COMPLEX TONE PROCESSING IN THE PRIMATE BRAIN:
BEHAVIORAL AND SINGLE UNIT EXPERIMENTS

by

ROGER WILLIAM WARD TOMLINSON

B.Sc., The University of Toronto, 1981
M.Sc., The University of Toronto, 1983

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS OF THE DEGREE OF
DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES
(Department of Physiology)

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

January 1989

© Roger William Ward Tomlinson, 1989
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of **Physiology**

The University of British Columbia  
1956 Main Mall  
Vancouver, Canada  
V6T 1Y3  

Date  **JAN 24 1989**
ABSTRACT

The mechanisms of complex tone processing important in pitch perception were investigated at the behavioral, neurophysiological and theoretical level in a nonhuman primate, the rhesus monkey (*Macaca mulatta*).

Four rhesus monkeys were trained to press a button when the fundamental frequencies (missing or present) of two complex tones in a tone pair matched. Both tones were based on a five component harmonic series. Zero to three of the lowest components could be missing in the first tone, while the second (comparison) tone contained all five harmonics. The range of fundamentals tested varied from 200 to 600 Hz. Three monkeys learned to match tones missing their fundamentals to comparison harmonic complexes with the same pitch whereas the fourth monkey required the physical presence of the fundamental. Consideration of several cues available to the monkeys suggests that the animals could perceive the missing fundamental.

The responses of 476 single units in auditory cortex of three alert rhesus monkeys to pure tones and harmonic complex tones were surveyed. Several neuron classes were identified, although responses varied over a continuum. "Filter" neurons had no inhibitory sidebands and responded well when any component of a complex tone entered its pure tone receptive field. "Resolver" neurons had narrow tuning when compared with filter neurons and had a frequency selectivity which was sufficient to resolve peaks within harmonic complex tone spectra. This frequency selectivity occurred over a limited dynamic range. "Fundamental" neurons exhibited similar tuning to pure tones and complex tones with their fundamentals. When a complex tone series without its fundamental was presented, these neurons did not respond, except when the physically present components entered its pure tone receptive field. This selectivity was caused by a powerful lower inhibitory sideband. "Wide band" neurons responded to complex tones and noise but had comparatively weak pure tone responses. "Narrow band" neurons responded selectively to
pure tone stimuli, and poorly, if at all to complex tones and noise. The hypothetical function of arrays of these neurons was discussed.

The responses of hypothetical neurons to pure and complex tones was simulated using linear Gaussian filters to represent excitatory and inhibitory receptive fields. The effects on the neural response, of varying parameters of bandwidth, center frequency, and relative strength of inhibition versus excitation were tested. The behavior of each of the neuron types of the physiological experiment could be explained using various combinations of upper and lower inhibitory sidebands. It was found that the responses of filter neurons could be simulated using relatively broad excitatory receptive fields. Resolver neurons could be modelled with narrow (an octave or less in bandwidth) excitatory field, optionally flanked by narrow inhibitory sidebands, sharpening the filters. Fundamental neurons were simulated by the use of a powerful low frequency sideband. The response of narrow band neurons required powerful upper and lower inhibitory sidebands.

Properties of large groups of neurons were simulated computationally to investigate the effectiveness of non-topographic representations in processing pitch and representing complex tones. Two neural networks, consisting of 640 units arranged in three layers, were trained to represent excitation patterns of harmonic complex tones with fundamental frequencies from 100 to 6500 Hz. The networks were trained using the generalized delta-rule for the back-propagation of error. It was demonstrated that tonotopy and bandlimited responses for individual units are not necessary for processing complex tones. One of the networks was trained to make associations necessary to perform a pitch matching task. When tested with excitation patterns of two tone complexes and missing fundamental complexes it was found that the most important feature of the input for determining the pattern of the output was the low frequency edge, or lowest component of the input pattern.
ACKNOWLEDGEMENTS

I wish to express my gratitude to the following individuals who were involved in the course of my Ph.D. studies:

To my supervisor, Dr. D.W.F. Schwarz, for sharing his expertise in the field of otoneurology and for the years of continuous guidance and support he has given me.

To Mrs. E. To and Miss M. Lee for their expert technical and artistic assistance.

To Mr. W. Treurniet at the Communications Research Center, Department of Communications in Ottawa for making available his personal expertise, and for allowing the use and modification of his neural network software and VAX computer.

To Dr. F.E. Doujak, Dr. D.D. Greenwood, Dr. J.A. Manley-Toronchuk, and Miss H. Becker for their constructive and insightful comments on earlier drafts of this thesis.

To my family and friends, who have provided their emotional support throughout my years of study.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title page</td>
<td>i</td>
</tr>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>iv</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>v</td>
</tr>
<tr>
<td>List of Figures</td>
<td>viii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xi</td>
</tr>
<tr>
<td>Preface</td>
<td>1</td>
</tr>
<tr>
<td>Chapter 1</td>
<td>3</td>
</tr>
</tbody>
</table>

## INTRODUCTION

- Psychology and Physics of Tone Perception: 3
  - Definitions 3
  - Nonlinear Distortion Arising in the Cochlea 7
  - Human Psychophysics 7
  - Animal Psychophysics 10

- Auditory Neurophysiology and Anatomy Relevant to Pitch Perception: 15
  - Brainstem and Midbrain 15
  - Thalamus 16
  - Cortex 17
  - Cortical Single Unit Data 24

- Theory and Models for Complex Tone Perception: Place versus Periodicity Coding. 27

## Chapter 2

### MISSING FUNDAMENTAL PERCEPTION IN RHESUS MONKEYS

- INTRODUCTION 30
- METHODS 31
  - Subjects 31
  - Stimulus Generation 31
  - Training 32
  - Experiment 33
- RESULTS 37
- DISCUSSION 44
## Chapter 3

**RESPONSES OF SINGLE CORTICAL NEURONS TO HARMONIC COMPLEX TONES**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRODUCTION</td>
<td>53</td>
</tr>
<tr>
<td>METHODS</td>
<td>54</td>
</tr>
<tr>
<td>Surgical Procedures</td>
<td>57</td>
</tr>
<tr>
<td>Recording Procedures</td>
<td>57</td>
</tr>
<tr>
<td>RESULTS</td>
<td>59</td>
</tr>
<tr>
<td>Histology</td>
<td>59</td>
</tr>
<tr>
<td>Electrophysiology</td>
<td>72</td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>99</td>
</tr>
<tr>
<td>Neuron Classes</td>
<td>99</td>
</tr>
<tr>
<td>Filter neurons</td>
<td>99</td>
</tr>
<tr>
<td>Resolver neurons</td>
<td>106</td>
</tr>
<tr>
<td>Fundamental neurons</td>
<td>109</td>
</tr>
<tr>
<td>Wide band neurons</td>
<td>110</td>
</tr>
<tr>
<td>Narrow band neurons</td>
<td>111</td>
</tr>
<tr>
<td>Other responses</td>
<td>112</td>
</tr>
<tr>
<td>Site of Inhibitory Interaction</td>
<td>114</td>
</tr>
<tr>
<td>Off Responses in Cortical Neurons</td>
<td>115</td>
</tr>
<tr>
<td>Complex Tone Processing By Auditory Cortex</td>
<td>115</td>
</tr>
<tr>
<td>Are There Neurons Selective For Harmonic Complexes?</td>
<td>116</td>
</tr>
<tr>
<td>Topographic Representation of Pure Tones in Alert Cortex</td>
<td>120</td>
</tr>
<tr>
<td>Topographic Representation of Complex Tones and Pitch in Alert Cortex</td>
<td>120</td>
</tr>
<tr>
<td>Summary: Complex Tone Processing by Cortical Neurons</td>
<td>122</td>
</tr>
</tbody>
</table>

## Chapter 4

**EFFECTS OF INHIBITORY SIDEBANDS ON THE NEURAL RESPONSES TO COMPLEX TONES.**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRODUCTION</td>
<td>123</td>
</tr>
<tr>
<td>METHODS</td>
<td>124</td>
</tr>
<tr>
<td>RESULTS AND DISCUSSION</td>
<td>125</td>
</tr>
<tr>
<td>Summary</td>
<td>140</td>
</tr>
</tbody>
</table>
Chapter 5
THE INTERNAL REPRESENTATION AND PROCESSING OF HARMONIC COMPLEX TONES IN PARALLEL DISTRIBUTED PROCESSING NETWORKS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRODUCTION</td>
<td>141</td>
</tr>
<tr>
<td>METHODS</td>
<td>142</td>
</tr>
<tr>
<td>Simulating Parallel Distributed Processing Networks</td>
<td>142</td>
</tr>
<tr>
<td>Excitation Patterns Used For Inputs and Target Outputs</td>
<td>147</td>
</tr>
<tr>
<td>Two Simulation Paradigms</td>
<td>148</td>
</tr>
<tr>
<td>Autoassociation</td>
<td>149</td>
</tr>
<tr>
<td>Pitch association</td>
<td>149</td>
</tr>
<tr>
<td>RESULTS</td>
<td>150</td>
</tr>
<tr>
<td>Autoassociation</td>
<td>150</td>
</tr>
<tr>
<td>Pitch Association</td>
<td>151</td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>164</td>
</tr>
<tr>
<td>Autoassociation</td>
<td>164</td>
</tr>
<tr>
<td>Pitch association</td>
<td>165</td>
</tr>
<tr>
<td>Summary</td>
<td>167</td>
</tr>
</tbody>
</table>

Chapter 6
SUMMARY AND CONCLUSIONS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>FUTURE STRATEGIES AND QUESTIONS</td>
<td>172</td>
</tr>
<tr>
<td>Complex Tone Perception in Monkeys</td>
<td>172</td>
</tr>
<tr>
<td>Temporal Processing of Complex Sound Stimuli</td>
<td>173</td>
</tr>
<tr>
<td>Internal Representation of Complex Sound in Cortex</td>
<td>174</td>
</tr>
</tbody>
</table>

References                                      | 175  |

Appendix                                       | 194  |

THE RESPONSE OF A NETWORK, WHICH HAS BEEN TRAINED WITH AN AUTOASSOCIATION PARADIGM, TO NOVEL INPUTS
## LIST OF FIGURES

### Chapter 1

<table>
<thead>
<tr>
<th>Fig.</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pure and complex tone waveforms and spectra</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>Diagram of mammalian auditory central nervous system</td>
<td>13</td>
</tr>
<tr>
<td>3</td>
<td>Maps of cat auditory cortex</td>
<td>19</td>
</tr>
<tr>
<td>4</td>
<td>Maps of rhesus monkey auditory cortex</td>
<td>21</td>
</tr>
</tbody>
</table>

### Chapter 2

<table>
<thead>
<tr>
<th>Fig.</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>The behavioral paradigm</td>
<td>34</td>
</tr>
<tr>
<td>6</td>
<td>Complex tone pitch matching in monkeys</td>
<td>40</td>
</tr>
<tr>
<td>7</td>
<td>Predicted response of component matching strategy</td>
<td>47</td>
</tr>
<tr>
<td>8</td>
<td>The use of an absolute frequency strategy to complete the experimental task.</td>
<td>49</td>
</tr>
</tbody>
</table>

### Chapter 3

<table>
<thead>
<tr>
<th>Fig.</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>Diagram of recording electrode’s approach to auditory cortex.</td>
<td>55</td>
</tr>
<tr>
<td>10</td>
<td>Distribution of characteristic frequencies observed in auditory cortex.</td>
<td>60</td>
</tr>
<tr>
<td>11</td>
<td>Maps of neuron types and best frequencies found in auditory cortex of subject M2</td>
<td>62</td>
</tr>
<tr>
<td>12</td>
<td>Maps of neuron types and best frequencies found in auditory cortex of subject M3</td>
<td>65</td>
</tr>
<tr>
<td>13</td>
<td>Maps of neuron types and best frequencies found in auditory cortex of subject M4</td>
<td>68</td>
</tr>
<tr>
<td>14</td>
<td>Depth of units in cortex</td>
<td>76</td>
</tr>
<tr>
<td>15</td>
<td>Filter neuron responses</td>
<td>79</td>
</tr>
<tr>
<td>16</td>
<td>Resolver neuron responses</td>
<td>81</td>
</tr>
</tbody>
</table>
### Chapter 3 (cont)

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fig. 17</td>
<td>Fundamental neuron responses</td>
<td>88</td>
</tr>
<tr>
<td>Fig. 18</td>
<td>Wide band neuron &quot;off&quot; responses</td>
<td>91</td>
</tr>
<tr>
<td>Fig. 19</td>
<td>Narrow band neuron responses</td>
<td>93</td>
</tr>
<tr>
<td>Fig. 20</td>
<td>Unclassified neuron responses</td>
<td>95</td>
</tr>
<tr>
<td>Fig. 21</td>
<td>Unclassified neuron responses</td>
<td>97</td>
</tr>
<tr>
<td>Fig. 22</td>
<td>Predictions of complex tone response from pure tone response</td>
<td>100</td>
</tr>
<tr>
<td>Fig. 23</td>
<td>Resolver neuron response to eight component harmonic complexes</td>
<td>102</td>
</tr>
<tr>
<td>Fig. 24</td>
<td>A plot of the peak frequencies in histograms of resolver neurons compared with subharmonics of the best frequencies.</td>
<td>104</td>
</tr>
</tbody>
</table>

### Chapter 4

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fig. 25</td>
<td>Response of Gaussian filters of varying bandwidths to complex tones</td>
<td>127</td>
</tr>
<tr>
<td>Fig. 26</td>
<td>Responses of a filter combination consisting of a narrow excitatory band with lower inhibitory sidebands</td>
<td>130</td>
</tr>
<tr>
<td>Fig. 27</td>
<td>Responses of a filter combination consisting of a broad excitatory band with lower inhibitory sidebands</td>
<td>132</td>
</tr>
<tr>
<td>Fig. 28</td>
<td>Responses of a filter combination consisting of an excitatory band with upper inhibitory sidebands</td>
<td>134</td>
</tr>
<tr>
<td>Fig. 29</td>
<td>Responses of a filter combination consisting of an excitatory band with both upper and lower inhibitory sidebands</td>
<td>136</td>
</tr>
</tbody>
</table>
Chapter 5

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fig. 30</td>
<td>Diagram of neural network used in the simulations</td>
<td>143</td>
</tr>
<tr>
<td>Fig. 31</td>
<td>Results of autoassociation training in a neural network</td>
<td>152</td>
</tr>
<tr>
<td>Fig. 32</td>
<td>Responses of hidden units in an autoassociative network</td>
<td>154</td>
</tr>
<tr>
<td>Fig. 33</td>
<td>Response of an autoassociative network to harmonic complex stimuli missing their fundamentals</td>
<td>156</td>
</tr>
<tr>
<td>Fig. 34</td>
<td>Results of pitch association training in a neural network</td>
<td>158</td>
</tr>
<tr>
<td>Fig. 35</td>
<td>Responses of pitch association network to missing fundamental stimuli</td>
<td>160</td>
</tr>
<tr>
<td>Fig. 36</td>
<td>Responses of pitch association network to two tone stimuli</td>
<td>162</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Chapter 2

Table I  Stimulus frequencies for test and comparison tones with 400 Hz fundamentals 36

Table II  Significance of differences in response probabilities between the case of matching fundamentals and other frequency ratios for each spectral pattern 42

Table III  Significance of differences in response probabilities between the case of matching fundamentals and other frequency ratios for each test frequency and pattern 43

Chapter 3

Table IV  Results from linear regressions of the unit CF (dependent variable, logarithmic frequency scale) versus the unit’s antero-posterior coordinate (independent variable, setting on x-y positioner in mm) for each subject. 71

Table V  Summary of criteria for unit classification 74

Table VI  Population and range of characteristic frequencies of functionally classified cortical units 75

Table VII  Maximum number of components and dynamic range of resolution of resolver neurons. 83

Table VIII  The distribution of characteristic frequency of resolver neurons. 84
Preface

This thesis describes various aspects of complex tone perception in a non-human primate, the rhesus monkey. These investigations were undertaken to provide a further understanding of the mechanisms underlying pitch perception in humans, which seems to be closely interrelated with the processing of harmonic complex tones (cf. de Boer, 1976).

The sensation of pitch is most salient when one listens to and appreciates music. It also serves more basic functions. The term "intonation" denotes the variation of pitch over a sentence. In "intonation languages" such as English, pitch carries linguistic information, aiding the grammatical segmentation of spoken sentences (Lehiste, 1970). In "tone languages", such as Mandarin Chinese, the pitch contour within a word can play a significant role in phoneme identification and differentiation (McCawley, 1978). Intonation can also carry nonlinguistic information, such as the emotional state of the speaker. Pitch cues are not only important in the processing of language and other vocalizations, but also assist in the identification of environmental sounds.

A complete understanding of how pitch is perceived requires explanation of diverse phenomena. For instance, what is the basis of missing fundamental pitch? How is pitch represented internally? Can the firing of a single neuron uniquely signal which pitch is heard? What mechanisms might evoke a perception of pitch when the peripheral auditory sensory apparatus consists of a hearing prosthesis? This thesis attempts to address some of these questions.

The thesis is divided into six chapters. The first chapter reviews the psychophysics of complex tone perception and tone representations in the auditory central nervous system (CNS). The second chapter describes the behavioral experiments that demonstrate the ability of rhesus monkeys to make pitch judgements.
In the third chapter, the responses of single units in alert rhesus monkey auditory cortex to tonal stimuli with controlled pitches are described and analyzed. In the fourth and fifth chapters, various numerical models are used to test the principles of complex tone processing in neuronal systems. The final chapter summarizes the data and suggests avenues for further research.
"You don't know how deep a puddle is until you step in it".

-Anon.

CHAPTER 1

INTRODUCTION

Psychology and physics of tone perception.

Definitions: Complex tones are comprised of "pure tones". Pure tones are sinusoidal oscillations in air pressure completely characterized by their period (P), amplitude (A) and phase (Ph) and containing only one component in their amplitude and phase spectrum. Whistling or flute sounds are examples of nearly pure tones. The Fourier transforms of complex tones yield line spectra with more than one component at discrete frequencies. Although any combination of pure tones or "partials" constitutes a complex tone, only "harmonic" complex tones will be discussed here. Frequencies within harmonic complexes (Figure 1B and C) are integral multiples of the lowest component, the "fundamental" tone (when the components of a harmonic complex are given numbers here, the fundamental shall be component no. 1). Most tonal sounds in nature are generated by vibrating strings or columns of air and therefore have a harmonic structure.

In mathematical terms, pure and complex tones are periodic functions of time. The period of a pure tone is equal to the inverse of its frequency. (i.e. a pure tone of 100 Hz has a period of 0.01 sec.). Harmonic complex waveforms have a period equal to the inverse of their fundamental frequency (which determines their pitch). This is true even when the fundamental component is not physically present.
FIG. 1 Waveforms and amplitude spectra for pure and complex tones with the same period, P.

(a) left: the sinusoidal waveform of a pure tone; right: the spectrum of a pure tone with one component, with linear axes of intensity and frequency.

(b) left: the waveform of a harmonic complex tone with components 1 to 6 (component 1 is the fundamental). The components are all in sine phase (see text for details). right: the amplitude spectrum of the complex tone, with the same axes and scale as (a).

(c) left: the waveform of a harmonic complex tone with components 1 to 6. The components are in random phase with respect to one another; right: The amplitude spectrum of this complex tone, identical to the spectrum in (b)
WAVEFORM

PURE TONE

6 COMPONENT HARMONIC COMPLEX

SPECTRUM
Each component of a complex tone has its own amplitude and phase. The complex tones in Fig. 1 all have components of equal intensity. When all of the components start out with their phase equal to zero, the complex is said to be in sine phase (Fig. 1b). When the phases of the components are initially randomly distributed, the resulting amplitude spectrum, period, and pitch of the tone are identical to the sine phase version but the waveform differs (Fig. 1c). Noise stimuli are aperiodic and have continuous spectra. White noise has an equally distributed average amplitude over the audible frequency range.

Pitch is, according to the American Standards Association (1960), "that attribute of auditory sensation in terms of which sounds can be ordered on a musical scale". Pitch remains the same when a given note is played on different musical instruments, the differences are "timbres" of the sound. If different complex tones are compared, it is sometimes difficult to decide between timbre and pitch differences. Timbre can be ordered along many dimensions, whereas pitch is conventionally organized along one dimension, given by comparison with pure tones. Pure tone pitch varies monotonically with frequency and provides the reference scale for complex tone pitch. Pure tone pitch can shift slightly with amplitude, therefore its intensity should also be specified when making comparisons. In pitch matching experiments subjects generally equate pitches of harmonic complex tones with pure tones of the fundamental frequency (de Boer, 1976).

Some discrepancy exists as to what constitutes a "musical pitch". Certain sounds with periodicities below 200-300 Hz are said to have a "rate pitch" (Loeb et al., 1983) where perceptual continuity is interrupted. Missing fundamental pitch exists for fundamentals above 50 Hz (Ritsma, 1962). In occidental music, tones with fundamental frequencies above 5000 Hz are rarely found and are generally regarded as having ill-defined pitches. For the present discussion, musical pitch is assumed to exist over the fundamental frequency range between 50 and 5000 Hz.
Nonlinear distortion arising in the cochlea: When a pure tone is passed through a linear system, only its amplitude may change. In a non-linear system though, distortion can result in new frequencies being present in the output which were not present in the input. "Harmonic distortion" as the name implies, introduces new frequencies at integral multiples of the original tone. Products of "intermodulation distortion" arise when two frequencies interact, producing a series of new tones at various sums and differences of the frequencies of the original two tones. The new frequencies are called combination tones and some may be predicted by the formulas:

\[ f_k = f_1 \cdot k(f_2 - f_1) \]
\[ f_k = k(f_2 - f_1) \]

where \( k \) is an integer, and \( f_1 \) and \( f_2 \) are the frequencies of the primary tones, such that \( f_2 \) is greater than \( f_1 \). When experimenting with harmonic complex tones with missing fundamentals, the new frequencies are important because, nonlinear distortion products may occur at the fundamental frequency. Empirically, the most prominent of the distortion tones are referred to as the "cubic difference tone" (\( 2f_1 - f_2 \)) and the "simple difference tone" (\( f_2 - f_1 \)). Under most conditions, though by no means all, the perceptibility of combination tones is poor for primary tone amplitudes less than 50 dB sound pressure level (SPL) (Plomp, 1965; Smoorenburg, 1972).

