

On The Processing Of Vowels In The Mammalian Auditory System

Christian Honey
St. John's College
Oxford

Thesis submitted in partial fulfilment of the requirements

for the degree of Doctor of Philosophy

at the University of Oxford

ABSTRACT - ON THE PROCESSING OF VOWELS IN THE MAMMALIAN AUDITORY SYSTEM

This thesis was submitted to the Examination Schools at Oxford University in Michaelmas Term 2012, by Christian Honey, of St. John's College, in partial fulfilment of the requirements for the degree of Doctor of Philosophy at the University of Oxford.

The mammalian auditory system generates representations of the physical world in terms of auditory objects. To decide which object class a particular sound belongs to, the auditory system must recognise the patterns of acoustic components that form the acoustic "fingerprint" of the sound's auditory class. Where in the central auditory system such patterns are detected and what form the neural processing takes that underlies their detection are unanswered questions in sensory neurophysiology. In the research conducted for this thesis I used artificial vowel sounds to explore the neural and perceptual characteristics of auditory object recognition in rats. I recorded cortical responses from the primary auditory cortex (A1) in anaesthetised rats and determined how well the spiking responses, evoked by artificial vowels, resolve the spectral components that define vowel classes in human perception. The recognition of an auditory class rests on the ability to detect the combination of spectral components that all member sounds of the class share. I generated and evaluated models of the integration by A1 responses of the acoustic components that define human vowels classes. The hippocampus is a candidate area for neural responses that are specific to particular object classes. In this thesis I also report the results of a collaboration during which we investigated how the hippocampus responds to vowels in awake behaving animals. Finally, I explored the processing of vowels behaviourally, testing the perceptual ability of rats to discriminate and classify vowels and in particular whether rats use combinations of spectral components to recognise members of vowel classes. For the behavioural training I built a novel integrated housing and training cage that allows rats to train themselves in auditory recognition tasks. Combining the results and methods presented in this thesis will help reveal how the mammalian auditory system recognises auditory objects.

Declaration of authorship

I hereby declare that this document is completely and wholly my own work. Contributions by others due to collaborative work are acknowledged where appropriate.

“It is impossible, by the way,

by picking one example of anything,

to avoid picking one which is

atypical in some sense

- that's the wonder of the world.”

(Richard Feynman)

TABLE OF CONTENTS

CHAPTER 1 – GENERAL INTRODUCTION 11

1.1. Motivation 11

1.2. Auditory Object Recognition 13

1.2.1. Classical Perspectives: Gestalten And Streams 13

1.2.2. Relevant And Irrelevant Acoustic Dimensions In Class Formation 16

1.3. Perception Of Speech Sound Classes In Humans And Animals 17

1.3.1. Invariance Of Sound Class Perception 17

1.3.2. Vowel Classes As Auditory Objects 20

1.4. Hearing And Learning Of Speech Sounds In Rats 22

1.4.1. Hearing In Rats 22

1.4.2. Learning Of Speech Sounds In Rats 24

1.5. Neural Correlates Of Auditory Classification 25

1.5.1. Classification In Auditory Cortical Processing Streams? 25

1.5.2. Tuning To Combinations Of Acoustic Components In A1 26

1.5.3. Invariant Responses To Sound Class In Hippocampus 28

1.6. Hypotheses And Methods Explored In This Thesis 29

CHAPTER 2 – RESOLUTION OF FORMANT COMBINATIONS IN RAT AUDITORY CORTEX 33

2.1. Introduction 33

2.1.1. Behavioural Discrimination Of Speech Sounds And Tones In Various Species 33

2.1.2. Neurophysiological Evidence Of Discrimination In The Auditory Cortex 35

2.2. Methods 37

2.2.1. Surgery 37

2.2.2. Electrophysiology 38

2.2.3. Stimuli 40

2.2.4. Data Analysis 42

2.3. Results 43

2.3.1. Responses To Double-formant Vowels 43

2.3.2. Optimal And Sub-optimal Smoothing 49

2.3.3. Average Best Filter Width For Double-formant Vowels 51

2.3.4. Anatomical Distribution Of Optimal Filter Widths 54

2.4. Discussion 55

2.4.1. Cortical Resolution Of Formant Combinations 55

2.4.2. Anatomical Distribution Of Cortical Formant Resolution 56

2.5 Summary And Conclusions 58

CHAPTER 3 – BEHAVIOURAL EXPLORATIONS OF VOWELS BY RATS 59
3.1. Introduction 59
3.2. Methods 61
3.2.1. Animals 61
3.2.2. Stimuli 61
3.2.3. Training 63
3.2.4. Performance Analysis And Estimation Of Confidence Intervals 64
3.2.5. Confidence Intervals And Isocline Analysis For The Estimation Of Class Boundaries 65
3.3. Results 65
3.3.2. Predefined Classification 65
3.3.3. Spontaneous Classification 68
3.4. Discussion 75
3.4.1. Learning Of A Classification Based On Acoustic Similarity 75
3.4.2. Spontaneous Similarity Judgements After Prototype Discrimination 78
3.5 Summary & Conclusions 79

CHAPTER 4 – INTEGRATION OF FORMANT COMBINATIONS IN RAT AUDITORY CORTEX 81
4.1. Introduction 81
4.1.1. Tonotopy And Pure Tone Tuning 82
4.1.2. Linear And Nonlinear Models Of Integration Properties In A1 82
4.2. Methods 86
4.2.1. Surgery And Electrophysiology 86
4.2.2. Stimuli 86
4.2.3. Electrophysiological Recordings 86
4.2.4. Formant Receptive Fields And Detection Of Formant Tuning Peaks 87
4.2.5. Model Fits Of Double-formant Integration 87
4.2.6. Variance Explained By Different Models: Variance Decomposition And Exclusion Criterion 88
4.2.7. Measure Of Model Fit: Akaike Information Criterion 91
4.2.8. Measure Of Linearity: Linearity Index 91
4.3. Results 92
4.3.1. Structure Of Formant Receptive Fields 92
4.3.2. Models Of Double-Formant Responses 96
4.3.3. Improvement By Combining Additive And Multiplicative Models 101
4.3.4. Anatomical Clustering Of Formant Integration 104
4.3.5. Comparison Of FRFs For Onset And Offset Responses 105
4.3.6. Comparing Onset And Offset Responses 108
4.4. Discussion 110
4.4.1. Additive (linear) Versus Multiplicative (non-linear) Formant Integration 111
4.4.2. Diversity Of Integration Properties In A1 113
4.4.3. Offset Responses To Double-formant Vowels Are Less Linear Than Onset Responses 114
4.5. Summary & Conclusions 117

CHAPTER 5 – HIPPOCAMPAL REPRESENTATIONS OF SINGLE-FORMANT STIMULI 119
5.1. Introduction 120
5.2. Methods 122
5.2.1. Ethics Statement. 122
5.2.2. Animals. 122
5.2.3. Stimuli 122
5.2.3. Training Arena 123
5.2.4. Auditory Classification Task. 123
5.2.5. Surgery And Electrophysiological Recordings. 125
5.2.6. Statistical Procedures 127
5.2.7. Statistical Significance Of Response Direction Related Information. 130
5.2.8. Histology 133
5.2.9. Contributions By Different Members Of The Collaboration 134
5.3. Results 134
5.3.1. Neuronal Representation Of Sound Identity 134
5.3.2. Passive Sound Exposure 138
5.3.3. Neuronal Representation Of Response Direction. 142
5.3.4. Relationship Of Sensitivity Across Different Platforms 143
5.3.5. Relationship Of Sensitivity To Sound And Response Direction. 145
5.3.6. Firing Properties Of Putative Interneurons And Pyramidal Cells. 147
5.4. Discussion 148
5.5. Summary & Conclusions 153

CHAPTER 6 – AN AUTOMATED TRAINING ENVIRONMENT FOR RATS 157
6.1. Motivation 157
6.2. Design & Components Of The Training Cage 158
6.3. Functionality Of The Training Cage....162
6.3.1. Registration Of Animals 162
6.3.2. Assignment Of Tasks 162
6.3.3. Training 163
6.3.4 Monitoring Of Training And Behaviour 164
6.5. Chronic Implants - Proof Of Concept....164
6.6. Data Collection In The Training Cage....166
6.6.1. An Easy Discrimination Task 166
6.6.2. Recording Of Neural Data In Awake Behaving Animals 168
6.7. Summary & Conclusions 169

CHAPTER 7 – GENERAL CONCLUSIONS 171

REFERENCES 178

Acknowledgements

I would like to express my deepest thanks for all the support I received from lovely and wise people over the four years leading up this thesis. Thanks go to Stefanie Müller for her love and care for the grumpy bear I was at times, to Katharina Stein and Jyri Stein for giving parental reassurance and for always believing in me, to Paul Honey for his fatherly love, to Trevor Barton, for being a brother in this world, to my supervisor, Jan Schnupp, for his brilliant guidance and friendship, to Ben Willmore for his clear supervision and friendship, to Andrew King for his financial support, to all members of the Auditory Neuroscience Group for the great atmosphere they created, to Ros Gloag for the amazing help with correct spelling and to all my lovely friends in Oxford who were always there, giving me confidence when I lost it and for offering places and times to remember all the things in life that truly count.

CHAPTER 1 – GENERAL INTRODUCTION

1.1. Motivation

The diversity in architecture of sensory nervous systems and how they process sensory input reflects the diversity of environmental pressures under which different species evolved. And their different adaptations share one common function: to generate internal representations of the environment in terms of objects and their interactions. Where in the sensory nervous system object representations are generated and how the underlying neural processes operate remain key questions in sensory neurophysiology.

A number of influential paradigms of the cortical processing of objects have come from studies on vision. We now possess physiologically well-supported models of how retinal input is transformed into object-specific responses in higher visual cortices. A dual stream model of divergent processing for object identity and for location (Haxby et al., 1991) continues to enjoy popularity in the visual research community: *where* and *what* processing streams are thought to emanate from primary visual cortex, and "higher" visual areas, dedicated to the detection of increasingly complex visual features, have been identified. This fundamental paradigm of visual processing rests on the assumption that successive, and hierarchically ordered, visual cortical areas represent visual features of increasing complexity. In this model object classes are represented by the activation of neurons in high level cortical areas such as the hippocampus (Quian Quiroga et al., 2009) and the lateral prefrontal cortex (Freedman et al., 2001). A wealth of electrophysiological evidence to support this paradigm has been collected since Hubel and Wiesel first

described orientation selectivity and ocular dominance in the 1960s (Hubel and Wiesel, 1962; Wiesel and Hubel, 1963).

For the auditory system, analogous models of hierarchical and parallel cortical processing streams have recently been proposed (Lomber and Malhotra, 2008; Romanski et al., 1999). However the evidence for dual processing streams for auditory input is at best mixed (Schnupp et al., 2012), not least so because the receptive fields of auditory neurons are diverse in space and time (Nelken, 2008) and, even in primary auditory cortex (A1), do not appear to respond selectively to isolated sound features, such as pitch or timbre (Bizley et al., 2009). Moreover the "higher" auditory cortical areas like the superior temporal gyrus (STG) and auditory responsive ventral prefrontal cortex (vPFC) have been shown to tightly follow temporal modulations of communication calls in monkeys, while not being particularly selective to a class (i.e. type) of communication calls (Russ et al., 2008). Such evidence puts into question whether the paradigm of hierarchical processing in separate streams for identity and location, also describes the neural mechanisms that yield *auditory* object perception. To date we do not even possess sufficient evidence to judge whether A1 is directly involved in the detection and discrimination of auditory objects (as has been proposed by Nelken, 2008) or whether such "high level" mechanisms can only be provided at much later stages of auditory and multimodal processing, such as the hippocampal formation (Quian Quiroga et al., 2009).

This thesis discusses the representation of simplified vocal communication sounds in the central nervous system of rats and investigates how rats perceive these objects during behavioural tasks. To investigate how the acoustic components of auditory objects are processed in the cortical nervous system I used a large set of artificial

vowel sounds. Artificial vowels share important characteristics with sounds generated by a variety of animals and natural objects, both acoustically and perceptually. For recognition training, vowels can be grouped into classes based on the relative distances of a number of broad peaks in their amplitude spectrum called formants. Artificial vowels are easy to synthesise, so the researcher can tightly control and easily parameterise the features of the stimulus he is presenting. To explore the perception of auditory objects I trained rats in the discrimination and classification of artificial vowels. Furthermore I used a large set of the same type of stimuli to investigate the cortical integration of spectral components, recording extracellularly from the A1 of naive rats. Rats (*Rattus norvegicus)* are used extensively as a mammalian model in auditory research and have been shown to readily learn various types of phonetic discriminations and classifications.

In the following sections I present the background evidence that inspired the hypotheses tested in the remaining chapters of this thesis. First, I discuss the classical perspectives on auditory object recognition (1.2.), followed by a review of the perceptual properties of phonetic classes (1.3.). A separate section (1.4.) briefly summarises the hearing of rats and their ability to learn phonetic classes. The last section (1.5.) looks at the putative neural correlates of sound class formation in the candidate neural structures explored in this thesis: the primary auditory cortex and the hippocampus.

1.2. Auditory Object Recognition

1.2.1. Classical Perspectives: Gestalten And Streams

In analysing an auditory scene the auditory system has to master two key challenges: to determine how many objects are present and what their type or class is. To

determine the number of sound sources, the auditory system relies on the acoustic features available in the stream of incoming air vibrations. For instance, interaural time and level differences (ITDs and ILDs) are the primary cues to determine the direction of a sound's source. In this manner ITDs and ILDs provide the bases to find out how many sound objects there are in a given auditory scene (McAlpine, 2005). Other acoustic features of sounds, such as common onset, co-modulation and harmonic relation are strong grouping cues for spectral components of a sound indicate whether these components belong to the same external object (Warren, 2008). For example, the spectral components of a dog bark will be harmonically related (integer multiples of a common frequency), co-modulated (change together in frequency and sound level) and will have a common onset (arrive at one ear at the same time). The specific combination of sound components from a single source in turn allows the sensory auditory system to classify the source object.

Perceptual Grouping cues were first described in Gestalt Psychology as a basis for the perception of objects. They were shown to aid the perceptual separation of objects from their background and consequently to help decide how many objects are present in a sensory scene. Interestingly one of the earliest examples of 'grouping' as the basis of the formation of sound classes is melodies. The founders of Gestalt Psychology, von Ehrenfels and Wertheimer, wondered: "when we hear a melody and then, upon hearing it again, [...] what enables us to recognise the melody when it is played in a different key?" (Wertheimer, 1924). With this question they essentially were the first to ask how human auditory perception can recognise different sound waveforms as members of the same sound class - in Wertheimer's example the class of variations over the same melody.

Despite these musical beginnings Gestalt psychologists predominantly focused on visual grouping and derived from their psychophysical data a number of “Gestalt Principles” that the sensory nervous system may use as heuristics for the separation of the incoming acoustic stream into discrete sound sources. For instance, a simple Gestalt principle is grouping by proximity, whereby acoustic components that are physically similar or close to each in location, tend to group perceptually. Acoustic grouping cues like the ones mentioned above (e.g. common onset) were later integrated into the framework of Gestalt Principles (Bregman and Campbell, 1971).

Rigorous psychophysical studies on Gestalt-type auditory grouping were first conducted in the early 1950’s by researchers such as Miller and Heise who quantified the “trill threshold” (Heise and Miller, 1951). A musical trill ceases to be perceived as one trill “object” when the relative frequency separation between the two involved tones exceeds a certain threshold (+/- ~15%). After surpassing the threshold two separate sound objects (the constituent tones of the trill) are perceived. Albert Bregman (Bregman and Campbell, 1971) described the trill threshold as “a phenomenon in which a single rapid sequence of tones seems to "break up" perceptually into two or more parallel sequences, as if two or more different instruments were playing”. Bregman and Campbell (1971) called this separation of sensory input the “primary auditory stream formation”. They used the term ‘auditory stream’ instead of ‘auditory object’, but Bregman clarified in his most influential book “Auditory Scene Analysis” (Bregman, 1994) that “the auditory stream plays the same role in auditory mental experience as the object does in the visual”. Bregman’s studies on auditory stream formation showed that one important *streaming cue* (or principle of auditory grouping) is the *similarity in frequency* of different tones in tone sequences: tones that are similar in frequency tend to group perceptually. Timbre, the

distribution of energy over a broad spectrum of simultaneously emitted frequencies, was a little also shown to be a streaming cue for complex tones - the class of sounds that includes vowels (Wessel, 1979).

In summary, streaming cues are important heuristics for the auditory system to extract different sound sources out of the incoming acoustic stream. The separation of sound sources is deeply intertwined with their classification. When listening to an orchestra, for instance, streaming cues (such as common onset and harmonic relation) determine the number of different instruments we hear. Some streaming cues, like timbre and fundamental frequency are at the same time cues that allow the classification of individual instruments. In turn, for each separate instrument, it depends on potentially different streaming cues what class of musical objects we hear - trills or melodies.

1.2.2. Relevant And Irrelevant Acoustic Dimensions In Class Formation

Which class a sound belongs to is not *a priori* defined by acoustic cues. Whether we classify a sound as a vocalisation, a *bark*, or an *angry dog* depends on the context. Consequently, the circumstances decide which acoustic components of a sound are relevant for the attribution of a sound class. For example once we know we are dealing with the class of sounds called “vowels” we see, or rather hear, that the perception of a particular vowel class (/a/,/e/,/i/ ...) is insensitive to changes along a number of irrelevant acoustic dimensions, such as direction, loudness or fundamental frequency, while being very sensitive to changes along other, *relevant* acoustic dimensions: the positions in frequency of the vowel’s spectral peaks (I will discuss the structure of vowels in detail in the following section). In the next section on speech sound classification I will discuss the influence of irrelevant and relevant

acoustic dimensions on perceptual sound class formation, using examples from the literature on speech sound perception.

1.3. Perception Of Speech Sound Classes In Humans And Animals

1.3.1. Invariance Of Sound Class Perception

Good examples of relevant sound dimensions that allow for considerable acoustic variation while class perception does not change can be found among human phonemes. For the syllable classes /ba/ and /pa/, for example, the perceived syllable depends chiefly on voice onset time (VOT), the duration between the time the lips open to produce plosives, like ‘b’ or ‘p’, and the time the vocal folds start to vibrate (Lisker and Abramson, 1964). Although the perception of the syllable’s class varies with VOT, the perceptual change is not linearly related to the acoustic change of VOT: there are, for example, no perceptual intermediates between /ba/ and /pa/, and the change from a perceived /ba/ to /pa/ occurs within a narrow range of VOTs between the centres of these classes. Figure 1.1. is a reproduction of a graph from Leigh Lisker’s earliest study on VOT effects, showing how the perception of /pi/ in the word ‘rupee’ changes with increasing VOT into /bi/ in the word ‘ruby’. This example shows that the perception of some speech sounds is to an extent *invariant* in response to changes along acoustic dimensions that are relevant to class assignment, in this case VOT. Between the two syllables /bi/ and /pi/ in this example there is again no perceptual intermediate.

A crucial property of auditory object perception is that, although changes along a relevant sound dimension are often detectable by the listener, the change remains largely irrelevant to the perception of class membership.

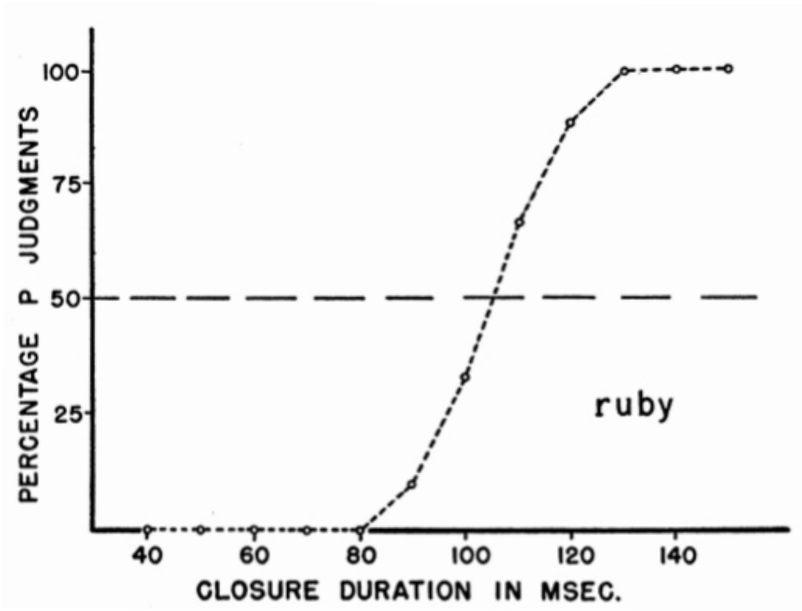


Figure: 1.1. Categorical Perception.

Perception of the syllables /bi/ and /pi/ in the words ruby and rupee. Change in perception from one syllable to the other is caused by the difference in voice onset time after the stop consonants 'b' and 'p'. No intermediates are perceived between the syllables. Reproduced from Lisker (1957). This effect was later named 'categorical perception' by Alvin Liberman (Liberman et al., 1961)

Invariance of perception in response to varying acoustic features leads to a particular tiling of continuous acoustic space into a mosaic of perceptual objects. Among speech sounds, a tiling of this kind is particularly prominent for vowels. Vowels are stacks of harmonically related tones (harmonics) above a fundamental frequency (their pitch), with a spectral envelope that imparts peaks of energy on the harmonics. The spectral peaks are called "formants" and the position of the first two formants (F1 and F2) is the most relevant dimension to the perception of human vowel classes. Shifts in the position of articulators (tongue, soft palate and lips) shift the position of the formants in frequency, which allows humans to produce a wide range of formant combinations. Figure 1.2. shows, in the top plot, the spectrograms of 4 natural human vowels within typical English syllables and, in the bottom plot, the distribution of

formant frequencies for English vowels first published by Peterson and Barney (1952).

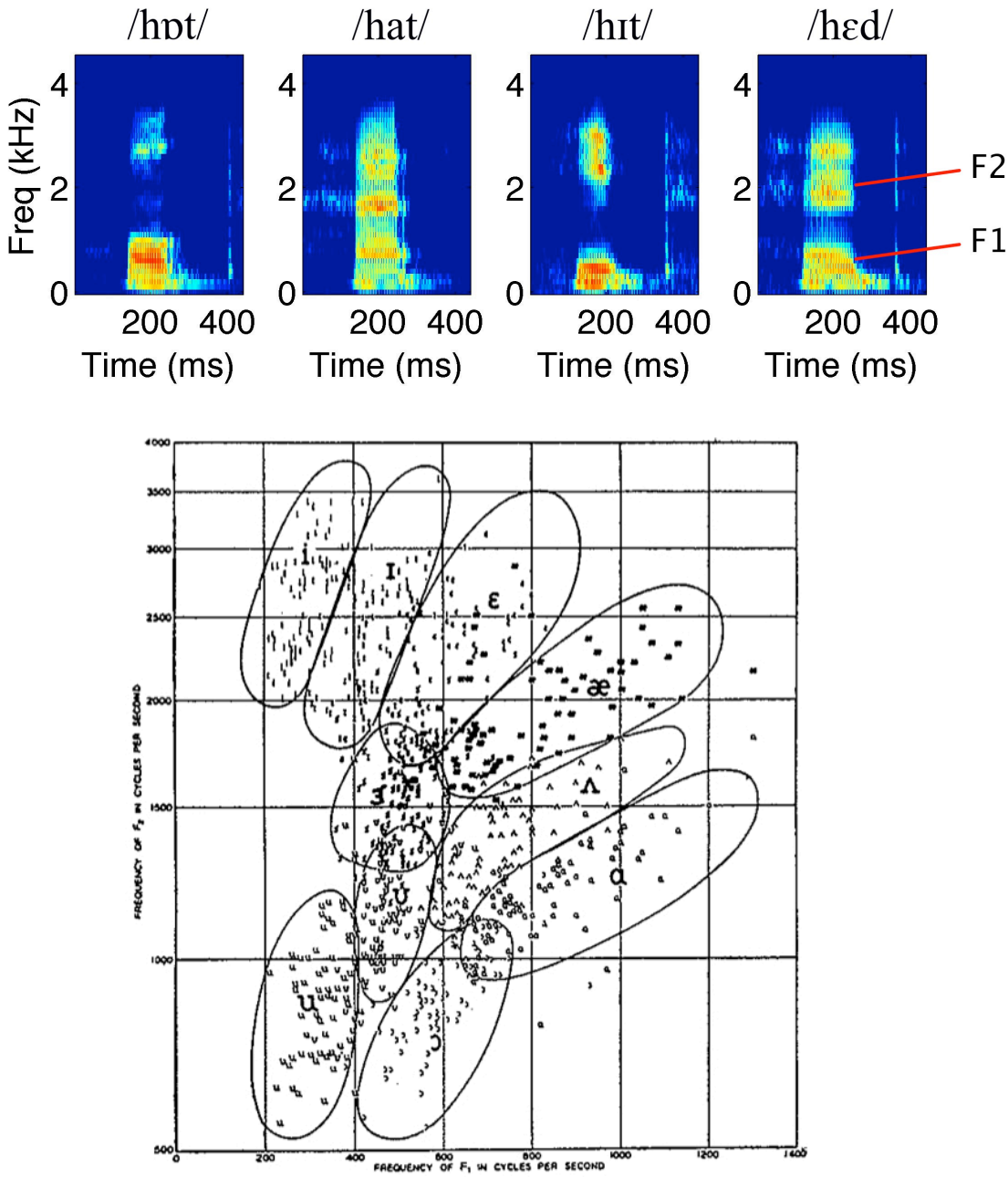


Figure 1.2. Formant Combinations Delineate Human Vowel Classes.

(Top) Spectrograms of four natural human vowels in the syllables /hɒt/, /hat/, /hɪt/ and /hɛd/ as in the British English words: hot, hat, hit and head. The locations of the formants F1 and F2 shift from vowel to vowel. (Bottom) The space of human English vowel formant frequencies F1 (on x-axis) and F2 (on y-axis). Note that there are no perceptual intermediates between vowel classes. From Peterson and Barney (1952)

Although humans perceive vowels as distinct classes with sharp perceptual boundaries and no perceptual intermediates they are, strictly speaking, not perceived *categorically*. The strict definition of *categorical perception* requires that phonemes near a class boundary should be easier to discriminate than phonemes near the centre of a class boundary (Liberman et al., 1961). This characteristic has however not been observed for vowels. Consequently, throughout the this thesis I will use the terms ‘class’ and ‘classification’ to refer to the perceptual grouping of sounds and I will qualify this the classification as *categorical* where appropriate, according to the classical definition of categorical perception.

1.3.2. Vowel Classes As Auditory Objects

Humans perceive vowels as members of distinct vowel classes. Patricia Kuhl (Kuhl, 1991) first explored whether human performance in vowel classification tasks depends on the experience of the listener. First, she presented a number of /i/ - type vowels (as in ‘peep’) to adult human listeners and asked them to judge how prototypical these vowels were regarding the /i/ sound they expected in the word ‘peep’. She then defined the /i/ - type vowel that was judged as the most typical example of an /i/ as the prototype (P) of the /i/ - class. Then she chose the vowel that was consistently judged as least prototypical /i/ as the non-prototype (NP) and generated overlapping distributions of tightly spaced /i/ - type vowels around the picked P and NP examples (see Figure 1.3.).

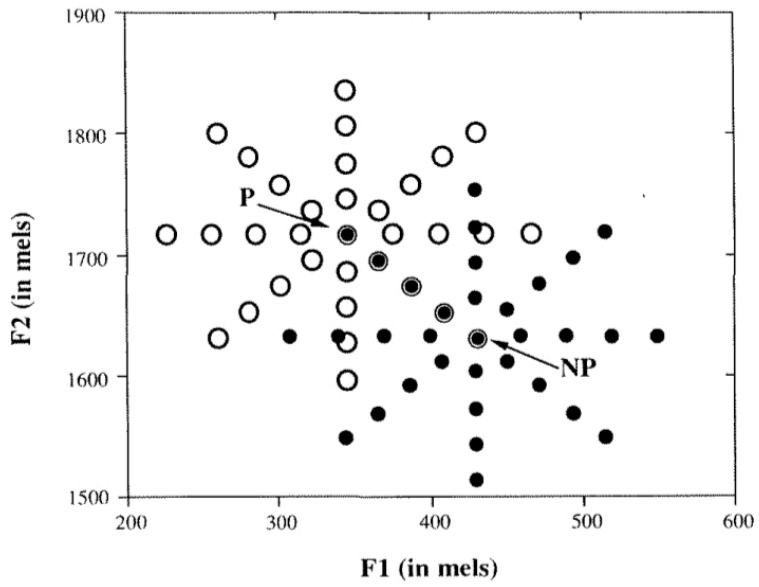


Figure 1.3. The Perceptual Magnet Effect.
Shown is a distribution of formant combinations around a prototype (P) and around a non-prototype (NP) /i/ - vowel as judged by human listeners. Listeners find it harder to discriminate vowels close to the prototypes of their mother tongue than to discriminate vowels close to non-prototypes. From (Kuhl, 1991)

Human listeners were then asked to detect the small changes between members within the P and NP classes. Interestingly, the listeners found it harder to discriminate between members of the prototypical class than between members of the non-prototypical class. This observation suggested that the similarity judgement of class membership depends on the amount of experience a listener has with that class. Kuhl called this tendency a “perceptual magnet effect” and later confirmed that at just 6 months of age, infants already find it harder to discriminate between vowels within the vowel classes of their mother tongue than between vowels of a different language (Kuhl et al., 1992).

Consequently, experience can alter the way auditory class membership is judged when acoustic components change along relevant dimensions: in the case of Kuhl’s study, along the first and second formant frequency. Whether the perceptual magnet

effect extends beyond human linguistic experience to other types of sound and other species is unknown. Nevertheless it has been established that nonhuman animals can be trained to classify vocalisations. This has been demonstrated for starlings (Kluender et al., 1998) and Japanese quail (Kluender et al., 1987; Lotto et al., 1997) using human vowels, for chinchillas (Kuhl and Miller, 1975; Sinnott and Brown, 1997) with the VOT-dependent syllable /ba/ and /pa/ classes and for rats with rise-time dependent /ʃɪp/ and /t͡ʃɪp/ (as in 'ship' and 'chip') classes (Reed et al., 2003) as well as with human vowels (Eriksson and Villa, 2006).

In summary, in humans and many other animal species, vowel classes exhibit the hallmark of auditory classification: invariance of perception in the presence of acoustic change along relevant acoustic dimensions. Auditory classes can be easily defined in the space spanned by the first two vowel formants (as e.g. Kuhl did, see Figure 1.3.). Therefore, vowels are a very useful type of stimulus to explore how animals perceive simple auditory objects, and how they get to be represented in the activity of the mammalian auditory cortex (AC). With suitably chosen vowels we can investigate how mammals classify auditory objects, and how the mammalian AC responds to such objects. What 'suitably chosen' means depends on the sound hearing abilities of the animal used as a model and on their performance in different types of tasks involving speech sounds.

1.4. Hearing And Learning Of Speech Sounds In Rats

1.4.1. Hearing In Rats

Rats hear in a frequency range of ~500Hz to ~64 kHz and their auditory sensitivity is highest between ~8 kHz and ~32 kHz in albino (Kelly and Masterton, 1977) and hooded rats (Heffner et al., 1994). Also the audiograms of both albino rats and

hooded rats are essentially the same between 1 kHz and 30 kHz (Heffner et al., 1994). Figure 1.4. shows the audiograms of several rat species, including those used in this study.

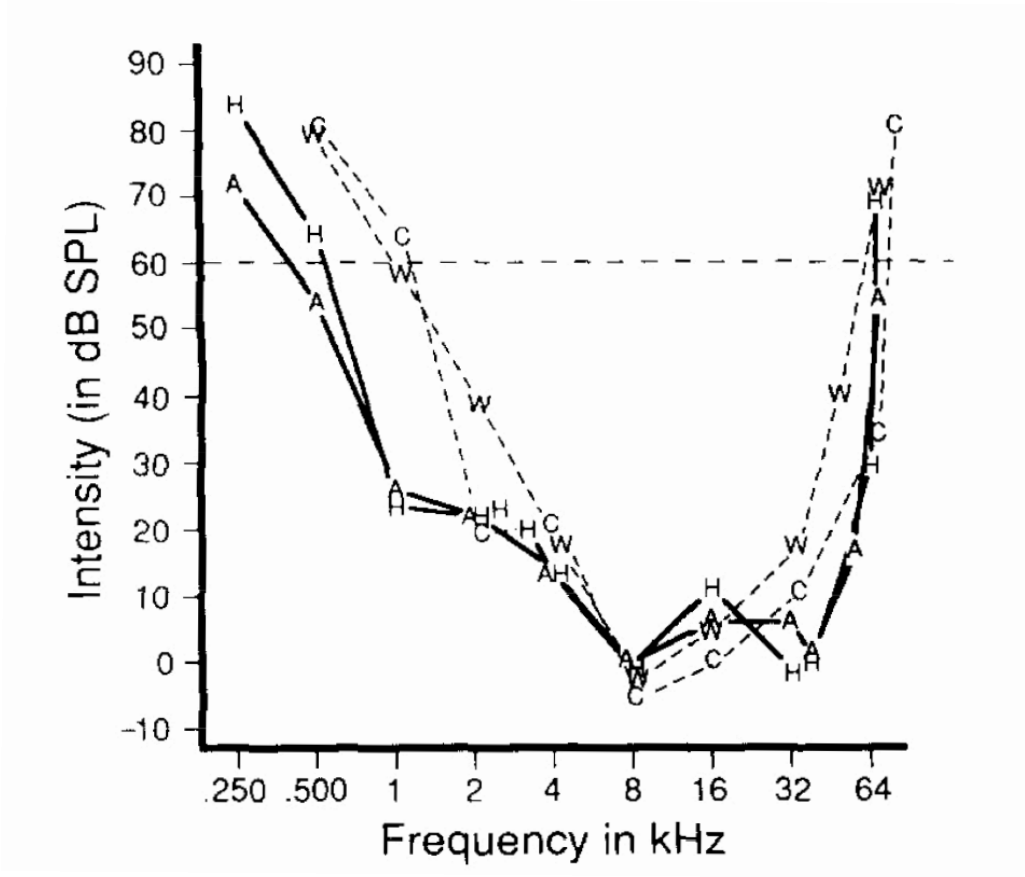


Figure 1.4. Average Audiograms Of 5 Different Rat Strains

The audiograms of hooded (H) and albino (A) rats (the strains used in this thesis) are very similar between 1 and 30 kHZ. The region of best hearing falls between 8 and 32 kHz. From Heffner et al. (1994). Note that all stimuli in this thesis were played at 75dB and had a formant frequency range of 1 - 30 kHz.

Talwar and Gerstein (1998) showed that albino rats trained to discriminate pure tones in a go-nogo task had a constant discrimination limens (DL) of 3.06% (Weber Fraction) in a frequency range of 5 kHz to 32 kHz and (Syka et al., 1996) reported a mean DL of 5.7% between 1 kHz and 64 kHz. All stimuli I used for cortical recordings had formant frequencies between 1 kHz and 31 kHz and stimuli for training covered the range of best hearing in rats.

1.4.2. Learning Of Speech Sounds In Rats

Eriksson and Villa (2006) published one of the few studies that used artificial human vowel classes to train rats in a classification paradigm. Their aim was to show how rats generalise over vowel variants with shifted fundamental frequency and slightly shifted formant frequencies. The animals were initially trained to discriminate two /ɛ/-class vowels from two /ɔ/-class vowels and learned this discrimination after 15 sessions with 27 correct answers per session. The two vowel classes differed in the position of the second formant of their member vowels and within each class vowels varied slightly in the position of their fundamental frequency and their three formants. The animals were then tested on a remaining set of eight vowels in each of the two classes and showed the ability to generalise over the new class members. In this manner Eriksson and Villa demonstrated that rats can use the same spectral cues as humans to form vowel classes. Similar evidence for the similarity of speech sound perception in humans and rats comes from a study by Reed et al. (2003). The authors trained rats to classify the sounds /tʃɑ/ and /ʃɑ/ (as in 'sha' and 'cha'). In this classification the relevant acoustic dimension is the rise time (the onset slope of the sound). Interestingly, after rats had successfully generalised over novel sounds from each class, they showed similar generalisation to non-speech sounds that were defined on the relevant stimulus dimension of rise time. Both speech and non-speech classes had formed sharp categorical boundaries, and the same observations were made in human subjects in a separate experiment, indicating that categorical perception may be a characteristic of auditory object perception that extends to non-speech classes.

In summary, speech sounds, and in particular vowels, can be classified by rats. This result supports my choice to study the representation of auditory object classes in rats using vowel sounds as a simplified object class.

1.5. Neural Correlates Of Auditory Classification

It is to date still unclear which parts of the auditory system generate the effects of class perception: invariance of class judgments in response to varying acoustic components, and the perception of similarities between sounds based on combinations of acoustic components. A starting point in this exploration is the attempt to find neural structures, the activity of which either represents such perceptual effects or is required for these effects to occur in behaving animals. Here I review the most important electrophysiological evidence for auditory class processing in candidate areas of the mammalian auditory system.

1.5.1. Classification In Auditory Cortical Processing Streams?

That separate streams for object recognition and localisation, similar to those in the visual system (Haxby et al., 1991), may exist in auditory system had first been proposed by Romanski and colleagues, based on tracer studies in the macaque (Romanski et al., 1999). The auditory cortex of higher mammals contains a number of tonotopic areas that are typically divided into a primary (core) and a number of higher (belt) regions with increasingly complex response properties (Rauschecker et al., 1997). More experimental support for separate processing streams was presented by Tian et al. (2001), who found that neurons in anterior belt regions respond more selectively to sound patterns, whereas cells in posterior belt regions display a stronger preference for sound source direction. Most recently, direct evidence has been brought forward for the separation of auditory processing into

dual streams in AC: Lomber and Malhotra (2008) showed that reversible deactivation of the anterior auditory field (AAF), a belt area of AC in the cat, impairs the ability of cats to discriminate (and therefore recognise) patterns of noise bursts while leaving their ability to locate such stimuli intact. Conversely, inactivation of the posterior auditory field (PAF) impairs the ability of cats to locate noise bursts while not affecting pattern recognition. Recent studies, analysing the human MEG M100 component, showed that speech sounds (mostly vowels) are represented tonotopically with respect to their main frequency components (Diesch & Luce, 1997; Obleser, Elbert, Lahiri, & Eulitz, 2003; Obleser, Lahiri, & Eulitz, 2003; Shestakova, Brattico, Soloviev, Klucharev, & Huotilainen, 2004).

1.5.2. Tuning To Combinations Of Acoustic Components In A1

Some studies show that neurons in all auditory fields of the macaque cortex (core, anterior and posterior belt) encode changes in spatial location (Harrington et al., 2008), and most cells in the analogue regions of the ferret exhibit sensitivities to combinations of source direction, pitch and timbre of artificial vowels (Bizley et al., 2009). These observations are evidence for sensitivity to combinations of acoustic components in A1 and argue for a role of A1 in the processing of auditory objects that goes beyond a simple representation of tonal frequency. This view is emphasised by studies on A1 responses to broadband stimuli. For instance Bar Yosef et al. (2002) found that many A1 neurons in the cat respond very differently to the main, narrowband, sound (e.g. a bird chirps) in a natural soundscape (e.g. a forest scene) when the very low energy background “noise” is removed from the presented sound. This shows that at least some A1 neurons integrate the spectral components of broadband sounds in a nontrivial (that is to say non-linear) manner.

Some researchers have suggested that not individual cells in A1 but the simultaneous activity of large populations of A1 neurons represents auditory objects. The first evidence in favour of this view was reported by Frank Ohl and Henning Scheich (Ohl et al., 2001). They trained Mongolian gerbils in a sound classification task using short pure tone sweeps that moved either upward in frequency or downward. Here the relevant acoustic dimension for classification was direction of movement in frequency. The gerbils were trained inside a conditioning cage, one half of which was a safe zone after upward tone sweeps and the other one as a safe zone following downward sweeps. If, after a tone sweep was played, the animal did not stay inside or move into the corresponding safe zone, it received a mildly aversive electric foot shock. The gerbils had a cortical surface EEG implant and EEG responses over the whole of A1 were monitored during training. The improvement in the animals' discrimination performance not continuous with increased trial numbers, but jumped rather abruptly from low to high after a few initial training sessions. This 'eureka' moment was accompanied by the emergence of a new response pattern in the recorded cortical EEG, which emerged roughly a second after the onset response to the tone sweeps. These late EEG patterns turned out to be highly informative about the class of the played tone sweeps. Distributed response patterns, which appear to be associated with particular stimulus classes, have also been described in spiking data of single neurons recorded in layer 5 of A1 of anaesthetised rats (Luczak et al., 2009). Here, a 4-shank probe with tightly spaced recording sites was used to record spontaneous and driven responses at four different cortical positions simultaneously in response to a variety of pure tone and cricket chirp stimuli. Luczak and his colleagues found that the recorded neurons fired in a stereotypical order that was preserved in both the spontaneous and the sound-driven regime. The firing rates over the window of activation however changed from stimulus to stimulus in such a

way as to show persistent activity patterns over the whole population. These patterns were predictive of both individual stimuli (e.g. tones of different frequency) and stimulus class (chirp or tone).

Despite these studies on simultaneously recorded neural populations, there is little evidence to suggest that individual cells in A1 respond invariantly to acoustically different members of the same auditory object class. One rare example comes from a study in echolocating bats. Firzlaff et al. (2007) report that about 13% of the cells they recorded extracellularly in a high frequency part of A1 responded invariantly to changes in the size of an object presented to their bats. Although this is evidence for invariant neural responses to a stimulus class, the changes to the stimulus were not along a dimension that is relevant to the class, and comparable to for instance changes in direction which do not affect neural responses characteristics such as frequency tuning in A1. Also, echolocation for form detection might involve different ways of sound analysis than passive sound classification. Therefore such properties of A1 cells may be limited to echolocating animals.

1.5.3. Invariant Responses To Sound Class In Hippocampus

Invariant responses to very different representations of particular objects have been observed in neurons of the CA1 region of the human hippocampus (Quian Quiroga et al., 2009). Quian-Quiroga recorded from the hippocampi of awake human patients. He presenting various persons and urban scenes to the patients while they were lying in their hospital bed. Some hippocampal neurons responded for instance only to representations of a particular known person or landmark, irrelevant of whether the person or landmark was represented by an image or a sound (their spoken name) or by letters (their written name). Similar results, were also reported for neurons in the

inferior-temporal cortex of monkeys (Reddy and Kanwisher, 2006) for visual objects. Hippocampal CA1 neurons are also known to represent an animal's location in space, an observation first made by John O'Keefe in the 1970s (O'Keefe, 1976), and this property of hippocampal responses has since become a major field of research (Neves et al., 2008) and it has been argued that hippocampal activity may integrate representations of objects in places (Parkinson et al., 1988; Gilbert and Kesner, 2004).

In summary A1 single units respond to various combinations of sound components, can be extremely sensitive to small variations in their broadband input and may therefore integrate a wide spectrum of frequency components in a non-linear way. Also, activity patterns of larger populations of neurons in A1 faithfully represent individual auditory stimuli and their class. The hippocampal CA1 region houses invariant object representations, including representations of individual auditory objects. These two brain areas (A1 and hippocampus) are therefore candidates for the exploration of neural correlates of auditory class formation.

1.6. Hypotheses And Methods Explored In This Thesis

To my knowledge, no one has probed the discrimination and integration capabilities of rat A1 in response to formant combinations. The research presented in this thesis takes a first step in determining how neural responses in the rat A1 discriminate between double-formant vowels and how they integrate the energy contained in the two formants of these stimuli. The cortical data I present comes from spike recordings of cortical multi-units, collections of neurons surrounding a recording site. I also verified that rats can learn to classify the artificial vowels I use. Rats were trained in a training environment I built, the capabilities of which I will present. Finally, I also

report the results of a collaborative research project into hippocampal representations of vowel sounds in spatial locations. In the following, I briefly state the hypotheses and methods explored in the remaining chapters of this thesis.

Chapter 2:

Humans and animals can learn the classification of sounds that are *discriminable* for them. In Chapter 2 I explore the question of how fine multi-units in rat A1 can discriminate between different formant combinations. I hypothesise that multi-units in A1 will reliably discriminate between formant combinations whose frequency shifts are at a level near the behavioural discriminability thresholds for pure tones.

Chapter 3:

One shortcoming of behavioural studies involving classification training is that class membership is often defined based on the combination of acoustic features that are experienced as relevant by the researcher. With this approach it remains unclear whether rats would use the information available in *combinations of sound components* to judge class membership. I hypothesise that rats can learn to classify double-formant vowels by acoustic similarity of formant combinations and that they will make use of the combination of formants as a cue for similarity judgements, rather than relying on one formant alone.

