantity that has relevance for characterising the statistics of natural auditory scenes. The initial motivation for using this particular measure came squarely from the visual system. In vision, much work has been devoted to understanding the contrast statistics of natural scenes (e.g. [Laughlin, 1981; van Hateren, 1993; Ruderman and Bialek, 1994; van der Schaaf and van Hateren, 1996; Tadmor and Tolhurst, 2000; Schwartz and Simoncelli, 2001; Mante et al., 2005], to name a few). In particular, it appears that contrast and mean luminance vary independently of one another in patches of natural visual scenes [Mante et al., 2005]. While, as mentioned above, there has been some characterisation of the statistics of natural auditory scenes, the question of how these statistics look from within the receptive field of a single auditory neuron has not been addressed.

Perhaps it would be informative to ask whether contrast-invariance makes sense given the ways that the same sound sources would produce different signals in

different acoustic environments. For example, we might ask what kinds of natural acoustic conditions would reduce the contrast of a given sound. Two possibilities come to mind: superimposing the sound onto a noisy background; or presenting it within a highly reverberant environment. Both manipulations reduce the magnitude of sound level fluctuations compared with the unmasked, anechoic condition. However, both these scenarios also produce a change in other sound statistics, such as the mean sound level and the spectral sharpness, which may also impact the coding of the stimuli, as well as subjects' performance on any behavioural tasks. Nevertheless, as mentioned in Section 4.4.1, listeners are better able to discriminate reverberant words when embedded within a reverberant context sentence, than when embedded within an anechoic context [Watkins, 2005]. This effect can be partially reproduced when the context is provided by a noise stimulus with controlled statistics [Watkins and Makin, 2007]. Gain control might therefore provide a means of ameliorating the contrast-reducing effects of reverberation.

8.5 Further experiments

There are several further avenues for extending the work in this thesis. While contrast-dependent gain changes were initially observed in Chapter 4 for both anaesthetised and awake animals, the rest of the data were collected in the anaesthetised preparation only. It would be a simple, though time-consuming process to verify the remaining results in the awake ferret. More valuable, however, would be to investigate the contrast-dependent coding changes at other levels of the auditory pathway — particularly in the auditory thalamus, which lies midway between the IC and A1/AAF. This would likely help unpick how contrast is partially integrated out by the level of cortex.

The contrast kernel models developed here do not represent an end point, but

rather part of an ongoing challenge to improve our understanding of what sound features stimulate auditory neurons to fire. Some extensions to this model might warrant attention. The explicit dependence of neural gain on the recent history of contrast, σ_{tfh} , may turn out to be a proxy for a divisive normalisation network [Heeger, 1992]. In such a formulation, the gain of a neuron would be directly modulated by the output of a pool of other neurons. This suggests an extension of the models in which the contrast signals, ζ_t (Equations (5.11) – (5.14)), are passed through a static sigmoidal output nonlinearity, simulating the threshold and saturation properties of the gain pool. This approach might explain the fitted exponent of $n = 1.31$ for Equation (4.9). Beyond the specifics of the model parameterisations used here, there are likely to be connections between the contrast kernel models and the context models of Ahrens et al. [2008], as well as with GLMs; unpicking these might reveal novel directions for improving models for cortical neurons. Finally, the fact that stimuli and responses were discretised into 25 ms blocks limits the capacity of the models I used to investigate how contrast affects the fine temporal patterns of neural spiking. These are likely to be important given the precise temporal coding for which the auditory system is famous.

Appendices

Appendix A

Contrast kernel models: derivatives

A.1 Model parameters

We start with the following representation of the contrast-dependent output nonlinearity model:

$$\hat{y}_t = a_t + b_t \cdot (1 + \epsilon_t)^{-1} \quad (\text{A.1})$$

$$\epsilon_t = e^{-z'_t} \quad (\text{A.2})$$

$$z'_t = (z_t - c_t)/d_t \quad (\text{A.3})$$

$$\theta_t = \theta_{low} + (\theta_{high} - \theta_{low}) \cdot \zeta_t^{(\theta)} \quad (\text{A.4})$$

$$\zeta_t^{(\theta)} = \sum_f \kappa_f^{(\theta)} \cdot \sigma_{tf} \quad (\text{A.5})$$

$$\kappa_f^{(\theta)} = \frac{\lambda_f^{(\theta)}}{\sum_{f'} \lambda_{f'}^{(\theta)}} = \frac{\lambda_f^{(\theta)}}{L^{(\theta)}} \quad (\text{A.6})$$

We will start by computing derivatives at the deepest level (Equation (A.6)), and work our way via the chain rule back to $\hat{y}_t$.

Firstly, dropping the superscript (θ) , we note for frequency indices f and g that

$\partial L / \partial \lambda_g = 1$, and that $\partial \lambda_f / \partial \lambda_g = \delta_{fg}$. From the quotient rule, this gives:

$$\frac{\partial \kappa_f}{\partial \lambda_g} = \frac{\delta_{fg} L - \lambda_f}{L^2} \quad (\text{A.7})$$

Inserting this into Equation (A.5):

$$\begin{aligned} \frac{\partial \zeta_t}{\partial \lambda_g} &= \frac{1}{L^2} \sum_f ((\delta_{fg} L - \lambda_f) \cdot \sigma_{tf}) \\ &= \frac{\sigma_{tg}}{L} - \frac{1}{L^2} \sum_f (\lambda_f \cdot \sigma_{tf}) \\ &= \frac{\sigma_{tg} - \zeta_t}{L} \end{aligned} \quad (\text{A.8})$$

The partials of θ_t w.r.t. the contrast kernel coefficients then follow:

$$\begin{aligned} \frac{\partial \theta_t}{\partial \lambda_g^{(\theta)}} &= \frac{1}{L} \cdot (\theta_{high} - \theta_{low}) \cdot (\sigma_{tg} - \zeta_t^{(\theta)}) \\ &= \frac{\theta_1}{L} (\sigma_{tg} - \zeta_t^{(\theta)}) \end{aligned} \quad (\text{A.9})$$

The partials of θ_t w.r.t. the nonlinearity parameters are straightforward:

$$\frac{\partial \theta_t}{\partial \theta_{low}} = 1 - \zeta_t^{(\theta)} \quad (\text{A.10})$$

$$\frac{\partial \theta_t}{\partial \theta_{high}} = \zeta_t^{(\theta)} \quad (\text{A.11})$$

Now the partials of $\hat{y}_t$ w.r.t. the parameters $\theta_t \in \{a_t, b_t, c_t, d_t\}$ need to be calculated. We start by considering Equation (A.2):

$$\frac{\partial \epsilon_t}{\partial c_t} = \frac{\epsilon_t}{d_t} \quad (\text{A.12})$$

$$\frac{\partial \epsilon_t}{\partial d_t} = \epsilon_t \cdot \frac{(z_t - c_t)}{d_t^2} = \frac{\epsilon_t z'_t}{d_t} \quad (\text{A.13})$$

This gives partials $\partial \hat{y}_t / \partial \theta_t$ as:

$$\frac{\partial \hat{y}_t}{\partial a_t} = 1 \quad (\text{A.14})$$

$$\frac{\partial \hat{y}_t}{\partial b_t} = (1 + \epsilon_t)^{-1} \quad (\text{A.15})$$

$$\frac{\partial \hat{y}_t}{\partial c_t} = -b_t (1 + \epsilon_t)^{-2} \cdot \frac{\epsilon_t}{d_t} \equiv \nu_t \quad (\text{A.16})$$

$$\frac{\partial \hat{y}_t}{\partial d_t} = -b_t (1 + \epsilon_t)^{-2} \cdot \epsilon_t \cdot \frac{(z_t - c_t)}{d_t^2} = \nu_t z'_t \quad (\text{A.17})$$

Note that in models where any of the nonlinearity parameters θ_t were set as contrast-independent (so that $\theta_t = \theta$), then derivatives for the respective constant parameters θ are provided from Equations (A.14) – (A.17).

For the contrast-dependent parameters, the partials of $\hat{y}_t$ w.r.t. θ_{low} , θ_{high} , and the (frequency) coefficients of the unnormalised contrast kernels, $\lambda_g^{(\theta)}$, are obtained by multiplying the respective terms from Equations (A.14) – (A.17) with those from Equations (A.9) – (A.11).

A.2 Log likelihood

For simplicity, i.i.d. Gaussian errors were assumed, such that $y_t \mid \hat{y}_t \sim \mathcal{N}(0, \sigma_y^2)$.

Denoting the residuals as $r_t = y_t - \hat{y}_t$, the likelihood of the whole dataset is:

$$P(y \mid \hat{y}) = \prod_t \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{r_t^2}{2\sigma_y^2}\right) \quad (\text{A.18})$$

With the log likelihood:

$$\begin{aligned} \mathcal{LL} &\equiv \log(P(y \mid \hat{y})) \\ &= -\frac{N_T}{2} \log(2\pi\sigma_y^2) - \frac{1}{2\sigma_y^2} \sum_t r_t^2 \end{aligned} \quad (\text{A.19})$$

For parameter x , we have $\frac{\partial r_t}{\partial x} = -\frac{\partial \hat{y}_t}{\partial x}$, giving:

$$\frac{\partial \mathcal{LL}}{\partial x} = \frac{1}{\sigma_y^2} \sum_t r_t \frac{\partial \hat{y}_t}{\partial x} \quad (\text{A.20})$$

While σ_y can also be optimised in this way, it was easier and quicker to fix σ_y as the standard deviation of the residuals from the linear-nonlinear (NL) model. This does not impact on the maximum likelihood (ML) estimates of the other parameters. However, it does have an impact on their MAP estimates, as it underestimates the precision of the data, and therefore weights the priors on these parameters more strongly. However, as priors were broad, and the number of data points N_T very large, this problem proved negligible.

A.3 Log posterior

Using Bayes' Rule, we have the log posterior on parameters θ as:

$$\mathcal{LP} = \mathcal{LL} + \log P(\theta) + \text{const} \quad (\text{A.21})$$

Priors on the nonlinearity parameters are given in Equations (5.32) – (5.35); priors on the unnormalised contrast kernels are given in Section 5.4.1.9. All of these were either exponential or normal distributions.