Human psychophysics: In 1843, Ohm applied the then new Fourier transform to the analysis of sound waveforms and established that a tone with the period \( P \) may be, and usually is, composed of a number of Fourier components which can be perceived as either one unified sound, or as a number of simultaneous simpler tones. Helmholtz in 1867, drawing on Ohm's work, suggested a biophysical mechanism that supported his theory of music perception. Helmholtz performed a series of experiments using resonators to verify the physical existence of the Fourier components. Although the lower partials were easily discriminated, increased difficulty was experienced with distinguishing the higher partials.
However, with the aid of mechanical resonators, he managed to observe partials up to the 16th overtone. Helmholtz put forth the idea that the ear is composed of an array of tiny resonators which analyze sound into its component frequencies; a particular frequency would excite a specific region in the cochlea and thus provide the brain with a frequency representation of sound. He proposed that the pitch of a tone was determined by the lowest frequency (the fundamental) and that the timbre was determined by the overtones.

In 1843, before the time of von Helmholtz, Seebeck performed experiments on pitch with acoustic sirens. An acoustic siren is a disk with holes punched in it. A pipe carrying a stream of pressurized air is positioned over the holes. By spinning a disk with equally spaced holes at a constant rate, one can produce periodic puffs of air, creating a tone with a rich collection of overtones. By setting the holes alternately closer and farther apart, Seebeck managed to produce tones without fundamental components, the lowest component being the second Fourier component. The pitch produced by this stimulus remained the same as the stimulus with its fundamental. Therefore he argued that the period of the stimulus and not the lowest frequency component, determined the pitch of his complex tones. These results were dismissed as erroneous by his contemporaries since the tones produced by the siren were fairly weak and contained significant amounts of noise. Later, Helmholz argued that tones without fundamentals will be transformed in the ear by intermodulation distortion into full complexes with fundamentals, thus accounting for the pitch of Seebeck’s stimulus. Ohm’s Law would still account for Seebeck’s results.

In the 1920's, interest in the phenomenon of missing fundamental pitch arose again, owing to the newly emerging communications industry. Telephone circuits carry the frequencies between about 300 Hz and 3000 Hz. The fundamentals of male voices typically have lower frequencies of around 120 Hz, and are effectively filtered out over the telephone. It is a common experience that the pitch of voices over the telephone appears normal even at low intensities at which cochlear distortion is negligible. Thus "the case of the missing
fundamental" (Fletcher, 1929) aroused new attention. However, the existence of distortion products, whether in the ear or in the telephone itself could not be determined experimentally, and interest declined once more.

The advent of electronic amplifiers and better sound transducers enabled Schouten in 1940 to generate more controlled sounds with improved signal-to-noise ratios. Using an optical version of Seebeck's siren, Schouten could produce controlled harmonic complex tones with fundamental frequencies either present or absent. He distinguished two kinds of pitches at the fundamental frequency: the first was produced by the physically present fundamental frequency itself, and "something else" produced by the remaining overtones, which he termed the "residue pitch". The periods of the fundamental frequency alone, and a set of overtones without the fundamental are identical. Since the residue pitch did not depend upon the lowest Fourier component, he concluded that the pitch was dependent upon the period of the complex tone.

Licklider (1954) demonstrated that non-linear distortion products could not account for the perception of the missing fundamental. He added a band of low frequency noise which masked the frequency region containing the fundamental. Although the fundamental frequency was masked, the residue pitch associated with the overtones remained unaffected by the noise.

Ritsma, a student of Schouten, in 1962 investigated conditions under which residue pitch could exist. Complex tones were generated employing amplitude modulated sine waves. When the amplitude of a sine wave carrier (frequency $f_c$) is modulated by another sine wave (frequency $f_m$), a three component stimulus is produced, with the following frequencies: $f_c$, $f_c + f_m$, and $f_c - f_m$. He set the frequencies $f_c$ and $f_m$ to obtain any three successive components of a desired complex without fundamental (e.g. to produce components at 1200, 1400, and 1600 Hz for the fundamental of 200 Hz: $f_c = 1400$ Hz, and $f_m = 200$ Hz). Ritsma found that residue pitch exists when:
1) the complex tones have fundamentals between 50 and 800 Hz.

2) component frequencies do not exceed 5000 Hz.

3) partials do not exceed the 20\textsuperscript{th} harmonic.

Plomp in 1967 determined which partials were most important for the sensation of missing fundamental pitch. He found that components 3, 4 and 5 from one harmonic series would dominate over other components from other simultaneously presented series. The dominant partials (3, 4, 5) could be 10-15 dB weaker than others and still be effective in determining the pitch. A minimum of three components from the "dominant region" are required for their fundamental to be perceived over fundamentals from other components. The most surprising finding was that the fundamental itself is not a dominant component.

Houtsma and Goldstein (1972) performed a series of experiments with stimuli consisting of 2-component complex tones to elicit the sensation of the missing fundamental. Since with only two components this percept is relatively weak, they made it more salient by presenting their stimuli such that the pitches of the absent fundamentals formed melodies. The component pairs were randomly selected for each note, subject to the constraint that the components be adjacent (i.e. components 3,4 versus 6,7, etc.); perception of the melody thus hinged on the perception of the missing fundamentals. Both tones were presented together in both ears, with the result that the subjects successfully perceived the melodies. Subsequently, one component of the complexes was presented to each ear and the subjects were still able to perceive the melodies. This experiment conclusively showed that interaction of the overtones in the cochlea is not required for the percept of the missing fundamental and that the percept can be reconstructed within the central nervous system.

Animal Psychophysics: To investigate the neurophysiological aspects of perception, work with animals is needed. A prerequisite to this is some understanding of the perceptual capabilities of the experimental animals. The first requirement for selecting an animal is that
it be able to hear in the frequency range of musical pitch in man, between ca. 50 and 5000 Hz. The frequency range of an animal's vocalizations may be a particularly useful guide, and in addition, the usually harmonic structure of vocalizations might suggest that the animal can process harmonic complexes. A sensitivity to human speech (following spoken commands) could also suggest the same capacities. Most small animals such as the guinea-pig and rat are not appropriate since their hearing is poor below 1000 Hz (Kelly and Masterton, 1977; Dallos, 1970). The monkey's hearing range extends from human low frequency limits to an upper limit of 35,000 to 40,000 Hz (Stebbins et al., 1966). The cat is similar in this regard (Miller et al., 1963). Both of these species use harmonically structured sounds for communication.

Many birds can imitate human speech. The Mynah bird is quite notable in this respect (Klatt and Stefanski, 1974). Their natural vocalizations (birdsong) are regarded as having a musical character. Hulse and Cynx (1985) tested the ability of the European Starling (Sturnus vulgaris), a mimicking species of song bird, to distinguish ascending and descending tone sequences on the basis of pitch. The birds could successfully make the distinction, provided the testing and training frequency range were similar. They displayed difficulty in generalizing the task to other frequency ranges, however. When testing stimuli which were far outside the frequency range of the training set, the bird's ability to make the discriminations fell rapidly. In a preliminary report in 1986, Cynx also reported on the ability of the Starling to discriminate between two complex tones with missing fundamentals. The birds made the same responses when components were changed, leaving the fundamentals the same.

Although cats are not renowned for their ability to follow spoken commands, their own vocalizations can possess richly harmonic and tonal qualities (Brown et al., 1977). In 1976, Heffner and Whitfield published a report on the ability of cats to perceive the missing fundamental. They trained their animals to discriminate between tone pairs of ascending or
descending pitch. After a stimulus with one of these directions the animal received a mild
shock; the other direction was safe and the animal received a reward. During training both
the absolute component frequencies and the fundamentals were changed in the same
direction. In the test phase, the fundamental frequency and absolute frequencies changed in
opposite directions. The cats responded to the tone pairs in accordance with the direction of
the missing fundamental transition. In a subsequent experiment, Whitfield (1980) trained
cats in the same task and then ablated the primary auditory cortex bilaterally. The cats no
longer retained, and could not relearn, the discrimination. Perception of the missing
fundamental is impaired in human patients with surgical ablation of Heschel’s gyri in the
superior temporal plane (Zatorre, 1988). This implicates the auditory cortex in processing the
missing fundamental, although ablation experiments, in general, cannot localize a complex
perceptual function such as pitch discrimination. The lesioned subjects may still retain the
relevant sensory capacity, but may be unable to express that capacity in the performance of
some motor responses.

Monkeys are relatively easy to train (Rumbaugh and Gill, 1975; Khalsa et al., 1987).
They can discriminate among complex acoustic stimuli such as human vowels (Dewson et al.,
1986; Sinnott and Kreiter, 1986), stop consonants (Waters and Wilson, 1976), and their own
rich repertoire of species-specific vocalizations (Winter et al., 1966 for squirrel monkeys;
Rowell and Hinde, 1962 for rhesus monkeys; Green, 1975 for Japanese macaques) which
makes them ideally suited for auditory research. However, no experiments have yet been
carried out on their ability to perceive pitch in general and missing fundamentals in
particular. A behavioral experiment demonstrating missing fundamental perception in
monkeys is the subject of Chap. 2.
FIG. 2 Diagram of auditory central nervous system:

Abbreviations: MGB, medial geniculate body; IC, inferior colliculus; SOC, superior olivary complex (comprised of LSO, MSO, and MNTB); VCN, ventral cochlear nuclei (made up of AVCN plus PVCN); DCN, dorsal cochlear nucleus; VIII nerve (auditory nerve).
Auditory Neurophysiology and Anatomy Relevant to Pitch Perception.

Chapter 3 describes experiments at the single neuron level on complex tone representation in the cortex of the alert rhesus monkey. The neural responses in the cortex are determined not only by cortical interactions, but also by the processing of the input at lower stations. Fig. 2 schematically illustrates the central auditory pathways.

Information relevant for pitch perception is represented in the auditory system in two ways: 1) There is a spatial representation of frequencies on the basilar membrane known as tonotopy which is preserved at higher levels in topographic projection patterns (Aitkin, 1976); 2) The phase-locking of spikes to waveform peaks also provides a temporally coded representation of acoustic stimuli (Rose et al., 1967; Young and Sachs, 1979) in the auditory nerve discharge pattern.

Brainstem and Midbrain: The first stage of central auditory processing takes place in the cochlear nuclei. The cochlear nuclei are comprised of the dorsal cochlear nucleus (DCN) and the ventral cochlear nucleus, which is subdivided into the anteroventral (AVCN) and the posteroventral (PVCN) cochlear nuclei. Auditory nerve axons form collaterals and terminate in all subdivisions of the cochlear nuclei. In the AVCN, large synaptic boutons, the end bulbs of Held, provide very secure synaptic transmission of the temporal information (Molnar and Pfeiffer, 1968).

The first major convergence point for information from both ears is the superior olivary complex (SOC). The SOC and the DCN project up to the inferior colliculus (IC) via the lateral lemniscus. It is at this point that a distinction between "lemniscal-line systems" and "lemniscal-adjunct systems" can be made. This division of ascending sensory pathways was suggested by Graybiel in 1973, drawing from anatomical data in the somatosensory and auditory pathways. The hypothesis has been extended by physiological data from the auditory system by Irvine and Phillips in 1982. The
lemniscal-line system carries precise sensory-discriminative information along core projections up to the thalamus and sensory cortex maintaining a tight topographic representation at each stage. The lemniscal-adjunct system carries less specific information up pathways which are convergent and polymodal. The adjunct system projects via pathways separate from the lemniscal-line system to the peripheral or "belt" areas of sensory cortex.

There are at least three subnuclei within the inferior colliculus. The central nucleus (ICC) receives a topographic projection from the lower nuclei and contains a tonotopic array of sharply tuned neurons, typical of a lemniscal-line projection (Aitkin et al., 1975). The central nucleus of the inferior colliculus projects to the ventral division of the medial geniculate body (MGB). The external (ICX) and pericentral (ICP) nuclei of the colliculus receive more diffuse projections. ICX receives ascending convergent fibers from the somatosensory modality (Aitkin et al., 1978).

**Thalamus:** The main and "specific relay" nucleus for auditory sensation in the thalamus is the medial geniculate body (MGB). The MGB can be divided into three divisions: ventral, medial and dorsal, which are distinguishable in Nissl stained sections (Morest, 1964).

The lateral part of the ventral division has the characteristics of a leminiscal-line system as it contains sharply tuned neurons in a tonotopic array (Aitkin and Webster, 1972). The ventrolateral part of the ventral division contains cells that are poorly tuned and respond poorly to sound (Calford, 1983) and projects to the secondary auditory cortex (AII) (Morel and Imig, 1987).

The medial division of the MGB is populated with sparsely distributed magnocellular neurons, in addition to a large number of smaller cells. The division contains many cells which are poorly tuned (Morel and Imig, 1987), and projects
diffusely to the secondary cortex. The magnocellular division receives converging input from the somatosensory modality (Wepsic, 1966).

The dorsal division of medial geniculate body can be divided into a superficial and a deep portion. The dorsal division receives projections from ICP (Calford and Aitkin, 1983) and projects diffusely to the field AII (Morel and Imig, 1987). Cells in the dorsal division are poorly tuned and exhibit long latencies to pure tones (Calford, 1983).

Parts of the posterior thalamic nuclei (PO) receive auditory input as well. In the lateral region (PO_l), there are sharply tuned neurons (Phillips and Irvine, 1979) which are arrayed in a tonotopic fashion (Imig and Morel, 1985). This part of the posterior nucleus receives projections from the ICC (Andersen et al., 1980) and projects to both the anterior auditory field (AAF) and the primary auditory cortex in the cat (Imig and Morel, 1987). The intermediate part of the posterior nucleus (PO_i) contains broadly tuned neurons, and it also projects to the primary auditory cortex.

The nucleus medialis dorsalis and the intralaminar nuclei (parafascicular and centromedian) contain cells that are driven by auditory input (Irvine, 1980). The responses in these nuclei are broadly tuned and usually polysensory, with input from visual, somatosensory, and/or auditory modalities. The main efferent projection is thought to terminate in suprasylvian auditory association cortex (e.g. Itoh and Mizumo, 1980), although there is limited evidence for a specific projection.

**Cortex:** Each of the 6 cortical layers has certain specific input and output projection patterns. The main (lemniscal line) thalamic inputs from MGB_v terminate in layer IV and the lower part of layer III (Jones and Burton, 1976). There is a reciprocal connection back to thalamus originating in layer VI (Mitani and Shumokouchi, 1985). Layers II and III project to other cortical fields within the same hemisphere and to the contralateral hemisphere via the corpus callosum (Code and Winer, 1986). Layer V
contains the main efferent projection from the cortex to subcortical structures. Layer V pyramidal cells in auditory cortex project to the inferior colliculus (ICP) (Kelly and Wong, 1981).

Since primary auditory cortex receives a dense projection from the thalamus (MGB), layer IV is greatly enlarged to receive it. This can be seen with both Nissl and myelin stain. Layers II and III are also enlarged as they provide heavy intracortical connections originating from this site. Layer VI projecting back to the thalamus is also dense, however, layer V is relatively light and sparsely populated.

Cortical areas responsive to acoustic stimuli were first electrophysiologically mapped using cortical surface electrodes. Cochlear nerve fibers were stimulated electrically at the exposed spiral osseous lamina (Woolsey and Walzl, 1942), or with pure tones in an intact cochlea (Licklider and Kryter, 1942; McCulloch et al., 1942). The areas which were activated with the shortest latency after stimulation corresponded to primary sensory cortex (koniocortex) identified from cytoarchitectonics. In humans, this region is on Heschel's transverse gyri which are buried within the Sylvian fissure on the superior surface of the temporal lobe. In macaques, it occupies approximately the same position, but in cats, the primary auditory cortex is exposed on the lateral surface. Using evoked potentials, Woolsey and Walzl (1942, 1944) found multiple complete topographic representations of the cochlea on the surface of the monkey and cat auditory cortex. The cortex adjacent to the tonotopic fields responded to auditory stimuli but with longer latencies.
FIG. 3  Maps of cat auditory cortex based on physiological criteria (from Imig and Reale, 1980).  (A) lateral view of cat brain and (B) an exploded view showing the depths of the sulci of auditory cortex with the following fields: primary auditory (AI), secondary auditory (AII), anterior auditory (A or AAF), posterior, (P), ventral, (V), ventroposterior (VP), dorso-posterior (DP) and temporal (T).  The sulci have the following abbreviations: sss, suprasylvian sulcus; aes, anterior ectosylvian sulcus; pes, posterior ectosylvian sulcus; and pss, pseudo-sylvian sulcus.
FIG. 4 Maps of rhesus monkey auditory cortex:

A) Cytoarchitectonic map of temporal cortex. The anterior pole is to the left. Top panel: lateral view of rhesus brain showing lateral temporal fields. Bottom panel: top view of superior temporal plane joined to lateral view of superior temporal gyrus showing cytoarchitectonic subdivisions (from Pandya and Sanides, 1973).

Abbreviations:

- Ts1:
- Ts2:
- Ts3:
- Kam/Kal: medial and lateral koniocortex
- paAr: rostral parakoniocortex
- paAc: caudal parakoniocortex
- paAlt: lateral parakoniocortex
- proA: prokoniocortex
- Pro: proisocortex
- pal: parainsular
- reIpt: retroinsular parietal
- reIt: retroinsular temporal
- Tpt: temporoparietal
- proS: somatic prokoniocortex

B) Physiological map of superior temporal plane of rhesus monkey (from Merzenich and Brugge, 1973). Abbreviations: Primary auditory cortex (AI), rostrolateral field (RL), lateral field (L), Caudomedial field (CM), auditory responsive cortex (a,b,c).
Evoked potentials represent the summed activity of large numbers of neurons over a relatively large cortical area around the electrode. Multi-unit recording with intracortical electrodes allows a higher spatial resolution. One such study (Reale and Imig, 1980) has revealed 5 tonotopic representations in the cortex of the cat with several other regions responding less specifically to auditory stimuli (Fig. 3). In the primary auditory field (AI), frequencies are spatially organized extending from low tones posteriorly to high tones anteriorly. The anterior auditory field (field A), which lies rostral to AI also contains a tonotopic array which is continuous in the high frequency region with AI with its low frequencies represented anteriorly. In addition, the posterior (P) and the ventral posterior (VP) fields also contain complete tonotopic maps. The ventral field (V), field AII, and the temporal and insular fields respond to sound but have a blurred or undetectable tonotopy.

The auditory cortex of the rhesus monkey (Macaca mulatta) was physiologically mapped using pure tone stimuli by Merzenich and Brugge (1973). The macaque’s auditory cortex is buried within the Sylvian fissure (Fig. 4) whereas in the squirrel monkey it extends onto the lateral surface of the superior temporal gyrus. Merzenich and Brugge identified two main tonotopic fields with three other secondary surrounding fields (Fig. 4B). In the primary auditory cortex, low frequencies were represented anteriorly with high frequencies extending into the posterior region. The region rostral and lateral (RL) to this is contiguous with AI in the low frequency region, with its high frequencies appearing anteriorly. The secondary fields, seen on Fig. 4, were centromedian (CM), lateral (L) and anterior (a). The electrophysiological maps were correlated with maps based on cytoarchitecture.

Pandya and Sanides (1973) and later Galaburda and Pandya (1983) investigated the anatomy and projections of the superior temporal plane, and surrounding cortex (see Fig 4A). The extent of anatomically defined primary sensory cortex (koniocortex) roughly parallels the physiological extent of AI. Pandya divided auditory koniocortex (KA) into
medial and lateral parts (Kam, Kalt). Three parakonial fields surround the core primary field laterally (paAlt), rostrally (paAr) and caudally (paAc). A fourth prokoniocortical field (proA) separates the koniocortex from the insular cortex. The four fields (paAc, paAr, paAlt, and proA) comprise the "belt" cortex and form a transition zone between koniocortex and temporal or insular association cortex. A similar, although less elaborate parcellation scheme was proposed by Rose (1949) from a cytoarchitectonic study of the cat auditory cortex.

**Cortical Single Unit Data:** Electrophysiological characterization of auditory cortical neurons was initially carried out in anesthetized cats with click and tone burst stimuli (Erulkar et al., 1956). Although relatively few neurons (about 14%) responded securely to sound, onset responses and the inhibition of spontaneous activity with tonal stimulation were described.

Mountcastle (1957) in his classic study of somatosensory cortex described a distinct columnar organization. The cells located within the same column perpendicular to the surface of cortex have receptive fields located in the same region of the body, and respond to the same submodality (eg. light touch or deep pressure). Similar findings have been reported for the visual cortex where cells in a given column have the same orientation selectivity and ocular dominance (Hubel and Wiesel, 1963). There is, at present, overwhelming evidence that columnar organization is a general attribute of all neocortex (Hubel and Wiesel, 1977).