Chapter 4:

A central question in auditory cortical research is how acoustic components of auditory objects are integrated (or combined) in the responses of cortical neurons. We will see that sensitivity to combinations requires a certain amount of nonlinear integration of sound features. If A1 is involved in the classification of auditory objects,

its neural responses should not be explainable by a purely additive (i.e. linear) model of responses to formant combinations. I hypothesise that the multi-unit responses of rat A1 cannot be exhaustively explained by an additive model of formant integration.

Chapter 5:

The CA1 region of the hippocampus has been reported to house object selective cells, which respond invariantly to different representations (images, sounds, etc.) of the same object. CA1 units also represent locations in space in awake, behaving animals. I hypothesise that in rats, trained to discriminate single-formant vowels, CA1 units in the hippocampus will exhibit sensitivity to different formant frequencies when rats experience these differences in one and the same location.

Chapter 6:

Auditory object classes are often not defined by acoustic similarity alone, as their perception is dependent on experience and context. Training animals in the discrimination and classification of sounds is a frequent approach to elucidate the properties of auditory class learning, both perceptually and neurally. Exploring neural representations of auditory classes and their changes during learning ideally requires the monitoring of both task performance and neural responses to stimuli in awake behaving animals. I here present a novel automated training environment for efficient behavioural training and electrophysiological recordings in rats.

CHAPTER 2 – RESOLUTION OF FORMANT COMBINATIONS IN RAT AUDITORY CORTEX

2.1. Introduction

This chapter explores the ability of multi-units in rat A1 to discriminate between vowels that differ only in the frequencies of their first two formants. How well will cortical responses resolve the acoustic dimensions that are relevant for recognition of vowel classes? This question is important because perceptual discriminability must have a neural correlate and it is not clear how closely A1 will follow changes of the spectra of broadband sounds, such as vowels. Knowledge of the resolution of A1 for double-formant vowels will allow comparisons to perceptual discriminability. These comparisons will in turn inform us about the extent to which A1 responses represent perceptual features such as invariance of a perceived sound class in the face of acoustic changes.

2.1.1. Behavioural Discrimination Of Speech Sounds And Tones In Various Species

The first experiments on the ability of non-human animals to discriminate speech sounds were reported by Shepherd on raccoons and cats in 1911 and 1912 respectively (Shepherd, 1911; 1912). Shepherd trained his animals to respond to their own name but not to that of a number of other animals present in the same room. Since then many different species have been shown to possess the ability to discriminate between human speech sounds, including cats (Dewson, 1964; Hienz, 1996), chinchillas (Burdick and Miller, 1975; Kuhl and Miller, 1975), macaques (Kuhl, 1991; Sinnott, 1991; Sinnott and Brown, 1997; 1998), baboons (Hienz, 1988), budgerigars (Brown, 1988; Dooling et al., 1989), blackbirds and pigeons (Hienz et al.,

1981), Japanese quail (Kluender et al., 1987), starlings (Kluender et al., 1998) and rats (Reed et al., 2003; Toro et al., 2003; Eriksson and Villa, 2006).

However most of these studies were interested in demonstrating the particular species' capacity to learn a speech sound discrimination, rather than in measuring thresholds for just noticable differences. So far, discrimination thresholds for vowel formant changes have only been reported in humans (Kewley-Port and Watson, 1994), macaques (Sinnott, 1991; Sommers, 1992) and cats (Hienz, 1996). In contrast, discrimination thresholds for pure tone stimuli are now well established for many species, including humans (Harris, 1952; Wier et al., 1977), cats (Hienz et al., 1993), (even) goldfish (Fay, 1970), blackbirds and pigeons (Sinnott et al., 1980), macaques (Sinnott, 1985; Prosen, 1990) and rats (Syka et al., 1996; Talwar and Gerstein, 1998; Sloan et al., 2009).

The value of investigating neural and behavioural sound discrimination in the context of classification becomes obvious when we remember that this ability depends crucially on the sound type and on the studied species. A species that performs well in tone discrimination does not necessarily perceive differences between vowel formants and vice versa. For instance humans discriminate tone frequency with Weber Ratios (WR: the just noticeable frequency difference at a given frequency) of ~0.3% in the frequency region near 125 Hz (Harris, 1952; Wier et al., 1977) and of ~1% near 8 kHz (Wier et al., 1977). In contrast humans discriminate shifts of the formant frequencies in double-formant vowels at just ~1.4%, in the range between 250 Hz and 2.9 kHz (Kewley-Port and Watson, 1994). Interestingly many non-human animals are considerably *worse* in discriminating between pure tones than between vowel formants. For example macaques have been reported to train to a peak WR of

1.6% for vowels but only to 4.2% for tones near 1.4 kHz (Sommers, 1992). Sinnott (1991) reported that the macaque discrimination thresholds for vowel discrimination were between 2.4% and 3.2% (range: 1 kHz and 2 kHz), but for pure tones thresholds were between 1% and 3%, (in the same range). So the macaques discrimination ability for tones and formants is much more similar than observed in humans. Finally, cats are considerably better at discriminating formants (2.2% near 1.7 kHz) than at discriminating pure tones (8.9%), according to Hienz (1996).

In hooded rats, Weber ratios for pure tone discrimination range between 3.7% and 7.3% for frequencies between 1 kHz and 64 kHz with a mean of 5.7% over that range (Syka et al., 1996). These values are similar to those observed in albino rats (3.06 +/- 0.44%) (Kelly and Masterton, 1977; Talwar and Gerstein, 1998). Formant discrimination thresholds have not yet been reported for rats, neither for behavioural nor for neural discrimination. The research in this chapter sets out to test whether the A1 of rats has the capacity to discriminate the formant combinations I used during classification training. Knowledge of the neural discriminability will aid defining “well-chosen” vowel classes behavioural training in rats, with class members that are, at the least, neurally discriminable. In the following section I summarise what is known about the discrimination capacities of AC for different sound types in humans and other species.

2.1.2. Neurophysiological Evidence Of Discrimination In The Auditory Cortex

Pavlov (Pavlov, 1927) was the first to suggest that slightly different sounds will activate the same cortical representations during a perceptual failure to discriminate. In this manner Pavlov attempted to explain his observation that dogs generalise over tones that are close in frequency to a prototype tone the dogs had learned previously.

Early electrophysiological studies on the relationship of auditory cortical activity and perceptual discrimination were limited to ablation studies, mostly in cats (Meyer and Woolsey, 1952) and dogs (e.g. Allen, 1945). These studies produced conflicting results, with some of them indicating that animals could learn sound discrimination after removal of AC (Allen, 1945; Meyer and Woolsey, 1952). Thompson (1960) later showed that the type discrimination task was a decisive variable in studies of the effect of the ablation of AC areas behavioural discrimination of sounds: change detection tasks (with aversive reinforcement after a missed tone change) were not affected while operant conditioning tasks (with positively and negatively reinforced stimulus pairs) were strongly affected. Thompson (1965) was also the first to accurately predict the extent to which cats experienced two tones as the same stimulus (Thompson, 1958) from data on the overlap of the receptive fields of neighbouring cells in A1 (Hind et al., 1960).

Much later, Orduña et al. (2005) reported that “perceptual sensitivities to complex sounds can be related to the degree of similarity in the cortical representations of those sounds.” The authors trained albino rats to discriminate between overlapping FM (frequency modulated) tone sweeps with different rise times. They found that rats were significantly better at discriminating between the slowest FM sweeps than between the fastest FM sweeps. In a separate study they showed that the temporal patterns of activity of multi-units in rat AC reflected this asymmetry in discrimination ability. Neurons in AC responded more dissimilarly in responses to different slow FM sweeps than to different fast sweeps. Other recent studies support the view and in temporal patterns of responses in A1 (and in its bird homologue) reflect the discrimination of speech sounds in zebra finches (Wang et al., 2007), ferrets (Walker et al., 2008) and rats (Engineer et al., 2008).

Keeling et al. (2008) report that well-trained cats detect shifts of ripple spectra at 0.06 octaves (i.e. 4.2% Weber fraction). Ripple spectra are stacks of harmonics with a sinusoidal spectral envelope, similar to envelopes of vowel spectra. That auditory cortical responses vary systematically with shifting ripples had already been reported in the cat (Calhoun and Schreiner, 1998) and ferret (Kowalski et al., 1996; Klein et al., 1998). Keeling et al. (2008) also recorded A1 responses to the same stimuli in trained and untrained cats. Ripple shifts produced systematic changes in the responses of A1 single and multi-units of all, trained and untrained, animals. Additionally A1 responses in trained cats had sharper tuning to pure tones and narrower transfer functions (spectro-temporal receptive fields) in response to ripple stimuli. The authors did not, however, estimate the neural *resolution* (discrimination thresholds) to compare to the animals perceptual resolution of formant shifts.

None of the previous studies on neural or behavioural discrimination of speech sounds, (dynamic) ripple or formants, *quantified* neural discriminability. This aspect of sound discrimination was reserved to behavioural discrimination. In this chapter I quantify how well multi-unit responses in A1 of the anaesthetised rat discriminate formant shifts in double-formant artificial vowels. I estimate the resolution of multi-unit responses and compare it to the behavioural resolution of various sounds observed in rats and other species.

2.2. Methods

2.2.1. Surgery

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. Four naive adult female rats (240-265 g), were used

for extracellular recordings from A1 under terminal general anaesthesia. Anaesthesia was induced by intraperitoneal injection of ketamine (Ketaset; Fort Dodge Animal Health, Overland Park, Kansas, USA) and medetomidine (Domitor, Pfizer, Walton Oaks, Surrey, UK), and maintained with intraperitoneal infusions of ketamine, medetomidine and butorphanol tartrate (Torbugesic; Fort Dodge Animal Health, Overland Park, Kansas, USA), at typical rates of 4.0 mg/kg/h, 16 µg/kg/h and 0.5 mg/kg/h, respectively. Medical Oxygen was administered at 0.5l/h to the animals' snout. The animals core temperature was monitored and kept at 38°C using a rectal temperature probe attached to homeothermic blanket. Heart rate was monitored throughout the experiment and the depth of anaesthesia was confirmed with regular paw and eye-lid reflex tests.

After inducing the anaesthesia, the skull was exposed and a craniotomy of 3-4 mm in diameter was performed over A1 at coordinates (depending on the animals weight) corresponding to bregma -5mm and 7.2mm lateral (right side of the scull) and 4.5 mm ventral in the Paxinos Atlas (Paxinos and Watson, 2005). An additional tiny hole was made over a frontal brain area on the left side of the scull and a silver wire was placed into the subdural space as a reference spike recordings. An example of a craniotomy I performed is shown in Figure 2.1. After removal of the dura mater, recordings from A1 started as described below.

2.2.2. Electrophysiology

For recording from A1 all animals were placed in a soundproofed recording chamber. Electrophysiological responses were recording using a TDT (Tucker Davies Technologies, Alachua, FL) RZ5 digital signal processor connected to neural recording probes via an TDT RA16 preamplifier. Probes were Neuronexus

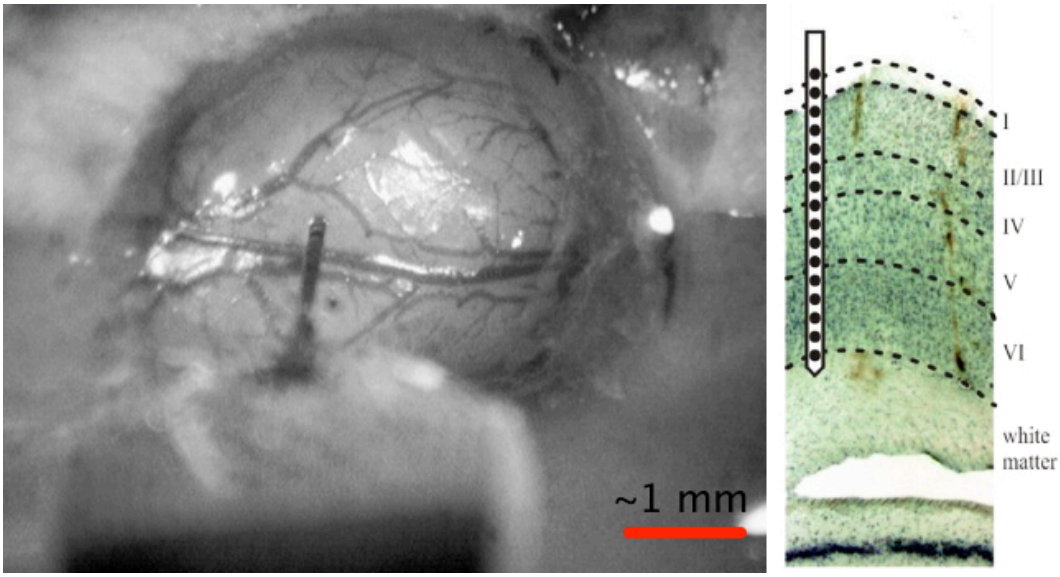


Figure 2.1. Exposed Rat A1 With An Inserted Probe And Schematic Of Electrode Track Through Cortical Layers.

*(**Left**) The exposed A1 of one of the four animals used for this study. The width of the craniotomy is ~3.5mm. (**Right**) Schematic of an electrode track through the layers of auditory cortex using the same type of silicon probe as used in my recordings. The schematic is reproduced from* (Szymanski et al., 2009).

(Neuronexus, AnnArbour, Michigan) A-style silicon probes (1 shank, 16 channels at 100 micron spacing and a surface area of 177 micrometers). Stimuli were played with and RX8 (TDT) digital signal processor and through a Visaton FRS 5 free field loud speaker. Neural signals were recorded with full bandwidth at 24 kHz sampling rate and monitored online with Jan Schnupp's Brainware (TDT). The position of the craniotomy over A1 was confirmed in Brainware based on response latencies and frequency response areas. After the acute recordings all raw data was exported to Matlab (The MathWorks, Inc., Natick, MA, USA) for offline analysis.

2.2.3. Stimuli

Artificial vowel stimuli were generated in Matlab, using Malcolm Slaney's Auditory Toolbox (http://cobweb.ecn.purdue.edu/malcolm/interval/1998-010/). Double-formant vowels were generated with a fundamental frequency of 151 Hz and formant centres at 10 logarithmically spaced positions between 10 kHz and 20 kHz with a spacing of 0.1 octaves between the centres. This resulted in a set of 55 double-formant combinations where the second formant is always larger than the first. Positions of formant centres are shown in Figure 2.2. (left panel) along with the amplitude spectrum of an example stimulus (right panel). All double-formant stimuli were normalised so as to have the same root mean squared amplitude. Each stimulus was played 20 times in pseudorandom order. Stimuli were played at a sound level of 75 dB SPL and sound levels were confirmed using the same condenser microphone (B&K). The same microphone was used to record impulse responses of the speaker to broadband clicks. These were then used to generate compensation filters to compensate for nonlinearities in the loudspeaker's frequency response.

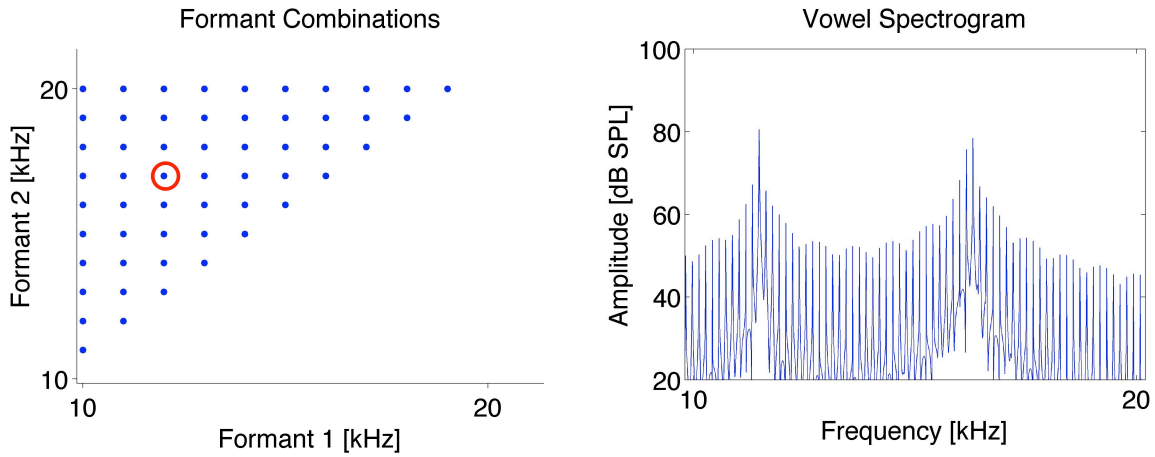


Figure 2.2. Formant Centres Of Artificial Vowel And Example Spectrum.

*(**Left**) Distribution of the 55 formant centres in the double-formant space between 10 and 20 kHz at 0.1 octave spacing. (**Right**) Amplitude Spectrum of the example stimulus marked with a red circle in A. vertical lines. Vertical lines are the harmonics of the fundamental frequency (F0) of 151 Hz.*

Figure 2.3. shows four example waveforms for double formant vowels used in this thesis. The stimulus in the top left panel is the one with lowest F1 (1088 Hz) and F2 (1436 Hz), the one on the bottom right panel is the one with highest F1 (22974 Hz) and F2 (30314 Hz), and the remaining two are intermediate stimuli with F1 (2500 Hz) and F2 (7579 Hz) in the top right panel and F1 (13195 Hz) and F2 (17411 Hz). It becomes obvious that all stimuli beat at the same fundamental frequency and that the first stimulus has the lowest frequency content.

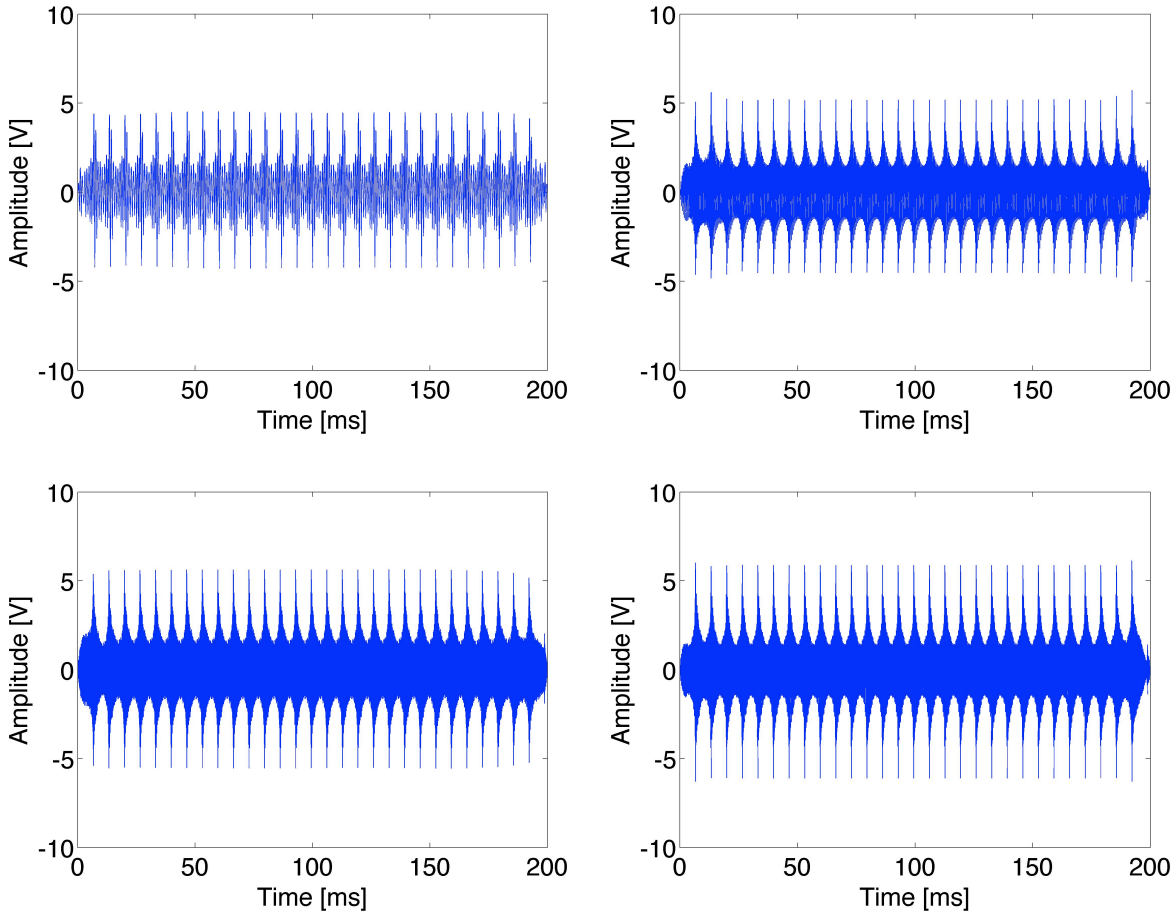


Figure 2.3. Example Waveforms Of The Double-Formant Stimuli Used In This Thesis
From top left to bottom right the frequencies of both F1 (1088; 2500; 13195; 22974 Hz) and F2 (1436; 7579; 13195; 30314 Hz) increase. The stimulus in the top left panel is the one with lowest F1 and F2, the stimulus in the bottom right panel is the one with highest F1 and F2 used in this thesis.

2.2.4. Data Analysis

Neural signals, once exported into Matlab, were bandpass filtered for spike detection between 200Hz and 2 kHz. Spikes were detected in each response trace by finding the negative peaks in the filtered voltage traces and then removing the peaks that did not fall under a level of the median -3.5 times the standard deviation of filtered trace. Spikes were counted in an onset response window (20-100 ms) and converted into onset spike rates. The median spike rate in the baseline window (400-480 ms) over all stimuli and repeats, was chosen as an estimate of the baseline response of the recorded multi-untis. Receptive fields for formant combinations were constructed by average the spike rates of the 20 trials per stimulus. The underlying receptive fields for response to double-formant vowels were approximated by smoothing the observed receptive fields with 2-dimensional Gaussian filters. The optimal strength of smoothing was determined using a cross-validation method that will be explained in the results section for the sake of clarity. Optimal filter width also describes the discrimination ability between neighbouring stimuli. Filter widths, expressed by the standard deviation of the Gaussian filter, were collected for all analysed multi-units. Smoothing and cross-validation methods are described using concrete examples in the presentation of the results.

2.3. Results

2.3.1. Responses To Double-formant Vowels

Figure 2.4. shows a raster plot of spikes recorded in 20 repeats to the same double-formant stimulus and a histogram of the pooled spikes, evoked by all double-formant stimuli. For each trial, subtracting the baseline spike rate from the onset spike rate yields the response strength in Hz. As the response measure for given double-formant stimulus I used the mean spike rate evoked by all repeats of the stimulus. All responses reported here are average spiking rates of multi-units, one per recording site on my silicon probes.

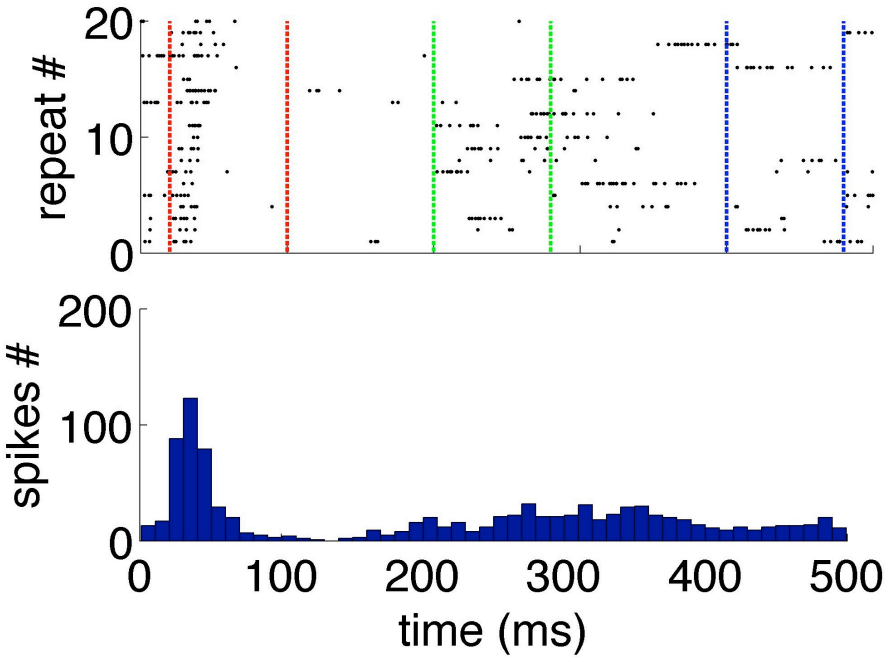


Figure 2.4. Trial-To-Trial Variability And Pooled Responses.

*(**Top**) A raster plot of spike recorded in response to 20 repeats of the same double-formant stimulus at one of the 16 recording sites during one penetration in A1 of the anaesthetised rat. Red line: onset window (20-100 ms), green lines: offset window (200-280 ms), blue lines: baseline window (400-480 ms). (**Bottom**) A histogram of spikes in response to all stimuli at the same recording site and penetration. Note the high trial-to-trial variability in spike count in the onset window.*

From the average spike rates for each double-formant stimulus, I constructed formant receptive fields (FRFs) for each recorded multi-unit (480 in total over four animals). Figure 2.5. shows a number of typical FRFs constructed from the average multi-unit responses to formant combinations. Each pixel in the plots in Figure 2.5. represents the average spike rate inside the onset window for a particular formant combination. Because the frequency of formant 2 is always higher than the frequency of formant 1 the FRF has a triangular shape. In total I recorded responses from 480 recording multi-units from 4 animals. We see that FRFs for formant combinations between 10 and 20 kHz are diverse both in their shape and in the amount of apparent structure.

Response differences between neighbouring stimuli can fluctuate considerably, and it is not clear to what extent these fluctuations represent underlying receptive fields of multi-units or neural noise. Consequently, to arrive at an estimate for the resolution of formant combinations the trial-to-trial variability needs to be taken into account. Trial-To-Trial variability of spike counts in onset responses was high (see Figure 2.4.). Because FRFs are constructed from mean responses over trials the structure of observed FRFs is contaminated with neural noise. Smoothing FRFs with an optimal 2-dimensional smoothing kernel can compensate for fluctuations of mean responses that are due to neural noise and in that manner yield a better representation of the underlying physiological sensitivity function (Hastie et al., 2009). Figure 2.6. shows the effect of smoothing with a 2-dimensional Gaussian filter of a particular width on the structure of the FRF in Figure 2.5. marked with a red star.

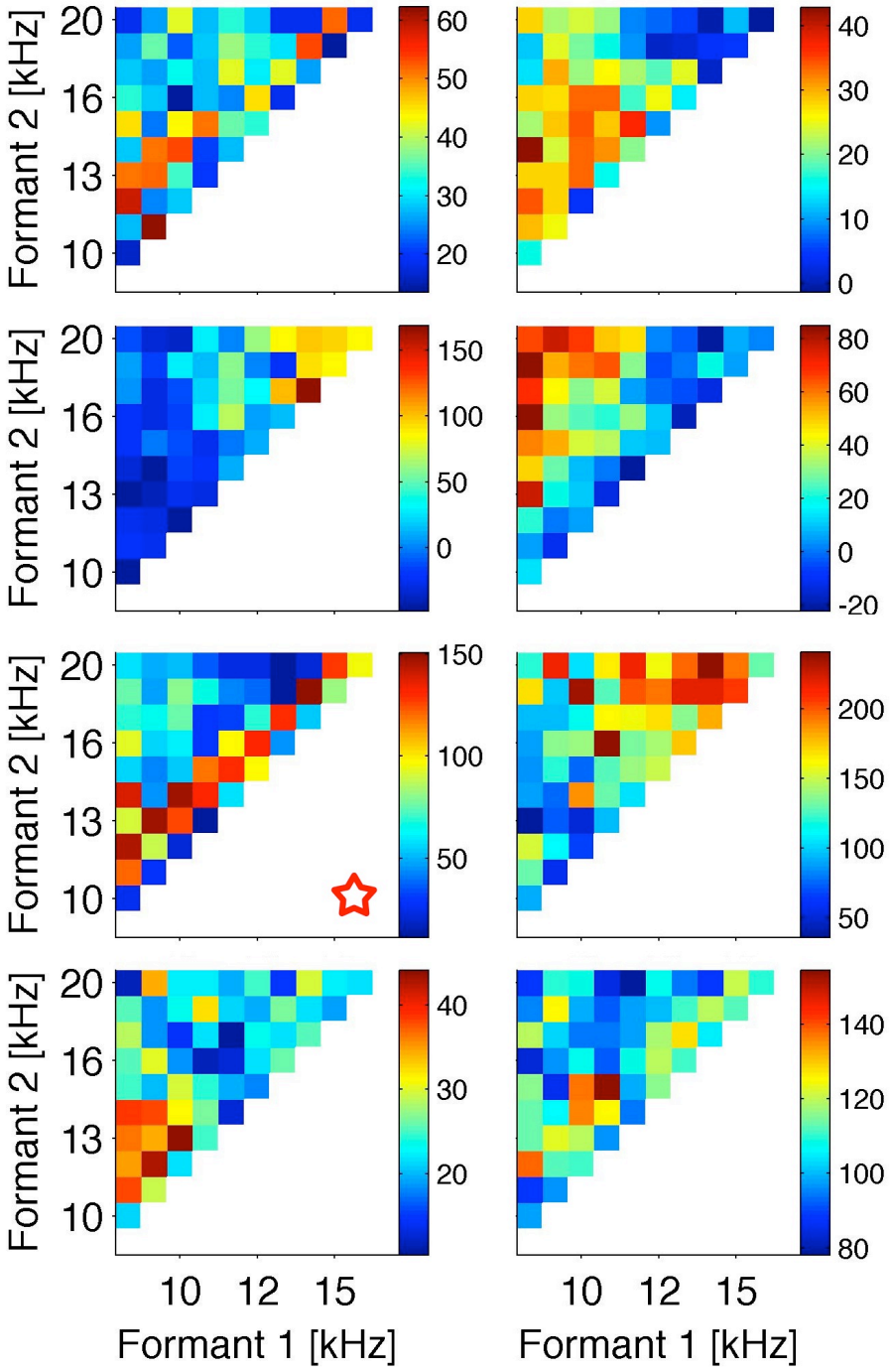


Figure 2.5. Example FRFs For Double-formant Vowels With Formant Centres Between 10 And 20 KHz.

Each pixel is the mean response over 20 trials for that formant combination. Each ***row*** *shows two example FRFs from one of the four animals I recorded from. FRF structure is diverse within and between animals and noise levels appear to differ. I chose the example marked with a star for an illustration of FRF smoothing in Figure 2.6.*

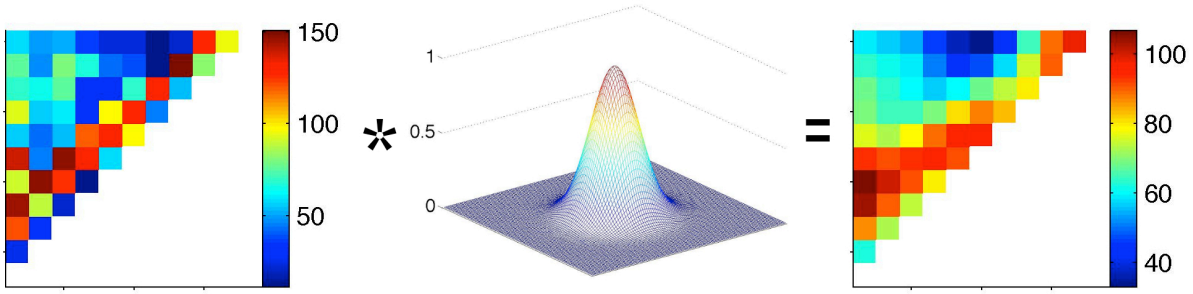


Figure 2.6. Smoothing Of An FRF Using 2-dimensional Gaussian Smoothing Filter. *(**Left**) The original FRF is the one marked with a red star in Figure 2.5. Color coded response values are spike rates in the onset window in Hertz. (**Middle**) A 2-dimensional Gaussian filter with σ = 0.1 octaves standard deviation. (**Right**) The result of convolving the FRF with the smoothing filter. Note how neighbouring pixels blur into each other.*

The high trial-to-trial variability of responses to the same formant combination can lead to differences of mean responses for very similar stimuli. If an unlimited number of trials per stimulus were available these differences would “smooth out” and approach the underlying physiological difference between neighbouring stimuli. Smoothing hopes to achieve a similar result: compensating for the fluctuation of mean responses due to neural noise. In one hypothetical extreme case, trial-to-trial variability would be zero. Differences between mean responses would all be genuine. In this case a filter that is wider than the distance between neighbouring stimuli would wash out the underlying physiological response patterns to different formant combinations and the optimal Gaussian filter would have a width of zero. In another hypothetical extreme case all apparent structure would be due to neural noise. In this case the underlying physiological receptive field would be flat and the optimal Gaussian filter width would be infinity.

If the responses I recorded yield response distributions between these extreme cases I should therefore be able to find an optimal filter with a width that compensates for

the fluctuations of mean responses due to neural noise while not smoothing out the genuine receptive field that would appear in the case of a very high trial number. The width of the optimal filter will then match the width of “genuine” receptive field feature and can therefore serve as a proxy for the resolution with which the neural spike rate code represents changes in formant frequency.

I used a cross-validation method to estimate the optimal filter width based on the trial-to-trial-variability in my recordings. The method is depicted in Figure 2.7. In this method one randomly chosen trial is removed from the sets of repeats of the presented stimuli. Then one FRF is generated from the average responses of the remaining 19 repeats (FRF1) and one from the single removed trials (FRF2). FRF1, smoothed to different degrees, can be thought of as a set of models of the underlying receptive field that can be used to “predict” the outcome observed in single trials (FRF2). The residuals (FRF1-FRF2) quantify the quality of the prediction and I can find the optimal filter width that minimises the residuals over a large number of trials. I used 5000 cross-validation steps and tested the following filter widths (width in octaves: 0.02, 0.04, 0.08, 0.16, 0.32, 0.64, 1.28, 2.56). The cross-validation process, as described in Figure 2.7., can therefore estimate the optimal filter width. The optimally smoothed FRF is our best estimate of the receptive field that would be observed after a very large number of real trials that cancel out the trial-to-trial variability.

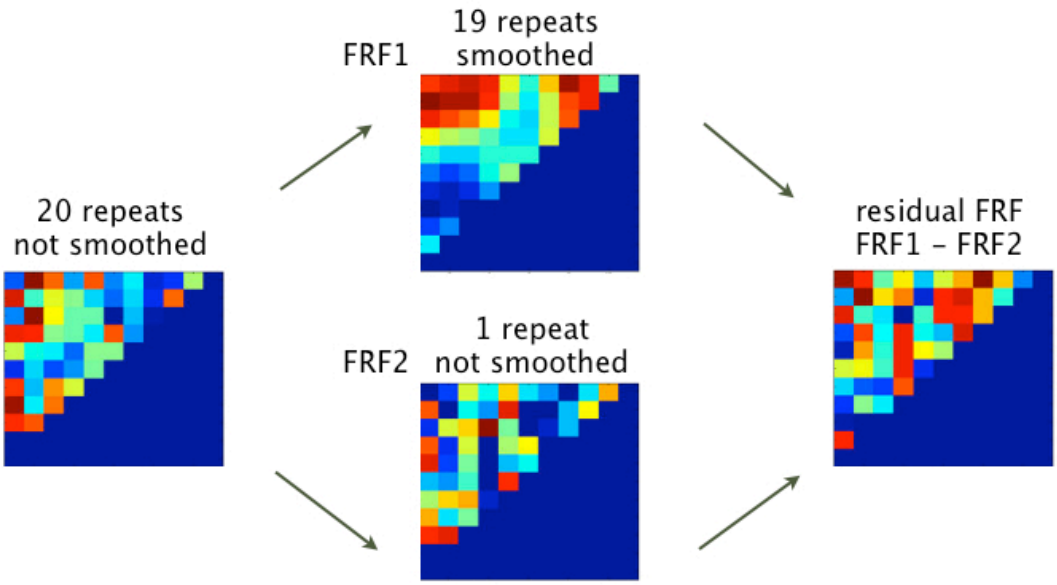


Figure 2.7. The Cross-validation Method.

*(**Left**) Each pixel in the original FRF is the result of averaging over 20 repeats. (**Middle**) The original set of repeats for all stimuli is divided at random in to one set containing 19 repeats per stimulus and one containing one repeat and average FRFs are generated (FRF1 and FRF2). The FRF generated from 19 repeats is smoothed, in this example the smoothing level is at 0.1 octaves. The residual FRF is the difference between the smoothed FRF1 and the single trial FRF2. The distance between FRF1 and FRF2 can be expressed as a cost function. We chose the standard deviation of the residual FRF as a cost function. This process was repeated 5000 times per filter width and the distributions of cost values for each filter were collected. By changing levels of smoothing the residual cost function can be minimized.*

2.3.2. Optimal And Sub-optimal Smoothing

Figure 2.8. shows the FRFs of 4 multi-units at the same penetration with different optimal smoothing (filter widths). In the top row of Figure 2.8.A the observed FRFs are plotted, below we see the same multi-units but smoothed at their respective optimal filter width. Figure 2.8.B shows the development of residuals from the cross-validation with different filter widths. The optimal filter width varies between multi-units. We can see how for all channels in this example residuals first decrease and then increase with increasing filter width. The fact that we can reduce the residuals by smoothing means that we can estimate the underlying FRF of this multi-unit. The filter that yields the lowest residuals is also an estimator of the response differences of neighbouring stimuli. Its width (in sigma) therefore describes the how strongly the analysed multi-unit discriminates between neighbouring stimuli.

Multi-units like those shown in the left-most panel in Figure 2.8.B do not show an obvious increase in residuals after the estimated minimum is reached. In these multi-units the trial-to-trial variability is so large for most stimuli, that the best estimate of single trial outcomes is the grand average over all stimuli and repeats. For the 41 multi-units on which I observed this effect I set the optimal filter width accordingly to the maximum width used in the cross validation method (2.56 octaves). In contrast, for multi-units like the one shown in the third column in Figure 2.8., the initial decrease of residuals is smaller indicating that this multi-unit already discriminates strongly almost at the smallest filter width. This is also apparent in the absence of a visible smoothing effect in 2.8.A. (third column).

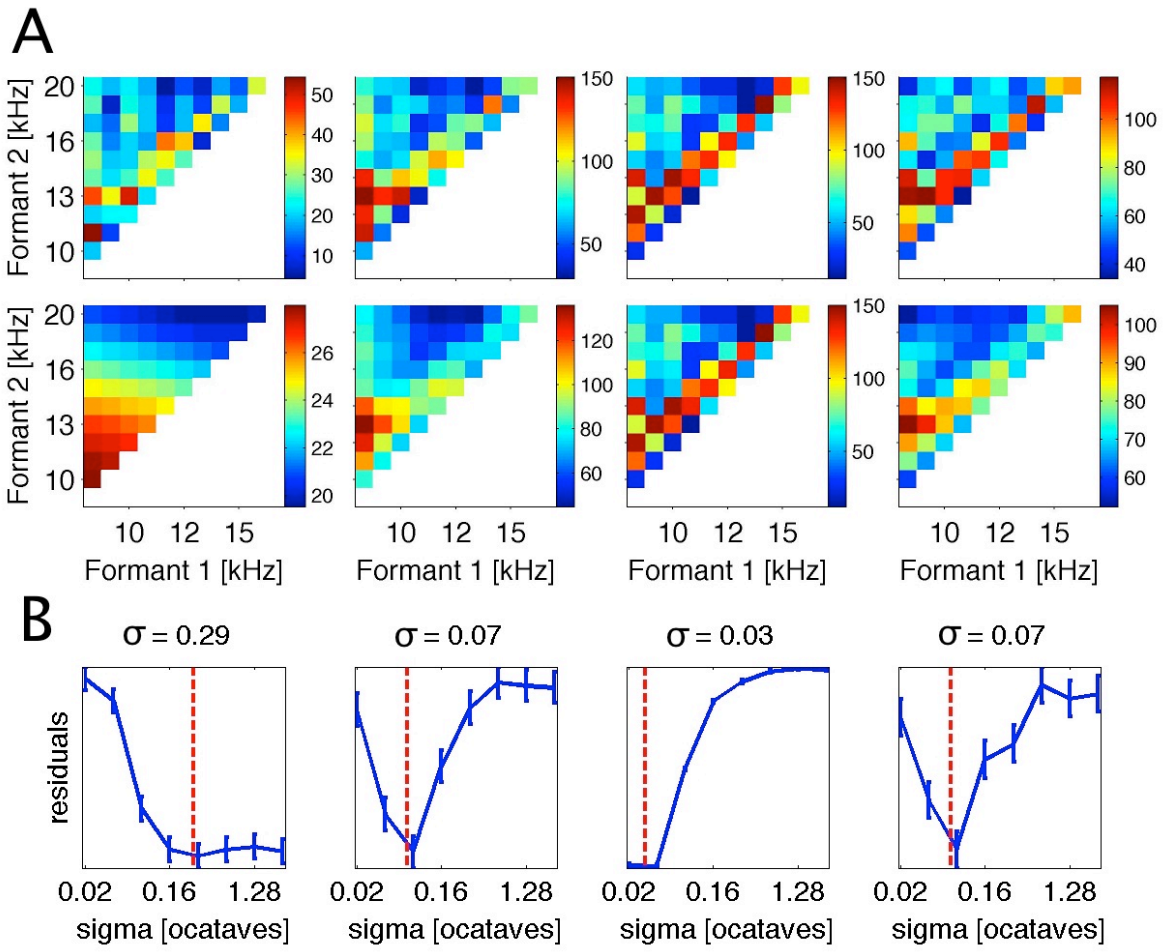


Figure 2.8. Effects Of Smoothing And Determination Of Optimal Filter Widths For Different Multi-units.

(A) *Four example multi-units without optimal smoothing (top row) and with smoothing (bottom row).* ***(B)*** *Results of the cross-validation at the sites in (A). Red lines show the estimated optimal smoothing used in the bottom row in (A). Error bars: s.e.m.*

2.3.3. Average Best Filter Width For Double-formant Vowels

Figure 2.9. shows histograms of the optimal filter widths of all 480 recorded multi-units in the four animals I recorded from. It is obvious that for a majority of multi-units (336/480, or 70%), the optimal filter width settled around 0.1 octaves. The median best filter width over all multi-units is 0.1063 octaves. This corresponds to a difference in formant frequency of roughly 7.6%. The peak of 41 multi-units at the largest filter width (2.56 octaves) represents units for which no optimal filter could be determined with the cross-validation method.

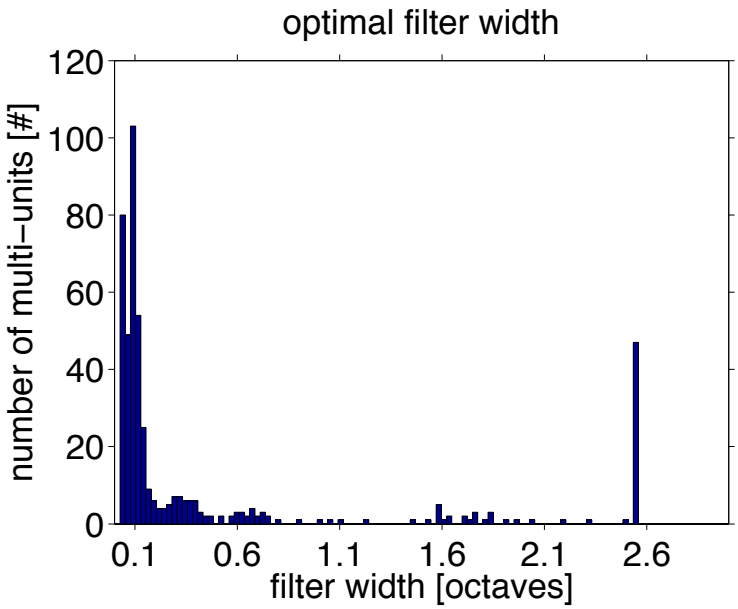


Figure 2.9. Optimal Filter Widths For Smoothing FRFs.

This histogram shows the number multi-units for which smoothing filters were optimal at a particular filter width. 480 multi-units were analysed. The filter width for those multi-units for which no optimal filter width could be determined (41 multi-units) was set to the maximum value (2.56).

I was under the impression that multi-units with a higher average spike rate response also resolved formants more finely, and hence have smaller optimal smoothing bandwidths, compared to less responsive multi-units. Therefore I determined whether there was an obvious relationship between optimal filter width and the median response strength for all sites in all animals. Figure 2.10. shows the relationship between spike rate in the onset window (20-100 ms) and the optimal filter for all 480 multi-units analysed. My impression was wrong; even weakly responsive multi-units had FRFs with small optimal smoothing filters.