For an exponential random variable $X \sim \text{Exp}(\beta)$, where β is the scale parameter:

$$\frac{\partial \log P(X)}{\partial X} = -1/\beta \quad (\text{A.22})$$

For a normal random variable $X \sim \mathcal{N}(\mu, \sigma^2)$:

$$\frac{\partial \log P(X)}{\partial X} = -\frac{1}{\sigma^2}(X - \mu) \quad (\text{A.23})$$

The derivatives of the log posterior w.r.t. the nonlinearity parameters are there-

fore:

$$\frac{\partial \mathcal{LP}}{\partial a_{low}} = \left(\frac{1}{\sigma_y^2} \sum_t r_t (1 - \zeta_t^{(a)}) \right) - \frac{1}{\beta_a} \quad (\text{A.24})$$

$$\frac{\partial \mathcal{LP}}{\partial a_{high}} = \left(\frac{1}{\sigma_y^2} \sum_t r_t \zeta_t^{(a)} \right) - \frac{1}{\beta_a} \quad (\text{A.25})$$

$$\frac{\partial \mathcal{LP}}{\partial b_{low}} = \left(\frac{1}{\sigma_y^2} \sum_t r_t (1 + \epsilon_t)^{-1} (1 - \zeta_t^{(b)}) \right) - \frac{1}{\beta_b} \quad (\text{A.26})$$

$$\frac{\partial \mathcal{LP}}{\partial b_{high}} = \left(\frac{1}{\sigma_y^2} \sum_t r_t (1 + \epsilon_t)^{-1} \zeta_t^{(b)} \right) - \frac{1}{\beta_b} \quad (\text{A.27})$$

$$\frac{\partial \mathcal{LP}}{\partial c_{low}} = \left(\frac{1}{\sigma_y^2} \sum_t r_t \nu_t (1 - \zeta_t^{(c)}) \right) - \frac{(c_{low} - \mu_c)}{\sigma_c^2} \quad (\text{A.28})$$

$$\frac{\partial \mathcal{LP}}{\partial c_{high}} = \left(\frac{1}{\sigma_y^2} \sum_t r_t \nu_t \zeta_t^{(c)} \right) - \frac{(c_{high} - \mu_c)}{\sigma_c^2} \quad (\text{A.29})$$

$$\frac{\partial \mathcal{LP}}{\partial d_{low}} = \left(\frac{1}{\sigma_y^2} \sum_t r_t \nu_t z'_t (1 - \zeta_t^{(d)}) \right) - \frac{1}{\beta_d} \quad (\text{A.30})$$

$$\frac{\partial \mathcal{LP}}{\partial d_{high}} = \left(\frac{1}{\sigma_y^2} \sum_t r_t \nu_t z'_t \zeta_t^{(d)} \right) - \frac{1}{\beta_d} \quad (\text{A.31})$$

When real-valued contrast kernels are admitted, the derivatives of the log posterior w.r.t. the unnormalised contrast parameters are:

$$\frac{\partial \mathcal{LP}}{\partial \lambda_g^{(a)}} = \left(\frac{a_1}{L^{(a)} \sigma_y^2} \sum_t r_t (\sigma_{tg} - \zeta_t^{(a)}) \right) - 100 \lambda_g^{(a)} \quad (\text{A.32})$$

$$\frac{\partial \mathcal{LP}}{\partial \lambda_g^{(b)}} = \left(\frac{b_1}{L^{(b)} \sigma_y^2} \sum_t r_t (1 + \epsilon_t)^{-1} (\sigma_{tg} - \zeta_t^{(c)}) \right) - 100 \lambda_g^{(b)} \quad (\text{A.33})$$

$$\frac{\partial \mathcal{LP}}{\partial \lambda_g^{(c)}} = \left(\frac{c_1}{L^{(c)} \sigma_y^2} \sum_t r_t \nu_t (\sigma_{tg} - \zeta_t^{(c)}) \right) - 100 \lambda_g^{(c)} \quad (\text{A.34})$$

$$\frac{\partial \mathcal{LP}}{\partial \lambda_g^{(d)}} = \left(\frac{d_1}{L^{(d)} \sigma_y^2} \sum_t r_t \nu_t z'_t (\sigma_{tg} - \zeta_t^{(d)}) \right) - 100 \lambda_g^{(d)} \quad (\text{A.35})$$

When contrast kernels are constrained to be positive, these become:

$$\frac{\partial \mathcal{LP}}{\partial \lambda_g^{(a)}} = \left(\frac{a_1}{L^{(a)} \sigma_y^2} \sum_t r_t \left(\sigma_{tg} - \zeta_t^{(a)} \right) \right) - 10 \quad (\text{A.36})$$

$$\frac{\partial \mathcal{LP}}{\partial \lambda_g^{(b)}} = \left(\frac{b_1}{L^{(b)} \sigma_y^2} \sum_t r_t (1 + \epsilon_t)^{-1} \left(\sigma_{tg} - \zeta_t^{(c)} \right) \right) - 10 \quad (\text{A.37})$$

$$\frac{\partial \mathcal{LP}}{\partial \lambda_g^{(c)}} = \left(\frac{c_1}{L^{(c)} \sigma_y^2} \sum_t r_t \nu_t \left(\sigma_{tg} - \zeta_t^{(c)} \right) \right) - 10 \quad (\text{A.38})$$

$$\frac{\partial \mathcal{LP}}{\partial \lambda_g^{(d)}} = \left(\frac{d_1}{L^{(d)} \sigma_y^2} \sum_t r_t \nu_t z'_t \left(\sigma_{tg} - \zeta_t^{(d)} \right) \right) - 10 \quad (\text{A.39})$$

When shared contrast kernels between several parameter are used, this involves summing the likelihood terms with a single prior term, as:

$$\frac{\partial \mathcal{LP}}{\partial \lambda_g} = \left(\sum_{\theta} \frac{\partial \mathcal{LL}}{\partial \lambda_g^{(\theta)}} \right) + \frac{\partial P(\lambda_g)}{\partial \lambda_g} \quad (\text{A.40})$$

Appendix B

Cross-frequency kernel models: derivatives

B.1 Model parameters

We start with a representation of the cross-frequency contrast kernel model. First, the core nonlinearity is given as follows:

$$\hat{y}_t = a_t + b_t \cdot (1 + \epsilon_t)^{-1} \quad (\text{B.1})$$

$$\epsilon_t = e^{-z'_t} \quad (\text{B.2})$$

$$z'_t = \sum_f \frac{z_{tf} - c_t}{d_{tf}} \equiv \sum_f z'_{tf} \quad (\text{B.3})$$

$$z_{tf} = \sum_h X_{tfh} k_{fh} \quad (\text{B.4})$$

Next, the nonlinearity parameters change over time:

$$d_{tf} = d_{low} + (d_{high} - d_{low}) \cdot \zeta_{tf} \quad (\text{B.5})$$

$$c_t = c_{low} + (c_{high} - c_{low}) \cdot \zeta_t \quad (\text{B.6})$$

with the average over frequency:

$$\zeta_t = \frac{1}{N_F} \sum_f \zeta_{tf} \quad (\text{B.7})$$

For brevity, for each $\theta \in \{c, d\}$, we also have the definitions $\theta_0 = \theta_{low}$ and $\theta_1 = (\theta_{high} - \theta_{low})$.

Finally, the weighted contrast signal, ζ_{tf} , depends on a cross-frequency contrast kernel, C_{fg} :

$$\zeta_{tf} = \sum_g C_{fg} \sigma_{tg} \quad (\text{B.8})$$

$$C_{fg} = \frac{\mu_{fg}}{M} \quad (\text{B.9})$$

$$M = \frac{1}{N_F} \sum_{f,g} \mu_{fg} \quad (\text{B.10})$$

$$\mu_{fg} = S_{fg} \lambda_g \quad (\text{B.11})$$

$$S_{fg} = \exp\left(-\frac{\tau^{(\kappa)}}{2} D_{fg}\right) \quad (\text{B.12})$$

$$D_{fg} = (f - g)^2 \quad (\text{B.13})$$

The parameters to fit in this model are: the nonlinearity parameters a , b , c_{low} , c_{high} , d_{low} , d_{high} ; the (unnormalised) contrast kernel λ_g ; and the spectral spread of contrast-dependent effects, given by the precision (i.e. inverse variance) $\tau^{(\kappa)} = 1/\sigma_\kappa^2$. Three forms of this model were explored. In the *global effect model*, gain changes were applied to all frequency bands, i.e. $\tau^{(\kappa)} \rightarrow 0$, and $S_{fg} = 1$. As shown below, this reduces to the standard contrast frequency kernel model presented in Section 5.4.2.3. In the *local effect model*, gain changes were applied only to individual frequency bands, such that a high contrast in band f only reduces the gain for sounds in that band. Here, $\tau^{(\kappa)} \rightarrow \infty$, so that $S_{fg} = \delta_{fg}$. Finally, I allow $\tau^{(\kappa)}$ to vary, denoted the *regional effect model*.

Priors on parameters are as in the previous model, documented in Appendix A. Since the horizontal offset term, c_t , is effectively repeated N_F times in this representation, the mean and standard deviation of the Gaussian priors over c_{low} and c_{high} were adjusted proportionally.

B.2 Global effect model

As explained above, the global effect model takes the limit $\tau^{(\kappa)} \rightarrow 0$. From the equations above, the cross-contrast kernel, C_{fg} , reduces as follows:

$$\begin{aligned}
 S_{fg} &= 1 \quad \forall f, g \\
 \Rightarrow \mu_{fg} &= \lambda_g \\
 \Rightarrow M &= \sum_{g'} \lambda_{g'} \\
 \Rightarrow C_{fg} &= \frac{\lambda_g}{\sum_{g'} \lambda_{g'}} \equiv \kappa_g
 \end{aligned} \tag{B.14}$$

The cross-contrast kernel C_{fg} measures the influence of the contrast in band g on the gain of band f . From Equation (B.14), this is independent of f ; thus the contrast signal, ζ_{tf} , is independent of f :

$$\zeta_{tf} = \zeta_t = \sum_g \kappa_g \sigma_{tg} \tag{B.15}$$

The nonlinearity parameters are therefore:

$$d_{tf} = d_t = d_{low} + (d_{high} - d_{low}) \cdot \zeta_t \tag{B.16}$$

$$c_t = c_{low} + (c_{high} - c_{low}) \cdot \zeta_t \tag{B.17}$$

Finally, we get:

$$\begin{aligned}
 z'_t &= \sum_f \frac{z_{tf} - c_t}{d_t} \\
 &= \frac{z_t - c_t}{d_t}
 \end{aligned} \tag{B.18}$$

thus recovering the model of Section 5.4.2.3.