In 1970 Abeles and Goldstein used paralyzed unanesthetized cats to investigate whether columnar organization is also a property of the auditory cortex. By taking into account the angle of penetration of the track in cortex, they were able to conclude that characteristic frequency was the same within a column normal to the surface of cortex. Imig and Adrian (1977) showed that binaural properties of auditory neurons are also similar when penetrations are normal to the surface of the cortex.
Several studies have been undertaken in anesthetized cats on the tonal response properties of neurons in individual fields of auditory cortex. Cells in both AI and AAF respond with short latency (10-20 ms), have sharp tuning curves and typically have monotonic rate level functions with a 30-40 dB dynamic range (Phillips and Irvine, 1982). Field AII, which is ventral to these tonotopic regions, responded with longer latency and with broad tuning curves, and also had slightly higher thresholds (Schreiner and Cynader, 1984). The posterior field, P, contains neurons with medium to long latencies (20-50 ms) and narrowly tuned receptive fields (Phillips and Orman, 1984). Eighty-six percent of cells in field P have nonmonotonic rate-level functions. In the cat cortex three regions of "non-specific" auditory cortex have been identified, termed the medial suprasylvian area (MSA), the anterolateral area (ALA) and the pericruciate area (PCA). These regions are characterized by cells with broad tuning curves, medium to long latencies (16-54 ms) and which exhibit polysensory convergence (Irvine and Phillips, 1982).

Cortical responses to amplitude modulated pure tones have also been investigated (Phillips and Hall, 1987). Schreiner et al. (1983) and Schreiner and Urbas (1986) systematically studied the responsiveness of cortical neurons to different rates of amplitude modulation. Neurons in several fields (AI, AII, AAF, P and VP) had responses which exhibited a preferred rate of modulation. It was suggested that AAF had the highest degree of selectivity to amplitude modulations. Best modulation frequencies (BMF) in AAF ranged from 2 to 100 Hz, with a median value of about 20 Hz.

Goldstein et al. (1971) suggested that the tonotopic organization of the cortex might reflect the organization of a more complex feature such as periodicity or complex tone pitch. Accordingly, they investigated the responses of single neurons in cortex with trains of clicks. They described three classifications of cells based on the ability of the neurons to respond to each click in the train: lockers, groupers and special responders. Most of the lockers were able to follow clicks up to 100 per second with the highest limiting rate being
1000 per second. However, it was concluded that none of the cells were selective to particular rates of clicks. Kiang and Goldstein (1959) used interrupted noise stimuli, and came to approximately the same conclusion.

Sideband inhibition has been found in the responses of cells at all levels of the auditory CNS in the anesthetized cat (Katsuki et al., 1959). Abeles and Goldstein (1972) found the same type of inhibition in cortical cells of unanesthetized cats. A detailed survey Shamma and Symmes (1985) described several classes of units in the cortex of the unanesthetized squirrel monkey according to the degree and type of two tone inhibition. Some neurons displayed strong inhibition from a second tone whose frequency was up to two octaves above or below the neuron's characteristic frequency. Other units exhibited little or no inhibition. Strong lateral inhibition was only seen in primary auditory field (AI) and possibly the rostral field (RL).

A large body of work has been devoted to the idea that the auditory cortex is responsible for processing and interpreting vocalizations. The squirrel monkey has a repertoire of approximately 12 well characterized stereotyped calls (Winter et al. 1966). Winter and Funkenstein (1973) recorded the response of single units in the auditory cortex of the awake squirrel monkey to tape-recorded samples of vocalizations. It was hoped that units might respond selectively to individual calls. However, the majority of neurons responded to most calls and no neuron could be proven to respond selectively to any specific call. It was found that responses of cells in the inferior colliculus to the same type of stimuli exhibited even less selectivity (Manley and Mueller-Preuss, 1981). The responses in various nuclei of the auditory CNS to species-specific vocalizations has been investigated in other animals, including cats (Watanabe and Katsuki, 1974), and birds (Bonke et al., 1979; Muller and Leppelsack, 1985) with similar results.

The vocalizations of humans and animals have many similar properties: they typically consist of a sequence of brief noise bursts, more tonal complex sounds, and
periods of silence. Vowels are relatively steady-state complex tones, lasting up to 300 ms and are more amenable to experimental manipulation. In the frequency domain, vowels consist of a harmonic line spectrum with a spectral envelope containing up to four peaks or "formants". Only the first two formants are required to identify most English vowels (Delattre et al., 1952). Single neurons in field L of the mynah bird (field L is the avian homologue of the auditory cortex) were investigated with both recorded and synthetic two-formant vowels (Langner et al., 1981). The neurons were excited only when a formant fell within a neuron's pure tone excitatory receptive field and were inhibited if a formant fell within an inhibitory sideband. The responses were entirely accounted for by the pure tone sensitivity of the units. No higher level feature extraction seems to have taken place.

In contrast to this apparent inability of single cortical neurons to extract and represent complex features of sound (i.e. vocalizations), there are reports that some cat cortical neurons are sensitive to the direction of movement of a sound source (Sovijarvi and Hyvarinen, 1974), and that neurons in field L of the white-crowned sparrow are specifically sensitive to particular sequences of tones (Margoliash, 1983). The auditory cortex of the mustached bat (Pteronotus parnellii) contains many examples of neurons which are highly selective to specific features of its echolocation calls (reviewed by Suga, 1984)

Theory and Models for Complex Tone Perception:

Place versus Periodicity Coding.

It is not clear how the CNS derives percepts such as pitch or the identity of a vowel from complex sounds. Two basic approaches are possible: temporal and spectral analysis. Signal processing can be performed in the time domain (extraction of periods and fine structure from the time course of the waveform), or in the frequency domain (using frequencies and amplitudes of all resolved Fourier components).
Spectral information is present in the spatial distribution of peaks in the firing rate and in synchronicity along the topographic array of fibers tuned to narrow frequency bands in the auditory nerve (Sachs et al., 1982), which project to all higher stations of the auditory pathway. Behavioral resolution of components has been observed only for the first six to eight partials of a harmonic series (Plomp and Mimpen, 1968). However, residue pitch can still be perceived when only high order unresolved components are presented (Ritsma, 1962).

Temporal information concerning the waveform is brought into the CNS by phase locking of the auditory nerve fiber discharges. The upper limit for phase locking is about 5000 Hz (Kiанг et al., 1965) which also corresponds to the upper frequency limit for the existence region of the residue pitch, musical tones and vowels. There is also physiological evidence for high resolution processing of the temporal code: onset units in the AVCN can entrain to a stimulus with greater precision than their auditory nerve inputs (Rose et al., 1974; Rhode and Smith, 1986), presumably by an averaging process.

In cochlear nerve fibers temporal parameters of tones are represented very faithfully in period histograms (Horst et al., 1986; Javel, 1980) if one takes into account certain non-linear transformations. These transformations include an instantaneous compressive non-linearity, gain control and bias, followed by rectification. These mathematical operations can replicate period histograms of primary neurons responding to harmonic complexes (Greenwood, 1988). These data showed that a temporal code of high quality is available in the population of cochlear nerve fibers even for rich stimulus complexes. It is therefore, not surprising that the waveform is replicated, as analogue field potentials in the avian laminar nucleus (Sullivan and Konoshi, 1986) and its homologue in mammals the medial superior olivary nucleus (Bojanowski, et al., 1988). The upper frequency limit of phase-locking of neural spike trains to stimulus periods appears to decrease in the more rostral parts of the auditory nervous system (e.g. Schreiner and
Urbas, 1988). This suggests that if this cue is to be available in the higher stations, it must be transformed into a rate or place code.

There are several mathematical models of pitch perception. Most of the theories, including those of Terhardt (1972), Goldstein (1973), and Wightman (1973) depend on resolution of the Fourier components. The pitch is equated with that of the closest fitting harmonic complex. In contrast, the theories of Langner (1983) and Raatgever and Bilsen (1986) depend on temporal cues, using an autocorrelation process to compute the pitch. The model of Srulovicz and Goldstein (1983) combines use of both place and temporal coding principles. Interval histograms from the auditory nerve are used to compute the frequency components which are then fed into a pattern matching algorithm to obtain the closest matching harmonic complex.

The preceding pages have reviewed the psychophysics of complex tone perception, the processing of tones in the auditory nervous system, and the theoretical models currently available for our understanding of pitch processing. The following chapters will describe and interpret experiments in complex tone perception in the primate, the rhesus monkey.
INTRODUCTION

The psychophysical parameters of pitch are now well enough understood that several sophisticated mathematical models have emerged (Terhardt, 1972; Goldstein, 1973; Wightman, 1973; Langner, 1983; Raatgever and Bilsen, 1986; Loeb et al., 1983). Pitch seems to be a basic stimulus dimension in auditory perception. The pitch of a given signal remains constant even when the original signal has been greatly degraded. For example, when the fundamental and several overtones are removed from a harmonic complex tone, the complex still retains its original pitch (the phenomenon of the missing fundamental; de Boer, 1976). Missing fundamentals have been shown to be extracted in the central nervous system, in the absence of stimulus energy or distortion tones at the frequency of the fundamental in the cochlea (Houtsma and Goldstein, 1972) or in the presence of low pass masking noise in the region of the fundamental (Licklider, 1951; Patterson, 1969). In order to directly study the neuronal elements responsible for these psychoacoustical observations, an animal model is required. This raises the question whether animals can also extract pitch from a sound. It has been suggested (Terhardt, 1972) that the perception of the missing fundamental arises as an epiphenomenon of human language acquisition and speech processing. However, Heffner and Whitfield (1976), and Whitfield (1980) have shown that cats can perceive the pitch of the missing fundamental. Cynx (1986) suggested that even the European Starling may be able to perform a task using this cue. While the results with cats and birds argue against the assumption that missing fundamental perception is a specifically human ability, brain mechanisms supporting this percept in these animals may differ from those in humans. Therefore, the perception of the missing fundamental was investigated in the rhesus
monkey, in order to establish a primate model for the study of neural pitch mechanisms likely to be similar to those in humans.

METHODS

Subjects: Four female rhesus macaques of ages between 2 1/2 to 4 years, were used in this study. All were colony-raised, psychophysically naive and had no evident ear disease or hearing deficit.

Stimulus generation: The tests and training took place within a sound attenuated chamber. Free field stimuli were presented from an overhead speaker (ADS 200) with the animal's head centered approximately in the speaker's axis at a distance of 1 m. Pure and complex tones were generated by a 16-bit digital synthesizer with a 35 kHz digital to analog conversion rate. The frequencies and amplitudes of up to 15 partials could be controlled independently. The digital waveform was sent through Krohn-Hite 3343 filter set to low pass with a 48 dB per octave slope and 15 kHz cut off frequency. Tones were gated on and off with a Coulbourne S84-04 gate with linear 10 ms ramps. Frequencies used ranged from 100 Hz to 8 kHz.

Amplitudes were calibrated with a 1/2 inch Bruel and Kjaer 4134 condenser microphone placed at a position corresponding to the center of the animal's head. Sound levels were measured by a Hewlett Packard 3582A spectrum analyser at octave intervals and stored by the PDP 11/23 computer, which controlled sound delivery, so that compensation could be made for the speaker's transfer function and room acoustics; intermediate values were interpolated. Each partial was set to the same amplitude, which was randomly varied from 30 to 50 dB SPL. The signal to noise ratio was 60 dB or greater, and the harmonic and intermodulation distortion products were not measurable in the experimental chamber over the frequency and amplitude ranges used.
Training: The monkeys were trained using a positive reinforcement protocol with a water reward. The training sessions lasted approximately one hour and took place every day for five days a week. The monkeys had free access to water for the remaining two days a week, but were supplemented at other times if performance was poor. Punishment consisted of a ten second time-out (waiting) period which was initiated by an alarm sound (one second burst from a buzzer).

Training proceeded in several phases of increasing difficulty: Phase A: The monkey was rewarded for simply pushing the button. There were no punishments. Phase B: The monkey was rewarded for pushing the button after the onset of a 200 ms pure tone between 10 and 600 ms after tone onset. Button pushes outside this time window were punished, but there was no punishment for failing to push. The tone's frequency was chosen randomly from a list between 100 to 600 Hz. The amplitude was varied at random between 50 and 30 dB SPL and the intertrial interval was also varied randomly between 1.3 and 4.3 seconds. Phase C: The monkey was rewarded for pressing the button after the onset of the second tone of a tone pair and punished for button pressing at other times (see Fig. 5). The tone parameters varied in the same fashion as in phase B.

Phase D: The monkey was rewarded for pressing the button when the frequencies of the two pure tones in the tone pair were identical (i.e. a pure tone matching task). The monkey was punished for pressing in response to different frequencies, and also for errors in timing. The first tone of the pair (the "unknown" tone) was initially set to 400 Hz; later extra frequencies were added as performance improved. The second (comparison) tone was varied at random in frequency and amplitude. Its fundamental frequency was related to that of the first tone by a set of fixed ratios. At the outset, the ratios were 2.0, 1.0 and 0.5; thus the possible frequencies for tone two in this case were 200, 400 and 800 Hz. The frequencies of both tones were equal for 50% of the presentations. The ratios were
gradually brought closer to 1.0 and the list of possible frequencies for the unknown tone expanded as the monkey's skill improved.

**Experiment**

The experiment was similar to training phase D, but complex tones were used in place of pure tones (Fig. 5). Comparison tones with a full set of harmonics were compared to "unknown" complexes with and without fundamentals and other lower components. The comparison tone was composed of the first five harmonics. Its fundamental frequencies were related to those of the first tone's fundamentals by fixed ratios, as in phase D (0.5, 0.67, 1.0, 1.5, 2.0).

Several combinations of frequency and spectral pattern for the unknown tone were tested. Tones containing all five components were used as well as those missing one, two or three of the lowest harmonics. Fundamental frequencies were chosen between 200 and 600 Hz. Table I lists the possible stimulus configurations for the fundamental of 400 Hz with examples of spectral patterns of the first and second tone. Components which could be deleted are shown in brackets. Approximately 400 tone pair presentations were made each session.

Four fundamental frequencies for the unknown tone were tested per day, two per half hour. Typically for a half-hour session, one complex tone contained a fundamental and the other did not. The computer recorded the number of presented stimulus tone pairs together with the number of responses to each pair. The ratio of responses to presentations per stimulus pair indicated the response probability, which served as an index for perceived tone similarity. The criterion that was used to determine that the monkeys could complete the task correctly was the occurrence of a response probability for two tones with matching fundamentals that was five times greater than the probability obtained for test tone fundamentals an octave above or below the matching value.
FIG. 5 The behavioral paradigm used with two tones. The lower part of the figure shows the timing of the two tones in phases C and D, and in the experiment. The scale bar represents 100 ms. The monkey must press the button during the "reward window" to receive its reward. Above each tone are displayed harmonic complex tone spectra. The fundamentals are marked as $f_1$ and $f_2$ respectively; $f_2$ is 2.0 times $f_1$ in this example. The dashed lines in tone one's spectrum show the components which might be removed in the course of the experiment.
The diagram represents a sound testing scenario with two tones, f1 and f2. The sequence begins with f1, followed by TONE 1 labeled as 'UNKNOWN', then f2, and finally TONE 2 labeled as 'COMPARISON'. The test concludes with a REWARD WINDOW after 100 ms.
Table I. Stimulus components for spectral patterns of first (unknown) tone and the set of second (comparison) tones for a 400 Hz fundamental. Components which could be missing are shown in brackets.

<table>
<thead>
<tr>
<th>TONE 1 COMPONENTS (Hz)</th>
<th>RATIO</th>
<th>TONE 2 COMPONENTS (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(400),(800),(1200),1600,2000</td>
<td>0.5</td>
<td>200,400,600,800,1000</td>
</tr>
<tr>
<td></td>
<td>0.67</td>
<td>266,533,800,1066,1333</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>400,800,1200,1600,2000</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>600,1200,1800,2400,3000</td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>800,1600,2400,3200,4000</td>
</tr>
</tbody>
</table>
RESULTS

Training progressed smoothly from phase A to phase C. However, all animals (M1 to M4) experienced difficulty in phase D. From 90 to 150 training sessions were required to bring the animals' performance to criterion for the full range of fundamentals. Progress was slow but steady as the range of possible unknown tone fundamentals was increased. The data listed in the results were obtained in 8 testing sessions. Subject M4 failed to learn the task when the fundamentals were missing and M3 learned it only for a narrower range of fundamental frequencies (200-450 Hz). In contrast, subject M1 generalized the task quickly to many fundamental frequencies once pitch matching was learned, and could readily recognize the missing fundamental.

Fig. 6 shows results from the three monkeys which were able to complete the task. The results are grouped by the spectral pattern of the unknown tone (rows) and by subject (columns). Each point on the graphs represents at least 30 stimulus pair presentations.

When the unknown and comparison tone contained all five harmonic components, and when their fundamentals matched, both were perceived by the monkey as being similar with a high probability (Fig. 6, top row). Response probability dropped to significantly lower levels when the fundamentals differed (Table II). Response probabilities and accuracy for the lower fundamental tones (250 Hz) appears poorer than for higher tones. Response probabilities are also lower for higher fundamental frequencies. The results in the top row of Fig. 6 might, of course, be expected since the stimuli judged most frequently to be the same were in fact identical.

However, when the two tones were not physically identical in the case when comparison tones contained fundamentals and the unknown tones did not (Fig. 6, second row), the similarity between the tones was still judged greatest when the (present and absent) fundamentals were identical (Table II), although response probability was
somewhat lower. The same was true, when the two lowest components were missing in the unknown tone (Fig. 6, third row). The further case in which the unknown tone consisted only of the 4th and 5th harmonic components, the two tones with the same fundamentals were no longer matched reliably (Fig. 6 bottom row).

Monkeys M1 and M2 performed the experimental task well, except that monkey M2 had more difficulty matching when two components were present in the unknown tone and only learned the task for a fundamental of 400 Hz under this condition. Monkey M3 was also able to perform the task correctly though it showed less selectivity to matching tones (i.e. more responses to incorrect frequency matches) and lower response probabilities overall.

Several trends are evident in the data. The curves from the data have in common a single peak, centered on the fundamental of the unknown tone. For a given spectral pattern, the peaks are highest for the middle range of frequencies used, namely 400 and 450 Hz, with a decrease in peak height with decreasing number of components in the unknown tone. In the bottom row of Fig. 6 only the 400 Hz fundamental shows a peak for matching fundamentals.

Grouping the responses by both fundamental frequency and spectral pattern (Table III) reveals some further details. When all the components are present in the unknown tone (spectral pattern 1-5), the differences between responses to matching versus nonmatching fundamentals are all significant. The differences remain significant when the fundamental and higher overtones were absent, at all fundamentals. The unknown tone requires as few as two components to be matched correctly when fundamentals of 400 Hz are used. Three and four components are required for fundamentals of 600 and 200 Hz respectively. The lowest significance levels were found for the frequency ratio 0.67
The statistical significance of differences between groups of response probability scores was determined by the Student's t test. This test requires the assumptions that the monkey's scores have an approximately Gaussian distribution and that the variances of each group of scores are of similar magnitude. With non-parametric statistical tests, these constraints are relaxed.

The Mann-Whitney U test, which can be used to determine the significance of the difference between populations, pools and then ranks the scores of two groups and establishes the degree of overlap or mixing of the two groups. If the scores of one group are ranked higher than all the scores of the second group, the difference between the two populations is maximally significant. Mixing of the scores of the two groups decreases the likelihood that the groups are, in fact, separate.

In the present study, all of the response probabilities for the condition of frequency ratio = 1.0 were greater than the response probabilities for any other condition. Thus, the difference of the scores with ratio = 1.0 from the other scores at the other conditions is maximally significant for the given degrees of freedom. Many frequency ratios had only two scores in one monkey, resulting in a significance level of $p = 0.25$, close to the level of random chance. However, when data from ratios not equal to 1.0 are pooled (i.e. the data with frequency ratios of 0.5 and 0.67 are considered to be one group), the difference is significant to levels of $p = 0.01$ and beyond.
FIG. 6. Results for subjects M1, M2 and M3, by columns. The numbers to the left give the range of harmonic components contained in the unknown tone (spectral pattern) for that row. Component 1 is the fundamental. Each curve represents responses from one unknown tone fundamental, f1.

Top row: (□) 250 Hz; (△) 450 Hz; (○) 550 Hz.

Bottom three rows: (□) 200 Hz; (△) 400 Hz; (○) 600 Hz.

The comparison tone frequency (f2) is plotted on the x-axis and is normalized by dividing it by the unknown tone frequency (f2/f1). The y-axis shows the response probability (see text). Each point represents the response probability to a specific tone pair.
Table II. Statistical significance (Student's t test) of differences in response probabilities between the case of matching fundamentals and those of other frequency ratios. Data are grouped by spectral pattern of unknown tone.

<table>
<thead>
<tr>
<th>Spectral pattern</th>
<th>Frequency Ratio (f2/f1)</th>
<th>2.0</th>
<th>1.5</th>
<th>0.67</th>
<th>0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-5</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>2-5</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>3-5</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>4-5</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>**</td>
</tr>
</tbody>
</table>

* - significant at a 99% level

** - significant at a 95% level
Table III. Statistical significance (Student's t test) of differences between responses to stimuli with matching frequency ratios \((f_2/f_1 = 1.0)\) and nonmatching frequency ratios \((f_2/f_1 \neq 1.0)\). Significance levels are grouped by fundamental of unknown tone and spectral pattern.