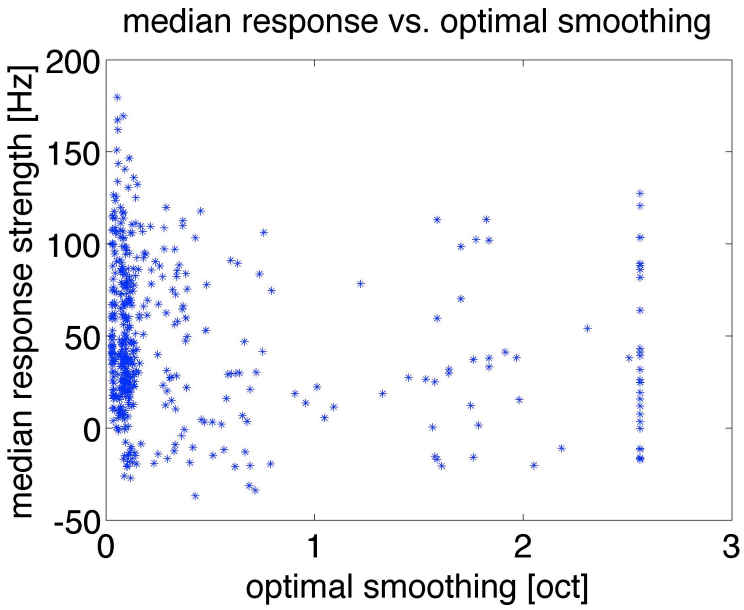


Figure 2.10. Relationship Of Optimal Filter Width And Response Strength.

Median response strength as a function of optimal smoothing I did not find an obvious or significant relationship between the strength of neural responses and the width of the best filter.

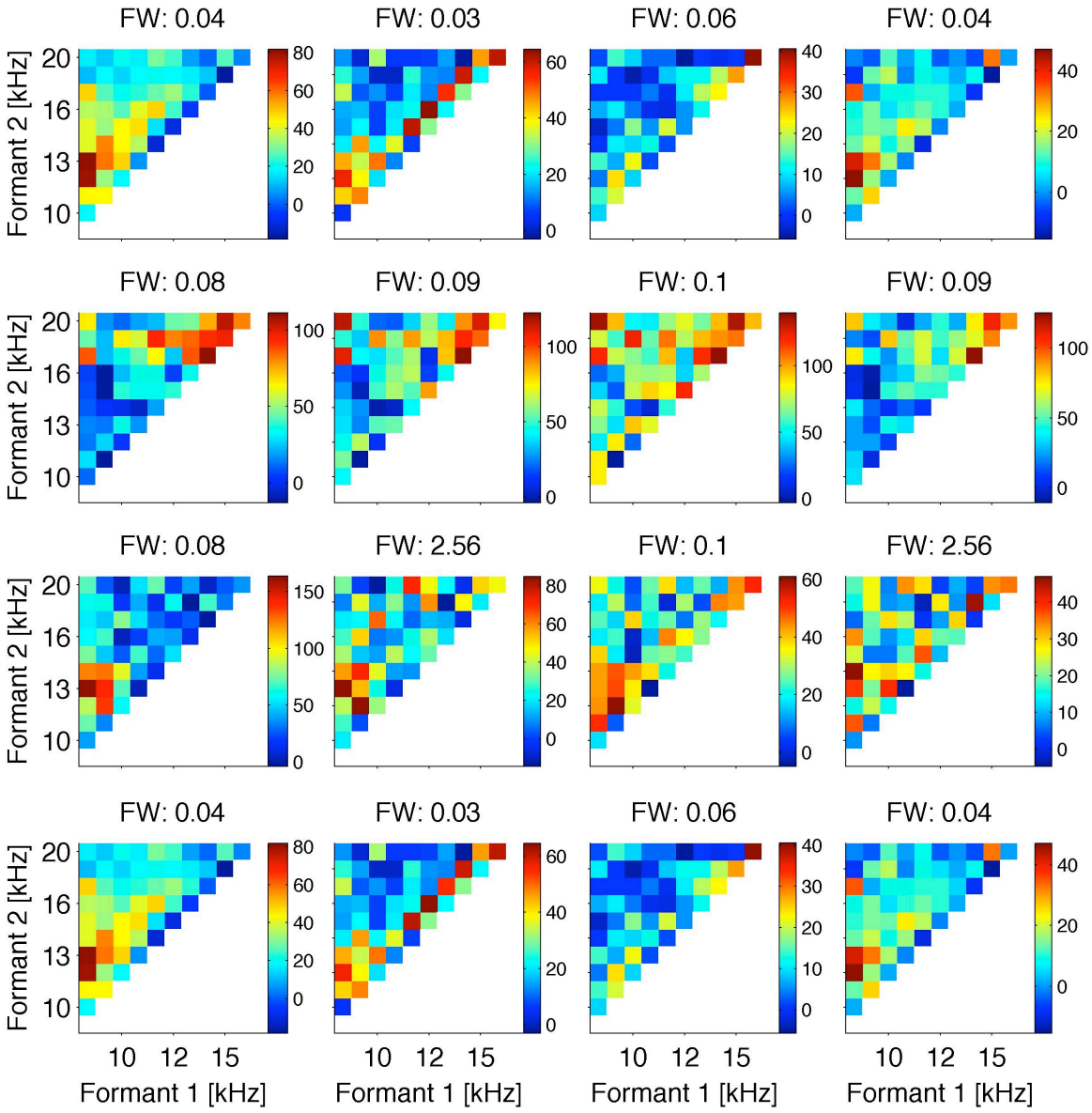


Figure 2.11. Development Of FRF Structure And Optimal Filter Width With Cortical Depth.

*Each **row** shows 4 FRFs (non-smoothed) from one penetration per animal. From left to right the **columns** represent cortical depths at 400, 800, 1200 and1600 microns. Optimal filter widths (FW) for smoothing are shown in the title of each plot. Note that the filter width for units for which the cross-validation produced no minimum was set to the maximum filter width (FW = 2.56). Although responses within penetration are similar there is considerable variation in FRF structure across penetrations. I do not find an obvious relationship between optimal filter width and cortical depth.*

2.3.4. Anatomical Distribution Of Optimal Filter Widths

Examples for FRFs of multi-units at four different cortical depths (columns 1-4: 400, 800, 1200 and 1600 microns) from four different penetrations are shown in Figure 2.11. Optimal filter widths (FW) are shown in the title of the plots. Qualitatively, the structures of FRFs from multi-units along the same penetration (rows) are often similar. However, the FRF structure was quite different between penetrations. I did not find an effect of cortical depth on optimal filter width. Figure 2.12. shows a plot of filter widths with the 41 multi-units that were set to the maximum filter width. These 41 multi-units were not systematically distributed with depth or penetration and this figure serves to show that even for those units for which an optimal filter could be determined with the cross-validation method, no obvious relationship of filter width and cortical depth becomes apparent.

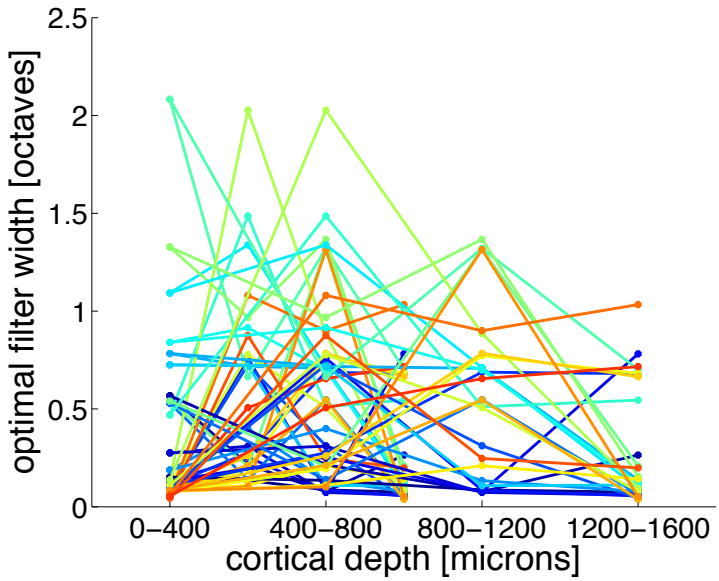


Figure 2.12. Development Of Optimal Filter Width With Cortical Depth.

Each line shows the distribution of average optimal filter width at 4 different depth regions (100-400, 400-800, 800-1200 and 1200-1600 microns). Penetrations were perpendicular to the cortex surface (see Figure 2.1. for illustration). Multi-units for which no optimal filter could be determined (41/480 multi-units) are not shown. No clear relationship is apparent between cortical depth and filter width.

2.4. Discussion

In this experiment I set out to determine the extent to which multi-units in the primary auditory cortex of the rat discriminate between formant combinations in double-formant vowels. A cross-validation method based on kernel smoothing of formant receptive fields (FRF) allowed us to estimate the resolution with which formant frequencies are represented by each multi-unit.

2.4.1. Cortical Resolution Of Formant Combinations

I found that the majority of optimal smoothing filters had a width of less than 0.2 octaves, with a median best filter width of 0.106 octaves (or 7.56% Weber ratio, interquartile range [5.18-25.6%]). The optimal filter width describes the width of a 2-dimensional Gaussian filter in units of standard deviation. A Gaussian filter of any width will drop to ~61% of its maximum height at one standard deviation distance from its centre. Therefore, the optimal filters for FRFs I determined describe prominent changes of the neural response when formants are shifted by the optimal filter's width. Consequently, formant shifts by in the order of ~8% or more should generate detectable changes in evoked neural firing rates in a substantial portion of cortical neurons, and we would therefore predict that animals would presumably find changes of this magnitude quite easy to detect, but might struggle if formants changes by less than ~5%.

The neural resolution of vowel formant shifts by A1 responses in the rat has, to my knowledge, not been quantified previously. Syka et al. (1996) report tone frequency discrimination in behaving rats at 5.7% with a range of 3.7%-7.3% (frequency range: 0.5-64kHz). These thresholds describe tone frequency shifts at which rats perform significantly above 50% in a go-nogo task and are therefore a rather sensitive

measure of just noticable differences. Talwar and Gerstein (1998), based on the same approach and criterion responses, found mean Weber ratios for pure tone discrimination of 3.06 ± 0.44% (frequency range: 5-32kHz).

Our results (median resolution 7.56%) lie at the upper end of the behavioural ranges for pure tone discrimination. However optimal filter widths do not represent just noticable differences for discrimination, but rather they are estimates of response changes between neighbouring stimuli. It is therefore likely that cortical discrimination thresholds for formant shifts are even closer to those seen in behavioural tone discrimination.

The extent to which smoothing revealed the underlying structure of the FRF varied among the multi-units I recorded. I also identified those multi-units for which the cross-validation method did not produce a significant local minimum of residuals (41/480). These multi-units were characterised by a lack of clear structure in their FRF, i.e. the average response over all stimuli was the best model of the underlying FRF.

2.4.2. Anatomical Distribution Of Cortical Formant Resolution

I did not find a clear relationship of optimal filter width and cortical depth in 480 multi-units in A1 of four different animals. As optimal filter width is a proxy for how finely a multi-unit resolves changes between stimuli, this result indicates that the discrimination ability of cortical multi-units does not depend on the cortical layer from which neural responses are recorded, at least for broadband stimuli such as my double-formant vowels.

That neural responses to tone frequency are similar along single perpendicular penetrations of A1 was demonstrated half a decade ago. Suga in 1965, when probing the A1 of echolocating bats with pure tones reported that “single units recorded within each puncture seemed to have similar best frequencies” (Suga, 1965). Similar results with pure tones were a little later also reported in cats (Abeles and Goldstein, 1970; Merzenich et al., 1975). Integration properties of cortical units in A1 other than frequency tuning were also reported to be consistent along radial penetrations, including intensity tuning (Clarey et al., 1994), frequency sweep tuning (Mendelson et al., 1993) as well as binaural integration properties (Brugge and Merzenich, 1973; Middlebrooks et al., 1994). Further, these and other response properties of neurons in A1 are known to vary with cortical location, including latency of responses (Mendelson et al., 1997), bandwidth of spectral integration (Schreiner and Mendelson, 1990) and the structure of frequency tuning curves (Sutter and Schreiner, 1991).

My result, that receptive field structures are qualitatively similar along individual penetrations is therefore in accordance with the literature reporting response properties of cortical units in A1. However the picture of essentially equal representations of acoustic features through different lamina of A1 begs the question of what the computational function of vertical networks of cortical cells may be. Differences in minimum response latencies of A1 neurons in different layers have been reported in cats (Phillips and Irvine, 1981), gerbils (Sugimoto et al., 1997) and recently in rats (Szymanski et al., 2009). However the literature on columnar information processing is still inconclusive today, 57 years after Mountcastle coined the term cortical column for his observations in cat V1. And this is potentially reflected

in my results that give no indication of differences in formant resolution across putative cortical columns.

2.5 Summary And Conclusions

In this chapter I characterised the ability of multi-units in the primary auditory cortex of the rat to discriminate between formant combinations in double-formant vowels. The cortical resolution of formant shifts I find lies at the upper end of the previously reported ability of rats to discriminate pure tones. *Discriminability of* and *sensitivity to combinations* of acoustic components is an important prerequisite in the ability to form auditory object classes. The formation of auditory classes during learning depends on the ability to discriminate between the variants along the acoustic dimension that are relevant to identity of the auditory class. Will rats hear differences between formant combinations near the resolution of A1 responses? Will they learn to classify the type of double-formant vowels I used here to explore cortical discrimination? If yes, will they use the information available in combinations of formants when judging how similar two vowels are? Or will they rely on the position of only one formant to discriminate vowels defined in F1-F2 space? These questions are the subject of the next chapter of this thesis.

CHAPTER 3 – BEHAVIOURAL EXPLORATIONS OF VOWELS BY RATS

We have seen in Chapter 2 that multi-units in the primary auditory cortex of rats discriminate formant shifts in double-formant vowels at a resolution of ~0.1 octaves. As discussed in the General Introduction, discriminability is one of the cornerstones of auditory class formation. In this chapter I explore whether rats can use formant combinations to make similarity judgements for double-formant vowels.

3.1. Introduction

To recognise a sound as member of a particular sound classification requires that the listener perceives the sound as similar to a set of other sounds. Similarity judgements can be based purely on the physical similarity of sounds but can also be based on abstract features, such as similar behavioural meaning of acoustically different sounds (Russ et al., 2007). That similarity judgements, based on common acoustic features of sounds, are a key element of sound classification has been demonstrated for bird song in birds (Doupe and Kuhl, 1999) and speech sounds in human infants (Eimas et al., 1971) and animals (Kuhl, 1981).

As described in the General Introduction (Chapter 1) for humans there are often no perceptual intermediates between two speech sound classes (Liberman et al., 1957). It has also been shown that human listeners judge two sounds that are acoustically close to the prototype of a sound class as more similar than two sounds that are close to the boundary between the classes. The strength of this perceptual magnet effect depends on linguistic experience in human listeners (Kuhl, 1983).

Consequently perceptual similarity judgements can be to an extent independent of the *acoustic* similarity of stimuli and subject to learning.

That invariant perception of speech sounds as members of a single class is not limited to humans has been demonstrated for many birds and non-human mammals, including chinchillas (Sinnott and Brown, 1997; Reed et al., 2003), cats and macaques (Kuhl and Padden, 1982). And some studies demonstrated that these species can use combinations of vowel formants to assign class membership to vowels: (Dooling and Brown, 1990) in budgeriars, (Ohms et al., 2010) in zebra finches, (Burdick and Miller, 1975; Kuhl and Miller, 1975) in chinchillas, (Hienz, 1996) in cats, (Hienz, 1988) in macaques and more recently in rats (Eriksson and Villa, 2006). Consequently, the acoustic similarity of formant combinations can be used by nonhuman species to form auditory classes. It is however unclear whether animals, when learning vowel classes, employ other strategies than expected, as for instance simply memorising stimulus-reward associations or using just one formant as an indicator to judge the similarity to one or the other prototype of two available double-formant classes.

Here I present evidence that rats do not obviously rely on one of two formants in spontaneous similarity judgements of double-formant vowels. I trained rats in two go-nogo tasks in making similarity judgements for formant combinations. In a first “predefined” classification task I show that rats can learn a vowel classification based on similarity of formant combinations. In a second “spontaneous” classification task I show that rats can spontaneously make similarity judgements using formant combinations when facing several possible and ambiguous ways of grouping by similarity.

3.2. Methods

3.2.1. Animals

All animal procedures were approved by the local ethical review committee and performed under license from the UK Home Office according to the Animal (Scientific Procedures) Act 1986. Seven adult female rats (Lister Hooded, 230-260 g) were trained in an automated training environment (described in detail in Chapter 6) to discriminate and classify vowels sounds. Two animals were trained in a predefined classification task, while the other 5 animals were trained in a spontaneous classification task. During training animals were housed inside the training cage. The delivery of drinking water was coupled to performing a task, and the animals performed training according to their own schedule and drinking habits.

3.2.2. Stimuli

All stimuli used were artificial double-formant vowels, generated as described in the Methods section of Chapter 2. All stimuli had a fundamental frequency of 151 Hz. All vowel stimuli were 200 ms long, cosine-ramped with 5 ms rise/fall time and played at a sound level of 75 dB SPL with a random jitter of up to ± 5 dB.

For the *predefined classification* task, 10 double-formant vowels were generated, 5 per vowel class, with formant centres of neighbouring vowels spaced at 1/6 of an octave, and formant centres between both classes spaced at 2/6 of an octave. The lowest F1 was at 7.4 kHz and the highest F2 at 19.7 kHz The distribution of formant combinations for the two vowel classes is shown in Figure 3.1. The group of vowels close to the main diagonal of the F1-F2 space was the rewarded vowel class and the one further away from the main diagonal was the non-rewarded class. The minimal distance of F1 to F2 was 0.24 octaves, the maximum distance was 1.41 octaves and

the two vowels closest and furthest from the main diagonal spanned 1 octave on both formant dimensions. The classes were defined along the perpendicular axis to the main diagonal such that moving from one class to the other involved shifting F1 and F2 further apart. The group stimuli closer to the main diagonal (blue shades in Figure 3.1. *left*) was the rewarded (go) class and the group of stimuli further away was the unrewarded (nogo) class in the *predefined classification task* described in the next section.

For the *spontaneous classification* task, 6 double-formant vowels were generated. Formant combinations were chosen such that all stimuli were at an equal distance from each other along both formant dimensions. In this way, no grouping cues based on physical proximity could be used to define a boundary between classes. The distance of the respective F1 and F2 in neighbouring vowels was 0.2 octaves and the lowest and highest F1 (and F2) spanned the octave between 10 and 20 kHz. The stimulus represented in green in Figure 3.1 (*right*) was the unrewarded (nogo) prototype and the stimulus in blue was the rewarded (go) prototype stimulus in the *spontaneous classification task* described in the next section.

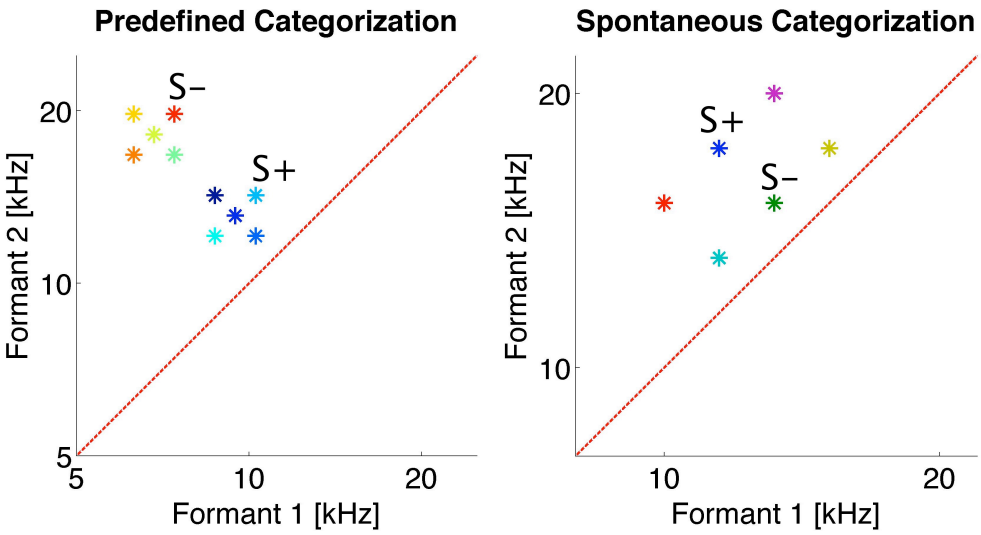


Figure 3.1. Stimulus Sets Used In The Two Classification Trainings.

*(**Left**) For the "predefined classification" I defined two sound classes by acoustic similarity of formant combinations. Formant 1 and Formant 2 in each combination are 1/6 of an octave away from the respective formant of their direct neighbour. The two classes are at a distance of 2/6 of an octave in both formant directions. (S+) the rewarded (go) class, (S-) the non-rewarded (nogo) class*

*(**Right**) The set of 6 formant combinations I used for the "spontaneous classification" training. All neighbouring combinations are at a distance of 0.2 octaves from one another in both formant directions. S+ is the rewarded (go) prototype stimulus, S- the non-rewarded (nogo) prototype stimulus.*

3.2.3. Training

All animals trained themselves on a go-nogo task in the automated training cage described in further detail in Chapter 6 below. Their task was always to initiate a trial by poking into a nose poke hole that contained the only water spout in the cage. This triggered one of the stimuli to be played. All stimuli in both tasks were presented in pseudo-random order. With the onset of the sound, a 2 s long response window started, during which animals could decide to stay and lick the water spout (go-response) or to withdraw from the nose poke (nogo-response). A "false alarm" evoked an air-puff inside the nose poke, a "hit" evoked 3 drops of water to be

delivered by the water spout. In this task, training blocks of 100 presentations of each stimulus were analysed.

In the *predefined classification task* animals had to recognise whether a presented stimulus belonged to the rewarded or non-rewarded class, based on acoustic similarity within the two classes and on the acoustic dissimilarities between the two classes. In the *spontaneous classification task* animals first had to learn to discriminate a rewarded prototype stimulus (S+ in Figure 3.1. left) and a non-rewarded prototype (S-). After reaching an average correct response over both stimuli of at least 80% in 3 consecutive training blocks the remaining 4 stimuli were presented in 10% of the trials. In this task, training blocks of 200 presentations of each stimulus were analysed. These stimuli are *catch* stimuli that are neither rewarded nor non-rewarded and were used to test the animals' spontaneous similarity judgements after successful prototype learning.

3.2.4. Performance Analysis And Estimation Of Confidence Intervals

I used the percentage of go-responses (%GO) for each stimulus as the main measure of performance, with the expectation that this should be close to 100% for the rewarded sound class and close to 0% for the unrewarded class if the animal understands the task and is able to perform it correctly. All stimuli were presented in pseudo-random order each stimulus presentation is an independent sample from all others. I chose to compare blocks of stimulus presentations rather than blocks of time. For instance the performance shown for a "block 1" with a block size of 100 trials compares the percentage of go-responses for the first 100 presentations of each stimulus used in the respective task.

3.2.5. Confidence Intervals And Isocline Analysis For The Estimation Of Class Boundaries

I determined confidence intervals for the comparison of performance during each training block. The upper and lower boundaries of the confidence intervals were derived from a cumulative binomial distribution based on block size and %GO responses in that block. To estimate class boundaries of spontaneous similarity judgements in the spontaneous classification task I used a cubic 2-dimensional to find putative performance values in the space between the 6 stimuli used. The boundary between the stimuli was then defined as the 50% performance line through that space. These methods are illustrated in the appropriate places in the Results section.

3.3. Results

3.3.2. Predefined Classification

In the predefined classification task two animals trained themselves in the automated training cage described in Chapter 6. They were exposed to five rewarded (S+) and five non-rewarded (S-) stimuli shown in Figure 3.2. (top). We have seen in Chapter 2 that rat A1 multi-units can discriminate between vowels with formant centre shifts of roughly a 1/10 of an octave. To render the task easy I chose to use formant shifts of 1/6 (or 0.16) of an octave within each vowel class. This is very close to the value deemed easily discriminable from the results of cortical discriminability described in Chapter 2 (8% or 0.11 octaves). The rewarded stimuli (S+) were associated with a go response and the non-rewarded stimuli with the nogo response. On average the animals performed 551 and 611 trials per 24 h respectively.

The results for the two animals I trained in this task are shown bottom half of Figure 3.2. The percentage of go responses to each formant combination is plotted and stimuli have the same colour code as in the top plot, showing the presented formant combinations for comparison. Confidence intervals on all performance plots were estimated from a cumulative binomial distribution with the block size as number of samples n and the %GO as probability p. From this distribution we can derive the probability of a particular number of observed outcomes under the assumption that p is the true probability for a go-response. Confidence intervals represent the number of outcomes that would be expected with a probability of 0.95 (lower bound) and 0.05 (upper bound).

I designed this task to test whether my rats could learn a class discrimination based on formant combinations with relatively tightly spaced class members. Already during the first block both animals responded differentially to both vowel classes. Both animals learned to withdraw for stimuli in the non-rewarded class, explaining the high percentage of go responses for the rewarded class. Within the non-rewarded class, both animals initially discriminated formant combinations with different degrees of certainty. Note, for example, that the stimulus shown in light turquoise, which is closest to the class boundaries, has among the highest numbers of false alarms (incorrect go responses). With increasing trial number this difference between members of the non-rewarded class declined (see Figure 3.3.).

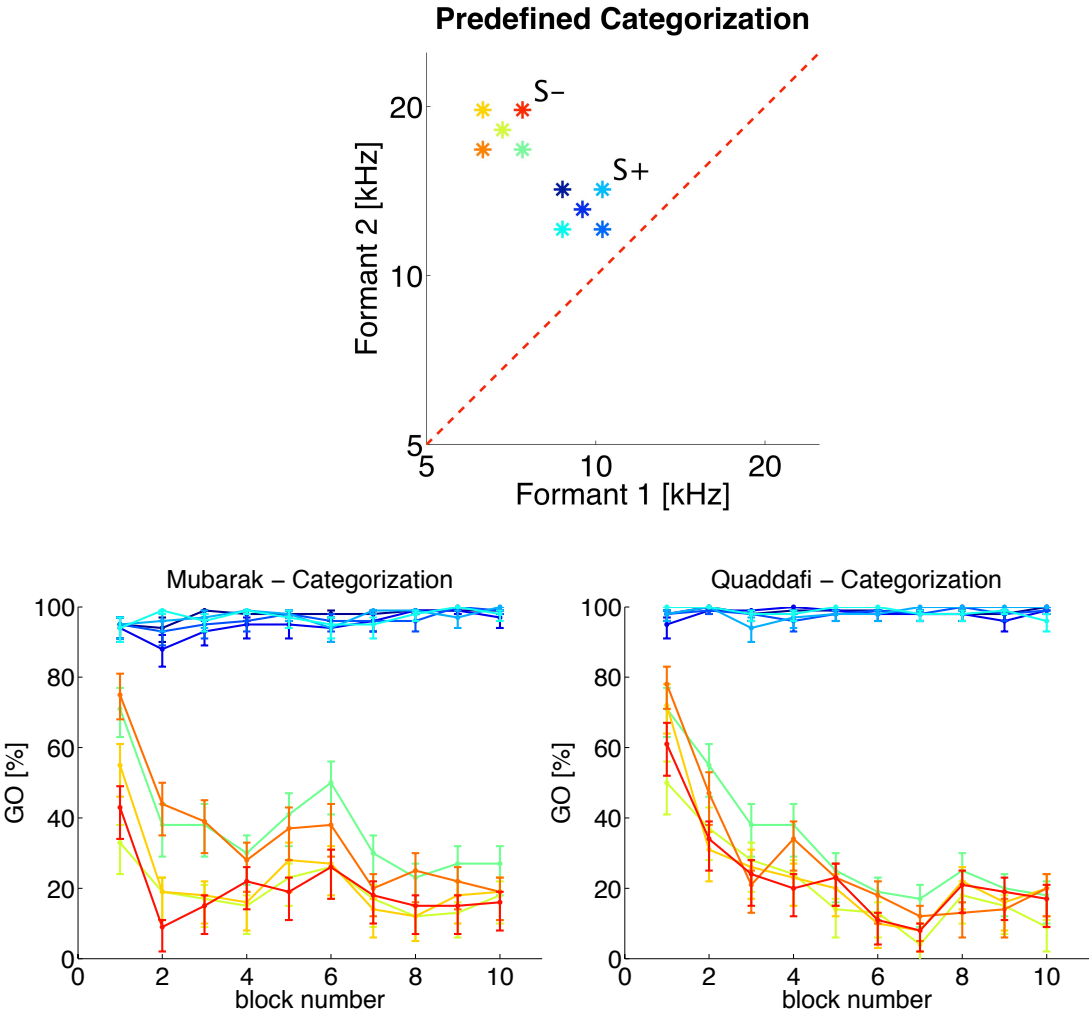


Figure 3.2. Performance Of Two Animals In The Predefined Classification Task

*(**Top**) The rewarded (S+) and non-rewarded (S-) formant combinations for comparison with performance. (**Bottom**) Performance in classification during the10 training blocks after the start of training. Stimuli are colour coded as shown in the top plot. Both animals immediately learned the class discrimination and improved performance during the first 7 blocks. Each block represents the performance after 100 trials for each of the stimuli. Error bars are the 90% confidence intervals of the observed performance based on a 100 trial binomial distribution. Initially for the S- class the performance still differs significantly between class members. This difference disappears over the subsequent 9 blocks of training.*

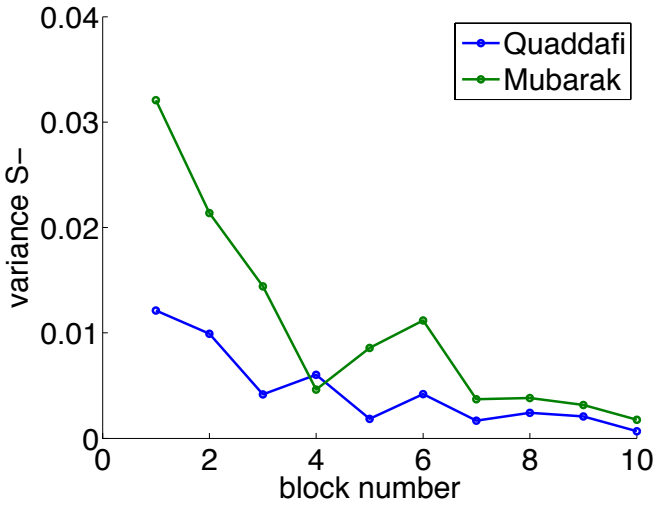


Figure 3.3. Decrease In Variance Of Within-class Discrimination With Training.
Within the non-rewarded the variance in go responses decreases over the first 10 block of training. This confirms the visual impression of fading difference between stimuli with the non-rewarded class with increased training. The block size is 100 presentations per stimulus

3.3.3. Spontaneous Classification

In the first experiment similarity judgements were provoked due to the acoustic similarity of stimuli within each group of vowels (the rewarded and non-rewarded groups). In the next experiment 5 animals trained themselves in the automated training cage described in Chapter 6. They were first trained to discriminate two prototype vowels (S+ an S-) and then tested on 4 novel test formant combinations that allowed several possible groupings.

I limited acoustic and behavioural grouping cues by presenting stimuli at equal formant distances and by neither positively nor negatively reinforcing test stimuli. I used this *spontaneous classification* task to find whether animals, without access to acoustic or behavioural grouping cues, would judge novel formant combinations more similar to one prototype than to the other, and if so, where the boundary between such similarity groups would lie. Three animals reached the criterion response. In the spontaneous classification tasks the three animals tested performed 501, 517 and

560 trials per 24 h respectively. Figure 3.4. shows the formant combinations used for training and testing and three possible ways of grouping novel stimuli with the prototypes based on similarity. In principle, double-formant vowels could be grouped by similarity of just one formant in novel sounds and known prototypes. The hypothetical class boundaries for this case are shown in the top panel of Figure 3.4. If however the combination of formants is used for similarity judgements then the diagonal class in the bottom panel in Figure 3.4. would be closer the perceived boundary.

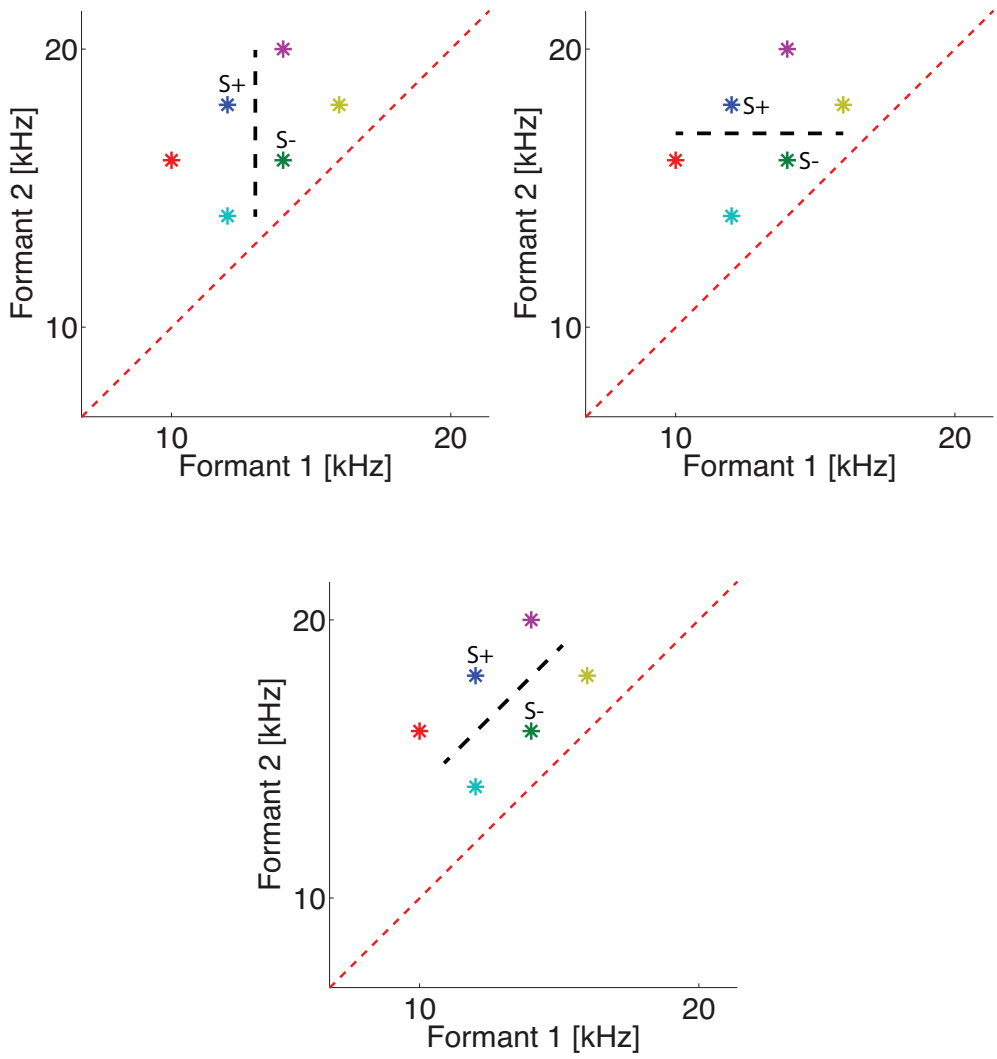


Figure 3.4. Hypothetical Class Boundaries In F1-F2 Space

Three possible ways of judging similarity between prototype stimuli (S+ and S-) and 4 novel stimuli. First animals are trained on the two prototypes. Then they perform the go-nogo task on all stimuli, including the 4 novel formant combinations. In the top two cases similarity would be judged based on only one formant. In the bottom case similarity would be judged on the relative distance of formants, i.e. on the way the formants are combined.

In Figure 3.5. I show, for two animals, the development of performance in prototype discrimination (left) and the development of spontaneous similarity judgements of 4 novel stimuli and prototypes during the first two test blocks (right).

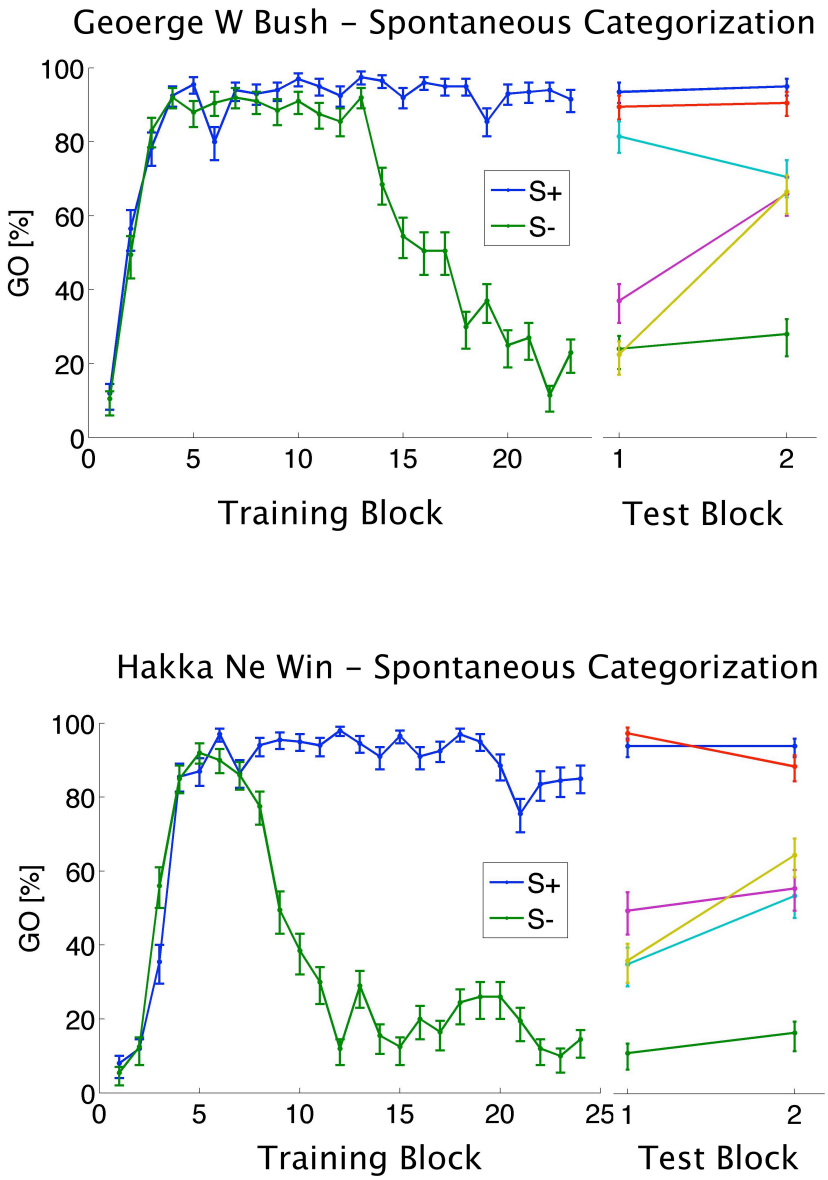


Figure 3.5. Prototype Learning And Spontaneous Classification.
*(**Left**) %GO responses of two animals in the discrimination of two reinforced prototype formant combinations. (**Right**) Spontaneous similarity judgements for prototypes and 4 novel, non-reinforced formant-combinations after at least 80% average performance in 3 consecutive blocks of the prototype discrimination. Each block analysed has a size of 200 trials per stimulus. Both animals spontaneously classified the novel examples as being perceptually closer to one prototype than to the other.*

The analysed trial blocks in all plots of Figure 3.5. are 200 trials wide. Both animals began with few go responses for both prototype stimuli and shifted to predominantly go responses within the first 5 blocks (1000 trials). In other words, it took the animals about 2 to 3 days to work out that the sounds are a cue to whether drinking water is

available. It then took them another 5 to 10 blocks (3 to 5 days) to learn to distinguish the two sounds and to withdraw (nogo-response) in response to the non-rewarded stimulus. Average correct response over both prototype stimuli in the last block of the initial training was 84% for Geoerge W Bush and 85% for Hakka Ne Win. Maximum average performance was 91% in block 22 for Geoerge W Bush and 92% in block 12 for Hakka Ne Win.

After the animals had reached the criterion performance of more than 80% correct response on average over both prototypes for 3 consecutive blocks, I introduced the non-reinforced test stimuli. We can see in the right plots in Figure 3.5. that both animals readily discriminated the novel examples, but not in exactly the same way. In the first block after introduction of the novel formant combinations, animals clearly judged the stimulus with lower F1 and F2 centres (represented in red) as being very similar (90% and 98% go responses) to the rewarded stimulus (blue). However they did not agree on the class membership of the turquoise stimulus: George W Bush judged it very similar to the rewarded stimulus (82% go response) while Hakka Ne Win judged it closer to the non-rewarded stimulus (35% go response).

The proportion of GO responses in the first block (200 trials) of the test phase, for the three animals that reached the 80% criterion during the training phase, are shown in Figure 3.6. The black lines in each plot are the 50% isocline lines that were interpolated using a method to determine iso-height lines on physical maps. The 50% isoclines represent the putative class boundaries in the spontaneous classification of all 6 stimuli. Stimuli lying on that boundary would all be expected to leave the animal undecided in its class judgement and would lead to 50% go- and nogo-responses. To calculate this 50% isocline line I generated a 2-D interpolation between the observed

stimuli in the F1-F2 plane. Isoclines can be understood as points of the same height as seen on physical maps. The class boundary in our F1-F2 map is the set of points where the 2-D interpolation predicts 50% performance.

Perceived class membership in the spontaneous judgement does not appear to be the same in the three animals. For Geoerge W Bush a boundary between two classes is estimated between the F1 values of the rewarded and non-rewarded stimuli. For Hakka Ne Win the class boundary comes to lie between the three stimuli near the main diagonal and the three stimuli further away from the main diagonal and runs straight through the stimulus with highest F2. For Adolf, all but the rewarded non-rewarded stimulus (green) would fall into one class.

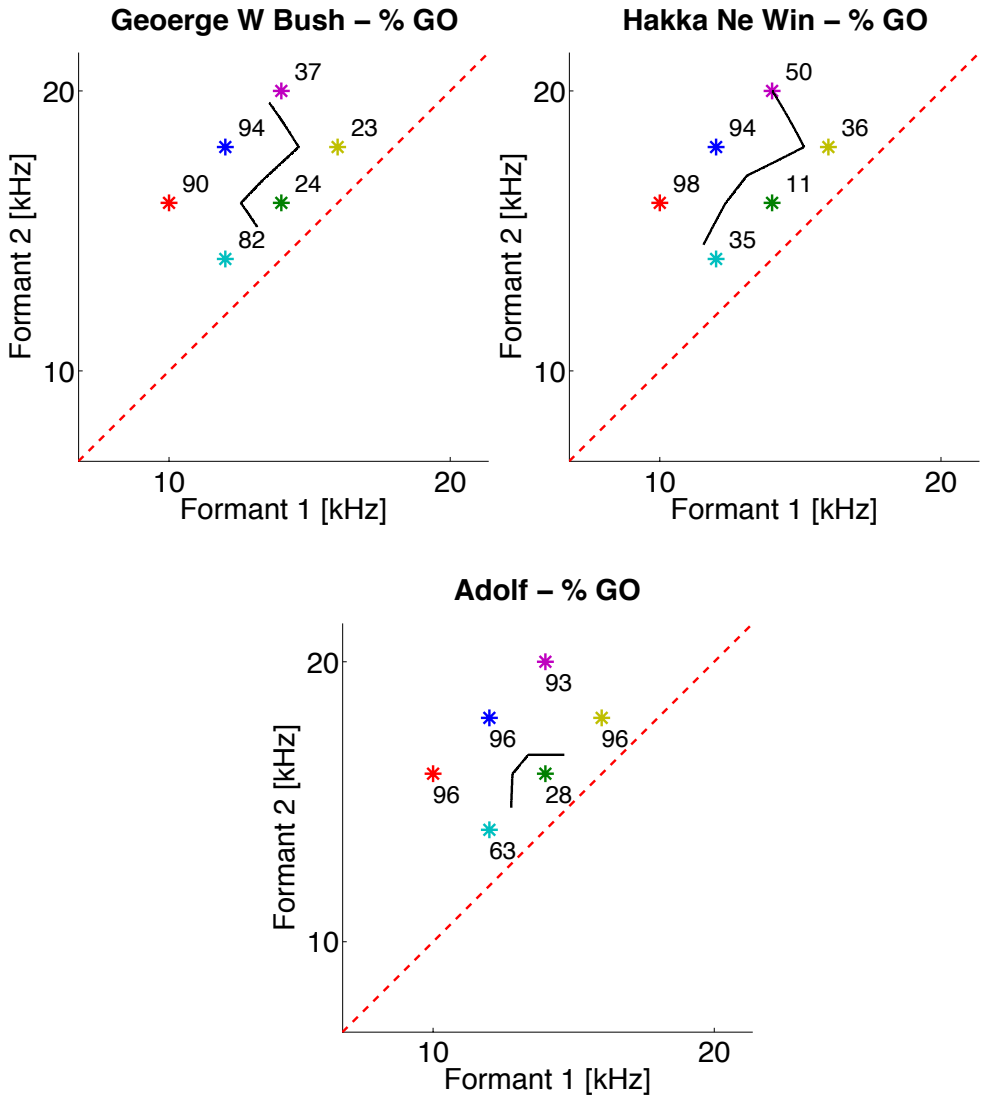


Figure 3.6. Performance of animals in the spontaneous classification

%GO responses in the 3 animals that reached the criterion performance. %GO for all 6 stimuli in the initial block of 200 trials per stimulus in the test phase. The black line is the estimation of the 50% class boundary. The similarity judgement differs between the animals, particularly for the stimulus with lower F2 but same F1 compared to the rewarded stimulus (turquoise). Also the class judgement for the stimuli with higher F1 (yellow) or F1 and F2 (magenta) differ between animals.