B.3 Local effect model

In the local effect model, the limit $\tau^{(\kappa)} \rightarrow \infty$ is taken, such that contrast changes in band g only affect the responses to level changes within that band g . The cross-contrast kernel, C_{fg} , reduces as follows:

$$\begin{aligned}
 S_{fg} &= \delta_{fg} \\
 \Rightarrow \mu_{fg} &= \lambda_g \cdot \delta_{fg} \\
 \Rightarrow M &= \frac{L}{N_F} \\
 \Rightarrow C_{fg} &= \frac{N_F \cdot \lambda_g \cdot \delta_{fg}}{L} = N_F \cdot \kappa_g \cdot \delta_{fg}
 \end{aligned} \tag{B.19}$$

where $\kappa_g = \lambda_g/L$ and $L = \sum_{g'} \lambda_{g'}$ are defined in the standard way.

Applying Equation (B.8), the contrast signal, ζ_{tf} thus depends only on the contrast in band f :

$$\zeta_{tf} = N_F \cdot \kappa_f \cdot \sigma_{tf} \tag{B.20}$$

$$\zeta_t = \sum_f \kappa_f \cdot \sigma_{tf} \tag{B.21}$$

B.3.1 Derivatives

We begin by computing derivatives w.r.t. the parameters of the unnormalised contrast kernel, λ_f :

$$\frac{\partial \kappa_f}{\partial \lambda_{f'}} = \frac{1}{L} (\delta_{ff'} - \kappa_f) \tag{B.22}$$

$$\frac{\partial \zeta_{tf}}{\partial \lambda_{f'}} = \frac{1}{L} (N_F \cdot \delta_{ff'} \cdot \sigma_{tf} - \zeta_{tf}) \tag{B.23}$$

$$\frac{\partial \zeta_t}{\partial \lambda_{f'}} = \frac{1}{L} (\sigma_{tf'} - \zeta_t) \tag{B.24}$$

$$\frac{\partial c_t}{\partial \lambda_{f'}} = \frac{c_1}{L} (\sigma_{tf'} - \zeta_t) \tag{B.25}$$

$$\frac{\partial d_{tf}}{\partial \lambda_{f'}} = \frac{d_1}{L} (N_F \cdot \delta_{ff'} \cdot \sigma_{tf} - \zeta_{tf}) \tag{B.26}$$

We also have the derivatives w.r.t. θ_{high} and θ_{low} :

$$\frac{\partial c_t}{\partial c_{low}} = 1 - \zeta_t \quad (\text{B.27})$$

$$\frac{\partial c_t}{\partial c_{high}} = \zeta_t \quad (\text{B.28})$$

$$\frac{\partial d_{tf}}{\partial d_{low}} = 1 - \zeta_{tf} \quad (\text{B.29})$$

$$\frac{\partial d_{tf}}{\partial d_{high}} = \zeta_{tf} \quad (\text{B.30})$$

To put these together, we take partials of z'_t with respect to d_{tf} and c_t from Equation (B.3):

$$\frac{\partial z'_t}{\partial d_{tf}} = -\frac{z'_{tf}}{d_{tf}} \equiv \phi_{tf} \quad (\text{B.31})$$

$$\frac{\partial z'_t}{\partial c_t} = -\sum_f \frac{1}{d_{tf}} \equiv \psi_t \quad (\text{B.32})$$

Such that the partials of z'_t with respect to an arbitrary variable θ is given by:

$$\frac{\partial z'_t}{\partial \theta} = \left[\sum_f \phi_{tf} \frac{\partial d_{tf}}{\partial \theta} \right] + \psi_t \frac{\partial c_t}{\partial \theta} \quad (\text{B.33})$$

Finally, we get:

$$\begin{aligned} \frac{\partial z'_t}{\partial \lambda_{f'}} &= \frac{d_1}{L} \left[N_F \cdot \phi_{tf'} \cdot \sigma_{tf'} - \sum_f \phi_{tf} \cdot \zeta_{tf} \right] \\ &\quad + \frac{c_1 \psi_t}{L} \left[\sigma_{tf'} - \zeta_t \right] \end{aligned} \quad (\text{B.34})$$

And:

$$\frac{\partial z'_t}{\partial c_{low}} = \psi_t (1 - \zeta_t) \quad (\text{B.35})$$

$$\frac{\partial z'_t}{\partial c_{high}} = \psi_t \cdot \zeta_t \quad (\text{B.36})$$

$$\frac{\partial z'_t}{\partial d_{low}} = \sum_f \phi_{tf} (1 - \zeta_{tf}) \quad (\text{B.37})$$

$$\frac{\partial z'_t}{\partial d_{high}} = \sum_f \phi_{tf} \cdot \zeta_{tf} \quad (\text{B.38})$$

These are combined with:

$$\frac{\partial \hat{y}_t}{\partial a} = 1 \quad (\text{B.39})$$

$$\frac{\partial \hat{y}_t}{\partial b} = (1 + \epsilon_t)^{-1} \quad (\text{B.40})$$

$$\frac{\partial \hat{y}_t}{\partial \theta} = b(1 + \epsilon_t)^{-2} \epsilon_t \frac{\partial z'_t}{\partial \theta} \quad (\text{B.41})$$

and the same procedures as in Appendix A.2 and A.3 are followed.

B.4 Regional effect model

The region effect model allows $\tau^{(\kappa)}$ to take on any value. I placed an exponential prior on $\tau^{(\kappa)}$, with a scale parameter $\beta = 1$. The log probability of the prior is particularly flat over the plausible range of $\tau^{(\kappa)}$, and, given the volume of data, the priors were dwarfed by $\tau^{(\kappa)}$ -dependent changes in the log likelihood.

We start by computing derivatives w.r.t. coefficients of the unnormalised contrast kernel:

$$\frac{\partial \mu_{fg}}{\partial \lambda_{g'}} = S_{fg} \delta_{gg'} \quad (\text{B.42})$$

$$\frac{\partial M}{\partial \lambda_{g'}} = \frac{1}{N_F} \sum_f S_{fg'} \equiv S_{g'} \quad (\text{B.43})$$

$$\frac{\partial C_{fg}}{\partial \lambda_{g'}} = \frac{1}{M} [S_{fg} \cdot \delta_{gg'} - C_{fg} \cdot S_{g'}] \quad (\text{B.44})$$

$$\frac{\partial \zeta_{tf}}{\partial \lambda_{g'}} = \frac{1}{M} [S_{fg'} \cdot \sigma_{tg'} - \zeta_{tf} \cdot S_{g'}] \quad (\text{B.45})$$

$$\frac{\partial \zeta_t}{\partial \lambda_{g'}} = \frac{1}{M} S_{g'} [\sigma_{tg'} - \zeta_t] \quad (\text{B.46})$$

This gives:

$$\begin{aligned} \frac{\partial z'_t}{\partial \lambda_{g'}} &= \frac{d_1}{M} \sigma_{tg'} \left[\sum_f \phi_{tf} \cdot S_{fg'} \right] \\ &\quad - \frac{d_1}{M} S_{g'} \left[\sum_f \phi_{tf} \cdot \zeta_{tf} \right] \\ &\quad + \frac{c_1 \psi_t}{M} S_{g'} \left[\sigma_{tg'} - \zeta_t \right] \end{aligned} \quad (\text{B.47})$$

with ϕ_{tf} and ψ_t defined as in Equations (B.31) and (B.32).

Next, derivatives w.r.t. $\tau^{(\kappa)}$:

$$\frac{\partial S_{fg}}{\partial \tau^{(\kappa)}} = -\frac{1}{2} S_{fg} D_{fg} \quad (\text{B.48})$$

$$\frac{\partial \mu_{fg}}{\partial \tau^{(\kappa)}} = -\frac{1}{2} D_{fg} \mu_{fg} \quad (\text{B.49})$$

$$\frac{\partial M}{\partial \tau^{(\kappa)}} = -\frac{1}{2 N_F} \sum_{fg} D_{fg} \mu_{fg} \quad (\text{B.50})$$

$$\frac{\partial C_{fg}}{\partial \tau^{(\kappa)}} = C_{fg} \left[-\frac{1}{2} D_{fg} - \frac{1}{M} \frac{\partial M}{\partial \tau^{(\kappa)}} \right] \quad (\text{B.51})$$

$$\frac{\partial \zeta_{tf}}{\partial \tau^{(\kappa)}} = -\frac{1}{2} \left[\sum_g C_{fg} D_{fg} \sigma_{tg} \right] - \frac{1}{M} \frac{\partial M}{\partial \tau^{(\kappa)}} \zeta_{tf} \quad (\text{B.52})$$

$$\frac{\partial \zeta_t}{\partial \tau^{(\kappa)}} = -\frac{1}{2 N_F} \left[\sum_{fg} C_{fg} D_{fg} \sigma_{tg} \right] - \frac{1}{M} \frac{\partial M}{\partial \tau^{(\kappa)}} \zeta_t \quad (\text{B.53})$$

$$\frac{\partial z'_t}{\partial \tau^{(\kappa)}} = d_1 \left[\sum_f \phi_{tf} \frac{\partial \zeta_{tf}}{\partial \tau^{(\kappa)}} \right] + c_1 \psi_t \frac{\partial \zeta_t}{\partial \tau^{(\kappa)}} \quad (\text{B.54})$$

These, together with Equations (B.35) – (B.41) provide the required terms.