<table>
<thead>
<tr>
<th>Fundamental Frequency (Hz)</th>
<th>Frequency Ratio ((f_2/f_1))</th>
<th>2.0</th>
<th>1.5</th>
<th>0.67</th>
<th>0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fundamental Frequency</td>
<td>Spectral Pattern</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>250</td>
<td>1-5</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>200</td>
<td>2-5</td>
<td>**</td>
<td>**</td>
<td>***</td>
<td>*</td>
</tr>
<tr>
<td>200</td>
<td>3-5</td>
<td></td>
<td></td>
<td></td>
<td>**</td>
</tr>
<tr>
<td>450</td>
<td>1-5</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>400</td>
<td>2-5</td>
<td>*</td>
<td>**</td>
<td>**</td>
<td>*</td>
</tr>
<tr>
<td>400</td>
<td>3-5</td>
<td>*</td>
<td>**</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>400</td>
<td>4-5</td>
<td>*</td>
<td>*</td>
<td>***</td>
<td>*</td>
</tr>
<tr>
<td>550</td>
<td>1-5</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>600</td>
<td>2-5</td>
<td>**</td>
<td>**</td>
<td></td>
<td>**</td>
</tr>
<tr>
<td>600</td>
<td>3-5</td>
<td>*</td>
<td>**</td>
<td>**</td>
<td>*</td>
</tr>
</tbody>
</table>

* - significant at 99% level

** - significant at 95% level

*** - significant at 90% level
DISCUSSION

The results of these experiments show that rhesus monkeys can match complex tone stimuli with and without fundamentals. The design of this experiment was to make the experimental task dependent on the subjective attribute of pitch. There are, however, other cues and strategies that the monkeys could have used with some degree of success.

The implicit assumption in the following discussion is that pitch perception is not dependent on the presence of the fundamental component as a combination tone. It is well known that simple difference tones \((f_2 - f_1)\) become psychoacoustically prominent when primary tone amplitudes are ca. 51-57 dB (Plomp, 1965). Tone amplitudes above 50 dB SPL were not used for this reason. According to the study of Smoorenburg (1972), 33% of human subjects would be able to hear out the cubic difference tone \((2f_1 - f_2)\) under the following restricted conditions in this experiment: unknown tone fundamental frequencies of 200 or 400 Hz; the second to fifth harmonics present in the stimulus; amplitudes of 45 dB SL or greater. This condition arose for only a fraction of the trials in the second row of Fig. 6 (approx. 26%; the percentage of trials that had primary tone amplitudes greater than or equal to 45 dB at 200 and 400 Hz). Thus, the use of cubic difference tones would be important in 9% of the trials and probably made only a minor contribution to the monkey’s scores. It is further unlikely that distortion products could have replaced the missing fundamentals since cortical neurons of the same monkeys, examined in the same laboratory, and tuned to relevant pure tone frequencies, did not respond when the frequency of the missing fundamental of harmonic complexes of up to a 60 dB, fell within the neuron’s pure tone excitatory receptive field (Schwarz and Tomlinson, 1987).

One obvious strategy which the monkeys might have used to perform the task employs component matching. Since only the first five harmonics were used in the complex tones, human psychophysical data indicates that all of the partials could have been resolved (Plomp and Mimpfen, 1968), and might serve as a basis for comparison.
The monkeys were initially trained with pure tones (i.e. fundamentals only). Subsequent training utilized complete complex tones. It is possible that the monkeys learned to compare the fundamentals of either class of tone exclusively, and ignore any overtones in completing their task. Since they were still able to match complex tone pairs when one tone had no fundamental, another strategy is required to account completely for their success. However, presence of the fundamental did improve the monkeys' performance (response probability) and is, therefore, not unimportant. This is not surprising since two stimuli with fundamentals resemble each other more strongly than one stimulus with and the other without fundamental.

Another potential strategy utilizes the comparison of one or more partials between the two tones: the more components in the two tones that match, the more likely the monkey would be to push the button. The best match would occur when unknown and comparison complexes have the same fundamental (i.e. all components of the unknown tone would have corresponding components in the comparison tone). The next best match would occur when the fundamentals were one octave apart. Thus, monkeys employing this strategy should also have matched the two tones at fundamental frequency ratios of 0.5 and 2.0 with greater probability than at intermediate values.

The predicted results for such a component matching strategy are illustrated in Fig. 7. For each stimulus pair the following calculation was made: the number of matching partials in the unknown and comparison tones were counted; the results were then grouped according to spectral patterns of unknown tones, and normalized. For example, taking the case of frequency ratio 0.5, a fundamental frequency of 100 Hz, and unknown tone composed of components 2, 3, 4, and 5, one finds the following frequencies:

Unknown tone frequencies (Hz): 200, 300, 400, 500.

Comparison tone frequencies (Hz): 50, 100, 150, 200, 250.
There are 2 components in common between the unknown tone and comparison tone out of a possible maximum of 4 (when the frequency ratio is 1.0). The normalized response probability is therefore $2/4 = 0.5$.

The plots in Fig. 7 assume the simplest case of equal perceptual weighting for each partial in the complexes. The curves do not change substantially even if the weights are allowed to vary by as much as 50%. The curves in Fig. 7 contain up to three peaks, but only one central peak is evident in the subjects' responses. Animals adhering strictly to a component matching strategy would have had up to three peaks in their responses, and in all cases, elevated response levels at the upper octave. The absence of these peaks in the subjects' responses implies that the animals have probably not employed this strategy.

The large number of matching components of two harmonic complex tones whose fundamentals are an octave apart leads one to the phenomenon of "octave generalization". To a human observer two musical notes an octave apart bear a certain similarity (Deutsch, 1982). This similarity leads to octave errors in a pitch matching task, whereby pitches which are at octaves above or below the target tone may be selected erroneously. Such errors are not surprising when working with harmonic complex tones, since at such intervals, the target tone and the matched tone will share some components. Albino rats have demonstrated signs of octave generalization to pure tone stimuli (Blackwell and Schlosberg, 1943), but the frequencies tested were well over 5000 Hz, and the magnitude of harmonic and intermodulation distortion tones was not adequately measured. The subjects in this experiment showed no tendencies towards octave generalization, which as previously mentioned could also indicate the use of a component matching strategy.
FIG. 7. The expected response for a component matching strategy (see text for explanation). The numbers to the left of each panel give the spectral pattern for the unknown tone. The axes are as in Fig. 6.
FIG. 8. Data of Fig. 6 (bottom row, left most panel) replotted along a horizontal axis representing the absolute frequency of comparison tone fundamental. Fundamental frequencies of tone one are: (□) 200 Hz; (△) 400 Hz; (○) 600 Hz.
One method the monkeys could employ does not involve a comparison between the
two tones at all. The monkeys could completely ignore the unknown tone and press the
button when the frequencies of the comparison tone fell within a particular absolute
frequency range. This strategy would be successful if only a small range of unknown tone
frequencies was used. The results in the bottom left panel of Fig. 6 show evidence of this
strategy as can be seen when the data are replotted along an abscissa representing the
absolute fundamental frequency of the comparison tone. Examination of Fig. 8 shows that
the monkey's response probabilities are maximal when the fundamental of the comparison
tone is 400 Hz, irrespective of unknown tone frequency. The one exception to this trend
emerges at the lower peak on the 200 Hz curve, and in this case the two tone complex
starts at 400 Hz. The preference for the fundamental of 400 Hz probably arises as a
consequence of the monkey's training since they were first rewarded for 400 Hz pure tones
and complexes. One interpretation is that the monkey found the matching task to be too
difficult when only two components are available in the unknown tone and resorted to this
strategy instead.

Intensity is also an available cue. A monkey might attend to frequency-dependent
loudness cues and only push the button when tone one and tone two appeared to occur at a
certain loudness ratio. Monkeys discriminate intensities well (Sinnott et al, 1985) and
might find this easier than attending to pitch cues. Since the amplitudes of the tones were
randomly varied over 20 dB, this cue was probably not useful to the animal.

In the preceding paragraphs several alternate strategies have been considered that
the monkeys may have used but which do not require reconstruction of a missing
fundamental. It has been shown that these strategies would not be effective under these
experimental conditions. The simplest remaining strategy which can generally account for
these results requires the animals' perception of the missing fundamental. The test stimuli
would be most similar when the fundamentals, both missing and present, matched. In the
case with only two components in the unknown tone (Fig. 6, bottom row), the fundamental was difficult to reconstruct, and the monkeys may have had to fall back on another more simple, but less effective strategy.

It is concluded that monkeys must be able to reconstruct the pitch of the missing fundamental. This implies that machinery for processing pitch information is present in the brains of lower primates. One may now begin to address these questions at a new level and investigate possible neurophysiological mechanisms as to how a missing fundamental might be reconstructed as well as attempt to determine an internal representation of the pitch of complex tones in an animal which is phylogenetically similar to man.
CHAPTER 3

RESPONSES OF CORTICAL NEURONS TO
HARMONIC COMPLEX TONES

INTRODUCTION

Monkeys can perform a task which involves perception of a missing fundamental of a harmonic complex tone stimulus (Chap. 2; Tomlinson and Schwarz, 1988). In humans, the pitch of a harmonic complex tone is identical with the pitch of a pure tone at the frequency of the fundamental, even when that fundamental component is not physically present (cf. de Boer, 1976). Missing fundamental perception is mediated, at least in part, by the interactions occurring solely within the central nervous system, (Houtsma and Goldstein, 1973). Yet the neural mechanisms involved in complex tone processing remain poorly understood.

Though single unit studies have been performed using stimuli with harmonic line spectra (amplitude modulated tones by Schreiner et al., 1983; Schreiner and Urbas, 1986; repetitive click trains by de Ribaupierre et al., 1972; Goldstein et al., 1971) the focus has been on the sensitivity of the cortical neurons to temporal parameters of the stimulus, particularly the stimulus period. Experiments employing interrupted noise, which has a flat long term spectrum, found no representation of periodicity in cat cortex (Kiang and Goldstein, 1959). The experiments described in this chapter represent an attempt to determine how the spectral parameters of harmonic complex tones are represented in the auditory cortex of unanesthetized monkeys, with the aim of uncovering possible physiological mechanisms of pitch perception.
METHODS

The subjects were three juvenile female rhesus monkeys (*Macaca mulatta*) aged from 3 1/2 to 4 1/2 years old, weighing between 4.5 and 6.5 kg. All subjects (M2, M3, M4) had been previously trained in a psychophysical pitch matching task (Chap. 2; Tomlinson and Schwarz, 1988) and two of them (M2, M3) could apparently perceive the missing fundamental.

The pure and complex tone stimuli were generated by a specially built digital synthesizer. The synthesizer had a 16-bit digital-to-analogue converter (DAC) driven at 72 kHz. The signal was then fed into a 12-bit multiplying DAC and through an eight-pole elliptical filter with a cut-off frequency of 20 kHz. Up to fifteen sinusoidal components of a specified frequency and amplitude could be digitally added together into a complex tone. The white noise used for search stimuli and unit testing was produced by a Coulbourne S81-02 noise generator. The signals were switched by a Coulbourne S84-01 rise/fall gate with linear 5 ms ramps. Sounds were presented to the animal within an electrically shielded, sound treated booth from an overhead speaker (ADS 200 or ADS L520). The monkey's head was positioned on the speaker's axis, approximately 1 meter from the drivers.

The tones generated by the digital synthesizer were automatically calibrated by computer. A 0.5 inch Bruel and Kjaer condenser microphone was placed in the center of the space normally occupied by the animal's head. The amplitude of ninety one pure tones spaced equally along a logarithmic scale from 80 to 18000 Hz were measured by an HP 3582A spectrum analyser and stored in a table in the computer to be used as a reference for the production of tones of a known amplitude sound pressure level (dB SPL). Tone amplitudes at intermediate frequencies were interpolated from the table. No harmonic or intermodulation distortion was measurable for the range of amplitudes and frequencies used.
FIG. 9 A cross section in the coronal plane showing a microelectrode penetrating auditory cortex. The sylvian fissure is shown by the arrow. The stainless steel chamber is secured to the skull with dental acrylic and anchoring screws. The horizontal and vertical scales represent respectively the dorso-ventral \((D-V)\) and medio-lateral \((M-L)\) stereotaxic coordinates with respect to ear bar zero.
Surgical procedures: A stainless steel chamber was surgically implanted vertically over the auditory cortex. Aseptic procedures were observed at all times. The animals were anesthetized for surgery (sodium pentobarbital at 35 mg/kg intraperitoneally, supplements as required) and the head placed in a stereotaxic frame. A midline incision was made and the skin and temporal muscle overlying the skull reflected. A 2 cm. hole was drilled in the skull with its center being located 5.9 mm anterior to ear bar zero and 20.5 mm lateral from the midline (coordinates adapted from Pfingst and O'Conner, 1980), and stainless steel anchoring screws were placed around the chamber in the skull bone. The chamber was set in place with a stereotaxic positioning device. The meninges were left intact. The entire apparatus was cemented together with dental acrylic and a threaded Teflon cap screwed onto the chamber to seal it. The animals were allowed at least a week to recover before the recording sessions began.

Recording procedures: Epoxylite-coated, etched tungsten microelectrodes (0.8-2.0 mega-ohm impedance) were used to record extracellular potentials from the animal's cortex. The tips of these electrodes were protected when penetrating the dura mater by a 22 guage stainless steel guide tube. The electrode was attached to the microdrive so that it was recessed 1.2 mm within guide tube when the microdrive was fully retracted.

The monkey was positioned in the chair and its head was immobilized by bolting a stainless steel bar implanted during surgery on the skull to a socket attached to the chair. A Trent Wells x-y positioner was affixed to the chamber and the track coordinates set on a 1.0 x 1.0 cm grid. The guide tube and electrode were inserted into parietal cortex above the superior temporal plane (Fig. 9) and the electrode was advanced into auditory cortex with the microdrive.

The amplified signals from the microelectrode were fed into a window discriminator and to a digital storage oscilloscope. Each discriminated spike waveform
was compared to a previously stored reference waveform. Pulses from the discriminator were conducted to the computer and displayed as points in a dot raster plot (e.g. Fig. 15). Data acquisition sweeps were 500 ms long with a 100 ms delay until tone onset with a binwidth of 2 ms. Histograms could be taken with respect to either the time or frequency axis from the dot rasters.

Irregularly spaced 200 ms noise pulses were used to find and isolate auditory units. Single units were tested with sets of pure and complex tones as well as white noise at several intensities. The pure tone frequencies and the complex tone fundamentals were taken directly from the table of calibrated frequencies. The tones were presented from the list in a pseudorandom sequence. The complex tones were composed of harmonic components 1 to 8, where the fundamental is component 1. Missing fundamental stimuli could contain up to component 11. Complex tones with frequencies over 18000 Hz were presented with those components absent. To obtain rate-level curves and find the threshold at the best frequency (BF), blocks of 40 identical BF tones were presented at several intensity levels. To assess for inhibitory sidebands, a series of BF tones was presented at 20-30 dB above threshold, together with a pure tone series, at the same or different intensities. Blocks of 40 noise bursts at different intensities were used to obtain the rate-level function to broad band stimuli.

At the end of the series of electrode tracks in an animal placed over a period of months, two or three small electrolytic lesions (10 microamps for 10 minutes, electrode positive) were made in the cortex. After a period of at least three days (to permit a gliosis to develop around the lesions), the animal was sacrificed by transcardiac perfusion while under deep barbiturate anesthesia. The brain was fixed by subsequent perfusion of 10% buffered formalin and allowed to post-fix for seven days, immersed in the same solution. The fixed brain was then embedded in celloidin (M4) or frozen (M2 and M3) and cut into 30 micron sections which were stained with Cresyl violet or Thionine. The
sections were then examined under the microscope to locate the electrolytic lesions, electrode tracks, and primary sensory cortex. The criteria and nomenclature of Pandya and Sanides (1973), and Galaburda and Pandya, (1983) were used to identify cortical fields.

RESULTS

A total of 476 units were isolated from 176 tracks in 3 monkey cortices. The characteristic frequency (CF) was estimated in 367 units and Fig. 10 shows their distribution over the population. The CF of a cell was estimated as the frequency with the greatest evoked rate of discharge (the best frequency, or BF) in the responses to the lowest amplitude set of pure tones obtained. For broadly tuned units, the frequency of the geometric center of the lowest intensity response area was used to estimate CF. The CF's observed within the population ranged from 100 Hz to over 18000 Hz with the highest proportion of units between 500 and 2000 Hz (Fig. 10). It is uncertain if this reflects the actual predominance of CF in monkey auditory cortex; preference was given to recording in regions where unit frequencies were below 5000 Hz, since frequencies above 5000 Hz were found (Ritsma, 1962) not to be important for perception of the missing fundamental in human beings.

Histology: Each cortex was examined histologically to identify the location of koniocortex and surrounding belt cortex. The sites of the electrolytic lesions and defects from electrode tracks allowed the registration of the cytoarchitectonic maps with the coordinates from the x-y positioner. Pfingst and O'Conner (1980) estimated the error in location of an electrode track at 0.7 mm with the vertical stereotaxic method, and recommended a track spacing of 1.5 mm for unambiguous localization of units. In this experiment, units were recorded with more narrowly spaced tracks to obtain a large sample of cells from each cortex.
FIG. 10 Histogram showing the distribution of characteristic frequency for 367 single neurons in the superior temporal cortex of the rhesus monkey.
FIG. 11 Maps of neuron types and best frequencies found in auditory cortex of subject M2.

A) Map of estimated characteristic frequency (CF) in the cortex of subject M2. The coordinates of the tracks (signified by dots) are taken from the settings on the x-y positioner. The numbers beside each dot show the CF, in kHz, of the units encountered in the track. Each square is one millimeter on each side.

B) Map of classified units in the cortex of subject M2. The scale is the same as in A. The symbols are defined in the following key:

- FILTER
- RESOLVER
- FUNDAMENTAL
- NARROW BAND
- WIDE BAND
- UNCLASSIFIED
FIG. 12 Maps of neuron types and best frequencies found in auditory cortex of subject M3.

A) Map of estimated characteristic frequency (CF) in the cortex of subject M3. Axes are as in Fig. 11.

(following page)

B) Map of classified units in the cortex of subject M2. The scale and symbols are the same as in Fig 11.
FIG. 13 Maps of neuron types and best frequencies found in auditory cortex of subject M4.

A) Map of characteristic frequency (CF) in the cortex of subject M4. Axes are as in Fig. 11.

(following page)

B) Map of classified units in the cortex of subject M2. The scale and symbols are the same as in Fig. 11.
Table IV. Results from linear regressions of the unit CF (dependent variable, logarithmic frequency scale) versus the unit's antero-posterior coordinate (independent variable, setting on x-y positioner in mm) for each subject.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Slope (octaves/mm)</th>
<th>Standard Error of Slope</th>
<th>Correlation Coefficient ($r^2$)</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>M2</td>
<td>-0.908</td>
<td>0.147</td>
<td>0.574</td>
<td>30</td>
</tr>
<tr>
<td>M3</td>
<td>-0.831</td>
<td>0.118</td>
<td>0.216</td>
<td>180</td>
</tr>
<tr>
<td>M4</td>
<td>-0.925</td>
<td>0.113</td>
<td>0.375</td>
<td>113</td>
</tr>
</tbody>
</table>
Maps of CF versus track coordinates (as set on the positioner) are shown for each animal in Figs. 11a, 12a and 13a. In each map there is a trend of higher CF in the posterior region, with low frequencies in the anterior region. Variation from this trend is common. In the anterior portion of M3 and M4's cortices, the trend shows signs of reversing. Linear regressions were performed of logarithm of the CF's versus the antero-posterior coordinate of the units on the x-y positioner for each hemisphere (see Table IV). The trend of high frequencies posteriorly and low frequencies anteriorly can be seen by the negative slope of the lines, however, the scatter of the points about each line was such that the slopes were not significantly different from zero (Student's t test, p = 0.05) in any hemisphere.

Due to the potential degree of overlap of tracks, only the following descriptive account is given of the location of cytoarchitectural fields on Figs. 3, 4 and 5. All tracks in subject M2 were found caudal to koniocortex, and probably fall within field paAc. Subject M3's koniocortex was judged to lie in the anterior 50% of Fig. 12, with the rostral field (paAr) appearing in the anterior 3%. In subject M4, koniocortex occupies the anterior 60% of Fig. 13, with the rostral field appearing in the anterior 15%. While electrode tracks were seldom observed lateral to koniocortex, several tracks were found in the medial edge of KA and paAr penetrating the depths of the inferior limiting sulcus and entering field proA.

**Electrophysiology:** The most prominent impression gained from analysis of the neural recordings was that there was no evidence for simple categories of neural behavior. This does not imply that the responses were homogeneous, rather that the neural behaviors seemed to be distributed along a continuum (the possible dimensions of which will be considered in Chap. 4.) Near points in this continuum there were groups of units with responses that showed distinctive or simple properties. It is these groups whose behavior is discussed in the following
paragraphs. The reader should keep in mind however, that a large proportion of
units exhibited characteristics that were intermediate between two or more of these
groups.