3.4. Discussion

In this Chapter I explored two questions: First, can rats learn a classification of double-formant vowels, based on acoustic similarity? Second, when neither acoustics nor behavioural relevance propose class membership, do rats still judge some novel stimuli as more similar to one prototype or the other, and if so, based on which criteria? I here use the term classification in the sense of *grouping stimuli according to perceived similarity*.

3.4.1. Learning Of A Classification Based On Acoustic Similarity

The aim of this training was to show that animals learn similarity judgements easily with double-formant vowels constructed in the same way as for my electro-physiological recording from rat A1. My animals almost immediately (within the first 100 trials) learned which of the 10 sounds belonged to the rewarded class and which to the non-rewarded class. Initially both animals judged the non-rewarded stimuli significantly differently, but with training over the next 900 trials this difference disappeared. The decrease of variance in the proportion of go-responses towards the non-rewarded stimuli indicates that the animals ceased to discriminate between the 5 stimuli of the non-rewarded formant combinations.

That rats can learn to classify vowels has previously been shown by Eriksson and Villa (2006). They generated artificial vowels for two vowel classes with 3 formants at positions taken from the distribution of human vowels reported by Peterson (Peterson and Barney, 1952). Eriksson and Villa generated the members of each vowel class by shifting both the fundamental frequencies and the formants. They initially trained their rats to discriminate 4 of these stimuli (2 from each class). Then they then demonstrated how their rats generalised over novel stimuli in each class (total of 5

stimuli per class). Their results illustrated that rats can classify vowels based on the similarity of formants, and in the presence of varying fundamental and formant frequencies. My predefined classification task replicates this result with a few variations in experimental design. First, my artificial vowels contain only 2 formants and only vary along the dimension that is relevant for human vowel classes, i.e. the position of the first two formants in frequency. Also, my stimuli were shifted into the rat's region of best hearing while Eriksson et al. chose relatively much lower formant frequency values in a range typical for human speech. Their two stimulus classes differed mainly in the positions of the second formant: in one class F1 ranged between ~850-1100Hz while in the other F1 ranged between 1100 and 1180 Hz. When considering the rat audiogram (e.g. Heffner et al., 1994) we can see that in the region between 500 and 2000 Hz the rat's hearing threshold declines somwhat steeply (with a change of 30 dB SPL), whereas between 7.4-19.7 kHz the rat audiogram is essentially flat (see Figure 1.4.). It is therefore possible that the rats in Erikksson et al.'s study heard stimuli from both classes at different perceived loudnesses. In contrast, my stimuli were played at sound levels which were randomly jittered between 70 and 80 dB SPL to eliminate similarity judgements based on loudness, and ensuring that the animals had to use the vowels' formants as the cues for classification.

To my knowledge the study by Eriksson and Villa (2006) is the only previous study on the rat's ability to classify vowels. Kluender et al. (1998), in an ingenious study, showed that starlings learn to classify artificial double-formant vowels. The birds were trained in a go-nogo task to peck at a target after hearing a vowel from the rewarded class and to withhold from pecking after a non-rewarded vowel. Not all birds learned the same classification. Each half of the cohort of birds was trained on a different,

orthogonal stimulus class, i.e. some pairs of vowels that were members of a single class for the first group of birds were members of the two separate classes for the second group. In that manner, Kluender et al. showed that the small difference in peck rates for vowel pairs in the same class were not due to the birds' inability to hear differences, but due to their judgements of similarity to the class prototypes. That my rats first exhibited performance differences for the non-rewarded stimuli argues for the fact they also could discriminate between members of that class, but learned over the course of training to no longer respond to this acoustic dissimilarity, in analogy to the behaviour Kluender et al. showed in starlings.

The decrease in discrimination of the 5 non-rewarded stimuli may indicate a "flattening" of the response for members of the non-rewarded stimulus class. Such flattening of the discrimination curve within a sound class is one of the characteristics of categorical perception and has been described for example in gerbils trained to classify tone sweeps (Ohl et al., 2001), chinchillas trained in VOT classification (Kuhl and Miller, 1975) and in starlings that learned to classify vowels (Kluender et al., 1987). However, these studies presented novel stimuli after a number of prototypes of a class were learned, whereas I presented all class members right from the beginning of training. I can therefore not exclude the possibility that my rats had learned the 5 non-rewarded stimuli by heart. Nevertheless the immediate classification I observed already during the first 100 trials indicates that my animals had judged the classes by similarity of their sound members. Consequently the reduced behavioural discrimination remains an indication of categorical perception in my rats.

3.4.2. Spontaneous Similarity Judgements After Prototype Discrimination

Where the predefined classification task shows that rats can learn a double-formant classification based on acoustic similarity, the spontaneous classification task provides a complementary perspective in 2 ways: animals are trained on two prototype stimuli and only later exposed to novel test stimuli. For the novel stimuli, the feedback mechanisms of the task do not define a "correct" classification - for them no response is reinforced. Typically the term “generalisation” describes the ability to group new stimuli correctly into a previously trained stimulus class. Therefore, if no cues for grouping are available, any apparent grouping should reflect the animal's guess based on "perceived similarity", which may reveal the features based on which an animal would divide the formant space after prototype learning.

All three animals that reached criterion performance in the prototype discrimination training immediately classified novel vowels, either as being similar to the rewarded prototype or as being more similar to the non-rewarded one. The similarity groups (or classes) my animals indicated by their go-nogo responses were, however, not the same for each animal. Consequently, during prototype learning, different animals must have relied on different features of the prototypes to discriminate them and used the same features for similarity judgements of the novel vowels. In one animal (Geoerge W Bush) the difference in go-nogo responses to the whole set of formant combinations could easily be explained by considering only changes of the second formant: when F2 was higher than in the rewarded stimulus, Geoerge considered stimuli as being similar to the non-rewarded prototype. In contrast, for Hakka Ne Win this feature (frequency of F2) would not explain the difference in go-nogo responses to the test stimuli. The estimated class boundary came to lie much closer to the diagonal through all stimuli and therefore closer to a similarity judgement based on

the relative distance of the formants, with similarly spaced formant combinations being judged as similar. Finally Adolf appears to have employed yet another strategy, judging all formant combinations different from the non-rewarded prototype as potentially rewarded, except for the stimulus with the lower F1 and F2. In this case the animal appears to have hesitated to judge that stimulus as similar to the rewarded prototype. In summary, rats, when trained in a prototype discrimination will use different features of double-formant vowels to classify novel stimuli, and are not limited in their choice of features as grouping cues, such as the position of one formant alone.

The study by Kluender et al. (1998) in starlings and classical studies on similarity judgements for vowels in adult and infant humans and monkeys (Kuhl, 1991) show that similarity to prototypes of phonetic classes is a key feature of phonetic classification. None of these studies however investigated explicitly which feature or combinations of features of vowels are used by animal listeners to make such similarity judgements. I find that different rats employ different strategies in their similarity judgements and that these strategies are not necessarily based on changes along one formant dimension. This observation is in accordance with the report by Kluender et al. (1998) that similarity judgements of starlings depend on the distance from the prototype in F1-F2 space, irrespective of the direction of change (whether towards or away from the second classes prototype).

3.5 Summary & Conclusions

The evidence I present here shows that rats learn vowel classes based on acoustic similarity, and that their similarity judgements do not obviously depend on variability along one formant dimension alone. The spontaneous classification task revealed

that different animals may solve classification challenges on potentially very different perceptions of what makes different stimuli similar. One important question in auditory neurophysiology is how receptive fields of sensory neurons map onto perceptual judgements. In the case of spontaneous classification, awake recordings from auditory cortex areas could answer the question of whether receptive fields are more representative of the acoustic structure of the stimulus or of the behavioural response and consequently whether they represent the perceived quality of a sound.

To perform such comparisons between neural and behavioural responses, reasonable metrics have to be employed and it is not clear whether Euclidean distance metrics, which I used to define distances in acoustic formant space, translate directly into perceived and neural metrics. Finding metrics that describe the relationship of perceptual and neural responses remains a central problem in auditory neurophysiology. The next chapter addresses a part of this problem by asking whether A1 responds to formant combinations in a way that is characteristic for the typical human perception of vowel classes.

CHAPTER 4 – INTEGRATION OF FORMANT COMBINATIONS IN RAT AUDITORY CORTEX

4.1. Introduction

In Chapter 2 of this thesis I have explored one cornerstone of the formation of auditory object classes: the ability to discriminate between sounds along stimulus dimensions that are relevant to the objects class. In the case of vowels these are the positions of the first two formants in frequency. We have seen that even at a small spacing (of 0.1 octaves) between formants, multi-units in the A1 discriminate between neighbouring vowel sounds with a resolution comparable to that seen in behavioural studies in rats. The cross-validation method also showed that, already on the small scale (10-20 kHz) of formant frequencies, formant receptive fields exhibited some structure.

As discussed in Chapter 1, the second cornerstone feature for the formation of auditory classes is *sensitivity to combinations* of those sound features that are relevant to an object's identity. In this chapter I investigate how neural activity in the A1 of the rat integrates the energy contained in particular combinations of the relevant sound features (the first two formants) of artificial vowels on a broad scale (1-30 kHz). To assess the extent to which auditory cortical multi-units in A1 are sensitive to formant combinations, I evaluate how well different models of formant integration describe responses to double-formant vowels. First, let us revisit some of the literature on response properties of A1 cortical units that inspired the models of formant integration which I tested here.

4.1.1. Tonotopy And Pure Tone Tuning

Ever since Woolsey and Walzl (Woolsey and Walzl, 1944) first reported the tonotopic organisation of the cat auditory cortex, response properties of mammalian A1 neurons have been characterised with their pure tone sensitivity. Tonotopy has been reported in AC of all mammalian species studied to date, including cats (Hind, 1953; Evans et al., 1965; Reale and Imig, 1980), dogs (Tunturi, 1950), rabbits (McMullen and Glaser, 1982), mice (Stiebler et al., 1997), bats (Suga and Jen, 1976), rats (Sally and Kelly, 1988; Kalatsky et al., 2005), gerbils (Thomas et al., 2006), guinea pigs (Hellweg et al., 1977), hedgehogs (Batzri-Izraeli et al., 1990), ferrets (Bizley et al., 2005), macaques (Brugge and Merzenich, 1973) and humans (Romani et al., 1982). Tonotopy is based on the extraction of a best or characteristic frequency for a given cortical unit from its responses to a set of pure tones that range over a broad frequency interval. The resulting receptive field, when probing the responses of an A1 neuron with pure tones at different sound levels, is a frequency-level curve or frequency response area FRA (the area above the curve), representing the strength of a neuron's response to different pure tone frequencies at different sound levels. If responses of A1 cortical units were sufficiently described by their FRA, the frequency level curve could be used as a linear filter to predict the responses to arbitrary broadband sounds.

4.1.2. Linear And Nonlinear Models Of Integration Properties In A1

However, given that both thalamic and cortical inputs into A1 are highly preprocessed (Nelken, 2004), that already auditory nerve fibres are tonotopically ordered, and that a large proportion of A1 units do not respond at all to very simple stimuli like pure tones (Hromádka et al., 2008; Sadagopan and Wang, 2008) it seems likely that simple linear models of spectral integration will fall short of capturing the response

properties of A1 units to spectrally rich stimuli. However some linear models have been shown to yield good estimates of A1 receptive field structures in response to stimuli with complex temporal and spectral features, including dynamic ripples (Kowalski et al., 1996; Klein et al., 2000; Miller et al., 2002), random chords (deCharms et al., 1998; Blake and Merzenich, 2002; Rutkowski et al., 2002; Linden et al., 2003) and binaural spectral cues (Schnupp et al., 2001). For speech sounds in particular, linear integration of spectral features have proven to be a useful model.

One class of linear models of A1 receptive fields takes into account how spectral components develop over time. These models of spectral integration estimate so called spectro-temporal receptive fields (STRFs) of neurons. STRFs are generated by averaging stimulus histories over set of time windows (of equal length) that are placed onto the stimulus spectrogram just before all the occurrences of action potentials in the train of neural responses to the same stimulus. The resulting average stimulus history is the STRF for that neuron. STRF models of spectral integration rest on the assumption that an “optimal” window of stimulus history will evoke a spiking response and a suboptimal history will not.

STRFs, when constructed from the best frequencies of neural populations in rat A1 (Engineer et al., 2008) and of individual cells (Mesgarani et al., 2008) in ferret A1, successfully represented the similarity of speech sounds and the discriminability of these sounds in behavioural experiments. STRFs are linear models of spectral-temporal integration and do not estimate possible nonlinear interactions between spectral or temporal components in stimulus histories. Yet, it has been reported that individual neurons in A1 and its homolog area L in birds can exhibit highly nonlinear integration of spectral components when driven by complex natural stimuli. For

instance Theunissen et al. (2000) reported that, for neurons in area L of the zebra finch, STRFs constructed from responses to natural stimuli were superior to STRFs derived from tone pips, in predicting area L responses to broad and narrowband sounds. However, Theunissen and colleagues emphasise that all STRFs turned out to be incomplete models of the neurons' responses, irrespective of the stimulus type used. The observation that STRFs are not sufficient in describing neural response properties was also made in individual neurons in layer 2/3 of rat the A1 (Machens et al., 2004). Strong nonlinear interactions of narrowband sounds and their low-energy broadband background have been observed in neural responses of A1 units. For instance, Bar Yosef et al. (2002) showed that A1 neurons respond with extreme sensitivity to the removal from natural bird chirps of the very quiet noise background. Explicitly multiplicative models for the integration of temporal features of sounds (such as intra-aural time and level differences) proved to match neural responses in the bird area L (Peña and Konishi, 2001) and such simple nonlinear models can be a good starting point for the exploration of nonlinear integration properties of A1 responses in mammals. Multiplicative integration of sound components can be modelled with a simple logical AND operation on spectral components, providing a straight forward comparison to additive models that are equivalent to a logical OR operation (Schnupp and King, 2001). Finally, regarding speech sounds Ohl and Scheich (1997) proposed a nonlinear model of formant interaction to account for regions of excitation in the FRFs of single units in the A1 of the Mongolian gerbil.

In summary, the literature paints a mixed picture of the way in which A1 integrates spectral features of sounds. While some studies suggest that models based on an assumption of an approximately linear integration of frequency components can provide a quite reasonable description of response properties in A1, other studies

find very significant non-linearites. Of course, if we ask whether cortical units might be selective for acoustic object classes that are defined along two or more spectral features of a sound (e.g. the formants of a vowel) then we would expect these cortical units to exhibit non-linear responses which make them sensitive to specific combinations of these features.

In this chapter I explore whether, concerning the class-relevant dimension of formant location, A1 neurons behave more like linear spectral filters, or more like feature combination detectors. Do A1 multi-units integrate the energy contained in the formants of double-formant artificial vowels in a linear, additive fashion, or in a nonlinear, combination sensitive fashion? The first of these types corresponds to an additive, or logical OR-like, operation, where the firing rate in response to a double format vowel would be reasonably well approximated by a weighted sum of the responses to each formant presented alone. The second operation would be more multiplicative, or logical AND-like, such that the responses to double-formant vowels would be best approximated by the weighted product of the responses to single-formant vowels.

4.2. Methods

4.2.1. Surgery And Electrophysiology

Surgery and recordings were conducted under exactly the same conditions as described in Chapter 2 (Methods).

4.2.2. Stimuli

Artificial single-formant and double-formant vowels were generated in Matlab using Malcolm Slaney's Auditory Toolbox. The 13 formant centres were spaced at 0.4 octaves between 1088-30314 Hz. The fundamental frequency was chosen at 151 Hz (same as in Chapter 2). First, single-formant vowels were generated for all 13 formant centres. To ensure that the double-formant vowels used in this study were acoustically the linear sum of the single-formant vowels, I generated the double-formant vowels by summing the amplitude spectra of the single-formant vowels (example in the right plot in Figure 4.1.) in the frequency domain, while holding the phase spectrum constant. In this manner I avoided interference patterns ("beating") which might introduce artefacts if the single-formant vowels were simply added in the time domain. This resulted in 78 double-formant stimuli for which the second formant frequency is always higher than the first formant frequency. The set of formant combinations in F1-F2 space I used is shown in Figure 4.1. (left).

4.2.3. Electrophysiological Recordings

The neural recordings were performed and neural response measures were extracted in the same way as described in Chapter 2 (Methods).

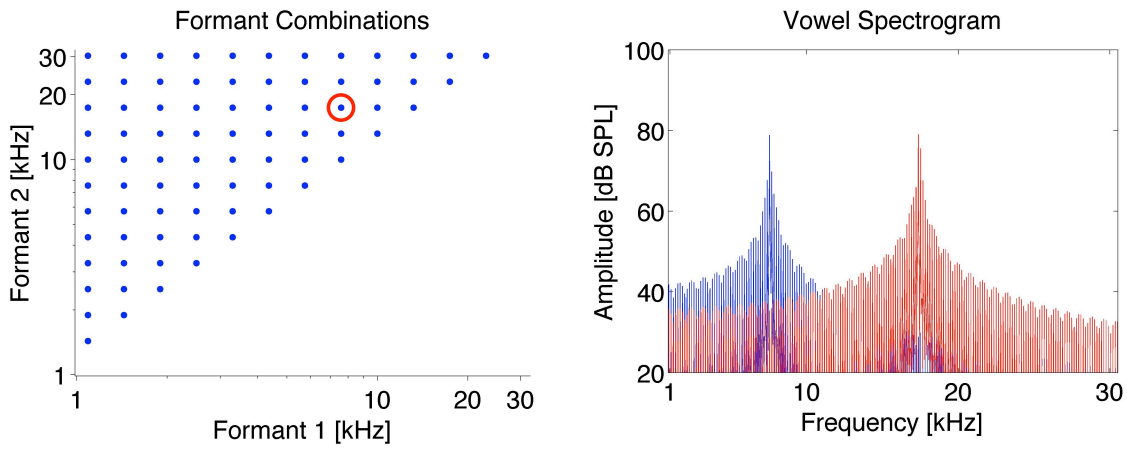


Figure 4.1. Formant Combinations And Example Spectrum

*(**Left**) The distribution of formant combinations sampled in recordings from rat A1. Each double-formant vowel was constructed as the sum of the sound vectors of two single-formant vowels. (**Right**) The spectrograms of two single-formant vowels that were combined form the double-formant vowel encircled in red in the left plot.*

4.2.4. Formant Receptive Fields And Detection Of Formant Tuning Peaks

Formant receptive fields (FRFs) for single and double-formant vowels were generated for both onset (20-100ms) and offset (200-280ms) responses. The measure used to describe response strength was average firing rate in the respective response window. The raw data was analysed and spike rates determined in the same way as described in Chapter 2 (Methods). FRFs for single and double-formant vowels are always plotted alongside for comparison in this chapter.

4.2.5. Model Fits Of Double-formant Integration

FRFs were fitted with a number of simple parametric models based on single-formant responses. Model parameters were estimated by fitting models to the observed double-formant responses using a least squares error criterion and gradient descent. The Matlab function fminsearch was used to implement this.

Three different models based on single-formant responses were evaluated this way:

a purely additive model (a linear combination)

$$R_{i,j} = a + b \cdot \left(F_i + F_j \right)$$

a purely multiplicative model:

$$R_{i,j} = a + b \cdot \left(F_i \cdot F_j \right)$$

a combined additive and multiplicative model:

$$R_{i,j} = a + b \cdot \left(F_i + F_j \right) + c \cdot \left(F_i \cdot F_j \right),$$

where R is the modelled FRF, a, b, c are model parameters, and F is the response to single-formant vowels with formant centre frequency i or j. To evaluate the "success" of each model, I calculated the proportion of neural response variance in the observed FRF data they could explain, and calculated Akaike information criteria, as described below.

4.2.6. Variance Explained By Different Models: Variance Decomposition And Exclusion Criterion

The evaluation of the success of each model is based on a variance decomposition method which assumes that the trial-to-trial variability in neural responses will be in part attributable to systematic variation in stimulus parameters, and in part due to physiological noise. Only the first part of the response variance is in principle "explainable" by a deterministic model built around stimulus parameters. To measure the success of a model I therefore first need to estimate how much of the response variance is "explainable", and then compare this to the amount of response variance captured by the model.

The total variation in a data set based on n different stimuli and m repeated presentations per stimulus can be expressed as the double sum of the total sum of squares

$$TSS = \sum_{i=1}^{n} \sum_{j=1}^{m} (x_{ij} - \overline{x})^2 ,$$

where x_{ij} is the response of the unit to j-th repeat of the i-th stimulus and $\bar{x}$ is the grand mean over all responses. The total variance can be divided into a sum of that part which describes the variance within the repeats of each stimulus (due to physiological noise) and the variance between the mean responses to each stimulus (which is attributable to stimulus effects and needs to be captured or explained by the model). It can be shown (Sokal and Rohlf, 1995) that the total sum of squares is equal to the summed deviations of repeats from their mean plus the deviations of each stimulus mean from the grand mean times the number of repeats:

$$TSS = \sum_{i=1}^{n} \sum_{j=1}^{m} (x_{ij} - \overline{x_i})^2 + m \cdot \sum_{i=1}^{n} (x_i - \overline{x})^2$$

$$TSS = NOISE + m \cdot ESS$$

where x_{ij} is the mean over all m repeats for each stimulus i. The left term in the equation describes the variation within each group of stimulus repeats (the NOISE part of the variation), the noise part of the variation, and the sum in the right term is the explainable sum of squares (ESS) part of the variation. Model fits can in general only approach the explainable variation in data.

The explainable sum of squares (ESS) can therefore be split into a part term that expresses how close the model is to the observed response (the model fit: MF) and a term that describes how much variation the model is able generate (the model variation: MV).

$$ESS = \sum_{i=1}^{n}(x_i - x_i^m)^2 + \sum_{i=1}^{n}(x_i^m - \overline{x})^2$$

$$ESS = MF + MV$$

Before introducing a term for the amount of variation explained by a particular model of the data it has to be determined whether there is a significant amount of explainable variance in the observed data. After all and it would not be sensible to debate the relative merits of different types of models for describing the formant sensitivity of a particular neural multi-unit in my sample if it turns out that this multi-unit is essentially uninterested in changes in formant frequencies. To decide whether the explainable variance in the neural responses from any one multi-unit is "sizeable" I used a yardstick measure based on the ANOVA definition of a large effect.

$$yardstick = \frac{NOISE}{n \cdot (m-1)}$$

Sufficient explainable variation in the observed data exists if the explainable sum of square is larger than the yardstick. So multi-units were excluded from the analysis if

$$ESS \leqslant yardstick$$

Given a large enough amount of explainable variation I defined the *percentage variation explained* as

$$VE = \frac{MV}{ESS} \cdot 100$$

4.2.7. Measure of model fit: Akaike Information Criterion

To quantify and compare the goodness of fit generated by different models I used the Akaike Information Criterion (AIC), which is defined as

$$AIC = -2 \cdot L + 2 \cdot (p + 1),$$

where L is the log-likelihood of observing the data under the assumption that the fitted FRF is an accurate description of the "true", underlying neural sensitivity function. *p* is the number of model parameters used to generate the fit. The log-likelihood is calculated under the assumption of Poisson-distributed spike counts over trials, where the FRF gives the Poisson rate for a particular combination of stimulus formants. The measure penalises increasing numbers of parameters and therefore takes into account potential over-fitting.

4.2.8. Measure Of Linearity: Linearity Index

To determine how strongly linear or nonlinear multi-units integrated the energy contained in the formants of a double-formant vowel, I calculated a linearity index from the amounts of variance explained by the additive and the multiplicative model using the following:

$$LIN = \log(\frac{VE_{add}}{VE_{mult}}),$$

where VE_{add} and VE_{mult} are the percentage of variance explained by the additive and the multiplicative models, respectively.

4.3. Results

4.3.1. Structure Of Formant Receptive Fields

I recorded the responses to 78 double-formant and 13 single-formant vowels from 448 multi-units in A1 of 4 rats. Double-Formant vowels were constructed by summing pairs of neighbouring single-formant vowels. Figure 4.2. shows a raster plot of evoked spikes in one multi-unit to 20 repeats of the same stimulus (top) and the peri-stimulus time histogram of evoked spikes by the same unit in response to all stimuli and all repeats (bottom). In this chapter I analyse the integration of formants in receptive fields generated from both onset (20-100 ms post stimulus) and offset responses (200-280 ms post stimulus). Figure 4.2. shows that spikes are evoked by both the onset (at 0 ms) and offset (at 200 ms) of formant combination stimuli.

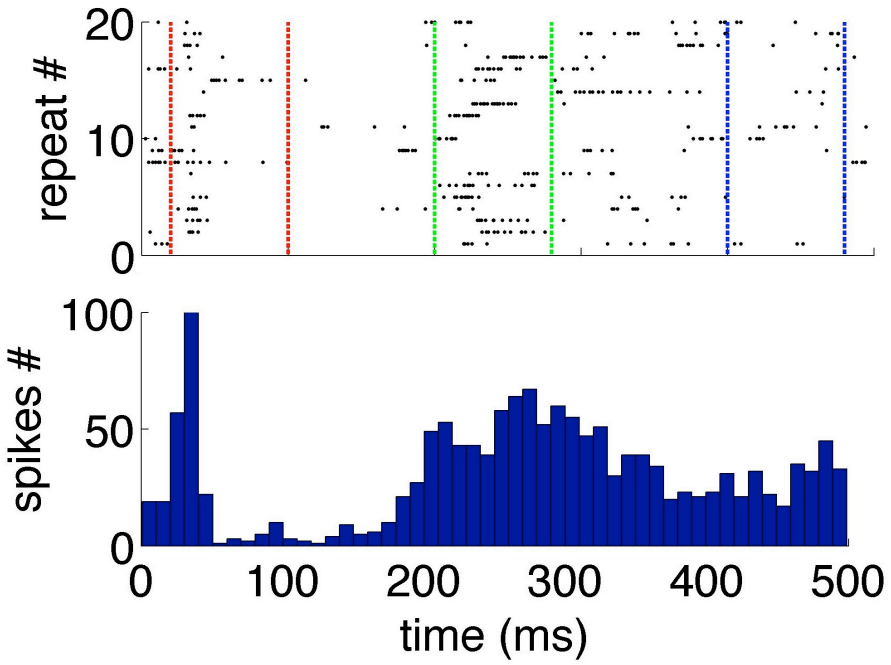


Figure 4.2. Repeats Of One Stimulus And Pooled Responses.

*(**Top**) A raster plot of spike recorded in response to 20 repeats of the same double-formant stimulus at one the 16 recording sites during one penetration in A1 of the anaesthetized rat. Red line: onset window, green lines: offset window, blue lines: baseline window. (**Bottom**) A histogram of spikes at the same recording site and penetration, in response to all 20 repeats of the 78 vowel stimuli used in this chapter.*

Figure 4.3.A shows three examples of FRFs generated from onset responses. We can see that single-formant responses are roughly aligned with double-formant responses but the extent to which a second formant influences the response varies between different multi-units. Figure 4.3.B shows an additive and a multiplicative model of for a hypothetical single-formant response distribution (shown on the axes).

In Figure 4.3.A, the FRF of the multi-unit on the right appears to have a distinct “hotspot“ of responses to formant combinations that are close together (near the main diagonal). This structure resembles more closely the structure generated by a multiplicative (nonlinear) model of formant integration. In contrast the FRF of the multi-unit in the middle maintains a strong response to a formant that falls on its single-formant peak over a wide range of second formant frequencies. Such structures resemble more closely the additive (linear) model of formant integration. The FRF on the left of Figure 4.3.A appears to be a mixed case.

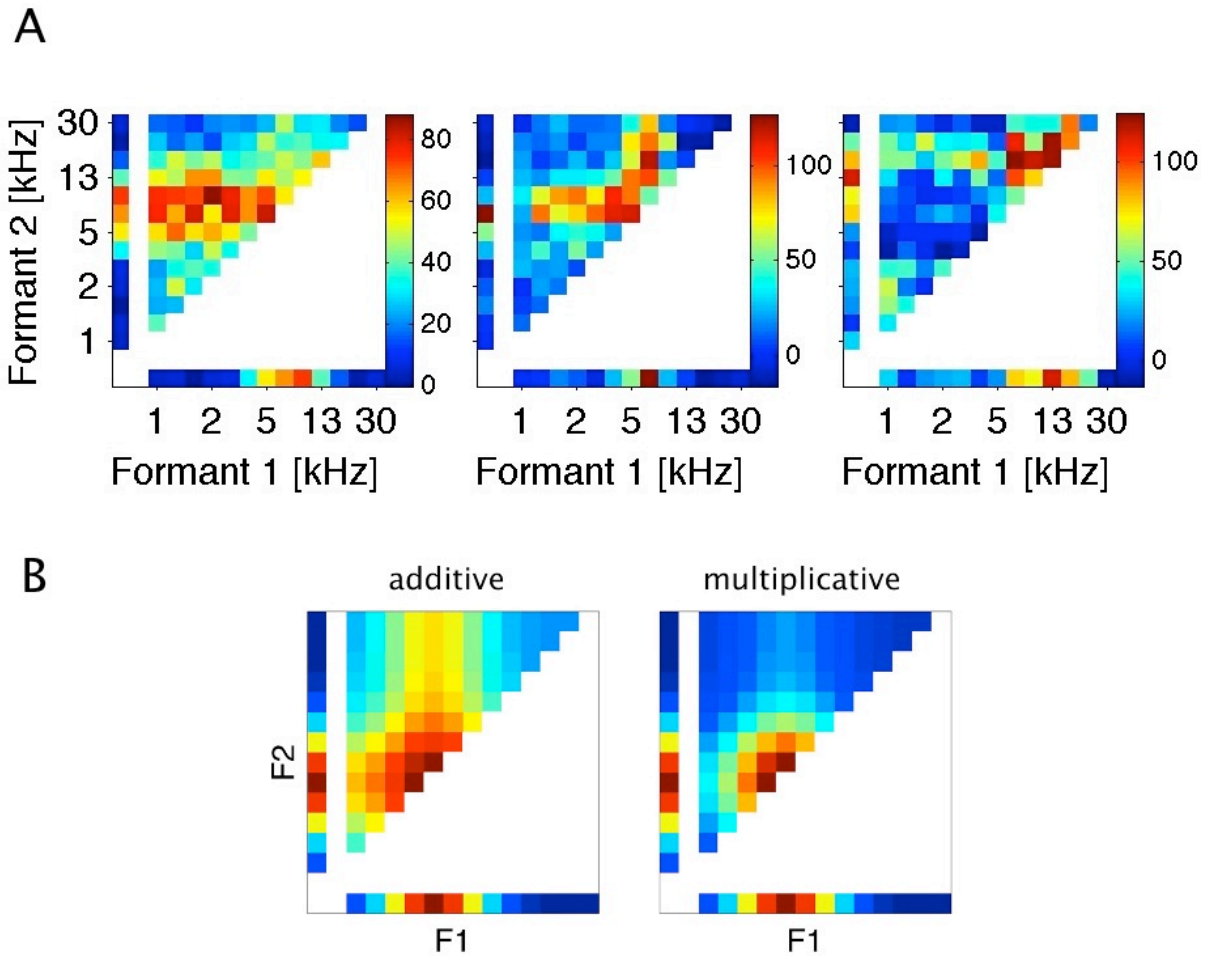


Figure 4.3. Examples Of Observed Formant Receptive Fields And Structure Of Standard Models Of Single-Formant Integration

*(**A**) Three typical examples of formant receptive field from three different animals. Onset responses to 78 double-formant vowels and 13 single-formant vowels with a spacing of 0.4 octaves between neighbouring stimuli. The responses to double-formant vowels are shown by the colour scale in the upper triangle of the matrix. Responses to the single-formant vowels are shown along both the x and y axes for comparison. All colour codes (z-axes) depict the spike rate in Hertz. (**B**) An additive and a multiplicative model of double-formant responses derived from hypothetical single-formant response.*

The great majority of single-formant responses had a single peak response as shown in Figure 4.4, indicating that the models shown in Figure 4.3.B are a useful starting point to investigate the properties of formant integration. Note, that even with multiple peaks these models would yield clear and distinct predictions of the expected structure in FRFs.

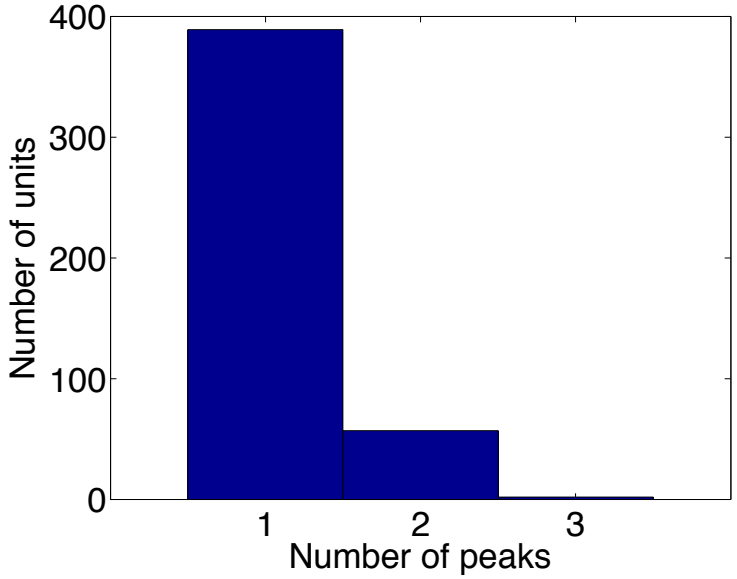


Figure 4.4. The Number Of Peaks In Onset Responses To Single-formant Stimuli
The great majority of multi-units showed one peak of maximal response. Single-peaked frequency responses to single formants generate models of formant integration as shown in the examples in Figure 4.3.B.

To determine the number of tuning peaks in single-formant responses the raw tuning curve was convolved with a 6-point hanning window (to smooth) and upsampled by 10 times. The resulting tuning curves (TC) were normalised by their maximum values. Extrema in TC were determined using the discrete derivative to get TC'. Taking the derivate again yields TC'' from which we can find those extreme points that are maxima. Peaks were defined as those maxima for which TC>0.5, TC'=0 and TC''<-0.5. This eliminates points where the peak is not very high, or not very "peaky".

4.3.2. Models Of Double-Formant Responses

The explainable variance in onset responses was never smaller than twice the "yardstick" explained above, which I used to decide whether neural responses were sufficiently sensitive to the stimulus parameters to make any form of modelling worthwhile. The median value was 13.6 times the yardstick level (interquartile range [8.7 21.79]). Therefore all multi-units qualified for model fitting. To determine whether the multi-units I recorded integrated the energy contained in vowel formants in a more linear or more multiplicative manner, I generated for each multi-unit an additive and a multiplicative fit, based on the single-formant responses of the same multi-unit. Figure 4.5. shows two of the observed FRFs described in Figure 4.3. together with a purely additive (middle plots) and multiplicative (right plots) models . For the definition and means of construction of these models please refer to section 4.2.5 of the Methods. The single-formant responses that these models were based on are plotted along the x- and y-axes, using the same colour coding. Colours describe the observed response rates in Hz in the onset time window (20-100ms).

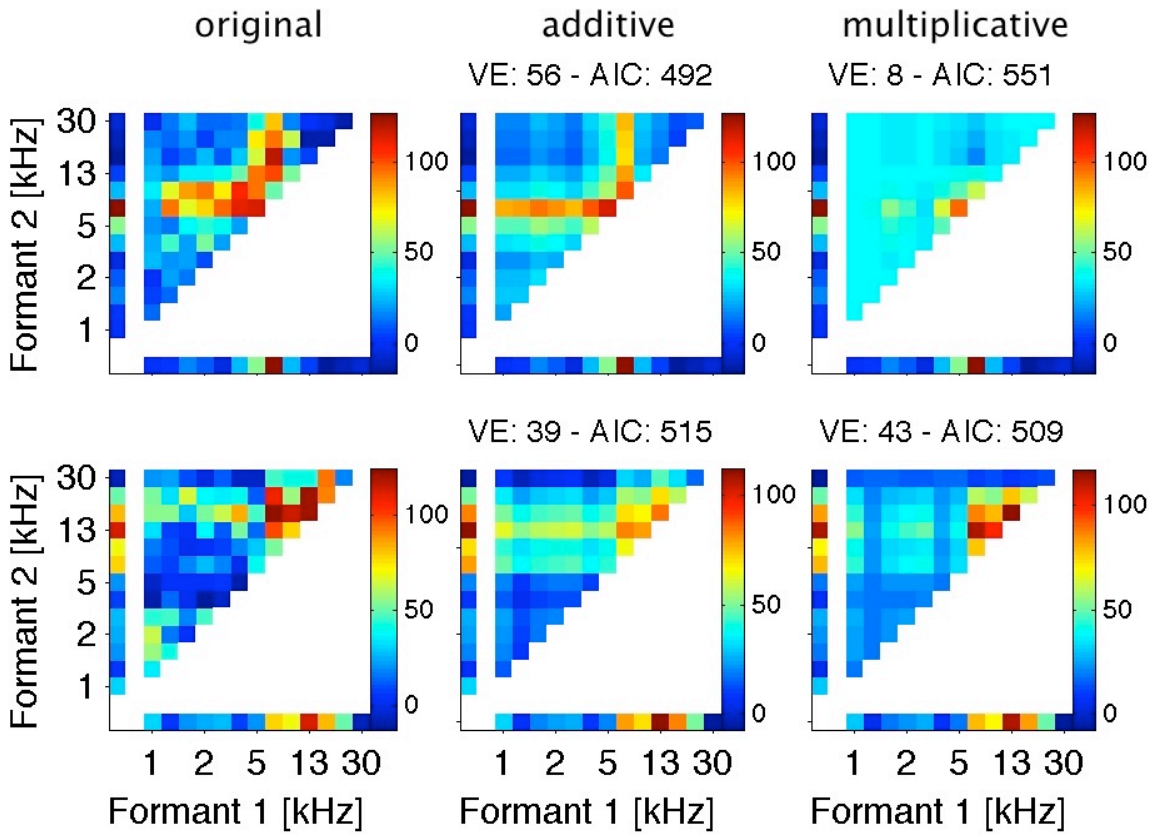


Figure 4.5. Comparison Of Observed FRFs With Additive And Multiplicative Fits Based On Observed Single-formant Responses

The observed FRF of the multi-unit presented in Figure 3.2 with three generated FRFs derived from fits of responses to single-formant stimuli. Model fits are additive, additive + multiplicative and purely multiplicative (see Methods). Responses to single formants are shown next to the x- and y-axes for comparison of single FRFs and double FRFs.

For the multi-unit in the top row of Figure 4.5. the additive model appears to give a much closer fit of the observed FRF than the multiplicative model. This impression is confirmed by %VE (additive: 56% vs. multiplicative: 8%) and AIC values (additive: 442 vs. multiplicative 551). In contrast, for the multi-unit described in the second row, the multiplicative models produces a slightly better fit (additive: VE 39%; AIC 515 vs. multiplicative: VE 43%; AIC 509).

Visual inspection of all recorded multi-units suggested that many multi-units would fall into at least two response classes as exemplified in Figure 4.5.: one that integrates the formant energy in an additive way and another that integrates formant energy in a multiplicative manner. Using the Akaike Information Criterion I determined how many of the 448 recorded multi-units were better fitted by either the additive or multiplicative model. Figure 4.6. shows the histogram of the number of multi-units that had lower AIC values for the type of model shown on the ordinate. A lower AIC value indicates a better model fit.

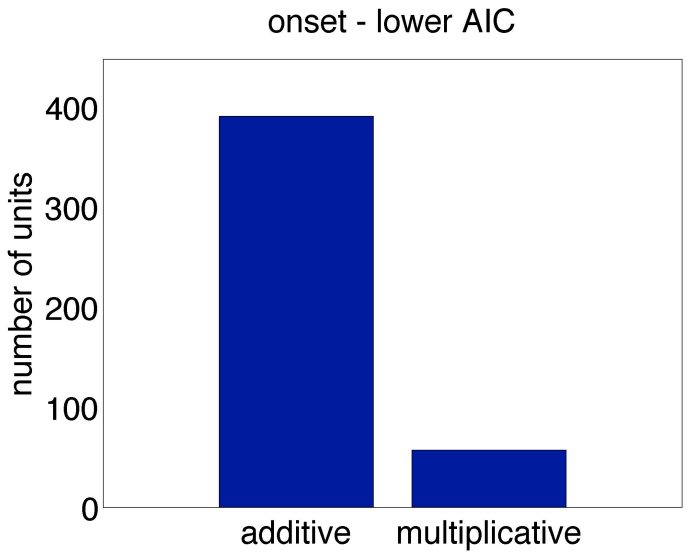


Figure 4.6. Number Of Multi-Units With Lower AIC Values For The Additive Or Multiplicative Fits Of Onset-FRFs

Number of multi-units for which the additive or multiplicative FRF fits based on single-formant responses produced lower AIC values. Lower AIC values describe better fits of an FRF model to the observed FRF. The additive model by far yields the higher number of better fitted FRFs.

The majority of FRFs (392/448 or 88%) were better modelled by additive integration of formant energy than by multiplicative integration. However, for a few multi-units, (56/448 or 12%) the multiplicative model provided the better fit. While this suggests perhaps two different classes of neural responses, that way of thinking about the

make-up of the cortical neural population is meaningful only if whichever model is found to be "better" does in fact account for a nontrivial amount of stimulus dependent response variance. In cases where neither the additive nor the multiplicative model provides a good fit to the FRF, asking which is better is moot. I therefore determined how well the additive and multiplicative models performed for these multi-units by comparing the amount of variance the respective models were able to explain (see Methods section 4.2.6 for the definition of variance explained). Figure 4.7. shows the distribution, for all multi-units, of the percentage of variance explained by the linear and the multiplicative model respectively.

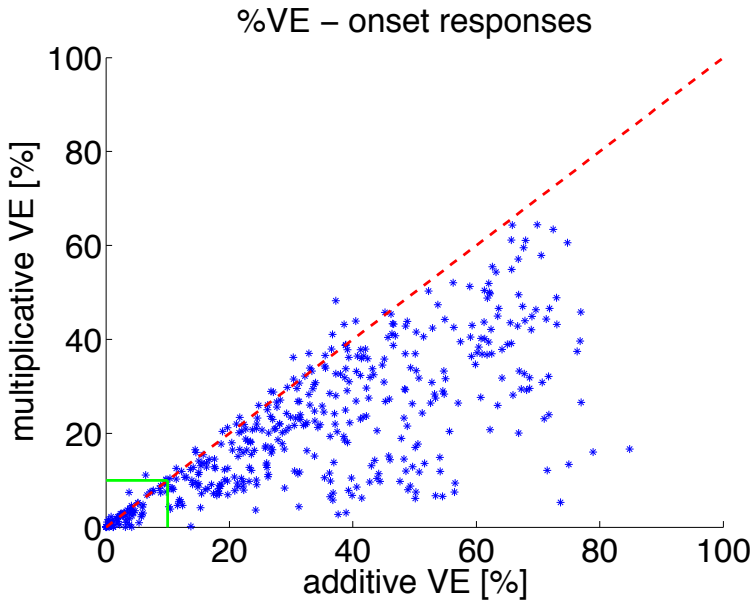


Figure 4.7. Variance In FRFs Explained By The Additive Model Versus The Multiplicative Model Of Formant Integration

Percentage of variance explained by the additive model plotted against the variance explained by the multiplicative model for all 448 recorded multi-units. Each blue dot represents the percentage variance explained by either of the two models. The green box near the origin contains the 72 multi-units for which neither model could explain more than 10% of the explainable variance. These multi-units were excluded for the estimation of linearity.