References

- Abbott, L., Chance, F. S. and V.A. Casagrande, R. G. [2005], Drivers and modulators from push-pull and balanced synaptic input, *in* ‘Cortical Function: a View from the Thalamus’, Vol. 149, Elsevier, pp. 147–155.
- Abolafia, J. M., Vergara, R., Arnold, M. M., Reig, R. and Sanchez-Vives, M. V. [2010], ‘Cortical auditory adaptation in the awake rat and the role of potassium currents’, *Cerebral Cortex* **21**(5), 977–990.
- Aertsen, A. M. H. J. and Johannesma, P. I. M. [1981], ‘The spectro-temporal receptive field’, *Biological Cybernetics* **42**, 133–143.
- Aertsen, A. M. H. J., Johannesma, P. I. M. and Hermes, D. J. [1980], ‘Spectro-temporal receptive fields of auditory neurons in the grassfrog II: analysis of the Stimulus-Event relation for tonal stimuli’, *Biological Cybernetics* **38**, 235–248.
- Ahrens, M., Linden, J. and Sahani, M. [2008], ‘Nonlinearities and contextual influences in auditory cortical responses modeled with multilinear spectrotemporal methods’, *The Journal of Neuroscience* **28**(8), 1929–42.
- Albrecht, D. G. and Hamilton, D. B. [1982], ‘Striate cortex of monkey and cat: contrast response function’, *Journal of Neurophysiology* **48**(1), 217–237.
- Andersen, R. A., Knight, P. L. and Merzenich, M. M. [1980], ‘The thalamocortical and corticothalamic connections of AI, AII, and the

- anterior auditory field (AAF) in the cat: evidence for two largely segregated systems of connections', *The Journal of Comparative Neurology* **194**(3), 663–701.
- Angelucci, A. and Bressloff, P. [2006], 'Contribution of feedforward, lateral and feedback connections to the classical receptive field center and extra-classical receptive field surround of primate V1 neurons', *Progress in Brain Research* **154**, 93–120.
- Angelucci, A., Levitt, J. B., Walton, E. J. S., Hupé, J., Bullier, J. and Lund, J. S. [2002], 'Circuits for local and global signal integration in primary visual cortex', *The Journal of Neuroscience* **22**(19), 8633–8646.
- Atick, J. [1992], 'Could information theory provide an ecological theory of sensory processing?', *Network: Computation in Neural Systems* **3**(2), 213–251.
- Attias, H. and Schreiner, C. [1997], Temporal low-order statistics of natural sounds, in 'Advances in Neural Information Processing Systems', Vol. 9, MIT Press, Cambridge, MA, pp. 27–33.
- Attneave, F. [1954], 'Some informational aspects of visual perception', *Psychological Review* **61**(3), 183–193.
- Baccus, S. and Meister, M. [2002], 'Fast and slow contrast adaptation in retinal circuitry', *Neuron* **36**(5), 909–919.
- Bacon, S. and Grantham, D. [1989], 'Modulation masking: effects of modulation frequency, depth, and phase', *The Journal of the Acoustical Society of America* **85**(6), 2575–2580.
- Bajo, V. M., Nodal, F. R., Moore, D. R. and King, A. J. [2010], 'The descending corticocollicular pathway mediates learning-induced auditory plasticity', *Nature Neuroscience* **13**(2), 253–260.

- Barbour, D. L. [2011], ‘Intensity-invariant coding in the auditory system’, *Neuroscience & Biobehavioral Reviews* **35**(10), 2064–2072.
- Barbour, D. L. and Wang, X. [2003a], ‘Auditory cortical responses elicited in awake primates by random spectrum stimuli’, *The Journal of Neuroscience* **23**(18), 7194–7206.
- Barbour, D. L. and Wang, X. [2003b], ‘Contrast tuning in auditory cortex’, *Science* **299**(5609), 1073–1075.
- Barlow, H. B. [1961], Possible principles underlying the transformation of sensory messages, *in* W. Rosenblith, ed., ‘Sensory Communication’, MIT Press, Cambridge, MA, pp. 217–234.
- Basso, M. A. and Wurtz, R. H. [1997], ‘Modulation of neuronal activity by target uncertainty’, *Nature* **389**(6646), 66–69.
- 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.
- Beck, J. M., Latham, P. E. and Pouget, A. [2011], ‘Marginalization in neural circuits with divisive normalization’, *The Journal of Neuroscience* **31**(43), 15310–15319.
- Benardete, E. A., Kaplan, E. and Knight, B. W. [1992], ‘Contrast gain control in the primate retina: P cells are not x-like, some m cells are’, *Visual Neuroscience* **8**(05), 483–486.
- Bendor, D. and Wang, X. [2005], ‘The neuronal representation of pitch in primate auditory cortex’, *Nature* **436**(7054), 1161–1165.
- Berens, P. [2009], ‘CircStat: a MATLAB toolbox for circular statistics’, *Journal of Statistical Software* **31**(10), 1–21.

- Bitterman, Y., Mukamel, R., Malach, R., Fried, I. and Nelken, I. [2008], ‘Ultra-fine frequency tuning revealed in single neurons of human auditory cortex’, *Nature* **451**(7175), 197–201.
- Bizley, J. K., Nodal, F. R., Nelken, I. and King, A. J. [2005], ‘Functional organization of ferret auditory cortex’, *Cerebral Cortex* **15**(10), 1637–1653.
- Bizley, J. K., Walker, K. M. M., King, A. J. and Schnupp, J. W. H. [2010], ‘Neural ensemble codes for stimulus periodicity in auditory cortex’, *The Journal of Neuroscience* **30**(14), 5078–5091.
- Bizley, J. K., Walker, K. M. M., Silverman, B. W., King, A. J. and Schnupp, J. W. H. [2009], ‘Interdependent encoding of pitch, timbre, and spatial location in auditory cortex’, *The Journal of Neuroscience* **29**(7), 2064–2075.
- Blake, D. T. and Merzenich, M. M. [2002], ‘Changes of AI receptive fields with sound density’, *Journal of Neurophysiology* **88**(6), 3409–3420.
- Bonin, V., Mante, V. and Carandini, M. [2005], ‘The suppressive field of neurons in lateral geniculate nucleus’, *The Journal of Neuroscience* **25**(47), 10844–10856.
- Bonin, V., Mante, V. and Carandini, M. [2006], ‘The statistical computation underlying contrast gain control’, *The Journal of Neuroscience* **26**(23), 6346–6353.
- Brady, N. and Field, D. J. [2000], ‘Local contrast in natural images: normalisation and coding efficiency’, *Perception* **29**(9), 1041–1055.
- 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’, *The Journal of Comparative Neurology* **155**(3), 251–299.
- Britten, K. H. and Heuer, H. W. [1999], ‘Spatial summation in the receptive fields of MT neurons’, *The Journal of Neuroscience* **19**(12), 5074–5084.
- Brown, S. and Masland, R. [2001], ‘Spatial scale and cellular substrate of contrast adaptation by retinal ganglion cells’, *Nature Neuroscience* **4**(1), 44–51.
- Budinger, E., Heil, P. and Scheich, H. [2000], ‘Functional organization of auditory cortex in the mongolian gerbil (*Meriones unguiculatus*). III. anatomical subdivisions and corticocortical connections’, *European Journal of Neuroscience* **12**(7), 2425–2451.
- Busse, L., Wade, A. R. and Carandini, M. [2009], ‘Representation of concurrent stimuli by population activity in visual cortex’, *Neuron* **64**(6), 931–942.
- Calabrese, A., Schumacher, J. W., Schneider, D. M., Paninski, L. and Woolley, S. M. N. [2011], ‘A generalized linear model for estimating spectrotemporal receptive fields from responses to natural sounds’, *PLoS ONE* **6**(1), e16104.
- Calford, M. and Aitkin, L. [1983], ‘Ascending projections to the medial geniculate body of the cat: evidence for multiple, parallel auditory pathways through thalamus’, *The Journal of Neuroscience* **3**(11), 2365–2380.
- Calhoun, B. M. and Schreiner, C. E. [1998], ‘Spectral envelope coding in cat primary auditory cortex: linear and non-linear effects of stimulus characteristics’, *European Journal of Neuroscience* **10**(3), 926–940.

- Campbell, R. A. A., Schulz, A. L., King, A. J. and Schnupp, J. W. H. [2010], ‘Brief sounds evoke prolonged responses in anesthetized ferret auditory cortex’, *Journal of Neurophysiology* **103**(5), 2783–2793.
- Carandini, M., Heeger, D. J. and Movshon, J. A. [1997], ‘Linearity and normalization in simple cells of the macaque primary visual cortex’, *The Journal of Neuroscience* **17**(21), 8621–8644.
- Carandini, M., Heeger, D. J. and Senn, W. [2002], ‘A synaptic explanation of suppression in visual cortex’, *The Journal of Neuroscience* **22**(22), 10053–10065.
- Casseday, J., Fremouw, T. and Covey, E. [2002], ‘The inferior colliculus: a hub for the central auditory system’, *Springer Handbook of Auditory Research: Integrative Functions in the Mammalian Auditory Pathway* **15**, 238–318.
- Chander, D. and Chichilnisky, E. J. [2001], ‘Adaptation to temporal contrast in primate and salamander retina’, *The Journal of Neuroscience* **21**(24), 9904–9916.
- Cheng, H., Chino, Y. M., Smith, E. L., Hamamoto, J. and Yoshida, K. [1995], ‘Transfer characteristics of lateral geniculate nucleus x neurons in the cat: effects of spatial frequency and contrast’, *Journal of Neurophysiology* **74**(6), 2548–2557.
- Chichilnisky, E. J. [2001], ‘A simple white noise analysis of neuronal light responses’, *Network: Computation in Neural Systems* **12**(2), 199–213.
- Christianson, G. B., Sahani, M. and Linden, J. F. [2008], ‘The consequences of response nonlinearities for interpretation of spectrotemporal receptive fields’, *The Journal of Neuroscience* **28**(2), 446–455.
- Chung, S., Li, X. and Nelson, S. B. [2002], ‘Short-term depression at thalamocortical synapses contributes to rapid adaptation of cortical sensory responses in vivo’, *Neuron* **34**(3), 437–446.

- Colwell, S. A. [1975], ‘Thalamocortical-corticothalamic reciprocity: a combined anterograde-retrograde tracer technique’, *Brain Research* **92**(3), 443–449.
- Dahmen, J. C., Hartley, D. E. H. and King, A. J. [2008], ‘Stimulus-timing-dependent plasticity of cortical frequency representation’, *The Journal of Neuroscience* **28**(50), 13629–13639.
- Dahmen, J. C., Keating, P., Nodal, F. R., Schulz, A. and King, A. J. [2010], ‘Adaptation to stimulus statistics in the perception and neural representation of auditory space’, *Neuron* **66**(6), 937–948.
- Dallos, P. [1992], ‘The active cochlea’, *The Journal of Neuroscience* **12**(12), 4575–4585.
- Dau, T. [1997], ‘Modeling auditory processing of amplitude modulation. i. detection and masking with narrow-band carriers’, *The Journal of the Acoustical Society of America* **102**, 2892–2905.
- David, S. V., Mesgarani, N., Fritz, J. B. and Shamma, S. A. [2009], ‘Rapid synaptic depression explains nonlinear modulation of spectro-temporal tuning in primary auditory cortex by natural stimuli’, *The Journal of Neuroscience* **29**(11), 3374–3386.
- David, S. V., Mesgarani, N. and Shamma, S. A. [2007], ‘Estimating sparse spectro-temporal receptive fields with natural stimuli’, *Network: Computation in Neural Systems* **18**(3), 191–212.
- Davis, K. A., Ramachandran, R. and May, B. J. [2003], ‘Auditory processing of spectral cues for sound localization in the inferior colliculus’, *Journal of the Association for Research in Otolaryngology* **4**, 148–163.
- Dean, I., Harper, N. S. and McAlpine, D. [2005], ‘Neural population coding of sound level adapts to stimulus statistics’, *Nature Neuroscience* **8**(12), 1684–1689.