Five simple neuron classes could be recognized, based on the unit's response
to pure tones, complex tones, and noise in a subpopulation of 251 units. There are:
1) filter neurons; 2) resolver neurons; 3) fundamental neurons; 4) wide band neurons;
and 5) narrow band neurons. Table V shows a summary of the criteria used to
classify the units. Table VI lists the unit types, together with their number and
the range of CF exhibited by each type. The narrow band neurons and
fundamental neurons have the most restricted CF ranges, and also comprise the
smallest populations. Unclassified neurons, which had intermediate response
behavior, occupy the largest proportion of units (58%) surveyed. The locations of
the classified units are displayed on maps in Figs. 11b, 12b and 13b. Each unit
class is represented by a symbol, and the symbols are plotted according to the
track coordinates on the x-y positioner. The distribution of neurons according to
depth within the cortical mantle is given in Fig. 14 for 115 units. The depths are
normalized with respect to the thickness of the responsive region. Due to the
curvature of cortex and the pooling of unit depths from different cytoarchitectonic
regions, the normalized depths used in the present study only show whether the
units occupy either superficial, middle, or deep layers in the cortical mantle. The
top part of the Fig. 14 shows the depth distribution of the entire set of neurons, the
majority of which were found in the middle and deep layers. The lower part of
Fig. 14 shows the depth distribution for each type of unit.
Table V. Summary of the criteria for unit classification.

<table>
<thead>
<tr>
<th>UNIT CLASS</th>
<th>UPPER FREQUENCY LIMIT OF CT RESPONSE</th>
<th>LOWER FREQUENCY LIMIT OF CT RESPONSE</th>
<th>RESPONSE TO WIDE BAND STIMULI</th>
<th>RESPONSE TO NARROW BAND STIMULI</th>
</tr>
</thead>
<tbody>
<tr>
<td>FILTER N.</td>
<td>$f_u$</td>
<td>$&lt;f_{l-2}$ oct.</td>
<td>good</td>
<td>good</td>
</tr>
<tr>
<td>RESOLVER N.</td>
<td>$f_u$</td>
<td>$&lt;f_{l-2}$ oct.</td>
<td>good</td>
<td>good</td>
</tr>
<tr>
<td>UNCLASSIFIED N.</td>
<td>$&lt;f_u$</td>
<td>$&gt;f_{l-2}$ oct.</td>
<td>good</td>
<td>good</td>
</tr>
<tr>
<td>FUNDAMENTAL N.</td>
<td>$f_u$</td>
<td>$f_l$</td>
<td>good</td>
<td>good</td>
</tr>
<tr>
<td>WIDE BAND N.</td>
<td>n/a</td>
<td>n/a</td>
<td>good</td>
<td>poor</td>
</tr>
<tr>
<td>NARROW BAND N.</td>
<td>n/a</td>
<td>n/a</td>
<td>poor</td>
<td>good</td>
</tr>
</tbody>
</table>

Abbreviations: $f_u$ - upper frequency limit of pure tone response; $f_l$ - lower frequency limit of pure tone response; N. - neuron; oct. - octave; CT - complex tones with components 1 to 8; > - less than; < - greater than; n/a - not applicable.

1 Accounts for 54% of these neurons.
Table VI. Population, and range of characteristic frequencies of functionally classified cortical units.

<table>
<thead>
<tr>
<th>UNIT CLASS</th>
<th>FREQUENCY RANGE (Hz)</th>
<th>NUMBER</th>
<th>PERCENTAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>FILTER</td>
<td>184-18000</td>
<td>52</td>
<td>21%</td>
</tr>
<tr>
<td>RESOLVER</td>
<td>330-11450</td>
<td>27</td>
<td>11%</td>
</tr>
<tr>
<td>FUNDAMENTAL</td>
<td>170-2800</td>
<td>8</td>
<td>3%</td>
</tr>
<tr>
<td>NARROW BAND</td>
<td>100-280</td>
<td>7</td>
<td>3%</td>
</tr>
<tr>
<td>WIDE BAND</td>
<td>248-9100</td>
<td>9</td>
<td>4%</td>
</tr>
<tr>
<td>UNCLASSIFIED</td>
<td>100-18500</td>
<td>147</td>
<td>58%</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td>251</td>
<td>100%</td>
</tr>
</tbody>
</table>
FIG. 14 Distribution of units according to normalized depth.

Top panel: Histogram of distribution of 75 units within the cortical mantle of the following composition: 25 filter neurons; 3 narrow band neurons; 5 fundamental neurons; 3 wide band neurons; 9 resolver neurons; and 30 unclassified neurons. Depth is expressed on a scale from 0.0 to 1.0, 0.0 being outermost. Lower group of panels: Proportional occupation of depth according to unit type. Each panel shows how a particular unit type (title to right of panel) is distributed with respect to cortical depth. For example, at level 0.0-0.1 there were 4 units (top panel), 25% of which were filter neurons (lower panel).
Four of the five neuron types exhibited complex tone responses which were either stronger or equal to the neuron's pure tone response. The simplest of these is the "filter" neuron, so named because its response is reminiscent of a simple bandpass filter. Units were classed as filter neurons if they responded well to pure tones and noise and displayed no evidence of strong sideband inhibition. Strong sideband inhibition blocks both spontaneous and evoked neural discharges. At a given intensity the pure tone response falls between an upper and lower frequency limit, \( f_u \) and \( f_l \), respectively. The frequency borders of the responses were measured from histograms, which were computed off-line (see Fig. 10). The fundamentals of the complex tone (1-8) response area covered a single range of frequencies extending from the upper edge of the pure tone response \( f_u \), to at least 2 octaves below the lower edge of the pure tone response \( f_l \). If the lowest fundamental of the complex tone response extended down to 84 Hz (the lowest frequency tested), but less than 2 octaves below \( f_l \), and there was no evidence of strong sideband inhibition, then the unit was also classified as a filter neuron.

Fig. 15 shows a representative example of the filter neuron class. The pure tone response (left column) decreases its bandwidth \((f_u - f_l)\) with decreasing intensity of stimulus. The complex tone response appears down to fundamentals three octaves below \( f_l \). The rate-level function to BF pure tones is monotonic as is the rate-level function to noise. Filter neurons are the most numerous (21%) class of cells and have been found in all fields in cortex (Figs. 11-13), distributed at all depths (Fig. 14).

A second type of cell is the "resolver" neuron. Like filter neurons, resolver neurons responded well to both pure and complex tones, and the complex tone response ranged over fundamentals from \( f_u \) to 2-3 octaves below \( f_l \). Unlike filter neurons, the cells had to exhibit two or more peaks in their complex tone response histograms (see Fig. 16) on at least one plot to be classified as resolver neurons. These distinct peaks are a result of a relatively high frequency resolving power, reflected also in the unit's pure tone response. Sideband inhibition could be present or absent.
FIG. 15 The responses of a filter neuron (unit code D-72-A) with an estimated characteristic frequency of 1000 Hz. The left column of rasters contains responses to pure tones and the right column of rasters shows the responses to complex tones (components 1-8, with 1 being the fundamental). Each dot in the rasters denotes an action potential. The y-axis of the rasters represents fundamental frequency of the stimulus (for pure tones it represents frequency) along a logarithmic scale in 91 equal steps from 84 to 18000 Hz. The frequencies are shown in kHz. The x-axis of the rasters represents time, measured in 2 millisecond bins. The bar denotes the presence of a 200 millisecond sound burst. The histograms to the right of each raster are computed by counting the dots between the two vertical lines ( cursors) in the middle of the raster. The y-axis of the histogram is the same as the raster. The x-axis in each graph represents the number of counted spikes. Two lower panels show the rate-level functions for pure tones at 1000 Hz and noise. The y-axis represents spike count and the x-axis represents stimulus amplitude in dB SPL. The evoked response to tones or noise (square symbols) was determined by counting the spikes which occurred from between 20 to 250 ms after onset of 40 repeated bursts. The spontaneous activity (cross symbols) was estimated by counting the spikes in the 100 ms pre-stimulus interval in the same stimulus presentation series and then scaling them to be comparable with the evoked counts.
"FILTER" NEURON

PURE TONES

COMPLEX TONES (1-8)

TONES RATE-LEVEL FUNCTIONS NOISE

SPECTRUM LEVEL (dB) SPONTANEOUS

SPECTRUM LEVEL (dB) EVOKED

D-72-A
FIG. 16 The responses of a resolver neuron (unit J-2-C) with an estimated CF of 4200 Hz. The left column contains the responses to pure tones and the right column shows the responses to complex tones with components 1-8. The axes of the rasters and histograms are as described in Fig. 15. The numbers to the left of each row designate the sound pressure level of the stimulus in dB SPL.
"RESOLVER" NEURON

PURE TONES

COMPLEX TONES 1-8

60

40

20

J-2-C
Table VII. Maximum number of components and dynamic range of resolution of resolver neurons.

<table>
<thead>
<tr>
<th>Max. components resolved</th>
<th>Number of Units</th>
<th>Dynamic Range (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>4</td>
<td>10-30</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>10-30</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>10-50</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>10-35</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>40</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>30</td>
</tr>
</tbody>
</table>
Table VIII. The distribution of the characteristic frequencies of resolver neurons.

<table>
<thead>
<tr>
<th>Frequency range (Hz)</th>
<th>Number of Resolver Neurons with CF</th>
</tr>
</thead>
<tbody>
<tr>
<td>125-250</td>
<td>0</td>
</tr>
<tr>
<td>250-500</td>
<td>5</td>
</tr>
<tr>
<td>500-1000</td>
<td>3</td>
</tr>
<tr>
<td>1000-2000</td>
<td>12</td>
</tr>
<tr>
<td>2000-4000</td>
<td>3</td>
</tr>
<tr>
<td>4000-8000</td>
<td>2</td>
</tr>
<tr>
<td>8000-16000</td>
<td>2</td>
</tr>
</tbody>
</table>
Fig. 16 shows an example of a resolver neuron with up to six bands in its complex tone response plots, visible in both the rasters and in the histograms. The bands on a given plot become narrower and less distinct with lower fundamental frequencies. Fewer bands are distinguishable in high amplitude complex tone plots, when the pure tone response bandwidth is wider.

Not all resolver neurons show the same degree of banding, as Table VII illustrates. The left column of Table VII shows the distribution of maximum observed number of bands that the units could resolve, together with the dynamic range over which at least two bands could be distinguished. Cells which had more bands, retained them over a greater dynamic range. Table VIII shows the distribution of CF in the population of resolver neurons, in the octaves from 125 Hz to 16000 Hz. The largest proportion (74%) of frequency resolution occurs for frequencies between 250 and 2000 Hz. Some resolver neurons showed suppression of spontaneous discharges in the pure tone response, indicating the presence of inhibitory sidebands. Resolver neurons respond well to noise stimuli, though some had nonmonotonic rate-level functions. Resolver neurons were found in the middle to deep layers (Fig. 14), in most fields of cortex (Figs. 11-13) but were not found in the anterior low frequency region in subject M4 in Fig 13b.

A third class of unit is the "fundamental" neuron. A neuron belongs to this class if the tuning to fundamentals of complex tones is similar to that for pure tone. That is, if the range of fundamentals in the complex tone (1-8) response does not extend below \( f \). An example is shown in Fig. 17. The pure tone response bandwidth of this neuron was relatively constant with respect to amplitude. The complex tone (1-8) response histograms to the right of each raster demonstrate a similar tuning for fundamentals of pure tones and for complex tones. A series of complex tones (4-11) missing their fundamental was presented, however there was no response to missing fundamental complexes. Two tone testing revealed a powerful low frequency inhibitory sideband in
this cell. Such sidebands were characteristic for fundamental neurons. The rate-level curves of the unit in Fig. 17 were monotonic with a dynamic range of 30-40 dB for BF pure tones, and nonmonotonic for white noise. Fundamental neurons were found in the middle layers of cortex (Fig. 14) which was either anterior or posterior to primary auditory cortex (Figs. 11-13).

A fourth class of cells, "wide band" neurons (Fig. 18), responds much better to wide band stimuli (both complex tones and noise) than to pure tones. This might be expected simply on the grounds that the total energy of wide band stimuli is greater than pure tones when both have the same spectrum level (component amplitude). An eight-component complex tone has eight times greater total energy than a pure tone with the same spectrum level, corresponding to an increase of 9 dB on a logarithmic scale (3 dB per doubling of energy). A cell was classified as a wide band neuron if it responded better to an eight-component complex tone series than to a pure tone series having a 10 dB greater spectrum level. Wide band neurons were observed mainly in deep layers of cortex (Fig. 14), and caudal to koniocortex (Figs. 11-13).

The previous four classes of cells responded as well or better to complex tones as they did to pure tones. The fifth class of cells responded better to pure tones than either complex tones or noise stimuli, and are called "narrow band" neurons. An example of this type is found in Fig. 19. Most of the narrow band neurons exhibited signs of strong sideband inhibition. Two-tone testing of the neuron in Fig. 19 revealed a powerful upper inhibitory sideband. The nonmonotonic nature of the pure tone (at CF) and noise rate-level functions in this cell show the effects of sideband inhibition.

The remainder of the 251 neurons formed the so-called "unclassified" group. The majority (54%) of these units had a single pure tone response band (from \( f_u \) to \( f_u \)), and a response to complex tones with fundamentals that extended from less than \( f_u \) to...
below $f_1$, but not to frequencies as low as $f_1 \cdot (2$ octaves$)$. The responses were similar to the filter neuron class, but had a more restricted complex tone response band. Even more restricted complex tone response bands were discerned in the fundamental neurons. The rest (46%) either had weak and diffuse responses to sound, showed pure tone responses with multiple excitatory domains, had responses that could be classed as either one type or another depending on what temporal part of the response was being regarded, or had responses that were inconsistent from trial to trial.

One example of a combination of response types (though not typical of all unclassified neuron responses) is seen in Fig. 20, containing both the filter neuron pattern and the fundamental neuron pattern in different temporal portions of its response. The response during the stimulus shows similar tuning for fundamentals of complex tones and pure tones (fundamental neuron type). The "off" portion, after termination of the stimulus, indicates responses for complex tone fundamentals at and below the pure tone best frequency (filter neuron type).

Some other neurons (unclassified) exhibited two excitatory bands in their responses to pure tones. Fig. 21a illustrates one such neuron whose excitatory region appears to be divided by a central inhibitory band. The neuron in Fig. 21b exhibits a narrowly tuned high frequency excitatory region lasting throughout the stimulus duration accompanied by a weaker and broader low frequency response near the stimulus onset.

Twenty-three cortical neurons were tested with missing fundamental stimulation. The characteristic frequency of these units varied from 170 to 3700 Hz and the amplitude of the stimuli ranged from 10 to 60 dB SPL. There was no evidence of responses to difference tones ($f_2 - f_1$ and $2f_1 - f_2$) at the frequency of the missing fundamental.
FIG. 17  The responses of a fundamental neuron (unit code D-24-C) with a CF of 173 Hz. The axes of the rasters and the histograms are described in Fig. 15. The numbers to the left of each row of rasters designates the stimulus amplitude in dB sound pressure level. The left column shows responses to pure tones. Responses to complex tones containing their fundamental (components 1-8) are in the right column.

(following page)
The left column of rasters shows responses to a series of complex tones missing their fundamentals (components 4-11). The right column contains a series of responses to two tone stimulation. The fixed tone had a frequency of 173 Hz (CF) and an amplitude of 40 dB. The numbers to the left of the rasters show the intensity of the second simultaneously present tone. The lower panels show the rate-level functions to 173 Hz pure tones (left) and white noise (right). The rate-level axes are described in Fig. 15.
"FUNDAMENTAL" NEURON

PURE TONES

COMPLEX TONES 1-8
"FUNDAMENTAL" NEURON

COMPLEX TONES 4-11

TONE 10.0

SPECTRUM LEVEL (dB)

SPONTANEOUS

EVOKED

D-24-B
FIG. 18  Responses to pure tones, complex tones and noise for a wide band neuron (unit code J-50-C) at 50 dB SPL. The axes of the rasters and the histograms are described in Fig. 15. The weak pure tone response was centered at 1.5 kHz.
"WIDE BAND" NEURON

PURE TONES

COMPLEX TONES (1-8)

NOISE

J-73-A
FIG. 19. Responses of a narrow band neuron (unit J-57-A) with a CF of 121 Hz. The axes of the rasters and the histograms are described in Fig. 15. The responses to pure and complex tones (1-8) with amplitudes of 40 dB SPL are shown in the top two dot raster plots. The lower dot raster plot shows inhibition of the evoked response to a 40 dB tone at 121 Hz by a second tone at 45 dB. The lower two panels contain rate-level functions for 121 Hz pure tones and noise. Axes are as described in Fig. 15.
"NARROW BAND" NEURON

PURE TONES

COMPLEX TONES (1-8)

TWO TONES

TONES RATE-LEVEL FUNCTIONS

NOISE

SPIKE COUNT

SPECTRUM LEVEL (dB)

SPONTANEOUS

EVOKED

J-57-A
FIG. 20  Unclassified neuron exhibiting a mix between filter neuron ("off" response) and fundamental neuron ("through" response). The unit exhibits a CF of 1250 Hz for the through response. The axes of the rasters and the histograms are described in Fig. 15. The top raster shows responses to pure tones and the bottom raster to complex tones (components 1-8) at 50 dB SPL. The left histograms were computed over the "through" response (according to the cursors in the rasters). The right histograms were computed over the "off" response (cursors not shown).
MIXED TYPE

PURE TONES

COMPLEX TONES 1-8

G-19-A
FIG. 21 Neurons with pure tone responses having two excitatory regions. The axes of the rasters and the histograms are described in Fig. 15. See text for discussion.
PURE TONE RESPONSES

A

B
DISCUSSION

Neuron classes

Cortical units have been seen to respond with a variety of patterns to complex tones. In the preceding section, various response classes were identified. How do these responses arise and what is their role in the processing of acoustic stimuli?

Filter neurons: Filter neurons were characterized by a strong response to complex tones (1-8) with fundamentals at and below the pure tone response area. How does this pattern originate? The top part of Fig. 22 illustrates four complex tones (1-8) in different positions with respect to the pure tone response area of a hypothetical neuron. When any component of the complex tone enters the excitatory region (between the two horizontal lines) the unit responds. The vertical bar on the right side of Fig. 22 shows the range of fundamentals for which this condition is true and predicts a response to fundamentals three octaves (eight components) below the lower edge of the pure tone response.

Filter neurons lack sideband inhibition (by definition). Sideband inhibition also exists at peripheral levels of the CNS as low as the cochlear nuclei (Greenwood and Maruyama, 1965; Evans and Nelson, 1973). If such inhibition is present at any stage in the ascending input to a cortical unit, it would most likely be reflected in the response of that cell. Thus filter neurons probably receive projections from cells which are relatively free from inhibitory sidebands. Cells of the AVCN rarely have inhibitory sidebands (Evans and Nelson, 1973) and could provide the first stage of this input. There are not sufficient experimental data to speculate on the identity of higher stages of such a pathway at this time.
FIG. 22 Prediction of complex tone response from knowledge of pure tone response for the following two types of harmonic complex tone: tones with components 1-8 (top panel) and tones with components 4-11 (bottom panel). Frequency is represented on the vertical dimension along a logarithmic scale. The hypothetical unit responds to pure tones with frequencies ranging between upper and lower limits $f_u$ and $f_l$ (two long horizontal lines). This range is illustrated by thick vertical bars on the left of the figure. Four examples of complex tones are shown in each panel, represented by sets of horizontal lines, one line for each component of the complex. The lowest lines in each set represent the fundamental of the complex. The dashed horizontal lines in the lower panel represent missing components (i.e. components 1-3) of the complexes. The four cases demonstrate the following conditions:

A) the complex tone is entirely below $f_l$.

B) the upper component of the complex tone is just above $f_l$.

C) the lowest physically present component of the complex tone is just below $f_u$.

D) all of the complex tone components are above $f_u$.

The vertical bars on the right show the range of fundamental frequencies which have components falling within the pure tone response area. In the top panel this bar extends from $f_u$ to three octaves below $f_l$, and in the lower panel, the bar is entirely below $f_l$. 
COMPLEX TONES  1-8

PURPOSE TONE RESPONSE

COMPLEX TONE RESPONSE
(FUNDAMENTALS)

log f

A  B  C  D

COMPLEX TONES  4-11

PURPOSE TONE RESPONSE

COMPLEX TONE RESPONSE
(FUNDAMENTALS)

log f

A  B  C  D
FIG. 23 An expanded plot of the 40 dB complex tone (1-8) response from the resolver neuron in Fig. 16. The axes for the raster and the histogram are as in Fig. 15. The left column of figures shows the frequencies measured from the histogram peaks and the right column shows the predicted peaks based on the subharmonics of 4200 Hz (i.e. 4200/1, 4200/2, 4200/3, . . . , 4200/6).
COMPLEX TONES
HARMONICS 1 - 8

<table>
<thead>
<tr>
<th>ACTUAL (Hz)</th>
<th>PREDICTED (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4200</td>
<td>4200</td>
</tr>
<tr>
<td>2115</td>
<td>2100</td>
</tr>
<tr>
<td>1408</td>
<td>1400</td>
</tr>
<tr>
<td>1130</td>
<td>1050</td>
</tr>
<tr>
<td>841</td>
<td>840</td>
</tr>
<tr>
<td>712</td>
<td>700</td>
</tr>
</tbody>
</table>

40 dB
FIG. 24 A plot of the peak frequencies in histograms of resolver neurons compared with subharmonics of the best frequencies. The neuronal best frequency (BF) is represented along the horizontal axis. The vertical axis shows the frequencies of the peaks, normalized by division by the best frequency, and plotted on a logarithmic scale. The horizontal lines show the locations of the subharmonics of the best frequency (see text for explanation).
Resolver neurons: The distinctive feature of resolver neurons is the banding of their complex tone response. This banding becomes more pronounced the narrower the pure tone response bandwidth is. A simple explanation which accounts for the banding is that the neurons are able to respond to individual components of a complex tone. Consider a neuron with an infinitesimally narrow frequency response area such that the cell responds only to tones at CF. As the fundamental frequency of an eight-component harmonic complex tone is increased, the first component to enter the receptive field will be the eighth component. The neural response, plotted at the fundamental frequency, will appear on the raster display at one-eighth of the CF. As the seventh component enters, a response will appear at a fundamental which is one-seventh of the CF, and so forth. The resulting response plot will have bands at the first through to the eighth subharmonics of the CF. Fig. 23 compares the subharmonics of 4200 Hz (the CF of the resolver neuron in Fig. 16) to the center frequencies of each band in the complex tone response at 40 dB. The strong agreement between the two lists shows that the banding is indeed caused by complex tone components discretely exciting the cell.