Three observations catch the eye. First, the amount of variance that the models are able to explain varies greatly from one multi-unit to the next, and in 72/448 (16%) of multi-units, less than 10% of the variance of FRFs is explained by either the additive or the multiplicative model of formant integration. Second, in many cases the additive and multiplicative models appear to explain a similar percentage of the variance (those multi-units near the main diagonal). And third, the data points lie either near or well below the main diagonal, indicating that the additive model is often clearly superior to, and never much behind, the multiplicative model in the amount of variance it can explain. In only 1 FRF the multiplicative model explained more than 10% more variance than the additive model. On average the additive model explained 33.6% (inter-quartile range: 16.4-49.3%) of the explainable variance and the multiplicative explained 21% (inter-quartile range: 8.6-31.2%), an average advantage for the additive model of 12.6%.

To further quantify the relative explanatory power of the additive and the multiplicative models, I calculated the linearity index for each multi-unit as:

$$LIN = \log(\frac{VE_{add}}{VE_{mult}})$$

This index is positive if the multi-unit is better described by an additive model, negative if it is better described by a multiplicative one.The 72 multi-units with a low (< 10) %VE were excluded from this analysis, given that ratios of numbers near zero do not produce stable and reliable statistical estimates. Figure 4.8. shows the histogram of the linearity indices for the 376 multi-units with %VE > 10. The distribution is overwhelmingly skewed to the positive side of the x-axis which is associated with a superiority of the linear over the multiplicative model.

Together, the observations made in this section argue rather strongly against the hypothesis that neurons in A1 might be conceived as "combination sensitive elements" which would develop vowel-class specific responses by combining sensitivity to individual formants in an AND-like, multiplicative manner.

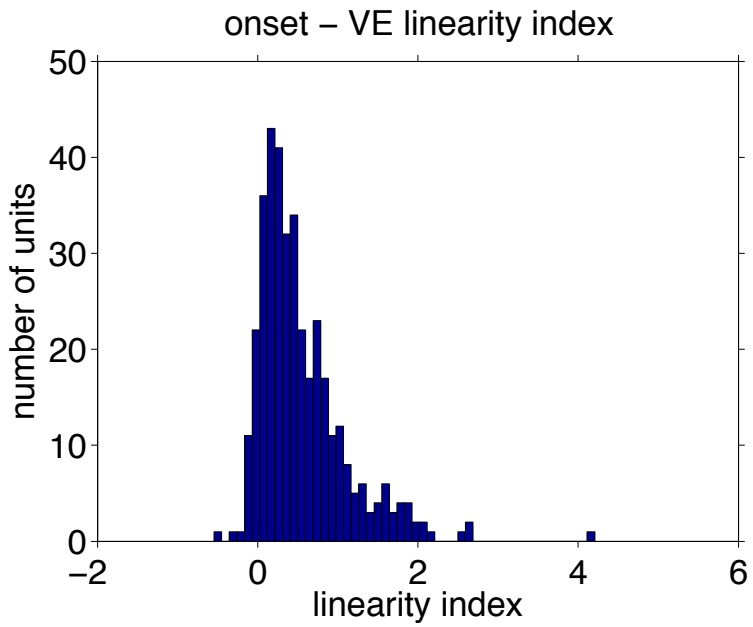


Figure 4.8. Linearity Of Formant Integration By A1 Multi-units.
Distribution of linearity indices for all multi-units with at least 10% variance explained by either the additive or the multiplicative models. Indices are predominantly positive and the peak of the distribution lies close to zero. The histogram does not show a clear clustering of indices into different classes.

4.3.3. Improvement By Combining Additive And Multiplicative Models

Although the additive model is almost always better than the multiplicative, when it comes to predicting responses to double-formant vowels from responses to single-formant vowels, the additive model nevertheless often left a sizeable proportion of the stimulus related variance unexplained, and the multiplicative model, albeit inferior, nevertheless often explained some variance. This raises the question of whether a model which combines additive and multiplicative factors could improve the amount

of variance explained over a purely additive one. To investigate this, I fitted combined additive-multiplicative models to each observed FRF, which took the form:

$$R_{i,j} = a + b \cdot \left(F_i + F_j \right) + c \cdot \left(F_i \cdot F_j \right)$$

This combination model has one more free parameter than the purely additive one, and for that reason alone would be expected to provide a better fit of the data. However, AIC values calculated for these models compensate for the number of parameters. Comparing the AIC values for the additive and combined models (see Figure 4.9.) shows that for 104/448 multi-units (24%) the additive model provides a better fit of the double-formant responses. 344/448 multi-units (76%) are better described by the combined model. Although the AIC analysis tells us which model is the better fit of the observed data, it does not quantify the size of the improvement of the combined over the additive model. This is addressed in Figure 4.10., which shows how much the %VE improved in the 344 multi-units in which the combined model was a better fit of FRF structure. For most multi-units (319/344, 93%) the combined model improved the %VE by only very modest amounts of less than 5%. For 37 multi-units (8.3%) the improvement was between 10%VE and 20%VE, 12 multi-units improved by 20-30%VE and 2 by 30-45%VE.

In summary, only in 51 multi-units (11.4%) could the addition of a multiplicative nonlinearity improve the amount of explained variance for formant integration by more than 10%. The improvement conferred by adding a multiplicative term to the model of the FRF is often statistically significant in terms of lowering AIC values, but only rarely physiologically significant since the improvements in model fits are sizeable only in a small subset of A1 multi-units.

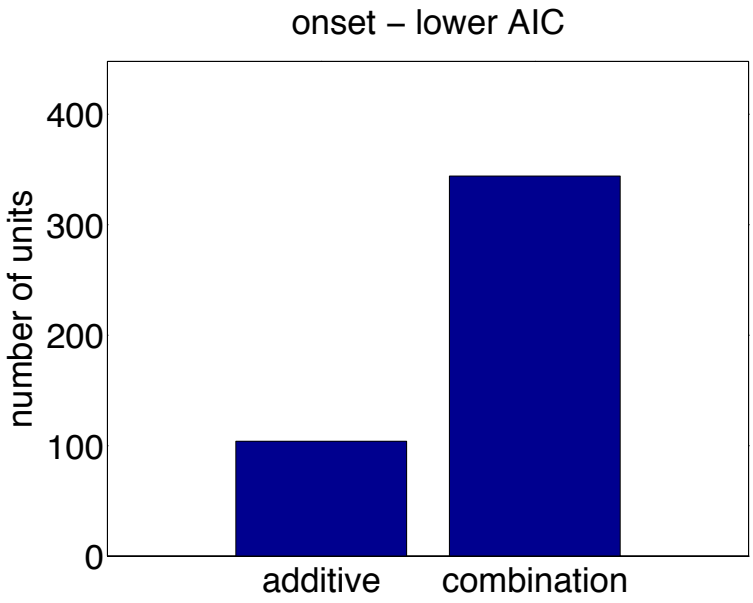


Figure 4.9. Number Of Multi-Units With Lower AIC Values For The Additive Or Combination Fits Of Onset-FRFs

Number of multi-units for which the additive or combination fits based on single-formant responses produced lower AIC values. Combination fits are generated by combining the additive and multiplicative model. Lower AIC values describe better fits of an FRF model to the observed FRF. The combination model yields the higher number of better fitted FRFs.

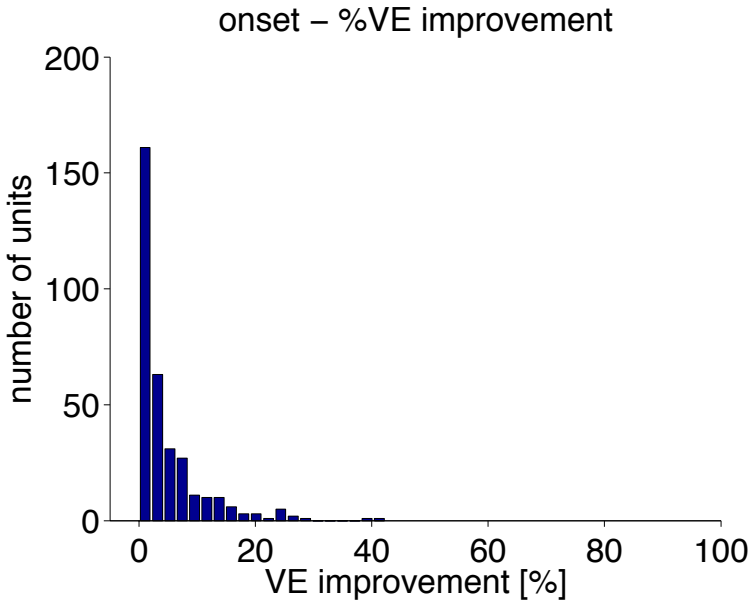


Figure 4.10. Improvement Of Variance Explained

Improvement of model predictions when combining the additive model with multiplicative model of formant integration. In only 11% of the unite the combination model could improve the variance explained by more than 10% compared to the additive model

4.3.4. Anatomical Clustering Of Formant Integration

This apparent diversity raises the question whether neural multi-units cluster anatomically with respect to how linearly they integrate formant energy in double-formant vowels? One could envisage a number of different scenarios. For example, classic studies by Hubel and Wiesel (1963) have described visual cortex as consisting of functional "columns", where neurons along a vertical penetration all share similar orientation or binocular tuning properties. By analogy, A1 might exhibit columns where all neurons along a vertical penetration integrate frequency channels in a similarly linear or non-linear manner. Alternatively, one could imagine that A1 response properties become increasingly non-linear as information is processed in cortical "canonical microcircuits" (Douglas and Martin, 1991) which run from the thalamorecipient, granular, middle layers first up to the superficial and then to the deep layers of cortex. Or it is of course possible that response linearity varies randomly throughout A1, without clear anatomical clustering.

To investigate this question I sorted all 448 linearity indices by penetration number (1-26) and cortical depth (1-16), and performed a two-way ANOVA to test for main effects of penetration and depth. This analysis revealed no significant effect of cortical depth ($F = 0.95$; $p = 0.51$) but a highly significant effect of penetration ($F = 6.62$; $p = 3.6 \cdot 10\text{-}17$). This clearly indicates that A1 response linearity follows a columnar arrangement. I also tested whether the improvement of %VE by the combined model is dependent on cortical depth or position in the cortex or both. A 2-way ANOVA again revealed no significant main effect of cortical depth ($F = 1.06$; $p = 0.39$), but a highly significant effect of penetration ($F = 2.96$; $p = 4.6 \cdot 10\text{-}6$), again indicative of a columnar structure.

4.3.5. Comparison Of FRFs For Onset And Offset Responses

Many multi-units in my dataset had clear offset responses to both single- and double-formant vowels (Figure 4.2.). Pronounced onset and offset responses are clearly visible. Offset spiking response were evoked by *both* single and double-formant stimuli. The number tuning peaks in offset responses to single-formant vowels was slightly different compared to the onset case and is shown in Figure 4.11 with the majority of multi-units still responding to only one formant frequency with a peak response. The explainable variance in offset responses was also never smaller than twice the yardstick level, and had a median of 6.4 times yardstick (interquartile range [5.4 10.3]). Therefore all multi-units qualified for model fitting. I analysed the offset responses in the same way as the onset responses to find whether the integration properties in both response phases notably differ. In particular, I was interested in the question whether the longer integration and neural processing time available to offset responses would favour a more nonlinear mode of formant integration.

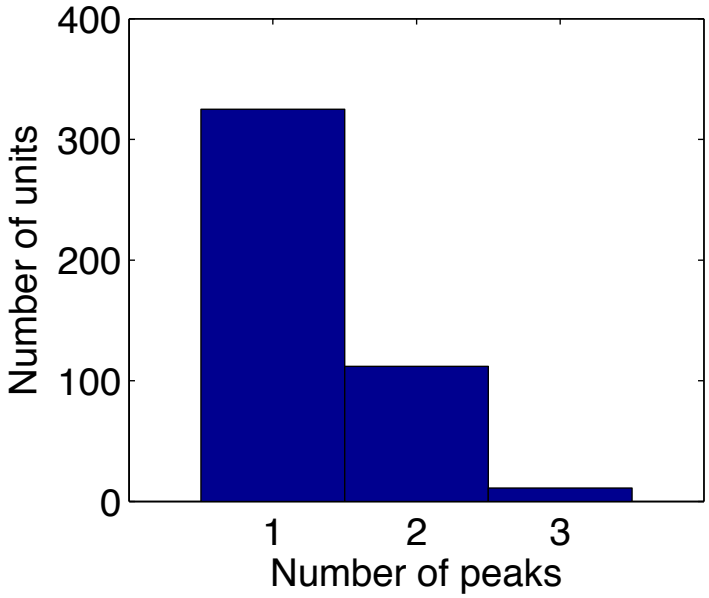


Figure 4.11. The Number Of Peaks In Offset Responses To Single-formant Stimuli

Number of tuning peaks in offset responses to single-formant vowels for all 448 multi-units. The number of double-peaked frequency tuning to single formants is slightly increased compared to onset responses.

Figure 4.12. shows four examples (one per animal) of onset- (left column) and offset- (right column) FRFs. Offset-FRFs generally appear to be somewhat negatively correlated to onset-FRFs. Parts of formant space which elicited strong onset responses (shown in yellow or red) often exhibited only weak offset responses or even inhibition (blue), and vice versa. I investigated this impression of a negative correlation of onset and offset-FRFs by calculating their cross-correlation in all the 448 multi-units. The results are shown in Figure 4.13. The distribution of correlation coefficients is almost exclusively negative with a mean of -0.26 (0.01 s.e.m., interquartile range [-0.41;-0.1]), confirming the visual impression of negatively correlated FRF structures for onset and offset responses in Figure 4.12.

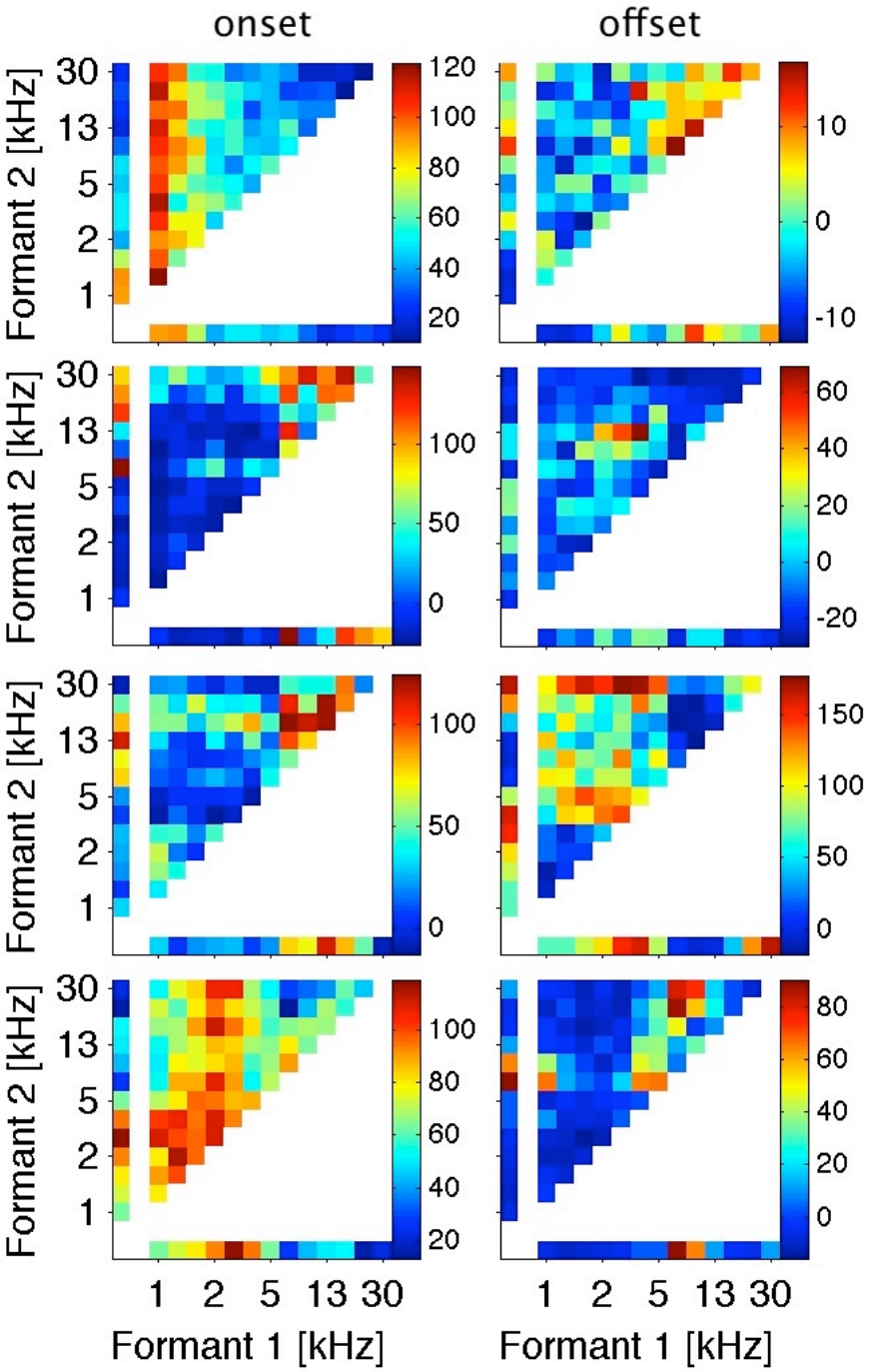


Figure: 4.12. Onset- And Offset-FRFs Appear Negatively Correlated.

Example onset and offset-FRFs (columns), one per recorded animal (rows). The structure of the FRFs in from onset and offset responses appears negatively correlated. Color code for spike rates in Hertz in the onset (20-100 ms) and offset (200-280 ms) windows. Note that scales of responses strength are different for onset and offset responses to emphasise FRF structure in both response periods.

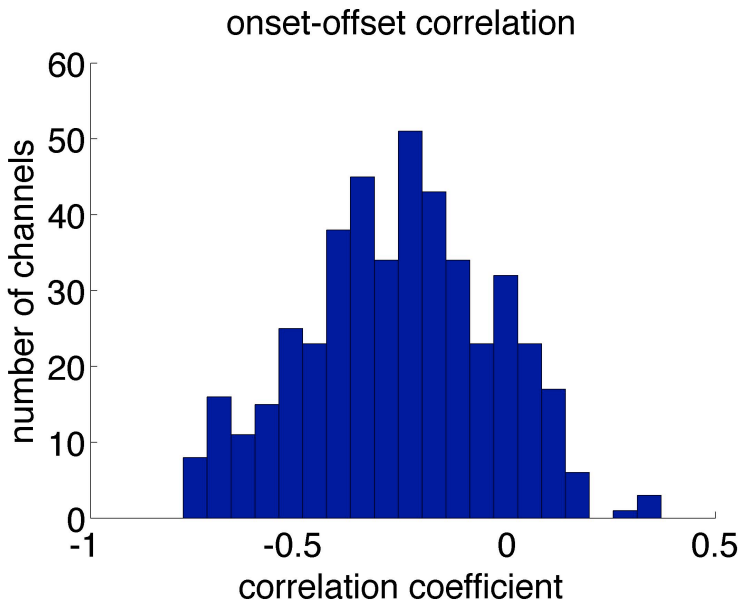


Figure 4.13. Onset- And Offset-FRFs Are Negatively Correlated

Onset and offset-FRFs are on the whole negatively correlated. The average negative correlation is moderate (R ~ -0.26). The moderate negative correlation confirms the visual impression of the relation between onset- and offset-FRF structure on Figure 4.12.

4.3.6. Comparing Onset And Offset Responses

Although there is a clear negative correlation between onset and offset receptive fields for double-formant vowels, the onset and offset-FRFs are not inverse images of each other. Correlation coefficients are relatively modest on average (R ~ -0.26). If onset and offset-FRFs were inverse images, then additive or multiplicative models should be equally good at fitting the onset and the offset-FRF. Looking at the numbers of minimal AIC values for models of offset-FRFs (Figure 4.14. left), we see that again the additive model fits the offset-FRFs better in 301/448 (67%) multi-units, compared to 88% in onset-FRFs. The proportion of multi-units which were better fit by the multiplicative model increased to 147/448 (32%), compared to only 12% in onset responses.

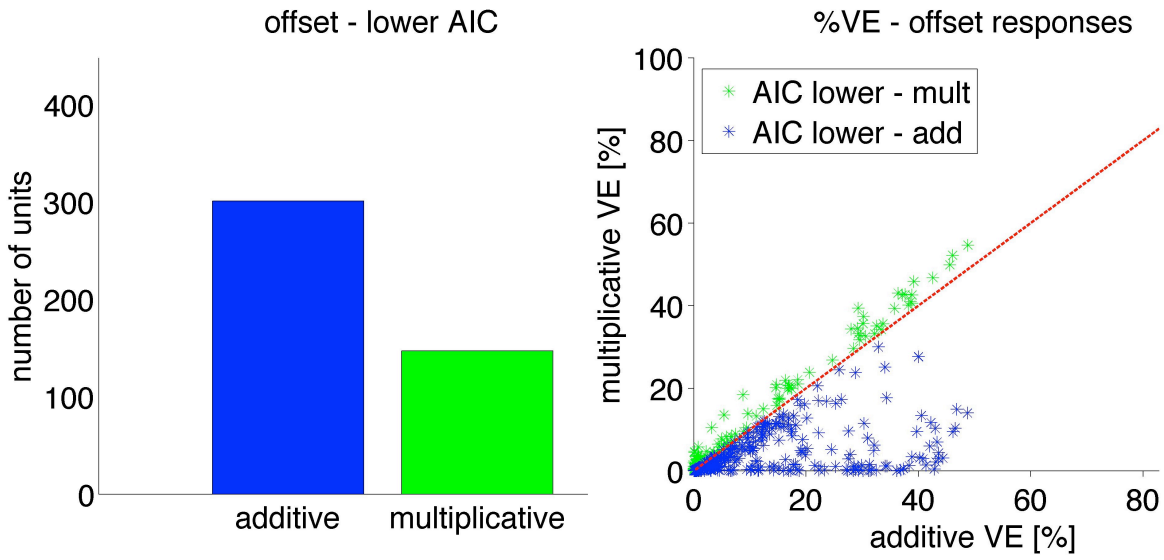


Figure 4.14. AIC Values And Percentage Variance Explained For Additive And Multiplicative Models Of Offset-FRFs

*(**Left**) Number of multi-units for which AIC values were lower when modelling their offset-FRFs with the additive or the multiplicative model. The additive model is the better fit of offset-FRFs for most multi-units (lower AIC values describe better fits). (**Right**) Amounts of variance explained by the additive versus multiplicative models in offset-FRFs. Each data point represents the %VE by either model in the offset-FRFs of one multi-unit. Units are colour coded with respect to whether additive model (blue) or the multiplicative model (green) produced lower AIC values as seen in the left plot: lower AIC values translate into higher variance explained.*

From the relative success of the additive model for both onset and offset responses in a majority of multi-units (269/448, 60%) one may assume that the properties of formant integration are similar in both response phases. However when we compare the amount of explained variance by either model in the two phases a different result emerges. Figure 4.14. shows, in the right panel, the amount of variance explained for all offset responses. Bars and marker in both plots are coloured depending on whether multiplicative or additive models produced lower AIC values for the represented multi-units. The distributions confirm that lower AIC values translate directly into higher amounts of explained variance.

In the offset-FRFs the additive model explained on average 11.6% of the explainable variance; in the onset-FRFs it explained 33.6%VE. Similarly, in the offset-FRFs the multiplicative model explained 6.6% of the explainable variance, in onset-FRFs it explained 21% in onset-FRFs. For the individual multi-units I find that in 340/448 (76%) both the additive and the multiplicative models explained less variance in the offset-FRFs than in the onset-FRFs. In only 39 multi-units (9%) both models explained more variance in the offset-FRFs. In the remaining 69 multi-units (15%) one of the models explained more variance in for the offset-FRFs (additive: 49 multi-units, multiplicative: 20 multi-units).

In conclusion, although the additive model still does better than the multiplicative model in accounting for FRF structure in offset responses, the amount of variance it can explain is substantially decreased in this response phase. This result indicates an overall decrease in linearity in offset responses relative to onset responses. However, this decrease in linearity cannot be captured by a simple multiplicative nonlinearity, which instead performs even worse in offset-FRFs.

4.4. Discussion

In this chapter I investigated the manner in which cortical multi-units in A1 of the rat integrate the frequency components contained in the spectrum of double-formant vowels. If multi-units in A1 were sensitive to formant combinations, they might be expected to perform a "logical AND-like" operation, which requires nonlinear interactions of a multiplicative type between frequency bands. I used a simple multiplicative model to explore whether nonlinear interactions describe the structure of formant integration more accurately than a purely additive integration model can.

4.4.1. Additive (linear) Versus Multiplicative (non-linear) Formant Integration

The AIC analysis showed that in the vast majority (87%) of the analysed onset-FRFs the additive model of formant integration delivers a better fit than the multiplicative model. When comparing the actual strength of either model for onset-FRFs, I found that the additive model explained on average 12.6% more variance. These results express an advantage of linear formant integration over multiplicative integration, however not an overwhelming dominance. The distribution of linearity indices for all multi-units, based on the relative variance explained by the additive and multiplicative models, demonstrates that almost all multi-units integrate the formants in double-formant vowels in a more additive than multiplicative manner.

Nevertheless, for onset responses of many multi-units the multiplicative model as well explained a substantial fraction of the variance in FRFs. This observation could indicate a mixed additive-multiplicative integration of formant energy. I found however that in only 11.4% of the analysed multi-units a combined model explained more than 10% more variance than the simple additive model alone. From this I conclude that A1 multi-units in the rat generally did not employ multiplicative, logical AND-like operations to combine inputs across frequency channels and that logical OR-operations alone can explain formant integration properties in onset-FRFs with some success.

Vowel classes can cover a large area of formant frequency combinations in the F1-F2 plane. The distance between the formant centres in human vowel classes are also often quite large (see the classical plot by Peterson and Barney, 1952, in Chapter 1, Figure 1.3.). For example the F2 centre of the prototype English vowel /i/ (Peterson and Barney, 1952) is roughly 4.5 octaves above the F1 centre. In an early publication

Langner et al., (1981) reported finding a number of neurons in the area L (the bird homolog of mammalian A1) of the Mynah bird, which were “selective” formant combinations with a relatively large distance between formants. Langner and his colleagues also generated additive models based on single-formant responses to fit observed responses to double formants. Their models described areas of formant combinations in the F1-F2 plane to which the birds’ L-neurons respond strongly. However, their models always produce a strong response near the F1=F2 diagonal but these formant combinations were not used as stimuli by the authors. To achieve *selectivity* for combinations of formants at the distances Langner et al. used, a nonlinear (inhibiting) formant interaction is required near the main diagonal of the F1-F2 space. Otherwise these models will predict strong responses for one formant alone, contradicting the claim of combination sensitivity. My combined model is able to produce some such suppression if the F1*F2 term takes a negative weight. However, adding a multiplicative term to the additive model did only very rarely lead to substantially improved model fits, rejecting the hypothesis of combination sensitivity in A1, through multiplicative suppression of formant combinations near the main diagonal of the F1-F2 space.

Although for some multi-units the additive model generated near perfect fits of their FRF, on average it explained only 33.6% of the explainable stimulus related variance. The variable, and sometimes very modest, success of this linear model in accounting for FRF structure implies that A1 responses often use substantially nonlinear interactions in their response to broadband sounds like vowels, even if these non-linearites do not appear to be multiplicative in nature. In a previous study of A1 of the gerbil, Ohl and Scheich (Ohl and Scheich, 1997) reported that a fourth order polynomial model of formant integration adequately described the responses to

double-formant vowels. They argue that inhibitory formant interaction may depend on the distance between the formants, produced by distance-dependent nonlinear interactions. Yet, they report essentially additive interactions between formants that are relatively close to each other, which would accord with my results.

4.4.2. Diversity Of Integration Properties In A1

I find that the linearity of responses varied strongly from one vertical electrode penetration to another. Furthermore, the (often small) improvements that the combined model was able to produce compared to the additive model also varied from one penetration to another. The patchwork-like organisation of A1 that this suggests is in good agreement with recent observations in ferret A1 reported by Bizley et al. (2009). These authors presented 64 different vowel sounds (4 different artificial vowels at 4 different voice pitches from 4 different directions) to anaesthetised ferrets while recording A1 single-unit responses. One main observation in this study was that A1 units varied greatly in their sensitivity to vowel type, were often sensitive to at least two dimensions of the presented stimuli (pitch, timbre, direction), and frequently “exhibited significant nonlinear interactions” that were “most pronounced for pitch and timbre”. However, while (Bizley et al., 2009) found significant differences of responses to different feature combinations with cortical depth as well as location of penetration, I did not find significant differences in the amount of variance explained as a function of cortical depth. Yet, my results agree with the observation of Bizley et al. and of many earlier studies of a columnar arrangement of integration properties (Suga, 1965; Abeles and Goldstein, 1970; Merzenich et al., 1975; Schreiner and Sutter and Schreiner, 1991; Clarey et al., 1994; Mendelson et al., 1993; Brugge and Merzenich, 1973; Mendelson, 1990;

Middlebrooks et al., 1994; Mendelson et al., 1997), and in agreement with my finding in Chapter 2 of a columnar arrangement of formant resolution in A1.

4.4.3. Offset Responses To Double-formant Vowels Are Less Linear Than Onset Responses

The perception of speech sounds relies on a processing or temporal patterns, where offset responses could play an important role. For example, the rate of speech can change drastically between speakers, and can change the perception of speech classes that depend on the length of voiced speech components (Miller and Volaitis, 1989). Changes in speech rate manifest in variation in the duration of phones, and affect identification of consonants in syllables with different vowel lengths (Pind, 1998). However, human listeners easily adapt to changes in speech rate (Pind, 1995), and presumably use information about temporal cues such as the offset times of vowels to do so.

It is well known that cortical neurons, unlike those in the brainstem or midbrain, hardly ever produce sustained responses to steady-state sounds (Phillips, 1985; Calford and Semple, 1995; Heil, 1997; Schnupp et al., 2001; DeWeese et al., 2003), but more typically mark their onsets, and sometimes also their offsets, with transient responses. Also, information about the identity of auditory stimuli can persist for some time after stimulus offset in the firing of A1 neurons in the cat (Volkov and Galazjuk, 1991; Moshitch et al., 2006; Qin et al., 2007), ferret (Garcia-Lazaro et al., 2006; Walker et al., 2011), and rat (Takahashi et al., 2004). Observations like these suggest that offset responses may carry important information about the temporal structure of speech sounds as well as about the identity of phonemes that were perceived a few hundred milliseconds in the past.

Here, I observed that FRFs derived from offset responses to double-formant vowels were usually negatively correlated with onset receptive fields, with a mean correlation coefficient of around -0.24 ([-0.4;-0.1] interquartile range). This moderately negative correlation shows that offset-FRFs are not simply the negative images of their corresponding onset-FRFs. When using offset responses to single-formant vowels to fit additive and multiplicative models, I found that for most of the units both models explained considerably less variance in the units' offset-FRFs compared to their onset-FRFs. This decreased explanatory power of the additive model indicates that offset responses integrate the acoustic energy in the two formants in a less linear manner than is the case for onset responses. And the putative non-linearity that would describe formant integration in offset responses is, just as for onset-FRFs, not a logical AND-like, multiplicative interaction.

A study by Qin et al. (2007) was the first to explicitly explore differences between onset and offset receptive fields of A1 units in awake cats. The authors used pure tone stimuli of 500 ms length to stimulate A1 in passively listening animals. They then compared onset and offset frequency receptive fields at the sound pressure level that elicited the strongest onset responses. The determined frequency-level curves (fRs) and report that "offset-fRF is quite different from onset-fRF in most cells [...] and A1 cells had diverse combinations of onset- versus offset-fRF." In most cases onset-FRFs were not the negative image of onset-responses, but rather best frequencies for offset responses were either higher or lower in frequency compared to onset responses. The fact that the offset-FRFs I observed in response to double-formant vowels are not clear negative images of onset-FRFs is in agreement with the result by Qin and colleagues.

One obvious difference between the study of Qin et al. and the experiment reported in this chapter is that Qin et al. recorded responses in awake animals while my rats were anaesthetised. Barbiturate-anaesthesia appears to render offset responses a relatively rare phenomenon, which was often interpreted as rebound from inhibition (Phillips et al., 2002). However ketamine-anaesthesia has been shown to allow for more complex temporal response properties of A1 neurons (e.g. Volkov and Galazjuk, 1991). I used a medetomidine-ketamine based anaesthesia for acute cortical recordings and this regime may explain why the offset responses I observed yielded sufficient explainable variance to fit models of formant integration. Also, responses to pure tones, as used by Qin et al., are not directly comparable to responses to broadband sounds, as in my experiment, especially if cortical units do not integrate frequency components completely linearly, as I conclude for offset responses in particular.

The only study I found that has investigated offset responses in the rat A1 was conducted by Takahashi et al. (2004). The authors used ketamine-xylazine anaesthesia and recorded auditory evoked potentials (AEPs) from the cortical surface of A1 in response to pure tones. Their general finding is that pure tone receptive fields in AEPs have peaks at the “fringes” of the single tuning peaks of onset receptive fields. This observation supports the perspective that offset responses are a kind of after image due to rebound of inhibition. This is however not supported by my observation that formant integration is even less linear in offset responses than in onset responses when using broadband double-formant vowels as stimuli. In summary my results from fitting offset-FRFs, based on offset responses to single formants, indicate that offset receptive fields are not mere after images of onset responses, at least not for more complex stimuli than pure tones, and that the

receptive fields I observe are a consequence of nonlinear interactions between bottom up inputs (or their cessation) and top down cortical inputs into A1.

4.5. Summary & Conclusions

In this chapter I investigated the second cornerstone (the first being discriminability) of auditory class processing: i.e. the way multi-units in A1 of the rat integrate the information present in combinations of acoustic features.

A1 multi-units in the anaesthetised rat integrated combinations of two formants in a predominantly additive, that is linear, manner. Nevertheless most receptive fields had substantial amount of unexplained variance in the way they integrate the energy contained in two formants, arguing for a the ability to A1 cortical multi-units to represent more diverse features than simple Boolean combinations of individual acoustic components. This was even more the case for offset responses.

Human vowel classes are perceptually defined by combinations of formants. Neural FRFs, to match human vowels classes, would require both positive interactions among frequencies as well as suppression of responses with increasing distance from the “preferred” formant combination. The results presented here clearly show that neural responses in A1 do not exhibit such response properties, at least in the naive, anaesthetized rat. The situation could of course be very different either in awake animals which have been trained to distinguish vowel classes and are engaged in such a classification task, or in other, higher order brain areas. The results I present here thus present a baseline against which further studies can be compared.

One higher order brain structure which may deserve attention in this context and which has attracted much attention as a potential seat of declarative memory is the hippocampal formation. Highly abstracted representations of sensory "objects" have been described for responses in CA1 pyramidal cells in humans and animals. Consequently this area is also a potential candidate for the neural representation of auditory object classes, including vowels. In the next chapter I describe results from a study which I carried out in collaboration with colleagues in Trieste, Italy, and which aimed to find how CA1 cells respond to simple vowel sounds taking into account the strong response of these units to spatial information.

CHAPTER 5 – HIPPOCAMPAL REPRESENTATIONS OF SINGLE-FORMANT STIMULI

We have seen in Chapter 4 that multi-units in the rat A1 integrate formant energy in a fairly linear, additive manner, and in as far as the responses are non-linear, the non-linearities are not of the multiplicative, Boolean AND type that one might expect to generate responses which are truly selective to formant combinations.

A known candidate area for abstract representations of objects is the hippocampal formation. In a collaboration with a laboratory specialised in hippocampal research at the 'Scuola Internazionale Superiore di Studi Avanzati' (SISSA) in Trieste, Italy, we conducted a study to find whether hippocampal single units in awake rats discriminate vowel stimuli after discrimination training. A natural question in this endeavour was also how the well-studied spatial response properties of hippocampal cells would interact with sound related inputs.

Ideally, to facilitate the comparison between A1 and hippocampus, the studies should have used the same types of double-formant vowels, However, for a variety of historical and practical reasons, we decided to use single-formant vowels for the hippocampal work described here. This was done in part to keep the complexity of behavioural tasks and their analysis at a minimum,.The behavioural setup at SISSA uses a 2 alternative forced choice task with 2 different response locations. As will be described further below, many hippocampal neurons exhibit "place cell behaviour", i.e. their firing rate may be modulated by the animal's changing its position in space. To disambiguate space related from sound related firing we therefore had to teach the animal to associate several distinct sounds with each response location. We also

needed to be able to collect a large number of trials per stimulus. And we wished to keep the experimental design essentially similar to a previous study from the SISSA group which had investigated tactile responses to different textures in hippocampus (Itskov et al., 2011). These factors constrained us to use a relatively simple stimulus set, and we settled for a set of four distinct single-formant vowels. Yet even though this study of hippocampal responses used a somewhat different set of stimuli and stimulus classes from those used in the previous chapters, they nevertheless create an interesting complementary view on how different parts of the central nervous system represent complex sounds.

The outcome of the collaboration described in this chapter has been published in the Journal of Neurophysiology (Itskov et al., 2012), with me as a co-author. Large parts of the following text and all figures in this chapter are adapted from this publication. I acknowledge the respective contributions of different members of this scientific collaboration at the end of the Methods section.

5.1. Introduction

Responses of hippocampal neurons represent an animal's location in the environment (O'Keefe, 1976; Wilson and McNaughton, 1993; Leutgeb, 2004; Leutgeb et al., 2005; Leutgeb, 2005; Wills, 2005). A growing body of work documents that hippocampal neurons also represent individual sensory stimuli and other non-spatial events (Christian and Deadwyler, 1986; Eichenbaum et al., 1987; Wiener et al., 1989; Moita et al., 2004; Hok et al., 2007; Lenck-Santini et al., 2008; Manns and Eichenbaum, 2009; Itskov et al., 2011). Since hippocampus contains a prominent representation of location in space, one can describe and quantify a putative stimulus representation only after teasing it apart from head and body position-dependent

responses associated with the stimulus-guided behaviour. Despite the body of work demonstrating the representation of non-spatial, sensory stimuli in the hippocampus, few studies have focused on the distinction between spatial representation and stimulus representation.

Recently (Itskov et al., 2011) demonstrated that hippocampus contains a robust representation of tactile stimuli: many neurons fired at different rates when the rat encountered surfaces of different roughness even when the action associated with those stimuli was identical. Still, in separate locations within the same experimental room, independent populations of neurons responded to the same tactile stimuli, indicating that responses to the tactile stimuli were modulated by the location of the animal. They also found that the explicit behaviour of the animal (the direction of turn associated with reward) and tactile stimuli were represented by independent populations of neurons.

The goal of this study was to confirm and extend the previous results from tactile stimuli to sound stimuli and to further elaborate on principles underlying the neuronal representation of sounds in the hippocampus. We therefore trained animals to associate four different sounds with two possible response directions. Because the behavioural responses to the two different stimuli associated with the same response direction were identical, any difference in the neuronal activity evoked by these two stimuli must reflect the coding of stimulus identity, rather than some aspect of behaviour. The experimental design allowed us to “isolate” stimulus-specific responses to vowels and to examine how such sensory responses depend on spatial and behavioural context.

5.2. Methods

5.2.1. Ethics Statement.

All experiments were conducted in accordance with National Institutes of Health, international, and institutional standards for the care and use of animals in research and were approved by the Bioethics Committee of the International School for Advanced Studies (permit n.5575-III/14) and supervised by a consulting veterinarian.

5.2.2. Animals.

Three Wistar male rats weighing about 350 g were housed together and maintained on a 14:10-h light-dark cycle. To ensure that the animals did not suffer dehydration because of water restriction, they were allowed to continue the behavioural testing to satiation (250–350 trials per session) and were given access ad libitum to drinking water for 1 h after the end of each training session. The animals' body weight and general state of health was monitored throughout.

5.2.3. Stimuli

Single-Formant artificial vowels were generated using binary click trains at the fundamental frequency of 152 Hz and 200-ms duration and were bandpass filtered with a bandwidth equal to 1/50 of the formant frequency using Malcolm Slaney's Auditory Toolbox (http://cobweb.ecn.purdue.edu/malcolm/interval/1998-010/). Formant centre frequencies were 1,520, 2,280, 3,420, 5,130 Hz, respectively. The spectra of all four stimuli are shown in Figure 5.1.B. The formant frequencies chosen here lie below the point of maximum sensitivity in the rat's audiogram (Heffner et al., 1994), but previous work (Eriksson and Villa, 2006) has shown that rats can easily learn to classify vowels that are distinguished by formant peaks entirely below 4 kHz. The stimuli were ramped on and off using a Hanning window and played three times in succession, such that the stimuli presented to the animals were 600 ms long. The stimuli were presented at a sound level of 70 dB SPL simultaneously from two

speakers located symmetrically on the two sides of the platforms of the training arena (see next subsection), so the sound level at the position of the animal's ears was equal when the animal was in the "nose poke" position at the start of the trial (see Figure 5.1.D).

For the training and for analysis of neuronal responses, it was crucial for the sounds to be easily distinguishable. For this reason, we ensured that the distance between formant peaks (0.585 octave) was well above the frequency difference at which rats reach plateau performance in tone discrimination (Sloan et al., 2009). Sounds were presented through Visaton FRS 8 speakers, which have a flat frequency response (less than +/- 2 dB) between 200 Hz and 10 kHz. The speakers were driven through a standard personal computer sound card controlled by LabView (National Instruments, Austin TX).

5.2.3. Training Arena

The arena consisted of two platforms, each with a layout as illustrated in Figure 5.1.C. The platforms were elevated 40 cm to encourage the animals to remain on the platform during the experiment. A small wall with a nose-poke hole was positioned at one end of each platform. Reward water spouts were located on either side of the nose poke, 3.5 cm from the lateral platform edge on the level of the platform. The two platforms faced each other with their nose-poke walls at a distance of 25 cm. The two platforms differed only by the texture of the floor and their spatial position and orientation. The whole setup was in a Faraday room and was illuminated with infrared light (880 nm) during testing sessions.

5.2.4. Auditory Classification Task.

The day before behavioural testing, animals were placed on a water-restricted dry food diet. They were trained to initiate each trial with a nose poke. This triggered the

presentation of one of the four stimuli. Different sounds were presented in random order. The animals did not have to wait for the end of a stimulus presentation but were free to respond at any time after stimulus onset. Right turns after presentation of the 1,520 or 3,420 Hz formants were considered correct responses, as were left turns after presentation of 2,800 or 5,130-Hz formants. Correct responses triggered delivery of a small water reward (~50 microlitres) from the spout on the corresponding side. Incorrect responses triggered no water delivery but caused a 3-s timeout during which no new trial could be initiated. Each session lasted until the rat stopped doing the task.

The reaction times (from sound onset to the approach to the reward site) were measured. Since our setup required the animal to move just 13 cm from nose poke to reward spout, reaction times are a better measure of behaviour than video tracking, where errors can exceed +/- 2 cm (Jensen and Lisman, 2000). Before and after the training session, the room was illuminated by visible light, so the rats had the opportunity to use visual cues to orient themselves relative to the layout of the experimental chamber and apparatus.

One key aim of this study was to dissociate neuronal coding of stimulus identity from any aspect of explicit behaviour (either ongoing or planned), for example, when the animals' response causes them to move into or out of hippocampal place fields. We therefore considered a neuron to be sensitive to the acoustic stimuli only if its response patterns differentiated cases where the stimuli differed even though the animal's behavioural responses were the same, i.e., it turned in the same direction, and did so with comparable speed, for both stimuli. By disregarding all the sessions in which the reaction times for the two stimuli associated with either response direction on either platform differed significantly ($P < 0.05$, Wilcoxon rank sum test,

14% of cases), we excluded the possibility that any systematic variability in the animals' motor response, which might be differentially encoded by hippocampal place cells, could be misinterpreted as stimulus related neuronal activity.

5.2.5. Surgery And Electrophysiological Recordings.

Once the rats reached >70% correct performance on both platforms, they were implanted with chronic recording electrodes. To do so, they were anaesthetised with a mixture of Zoletil (30 mg/kg) and xylazine (5 mg/kg) delivered intraperitoneally. A craniotomy was made above left dorsal hippocampus, centred 3.0 mm posterior to bregma and 2.5 mm lateral to the midline. A microdrive loaded with 12 tetrodes (25-µm-diameter Pl/Ir wire; California Fine Wire) was mounted over the craniotomy.