- Dean, I., Robinson, B. L., Harper, N. S. and McAlpine, D. [2008], ‘Rapid neural adaptation to sound level statistics’, *The Journal of Neuroscience* **28**(25), 6430–6438.
- deCharms, R. C., Blake, D. T. and Merzenich, M. M. [1998], ‘Optimizing sound features for cortical neurons’, *Science* **280**(5368), 1439–1444.
- Deneve, S., Latham, P. E. and Pouget, A. [1999], ‘Reading population codes: a neural implementation of ideal observers’, *Nature Neuroscience* **2**(8), 740–745.
- 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. and Zador, A. [1998], ‘Asymmetric dynamics in optimal variance adaptation’, *Neural Computation* **10**(5), 1179–1202.
- DiMattina, C. and Wang, X. [2006], ‘Virtual vocalization stimuli for investigating neural representations of species-specific vocalizations’, *Journal of Neurophysiology* **95**(2), 1244–1262.
- Dobrunz, L. E., Huang, E. P. and Stevens, C. F. [1997], ‘Very short-term plasticity in hippocampal synapses’, *Proceedings of the National Academy of Sciences* **94**(26), 14843–14847.
- Donoho, D. L. and Johnstone, J. M. [1994], ‘Ideal spatial adaptation by wavelet shrinkage’, *Biometrika* **81**(3), 425–455.
- Dunn, F. A., Lankheet, M. J. and Rieke, F. [2007], ‘Light adaptation in cone vision involves switching between receptor and post-receptor sites’, *Nature* **449**(7162), 603–606.

- Eggermont, J. J. [1991], ‘Rate and synchronization measures of periodicity coding in cat primary auditory cortex’, *Hearing Research* **56**(1-2), 153–167.
- Eggermont, J. J. [1994], ‘Temporal modulation transfer functions for AM and FM stimuli in cat auditory cortex. effects of carrier type, modulating waveform and intensity’, *Hearing Research* **74**(1-2), 51–66.
- Elliott, D. and McGee, T. [1965], ‘Effect of cochlear lesions upon audiograms and intensity discrimination in cats’, *The Annals of Otology, Rhinology, and Laryngology* **74**, 386–408.
- Escabi, M. A., Miller, L. M., Read, H. L. and Schreiner, C. E. [2003], ‘Naturalistic auditory contrast improves spectrotemporal coding in the cat inferior colliculus’, *The Journal of Neuroscience* **23**(37), 11489–11504.
- Escabi, M. A. and Schreiner, C. E. [2002], ‘Nonlinear spectrotemporal sound analysis by neurons in the auditory midbrain’, *The Journal of Neuroscience* **22**(10), 4114–4131.
- Fairhall, A. L., Lewen, G. D., Bialek, W. and de Ruyter van Steveninck, R. R. [2001], ‘Efficiency and ambiguity in an adaptive neural code’, *Nature* **412**(6849), 787–792.
- Finn, I. M., Priebe, N. J. and Ferster, D. [2007], ‘The emergence of contrast-invariant orientation tuning in simple cells of cat visual cortex’, *Neuron* **54**(1), 137–152.
- Fishbach, A., Nelken, I. and Yeshurun, Y. [2001], ‘Auditory edge detection: a neural model for physiological and psychoacoustical responses to amplitude transients’, *Journal of Neurophysiology* **85**(6), 2303–2323.
- Florentine, M. [1987], ‘Level discrimination as a function of level for tones from 0.25 to 16 kHz’, *The Journal of the Acoustical Society of America* **81**, 1528–41.

- Freeman, T. C., Durand, S., Kiper, D. C. and Carandini, M. [2002], 'Suppression without inhibition in visual cortex', *Neuron* **35**(4), 759–771.
- Fritz, J. B., Elhilali, M. and Shamma, S. A. [2005a], 'Differential dynamic plasticity of a1 receptive fields during multiple spectral tasks', *The Journal of Neuroscience* **25**(33), 7623–7635.
- Fritz, J. B., Elhilali, M. and Shamma, S. A. [2007], 'Adaptive changes in cortical receptive fields induced by attention to complex sounds', *Journal of Neurophysiology* **98**(4), 2337–2346.
- Fritz, J., Elhilali, M. and Shamma, S. [2005b], 'Active listening: Task-dependent plasticity of spectrotemporal receptive fields in primary auditory cortex', *Hearing Research* **206**(1-2), 159–176.
- Fritz, J., Shamma, S., Elhilali, M. and Klein, D. [2003], 'Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex', *Nature Neuroscience* **6**(11), 1216–1223.
- Furukawa, T., Hayashida, Y. and Matsuura, S. [1978], 'Quantal analysis of the size of excitatory post-synaptic potentials at synapses between hair cells and afferent nerve fibres in goldfish.', *The Journal of Physiology* **276**(1), 211–226.
- Gaese, B. H. and Ostwald, J. [2001], 'Anesthesia changes frequency tuning of neurons in the rat primary auditory cortex', *Journal of Neurophysiology* **86**(2), 1062–1066.
- Garcia-Lazaro, J., Ahmed, B. and Schnupp, J. [2006], 'Tuning to natural stimulus dynamics in primary auditory cortex', *Current Biology* **16**(3), 264–271.
- Geweke, J. [1992], Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments, in 'Bayesian Statistics', Vol. 4, Oxford University Press, Oxford, pp. 169–193.

- Gill, P., Woolley, S. M. N., Fremouw, T. and Theunissen, F. E. [2008], ‘What’s that sound? auditory area CLM encodes stimulus surprise, not intensity or intensity changes’, *Journal of Neurophysiology* **99**(6), 2809–2820.
- Gill, P., Zhang, J., Woolley, S., Fremouw, T. and Theunissen, F. E. [2006], ‘Sound representation methods for spectro-temporal receptive field estimation’, *Journal of Computational Neuroscience* **21**(1), 5–20.
- Golub, G. and Reinsch, C. [1970], ‘Singular value decomposition and least squares solutions’, *Numerische Mathematik* **14**(5), 403–420.
- Gourévitch, B., Noreña, A., Shaw, G. and Eggermont, J. J. [2009], ‘Spectrotemporal receptive fields in anesthetized cat primary auditory cortex are context dependent’, *Cerebral Cortex* **19**(6), 1448–1461.
- Goutman, J. D. and Glowatzki, E. [2007], ‘Time course and calcium dependence of transmitter release at a single ribbon synapse’, *Proceedings of the National Academy of Sciences* **104**(41), 16341–16346.
- Green, D. M. [1993], Auditory intensity discrimination, in ‘Human psychophysics (ed. Yost, W.A., Popper, A.N., Fay, R.R.)’, Vol. 3, Springer, New York.
- Grothe, B., Pecka, M. and McAlpine, D. [2010], ‘Mechanisms of sound localization in mammals’, *Physiological Reviews* **90**(3), 983–1012.
- Hack, M. H. [1971], ‘Auditory intensity discrimination in the rat’, *Journal of Comparative and Physiological Psychology* **74**(2), 315–318.
- Heeger, D. J. [1992], ‘Normalization of cell responses in cat striate cortex’, *Visual Neuroscience* **9**(2), 181–197.

- Heeger, D. J., Simoncelli, E. P. and Movshon, J. A. [1996], ‘Computational models of cortical visual processing’, *Proceedings of the National Academy of Sciences* **93**(2), 623–627.
- Heil, P. and Irvine, D. R. [1996], ‘On determinants of first-spike latency in auditory cortex’, *Neuroreport* **7**(18), 3073–3076.
- Heil, P. and Neubauer, H. [2003], ‘A unifying basis of auditory thresholds based on temporal summation’, *Proceedings of the National Academy of Sciences* **100**(10), 6151–6156.
- Heuer, H. W. and Britten, K. H. [2002], ‘Contrast dependence of response normalization in area MT of the rhesus macaque’, *Journal of Neurophysiology* **88**(6), 3398–3408.
- Hosoya, T., Baccus, S. A. and Meister, M. [2005], ‘Dynamic predictive coding by the retina’, *Nature* **436**(7047), 71–77.
- Huang, C. L. and Winer, J. A. [2000], ‘Auditory thalamocortical projections in the cat: Laminar and areal patterns of input’, *The Journal of Comparative Neurology* **427**(2), 302–331.
- Huffman, R. F. and Henson Jr., O. [1990], ‘The descending auditory pathway and acousticomotor systems: connections with the inferior colliculus’, *Brain Research Reviews* **15**(3), 295–323.
- Jeffreys, H. [1961], *Theory of probability*, Clarendon Press.
- Jones, E., Oliphant, T., Peterson, P. et al. [2001], *SciPy: Open source scientific tools for Python*.
- Joris, P. X., Schreiner, C. E. and Rees, A. [2004], ‘Neural processing of amplitude-modulated sounds’, *Physiological Reviews* **84**(2), 541–577.