In further support of this hypothesis, Fig. 24 shows the correlation between actual and predicted frequency values for the bands in the complex tone responses of all resolver neurons. The neuronal best frequency is represented along the horizontal axis. The vertical axis shows the frequencies of the peaks, normalized by division by the best frequency. The horizontal lines show the locations of the subharmonics of the best frequency. Note the close clustering of the measured values about the lines. A chi-squared test verifies the closeness-of-fit of the measured values to the values predicted by subharmonics of the best frequency to a $p=0.05$ level (chi-squared value of 69.0 on 58 degrees of freedom). This result shows that a population of cortical neurons can respond discretely to components of harmonic complex tones.
This neural resolution of components of complex tones in cortex shows some interesting parallels with human psychophysical experiments. Up to the sixth harmonic could be distinguished in the neural responses and lower harmonic numbers are resolved with the greatest security. The observed dynamic range of resolution varied from 10 to 50 dB. Trained human subjects can typically hear out the first six to eight partials of a harmonic complex tone (Plomp and Mimpen, 1968), resolving the lower components better than the higher ones, similar to the maximum neural resolution that was observed in monkey cortex. Thus, aspects of analytical complex tone perception (Terhardt, 1972) are reflected in the physiological behavior of cortical neurons. This correspondence has some limits, however. Humans can resolve components of complex tones up to levels of 80 dB SPL (Pick, 1977; Scharf and Meisleman, 1977), above which performance deteriorates. It has also been shown (Bilsen and Ritsma, 1970; Bilsen and ten Kate, 1977) that the minimum peak-to-valley ratio of comb-filtered noise necessary to perceive a pitch is almost unchanged when the overall amplitude of the noise ranges over the interval from 30 to 120 dB SPL. The best cortical neurons that have been seen can resolve 6 components over a dynamic range of approximately 40 dB and the majority much less. Only part of the dynamic range of human performance can be accounted for by characteristics of monkey resolver neurons. The dynamic range in the dorsal cochlear nucleus (DCN) is, however, much larger. Evans (1977) recorded from type 3 and 4 cells (with strong sideband inhibition) in cat DCN and stimulated them with comb-filtered noise with a peak-to-valley ratio of 21 dB. Comb-filtered noise is spectrally similar to harmonic complex tones, and has a quality of "repetition pitch" (Bilsen, 1966). These cells were able to resolve peaks of the stimulus, by means of modulations in their firing rate, over a range of 100 dB. In the same paradigm, auditory nerve fibers demonstrated dynamic ranges of 30-40 dB. No auditory nerve fiber showed resolution at levels greater than 70 dB SPL. The difference between auditory nerve and DCN was attributed to the effect of sideband inhibition in the input to DCN neurons. An experiment in cat anteroventral cochlear nucleus (AVCN) by
Smoorenburg and Linschoten (1977) observed resolution of components of a harmonic complex tone of up to the 13th harmonic. The resolution in the cat cochlear nuclei of components of complex tones and comb-filtered noise seems to be superior to both the resolution seen in cortex and demonstrated human psychophysical abilities. Species-specific differences may help to account for these findings.

The cochlear nuclei of primates are less well differentiated anatomically than those of felines (Moore, 1980), reflecting "encephalization" of the primate auditory system (Heiman-Patterson and Strominger, 1985). The cat cochlear nuclei may perform more extensive processing of the input signal than the primate cochlear nuclei, though we must consider also that such a high dynamic range may not be present in any single units in the primate brain. Even if the dynamic range and frequency resolving power were substantially greater in the primate cochlear nuclei than in cortex, convergence of frequency channels, and saturation of responses at high levels could still explain the lower range and resolution observed in cortical neurons.

Humans can hear out six to eight partials in a harmonic complex tone, whereas most resolver neurons could only resolve up the third component in a complex tone. There may be differences between human and monkey psychophysical abilities to hear out components of complex tones which is consistent with the low resolution of the physiological representation in monkey cortex. Although monkey psychophysical tuning curves are as sharp as in humans (Serafin et al., 1982; indicating that the filters in the periphery are equally sharp), the frequency discrimination for pure tones (Sinnott et al., 1987) and for formant frequencies (amplitude peaks in the spectrum of a complex tone; Sinnott and Kreiter, 1986) are described as poorer in monkeys by a factor of at least three. On this basis one might expect a poorer frequency representation in monkey cortex. If, however, the primate psychophysical ability to resolve components in a complex tone is on a par with humans, one must conclude that the firing rate of a single neuron in cortex in insufficient to
represent frequency components of tonal stimuli in the brain, and that frequency information may be encoded in a way not addressed by these experiments.

The range of BF in the population of resolver neurons extends from 330 to 11450 Hz (Table VII). Numerous studies of auditory cortex and other auditory regions (e.g., Phillips and Irvine, 1981) have found that the sharpness of tuning, measured by $Q_{10\text{dB}}$, in the responses of auditory neurons increases with their characteristic frequency. Since frequency resolving power seems to be the prime requisite for the formation of the resolver neuron response, one would expect that proportionately more high-CF neurons would exhibit this response than low-CF neurons. It is likely that one would have found a greater percentage of resolver neurons than reported here if more recording had been done in high-CF regions of cortex.

Several models of complex tone pitch perception require as inputs, the frequencies of the complex tone components (Goldstein, 1973; Terhardt, 1972; Wightman, 1973). A relatively high resolution of the peripheral activation pattern (compared with filter neurons) can be found in cortex in the population of resolver units. Although the resolution of components in cortex may not be necessary for pitch processing, the presence of such a frequency representation may reflect pitch processing in lower stages of the auditory system.

Fundamental neurons: This population of neurons was distinguished by tuning of the complex tone fundamentals which was similar to the tuning to pure tones (Fig. 17). Incomplete harmonic complexes were applied to test the possibility that the neuron might respond to missing fundamentals, which would indicate a sensitivity to the feature of pitch and not simply frequency. The lack of response to missing fundamental stimuli, coupled with the strong lower inhibitory sideband supports a more conventional explanation of fundamental neuron behavior. Complex tones with fundamentals below the CF would activate a lower inhibitory sideband and cause the response of the unit to be suppressed.
Only when all of the energy is at and above $f_1$, the lower bound of the pure tone excitatory response area, can the neuron be activated. The lower part of Fig. 22 explains the (lack of) response to missing fundamental stimuli. The range of fundamentals which are excitatory is shown by the bar on the right side of the figure. It is entirely below $f_1$, and also below the lower border of the dot raster plot of the cell illustrated in Fig. 17. As their name implies, fundamental neurons do seem to be tuned to fundamentals of complex tones, however the physical presence of the fundamental is still required.

Aural combination tones might have arisen which would have re-introduced the fundamental component when harmonic complex tones missing their fundamentals were presented. In humans, the simple difference tone $(f_2 - f_1)$ is audible with primary tone amplitudes above 51-57 dB SPL (Plomp, 1965). However the cubic difference tone $(2f_1 - f_2)$ would not be audible in complex tone stimuli with components 4-11 when primary tone levels are 60 dB SPL or less (Smoorenburg, 1972). Similar indications come from electrophysiological experiments in cat anteroventral cochlear nucleus (Smoorenburg et al., 1976). Changes were found in the firing rate of AVCN neurons when difference tones $(f_2 - f_1$ and $2f_1 - f_2$) entered their pure tone response area with primary tone levels as low as 40 dB SPL. Interval and period histograms of the firing of AVCN neurons showed temporal representation of combination tones at even lower amplitudes of primary tone. These results indicate that manifestations of distortion tones might be found in cortical responses with tone amplitudes over 40 dB. However, in the 23 units in which missing fundamental stimulation was used, no evidence of a response to missing fundamentals was found. This indicates that difference tones, if present in the cochlea, played little role in the rate representation of tonal stimuli in cerebral cortex in these experiments.

Wide band neurons: In this population of neurons, an increase in the bandwidth of the stimulus was more effective than an increase in pure tone amplitude in increasing the firing rate. One possible mechanism to account for this involves the convergence on these
neurons of many frequency channels whose firing rates saturate. A narrow band of energy concentrated at one frequency, whose amplitude is increased, will eventually saturate the channels devoted to that frequency. Further increases in convergent activation of wide band neurons would be due to the recruitment of other channels, for example, by spread of activation along the basilar membrane. On the other hand, a broad band stimulus would recruit the other channels from the outset, and thus be a more effective stimulus. Since saturating rate-level functions are commonplace in the CNS (e.g. Irvine, 1986), such a mechanism is plausible.

Type IV cells in the DCN of the unanesthetized cat also respond better to noise than to pure tones (Young and Brownell, 1976). Type IV cells receive inhibition from type II/III cells (Voigt and Young, 1980), which respond weakly or not at all to noise (Young and Voigt, 1982). Thus the response of wide band neurons may be accounted for by local inhibition from cortical neurons responding like type II/III cells, or wide band neurons may receive their input from DCN type IV cells.

The acoustic environment of monkeys contains several wide band sounds which might be processed by these cells. Grunts (Winter et al., 1966, squirrel monkeys) and shrieks (Rowell and Hinde, 1962, rhesus monkeys) are examples of monkey vocalizations with broad spectra. Although wide band neurons would be powerfully activated by such stimuli, experiments using vocal stimuli found no cells selective to broad spectrum vocalizations (Winter and Funkenstein, 1973; Newman and Wollberg, 1973).

*Narrow band neurons:* Most narrow band neurons had low characteristic frequencies and either one upper inhibitory sideband, or both upper and lower inhibitory sidebands. As a result, this class of cells responded preferentially to narrow band stimuli, and poorly to broad band complex tones and noise. The response pattern of narrow band neurons exhibits similarities to that of some type II/III neurons of DCN (Young and Brownell, 1976). Type II/III cells have excitatory responses to pure tones and can respond
poorly or not at all to noise. They are thought to be local inhibitory interneurons (Young and Voigt 1982) and therefore probably cannot provide ascending inputs to narrow band neurons in cortex. One naturally occurring variety of stimulus having energy concentrated in narrow frequency bands are the vowel-like vocalizations with formants. Neurons have been found in field L of trained mynah birds which respond selectively to human vowels based on formant frequency peaks in the vowel spectra (Langner et al., 1981). The formants of effective vowels fell within the neuron’s pure tone excitatory band. Vowels which were not effective in exciting the cells had formants which either fell completely outside the excitatory area, or fell within an inhibitory sideband. The long low coo of the Japanese macaque (Macaca fuscata) has a steady low frequency formant with its energy centered at about 450 Hz (Green, 1975). Narrow band neurons might conceivably function as low frequency formant detectors.

*Other responses:* A large proportion (58%) of the 251 classified cells did not fit into the strictly defined types described in the results. The majority of these cells (54%) behaved in the manner of filter neurons, having a more restricted band of complex tone fundamental frequencies, but not as restricted as either fundamental neurons or narrow band neurons. It is hypothesized that a moderate sideband inhibition causes these restrictions. Relative to filter neurons, upper sideband inhibition would cause a reduction in the upper boundary of the complex tone response, and lower sideband inhibition would cause an increase in the frequency of the lower boundary of the complex tone response area. Chapter 4 explores the effects of sideband inhibition on the complex tone response in more detail.

The other 46% of these cells either had weak and diffuse responses to sound, or showed pure tone responses with multiple excitatory domains, and correspondingly intricate complex tone responses.
Some cells (in the remaining 46%) had responses that varied from trial to trial. Part of this variability may stem from the fact that the subjects were unanesthetized. The freedom from anesthetics allows the potential for the brain to function entirely normally, which is not the case with pharmacological intervention (i.e. paralysis, sedation, and/or anesthesia). The level of activity and the number of units one encounters is dramatically increased in alert versus anesthetized preparations (Evans and Whitfield, 1964). Sustained responses to tonal stimulation are also more prevalent in the unanesthetized preparation (Evans and Whitfield, 1964).

On the other hand, there are additional variables which must be allowed for when dealing with unanesthetized animals. The usual brain motions caused by pulsation of the heart and respiratory action (Miller and Sutton, 1976) are joined by occasional gross movements of the animal's limbs and torso which deleteriously affects the stability of the recordings. Movements and vocalizations both introduce extraneous sound, and evoke a middle-ear reflex (Carmel and Starr, 1963) that attenuates airborne stimuli by up to 30 dB. It has been shown that the level of consciousness, anesthetic state and attentiveness of the animal to the stimuli can all affect the responsiveness of some cortical neurons (Hubel et al. 1959; Pfingst et al. 1977; Benson and Hienz, 1978; Hocherman et al., 1976; Beaton and Miller, 1975), even in meaningful stimuli such as vocalizations (Manley and Mueller-Preuss, 1978). In one estimate, 20 percent of units exhibited increased firing rates to attended versus non-attended stimuli (Benson and Hienz, 1978). The mechanisms for such effects include the direct influence on cortical neurons by the non-thalamic projections to auditory cortex from brainstem nuclei such as the locus ceruleus, the raphe nuclear complex and the nucleus basalis of Meynert (Campbell et al., 1987; Sato et al., 1987a,b; Foote and Morrison, 1987), or indirect mechanisms acting below the level of the cortex (e.g. the "thalamic gate hypothesis" of Skinner and Yingling, 1977).
In this study the subjects were not anesthetized, and were awake at the time of recording. Since no measures were taken during the experiment to assure that the animal was actively attending to the stimuli, some variability in the neuronal responses is to be expected. However, all animals were trained and had extensive experience discriminating pure tones, harmonic complex tones, and missing fundamental stimuli, and one cannot discount the possibility, that the relative proportion of the unit classes would be different in untrained animals. The training may also have helped to reduce variability in the neural responses, due to their previous importance to the animal.

**Site of Inhibitory Interaction**

Sideband inhibition has been observed as peripherally as the dorsal cochlear nucleus (Voigt and Young, 1980), but is also present in inferior colliculus (Ryan and Miller, 1978) and medial geniculate (Aitkin and Webster, 1972). Inhibition is also observed in binaural interactions, where sound coming from one ear inhibits the responses of sound to the other ear. Binaural inhibition is present at every level of the auditory CNS (Guinan et al., 1972; Roth et al. 1978; Aitkin and Webster, 1972). With extracellular recording it is impossible to tell whether the observed inhibitory sidebands arise within the cortex or result from some interaction taking place at an earlier point in the auditory nervous system. It is clear, however, that neural substrate for inhibitory interactions exists in the cortex. Inhibitory post-synaptic potentials have been recorded intracellularly from auditory cortical cells (de Ribaupierre et al., 1972) and glutamic acid decarboxylase, the synthesizing enzyme for the inhibitory neurotransmitter, GABA and has been localized immunohistochemically in neurons intrinsic to cortex (e.g. Houser et al., 1984). Therefore, at least some of the inhibitory interactions observed may have taken place within the cortex.
Off Responses in Cortical Neurons

Responses at the end of a stimulus, "off" responses, are prevalent in cortex and can be quite pronounced (e.g. Fig. 20). The mechanism for this kind of behavior, namely a neural response in the absence of a stimulus, frequently involves the termination of some stimulus-evoked inhibition. At the end of the inhibition, the cell's membrane potential rises sufficiently to cause a train of action potentials, in what is called "rebound excitation". Off responses can also be found in the visual system, from the photoreceptor cells in the retina, to neurons in visual cortex. The relation between cortical off responses and pitch processing is not likely to be strong, since one does not have to wait until a tone stops to know its pitch. However, the occurrence of a tone may have subtle, but reproducible, effects on the pitch of a closely following tone (Rakowski and Hirsh, 1980). The functional significance of off responses is not currently known, nevertheless, one may speculate that they are likely to have greater significance for processing sequences of auditory stimuli (i.e. speech signals, echolocation signals, etc.) than they are for steady-state complex tones.

Complex Tone Processing By Auditory Cortex

Two-tone inhibition has been systematically explored in the cortex of the alert squirrel monkey (Shamma and Symmes, 1985). Of the classified units, 30% (type A) exhibited lateral inhibition with upper, lower, or both upper and lower sidebands, and 20% (type C) displayed no sideband inhibition, showing only summation to tone inputs. The type A units are similar to the fundamental and narrow band neurons of the present study, but were found only in primary cortex (AI) and rostral field. The type C units are similar to filter neurons and were distributed over all cortical fields. The stimuli of Shamma and Symmes were presented only to one ear, excluding binaural inhibitory and excitatory interactions and the onset time of the two tones was staggered to maximize inhibition. In contrast, the present study used free field presentation of sound and simultaneous gating.
during two-tone testing. Differences in the incidence of sideband inhibition from the present study may be explained by differences in the experimental paradigm used to collect data.

Marr (1980) described a scheme for image processing in the visual system which simultaneously analyzed the visual scene in terms of large, intermediate and small scale features, by means of multiple representations with different spatial resolutions. Large scale features would be represented most economically in the low spatial resolution channel which receives input from large retinal receptive fields. A range of resolutions (receptive field sizes) is necessary to allow integration of information about large and small scale features. In the auditory system, a purely narrow band representation could theoretically encode any spectral feature broader than its bandwidth. Progressive convergence of inputs would eventually permit the integration of information over the whole spectrum and allow the detection of large scale features. Marr’s scheme requires that both high and low resolution representations of the input spectrum to be present at the same level. Coexistence of broadly and narrowly tuned units in the auditory cortex have been observed in the alert chinchilla (Tomlinson, 1983; Tomlinson et al., 1986) and the alert cat (Goldstein et al., 1968). Arrays of narrowly tuned resolver neurons, more broadly tuned filter neurons, and wide band neurons observed in monkey cortex might represent such sets of large and small scale representations.

Are There Neurons Selective For Harmonic Complexes?

Voiced sounds (harmonically structured vocalizations) are a common part of the communications repertoire of monkeys (Green, 1975; Winter et al., 1966; Rowell and Hinde, 1962), and humans (e.g. Flanagan, 1972). It is possible that the auditory system of these animals is specialized to process harmonically structured sounds, taking advantage of the regularities in the sound to aid in signal processing. Psychophysical theories concerning human pitch perception provide some support for this hypothesis. Terhardt’s (1972) theory of pitch perception postulates that each spectral component in the stimulus generates a
series of "virtual" pitches, which fall at subharmonics of the spectral component, and which are used by the brain to determine the perceived pitch. In Goldstein's (1973) model of pitch perception, the determination of the pitch of any complex tone is based on the fundamental of the closest matching harmonic series. Both of these theories, which successfully account for many pitch phenomena, are meant to reflect underlying mechanisms used by humans for processing sound to extract its pitch. Both theories place a high significance on the sensitivity of the human auditory system to harmonic complexes. Since monkeys have also been shown to be able to perform a task dependent on the pitch of the missing fundamental (Chap. 2; Tomlinson and Schwarz, 1988), one might assume that the same or similar neural machinery is present in their brains.

An example of apparent sensitivity to harmonic complexes has been found in the auditory cortex of the mustached bat, which contains a large population of neurons which are preferentially sensitive to the bat's echolocation cry (Suga, 1984; Neuweiler, 1983). The cry consists of four harmonics (H_1-H_4). Neurons in one field of auditory cortex were selectively responsive to combinations of components from both the emitted cry and the doppler-shifted echo (i.e. H_2 from the cry and H_3 from the echo; Suga et al., 1983). With synthetic stimuli, the neurons responded best to a particular frequency pair (H_2 + shifted H_3), weakly to either frequency alone and not at all to other frequencies presented alone. (In rhesus cortex some neurons were encountered with two frequency peaks in the pure tone response area (Fig. 21), however the peaks were not harmonically related.) Even minor deviations from the optimal combination of frequencies caused a drastic reduction in response. These cells were termed "harmonic-sensitive" neurons (Suga et al., 1979), later called "combination-sensitive" neurons (Suga et al. 1983), since doppler-shifting changes the fundamental frequency of the echo, breaking the harmonic relationship. The frequency specificity of the combination-sensitive response is aided by the specialization of the bat's cochlea which provides extraordinarily sharp tuning in the region of 60 kHz, the frequency of the H_2 component (Suga and Jen, 1977).
The idea that some neurons ought to possess harmonic-sensitivity is a compelling one, but not supportable from the evidence observed in primate cortex. Although most neurons responded better to complex tones than pure tones, those that did so also responded well to noise. On this basis one must conclude that while cortical neurons respond well to various parameters of complex tones, they exhibit no special selectivity to the harmonic complex. The combination-sensitivity observed in bat auditory cortex is probably linked to the high degree of specialization of its auditory system necessary for echolocation.

It must be noted that the complex tones used in these experiments were comprised of eight successive harmonics, whose amplitudes were all equal. Most vowels contain more than eight harmonic components, have spectra with from one to four formant peaks and a gradual high frequency roll off (Flanagan, 1972). Tonal vocalizations of monkeys show similar features. The possibility that cortical neurons might exhibit harmonic-sensitivity with more vowel-like spectra cannot be conclusively dismissed.