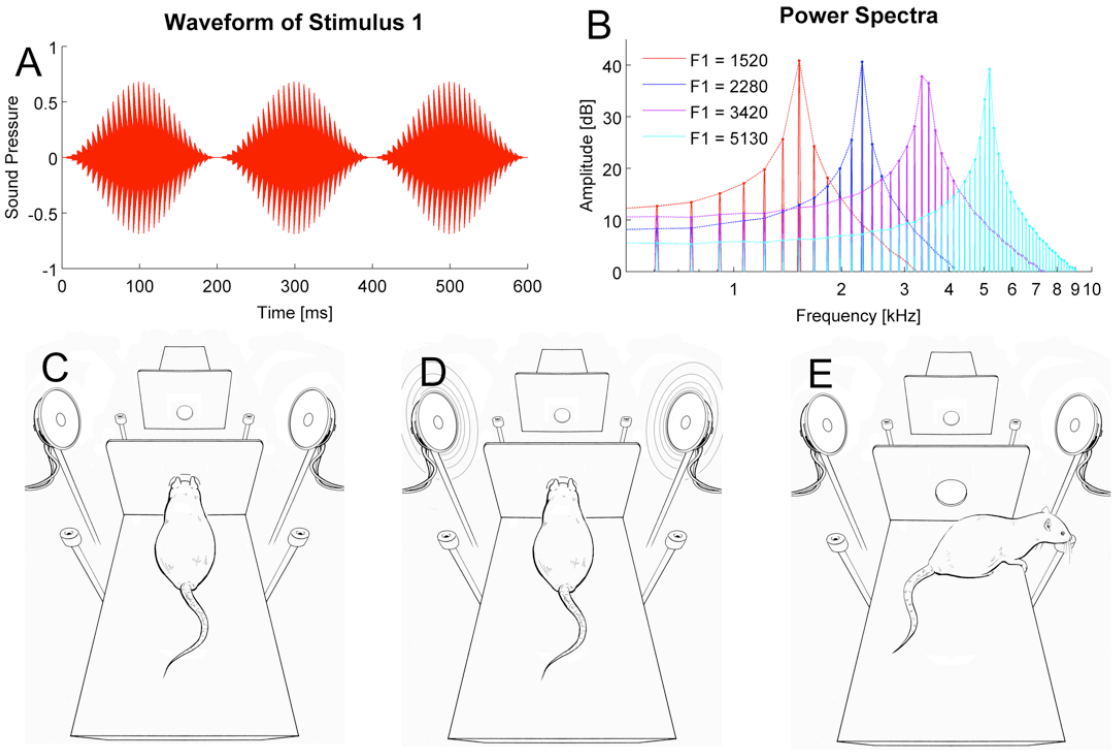


Figure 5.1. Stimuli And Behavirourl Task For Vowel Discrimination Training

*(**A**) Raw sound vector of one single-formant vowel used in training and during cortical recordings. All stimuli were played as triplets. (**B**) Spectra of each of the four single-formant vowels used. (**C-E**) sequence of steps in the training task: the animal has triggers by sticking its nose into a nose poke hole (**C**), a single-formant vowel is played through both speakers (**D**) and the animal decides which side the sounds associated with to collect a water reward (**E**).*

Tetrodes were positioned perpendicular to the brain surface and individually advanced in small steps of 40–80 µm per day until they reached the CA1 area, indicated by the amplitude and the shape of the sharp wave/ripples. All recordings were performed at a depth of 2,000–2,500 µm below the brain surface. After surgery, animals were given antibiotic (Baytril; 5 mg/kg delivered through the water bottle) and the analgesic caprofen (Rimadyl; 2.5 mg/kg by subcutaneous injection) for postoperative analgesia and prophylaxis against infection, and they were allowed to recover for 1 wk to 10 days after surgery before testing resumed.

After recovery from electrode implantation, the animals were retrained for one to two sessions to get them accustomed to performing the task with a microdrive and cable attached. They then performed regular behavioural tests with the classification paradigm described above, during which neuronal responses were recorded from the tetrode array using Neuralynx data acquisition equipment (analog-wired Cheetah acquisition system). Tetrodes had impedances between 100 and 500 kΩ. We usually recorded between one and six single units per tetrode. Spikes were pre-sorted automatically using KlustaKwik, with the use of waveform energy on each of the four channels of the tetrode as coordinates in a four-dimensional feature space. The result of the automatic clustering was inspected visually after the data were imported into MClust. Only well-isolated single units that exhibited a clear refractory period were included in further analysis. In addition, we used two measures of spike separation, the L-ratio and the isolation index, to define the separation between the clusters in the multidimensional space (Schmitzer-Torbert et al., 2005). In the vast majority of cases, each single unit was recorded only in one session. In rare cases, the same unit could be distinguished on 2 consecutive days (judged by the spike waveform, distribution of the inter-spike intervals, and the response pattern), in which case only the single session with better isolation quality was considered in the analysis.

5.2.6. Statistical Procedures

Firing rates usually varied over time, and the latency of the response varied across individual cells. Therefore, to quantify the effect of sound identity (or response direction) on neuronal activity, we used a method that would 1) take into account multiple time points, 2) provide a single measure of statistical significance of the effect, and 3) quantify the strength of the effect in each case. From among the many statistical methods that could be applied (ANOVA, classification techniques such as

linear discriminant analysis, etc.), we opted to use information theory measures because they provide a uniform scale that can be used to compare results from different experiments, their properties and biases have been extensively studied analytically, and robust bias correction procedures have been described and successfully implemented (Panzeri and Treves, 1999; Pola et al., 2003; Panzeri et al., 2007).

Shannon mutual information (Shannon, 1949) was used to quantify the statistical interdependence between trial parameter X (either stimulus identity or response direction) and neuronal spike count in a given temporal window. The information measures were computed in Matlab (MathWorks) using the information breakdown toolbox (Magri et al., 2009). Mutual information is given by the following equation *(Eq. 1)*:

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log_2 \left[\frac{p(x,y)}{p(x)p(y)} \right]$$

This quantifies the reduction of uncertainty about the trial parameter (X) gained by a single-trial observation of the spike count (Y). The probabilities in the above formulas are not known a priori and must be estimated empirically from the observed frequencies in a limited number of trials N. Inevitable inaccuracies in the resulting estimates of response probabilities can lead to an upward bias in the estimate of mutual information (Panzeri and Treves, 1999; Pola et al., 2003; Panzeri et al., 2007). An approximate expression for the bias has been formulated (Panzeri and Treves, 1999) and can be subtracted from the information estimates calculated with Eq. 1 above, provided that N is at least two to four times greater than the number of different stimuli or behavioural responses (Pola et al., 2003). In our data, N depended on the number of trials performed by the animals in a particular session but was

never less than 12 in any condition, whereas the number of classes was always 2 (corresponding to either the 2 response directions or the 2 different sound stimuli associated with 1 response direction). This allowed us to use the above-mentioned bias correction procedure (Panzeri and Treves, 1999). The information carried by spike counts was calculated in a 200-ms-wide window sliding in steps of 25 ms along the whole duration of the trial, from the sound onset to 1 s after. Since for each neuron we took into account many time points, multiple comparisons had to be accounted for. This was done using a permutation test, which tested the null hypothesis that the mean value of information obtained across all the time points could be expected by chance (see below).

Sound related information quantified how reliably a neuron's firing rate could distinguish between the two stimuli associated with the same response direction. First, the information about sound identity was measured independently at every time point from the sound onset until 400 ms after the sound offset. Each integration window was 200 ms long; the step size was 25 ms, yielding 41 sequential information values.

Next, to determine whether the mutual information carried by the neuron's firing rate was statistically significant, we used a permutation test that compared the observed mean information value (averaged across the 1-s analysis window) to a distribution of simulated values obtained after sound labels had been randomly reassigned to neuronal responses and the mean mutual information for these shuffled data had been computed. This resampling, repeated 200 times, provided an estimate of the distribution of average information values expected by chance if sound stimuli had no systematic influence on neuronal firing. P-values for the significance of sound related information were then calculated by comparing the information value for the observed

data to a Gaussian curve fitted to the distribution of 200 information values obtained by resampling. Neurons that were found to carry significant information about stimulus identity for at least one of the possible reward locations are referred to as “sound-sensitive neurons.”

Meta-test for the significance of sound related information in the neuronal population as a whole. Although the permutation test just described yielded small (highly significant) P values for many neurons in our data set, it was important to ask whether the number of significant outcomes observed across the population exceeded that expected from possible “false positives” that occur within any large sample. To ascertain this, we compared the distribution of sound information P values obtained across the entire data set (all 261 neurons) to a distribution of P values obtained from simulated neuronal responses created by 200 shuffles of the stimulus labels. We found that the distribution of observed P values was significantly smaller (Wilcoxon rank sum test, P = 0.0013) than that expected by chance. This means that the neuronal population carried a far stronger sound identity signal than could be expected based on firing rate fluctuations unrelated to the stimuli.

5.2.7. Statistical Significance Of Response Direction Related Information.

Response direction is the side to which the rat turned to obtain water. When the “turn right” and the “turn left” trials are compared, any difference in firing rate could be due to the place field of the cell, head direction, or predictive coding of the direction of the turn. The statistical significance of such differences was measured in the same way as was done for sound identity related differences in activity: at each time point, the firing rate in the trials with correct right turn was compared with the firing rate at the same time point in the trials with correct left turn. The information across all time points was averaged and compared with the distribution of average information

values obtained by random shuffling of the trial labels. Neurons carrying a significant quantity of information about the animal's response direction on at least one of the two test platforms are referred to as "response direction-sensitive neurons."

In theory, the difference between right-turn and left-turn trials could also be due to sound-dependent responses: for instance, if a neuron strongly responds to sound 1 (right) but does not respond to sounds 2, 4 (left), or 3 (right), it could result in spurious significant response direction information. To exclude cases in which soundrelated responses might cause an apparent response direction response, a neuron was classified as sensitive to response direction only if it showed a significant left vs. right firing difference for all four left vs. right stimulus pairings: sound 1 vs. sound 2, sound 1 vs. sound 4, sound 3 vs. sound 2, and sound 3 vs. sound 4.

Context dependence of sound sensitivity. The design of the experiment allowed us to test for unambiguous neuronal sensitivity to differences in sounds in four different conditions: discrimination of sounds 1 and 3, both of which directed the rat to the right water reward spout, and discrimination of sounds 2 and 4, both of which directed the rat to the left spout. Both discriminations were made on platforms 1 or 2. Consequently, one recorded neuron could exhibit sensitivity to sound identity (that is, carry a significant quantity of information about the sound stimulus) under a total of 0, 1, 2, 3, or 4 conditions. We compared the observed instances of sound sensitivity of each neuron to the number expected by chance if the sound sensitivity under each condition were independent.

The distribution expected by chance was calculated as follows. Under the assumption of independence, the number of sound-sensitive neurons associated with all four response directions is expected to equal Np^4, where N is the number of neurons in the sample and p is the probability that a neuron will carry a significant

quantity of information about sound at one of the four possible reward locations. Similarly, the number of neurons expected not to be sensitive associated with any of the four response directions equals $N(1-p)^4$. There are four different ways in which a neuron can be sensitive to sound associated with only one response direction (i.e., sensitive to the sound associated with the 1st, 2nd, 3rd, or 4th spout), and each of these possible outcomes has a probability of $p(1-p)^3$, yielding an expected number of neurons sensitive at just one spout equal to $4Np(1-p)^3$. Similarly, the expected number of neurons sensitive at three spouts equals $4Np^3(1-p)$. Finally, there are six different ways in which a neuron can be sensitive to sound at two spouts ([1 1 0 0], [1 0 1 0], [1 0 0 1], [0 1 0 1], [0 0 1 1], and [0 1 1 0]), each with a probability of $p^2(1-p)^2$, yielding an expected number of neurons sensitive at two spouts equal to $6Np^2(1-p)^2$. Once the expected values were calculated in this manner, their confidence intervals were obtained from the critical values of the inverse binomial distribution (Matlab function “binoinv”).

Separation and comparison of putative interneurons and pyramidal cells. In keeping with established methods, three criteria were used to distinguish neuronal cell types according to physiological criteria: mean firing rate (Ranck, 1973; Fox and Ranck, 1981), spike duration (Skaggs et al., 1996), and the autocorrelation function (Csicsvari et al., 1999). Firing rate was measured over the whole session. Spike duration was measured at 25% of maximum spike amplitude (Csicsvari et al., 1999). The autocorrelation-derived index is the difference between the numbers of spikes that occurred in a 3-5 ms post-spike window and a 20-80 ms post-spike, divided by the sum of the two. Spike duration and the autocorrelation-derived index yielded

bimodal distributions. On the basis of firing rate, spike duration, and the autocorrelation-derived index, the complete set of neurons was clustered into two classes using a K-means algorithm. In keeping with the conventions established by others (Fox and Ranck, 1981; Csicsvari et al., 1999; Wiebe and Staubli, 2001), we refer to the units with the broader action potentials and lower firing rates as putative pyramidal cells and to those with the narrower spikes and higher firing rates as putative interneurons. We stress that a definitive classification of cell types requires information beyond that which can be obtained with extracellular recordings alone. We found that narrow spike-firing, putative interneurons exhibited a broader distribution of spike widths and firing rates than putative pyramidal cells. This may be attributable to the greater diversity of the interneurons compared with pyramidal cells (Csicsvari et al., 1999; Klausberger and Somogyi, 2008).

For the two physiologically distinguished cell types, the statistical significance of the information about sound stimulus identity or response direction was determined using the permutation test described above. To quantify the amount of information carried by each neuron, we took the peak value of the information across time among those neurons that passed the significance test. These values were used to compare sound and response direction coding differences between the putative pyramidal cells and interneurons.

5.2.8. Histology

After conclusion of the recording experiments, the animals were overdosed with intraperitoneal injection of the anaesthetic urethane and transcardially perfused with 10% formalin. Brains were sectioned in the coronal plane and stained with cresyl violet.

5.2.9. Contributions By Different Members Of The Collaboration

This chapter developed out of a collaboration with the group of Prof Mathew Diamond in Trieste (Italy). The tasks involved were distributed between Pavel Itskov, Katya Vinnik and Mathew Diamond (of the Trieste group) and Jan Schnupp and myself at Oxford. All of us participated in the design of the experiment. The stimuli were designed and generated by myself. Training of animals was done by Pavel Itskov and myself. Surgery and electrophysiological recordings were performed by Pavel Itskov and Katya Vinnik. Interpretation and illustration of the data, as well as drafting text for publication, were performed by all participants.

5.3. Results

5.3.1. Neuronal Representation Of Sound Identity

In this study we set out to test how neurons in rat hippocampus represent auditory stimuli in the context of a behavioural task. Rats triggered a sound stimulus (Figure 5.1., A and B) by extending their head through an infrared light barrier placed at the front of the platform (the “nose-poke” position, see Figure 5.1C). When the light barrier was crossed, one of the four artificial vowel triplets (Figure 5.1.D) was played bilaterally. Depending on stimulus identity, animals had to turn either right or left to receive a water reward (Figure 5.1.E). Two sounds were associated with the reward on the right, and two other sounds with the reward on the left. Since the animals performed the same action for both sounds associated with the same reward spout, this design allowed us to look for stimulus-specific neuronal responses. The rats were initially trained on one of the two nearly identical training platforms until they performed at >70% correct for all four sounds. This took between 1.5 and 2 mo. Next, they were trained in the same task on the second platform with the same stimulus-response associations (in rat-centred coordinates). Although the behavioural task

and sound stimuli were identical, the transfer of the task to the second platform was not immediate. About 3–5 days (~400 trials) of additional training were required for the animals to become competent, and for the remaining sessions, they persistently achieved lower performance on the second platform (median performance across all recording sessions: 80.2% on first platform and 70.6% on second platform; Wilcoxon sign rank test, $P < 10^{-5}$). The lack of immediate transfer of reward contingencies suggests the animals may have formed two distinct representations of the task, one for each of the two training platforms. Alternatively, it may stem from the conflict between a room-centred representation of the arena and a self-centred representation.

To quantify the representation of sound identity, we measured the information carried by neuronal firing rate on correct trials (see Methods). To ensure that any observed differences in firing for the two different sounds associated with each response direction reflected differences in sound, and not differences in behaviour (see Methods), we excluded recording sessions in which reaction times differed between the two sounds tested at any of the four response directions (left or right on either platform 1 or 2). Accordingly, 261 of the total of 429 recorded neurons were included in further analysis.

Examples of responses from sound-sensitive neurons are shown in Figure 5.2.A total of 55 of 261 tested neurons (21%) exhibited significant sensitivity to sound identity. The median quantity of information about sound identity carried by neurons with sound sensitivity was 0.12 bits (interquartile range 0.08 – 0.16 bits). The upper limit, in a discrimination between two sounds, would be 1.0 bit. A meta-test (see Methods) confirmed that the observed sensitivity across the population was statistically significant (Wilcoxon rank sum test, P = 0.0013).

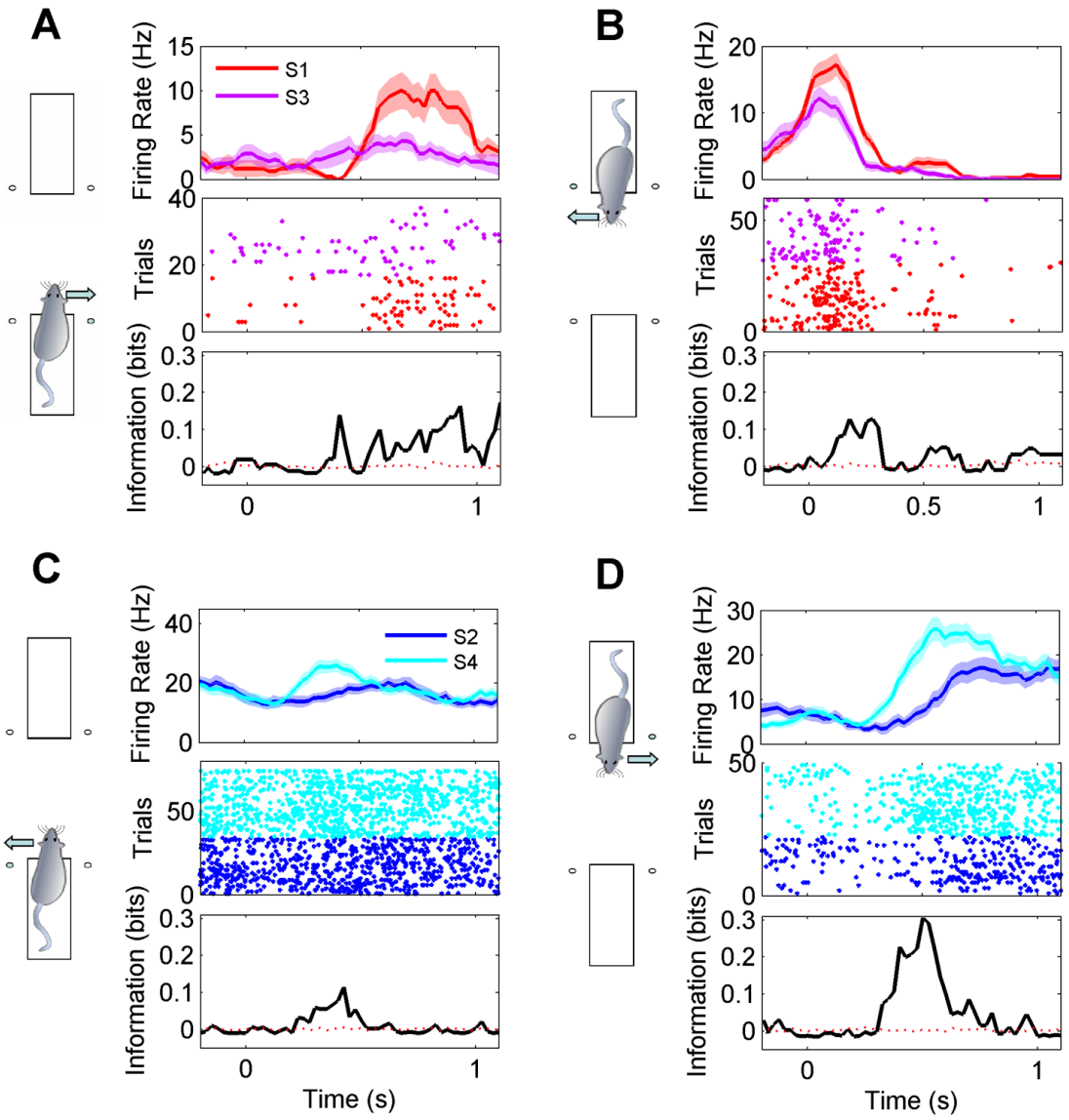


Figure 5.2. Examples Of Sound-sensitive Responses Recorded From 4 Neurons *(A–D). Diagrams at left indicate on which platform and with which reward direction the activity was recorded. Top plots show mean firing rates (+/- SE) after stimulus onset. Responses to the 2 different sound stimuli (S1 or S3; S2 or S4) tested for each response direction are colour coded as indicated. Middle plots show the responses as dot rasters (each dot corresponds to the time of a spike; sequential trials of 1 stimulus are plotted on separate rows). Bottom plots quantify the stimulus dependence of firing rate over time as bits of mutual information in 200-ms sliding time windows. Solid black lines indicate actual information quantity; red dashed lines represent average information across 200 permutations. Bin size = 200 ms, step = 25 ms.*

To evaluate the latencies of sound-specific neuronal responses, we analysed the distribution of the first bin after stimulus onset carrying a significant quantity of information about sound identity ($P < 0.05$). Neurons started to carry sound information at a median latency of 310 ms after the sound onset (Figure 5.3.A). There were insufficient numbers of error trials to permit an analysis of whether the representation of specific sounds was equivalent on correct and incorrect trials.

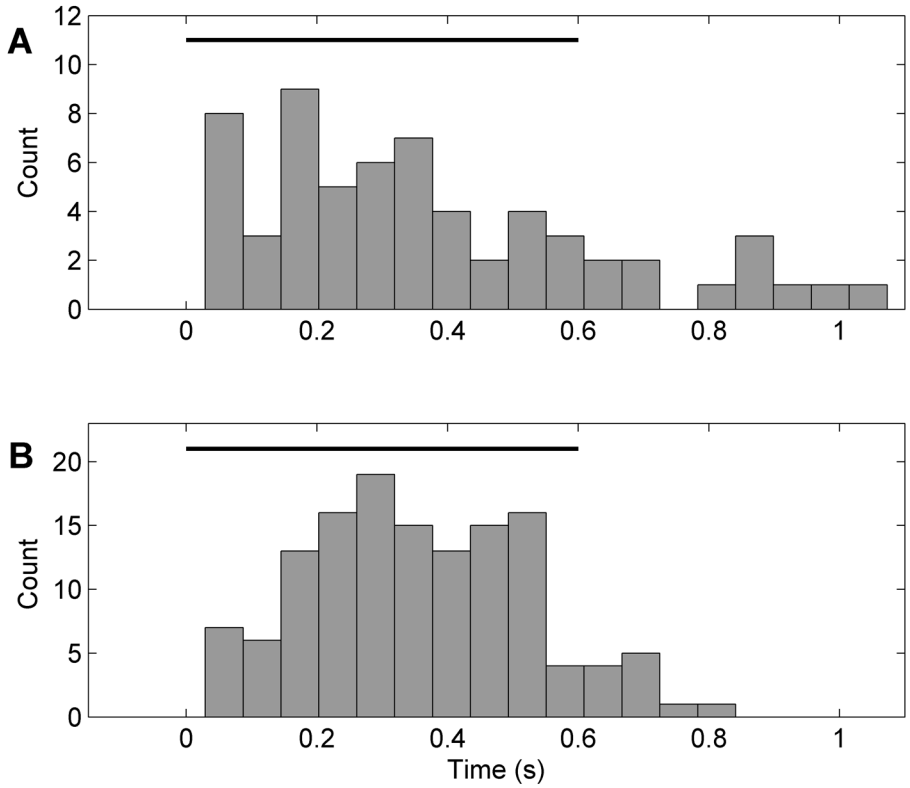


Figure 5.3. Latencies Of CA1 Neural Responses To Sound And Location

*(**A**) Latencies of sound-specific responses. Neurons started carrying sound related information after a median latency of 310 ms. (**B**) Latencies of direction-specific responses. Neurons started carrying location related information after a median latency of 350 ms.*

5.3.2. Passive Sound Exposure

The use of sounds as stimuli allowed us to ask whether their representation in CA1 was related to the stimuli as mere physical signals or as meaningful events dependent on behavioural context. To date, it has not been possible to show unambiguously whether rat hippocampal neurons truly encode non-spatial sensory stimuli or whether they fire to some particular aspect of the behaviour associated with the discrimination. To answer this question, we recorded neuronal activity just after the animals had completed the test session and were resting on the second platform. These data were obtained for 282 of the total of 429 cells in our data set. The animals were exposed to the same sounds as during the immediately preceding behavioural task. The sounds were triggered with no relation to the animal's posture or movement, in a random sequence lasting ~15 min with a random inter-stimulus interval ranging from 0.8 to 1.5 s. The rats did not respond to these sounds with head turns toward the water spouts or any other overt behaviour. We analysed the neuronal discharges in relation to sounds by the same methods as when the sounds were presented during the discrimination behaviour. Seventeen of 282 neurons (6%) recorded during the passive conditions were classified as sound sensitive at a $P < 0.05$ significance level, a number close to the expected "false alarm" rate of 5%. We compared the frequency of such responses across the population using the stimulus label shuffling meta-test (see Methods) and found no evidence for statistically significant sound coding across the population as a whole ($P = 0.39$). Moreover, among the neurons that did exceed the criterion, the median quantity of sound information was significantly smaller than the quantity found during the active task (0.05 bits in the passive condition vs. 0.12 bits for the active task; Wilcoxon rank sum test, $P < 10^{-7}$).

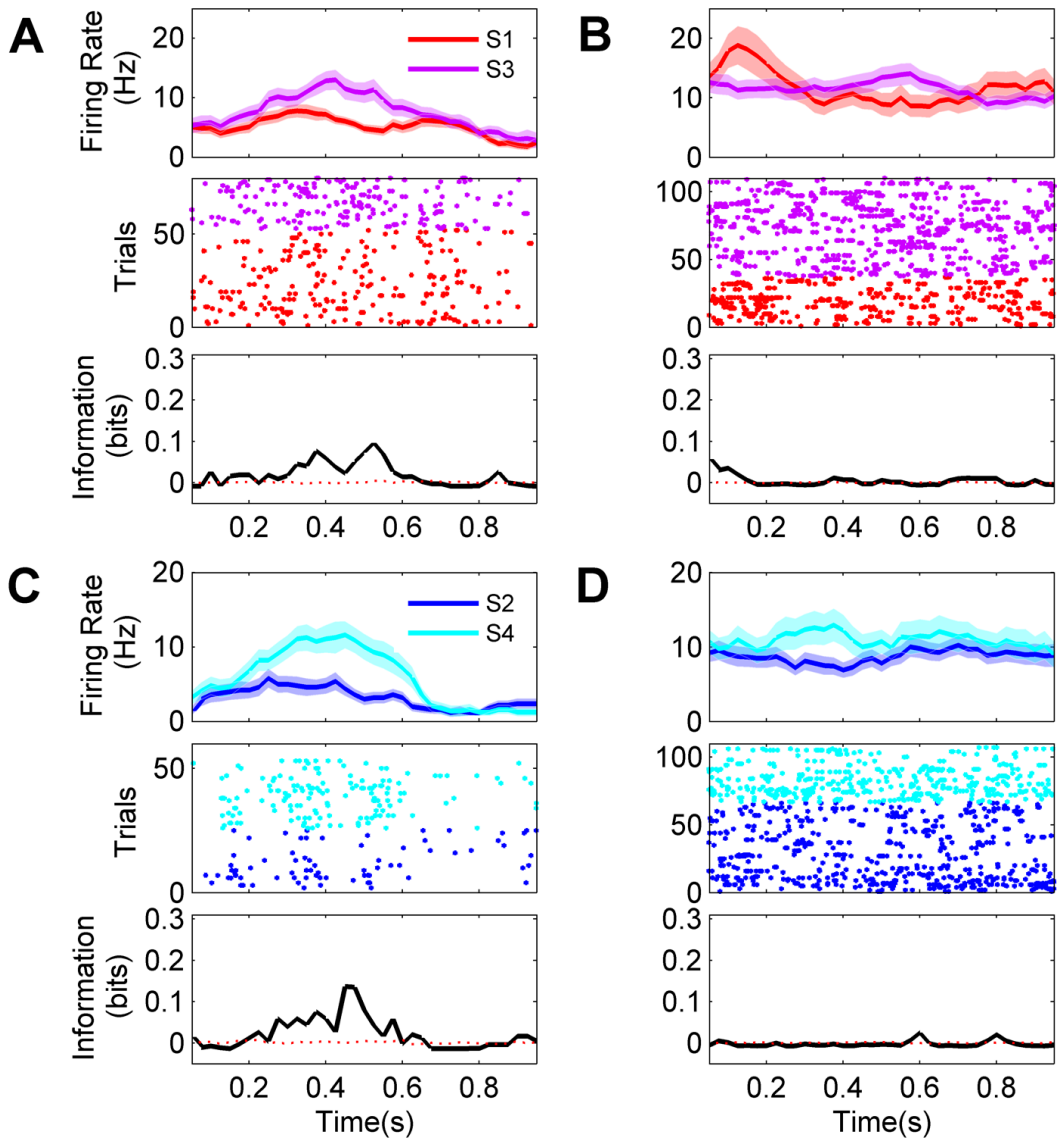


Figure 5.4. Examples Of Responses For 2 Neurons Sensitive To Sound Identity During The Behavioural Task But Not During Passive Listening.

*Top plots show mean firing rates (+/-SE) after stimulus onset. Middle plots show the responses as dot rasters. Bottom plots quantify the stimulus dependence of the firing rate over time as bits of mutual information in 200-ms sliding time windows. Black lines indicate actual information quantity; red dashed lines represent average information across 200 permutations. Bin size = 200 ms, step = 25 ms. The first neuron fired more vigorously to S3 than to S1 during the discrimination task (**A**) but did not show significant sound coding when the same stimuli were presented passively (**B**). The second neuron was excited by the onsets of each of the sounds during the discrimination task (the triplets of bursts in **C** are aligned to the triplets of sounds) but showed no stimulusrelated modulation during passive listening (**D**).*

By visual inspection, none of the 282 neurons recorded in the passive condition showed convincing, reproducible sound selectivity. Two examples of neurons that distinguished between two sounds within the behavioural task but not in the passive condition are shown in Figure 5.4.

Together, these results indicate that under our experimental conditions, neurons in the CA1 region of hippocampus were not sensitive to differences in acoustic stimuli presented in isolation, even if the neurons showed clear discrimination between the same acoustic stimuli in the context of sound-guided behaviour.

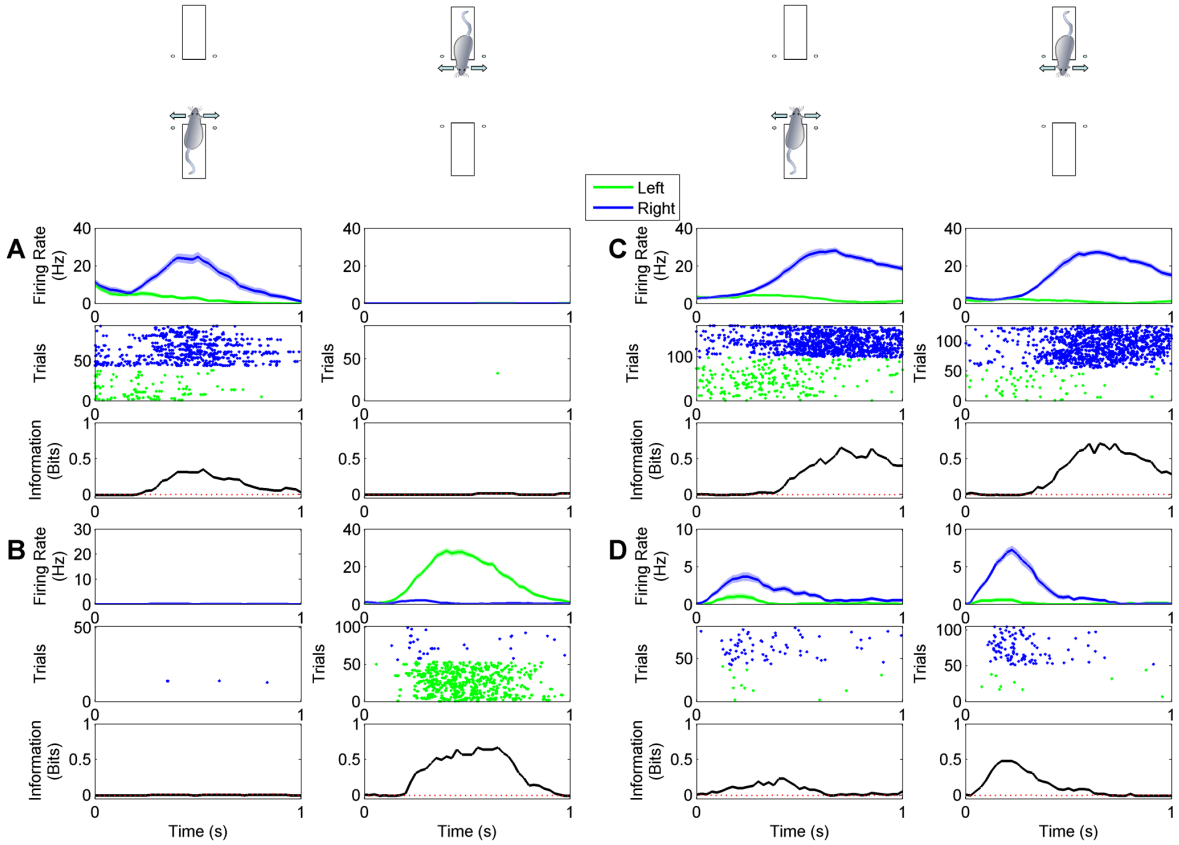


Figure 5.5. Examples Of Responses From 4 Neurons Sensitive To Response Direction. *(**A-D**) Responses recorded during right turn trials are shown in blue, and responses recorded during left turn trials are shown in green. (**Top plots**) Mean firing rates (+/-SE) after stimulus onset. (**Middle plots**) Responses as dot rasters. (**Bottom plots**) Quantification of the dependence of the firing rate over time on stimulus identity as bits of mutual information in a 200-ms sliding time window. Black lines indicate actual information quantity; red dashed lines represent average information across 200 permutations. Bin size = 200 ms, step = 25 ms. For each neuron, responses recorded on both platforms are shown, as indicated by the diagrams above the plots. **A** and **B** show examples of neurons that responded when the rat was on only 1 of the 2 platforms. **C** and **D** show examples of neurons that responded to left or right turns in similar ways regardless of when the rat was on platforms 1 and 2.*

5.3.3. Neuronal Representation Of Response Direction.

A wealth of literature documents that many neurons in rat hippocampus are highly sensitive to the animals' location in space, and one might therefore expect a number of neurons in our sample to be sensitive to whether the animal turned left or right to collect water reward. We tested for such sensitivity by applying the same mutual information measures used to test for stimulus coding (see Methods). Examples of neurons that showed response direction-dependent firing on one of the two platforms only are shown in Figure 5.5.A and B, whereas Figure 5.5, C and D, shows examples of neurons that fired more during turns toward the reward spout on the left regardless of platform.

A total of 113 of 261 tested neurons (43%) exhibited significant sensitivity to response direction. Neurons with response direction sensitivity carried a median value of 0.22 bits of information about response direction (interquartile range 0.15–0.38 bits). Neurons that encoded response direction on one platform only (67 putative pyramidal cells and 24 putative interneurons) were more common than those that encoded response direction on both platforms (12 putative pyramidal cells and 10 putative interneurons). Neurons started to carry response direction information at a median latency of 350 ms after the sound onset (Figure 5.3.B).

Whenever there were enoughincorrect trials to characterise neuronal firing on those trials, the firing of the response direction-sensitive neurons was found to be associated with the animals' actions on those trials. The simplest interpretation is that these neurons discriminated the actual direction of the movement rather than the stimulus class.

5.3.4. Relationship Of Sensitivity Across Different Platforms

The fact that animals were tested on two nearly identical but differently located platforms enabled us to investigate whether sound related information present in the neuronal discharges depended on the animal's spatial location, and how coding for sound identity and place interacted. Our data indicate that in the large majority of the cases, sound related information was observed only when the rat performed the task on one of the two platforms, suggesting that sound coding depended on spatial context.

The relationship between sound coding and spatial location was assessed as follows. When tested on two platforms, each with two response spouts, a neuron could in principle show sound sensitivity in its firing associated with each of the four possible response directions (i.e., during both left and right turns on each of the 2 platforms), or it might be sensitive to sound in association with only a subset of the four response directions or at none at all. We found that sensitivity to sound was linked to the position of the animal within the room, i.e., the neuronal responses to sounds on one platform could not be predicted based on the sensitivity of the same neuron on the other platform. Of the 261 neurons in our sample, 48 gave significant sound identity information for only one of the four response directions, 7 gave significant information for two response directions, and none gave significant information for more than two response directions. The filled bars in Figure 5.6.A summarise these results. The open bars show the distribution that is predicted if one assumes that that the presence in a given neuron of a sound identity code for any one of the four tested reward spout locations does not predict whether the same neuron will discriminate between sounds for other response directions. The observed distribution matches the distribution predicted by the model of independence.

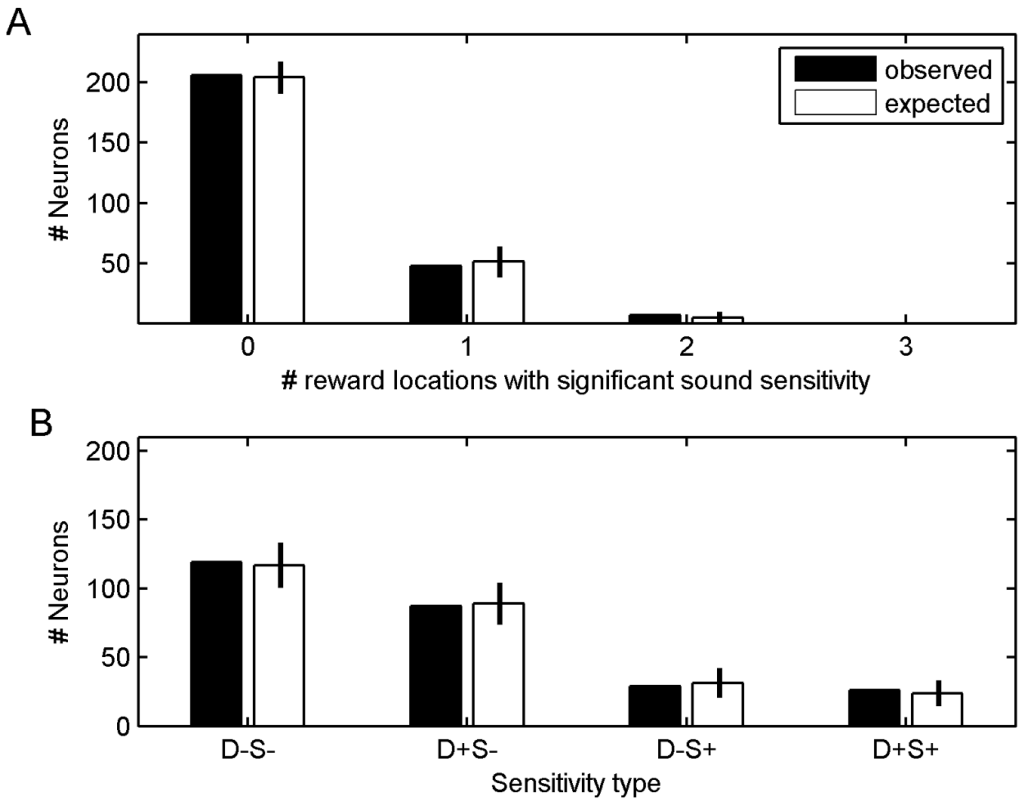


Figure 5.6. Observed And Expected Numbers Of Location And Sound Sensitive Neurons Under The Assumption Independent Coding

*(**A**) filled bars indicate numbers of neurons observed to be sensitive to the identity of the sound stimuli associated with **0, 1, 2, 3,** or **4** reward locations (response directions), respectively. Open bars represent the distribution expected if the probability of a neuron being sensitive to sound for each reward location were independent of the probability of it being sensitive to sound sensitivity for all other reward locations. (**B**) filled bars indicate numbers of neurons sensitive to neither response direction nor sound identity (**D-S-**), sensitive to response direction but not sound identity (**D+S-**), not sensitive to response direction but sensitive to sound identity (**D-S+**), or sensitive to both response direction and sound identity (**D+S+**). Open bars represent distribution of sensitivity to response direction and sound identity expected under the assumption of independent coding of response direction and sound identity. Error bars show 95% confidence intervals for the expected distributions.*

5.3.5. Relationship Of Sensitivity To Sound And Response Direction.

In principle, a neuron could be sensitive to only response direction (left/right), only sound identity, or both. The sensitivity of the neuron to one of these factors may depend on the other factor, or, alternatively, sound sensitivity and response direction sensitivity could be independent properties. We derived a distribution of cell counts expected to occur in a population, given the assumption of this independence (see Methods). In the experiment sown in Figure 5.6.B, we investigated the relationship between sensitivity to sound stimulus identity and sensitivity to response direction across the sample of 261 neurons. A neuron was considered sound sensitive (S+) if in at least one of the four conditions (platform 1: sound 1 vs. sound 3; platform 1: sound 2 vs. sound 4; platform 2: sound 1 vs. sound 3; platform 2: sound 2 vs. sound 4) it carried a significant quantity of information about sound identity. Otherwise, it was considered sound insensitive (S-). Likewise, a neuron was considered sensitive to response direction (D+) if it distinguished left and right response direction on either platform. Otherwise, it was considered to be insensitive to response direction (D-). One hundred nineteen neurons were sensitive to neither response direction nor sound identity (D-S-), 87 were sensitive to response direction but not to sound identity (D+S-), 29 were sensitive to sound identity but not to response direction (D-S +), and 26 were sensitive to both response direction and sound identity (D+S+). The filled bars in Figure 5.6.B show these observed neuron counts, whereas the open bars show the counts of neurons that would be expected if we assume that response direction sensitivity and sound identity sensitivity are statistically independent of each other. Again, there is a close correspondence between the observed distribution and that expected by independence. The results indicate that whether a particular neuron was recruited for the representation of auditory stimuli did not predict whether it would also be recruited for the representation of response direction. Responses for a

neuron characterised as D-S+ are illustrated in Figure 5.7.A–C, whereas responses for a neuron characterised as D+S- are illustrated in Figure 5.7.D–F.

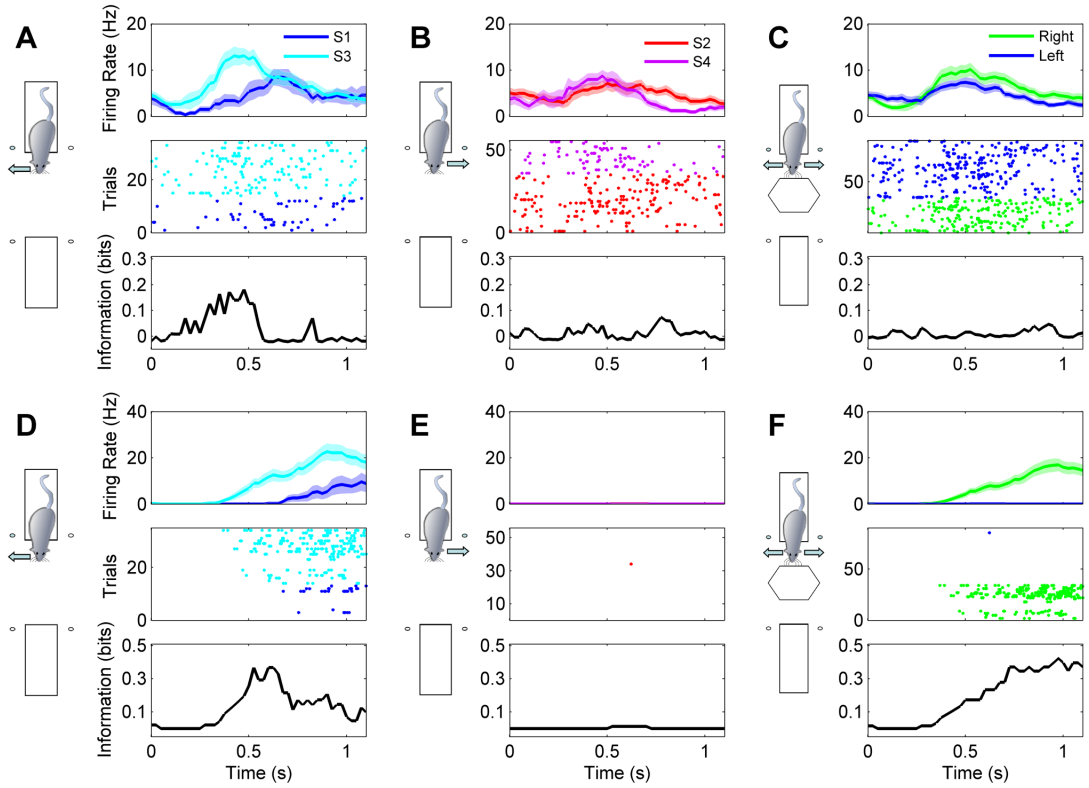


Figure 5.7. Responses Of 1 Neuron That Was Sensitive To Sound Identity But Not Response Direction.