- Joris, P. and Yin, T. [1992], ‘Responses to amplitude-modulated tones in the auditory nerve of the cat’, *The Journal of the Acoustical Society of America* **91**(1), 215–232.
- Kaplan, E., Purpura, K. and Shapley, R. M. [1987], ‘Contrast affects the transmission of visual information through the mammalian lateral geniculate nucleus.’, *The Journal of Physiology* **391**(1), 267–288.
- Kass, R. E. and Raftery, A. E. [1995], ‘Bayes factors’, *Journal of the American Statistical Association* **90**(430), 773–795.
- 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.
- Kim, K. J. and Rieke, F. [2001], ‘Temporal contrast adaptation in the input and output signals of salamander retinal ganglion cells’, *The Journal of Neuroscience* **21**(1), 287–299.
- Klein, D. J., Depireux, D. A., Simon, J. Z. and Shamma, S. A. [2000], ‘Robust spectrotemporal reverse correlation for the auditory system: optimizing stimulus design’, *Journal of Computational Neuroscience* **9**(1), 85–111.
- Kohn, A. [2007], ‘Visual adaptation: physiology, mechanisms, and functional benefits’, *Journal of Neurophysiology* **97**(5), 3155–3164.
- Kowalski, N., Depireux, D. A. and Shamma, S. A. [1996], ‘Analysis of dynamic spectra in ferret primary auditory cortex. i. characteristics of single-unit responses to moving ripple spectra’, *Journal of Neurophysiology* **76**(5), 3503–3523.
- Kowalski, N., Versnel, H. and Shamma, S. A. [1995], ‘Comparison of responses in the anterior and primary auditory fields of the ferret cortex’, *Journal of Neurophysiology* **73**(4), 1513–1523.

- 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.
- Kvale, M. N. and Schreiner, C. E. [2004], ‘Short-term adaptation of auditory receptive fields to dynamic stimuli’, *Journal of Neurophysiology* **91**(2), 604–612.
- Laughlin, S. B. [1981], ‘A simple coding procedure enhances a neurons information capacity’, *Z. Naturforsch* **36**, 910–912.
- Lee, C. and Middlebrooks, J. C. [2011], ‘Auditory cortex spatial sensitivity sharpens during task performance’, *Nature Neuroscience* **14**(1), 108–114.
- Lesica, N. A. and Grothe, B. [2008], ‘Dynamic spectrotemporal feature selectivity in the auditory midbrain’, *The Journal of Neuroscience* **28**(21), 5412–5421.
- Levitt, J. and Lund, J. [1997], ‘Contrast dependence of contextual effects in primate visual cortex’, *Nature* **387**(6628), 73–76.
- Lewicki, M. S. [2002], ‘Efficient coding of natural sounds’, *Nature Neuroscience* **5**(4), 356–363.
- Liang, L., Lu, T. and Wang, X. [2002], ‘Neural representations of sinusoidal amplitude and frequency modulations in the primary auditory cortex of awake primates’, *Journal of Neurophysiology* **87**(5), 2237–2261.
- Liberman, M. and Guinan, J. J. [1998], ‘Feedback control of the auditory periphery: Anti-masking effects of middle ear muscles vs. olivocochlear efferents’, *Journal of Communication Disorders* **31**(6), 471–483.
- 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, B.-h., Wu, G. K., Arbuckle, R., Tao, H. W. and Zhang, L. I. [2007], ‘Defining cortical frequency tuning with recurrent excitatory circuitry’, *Nature Neuroscience* **10**(12), 1594–1600.
- Luo, S. X., Axel, R. and Abbott, L. F. [2010], ‘Generating sparse and selective third-order responses in the olfactory system of the fly’, *Proceedings of the National Academy of Sciences* **107**(23), 10713–10718.
- Ma, X. and Suga, N. [2001*a*], ‘Corticofugal modulation of duration-tuned neurons in the midbrain auditory nucleus in bats’, *Proceedings of the National Academy of Sciences* **98**(24), 14060–14065.
- Ma, X. and Suga, N. [2001*b*], ‘Plasticity of bat’s central auditory system evoked by focal electric stimulation of auditory and/or somatosensory cortices’, *Journal of Neurophysiology* **85**(3), 1078–1087.
- Machens, C. K., Wehr, M. S. and Zador, A. M. [2004], ‘Linearity of cortical receptive fields measured with natural sounds’, *The Journal of Neuroscience* **24**(5), 1089–1100.
- Malone, B. J., Scott, B. H. and Semple, M. N. [2007], ‘Dynamic amplitude coding in the auditory cortex of awake rhesus macaques’, *Journal of Neurophysiology* **98**(3), 1451–1474.
- Malone, B. J., Scott, B. H. and Semple, M. N. [2010], ‘Temporal codes for amplitude contrast in auditory cortex’, *The Journal of Neuroscience* **30**(2), 767–784.
- Manookin, M. B. and Demb, J. B. [2006], ‘Presynaptic mechanism for slow contrast adaptation in mammalian retinal ganglion cells’, *Neuron* **50**(3), 453–464.
- Mante, V., Frazor, R. A., Bonin, V., Geisler, W. S. and Carandini, M. [2005], ‘Independence of luminance and contrast in natural scenes and in the early visual system’, *Nature Neuroscience* **8**(12), 1690–1697.

- May, B. J. [2000], ‘Role of the dorsal cochlear nucleus in the sound localization behavior of cats’, *Hearing Research* **148**(1-2), 74–87.
- McAlpine, D. [2005], ‘Creating a sense of auditory space’, *The Journal of Physiology* **566**(1), 21–28.
- McDermott, J. and Simoncelli, E. [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.
- Merzenich, M. M. and Reid, M. D. [1974], ‘Representation of the cochlea within the inferior colliculus of the cat’, *Brain Research* **77**(3), 397–415.
- Miller, E. K., Gochin, P. M. and Gross, C. G. [1993], ‘Suppression of visual responses of neurons in inferior temporal cortex of the awake macaque by addition of a second stimulus’, *Brain Research* **616**(1-2), 25–29.
- 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.
- Miller, R. L., Calhoun, B. M. and Young, E. D. [1999], ‘Contrast enhancement improves the representation of / ϵ /-like vowels in the hearing-impaired auditory nerve’, *The Journal of the Acoustical Society of America* **106**(5), 2693–2708.
- Missal, M., Vogels, R. and Orban, G. A. [1997], ‘Responses of macaque inferior temporal neurons to overlapping shapes.’, *Cerebral Cortex* **7**(8), 758–767.

- Moeller, C. K., Kurt, S., Happel, M. F. K. and Schulze, H. [2010], ‘Longrange effects of GABAergic inhibition in gerbil primary auditory cortex’, *European Journal of Neuroscience* **31**(1), 49–59.
- Moser, T. and Beutner, D. [2000], ‘Kinetics of exocytosis and endocytosis at the cochlear inner hair cell afferent synapse of the mouse’, *Proceedings of the National Academy of Sciences* **97**(2), 883–888.
- Mrsic-Flogel, T. D., King, A. J. and Schnupp, J. W. H. [2005], ‘Encoding of virtual acoustic space stimuli by neurons in ferret primary auditory cortex’, *Journal of Neurophysiology* **93**(6), 3489–3503.
- Nagel, K. I. and Doupe, A. J. [2006], ‘Temporal processing and adaptation in the songbird auditory forebrain’, *Neuron* **51**(6), 845–859.
- Nakamoto, K. T., Jones, S. J. and Palmer, A. R. [2008], ‘Descending projections from auditory cortex modulate sensitivity in the midbrain to cues for spatial position’, *Journal of Neurophysiology* **99**(5), 2347–2356.
- Nanou, E., Kyriakatos, A., Bhattacharjee, A., Kaczmarek, L. K., Paratcha, G. and El Manira, A. [2008], ‘Na⁺-mediated coupling between AMPA receptors and KNa channels shapes synaptic transmission’, *Proceedings of the National Academy of Sciences* **105**(52), 20941–20946.
- Nauhaus, I., Busse, L., Carandini, M. and Ringach, D. L. [2009], ‘Stimulus contrast modulates functional connectivity in visual cortex’, *Nature Neuroscience* **12**(1), 70–76.
- 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**, 397–406.
- Nelken, I., Rotman, Y. and Bar-Yosef, O. [1999], ‘Responses of auditory-cortex neurons to structural features of natural sounds’, *Nature* **397**(6715), 154–157.

- Nelson, P. C. and Carney, L. H. [2007], ‘Neural rate and timing cues for detection and discrimination of amplitude-modulated tones in the awake rabbit inferior colliculus’, *Journal of Neurophysiology* **97**(1), 522–539.
- Nomoto, M., Suga, N. and Katsuki, Y. [1964], ‘Discharge pattern and inhibition of primary auditory nerve fibers in the monkey’, *Journal of Neurophysiology* **27**(5), 768–87.
- Olsen, S. R., Bhandawat, V. and Wilson, R. I. [2010], ‘Divisive normalization in olfactory population codes’, *Neuron* **66**(2), 287–299.
- 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.
- Oswald, A. M. and Reyes, A. D. [2008], ‘Maturation of intrinsic and synaptic properties of layer 2/3 pyramidal neurons in mouse auditory cortex’, *Journal of Neurophysiology* **99**(6), 2998–3008.
- Otazu, G. H., Tai, L., Yang, Y. and Zador, A. M. [2009], ‘Engaging in an auditory task suppresses responses in auditory cortex’, *Nature Neuroscience* **12**(5), 646–654.
- Paninski, L. [2004], ‘Maximum likelihood estimation of cascade point-process neural encoding models’, *Network: Computation in Neural Systems* **15**(4), 243–262.
- Park, T. J., Klug, A., Holinstat, M. and Grothe, B. [2004], ‘Interaural level difference processing in the lateral superior olive and the inferior colliculus’, *Journal of Neurophysiology* **92**(1), 289–301.
- Patil, A., Huard, D. and Fonnesbeck, C. [2010], ‘PyMC: Bayesian stochastic modelling in Python’, *Journal of Statistical Software* **35**(4), 1–81.

- Patterson, R. D., Uppenkamp, S., Johnsrude, I. S. and Griffiths, T. D. [2002], 'The processing of temporal pitch and melody information in auditory cortex', *Neuron* **36**(4), 767–776.
- Percaccio, C., Engineer, N., Pruette, A., Pandya, P., Moucha, R., Rathbun, D. and Kilgard, M. [2005], 'Environmental enrichment increases paired-pulse depression in rat auditory cortex', *Journal of Neurophysiology* **94**(5), 3590–3600.
- 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.
- Pfingst, B. E. and O'Connor, T. A. [1981], 'Characteristics of neurons in auditory cortex of monkeys performing a simple auditory task', *Journal of Neurophysiology* **45**(1), 16–34.
- Phillips, D., Judge, P. and Kelly, J. [1988], 'Primary auditory cortex in the ferret (*Mustela putorius*): neural response properties and topographic organization', *Brain Research* **443**(1-2), 281–294.
- Phillips, D. P. and Irvine, D. R. [1981], 'Responses of single neurons in physiologically defined primary auditory cortex (AI) of the cat: frequency tuning and responses to intensity', *Journal of Neurophysiology* **45**(1), 48–58.
- Phillips, D. P. and Sark, S. A. [1991], 'Separate mechanisms control spike numbers and inter-spike intervals in transient responses of cat auditory cortex neurons', *Hearing Research* **53**(1), 17–27.
- Pienkowski, M. and Hagerman, B. [2009], 'Auditory intensity discrimination as a function of level-rove and tone duration in normal-hearing and impaired subjects: The mid-level hump revisited', *Hearing Research* **253**(1-2), 107–115.