In discussing the response selectivity of neurons to sounds with harmonic structure, one also addresses the theme of the coding of vocalizations. How well can vocalizations be represented by activity in single cells? Previous studies in squirrel monkeys (Winter and Funkenstein, 1973; Steinschneider et al., 1982; Newman and Lindsley, 1976; Glass and Wollberg, 1979; Mueller-Preuss, 1986) looked for simple representations of vocalizations in the firing of single cortical units. It was found that the majority of units responded to most calls (Newman and Wollberg, 1973), and the incidence of neurons selective to calls of any kind was low (2-3% of units exhibited selectivity to one or two acoustically related calls, Winter and Funkenstein, 1973). Even in the squirrel monkey, where the calls are acoustically distinct and stereotyped, no single unit representation was found.

Only low level feature detection (i.e. spectral peaks, spectral edges) is apparent in either the present study which used complex tone stimuli, or in the studies quoted above.
which employed vocalizations. It is possible that only low level feature analysis occurs in
single units in superior temporal cortex, serving to provide the input for some higher level
processor. If the information available from any single unit is small, then integration of
information over arrays of units becomes a primary consideration. What aspects of
complex tones might populations of cortical neurons be able to represent? As previously
mentioned, arrays of filter neurons could provide a low resolution representation of the
spectral input, with resolver neurons providing a somewhat higher resolution analysis. An
array of fundamental neurons could provide a specification of the fundamental frequency of
a harmonic complex, or the low frequency boundary of a non-harmonic stimulus. Wide
band and narrow band neurons could provide information on the bandwidth of the stimulus.
The question still remains, however, how does the CNS further process these
representations? At this point one may only speculate. The information may be fed into a
hierarchically oriented feature processor, with increasingly higher stages of feature
detection occurring with progressing degrees of integration. The convergence of center-
surround receptive fields on simple cells to form orientation-selective receptive fields in the
primary visual cortex (Hubel and Wiesel, 1962) exhibits just such behavior. At the highest
stage of such schemes one finds "gnostic" units, whose firing denotes the presence of very
specific stimulus patterns, an example of which is the hypothetical grandmother-cell (e.g.
Perrett et al., 1987). The antithetical alternative to such a scheme is one in which the
distributed pattern of activity over the population of neurons represents the information
(Hinton et al., 1986). In such a scheme, the information processing is spread diffusely
across the whole population of neurons. Discharge patterns of individual elements may
exhibit some selectivity to stimulus features (Jones and Hoskins, 1987), but not necessarily
to any great degree. The evidence gained thus far concerning the processing of complex
tones and pitch seems to favor neither gnostic units nor distributed representations, but
points to some processing scheme with elements of both.
Topographic Representation of Pure Tones in Alert Cortex

All three plots (Figs. 11, 12, 13) show a trend of high frequencies posteriorly and low frequencies anteriorly, similar to the trend seen for AI and the lateral field L, in the rhesus monkey (Merzenich and Brugge, 1973). Anteriorly in subject M4 (Fig. 18), there may be a reversal of this trend, indicating that some of the tracks may have penetrated rostrally in field RL. The large scatter in the data for unit CF vs. position, evident in the coefficients of correlation in Table IV, may be due to position errors resulting from use of the vertical stereotaxic method. There is, however, some evidence which supports the observation that large discrepancies in tonotopic maps can be found in data gathered from unanesthetized preparations (Whitfield, 1982). The original demonstration of tonotopic arrays of unit CF in the monkey cortex was performed in barbiturate anesthetized animals (Merzenich and Brugge, 1973). Most units observed in anesthetized cortex are found in the middle layers (Merzenich et al., 1975) which receive the main thalamic termination, and it is possible that the observation of tonotopy was based mostly on activity recorded from thalamic projections, since anesthetics preferentially affect polysynaptic projection routes. The more complex cortical responses found in other layers (Hubel and Wiesel, 1962) would be preferentially suppressed in such experiments. Mapping experiments in the unanesthetized preparation employing tangential penetrations have found significant deviations from a tonotopic organization in cat primary auditory cortex (Evans et al., 1965; Goldstein et al., 1970). Although the weak correlation between frequency and position observed is entirely consistent with previous investigations, it must be noted that the data in this study are not of sufficient precision to support or reject the hypothesis of a blurred tonotopy in the alert primary auditory cortex.

Topographic Representation of Complex Tones and Pitch in Alert Cortex

Each level in the anesthetized auditory CNS contains a tonotopic array of frequencies (Rose et al., 1960; Aitkin et al., 1975; Aitkin and Webster, 1972). To what extent this mapping is a topographic representation of the cochlea (cochleotopy), or a
reflection of the organization of more highly processed auditory information is not certain. In the cochlear nuclei the tonotopy is accounted for by the topographic projection pattern of the auditory nerve, whose fibers have a frequency sensitivity determined by their innervation of the basilar membrane (Liberman, 1982). In the cortex, one expects a more complex organization. Various workers have tried to determine if the cortical tonotopic map was actually a representation of stimulus periodicity by ablation experiments involving auditory cortex. Although the evidence is somewhat conflicting, it appears that simple frequency discrimination is spared after cortical lesions (Elliott and Trahoitis, 1972). However, lesions in the superior temporal plane of monkeys produces a deficit in their ability to discriminate interrupted noise at low rates (10-80 Hz) from noise with a 300 Hz interruption rate (Symmes, 1966). Whitfield (1980) found that cats trained in a task requiring discrimination of missing fundamentals, lost that ability after lesions to auditory cortex in fields AI, AII, and parts of Ep and the inferotemporal area. Perception of the missing fundamental is impaired in human patients with surgical ablation of Heschel's gyri in the superior temporal plane in humans (Zatorre, 1988). This evidence indirectly implicates the cortex in the processing of periodicity and pitch.

More direct tests of cortical periodicity coding have been made using the single unit recording technique. Kiang and Goldstein (1959) tested the sensitivity of single units in cat auditory cortex to periodically interrupted noise; no relation to the unit's CF was found. In 1972, de Ribaupierre and coworkers measured the limiting rates at which cortical neurons could phase-lock to repetitive click trains of up to 1000 impulses per second. Selectivity to specific repetition rates was not observed. Stimuli in both experiments had broad spectra and a relatively small selection of periods. Schreiner and Urbas (1986) used amplitude modulated sine waves and found cortical neurons in cat cortex which were tuned to the modulation frequency. The "modulation transfer functions" (demonstrating periodicity-selectivity) had shallow slopes, measured in terms of $Q_{2\text{dB}}$ (bandwidth at 2 dB below the modulation transfer function peak), and showed some correlation with the CF of the
location. Neurons sensitive to periodicities of amplitude modulated sine waves have been found in field L of the forebrain of trained mynah birds (Hose et al., 1987). The best modulation frequencies in field L were found to be arrayed topographically, but 95% of the best modulation frequencies were less than 100 Hz, indicating that this topographic array is probably most important for coding sound repetition rates (rhythm) rather than musical pitch.

In this study, the sensitivity of units to pitch has been directly tested and compared to the pure tone tuning curve. A small proportion of units (fundamental neurons) exhibit the same selectivity to complex tone fundamentals as they do to pure tones. However, the units require the physical presence of the fundamental. The response of the majority of classified units (filter neurons) reflects the activation pattern of cochlear nerve fibers.

Summary: Complex Tone Processing by Cortical Neurons.

The cortical processing of complex tones has been investigated at the single unit level in the alert monkey. The simplest type, the "filter neuron" maintained the representation of the cochlear activation pattern. Some units had a frequency selectivity sufficient to resolve up to the sixth partial in a harmonic complex tone, given moderate sound pressure levels. Certain cells responded preferentially to narrow band or wide band stimuli. "Fundamental" neurons seemed to be as well tuned to the fundamentals of harmonic complexes as to pure tones, when the fundamental component was physically present.

This classification scheme is based on the neurons' responses to spectral parameters of steady-state stimuli. From these responses, possible neural mechanisms for pitch processing were addressed. Pitch does not seem to be simply represented within the cortical tonotopic array, though some aspects of pitch processing may be inferred from the single unit responses to harmonic complexes. Processing of more complex signals (i.e. vocalization) was also addressed.
CHAPTER 4

EFFECTS OF INHIBITORY SIDEBANDS ON THE NEURAL RESPONSES TO COMPLEX TONES

INTRODUCTION

In the preceding chapter, the responses of cortical cells were classified according to their responses to pure and complex tones. The behavior of "resolver" neurons, "fundamental" neurons, "filter" neurons, and "narrow band" neurons was accounted for by combinations of inhibitory and excitatory frequency regions in the cells' response area. It was also indicated that the neural responses seemed to be distributed within a continuum of intermingled groups, much like vowels (Delattre et al., 1952) or monkey vocalizations (Green, 1975) are. What could be the nature of this continuum of neural responses? What are its dimensions? In Chap. 3, the parameters of time after onset of pure tone, and frequency and intensity of the pure tone were used to characterize the response of neurons. With 200 samples in time after tone onset, 91 frequencies, and up to 6 intensities, it is possible to construct a parameter space with 109,200 dimensions to describe a unit's response to pure tones. Clearly, this is too many for convenient use. Customarily, a smaller number of parameters is used to characterize the behavior of a given neuron. A prime example is the "characteristic frequency", denoting the frequency of the lowest intensity sound which will excite the neuron. As was seen in Chap. 3, neuronal responses can also be described by the bandwidth of the excitatory region, and the center frequency or frequencies of inhibitory bands and their bandwidths. The relative strength of the inhibition versus the excitation may also be considered.

If one can obtain estimates of each of these parameters, one may have a means to simply characterize the spectral behavior of the neurons. Such estimates were obtained
from the extracellular recordings of Chap. 3. Extracellular recordings, however, can provide only indirect measures of sideband inhibition, which may profoundly effect the responses of neurons to complex tones. The question arises, how can various parameters of inhibitory sidebands (center frequency, strength, and bandwidth) actually affect the responses of neurons to harmonic complex tones? As a corollary question one might ask which parameters of inhibitory sidebands are important for each of the response classes in Chap. 3? This chapter uses numerical simulations to examine the effects of inhibitory sidebands in hypothetical neurons.

METHODS

The experiments of Chap. 3 were simulated by representing the excitatory and inhibitory bands as filters and plotting the responses to the same complex and pure tone series used to assess the neural behavior. The simulations were implemented by programs written in BASIC on an IBM AT computer. The results of the simulations were displayed graphically in a format similar to the neural pure and complex tone histograms, which show the firing rate in response to iso-intensity tones of different frequencies.

Neurons in the cortex receive their inputs from brainstem auditory nuclei, which are in turn excited by the auditory nerve. Convergence of information from many frequency channels and from both ears is the rule, not the exception, in the ascending auditory pathway (Harrison and Howe, 1974). The use of single peripheral filter shapes (e.g. those of Patterson et al., 1982) to simulate cortical response bands is therefore inappropriate. Instead, one must consider the response in terms of a population of inputs, each with different thresholds and characteristic frequencies. Gaussian functions, which describe the responses of homogeneous populations, have additional virtues as filter functions of having a small number of parameters which can specify the center frequency (CF), bandwidth (BW) and relative strength (K) as in the following equation:

\[
G(f) = K \times e^{-((f-CF)/BW)^2}
\]
In this equation \( f \) is the frequency of the component being filtered and the double asterisk symbol (\( ** \)) represents exponentiation. The filter sharpness can be described by two metrics: the half-power bandwidth in Hz (Bendat and Piersol, 1971) equal to 1.18*BW, and the \( Q_{10dB} \) (CF divided by the bandwidth at 10 dB above threshold, Kiang et al., 1965) equal to \( \text{CF}/(2.15*\text{BF}) \). The unit’s net excitation to a given input frequency \( f \), \( H(f) \), was obtained by subtracting the inhibition from the excitation:

\[
(2) \quad H(f) = G(f)_{\text{ERF}} - G_{\text{ISB+}}(f) - G_{\text{ISB-}}(f)
\]

where ISB+ and ISB- are the upper and lower inhibitory sidebands (ISB’s) respectively, and the relative strength, \( K = 1.0 \) for the excitatory receptive field (ERF). The response of a neuron, \( R(f) \), to a harmonic complex tone with components 1 to 8 was derived by adding up the responses to each component, and plotting the response at the fundamental, \( f_{\text{FUND}} \):

\[
(3) \quad R(f_{\text{FUND}}) = \sum_{i=1}^{8} H(f_{\text{FUND}}*i)
\]

The use of non-linear compression or saturation functions and a logarithmic intensity scale would allow the models to achieve a closer fit to the histograms of neuronal response vs. frequency, but would also introduce extra parameters. Since the same points can be demonstrated employing a linear model without the need for additional parameters and assumptions, only simple Gaussian functions are used in these simulations.

RESULTS and DISCUSSION

Fig. 25 illustrates a set of filters without inhibitory sidebands. The filters have identical center frequencies but different bandwidths. The \( Q_{10dB} \)'s of the depicted filters fall between 1 and 6 which are comparable to the magnitudes of \( Q_{10dB} \)'s seen in cortical neurons in the anesthetized cat (Phillips and Irvine, 1981), although sharp tuning is less common in the alert preparation when compared to anesthetized animal. Tuning at lower
auditory stations may be somewhat sharper but of a similar magnitude (MGB, Aitkin and Webster, 1972; IC, Aitkin et al., 1975; AVCN, DCN, Rose et al., 1959).

The filters in Fig. 25 admit energy from complex tones with fundamental frequencies at and below the filter’s center frequency. Not surprisingly, the frequency response in "filter" neurons of Chap. 3 shows similar behavior. At broad filter bandwidths the complex tone response has a single peak, more intense than the maximum pure tone response. The peak occurs at the fundamental frequency for which the greatest number of complex tone components falls within the ERF. The difference in plotted amplitudes of complex tone vs. pure tone responses would be less pronounced on a logarithmic (dB) scale. The eight components have eight times the energy of a single tone, and would only result in a 9 dB increase in response (at 3 dB per doubling of energy). In addition, the rate-level curve of the auditory cortical neurons eventually saturates at high intensities (e.g. Phillips et al., 1985). The firing of cortical cells will not exceed this upper limit and thus the neurons will not necessarily show a much greater response to complex tones than to pure tones.

Narrower filters produce multiple peaks in the complex tone response. As the filter bandwidth decreases, the probability of the filter falling entirely between components of a complex increases. The peaks are produced by frequencies falling within the filter bandwidth. The resolution of components in a complex varies inversely with the filter bandwidth: the narrower the bandwidth, the greater the number of distinguishable peaks (components).
FIG. 25. Responses of a set of Gaussian filters of varying bandwidth to pure tones (right) and complex tones (left). The y-axis represents frequency in equal steps along a logarithmic scale from 84 Hz (at the origin) to 18000 Hz (at the 9th division). For complex tones, the y-axis represents fundamental frequency of the input complex tone. The x-axis represents energy (arbitrary units) admitted by the filter, along a linear scale. Positive values are plotted to the right and negative values are plotted to the left, negative values (not seen in this figure) are truncated at minus two divisions. Five divisions along the x-axis represents the energy of one component. The filter center frequencies, CF (equation 1), are all equal to 4200 Hz. The filter bandwidths, BW (equation 1) and $Q_{10\text{dB}}$'s are as follows:

<table>
<thead>
<tr>
<th>BW (Hz)</th>
<th>$Q_{10\text{dB}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top row: 2000</td>
<td>1.0</td>
</tr>
<tr>
<td>Second row: 1150</td>
<td>1.7</td>
</tr>
<tr>
<td>Third row: 700</td>
<td>2.8</td>
</tr>
<tr>
<td>Fourth row: 500</td>
<td>3.9</td>
</tr>
<tr>
<td>Bottom row: 350</td>
<td>5.6</td>
</tr>
</tbody>
</table>

N.B.: Inhibitory bands have exactly the same form as excitatory bands, except that they are reflected about the y-axis.
PURE EXCITATION

COMPLEX TONES (1-8)  PURE TONES

INTENSITY
As the fundamental frequency of a complex tone is increased, the first component to enter the filter passband will be the eighth component. The plot shows this response at the fundamental frequency of the complex, which would be one eighth of the filter center frequency. The next component to enter would be the seventh, whose fundamental occurs at one seventh of the center frequency. Thus peaks would fall at 1/8, 1/7, ..., and 1/1 of the center frequency in a subharmonic series. The peaks of the complex tone response of resolver neurons occur at subharmonics of the unit's best frequency just as predicted by the narrow filter model.

The effect of a lower inhibitory sideband on the pure and complex tone response is shown in Fig. 26 for a narrow band ERF ($Q_{10} = 3.25$) and in Fig. 27 for a broad band ($Q_{10} = 1.0$) ERF. The effect of an upper inhibitory sideband is seen in Fig. 28. The ERF functions have identical center frequencies and are shown without ISBs in the top row of each figure. A low frequency ISB (second row, Figs. 26, 27) attenuates energy from complex tones with fundamentals lower than the ERF center frequency, thus raising the lower boundary of its complex tone response area. Increasing the bandwidth of the lower ISB (third row, Figs. 26 and 27) further restricts the complex tone response bandwidth. Increasing the relative strength of the lower ISB attenuates responses to all fundamentals except those at the filter's center frequency. The filter's responses now resemble those of the "fundamental" neurons of Chap. 3. The figures demonstrate that such tuning for complex tones requires lower inhibitory sidebands that are broad and/or strong. Such extreme requirements on the ISB may be responsible for the low proportion of units (3%) observed to have fundamental neuron properties.
FIG. 26. Filter responses of an excitatory band (center frequency 4200 Hz, bandwidth 600 Hz, relative strength 1.0) is shown combined with various lower inhibitory sidebands. The axes are as in Fig. 25. The parameters for the ISB’s are the following:

Top row: no inhibitory sidebands

<table>
<thead>
<tr>
<th>CF</th>
<th>BW</th>
<th>K</th>
</tr>
</thead>
<tbody>
<tr>
<td>Second row: 3200 Hz</td>
<td>500 Hz</td>
<td>1.0</td>
</tr>
<tr>
<td>Third row: 2800 Hz</td>
<td>800 Hz</td>
<td>1.0</td>
</tr>
<tr>
<td>Bottom row: 2800 Hz</td>
<td>800 Hz</td>
<td>2.0</td>
</tr>
</tbody>
</table>
LOWER INHIBITORY SIDEBANDS

COMPLEX TONES (1-8)  PURE TONES

LOG F

INTENSITY
FIG. 27. Filter responses of a broad excitatory band (center frequency 4200 Hz, bandwidth 2000 Hz, relative strength 1.0) is shown combined with various lower inhibitory sidebands. The axes are as in Fig. 25. The parameters for the ISB's are the following:

Top row: no inhibitory sidebands

<table>
<thead>
<tr>
<th>CF</th>
<th>BW</th>
<th>K</th>
</tr>
</thead>
<tbody>
<tr>
<td>2600 Hz</td>
<td>1000 Hz</td>
<td>1.0</td>
</tr>
<tr>
<td>1500 Hz</td>
<td>1000 Hz</td>
<td>1.0</td>
</tr>
<tr>
<td>2600 Hz</td>
<td>1000 Hz</td>
<td>2.0</td>
</tr>
</tbody>
</table>
LOWER INHIBITORY SIDEBANDS

COMPLEX TONES (1-8)  PURE TONES

LOG F

INTENSITY
FIG. 28. Filter responses of an excitatory band (center frequency 4200 Hz, bandwidth 600 Hz, relative strength 1.0) is shown combined with various upper inhibitory sidebands. The axes are as in Fig. 25. The parameters for the ISB’s are the following:

- Top row: no inhibitory sidebands

<table>
<thead>
<tr>
<th>Row</th>
<th>CF</th>
<th>BW</th>
<th>K</th>
</tr>
</thead>
<tbody>
<tr>
<td>Second</td>
<td>6000 Hz</td>
<td>1000 Hz</td>
<td>1.0</td>
</tr>
<tr>
<td>Third</td>
<td>5000 Hz</td>
<td>700 Hz</td>
<td>1.0</td>
</tr>
<tr>
<td>Bottom</td>
<td>5000 Hz</td>
<td>600 Hz</td>
<td>1.0</td>
</tr>
</tbody>
</table>
UPPER INHIBITORY SIDEBANDS

COMPLEX TONES (1-8)  PURE TONES

LOG F

INTENSITY
FIG. 29. Filter responses of a broad excitatory band (center frequency 4200 Hz, bandwidth 2000 Hz, relative strength 1.0) are shown combined with various upper and lower inhibitory sidebands. The axes are as in Fig. 25.

Top row: no inhibitory sidebands

Middle row: upper ISB CF = 8500 Hz, BW = 2000 Hz, K = 1.0
          lower ISB CF = 2000 Hz, BW = 1000 Hz, K = 1.0

Bottom row: upper ISB CF = 8500 Hz, BW = 2000 Hz, K = 1.8
           lower ISB CF = 2000 Hz, BW = 1000 Hz, K = 1.8
UPPER AND LOWER
INHIBITORY SIDEBANDS

COMPLEXTONES (1-8)

PURETONES

LOG F

INTENSITY
In the visual system lateral inhibition can cause "contrast enhancement" of images, where the response near sharp contrast edges falling on the retina is augmented (Ratliff, 1968). Could inhibitory sidebands enhance the resolution of peaks in the complex tone response? The evidence from Figs. 26, 27, and 28 indicates that the improvements are only marginal. Figs. 26 and 28 have ERFs with the same center frequency and bandwidth. This bandwidth is comparable to a well tuned pure tone response bandwidth seen in cortical neurons of Chap. 3. By adding either an upper (Fig. 28) or a lower (Fig. 26) inhibitory sideband, the resolution can increase from 3 peaks to 4, which corresponds to a change in component spacing of 33%. This increase only occurs when the ISB is as narrowly tuned as the ERF. A broad ISB (Fig. 26, row 3) causes only the suppression of peaks.