*(**Top plots**) Mean firing rates (+/-SE) after stimulus onset. (**Middle plots**) Responses as dot rasters. (**Bottom plots**) Quantification of dependence of the firing rate over time on stimulus identity as bits of mutual information in a 200-ms sliding time window. Black lines indicate actual information quantity; red dashed lines represent average information across 200 permutations. Bin size = 200 ms, step = 25 ms. This neuron fired more vigorously to S3 than to S1 during right turn trials (**A**). but firing in response to left or right turns (combining the responses to different sounds; **C**) did not differ significantly. (**D–F**) Responses of a neuron sensitive to both sound identity and response direction. Plots are as described in **A–C**. This neuron fired only in trials where the animal turned to the right, and fired substantially more spikes in response to S3 than to S1.*

5.3.6. Firing Properties Of Putative Interneurons And Pyramidal Cells.

Using the physiological criteria described (see Methods) and illustrated in Figure 5.8.A, we classified 179 neurons as putative pyramidal cells and 82 as putative interneurons (Csicsvari et al. 1999; Fox and Ranck 1981; Itskov et al. 2011; Ranck 1973). We then asked whether the putative cell type was related to functional properties of neurons in the behavioural task. Putative interneurons were more likely to be sensitive to stimulus identity, and carried more sound related information, than putative pyramidal cells. Among putative interneurons, 33% were sound sensitive, whereas only 16% of putative pyramidal cells were sound sensitive (27 of the 82 putative interneurons and 28 of the 179 putative pyramidal cells). To quantify information in these cells, we took a peak value of sound identity information. The median value of the peak among all sound-sensitive putative interneurons was 0.15 bits (interquartile range 0.11– 0.19 bits), significantly higher than the value of 0.10 bits (interquartile range 0.07– 0.13 bits) in putative pyramidal cells (Wilcoxon rank sum test, $P = 0.0008$).

Seventy-nine of 179 (44%) putative pyramidal cells and 34 of 82 (41%) putative interneurons carried significant amounts of information about response direction. The median of the peak response direction information for putative pyramidal neurons (0.35 bits, interquartile range 0.17– 0.40 bits) was significantly higher than that of putative interneurons (0.19 bits, interquartile range 0.13 – 0.23 bits; Wilcoxon rank sum test, $P = 0.009$). Figure 5.8.B summarises values of information about stimulus identity and response direction for putative pyramidal and putative interneurons.

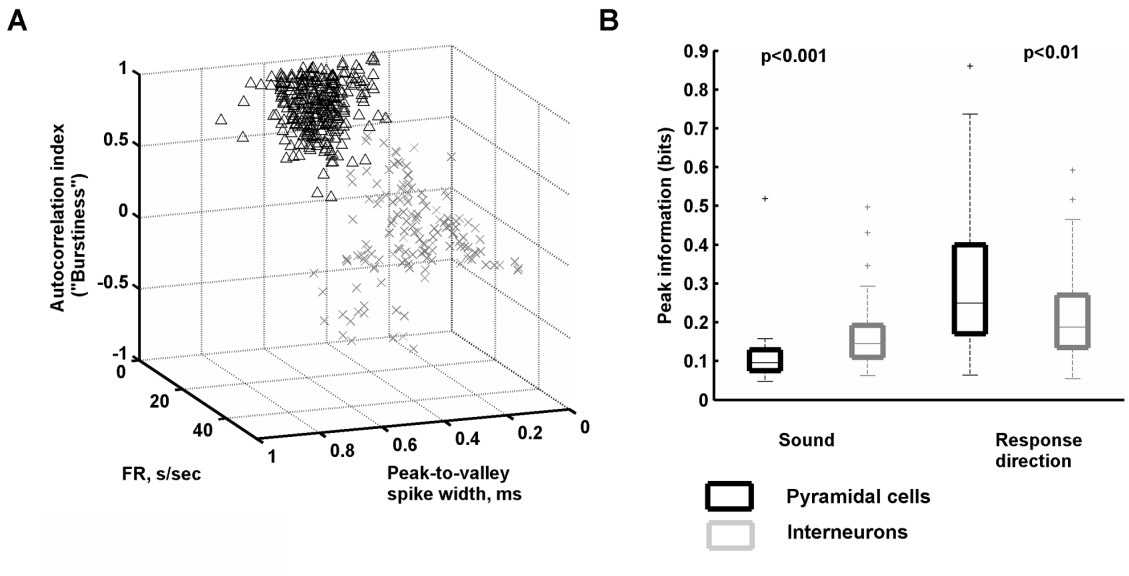


Figure 5.8. Functional Properties Of Putative Interneurons And Pyramidal Cells.
*(**A**) Neurons were classified as putative pyramidal neurons or interneurons according to physiological criteria including spike width, mean firing rate, and tendency to fire in bursts. Burst firing was quantified from each neuron's autocorrelation. FR, firing rate (s/sec, spikes per second). (**B**) putative pyramidal cells and interneurons exhibited systematic differences in the amounts of stimulus identity or response direction information. For each box, the central mark is the median, the edges of the box are the interquartile range, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually.*

5.4. Discussion

This work characterises the encoding of sound stimuli in rat hippocampus. Anatomic studies indicate that the hippocampus receives auditory input only very indirectly from perirhinal cortex through lateral entorhinal cortex (Burwell et al., 1995; Burwell and Amaral, 1998) and medial prefrontal areas via the nucleus reuniens of the thalamus (Vertes, 2006). The auditory information arriving in the hippocampus, like that of other modalities, is therefore highly processed, and the hippocampus is not a sensory structure as such. Nevertheless, a number of previous studies have demonstrated stimulus related firing in hippocampus for olfactory (Wiener et al., 1989; Eichenbaum et al., 1999; Komorowski et al., 2010), visual (Lever et al., 2002; Anderson and

Jeffery, 2003; Hampson et al., 2004; Leutgeb, 2005; Quian Quiroga et al., 2005), tactile (Itskov et al., 2011), auditory (Berger et al., 1976; 1983; Christian and Deadwyler, 1986; Sakurai, 1994; 1996; 2002; Moita et al., 2003; Quian Quiroga et al., 2009), and gustatory stimuli (Ho et al., 2011). Several groups have used operant conditioning to sound stimuli to study changes in hippocampal activity during learning (Segal et al., 1972; Berger et al., 1976; Christian and Deadwyler, 1986; Moita et al., 2003), and it appears from these studies that hippocampal neurons acquire responsiveness to the sound stimuli used for conditioning as the animal learns the conditioned behavioural response. After the acquisition of the conditioned response, hippocampal neurons started to fire to the sound onset, with a latency of around 80 ms, and unrewarded sounds caused smaller responses (Christian and Deadwyler, 1986; Moita et al., 2003). In our experiment, the onset of sound-specific firing can be seen in a wide distribution of latency values ranging from about 0.05 to 1.1 s (median value, 310 ms).

One potential difficulty in interpreting previous work is that when stimuli and conditioned responses are coupled, it is difficult to know whether neuronal activity recorded during an operant conditioning task encodes the stimulus or the animal's resultant behaviour. This can be particularly problematic if the response requires the animal to change its location or orientation in space, given that many neurons in rat hippocampus are well known to exhibit place related firing (O'Keefe, 1976; O'Keefe and Nadel, 1979; Wills, 2005). In the present study, we designed the experiment in a manner that can dissociate stimulus related firing from firing related to behavioural responses or spatial location, by training animals to associate more than one sound stimulus with each behavioural response. Moreover, we required rats to carry out identical stimulus-reward pairing experiments at different locations in the room. We

found that some CA1 neurons clearly distinguished between different acoustic stimuli, and because the actions associated with the paired sounds were identical, the sound sensitivity cannot be easily explained by any explicit behavioural correlate. Acoustic sensitivity was in most cases contingent on behavioural context and the animal's location in space. Neurons that discriminated between a sound pair when the animal was on one platform frequently failed to discriminate between the same sound pair when the animal was on the other platform. As a further demonstration of the role of context, we found that during passive listening, neurons did not discriminate between sounds. The modulation of neural responses to the presented sounds that is commonly found in hippocampus thus appears to be much greater in extent and qualitatively different from that normally seen either in AC as a function of changes in behavioural context (Fritz et al., 2007) or in the midbrain as a function of spatial variables (Campbell et al., 2006).

A degree of dependence on spatial context of sensory responses in hippocampus has been observed previously. For example, studies by Wood et al. (1999) and Komorowski et al. (2009) reported that hippocampal responses to odorant stimuli often differed at different locations, although some neurons appeared to "prefer" certain particular odours regardless of location. Our experiment was not designed to map sound preferences across space, but the results nevertheless confirm and extend the previous findings that coding for sensory stimuli and coding for space appear to interact in hippocampus. In particular, a neuron's tendency to differentiate two sound stimuli at one location in space did not predict whether the neuron would differentiate the same sounds at a different location. Interactions between coding for sensory stimuli and coding for spatial location do not of course preclude the

possibility that spatial or sensory coding may be achieved by somewhat separate, specialised subpopulations of hippocampal neurons.

A number of studies by Sakurai (Sakurai, 1994; 1996; 2002) have explored hippocampal population responses during various sound related tasks. Sakurai's experimental and analytical approach differs substantially from ours, but our work nevertheless confirms and extends several of Sakurai's findings. In one study (Sakurai 1996), animals were trained in three different tasks: auditory discrimination, visual discrimination, and a configural auditory plus visual discrimination. The populations of neurons that were engaged in each of these tasks were found to be partially overlapping. The author put forward the idea that these overlapping assemblies arose from independent coding. Independence here is to be understood in the statistical sense, meaning that the probability that any one neuron carries significant amounts of visual information is not influenced by whether the same neuron encodes auditory or spatial information. However, Sakurai did not demonstrate such statistical independence quantitatively. In comparison, our current experiment, as well as a previous experiment using tactile stimuli (Itskov et al., 2011), revealed distinct and statistically independent representations of sounds associated with different locations in the spatial arena (Figure 5.6.A) and also showed that neuronal populations involved in coding different aspects of the task, sound identity and the action associated with it, overlap in proportions consistent with statistical independence (Figure 5.6.B).

However, we observed some statistical dependence of coding for space or sound on cell type. At present, there is no consensus in the literature on the relative contributions made by putative hippocampal interneurons and putative pyramidal

cells to the representation of auditory stimuli or the animal's location in space. Whereas one study showed a clear functional dissociation between the two neuronal cell types (Christian and Deadwyler, 1986), another observed no clear differences in sound related firing (Moita et al., 2003). In our experiment, both types of neurons participated in the representation of sounds and places, but putative interneurons carried more stimulus related information, whereas putative pyramidal cells carried more response direction related information. These differences were statistically highly significant, but their neuroanatomical and functional significance remains difficult to evaluate at present, given that distinguishing anatomically defined neural populations on the basis of physiological criteria such as spike width and firing rate is not without difficulty (Vigneswaran et al., 2011) and given that these observed differences between the two physiologically distinguished cell types, although statistically highly significant, were nevertheless not very large, and it is important to remember that.

Although our results therefore agree with previous studies of rodent hippocampus, parallels with work on human hippocampus are harder to draw. Neurons in human hippocampus fire according to the stimulus class (Quian Quiroga et al., 2005; 2009),which suggests a highly abstracted representation of the semantics of stimuli (Gelbard-Sagiv et al., 2008) that is largely independent of context. This contrasts sharply with the strong spatial and behavioural context dependence of sensory coding that we observed in our experiments. The sounds in our experiments were identical on both platforms and in the active and passive conditions, so if the neurons had formed abstract and invariant representations of the sounds per se, they should have exhibited similar responses on both platforms and in the passive as well as active condition. An important difference in experimental design is that in the human

studies so far, the subjects were always tested in just a single location. Future studies in humans may reveal whether this representation of object classes is also influenced by the subject's position and behavioural context.

The spatial context dependence of sensory responses that we observed is certainly compatible with the hypothesis that hippocampus may create "object in place" representations by linking salient events to the spatial context in which they occur (Cohen and Eichenbaum, 1995; Eichenbaum et al., 1999; Wood et al., 1999; Eacott and Norman, 2004). This may also relate to the phenomenon referred to as "remapping" (Muller and Kubie, 1987; Bostock et al., 1991; Anderson and Jeffery, 2003; Colgin et al., 2008), whereby place fields of hippocampal neurons may change radically if the context in which an animal finds itself is altered or when the behavioural context is altered (Moita et al., 2004; Kennedy and Shapiro, 2009). The fact that different sensory events appear to be encoded by partially overlapping, statistically independent populations of neurons agrees with the predictions of theoreticians such as Marr (Marr, 1971), who considered this a desirable property in a memory system designed to collect "snapshots" of experience.

5.5. Summary & Conclusions

Hippocampus is a known candidate brain structure for the processing of object classes, and responses the CA1 section of hippocampus had been shown to represent particular objects independent of the modality (seeing or hearing) through which information about the object is received (Quian Quiroga et al., 2009). However, the activity of neurons in the CA1 region of hippocampus also represents the location of animals in physical space. In trying to determine whether neurons in CA1 can discriminate between different sounds it is therefore necessary to disentangle the

responses to spatial location from those to the auditory stimuli, and especially so in awake behaving animals.

In this study we achieved this goal by associating two sound stimuli with one reward location in a two alternative forced choice paradigm. In awake behaving animals who had learned to associate the two rewarded sounds per reward location, many hippocampal neurons recorded during the task responded in a stimulus-specific manner. The distribution of neurons which respond to a particular sound and location or both matched the distribution expected if the population of hippocampal neurons encoded sound and location in a statistically independent manner, indicating that a many hippocampal responses are sensitive to formant changes. This shows, that simple speech-like sounds evoke neural responses in CA1 of rat hippocampus, and carry information about the identity of the presented sound, inviting further investigation into the involvement of hippocampus in the processing of speech sounds in the mammalian hippocampus and confirming the validity of using vowel sounds to explore the neural representations of auditory classes.

Yet, the fact that CA1 neurons in passively listening animals did not exhibit sensitivity to sound also indicates that hippocampal representations of sounds are highly dependent on behavioural context. Furthermore, latencies of hippocampal responses far exceeded the known limits of fast class judgements in humans (VanRullen and Thorpe, 2001; Kirchner et al., 2006). It therefore likely that hippocampal processing *registers* auditory classes for episodic memory, that have already been detected somewhere else. The results from Chapter 4 of this thesis do not support the idea that A1 is the locus of this class-detection, but may provide the substrate for the detection of combinations of acoustic components that underly the perception of

auditory classes. I therefore speculate that higher auditory cortical and association areas that A1 feeds into and that in turn feed into hippocampus perform the required processing.

CHAPTER 6 – AN AUTOMATED TRAINING ENVIRONMENT FOR RATS

6.1. Motivation

An important goal in auditory cortical research is to explain the interplay between the perception of sounds and the cortical processes that generate sound perception. One key characteristic of auditory object recognition is the fact that object classes are often not innate but must be learned through experience. When learning to reliably classify a sound as a particular auditory object humans and animals learn to ignore variations of the sound along irrelevant auditory dimensions (like loudness and often direction) while detecting variations along relevant dimensions (depending on the object these include attack, decay, pitch and timbre among others). This shift in perception of acoustic variation must become observable in the behaviour of experimental animals to establish that the animal did not just learn to identify several similar sounds but formed a class based on similarity (invariance) of certain sound features (dimensions). Ideally the researcher would monitor the neural responses in the brain region he or she considers relevant throughout the learning process to try and establish a first correlation between changes in behavioural responses and changes in neural responses. A causal relation can then be sought with more advanced methods of, for instance with cortical stimulation or lesioning.

Another consideration comes from the observation that the manner in which, for example, A1 neurons integrate spectral components also depends on background sounds (Bar Yosef and Nelken, 2007) and on the type of task performed (Fritz et al., 2003). Therefore it is desirable to monitor changes in neural responses in the same

acoustic and behavioural environment that the animals get used to during the learning of an auditory task, both in active and passive listening conditions.

Keeping the environmental conditions as stable as possible in the transition between training and the animals' daily lives is not only desirable regarding experimental design but also from the point of view of animal welfare. For instance, reductions in the amount of stress that animals are exposed to can lead to quicker and more consistent performance in tasks that require normal hippocampal innervation and activity (Luine et al., 1994). And it has been consistently shown that environmental stress factors can impair long-term memory in spatial learning (Diamond et al., 1996; Park et al., 2001) and eye-blink conditioning tasks (Maeng et al., 2010).

I built an automated training environment for behavioural conditioning that overcomes these potential obstacles to successful auditory learning and to animal welfare during training and testing. Here I describe the design, components and functionality of the training environment and present preliminary data of neural recordings during training and passive listening. All behavioural data reported in Chapter 3 of this thesis were collected in this training environment.

6.2. Design & Components Of The Training Cage

Figure 6.1. shows a photo of the complete training setup. The training cage consists of a standard rat cage, modified to contain a nose poke hole (containing reinforcement components) and covered with removable perspex lid. The whole system has one central function: to run go-nogo operant conditioning tasks with water as a positive enforcer and air puffs as a negative enforcer.

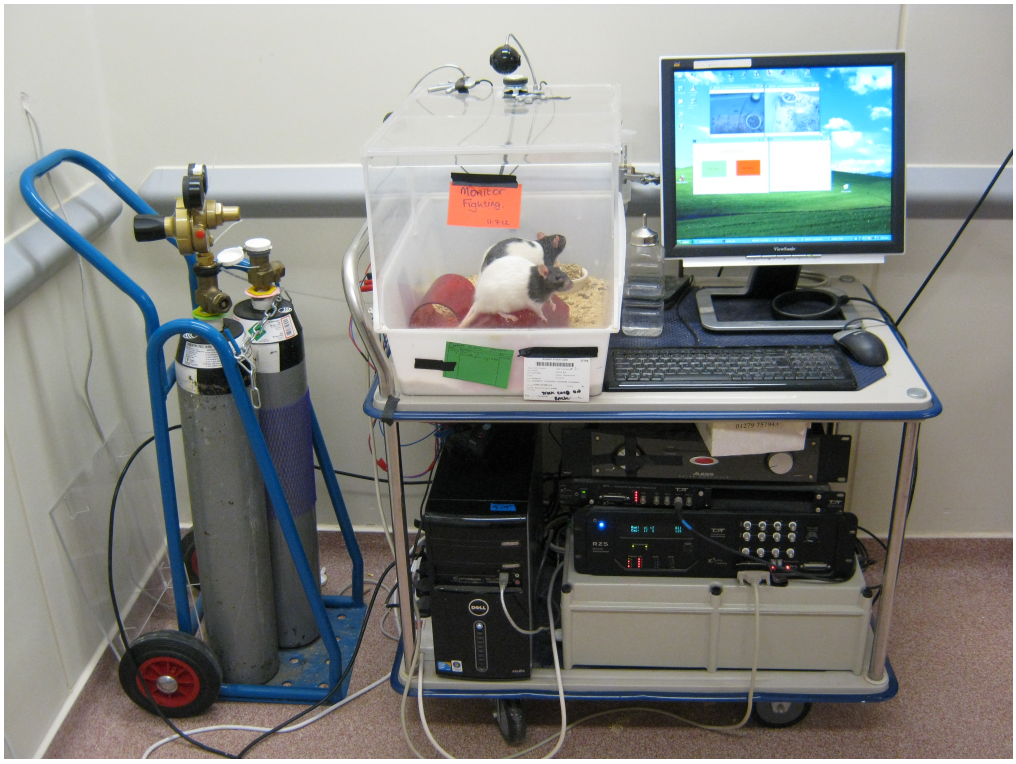


Figure 6.1. The Complete System For Automated Auditory Training In Rats
Left of the trolley are compressed air canisters that are now replaced by a medical air compressor. The controlling PC sits on the bottom platform of the trolley on the left and is connected to the RFID reader (bottom right) and the electrophysiological recording hardware (top right). Training can be monitored directly or remotely.

The nose poke hole is constructed such that it performs four basic functions: to detect the presence of an animal in its vicinity, to detect whether that animal attempts to drink water, to give water (positive reinforcer), and to administer air puffs (negative reinforcer). As I decided to train animals exclusively in go-nogo tasks, only one nose poke was required. The water and air-puff delivery are driven by solenoid valves attached to the back wall of the cage (Figure 6.2.). Air pressure is, since recently, generated by a miniature medical air compressor (Bambi Ht3), rendering the system more portable. Drinking water is kept in a small container, positioned above the spout.

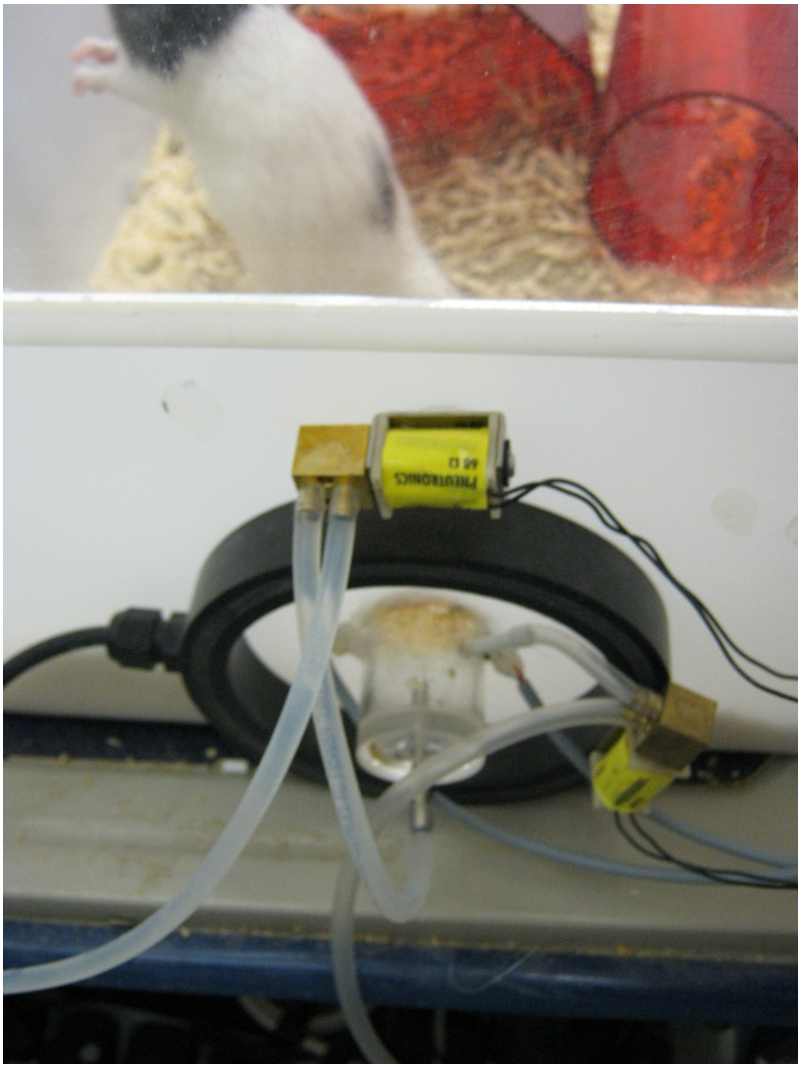


Figure 6.2. Back Side View Of The Reward Spout

We can see the ring antenna for RFID detection (black ring) the water and air puff solenoids (yellow), the infrared diodes for nose poke detections (ends of gray cables) the nose poke hole (perspex tube) with connected water and air puff tubes.

The cage operates by detecting animals through RFID (Radio Frequency IDentification). I implanted all my animals with miniature RFID tags (Trovan ID-100A Nano Transponder) for detection by the training system (see Figure 6.3.). An RFID antenna is attached to the cage and an RFID reader is connected to the controlling PC. Infrared barriers are used as triggers for training sessions, when an animal crosses the barrier (described in more detail below). All stimulation and reinforcement systems as well as infrared detection are directly operated by TDT recording and

stimulation hardware (Tucker Davis Technology, Austin, Texas, USA). Crossings of the IR barrier recorded by the TDT recording hardware (RZ5) and detection of RFID tags by the RFID readers are transmitted to the central controlling PC. Sounds are uploaded to an RX6 (TDT) and played when triggered from the controlling PC. RZ5 and RX6 are synchronised so sound onsets and offsets can be matched exactly with time points in ongoing neural recordings.



Figure 6.3. Recording, Stimulation And RFID Components.
The type of RFID tag implanted under the animals loose neck skin.

Along with running training automatically this training system allows the recording of neural data in animals with a chronically implanted neural probe. I chose to use Neuronexus chronic probes (A-style 32 channels with ZIF connector) to test the recordings during training. Probes are connected to the recording device using TDT ZIF-clip head-stages (ZIF32), powered by two 16 channels TDT preamplifiers (Medusa 16). As the recording device collects both training and cortical data, reaction times and reinforcement times can easily be aligned with neural data during analysis. To allow recording of neural responses in a more tightly controlled acoustic environment I built a soundproof chamber that can be slid into the main cage and creates a space directly underneath the loud speaker in front of the nose poke. When in use, a single animal can be placed in that space with or without a connected implant and can either be allowed do its task or kept there listen passively to sounds.

The controlling PC runs a purpose-written Matlab (Mathworks Inc.) software that responds to RFID and IR detection or absence thereof by starting, running and stopping training tasks. This involves playing sounds, controlling water and pressure air valves for the reinforcers. The PC collects and saves both training and neural data. Apart from these general abilities I wrote a toolbox of Matlab function that allows easy and successfully control of behavioural training. The whole training system fits on a small, two-levelled, trolley and can therefore easily be moved to a different room if desired.

6.3. Functionality Of The Training Cage

6.3.1. Registration Of Animals

As each RFID transponder sends out a unique identity code, when held into an appropriately changing magnetic field (produced by the RFID antennas), animals are easily identified during preparation and during training itself. Usually a training regime begins by choosing an animal that already carries an RFID chip under its neck skin and to name it (using for example names of tyrants). This is easily done by running one of the specifically written Matlab functions that detects any ID near an antenna reserved for registering an animal outside the training cage itself (it has an extra long cable and can be held into a neighbouring cage for instance).

6.3.2. Assignment Of Tasks

After registering an animal, a task is assigned to it. The beauty of using RFID tags is that they allow individual animals to be trained on different tasks during the same period and in the same training cage. The researcher can define an arbitrary number of tasks that will be saved in the system, and which will be available to be assigned to animals as required, and the researcher can easily switch an animal from one task to

another, also remotely by way of an internet link. A task usually consists of a set of stimuli and other parameters specific to the task, like number of water drops and air puffs used for reinforcement, response intervals, trigger delays etc, as well as algorithmic rules which determine the task paradigm. These items of information are conjoined into a single Matlab variable so that they can be easily assigned to an animal, but can also be recombined so that one stimulus set can be associated with as many different tasks as desired, while the training system can run these different tasks alongside one each other.

6.3.3. Training

When an animal approaches the nose poke, the RFID antenna mounted to the outer cage wall reads the animals ID tag. If the animal is registered and has a task assigned to it, the controlling software loads the stimulus set and reinforcement parameters of the identified task, and changes its state to wait for the animal to trigger a trial. If the animal then crosses the IR barrier, the systems state is changed again to active training, and the go-nogo task will run until the animal decides to leave the proximity of the RFID antenna.

A typical task could follow the sequence

Trigger 1➔Sound➔Go-Nogo Response➔Water/Air puff➔Withdraw ➔Trigger 2➔...

Because up to five animals live in the cage, I had to ensure that training is only ever allowed when just one animal is close to the nose poke. The RFID detection is therefore set up such that training is only started if only a single ID code is read for at least five seconds without a single detection of another code, and training only continues as long as this stays the case.

6.3.4 Monitoring Of Training And Behaviour

Behaviour and training are constantly monitored with two USB web cameras, one with an overhead view of the cage and the other with a close view from the side of the spout. Video snippets of animal behaviour can be recorded for demonstration purposes. The controlling PC can be connected to a network through an encrypted secure socket layer connection. With this capability one very helpful application is the ability of the system to send automated emails from the controlling software, for instance if the system runs into an error, or when a particular number of trials is reached. Generally this capability was used to download training data with a remote desktop application with the help of which the current state of the training cage of the behaviour of animals could be observed while sitting in the office in a different building, e.g. analysing the incoming data, as well as to help monitor animal welfare and the normal functioning of the system.

6.5. Chronic Implants - Proof Of Concept

With the current training system all chronic probes can be used for which a headstage exists that connects to TDT recording equipment. I used Neuronexus ZIF-clip probes for test implants and recordings. These types of probe are available in various configurations; with one or several shanks at various gaps between the shanks, with site diameters and distances depending on the purpose.

In this particular proof of purpose, attempting to record from awake freely moving animals during in the training cage, I used probes with 4 shanks at 6 mm length and 200 micron spacing between the shanks. Each shank contained 8 recording sites at 50 micron distance (see Figure 6.4.).

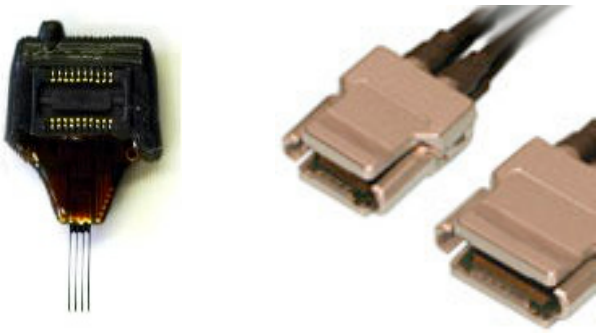
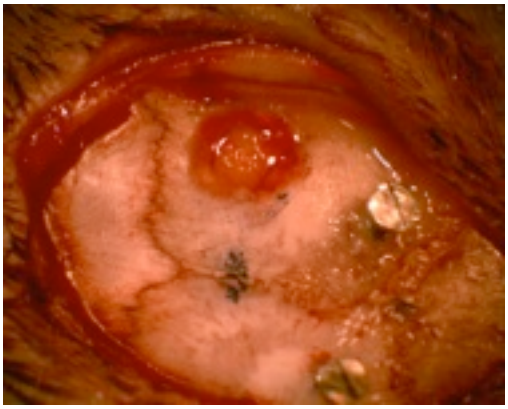


Figure 6.4. T Type Of Chronic Probe Used For Awake Recordings In The Training Cage. *(**Left**) The probe used as a chronic implant (Neuronexus Z-series). (**Right**) The ZIF-Clip headstage (TDT Z32) used with this type of probe for recordings in awake behaving animals in the training cage.*

Inspired by results from (Luczak et al., 2009) on the representation of auditory stimulus classes by patterns of distributed over rat A1, and considering what kind of surgery would have the smallest impact on the animals recovery, I attempted to surgically implant these probes into layer 5 of A1 tangentially from the top of the skull (Figure 6.5.). This type of surgery does not require the removal of the jaw muscle, a necessary procedure when accessing AC perpendicularly. The neural data presented in the next section was recorded from an awake behaving animal implanted as described in Figure 6.5.



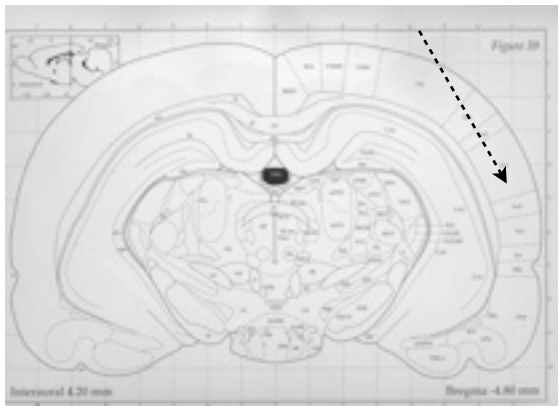


Figure 6.5. Craniotomy On And Schematic Of Track Of Chronic Probes

*(**Left**) View of the craniotomy from on the top left of the skull. Micro-screws are used for reference of electrical potentials. Marker specifies the position of the craniotomy (lambda - 5 mm posterior, 4 mm lateral). (**Right**) Schematic of the track of chronic probes from the position of the craniotomy to primary auditory cortex. Probes were implanted at a lateral angle of 35° from the vertical, as indicated by the arrow. The probes used during test recordings are shown in Figure 6.4.*

6.6. Data Collection In The Training Cage

6.6.1. An Easy Discrimination Task

To find whether rats could successfully learn a discrimination of double-formant vowels I trained three animals in a go-nogo task involving double-formant stimuli that had very different formant combinations. Figure 6.6. shows the performance of two animals in this task (the third animal performed at similar levels and is not shown for readability). The fundamental frequency of these stimuli was 151 Hz, and the centre frequencies of the first two formants in each vowel were F1 = 3.3 / F2 = 5.8 kHz for the appetitive stimulus and F1 = 16.8 / F2 = 19.7 kHz for the aversive stimulus. I chose these formant frequencies arbitrarily with the aim to keep the respective formants in one vowel far apart from those in the other vowel to render a discrimination easy. This discrimination task was the proof of concept that my animals could both use the training cage in the desired way, and they had no difficulty performing this easy vowel discrimination in that environment.

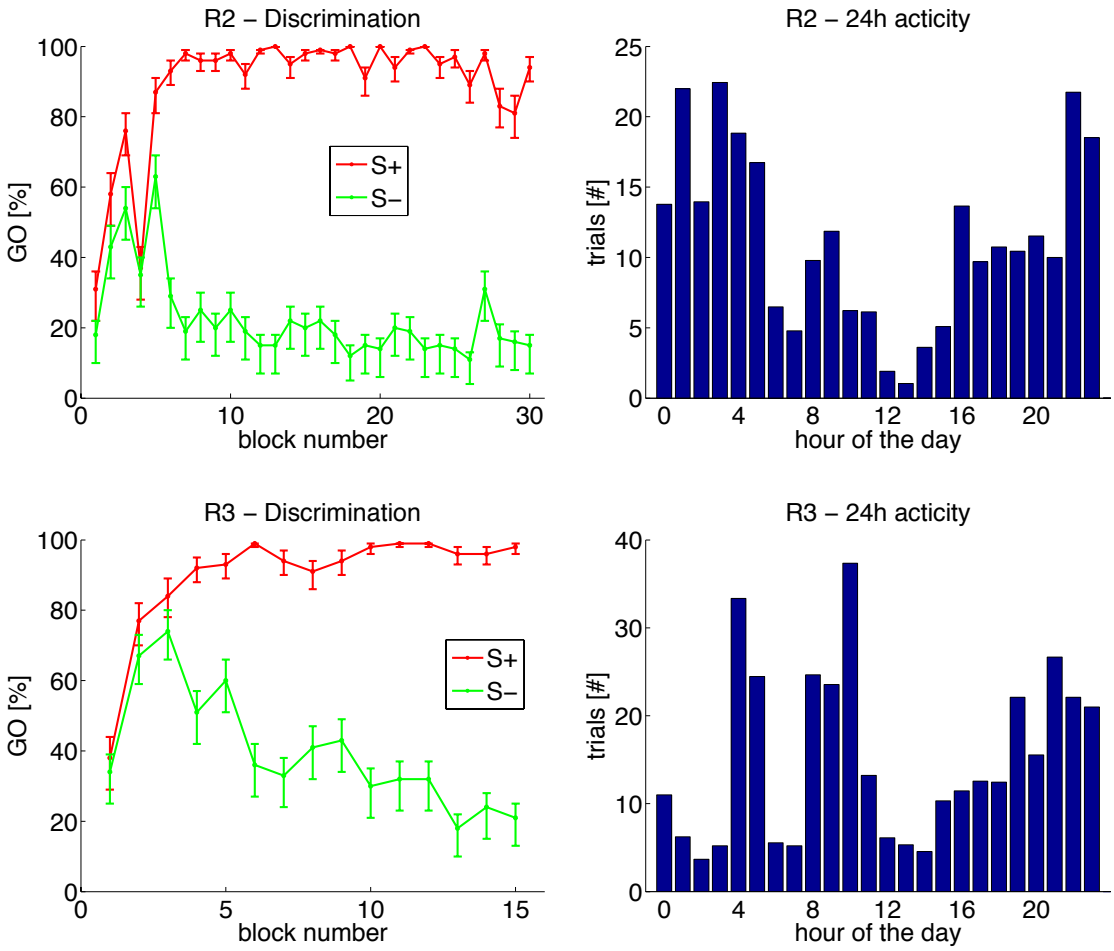


Figure 6.6. Performance Of Two Animals In A Simple Vowel Discrimination Task. *(**Left column**) %GO responses for rewarded (red) double-formant and non-rewarded (green) double-formant stimuli. All three animal employed a strategy of initial withdrawal, learning to stay inside the nose poke hole in response rewarded vowel (S+) with training. Error bars are confidence intervals estimated from a cumulative binomial distribution with mean performance values as probabilities and block size as number of trials (explained in detail in the Results of Chapter 3). Size of blocks is 100 trials. (**Right column**) Average number of trials over 24 hours for each animal (R1-R3) over the training periods shown in the left column. Note that animals are often performing the task when researchers are usually asleep.*

6.6.2. Recording Of Neural Data In Awake Behaving Animals

The data shown in Figure 6.7. serve purely to show the feasibility of awake neural recordings in the training cage, in animals implanted with chronic probes the way described in the previous section. Due to time constraints, I was unable to harness the full potential of the training and recording system for a fully developed study. Figure 6.7. shows histograms of A1 neural responses to broadband noise bursts recorded from an awake behaving rat that implanted with a chronic probe with the technique described in section 6.5. The data was collected inside the training cage using only the hardware described in this chapter.

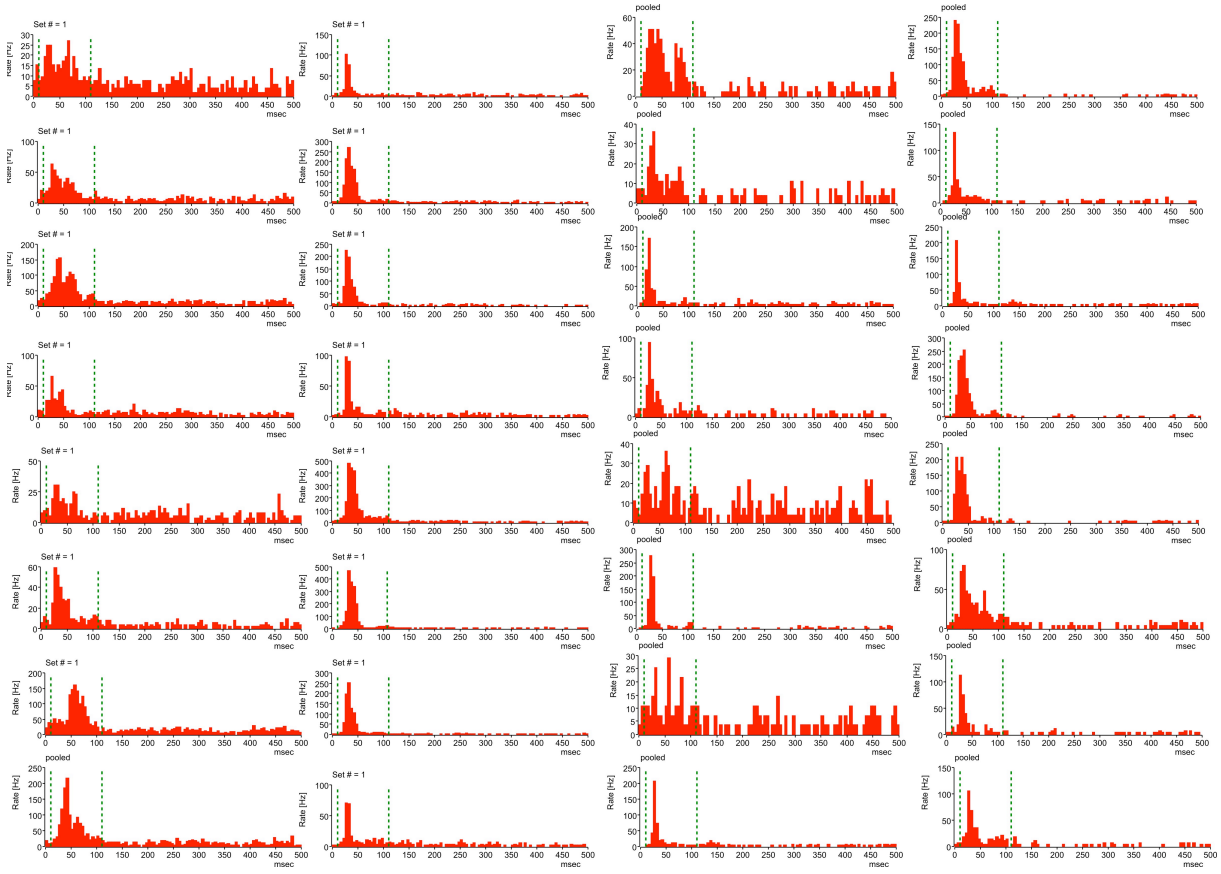


Figure 6.7. A1 Responses In An Awake Behaving Animal Inside The Training Cage

Histograms of the pooled auditory cortical activity (multi-unit spikes) in response to short (50 ms) broadband noise bursts recorded inside the training cage from an awake freely behaving animal. Each column represents one shank on the 4x8 probe shown in Figure 6.4. Each column represents one shank of the probe. The electrode was chronically implanted as shown in Figure 6.5. and recordings were made using the components of the training cage described in this chapter.

6.7. Summary & Conclusions

The training system I developed and used for behavioural training overcomes several limitations and difficulties often met in neural research that involves behavioural paradigms. Foremost the training system allows continuous training leading to a high daily number of trials. Natural activity phases of animals, which can be out of phase with those of the researcher, are not missed. The contact between animals and researchers is minimized to the advantage of both: animals are less stressed because they are not removed from their home cage for training and researcher spend less time in an allergen-rich environment. Along with training the cage allows the simultaneous recording of neural activity in response to presented stimuli from chronically implanted animals. The monitoring capabilities allow easy access to data and can lead to a tight control of the animals' health. This system has the potential to significantly improve the outcomes of behavioural training and awake recordings in rats.

CHAPTER 7 – GENERAL CONCLUSIONS

In the research I conducted for this thesis I explored neural and behavioural representations of a number of characteristics of perceptual auditory classes. To classify a sound as member of a particular sound class requires the recognition of the defining similarities that the sound shares with all other members of its class. The similarities that underlie a sound classification are often based on combinations of acoustic components which, within certain limits of variation, represent the “acoustic fingerprint” of the auditory class. For most auditory classes a large number of member sounds are perceptually discriminable and classification is thought to involve judging how similar a given sound to prototypes at the core of each available class.

Vowels are a useful tool for exploring auditory classes that are defined on combinations of acoustic components. The combination of the first and second formant determine the class membership of vowels, and the space of human vowels spanned by first and second formant frequencies is tiled with distinct regions of formant combinations: the vowel classes of human speech. The main characteristics of many phonetic classes are exemplified by vowels: Vowel classes are perceptually invariant regarding changes along irrelevant (e.g. loudness) and relevant (e.g. formant frequency) acoustic dimensions. Finally, prototypical vowels (the central members of a vowel class) are perceived as perceptual magnets by humans (Kuhl, 1991) and by birds (Kluender et al., 1998). I made use of these characteristics of phonetic classes, using artificial vowels to investigate how neural responses in candidate brain areas of auditory class processing respond to different combinations of formants, the components that define human vowel classes.

The primary auditory cortex (A1) is one of these candidate areas; the integration of spectral components by A1 single units and multi-units is diverse (Israel Nelken, 2008; Bizley et al., 2009), and linear models explain only part of the responses of A1 units to stimuli with rich spectral and/or temporal structure (Machens et al., 2004; Theunissen and Shaevitz, 2006; Bar Yosef and Nelken, 2007). The CA1 region of the hippocampus is another, much “higher order” candidate area. Object-selective responses of single units in the CA1 region of the hippocampus have been described in humans (Quian Quiroga et al., 2005; 2009) and monkeys (Reddy and Kanwisher, 2006) and a recent study in rats (Itskov et al., 2011) showed single CA1 single units represent different surface textures in awake behaving animals.