- Pillow, J. W., Paninski, L., Uzzell, V. J., Simoncelli, E. P. and Chichilnisky, E. J. [2005], ‘Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model’, *The Journal of Neuroscience* **25**(47), 11003–11013.
- Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E. J. and Simoncelli, E. P. [2008], ‘Spatio-temporal correlations and visual signalling in a complete neuronal population’, *Nature* **454**(7207), 995–999.
- Pollak, G. D., Burger, R. and Klug, A. [2003], ‘Dissecting the circuitry of the auditory system’, *Trends in Neurosciences* **26**(1), 33–39.
- Polley, D. B., Heiser, M. A., Blake, D. T., Schreiner, C. E. and Merzenich, M. M. [2004], ‘Associative learning shapes the neural code for stimulus magnitude in primary auditory cortex’, *Proceedings of the National Academy of Sciences* **101**(46), 16351–16356.
- Polley, D. B., Steinberg, E. E. and Merzenich, M. M. [2006], ‘Perceptual learning directs auditory cortical map reorganization through top-down influences’, *The Journal of Neuroscience* **26**(18), 4970–4982.
- Qin, L., Kitama, T., Chimoto, S., Sakayori, S. and Sato, Y. [2003], ‘Time course of tonal frequency-response-area of primary auditory cortex neurons in alert cats’, *Neuroscience Research* **46**(2), 145–152.
- Quiroga, R. Q., Nadasdy, Z. and Ben-Shaul, Y. [2004], ‘Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering’, *Neural Computation* **16**, 1661–1687.
- Rabinowitz, W., Lim, J., Braida, L. and Durlach, N. [1976], ‘Intensity perception. VI. summary of recent data on deviations from webers law for 1000-Hz tone pulses’, *The Journal of the Acoustical Society of America* **59**(6), 1506–1509.

- Recanzone, G. H., Wurtz, R. H. and Schwarz, U. [1997], ‘Responses of MT and MST neurons to one and two moving objects in the receptive field’, *Journal of Neurophysiology* **78**(6), 2904–2915.
- 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.
- Reiss, L. A. J., Bandyopadhyay, S. and Young, E. D. [2007], ‘Effects of stimulus spectral contrast on receptive fields of dorsal cochlear nucleus neurons’, *Journal of Neurophysiology* **98**(4), 2133–2143.
- Ringach, D. L. [2010], ‘Population coding under normalization’, *Vision Research* **50**(22), 2223–2232.
- Rodriguez, F. A., Chen, C., Read, H. L. and Escabi, M. A. [2010], ‘Neural modulation tuning characteristics scale to efficiently encode natural sound statistics’, *The Journal of Neuroscience* **30**(47), 15969–15980.
- Rose, H. J. and Metherate, R. [2005], ‘Auditory thalamocortical transmission is reliable and temporally precise’, *Journal of Neurophysiology* **94**(3), 2019–2030.
- Rosenzweig, M. [1946], ‘Discrimination of auditory intensities in the cat’, *The American Journal of Psychology* **59**(1), 127–136.
- Rothman, J. S., Cathala, L., Steuber, V. and Silver, R. A. [2009], ‘Synaptic depression enables neuronal gain control’, *Nature* **457**(7232), 1015–1018.
- Rotman, Y., Bar-Yosef, O. and Nelken, I. [2001], ‘Relating cluster and population responses to natural sounds and tonal stimuli in cat primary auditory cortex’, *Hearing Research* **152**(1-2), 110–127.
- Ruderman, D. L. and Bialek, W. [1994], ‘Statistics of natural images: scaling in the woods’, *Physical Review Letters* **73**(6), 814–817.

- Ruggero, M. and Rich, N. [1991], ‘Furosemide alters organ of corti mechanics: Evidence for feedback of outer hair cells upon the basilar membrane’, *The Journal of Neuroscience* **11**(4), 1057–1067.
- Rust, N. C., Schwartz, O., Movshon, J. A. and Simoncelli, E. P. [2005], ‘Spatiotemporal elements of macaque V1 receptive fields’, *Neuron* **46**(6), 945–956.
- Rutkowski, R. G., Shackleton, T. M., Schnupp, J. W., Wallace, M. N. and Palmer, A. R. [2002], ‘Spectrotemporal receptive field properties of single units in the primary, dorsocaudal and ventrostrual auditory cortex of the guinea pig’, *Audiology and Neuro-Otology* **7**, 214–227.
- Sachs, M. B. and Abbas, P. J. [1974], ‘Rate versus level functions for auditory-nerve fibers in cats: tone-burst stimuli’, *The Journal of the Acoustical Society of America* **56**, 1835–47.
- Sadagopan, S. and Wang, X. [2008], ‘Level invariant representation of sounds by populations of neurons in primary auditory cortex’, *The Journal of Neuroscience* **28**(13), 3415–3426.
- Sahani, M. and Linden, J. F. [2003*a*], Evidence optimization techniques for estimating stimulus-response functions, *in* ‘Advances in Neural Information Processing Systems’, Vol. 15, MIT Press, Cambridge MA, pp. 317–324.
- Sahani, M. and Linden, J. F. [2003*b*], How linear are auditory cortical responses?, *in* ‘Advances in Neural Information Processing Systems’, Vol. 15, MIT Press, Cambridge MA, pp. 125–132.
- Sakai, H. M. [1992], ‘White-noise analysis in neurophysiology’, *Physiological Reviews* **72**(2), 491–505.
- Schnupp, J. W. H., Morsic-Flogel, T. D. and King, A. J. [2001], ‘Linear processing of spatial cues in primary auditory cortex’, *Nature* **414**(6860), 200–204.

- Schwabe, L., Obermayer, K., Angelucci, A. and Bressloff, P. C. [2006], ‘The role of feedback in shaping the extra-classical receptive field of cortical neurons: a recurrent network model’, *The Journal of Neuroscience* **26**(36), 9117–9129.
- Schwartz, O. and Simoncelli, E. P. [2001], ‘Natural signal statistics and sensory gain control’, *Nature Neuroscience* **4**(8), 819–825.
- Schwarz, G. [1978], ‘Estimating the dimension of a model’, *The Annals of Statistics* **6**(2), 461–464.
- Sclar, G. [1987], ‘Expression of retinal contrast gain control by neurons of the cat’s lateral geniculate nucleus’, *Experimental Brain Research* **66**(3), 589–596.
- Sclar, G., Maunsell, J. H. and Lennie, P. [1990], ‘Coding of image contrast in central visual pathways of the macaque monkey’, *Vision Research* **30**(1), 1–10.
- 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.
- Shapley, R. and Enroth-Cugell, C. [1984], ‘Visual adaptation and retinal gain controls’, *Progress in Retinal Research* **3**, 263–346.
- Shapley, R. M. and Victor, J. D. [1978], ‘The effect of contrast on the transfer properties of cat retinal ganglion cells.’, *The Journal of Physiology* **285**(1), 275–298.
- Shapley, R. M. and Victor, J. D. [1981], ‘How the contrast gain control modifies the frequency responses of cat retinal ganglion cells.’, *The Journal of Physiology* **318**(1), 161–179.

- Sharma, J., Angelucci, A. and Sur, M. [2000], ‘Induction of visual orientation modules in auditory cortex’, *Nature* **404**(6780), 841–847.
- Simon, J. Z., Depireux, D. A., Klein, D. J., Fritz, J. B. and Shamma, S. A. [2007], ‘Temporal symmetry in primary auditory cortex: implications for cortical connectivity’, *Neural Computation* **19**(3), 583–638.
- Simoncelli, E. P. and Heeger, D. J. [1998], ‘A model of neuronal responses in visual area MT’, *Vision Research* **38**(5), 743–761.
- Simoncelli, E. P., Paninski, L., Pillow, J. and Schwartz, O. [2004], Characterization of neural responses with stochastic stimuli, *in* M. Gazzaniga, ed., ‘The Cognitive Neurosciences III’, MIT Press, Cambridge, MA, pp. 327–338.
- 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**, 3394–3411.
- Sinnott, J., Petersen, M. and Hopp, S. [1985], ‘Frequency and intensity discrimination in humans and monkeys’, *The Journal of the Acoustical Society of America* **78**, 1977.
- Smirnakis, S. M., Berry, M. J., Warland, D. K., Bialek, W. and Meister, M. [1997], ‘Adaptation of retinal processing to image contrast and spatial scale’, *Nature* **386**(6620), 69–73.
- Solomon, S. G., Peirce, J. W., Dhruv, N. T. and Lennie, P. [2004], ‘Profound contrast adaptation early in the visual pathway’, *Neuron* **42**(1), 155–162.
- Spassova, M. A., Avissar, M., Furman, A. C., Crumling, M. A., Saunders, J. C. and Parsons, T. D. [2004], ‘Evidence that rapid vesicle replenishment of the synaptic ribbon mediates recovery from short-term adaptation at the hair cell afferent synapse’, *Journal of the Association for Research in Otolaryngology* **5**, 376–390.

- Steeneken, H. J. M. and Houtgast, T. [1980], ‘A physical method for measuring speech-transmission quality’, *The Journal of the Acoustical Society of America* **67**(1), 318–326.
- Stratford, K. J., Tarczy-Hornoch, K., Martin, K. A. C., Bannister, N. J. and Jack, J. J. B. [1996], ‘Excitatory synaptic inputs to spiny stellate cells in cat visual cortex’, *Nature* **382**(6588), 258–261.
- Sur, M. and Rubenstein, J. [2005], ‘Patterning and plasticity of the cerebral cortex’, *Science* **310**(5749), 805–810.
- Sutter, M. L. [2000], ‘Shapes and level tolerances of frequency tuning curves in primary auditory cortex: quantitative measures and population codes’, *Journal of Neurophysiology* **84**(2), 1012–1025.
- Sutter, M. L. and Schreiner, C. E. [1995], ‘Topography of intensity tuning in cat primary auditory cortex: single-neuron versus multiple-neuron recordings’, *Journal of Neurophysiology* **73**(1), 190–204.
- Taaseh, N., Yaron, A. and Nelken, I. [2011], ‘Stimulus-specific adaptation and deviance detection in the rat auditory cortex’, *PLoS ONE* **6**(8), e23369.
- Tadmor, Y. and Tolhurst, D. [2000], ‘Calculating the contrasts that retinal ganglion cells and LGN neurones encounter in natural scenes’, *Vision Research* **40**(22), 3145–3157.
- Theunissen, F. E., Sen, K. and Doupe, A. J. [2000], ‘Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds’, *The Journal of Neuroscience* **20**(6), 2315–2331.
- Todorovic, A., van Ede, F., Maris, E. and de Lange, F. P. [2011], ‘Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: an MEG study’, *The Journal of Neuroscience* **31**(25), 9118–9123.