In Chapter 2 of this thesis I reported that cortical multi-units in A1 of anaesthetized rats discriminate shifts of formants in double-formant combinations at a *resolution* of roughly 7% difference between formants. This resolution lies at the upper end of the rats behavioural frequency discrimination threshold for pure tones (Syka et al., 1996; Sloan et al., 2009) and provides a baseline level of formant shifts for the definition of formant classes that A1 multi-units can still discriminate well. Discriminability between class members is an important characteristic of many phonetic classes. My results show A1 responses can follow acoustic changes in simplified auditory objects like vowels with a relatively high resolution. And in my behavioural chapter (Chapter) on the other hand I found that rats can actually learn to classify sounds that A1 multi-units did reliably discriminate. Therefore, A1 does not exhibit the invariance of responses that could be expected of neural correlates of object classification in general and categorical perception in particular.

In Chapter 3 I investigated vowel perception behaviourally. I trained rats in discrimination and classification tasks designed to probe two perceptual abilities: the ability to form perceptual classes, based on acoustic similarity, and the ability to spontaneously judge the similarity of combinations of sound components - in the case of vowel stimuli, of formant combinations. I used artificial vowel stimuli with formant distances marginally above the average cortical discrimination level found in Chapter 1. I trained animals to classify formant combinations with this distance between the formants of different class members. Rats quickly learned to classify formant combinations and proved to not rely on just one formant frequency as a cue for similarity judgements. That animals use formant combinations for similarity judgements means that their auditory system must display some degree of sensitivity for formant combinations. Whether this sensitivity is a feature of A1 multi-units responses was the central question of the research I carried out for Chapter 4.

In Chapter 4, I explored how A1 responses integrate the spectral components that define vowels classes. If vowel classes are defined by formant combinations, then it is pertinent to ask whether or in which manner A1 neurons might be “combination sensitive”. I discovered that a simple model based on formant integration in an additive (OR-like) manner captures the structure receptive fields of formant combinations with some success. In contrast a multiplicative (AND-like) model of formant integration was substantially less successful in describing the receptive field structure. Combining the additive with the multiplicative model did not notably improve the model fits. Consequently a substantial amount of stimulus related variance in the FRFs of A1 multi-units remained unexplained. This observation is especially true for offset responses to formant combinations, where both AND- and OR-like models perform worse in capturing FRF structure. From these observations I

conclude that additive integration of formant energy in A1 multi-unit responses is an approximation of the underlying physiological integration function and that a multiplication of formant energy does not adequately approximate the nonlinear components of formant integration in A1. Object selective responses, sensitive to particular combinations of stimulus components, have been reported in responses from the CA1 region of the hippocampus.

In Chapter 5 I presented the results of a collaborative study that focussed on the interplay of hippocampal responses to single-formant vowels and spatial location. We recorded responses from single cells from the CA1 region of the hippocampus of awake and freely moving rats while they were involved in a vowel discrimination task but also during passive listening. One result which was relevant to this thesis was that while many of CA1 neurons discriminated between single-formant vowels during task performance, they did not respond at all to the very same stimuli during passive listening. This result demonstrates that many hippocampal neurons are sensitive to formant changes but that this sensitivity only manifests itself in certain behavioural contexts, in line with perspectives of hippocampal function as building associations between objects and places (Gilbert and Kesner, 2004), probably to aid in the formation of episodic memory. Our results also show that hippocampal responses in rats are not *primarily* sensitive to stimulus identity but appear to integrate spatial information and stimulus related information is already available. The fact that response latencies in hippocampus were much longer than maximal speeds of classification in humans and animals supports this interpretation.

The distance of formants in this study was deliberately chosen to be large to render the discrimination task easy and to make a first attempt at detecting response

differences in rat hippocampus related to different vowels. So even though no direct conclusion on neural discriminability or classifications based on acoustic similarities can be drawn from this study it nevertheless set the stage for further exploration of speech sound classes in the hippocampus. This study also demonstrated the importance, when studying the neural correlates of class learning, to record putative neural correlates of sound classification in the behavioural context within which the auditory classes are learned.

In Chapter 6 I described an integrated housing & training environment, that I designed and built, which allows the training and recording of neural responses in one and the same environment. The automated training cage allows rats to train themselves in various auditory discrimination and classification tasks, and neural responses to the chosen acoustic stimuli can be recorded while animals perform their tasks. Because the training setup is an integral part of the animals' home cage, this also allows the monitoring of changes in neural responses as a function of learning. That neural responses in A1 change as a function of learning has been demonstrated for simple stimuli like pure tones (Weinberger, 2004) and more complex sounds, such as multi-tone patterns (Fritz et al., 2007). Neural responses recorded under anaesthesia can substantially differ from those recorded in awake behaving animals (Gaese and Ostwald, 2001; 2003) and the development of neural recording paradigms that can circumvent the effects of anaesthesia is central to deepening our understanding of the effects of learning in areas such as A1 and the hippocampus.

In conclusion, my results suggests that A1 is not the seat of highly invariant correlates of perceptual auditory objects. But its fine resolution of formant changes, as well as its nontrivial manner of integrating formants, indicate that A1 is involved in

the identification of sound features that act as a substrate for auditory object classification. Another candidate area for the formation of perceptually invariant auditory objects is the mammalian hippocampus and our results show that the CA1 region of the rat hippocampus can discriminate between vowel formants. Yet, we also observed that hippocampal responses are highly context dependent and occur at latencies that exceed behavioural limits of rapid class recognition. Therefore the hippocampus probably integrates class related information that is provided by other regions of the auditory processing network. Where in the auditory system combinations of acoustic features are detected to generate invariant auditory object representations should ideally be explored in awake behaving animals while they are learning auditory classification tasks. I demonstrated that rats are able to learn vowel classes based on acoustic similarity and that they will spontaneously use formant combinations to make similarity judgements. Combining behavioural training on auditory classes with electrophysiological recordings in awake behaving animals, for example in a training environment like the one I designed, will help solve the long-standing question of how the auditory system generates perceived objects from the patterns of sound waves that tickle our eardrums.

REFERENCES

Abeles, M., and Goldstein, M.H. (1970). Functional architecture in cat primary auditory cortex: columnar organization and organization according to depth. J Neurophysiol 33, 172–187.

Allen, W.F. (1945). Effect of destroying three localized cerebral cortical areas for sound on correct conditioned differential responses of the dog's foreleg. American Journal of Physiology.

Anderson, M.I., and Jeffery, K.J. (2003). Heterogeneous modulation of place cell firing by changes in context. Journal of Neuroscience 23, 8827–8835.

Bar Yosef, O., and Nelken, I. (2007). The effects of background noise on the neural responses to natural sounds in cat primary auditory cortex. Front Comput Neurosci 1, 3.

Bar Yosef, O., Rotman, Y., and Nelken, I. (2002). Responses of neurons in cat primary auditory cortex to bird chirps: effects of temporal and spectral context. Journal of Neuroscience 22, 8619–8632.

Bartley, S.H., and Nelson, T.M. (1963). Some Relations Between Sensory end Results and Neural Activity in the Optic Pathway. The Journal of Psychology 55, 121–143.

Batzri-Izraeli, R., Kelly, J.B., Glendenning, K.K., Masterton, R.B., and Wollberg, Z. (1990). Auditory Cortex of the Long-Eared Hedgehog (Hemiechinus auritus). Brain Behav Evol 36, 237–248.

Berger, T., Alger, B., and Thompson, R. (1976). Neuronal substrate of classical conditioning in the hippocampus. Science 192, 483–485.

Berger, T.W., Rinaldi, P.C., Weisz, D.J., and Thompson, R.F. (1983). Single-unit analysis of different hippocampal cell types during classical conditioning of rabbit nictitating membrane response. J Neurophysiol 50, 1197–1219.

Bizley, J.K., Nodal, F.R., Nelken, I., and King, A.J. (2005). Functional organization of ferret auditory cortex. Cereb Cortex 15, 1637–1653.

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. J Neurosci 29, 2064–2075.

Blake, D.T., and Merzenich, M.M. (2002). Changes of AI receptive fields with sound density. J Neurophysiol 88, 3409–3420.

Bostock, E., Muller, R.U., and Kubie, J.L. (1991). Experience-dependent modifications of hippocampal place cell firing. Hippocampus 1, 193–205.

Bregman, A.S. (1994). Auditory scene analysis (The MIT Press).

Bregman, A.S., and Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. Journal of Experimental Psychology 89, 244–249.

Brown, S.D. (1988). Speech perception by budgerigars (Melopsittacus undulatus): Spoken vowels. J Acoust Soc Am 83, S51–S51.

Brugge, J.F., and Merzenich, M.M. (1973). Responses of neurons in auditory cortex of the macaque monkey to monaural and binaural stimulation. Journal of Neurophysiology 36, 1138–1158.

Burdick, C.K., and Miller, J.D. (1975). Speech perception by the chinchilla: discrimination of sustained /a/ and /i/. J Acoust Soc Am 58, 415–427.

Burwell, R.D., and Amaral, D.G. (1998). Cortical afferents of the perirhinal, postrhinal, and entorhinal cortices of the rat. J Comp Neurol 398, 179–205.

Burwell, R.D., Witter, M.P., and Amaral, D.G. (1995). Perirhinal and postrhinal cortices of the rat: A review of the neuroanatomical literature and comparison with findings from the monkey brain. Science 5, 390–408.

Calford, M.B., and Semple, M.N. (1995). Monaural inhibition in cat auditory cortex. J Neurophysiol 73, 1876–1891.

Calhoun, B.M., and Schreiner, C.E. (1998). Spectral envelope coding in cat primary auditory cortex: linear and non-linear effects of stimulus characteristics. Eur J Neurosci 10, 926–940.

Campbell, R.A.A., Doubell, T.P., Nodal, F.R., Schnupp, J.W.H., and King, A.J. (2006). Interaural timing cues do not contribute to the map of space in the ferret superior colliculus: a virtual acoustic space study. J Neurophysiol 95, 242–254.

Christian, E.P., and Deadwyler, S.A. (1986). Behavioral functions and hippocampal cell types: evidence for two nonoverlapping populations in the rat. J Neurophysiol 55, 331–348.

Clarey, J.C., Barone, P., and Imig, T.J. (1994). Functional organization of sound direction and sound pressure level in primary auditory cortex of the cat. J Neurophysiol 72, 2383–2405.

Cohen, N.J., and Eichenbaum, H. (1995). Memory, Amnesia, and the Hippocampal System (The MIT Press).

Colgin, L.L., Moser, E.I., and Moser, M.-B. (2008). Understanding memory through hippocampal remapping. Trends Neurosci 31, 469–477.

Csicsvari, J., Hirase, H., Czurkó, A., Mamiya, A., and Buzsáki, G. (1999). Oscillatory coupling of hippocampal pyramidal cells and interneurons in the behaving Rat. J Neurosci 19, 274–287.

deCharms, R.C., Blake, D.T., and Merzenich, M.M. (1998). Optimizing sound features for cortical neurons. Science 280, 1439–1443.

DeWeese, M.R., Wehr, M., and Zador, A.M. (2003). Binary spiking in auditory cortex. J Neurosci 23, 7940–7949.

Dewson, J.H. (1964). Speech sound discrimination by cats. Science 144, 555–556.

Diamond, D.M., Fleshner, M., Ingersoll, N., and Rose, G. (1996). Psychological stress impairs spatial working memory: Relevance to electrophysiological studies of hippocampal function. Behav Neurosci 110, 661–672.

Diesch, E., & Luce, T. (1997). Magnetic fields elicited by tones and vowel formants reveal tonotopy and nonlinear summation of cortical activation. Psychophysiology, 34(5), 501–510.

Dooling, R.J., and Brown, S.D. (1990). Speech perception by budgerigars (Melopsittacus undulatus): spoken vowels. Percept Psychophys 47, 568–574.

Dooling, R.J., Okanoya, K., and Brown, S.D. (1989). Speech perception by budgerigars(Melopsittacus undulatus): The voiced-voiceless distinction. Percept Psychophys 46, 65–71.

Douglas, R.J., and Martin, K.A. (1991). A functional microcircuit for cat visual cortex. J Physiol (Lond) 440, 735–769.

Doupe, A.J., and Kuhl, P.K. (1999). Birdsong and human speech: common themes and mechanisms. Annu Rev Neurosci 22, 567–631.

Eacott, M.J., and Norman, G. (2004). Integrated memory for object, place, and context in rats: a possible model of episodic-like memory? Journal of Neuroscience 24, 1948–1953.

Eichenbaum, H., Dudschenko, P., Wood, E., Shapiro, M., and Tanila, H. (1999). The Hippocampus, Memory, and Place Cells: Is It Spatial Memory or a Memory Space? Neuron 23, 209–226.

Eichenbaum, H., Kuperstein, M., Fagan, A., and Nagode, J. (1987). Cue-sampling and goal-approach correlates of hippocampal unit activity in rats performing an odor-discrimination task. J Neurosci 7, 716–732.

Eimas, P.D., Siqueland, E.R., Jusczyk, P., and Vigorito, J. (1971). Speech perception in infants. Science 171, 303–306.

Engineer, C.T., Perez, C.A., Chen, Y.H., Carraway, R.S., Reed, A.C., Shetake, J.A., Jakkamsetti, V., Chang, K.Q., and Kilgard, M.P. (2008). Cortical activity patterns predict speech discrimination ability. Nat Neurosci 11, 603–608.

Eriksson, J.L., and Villa, A.E.P. (2006). Learning of auditory equivalence classes for vowels by rats. Behav Processes 73, 348–359.

Evans, E.F., Ross, H.F., and Whitfield, I.C. (1965). The spatial distribution of unit characteristic frequency in the primary auditory cortex of the cat. J Physiol (Lond) 179, 238–247.
Fay, R.R. (1970). Auditory frequency discrimination in the goldfish (Carassius auratus). J Comp Physiol Psychol 73, 175.
Firzlaff, U., Schuchmann, M., Grunwald, J.E., Schuller, G., and Wiegrebe, L. (2007). Object-oriented echo perception and cortical representation in echolocating bats. PLoS Biol 5, e100.
Fox, S.E., and Ranck, J.B. (1981). Electrophysiological characteristics of hippocampal complex-spike cells and theta cells. Experimental Brain Research. Experimentelle Hirnforschung. Experimentation Cerebrale 41, 399–410.
Freedman, D.J., Riesenhuber, M., Poggio, T., and Miller, E.K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. Science 291, 312–316.
Fritz, J., Shamma, S., Elhilali, M., and Klein, D. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. Nat Neurosci 6, 1216–1223.
Fritz, J.B., Elhilali, M., and Shamma, S.A. (2007). Adaptive changes in cortical receptive fields induced by attention to complex sounds. J Neurophysiol 98, 2337–2346.
Gaese, B.H., and Ostwald, J. (2001). Anesthesia changes frequency tuning of neurons in the rat primary auditory cortex. J Neurophysiol 86, 1062–1066.
Gaese, B.H., and Ostwald, J. (2003). Complexity and temporal dynamics of frequency coding in the awake rat auditory cortex. Eur J Neurosci 18, 2638–2652.
Garcia-Lazaro, J.A., Ahmed, B., and Schnupp, J.W.H. (2006). Tuning to natural stimulus dynamics in primary auditory cortex. Curr Biol 16, 264–271.
Gelbard-Sagiv, H., Mukamel, R., Harel, M., Malach, R., and Fried, I. (2008). Internally Generated Reactivation of Single Neurons in Human Hippocampus During Free Recall. Science 322, 96–101.
Gilbert, P.E., and Kesner, R.P. (2004). Memory for objects and their locations: the role of the hippocampus in retention of object-place associations. Neurobiol Learn Mem 81, 39–45.
Hampson, R.E., Pons, T.P., Stanford, T.R., and Deadwyler, S.A. (2004). Categorization in the monkey hippocampus: a possible mechanism for encoding information into memory. Proc Natl Acad Sci USA 101, 3184–3189.
Harrington, I.A., Stecker, G.C., Macpherson, E.A., and Middlebrooks, J.C. (2008). Spatial sensitivity of neurons in the anterior, posterior, and primary fields of cat auditory cortex. Hear Res 240, 22–41.
Harris, J.D. (1952). Pitch Discrimination. J Acoust Soc Am 24, 750.
Hastie, T., Tibshirani, R., and Friedman, J. (2009). The Elements of Statistical Learning (Springer).
Haxby, J.V., Grady, C.L., Horwitz, B., Ungerleider, L.G., Mishkin, M., Carson, R.E., Herscovitch, P., Schapiro, M.B., and Rapoport, S.I. (1991). Dissociation of object and spatial visual processing pathways in human extrastriate cortex. Proc Natl Acad Sci USA 88, 1621–1625.
Heffner, H., Heffner, R., Contos, C., and Ott, T. (1994). Audiogram of the hooded Norway rat. Hear Res 73, 244–247.
Heil, P. (1997). Auditory cortical onset responses revisited. II. Response strength. J Neurophysiol 77, 2642–2660.
Heise, G.A., and Miller, G.A. (1951). An Experimental Study of Auditory Patterns. The American Journal of Psychology 64, 68–77.
Hellweg, F.C., Koch, R., and Vollrath, M. (1977). Representation of the cochlea in the neocortex of guinea pigs. Experimental Brain Research. Experimentelle Hirnforschung. Experimentation Cerebrale 29-29.
Hienz, R.D. (1988). The acquisition of vowel discriminations by nonhuman primates. J Acoust Soc Am 84, 186.
Hienz, R.D. (1996). Vowel discrimination in cats: Thresholds for the detection of second formant changes in the vowel /ε/. J Acoust Soc Am 100, 1052.

Hienz, R.D., Sachs, M.B., and Aleszczyk, C.M. (1993). Frequency discrimination in noise: Comparison of cat performances with auditory-nerve models. J Acoust Soc Am 93, 462.
Hienz, R.D., Sachs, M.B., and Sinnott, J.M. (1981). Discrimination of steady-state vowels by blackbirds and pigeons. J Acoust Soc Am 70, 699.
Hind, J.E. (1953). An electrophysiological determination of tonotopic organization in auditory cortex of cat. Journal of Neurophysiology.
Hind, J.E., Rose, J.E., Davies, P.W., and Woolsey, C.N. (1960). Hind: Unit activity in the auditory cortex - Google Scholar (Neural Mechanisms of the …).
Ho, A.S., Hori, E., Thi Nguyen, P.H., Urakawa, S., Kondoh, T., Torii, K., Ono, T., and Nishijo, H. (2011). Hippocampal neuronal responses during signaled licking of gustatory stimuli in different contexts. Science 21, 502–519.
Hok, V., Lenck-Santini, P.-P., Roux, S., Save, E., Muller, R.U., and Poucet, B. (2007). Goal-related activity in hippocampal place cells. Journal of Neuroscience 27, 472–482.
Hromádka, T., DeWeese, M.R., and Zador, A.M. (2008). Sparse representation of sounds in the unanesthetized auditory cortex. PLoS Biol 6, e16.
Hubel, D.H., and Wiesel, T.N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol (Lond) 160, 106.
Hubel, D.H., and Wiesel, T.N. (1963). Shape and arrangement of columns in cat's striate cortex. J Physiol (Lond) 165, 559–568.
Israel Nelken, O.B.-Y. (2008). Neurons and Objects: The Case of Auditory Cortex. Frontiers in Neuroscience 2, 107.
Itskov, P.M., Vinnik, E., and Diamond, M.E. (2011). Hippocampal representation of touch-guided behavior in rats: persistent and independent traces of stimulus and reward location. PLoS ONE 6, e16462.
Itskov, P.M., Vinnik, E., Honey, C., Schnupp, J., and Diamond, M.E. (2012). Sound sensitivity of neurons in rat hippocampus during performance of a sound-guided task. J Neurophysiol 107, 1822–1834.
Jensen, O., and Lisman, J.E. (2000). Position reconstruction from an ensemble of hippocampal place cells: contribution of theta phase coding. J Neurophysiol 83, 2602–2609.
Kalatsky, V.A., Polley, D.B., Merzenich, M.M., Schreiner, C.E., and Stryker, M.P. (2005). Fine functional organization of auditory cortex revealed by Fourier optical imaging. Proc Natl Acad Sci USA 102, 13325–13330.
Keeling, M.D., Calhoun, B.M., Krüger, K., Polley, D.B., and Schreiner, C.E. (2008). Spectral integration plasticity in cat auditory cortex induced by perceptual training. Experimental Brain Research. Experimentelle Hirnforschung. Experimentation Cerebrale 184, 493–509.
Kelly, J.B., and Masterton, B. (1977). Auditory sensitivity of the albino rat. J Comp Physiol Psychol 91, 930–936.
Kennedy, P.J., and Shapiro, M.L. (2009). Motivational states activate distinct hippocampal representations to guide goal-directed behaviors. Proc Natl Acad Sci US A 106, 10805–10810.
Kewley-Port, D., and Watson, C.S. (1994). Formant-frequency discrimination for isolated English vowels. J Acoust Soc Am 95, 485–496.
Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. Vision research 46(11), 1762–1776.
Klausberger, T., and Somogyi, P. (2008). Neuronal Diversity and Temporal Dynamics: The Unity of Hippocampal Circuit Operations. Science 321, 53.
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, 85–111.
Klein, R.L., Meyer, E.M., Peel, A.L., Zolotukhin, S., Meyers, C., Muzyczka, N., and King, M.A. (1998). Neuron-specific transduction in the rat septohippocampal or nigrostriatal pathway by recombinant adeno-associated virus vectors. Exp Neurol 150, 183–194.

Kluender, K., Diehl, R., and Killeen, P. (1987). Japanese quail can learn phonetic classes. Science 237, 1195–1197.
Kluender, K.R., Lotto, A.J., Holt, L.L., and Bloedel, S.L. (1998). Role of experience for language-specific functional mappings of vowel sounds. J Acoust Soc Am 104, 3568–3582.
Komorowski, R.W., Manns, J.R., and Eichenbaum, H. (2010). Conjunctive Representations in the Hippocampus: What and Where? 30, 799–801.
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. J Neurophysiol 76, 3503–3523.
Kuhl, P. (1983). Perception of auditory equivalence classes for speech in early infancy*. Infant Behavior and Development 6, 263–285.
Kuhl, P., and Miller, J. (1975). Speech perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants. Science 190, 69–72.
Kuhl, P.K. (1981). Discrimination of speech by nonhuman animals: Basic auditory sensitivities conducive to the perception of speech-sound classes. J Acoust Soc Am 70, 340.
Kuhl, P.K. (1991). Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech classes, monkeys do not. Percept Psychophys 50, 93–107.
Kuhl, P.K., and Padden, D.M. (1982). Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques. Percept Psychophys 32, 542–550.
Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N., and Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. Science 255, 606–608.
Langner, G., Bonke, D., and Scheich, H. (1981). Neuronal discrimination of natural and synthetic vowels in field L of trained mynah birds. Experimental Brain Research. Experimentelle Hirnforschung. Experimentation Cerebrale 43, 11–24.
Lenck-Santini, P.-P., Fenton, A.A., and Muller, R.U. (2008). Discharge properties of hippocampal neurons during performance of a jump avoidance task. Journal of Neuroscience 28, 6773–6786.
Leutgeb, J.K., Leutgeb, S., Treves, A., Meyer, R., Barnes, C.A., McNaughton, B.L., Moser, M.-B., and Moser, E.I. (2005). Progressive Transformation of Hippocampal Neuronal Representations in “Morphed” Environments. Neuron 48, 345–358.
Leutgeb, S. (2004). Distinct Ensemble Codes in Hippocampal Areas CA3 and CA1. Science 305, 1295–1298.
Leutgeb, S. (2005). Independent Codes for Spatial and Episodic Memory in Hippocampal Neuronal Ensembles. Science 309, 619–623.
Lever, C., Wills, T., Cacucci, F., Burgess, N., and O'Keefe, J. (2002). Long-term plasticity in hippocampal place-cell representation of environmental geometry. Nature 416, 90–94.
Liberman, A., Harris, K.S., Eimas, P., Lisker, L., and Bastian, J. (1961). An effect of learning on speech perception: The discrimination of durations of silence with and without phonemic significance. Language and Speech 4, 175–195.
Liberman, A.M., Harris, K.S., Hoffman, H.S., and Griffith, B.C. (1957). The discrimination of speech sounds within and across phoneme boundaries. Journal of Experimental Psychology 54, 358–368.
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. J Neurophysiol 90, 2660–2675.
Lisker, L. (1957). Closure duration and the intervocalic voiced-voiceless distinction in English. Language 33, 42–49.
Lisker, L., and Abramson, A.S. (1964). A cross-language study of voicing in initial stops: acoustical measurements. Word 20, 384–422.
Lomber, S.G., and Malhotra, S. (2008). Double dissociation of 'what' and “where” processing in auditory cortex. Nat Neurosci 11, 609–616.
Lotto, A.J., Kluender, K.R., and Holt, L.L. (1997). Perceptual compensation for coarticulation by Japanese quail (Coturnix coturnix japonica). J Acoust Soc Am 102, 1134–1140.

Luczak, A., Barthó, P., and Harris, K.D. (2009). Spontaneous events outline the realm of possible sensory responses in neocortical populations. Neuron 62, 413–425.
Luine, V., Villegas, M., Martinez, C., and McEwen, B.S. (1994). Repeated stress causes reversible impairments of spatial memory performance. Brain Research 639, 167–170.
Machens, C.K., Wehr, M.S., and Zador, A.M. (2004). Linearity of cortical receptive fields measured with natural sounds. J Neurosci 24, 1089–1100.
Maeng, L.Y., Waddell, J., and Shors, T.J. (2010). The prefrontal cortex communicates with the amygdala to impair learning after acute stress in females but not in males. J Neurosci 30, 16188–16196.
Magri, C., Whittingstall, K., Singh, V., Logothetis, N.K., and Panzeri, S. (2009). A toolbox for the fast information analysis of multiple-site LFP, EEG and spike train recordings. BMC Neuroscience 10, 81.
Manns, J.R., and Eichenbaum, H. (2009). A cognitive map for object memory in the hippocampus. Learning & Memory 16, 616–624.
Marr, D. (1971). Simple memory: a theory for archicortex. Philos Trans R Soc Lond, B, Biol Sci 262, 23–81.
Mcalpine, D. (2005). Creating a sense of auditory space. J Physiol (Lond) 566, 21–28.
McMullen, N.T., and Glaser, E.M. (1982). Tonotopic organization of rabbit auditory cortex. Exp Neurol 75, 208–220.
Mendelson, J.R., Schreiner, C.E., and Sutter, M.L. (1997). Functional topography of cat primary auditory cortex: response latencies. J Comp Physiol A 181, 615–633.
Mendelson, J.R., Schreiner, C.E., Sutter, M.L., and Grasse, K.L. (1993). Functional topography of cat primary auditory cortex: responses to frequency-modulated sweeps. Experimental Brain Research. Experimentelle Hirnforschung. Experimentation Cerebrale 94, 65–87.
Merzenich, M.M., Knight, P.L., and Roth, G.L. (1975). Representation of cochlea within primary auditory cortex in the cat. J Neurophysiol 38, 231–249.
Mesgarani, N., David, S.V., Fritz, J.B., and Shamma, S.A. (2008). Phoneme representation and classification in primary auditory cortex. J Acoust Soc Am 123, 899–909.
Meyer, D.R., and Woolsey, C.N. (1952). Effects of localized cortical destruction on auditory discriminative conditioning in cat. J Neurophysiol 15, 149–162.
Middlebrooks, J., Clock, A., Xu, L., and Green, D. (1994). A panoramic code for sound location by cortical neurons. Science 264, 842–844.
Miller, J.L., and Volaitis, L.E. (1989). Effect of speaking rate on the perceptual structure of a phonetic class. Percept Psychophys 46, 505–512.
Miller, L.M., Escabí, M.A., Read, H.L., and Schreiner, C.E. (2002). Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex. J Neurophysiol 87, 516–527.
Moita, M.A.P., Rosis, S., Zhou, Y., LeDoux, J.E., and Blair, H.T. (2003). Hippocampal place cells acquire location-specific responses to the conditioned stimulus during auditory fear conditioning. Neuron 37, 485–497.
Moita, M.A.P., Rosis, S., Zhou, Y., LeDoux, J.E., and Blair, H.T. (2004). Putting fear in its place: remapping of hippocampal place cells during fear conditioning. Journal of Neuroscience 24, 7015–7023.
Moshitch, D., Las, L., Ulanovsky, N., Bar Yosef, O., and Nelken, I. (2006). Responses of neurons in primary auditory cortex (A1) to pure tones in the halothane-anesthetized cat. J Neurophysiol 95, 3756–3769.
Muller, R.U., and Kubie, J.L. (1987). The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells. J Neurosci 7, 1951–1968.
Nelken, I. (2004). Processing of complex stimuli and natural scenes in the auditory cortex. Curr Opin Neurobiol 14, 474–480.
Nelken, I. (2008). Processing of complex sounds in the auditory system. Curr Opin Neurobiol 18, 413–417.
Neves, G., Cooke, S.F., and Bliss, T.V.P. (2008). Synaptic plasticity, memory and the hippocampus: a neural network approach to causality. Nat Rev Neurosci 9, 65–75.

Obleser, J., Lahiri, A., & Eulitz, C. (2003). Auditory-evoked magnetic field codes place of articulation in timing and topography around 100 milliseconds post syllable onset. NeuroImage, 20(3), 1839–1847.
Obleser, J., Elbert, T., Lahiri, A., & Eulitz, C. (2003). Cortical representation of vowels reflects acoustic dissimilarity determined by formant frequencies. Brain Research. Cognitive Brain Research, 15(3), 207–213.
O'Keefe, J. (1976). Place units in the hippocampus of the freely moving rat. Exp Neurol 51, 78–109.
O'Keefe, J., and Nadel, L. (1979). The cognitive map as a hippocampus. Behavioral and Brain Sciences 2, 520–533.
Ohl, F.W., and Scheich, H. (1997). Orderly cortical representation of vowels based on formant interaction. Proc Natl Acad Sci USA 94, 9440–9444.
Ohl, F.W., Scheich, H., and Freeman, W.J. (2001). Change in pattern of ongoing cortical activity with auditory class learning. Nature 412, 733–736.
Ohms, V.R., Gill, A., Van Heijningen, C.A.A., Beckers, G.J.L., and Cate, ten, C. (2010). Zebra finches exhibit speaker-independent phonetic perception of human speech. Proceedings of the Royal Society B: Biological Sciences 277, 1003–1009.
Orduña, I., Mercado, E., Gluck, M.A., and Merzenich, M.M. (2005). Cortical responses in rats predict perceptual sensitivities to complex sounds. Behav Neurosci 119, 256–264.
Panzeri, S., and Treves, A. (1999). Analytical estimates of limited sampling biases in different information measures. Network: Computation in Neural Systems 7, 87–107.
Panzeri, S., Senatore, R., Montemurro, M.A., and Petersen, R.S. (2007). Correcting for the sampling bias problem in spike train information measures. J Neurophysiol 98, 1064–1072.
Park, C.R., Campbell, A.M., and Diamond, D.M. (2001). Chronic psychosocial stress impairs learning and memory and increases sensitivity to yohimbine in adult rats. Biological Psychiatry 50, 994–1004.
Parkinson, J.K., Murray, E.A., and Mishkin, M. (1988). A selective mnemonic role for the hippocampus in monkeys: memory for the location of objects. J Neurosci 8, 4159–4167.
Pavlov, I.P. (1927). Article Search Results - Periodicals Archive Online (New York).
Paxinos, G., & Watson, C. (2005). The Rat Brain in Stereotaxic Coordinates. Academic Press.
Peña, J.L., and Konishi, M. (2001). Auditory spatial receptive fields created by multiplication. Science 292, 249–252.
Peterson, G.E., and Barney, H.L. (1952). Control methods used in a study of the vowels. J Acoust Soc Am 24, 175.
Phillips, D.P. (1985). Temporal response features of cat auditory cortex neurons contributing to sensitivity to tones delivered in the presence of continuous noise. Hear Res 19, 253–268.
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. J Neurophysiol 45, 48–58.
Phillips, D.P., Hall, S.E., and Boehnke, S.E. (2002). Central auditory onset responses, and temporal asymmetries in auditory perception. Hear Res 167, 192–205.
Pind, J. (1995). Constancy and normalization in the perception of Voice Offset Time as a cue for preaspiration. Acta Psychol (Amst) 89, 53–81.
Pind, J. (1998). Auditory and linguistic factors in the perception of voice offset time as a cue for preaspiration. J Acoust Soc Am 103, 2117–2127.
Pola, G., Thiele, A., Hoffmann, K.P., and Panzeri, S. (2003). An exact method to quantify the information transmitted by different mechanisms of correlational coding. Network (Bristol, England) 14, 35–60.
Prosen, C.A. (1990). Frequency discrimination in the monkey. J Acoust Soc Am 88, 2152.
Qin, L., Chimoto, S., Sakai, M., Wang, J., and Sato, Y. (2007). Comparison between offset and onset responses of primary auditory cortex ON-OFF neurons in awake cats. J Neurophysiol 97, 3421–3431.

Quian Quiroga, R., Kraskov, A., Koch, C., and Fried, I. (2009). Explicit encoding of multimodal percepts by single neurons in the human brain. Curr Biol 19, 1308–1313.
Quian Quiroga, R., Reddy, L., Kreiman, G., Koch, C., and Fried, I. (2005). Invariant visual representation by single neurons in the human brain. Nature 435, 1102–1107.
Ranck, J.B. (1973). Studies on single neurons in dorsal hippocampal formation and septum in unrestrained rats. I. Behavioral correlates and firing repertoires. Exp Neurol 41, 461–531.
Rauschecker, J.P., Tian, B., Pons, T., and Mishkin, M. (1997). Serial and parallel processing in rhesus monkey auditory cortex. J Comp Neurol 382, 89–103.
Reale, R.A., and Imig, T.J. (1980). Tonotopic organization in auditory cortex of the cat. J Comp Neurol 192, 265–291.
Reddy, L., and Kanwisher, N. (2006). Coding of visual objects in the ventral stream. Curr Opin Neurobiol 16, 408–414.
Reed, P., Howell, P., Sackin, S., Pizzimenti, L., and Rosen, S. (2003). Speech perception in rats: use of duration and rise time cues in labeling of affricate/fricative sounds. J Exp Anal Behav 80, 205–215.
Romani, G.L., Williamson, S.J., and Kaufman, L. (1982). Tonotopic organization of the human auditory cortex. Science 216, 1339–1340.
Romanski, L.M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P.S., and Rauschecker, J.P. (1999). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. Nat Neurosci 2, 1131–1136.
Russ, B.E., Ackelson, A.L., Baker, A.E., and Cohen, Y.E. (2008). Coding of auditory-stimulus identity in the auditory non-spatial processing stream. J Neurophysiol 99, 87–95.
Russ, B.E., Lee, Y.S., and Cohen, Y.E. (2007). Neural and behavioral correlates of auditory categorization. Hear Res 229, 204–212.
Rutkowski, R.G., Shackleton, T.M., Schnupp, J.W.H., Wallace, M.N., and Palmer, A.R. (2002). Spectrotemporal receptive field properties of single units in the primary, dorsocaudal and ventrorostral auditory cortex of the guinea pig. Audiology and Neurotology 7, 214–227.
Sadagopan, S., and Wang, X. (2008). Level invariant representation of sounds by populations of neurons in primary auditory cortex. Journal of Neuroscience 28, 3415–3426.
Sakurai, Y. (1994). Involvement of auditory cortical and hippocampal neurons in auditory working memory and reference memory in the rat. J Neurosci 14, 2606–2623.
Sakurai, Y. (1996). Hippocampal and neocortical cell assemblies encode memory processes for different types of stimuli in the rat. J Neurosci 16, 2809–2819.
Sakurai, Y. (2002). Coding of auditory temporal and pitch information by hippocampal individual cells and cell assemblies in the rat. Neuroscience 115, 1153–1163.
Sally, S.L., and Kelly, J.B. (1988). Organization of auditory cortex in the albino rat: sound frequency. J Neurophysiol 59, 1627–1638.
Schmitzer-Torbert, N., Jackson, J., Henze, D., Harris, K., and Redish, A.D. (2005). Quantitative measures of cluster quality for use in extracellular recordings. Neuroscience 131, 1–11.
Schnupp, J.W., and King, A.J. (2001). Neural processing: the logic of multiplication in single neurons. Curr Biol 11, R640–R642.
Schnupp, J.W., Mrsic-Flogel, T.D., and King, A.J. (2001). Linear processing of spatial cues in primary auditory cortex. Nature 414, 200–204.
Schreiner, C.E., and Mendelson, J.R. (1990). Functional topography of cat primary auditory cortex: distribution of integrated excitation. J Neurophysiol 64, 1442–1459.
Segal, M., Disterhoft, J.F., and Olds, J. (1972). Hippocampal Unit Activity during Classical Aversive and Appetitive Conditioning. Science 175, 792–794.
Shannon, C.E. (1949). Communication in the presence of noise. Proceedings of the IRE 37, 10–21.
Shepherd, W.T. (1911). The discrimination of articulate sounds by raccoons. The American Journal of Psychology 22, 116–119.

Shepherd, W.T. (1912). The discrimination of articulate sounds by cats. The American Journal of Psychology 23, 461–463.
Shestakova, A., Brattico, E., Soloviev, A., Klucharev, V., & Huotilainen, M. (2004). Orderly cortical representation of vowel classes presented by multiple exemplars. Brain Research. Cognitive Brain Research, 21(3), 342–350.
Sinnott, J., and Brown, C. (1998). Effects of syllable duration on stop-glide identification in syllable-initial and syllable-final position by humans and monkeys. Perception & Psychophysics 60 (6), 1032-1043
Sinnott, J.M. (1985). Frequency and intensity discrimination in humans and monkeys. J Acoust Soc Am 78, 1977.
Sinnott, J.M. (1991). Differential sensitivity to vowel continua in Old World monkeys (Macaca) and humans. J Acoust Soc Am 89, 2421.
Sinnott, J.M., and Brown, C.H. (1997). Perception of the American English liquid /ra-la/ contrast by humans and monkeys. J Acoust Soc Am 102, 588–602.
Sinnott, J.M., Sachs, M.B., and Hienz, R.D. (1980). Aspects of frequency discrimination in passerine birds and pigeons. J Comp Physiol Psychol 94, 401–415.
Skaggs, W.E., McNaughton, B.L., Wilson, M.A., and Barnes, C.A. (1996). Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. Hippocampus 6, 149–172.
Sloan, A.M., Dodd, O.T., and Rennaker, R.L. (2009). Frequency discrimination in rats measured with tone-step stimuli and discrete pure tones. Hear Res 251, 60–69.
Sokal, R.R., and Rohlf, F.J. (1995). Biometry (W. H. Freeman).
Sommers, M.S. (1992). Formant frequency discrimination by Japanese macaques (Macaca fuscata). J Acoust Soc Am 91, 3499.
Stiebler, I., Neulist, R., Fichtel, I., and Ehret, G. (1997). The auditory cortex of the house mouse: left-right differences, tonotopic organization and quantitative analysis of frequency representation. J Comp Physiol A 181, 559–571.
Suga, N. (1965). Functional properties of auditory neurones in the cortex of echo-locating bats. J Physiol (Lond) 181, 671–700.
Suga, N., and Jen, P.H.S. (1976). Disproportionate Tonotopic Representation for Processing CF-FM Sonar Signals in the Mustache Bat Auditory Cortex. Science, New Series 194, 542–544.
Sugimoto, S., Sakurada, M., Horikawa, J., and Taniguchi, I. (1997). The columnar and layer-specific response properties of neurons in the primary auditory cortex of Mongolian gerbils. Hear Res 112, 175–185.
Sutter, M.L., and Schreiner, C.E. (1991). Physiology and topography of neurons with multipeaked tuning curves in cat primary auditory cortex. J Neurophysiol 65, 1207–1226.
Syka, J., Rybalko, N., Brožek, G., and Jilek, M. (1996). Auditory frequency and intensity discrimination in pigmented rats. Hear Res 100, 107–113.
Szymanski, F.D., Garcia-Lazaro, J.A., and Schnupp, J.W.H. (2009). Current source density profiles of stimulus-specific adaptation in rat auditory cortex. J Neurophysiol 102, 1483–1490.
Takahashi, H., Nakao, M., and Kaga, K. (2004). Cortical mapping of auditory-evoked offset responses in rats. Neuroreport 15, 1565.
Talwar, S.K., and Gerstein, G.L. (1998). Auditory frequency discrimination in the white rat. Hear Res 126, 135–150.
Theunissen, F.E., and Shaevitz, S.S. (2006). Auditory processing of vocal sounds in birds. Curr Opin Neurobiol 16, 400–407.
Theunissen, F.E., Sen, K., and Doupe, A.J. (2000). Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. Journal of Neuroscience 20, 2315–2331.
Thomas, H., Tillein, J., Heil, P., and Scheich, H. (2006). Functional organization of auditory cortex in the Mongolian gerbil (Meriones unguiculatus). I. Electrophysiological mapping of frequency representation and distinction of fields. European Journal of Neuroscience 5, 882–897.

Thompson, R.F. (1958). Primary stimulus generalization as a function of acquisition level in the cat. J Comp Physiol Psychol 51, 601–606.
Thompson, R.F. (1960). Function of auditory cortex of cat in frequency discrimination. J Neurophysiol 23, 321–334.
Tian, B., Reser, D., Durham, A., Kustov, A., and Rauschecker, J.P. (2001). Functional specialization in rhesus monkey auditory cortex. Science 292, 290–293.
Toro, J.M., Trobalon, J.B., and Sebastián-Gallés, N. (2003). The use of prosodic cues in language discrimination tasks by rats. Anim Cogn 6, 131–136.
Tunturi, A.R. (1950). Physiological determination of the arrangement of the afferent connections to the middle ectosylvian auditory area in the dog. Am. J. Physiol. 162, 489–502.
VanRullen, R., & Thorpe, S. J. (2001). The time course of visual processing: from early perception to decision-making. Journal of cognitive neuroscience, 13(4), 454–461.
Vertes, R.P. (2006). Interactions among the medial prefrontal cortex, hippocampus and midline thalamus in emotional and cognitive processing in the rat. Neuroscience 142, 1–20.
Vigneswaran, G., Kraskov, A., and Lemon, R.N. (2011). Large identified pyramidal cells in macaque motor and premotor cortex exhibit “thin spikes”: implications for cell type classification. Journal of Neuroscience 31, 14235–14242.
Volkov, I.O., and Galazjuk, A.V. (1991). Formation of spike response to sound tones in cat auditory cortex neurons: interaction of excitatory and inhibitory effects. Neuroscience 43, 307–321.
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, 135–152.
Walker, K.M.M., Bizley, J.K., King, A.J., and Schnupp, J.W.H. (2011). Multiplexed and robust representations of sound features in auditory cortex. Journal of Neuroscience 31, 14565–14576.
Wang, L., Narayan, R., Graña, G., Shamir, M., and Sena, K. (2007). Cortical discrimination of complex natural stimuli: can single neurons match behavior? Journal of Neuroscience 27, 582–589.
Warren, R.M. (2008). Auditory Perception (Cambridge University Press).
Weinberger, N.M. (2004). Specific long-term memory traces in primary auditory cortex. Nat Rev Neurosci 5, 279–290.
Wertheimer, M. (1924) Über Gestalttheorie [an address before the Kant Society, Berlin, '7th December, 1924], Erlangen, 1925. In the translation by Willis D. Ellis published in his "Source Book of Gestalt Psychology," New York: Harcourt, Brace and Co, 1938.
Wessel, D.L. (1979). Timbre space as a musical control structure. Computer Music Journal 3, 45–52.
Wiebe, S.P., and Staubli, U.V. (2001). Recognition memory correlates of hippocampal theta cells. Journal of Neuroscience 21, 3955–3967.
Wiener, S.I., Paul, C.A., and Eichenbaum, H. (1989). Spatial and behavioral correlates of hippocampal neuronal activity. J Neurosci 9, 2737–2763.
Wier, C.C., Jesteadt, W., and Green, D.M. (1977). Frequency discrimination as a function of frequency and sensation level. J Acoust Soc Am 61, 178–184.
Wiesel, T.N., and Hubel, D.H. (1963). Single-cell responses in striate cortex of kittens deprived of vision in one eye. J Neurophysiol 26, 1003–1017.
Wills, T.J. (2005). Attractor Dynamics in the Hippocampal Representation of the Local Environment. Science 308, 873–876.
Wilson, M.A., and McNaughton, B.L. (1993). Dynamics of the hippocampal ensemble code for space. Science 261, 1055–1058.
Wood, E.R., Dudchenko, P.A., and Eichenbaum, H. (1999). The global record of memory in hippocampal neuronal activity. Nature 397, 613–616.
Woolsey, C.N., and Walzl, E.M. (1944). Cortical Representation of the Cochlea in the Cat. The Journal of Nervous and Mental Disease *100*, 82.