- Tollin, D. J. and Yin, T. C. T. [2002], ‘The coding of spatial location by single units in the lateral superior olive of the cat. i. spatial receptive fields in azimuth’, *The Journal of Neuroscience* **22**(4), 1454–1467.
- Turner, R. E. [2010], Statistical models for natural sounds. PhD thesis, University College London.
- Ulanovsky, N., Las, L., Farkas, D. and Nelken, I. [2004], ‘Multiple time scales of adaptation in auditory cortex neurons’, *The Journal of Neuroscience* **24**(46), 10440–10453.
- Ulanovsky, N., Las, L. and Nelken, I. [2003], ‘Processing of low-probability sounds by cortical neurons’, *Nature Neuroscience* **6**(4), 391–398.
- Valentine, P. A. and Eggermont, J. J. [2004], ‘Stimulus dependence of spectro-temporal receptive fields in cat primary auditory cortex’, *Hearing Research* **196**(1-2), 119–133.
- van der Schaaf, A. and van Hateren, J. [1996], ‘Modelling the power spectra of natural images: Statistics and information’, *Vision Research* **36**(17), 2759–2770.
- van Hateren, J. [1993], ‘Spatiotemporal contrast sensitivity of early vision’, *Vision Research* **33**(2), 257–267.
- van Veen, T. M. and Houtgast, T. [1985], ‘Spectral sharpness and vowel dissimilarity’, *The Journal of the Acoustical Society of America* **77**(2), 628–634.
- Varela, J., Sen, K., Gibson, J., Fost, J., Abbott, L. and Nelson, S. [1997], ‘A quantitative description of short-term plasticity at excitatory synapses in layer of rat primary visual cortex’, *The Journal of Neuroscience* **17**(20), 7926–40.

- Versnel, H., Mossop, J. E., Mšic-Flogel, T. D., Ahmed, B. and Moore, D. R. [2002], ‘Optical imaging of intrinsic signals in ferret auditory cortex: responses to narrowband sound stimuli’, *Journal of Neurophysiology* **88**(3), 1545–1558.
- Viemeister, N. [1983], ‘Auditory intensity discrimination at high frequencies in the presence of noise’, *Science* **221**(4616), 1206–1208.
- Vinje, W. E. and Gallant, J. L. [2002], ‘Natural stimulation of the nonclassical receptive field increases information transmission efficiency in V1’, *The Journal of Neuroscience* **22**(7), 2904–2915.
- Walker, K. M. M., Ahmed, B. and Schnupp, J. W. H. [2008], ‘Linking cortical spike pattern codes to auditory perception’, *Journal of Cognitive Neuroscience* **20**(1), 135–152.
- Wallace, M. N., Kitzes, L. M. and Jones, E. G. [1991], ‘Intrinsic inter- and intralaminar connections and their relationship to the tonotopic map in cat primary auditory cortex’, *Experimental Brain Research* **86**(3), 527–544.
- Wang, X., Lu, T., Snider, R. K. and Liang, L. [2005], ‘Sustained firing in auditory cortex evoked by preferred stimuli’, *Nature* **435**(7040), 341–346.
- Watkins, A. J. [2005], ‘Perceptual compensation for effects of reverberation in speech identification’, *The Journal of the Acoustical Society of America* **118**(1), 249–262.
- Watkins, A. J. and Makin, S. J. [2007], ‘Perceptual compensation for reverberation in speech identification: Effects of single-band, multiple-band and wideband noise contexts’, *Acta acustica united with Acustica* **93**(3), 403–410.
- Watkins, P. V. and Barbour, D. L. [2008], ‘Specialized neuronal adaptation for preserving input sensitivity’, *Nature Neuroscience* **11**(11), 1259–1261.

- Webb, B. S., Dhruv, N. T., Solomon, S. G., Tailby, C. and Lennie, P. [2005], ‘Early and late mechanisms of surround suppression in striate cortex of macaque’, *The Journal of Neuroscience* **25**(50), 11666–11675.
- Webb, B. S., Tinsley, C. J., Barraclough, N. E., Parker, A. and Derrington, A. M. [2003], ‘Gain control from beyond the classical receptive field in primate primary visual cortex’, *Visual Neuroscience* **20**(3), 221–230.
- Wehr, M. and Zador, A. M. [2005], ‘Synaptic mechanisms of forward suppression in rat auditory cortex’, *Neuron* **47**(3), 437–445.
- Wen, B., Wang, G. I., Dean, I. and Delgutte, B. [2009], ‘Dynamic range adaptation to sound level statistics in the auditory nerve’, *The Journal of Neuroscience* **29**(44), 13797–13808.
- Williamson, R. S., Anderson, L. A., Christianson, G. B., Sahani, M. and Linden, J. F. [2011], Stimulus density dependence in the auditory thalamus, CoSYNe abstract, Salt Lake City, UT.
- Winer, J. [1992], The functional architecture of the medial geniculate body and the primary auditory cortex, in ‘The mammalian auditory pathway: Neuroanatomy’, Vol. 1, Springer, pp. 222–409.
- Winer, J. A., Larue, D. T., Diehl, J. J. and Hefti, B. J. [1998], ‘Auditory cortical projections to the cat inferior colliculus’, *The Journal of Comparative Neurology* **400**(2), 147–174.
- Winer, J. A., Miller, L. M., Lee, C. C. and Schreiner, C. E. [2005], ‘Auditory thalamocortical transformation: structure and function’, *Trends in Neurosciences* **28**(5), 255–263.
- Winkler, I., Denham, S. L. and Nelken, I. [2009], ‘Modeling the auditory scene: predictive regularity representations and perceptual objects’, *Trends in Cognitive Sciences* **13**(12), 532–540.

- Winkler, I., Teder-Sälejärvi, W. A., Horváth, J., Näätänen, R. and Sussman, E. [2003], ‘Human auditory cortex tracks task-irrelevant sound sources’, *Neuroreport* **14**(16), 2053–2056.
- Winter, I. M. [1991], ‘Intensity coding in low-frequency auditory-nerve fibers of the guinea pig’, *The Journal of the Acoustical Society of America* **90**, 1958–67.
- Winter, I. M., Robertson, D. and Yates, G. K. [1990], ‘Diversity of characteristic frequency rate-intensity functions in guinea pig auditory nerve fibres’, *Hearing Research* **45**(3), 191–202.
- Woolley, S. M. N., Fremouw, T. E., Hsu, A. and Theunissen, F. E. [2005], ‘Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds’, *Nature Neuroscience* **8**(10), 1371–1379.
- Woolley, S. M. N., Gill, P. R. and Theunissen, F. E. [2006], ‘Stimulus-dependent auditory tuning results in synchronous population coding of vocalizations in the songbird midbrain’, *The Journal of Neuroscience* **26**(9), 2499–2512.
- Wu, G. K., Arbuckle, R., Liu, B.-h., Tao, H. W. and Zhang, L. I. [2008], ‘Lateral sharpening of cortical frequency tuning by approximately balanced inhibition’, *Neuron* **58**(1), 132–143.
- Wu, M. C., David, S. V. and Gallant, J. L. [2006], ‘Complete functional characterization of sensory neurons by system identification’, *Annual Review of Neuroscience* **29**, 477–505.
- Xiong, Y., Zhang, Y. and Yan, J. [2009], ‘The neurobiology of sound-specific auditory plasticity: A core neural circuit’, *Neuroscience & Biobehavioral Reviews* **33**(8), 1178–1184.

- Yan, J. and Ehret, G. [2002], ‘Corticotugal modulation of midbrain sound processing in the house mouse’, *The European Journal of Neuroscience* **16**(1), 119–128.
- Yan, J., Zhang, Y. and Ehret, G. [2005], ‘Corticotugal shaping of frequency tuning curves in the central nucleus of the inferior colliculus of mice’, *Journal of Neurophysiology* **93**(1), 71–83.
- Yates, G. [1995], Cochlear structure and function, in B. Moore, ed., ‘Hearing’, Academic Press, San Diego, pp. 41–74.
- 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.
- Yu, J. J. and Young, E. D. [2000], ‘Linear and nonlinear pathways of spectral information transmission in the cochlear nucleus’, *Proceedings of the National Academy of Sciences* **97**(22), 11780–11786.
- Zaghloul, K. A., Boahen, K. and Demb, J. B. [2005], ‘Contrast adaptation in subthreshold and spiking responses of mammalian y-type retinal ganglion cells’, *The Journal of Neuroscience* **25**(4), 860–868.
- Zhang, T. [2005], ‘Boosting with early stopping: Convergence and consistency’, *The Annals of Statistics* **33**(4), 1538–1579.
- Zhou, X. and Jen, P. H. [2005], ‘Corticotugal modulation of directional sensitivity in the midbrain of the big brown bat, *Eptesicus fuscus*’, *Hearing Research* **203**(1-2), 201–215.
- Zhu, C., Byrd, R., Lu, P. and Nocedal, J. [1997], ‘Algorithm 778: L-BFGS-B: fortran subroutines for large-scale bound-constrained optimization’, *ACM: Transactions on Mathematical Software* **23**(4), 550–560.

- Zoccolan, D., Cox, D. D. and DiCarlo, J. J. [2005], ‘Multiple object response normalization in monkey inferotemporal cortex’, *The Journal of Neuroscience* **25**(36), 8150–8164.
- Zurita, P., Villa, A., de Ribaupierre, Y., de Ribaupierre, F. and Rouiller, E. [1994], ‘Changes of single unit activity in the cat’s auditory thalamus and cortex associated to different anesthetic conditions’, *Neuroscience Research* **19**(3), 303–316.