When the excitatory region is broad with a single peak in its complex tone response (Fig. 27), lateral inhibition is not very effective in producing higher resolution. The peaks seen in Fig. 27, row 2, result from the negative image of the ISB seen against the background of the ERF.

The addition of both upper and lower ISB's to a broad excitatory region (Fig. 29, row 2) allows one or perhaps two components to be resolved. Increasing the strength of the ISB's (Fig. 29, row 3) suppresses all response to complex tones, producing a unit selective to narrow band stimuli. Strong inhibition is required before this selectivity is manifested.

Fifty-four percent of unclassified neurons in Chap. 3 showed only weak inhibition, and had responses that were intermediate between those of the filter neuron, fundamental neuron, resolver neuron and the narrow band neuron. Figs. 26, 27, 28, and 29 illustrate that weak inhibitory sidebands can produce quite varied complex tone responses. The cortical units with weak sidebands exhibited the same varied nature of response, and
many unclassified cortical cell responses could also be described by combinations of filters representing ERFs with weak ISB's. A moderate upper inhibitory sideband will restrict the upper boundary of the complex tone response, and lower sideband inhibition raises the lower boundary of the complex tone response.

**Summary:** The behavior of the neuron classes of Chap. 3 were largely accounted for by various configurations of ERF and ISB, specifically:

1. Filter neurons can be simulated with a single pure excitatory receptive field with $Q_{10dB}$'s of less than about 1.5-2.0;

2. Fundamental neurons can be simulated with combination of an excitatory receptive field with a powerful ($K > 1.0$) lower inhibitory sideband;

3. Narrow band neurons can be simulated with an excitatory receptive field combined with powerful upper and lower inhibitory sidebands;

4. Resolver neurons can be simulated with narrow excitatory receptive fields ($Q_{10dB} >$ approx. 1.5-2.0), optionally with weak or narrow upper or lower inhibitory sidebands.

All of these classes may be interpreted as special cases along a continuum of possible receptive field interaction. Intermediate neuron classes, as frequently observed in monkey cortex, are easily simulated by simple parameter adjustments of the same filters which account for the defined classes.

It was seen that the inhibitory sidebands were restricted in their ability to improve resolution of components in a harmonic complex tone. Little improvement can be made in the resolution due to the constraints imposed by the bandwidth of the ERF. Inhibitory sidebands can cause the neuron's response to become selective to features of complex tones, for example, fundamental frequency or bandwidth. The responses become more
selective with strong ISBs. Since all observed response patterns can be modelled with interacting filters, it seems justified to assume that combinations of excitatory response areas and sideband inhibition is entirely responsible for the spectral response patterns to harmonic complexes in the alert monkey's cortex.
CHAPTER 5

THE INTERNAL REPRESENTATION AND PROCESSING
OF HARMONIC COMPLEX TONES IN
PARALLEL DISTRIBUTED PROCESSING NETWORKS

INTRODUCTION

The experiments in Chapter 2 demonstrated that monkeys can analyze complex tones and can perform a pitch matching task with harmonic complexes. The responses of cortical neurons to complex tones were examined in these same monkeys (Chap. 3). Most neurons were activated by complex tone stimuli, but few of the units displayed specific selectivity to harmonic complexes. The response patterns of most units could be explained in terms of linear combinations of pure tone excitatory or inhibitory response areas (Chap. 4). The lack of specificity in the single neuron responses suggests that complex tone analysis might be accomplished by a process which is distributed across the population of cortical neurons.

Recent theoretical advances have produced computational models of arrays of neuron-like units operating in parallel (e.g. Rumelhart and McClelland, 1986). The processing schemes used in such models are termed "parallel distributed processes" (Hinton et al., 1986), and are thought to have corresponding processes in neural systems. For example, in the cerebral cortex, the basic cellular architecture is formed by arrays of columnar functional units with similar input and output connection patterns (Mountcastle, 1957), suggesting some kind of parallel processing scheme.

Parallel distributed processing (PDP) networks process information through the interactions of a large number of simple elements called units, each of which send excitation and inhibition to other units. Units in PDP networks bear strong similarities to neuronal
units; the activity level of each unit is determined by the weighted sum of their excitatory and inhibitory inputs and this activity is passed on to other units via output connections. The strength, or weights, of the inter-unit connections can vary and (in the training phase) changes with "experience". Hebb (1949) put forth a hypothesis that experience can modify the strength of synaptic connections and perhaps form the basis of learned behavior in biological systems. This hypothesis has received strong support from the work of Kandel et al. (1983), Klein and Kandel (1978), and Abrams et al. (1983), which shows a direct correlation between learning and changes in synaptic strength in the invertebrate Aplysia californica. Espinosa and Gerstein (1988) have investigated the behavior of assemblies of cells in cat auditory cortex, and have found changes in the connectivity among the observed neurons for different tonal stimuli.

Could complex tones and pitch be encoded and processed by a distributed representation? Does missing fundamental pitch emerge from a parallel distributed processing (PDP) system which represents harmonic complex tones? These questions are pursued in this chapter using recent implementations of PDP networks (Rumelhart and McClelland, 1986).

METHODS

**Simulating Parallel Distributed Processing Networks**

The following discussion will be confined to networks which use "back-propagation", a phrase which describes the particular learning procedure (Rumelhart et al., 1986a,b). A back-propagation network with three layers of units is illustrated in Fig. 30. There is an input layer, an output layer, and a hidden layer whose units have no direct contact with the outside world. Hidden units can make networks more efficient by reducing the total number of connections required (Jones and Hoskins, 1987). Each input unit is connected with every hidden unit, and each hidden unit is connected to every output unit. The connections of one unit in each layer are shown in Fig. 30.
FIG. 30 The neural network used for both the autoassociation and pitch association training paradigms. The network consists of a layer of input units, a layer of output units and a layer of hidden units in-between. This network has only feed-forward connections; there are no connections between units in the same layer, or feed-back connections towards the input layer. The connections of only one of the input units and one of the hidden units are shown.
Each unit can have an activity level anywhere between 0 and 1. In the case of the units on the input layer, the activity levels are determined by the input stimulus. For the hidden and output layers, the activation level is the weighted sum of all the inputs from the previous layer. The activity in the \( i^{th} \) unit, \( U_i \), is:

\[
(1) \quad U_i = \text{squash}( \sum_{j} W_{ij} \cdot V_j )
\]

where

\[
(2) \quad \text{squash}(x) = (1 + e^{-x})^{-1}
\]

\( V_j \) is the activity of the unit in the previous layer, \( W_{ij} \) is the weight of the connection between the \( i^{th} \) and \( j^{th} \) units. The strengths of connections (weights) can be changed and may have any positive or negative value, though in the present simulations they were normally less than 1.0. The squashing function, \( \text{squash}(x) \), keeps the weighted sum between 0 and 1.

Such networks can be trained. At the start of training, all of the weights of the connections are assigned random values between -1.0 and 1.0. The "training set" consists of pairs of specified input and output patterns. The object of training is to get the network to produce a given (target) output pattern on presentation of a given input pattern. Training is accomplished by iteratively presenting the input, comparing the actual output to the target output to find the error, and altering the strengths of the inter-unit connections. The iterations continue until the magnitude of the error has been reduced to acceptable levels (the "criterion").

The formula for calculating the change of the connection strengths in networks with hidden units is called the "generalized delta-rule" (Stone, 1987; Jones and Hoskins, 1987) and takes the following form:
The weight of the connection between the \(i\)th output unit \((O_i)\) and the \(j\)th hidden unit \((H_j)\) is \(W_{ij}\). The change in weight, \(\Delta W_{ij}\), is proportional to the error between the target output \((O^t_i)\), the actual output \((O_i)\), and the learning rate \((L)\). Equations (3) and (4) are used to change the connection strengths between the output layer and the hidden layer. Another step with similar equations is required to change the weights of the connections between the hidden layer and the input layer. One may see how the term for this algorithm, "backpropagation of error", refers to the way that the error is propagated backward through the network to change the connection strengths (Rumelhart et al., 1986a,b).

Networks are said to associate the inputs with their corresponding outputs. The information representing the associations between inputs and outputs is stored in the matrices of connection strengths (e.g. \(W_{ij}\)). Since this information is distributed over the entire set of connections, the activity levels of a single hidden unit may not necessarily reflect the overall processing of the input.

Autoassociation, also known as identity mapping, is a training paradigm in which the target output is same as the input (Kohonen et al., 1976). When the network must replicate the input at its output, it must develop an internal representation of the stimulus set in the intervening hidden units. The hidden unit representation may then be examined for evidence of feature analysis.

PDP networks may sometimes exhibit the phenomenon of "pattern completion". When a network is presented with a fragment of a previously learned input, the network may still respond as though the entire input had been presented (Kohonen et al., 1981). Dramatic examples of pattern completion in video images of faces have been shown by Kohonen and coworkers (1981). The networks were trained to reproduce at the output the
images of faces supplied as input (autoassociation). After the training phase was concluded, part of one of the original faces was presented to the network. The complete face was reproduced at the output: the pattern was completed. If complex tone pitch is processed in this way, presentation of fragmentary complex tones (missing fundamentals) may still give rise to the appropriate pitch due to pattern completion.

In the following simulation, back-propagation networks were trained with arrays representing the excitation patterns of harmonic complex tones containing their fundamental. The responses of the hidden units were analyzed to characterize the internal representation. Arrays representing missing fundamental complexes and two tone stimuli were applied to test the networks' ability to process novel stimuli and to simulate missing fundamental pitch perception.

Excitation Patterns Used For Inputs and Target Outputs

The inputs and outputs for training and testing the PDP networks were excitation patterns produced by harmonic complex tones as derived from Moore and Glasberg's (1987) model; this paper included a computer program to generate the peripheral excitation patterns from a list of frequencies and amplitudes of discrete spectral components. The effect of the sensory threshold and peripheral filter functions were taken into account by this model. Examples of these excitation patterns can be seen in Fig. 31 for pure and complex tones. Moore and Glasberg used a frequency scale based on the equivalent rectangular bandwidth (Patterson, 1976) of the peripheral filters according to the following equation:

\[
E = 11.17 \times \log\left(\frac{f+0.312}{f+14.675}\right) + 43.0
\]

where \( f \) is frequency in kHz, and \( E \) is the number of equivalent rectangular bandwidths. The program calculates the excitation patterns, producing 270 amplitudes spaced equally on the interval 3.0 to 30.0 \( E \) units (100 to 6500 Hz). The amplitudes have units of dB and are scaled so that the range from 0 to 70 dB is mapped into the interval between 0.0 and 1.0.
(Recall that the units can only have activation levels between 0 and 1.) Input is provided to a network by setting the activity levels of the array of input units equal to the scaled values of an excitation pattern calculated from equation (5). The frequency scale from $E=3.0$ to $E=30.0$ was mapped sequentially onto the array of input units from unit 1 (left most on the scales of the Figs.) to unit 270 (right most). The same relation holds for the excitation patterns of the target outputs and the array of output units. The input activity is then propagated forward using equations (1) and (2), once for each set of connections. The result is an array of 270 output values.

For clarity, the arrays of 270 points which represent the excitation patterns of pure tones and complex tones shall be referred to in italics as pure tones or complex tones. A full or complete complex tone is one which contains a component at its fundamental frequency.

**Two Simulation Paradigms**

Two networks were trained, each composed of 270 input units, 100 hidden units, and 270 output units (Fig. 30). The hidden units reduce the number of total connections from input to output by 35% in this case. There was one "bias unit" at the input level and one at the hidden unit level, which always had an activity level of 1.0. The inter-unit connections of bias units were treated in the same fashion as any other unit's during the training phase. Training involved a large amount of calculation, requiring overnight runs on a VAX computer equipped with an array processor. Testing of the trained networks, on the other hand, could be accomplished on a personal computer (IBM AT compatible), with 40 megabytes of mass storage.

The networks were trained with complete harmonic complex tones with from 1 to 9 components set at 50 dB SPL per component. Fundamental frequencies varied from 120 Hz to 6500 Hz in 66 increments, spaced equally along a logarithmic scale. Complex tones with components over 6500 Hz were not used. The networks were trained until each of the
output points came within 3 dB of the target output points. The two networks were trained with one of the following paradigms:

1) Autoassociation: The full set of complex tones was used for autoassociation training, resulting in 386 input-output pairs. The autoassociation paradigm was used in order to investigate the internal representation of pure and complex tones in the hidden units of a PDP network. In the mammalian cortex, the "hidden units" show frequency tuning in their responses to pure tones (see Chap. 3). Is this property merely a by-product of the topographic projection from the cochlea, or are tuned responses required to represent and process tonal stimuli? Artificial neural networks need no topographic specificity in their connections (e.g. Fig. 30). If tuning develops in the responses of the network’s hidden units, then it might be proposed that tuning is actively produced in order to process complex tones.

Networks trained with the autoassociation paradigm are also particularly suited for tests of pattern completion. A degraded version of one of the inputs from the training set can be presented to the network and the resulting output compared with the original undegraded stimulus. The autoassociative network used in this experiment was trained only with complete harmonic complex tones. Missing fundamental complexes can be regarded as degraded versions of these complete complexes. If the network exhibits pattern completion with missing fundamental complexes, then one expects that the network will respond as if an appropriate complete harmonic complex tone was presented.

2) Pitch association: In this paradigm, complete harmonic complex tones (as inputs) were associated with pure tones at the fundamental frequency of the complex (as outputs). The pure tone represents the pitch of the complex tone. For example, the input: a 5-component harmonic complex tone with a fundamental of 215 Hz; and the output: a 215 Hz pure tone. If the network is provided with the associations between complex tones and their fundamental components, it may generalize this association and recognize the pitches of missing fundamental complex tones. The range of frequencies and fundamentals in the
training set was meant to mimic the range of musical pitch found in humans. *Complex tones* with fundamentals over 1500 Hz were not used, since they were not found by Ritsma (1962) to have a significant residue pitch. This restriction reduces to 319 the number of input-output pairs in the training set for pitch association.

In the autoassociation paradigm one input is paired with one output. In the pitch association paradigm this is no longer true. Each of the nine *complex tones* having the same fundamental is paired with the same output, namely the *pure tone* at the fundamental frequency. Networks which are required to learn this type of association (multiple inputs to one output) must find a common feature in the input set to trigger the correct output pattern. Can back-propagation networks be trained to make pitch associations? If so, what input features do they rely upon to produce their output? How do these networks respond when faced with missing fundamental *complex tones*?

**RESULTS**

*Autoassociation:* Fig. 31 demonstrates that the training phase was successful. Sample input patterns from the training set were presented to the trained network and the output pattern obtained. Comparing the input (left column) and output (right column) patterns, one observes that they agree to within the training criterion of 3 dB. However, when the local peaks on the input pattern are less than 3 dB high, they may be merged in the output pattern (bottom row, Fig. 31).

The frequency response of each hidden unit was obtained by recording the responses of the array of hidden units to a series of 66 *pure tones* over the range of 120 to 6500 Hz. The frequency response of all 100 hidden units was examined. Plots of response versus frequency are shown in Fig. 32 for six hidden units. Units were found showing selectivity to a small band of *pure tone* frequencies (Fig. 32, top row), to multiple ranges of frequency (Fig. 32, middle row), and to a broad band of frequencies (Fig. 32, bottom row). Most units displayed more irregular response curves than are shown in Fig. 32. The arrangement of
maximum response frequency to position in the array was not topographic (i.e., there was
no equivalent of tonotopy). All hidden units had gently curved frequency responses,
exemplified by the units in Fig. 32

The network’s response to missing fundamental complex tones was tested and typical
results are shown in Fig. 33. A series of 3-component harmonic complexes (i.e., with
components 2-4, 3-5, . . . , 7-9) were presented. If the phenomenon of pattern completion
played a role in the responses of the network, the network might return the complete
harmonic complex tone at the output. Instead, as with the training set, the output complex
tone strongly resembled the input complex tone. The number of peaks were similar as were
the high and low frequency borders. However, the overall amplitude of the output pattern
was always less than that of the input pattern. In no instance did the network respond with
a complete harmonic complex tone.

*Pitch Association:* The results of training, seen in Fig. 34, demonstrate the network
was successful in learning to associate full harmonic complex tones with pure tones at their
fundamental frequency. The training took more iterations than that for the autoassociative
network, and a small fraction of output points did not reach the criterion.

Fig. 35 illustrates the typical response of the network to the same 3-component
harmonic complexes used to test the autoassociative network. The pure tone (the pitch)
produced in response to the complexes was at or near the lowest component of the complex
tone. The only instances of a pure tone being produced which was below the lowest
component in the 3-component complex is when the component frequencies were all above
1500 Hz (Fig. 35, bottom right). Since the network was only trained to produce output pure
tones with frequencies less than 1500 Hz, it could not produce pure tones with any higher
frequencies.
FIG. 31 Results of autoassociation training. Three examples of input arrays (left column) used to train the network, which were subsequently presented to the trained network. The corresponding output arrays are shown in the right column. Each array consists of a set of 270 points, plotted along the horizontal scale, whose values range from 0.0 to 1.0 on the vertical scale. The input arrays are the excitation patterns generated by the program of Moore and Glasberg (1987). The amplitude of each frequency component used to generate the excitation patterns was set at 50 dB SPL. The horizontal scale goes from 3.0 (left) to 30.0 E units (right) covering the frequency range from 100 to 6500 Hz (see text for explanation).

Top row: fundamental of 215 Hz, 1 component
Middle row: fundamental of 215 Hz, components 1-3
Bottom row: fundamental of 215 Hz, components 1-8
AUTOASSOCIATION
TRAINING SET

INPUT

OUTPUT

INTENSITY

FREQUENCY (E)

FREQUENCY (E)
FIG. 32 Responses to pure tones of six hidden units in the autoassociative network. Each panel shows the response of a single hidden unit to a series of pure tones. The vertical axis represents the activity level (from 0.0 to 1.0) of the hidden unit in response to the excitation pattern of a pure tone. The horizontal axis represents the frequency of the pure tones along a logarithmic scale from 100 (left most) to 6500 Hz (right most).
HIDDEN UNIT TUNING CURVES

INTENSITY vs. LOG F
FIG. 33 Response of the autoassociative network to novel complex tones missing their fundamentals. See Fig. 31 for axes. The arrays for the 3-component harmonic complex tones missing their fundamentals are plotted as solid lines. The arrays for the output patterns produced by the network are shown as dashed lines.

Top panel: input fundamental, 215 Hz, components 3-5
Middle panel: input fundamental, 215 Hz, components 3-5
Bottom panel: input fundamental, 408 Hz, components 5-7
3-COMPONENT COMPLEXES  MISSING FUNDAMENTAL

INTENSITY

FREQUENCY (E)
FIG. 34  Sample inputs and outputs from a network trained with pitch association paradigm. See Fig. 31 for a description of the axes. The outputs (dashed lines) of the network are shown in response to inputs (solid lines) previously used to train the network.

Top panel:  input fundamental 215 Hz, components 1-2
Middle panel:  input fundamental 215 Hz, components 1-5
Bottom panel:  input fundamental 215 Hz, components 1-8
PITCH ASSOCIATION
TRAINING SET

INTENSITY

FREQUENCY (E)
FIG. 35  Typical responses of the pitch association network to novel complex tones missing their fundamentals. See Fig. 31 for a description of the axes. The excitation patterns of the 3-component harmonic complexes are shown as solid lines, the resulting output patterns are depicted as dashed lines. See Fig. 31 for a description of the axes. The excitation patterns of the 3-component harmonic complexes are shown as solid lines, the resulting output patterns are depicted as dashed lines.

Top left:    input fundamental 120 Hz, components 2-4
Top right:  input fundamental 215 Hz, components 2-4
Bottom left: input fundamental 151 Hz, components 5-7
Bottom right: input fundamental 486 Hz, components 5-7
3-COMPONENT COMPLEXES
MISSING FUNDAMENTAL
FIG. 36 Responses of the pitch association network to novel two-tone complexes. See Fig. 31 for a description of the axes. The excitation patterns of the two-tone complexes are shown as solid lines, the resulting output patterns are depicted as dashed lines.

Frequencies of input tones:

- Top-left: 150, 400 Hz
- Middle-left: 200, 400 Hz
- Bottom-left: 400, 515 Hz
- Top-right: 400, 800 Hz
- Middle-right: 400, 1500 Hz
TWO TONE TESTING

Intensity vs. Frequency (E)

Intensity vs. Frequency (E)

Intensity vs. Frequency (E)

Intensity vs. Frequency (E)
The response of the network was then tested with two tone complexes. Fig. 36 shows a set of two tone complexes with one frequency fixed at 400 Hz. The network is seen to respond with output pure tones at or near the lowest component of the two tone complex. The output pure tone overlaps completely with the lowest component of the input when is the two tones of the input complex are components 1 and 2 of a harmonic series (Fig. 36, middle-left and upper-right). In the majority of cases the network appears to treat the two-tone complex as though it were a single harmonic complex. The network responds with a second pure tone (low amplitude) at the output when the components in the input complex are far apart (Fig. 36, middle-right).

Together, Figs. 35 and 36 indicate that the most important aspect of the input complex determining the response of the pitch association network is the low frequency edge and/or component of the input.

**DISCUSSION**

*Autoassociation:* The autoassociation network provides an example of a distributed representation of full harmonic complex tones. Primary auditory cortical neurons (Phillips and Irvine, 1981) in particular and the lemniscal-line system (Graybiel, 1973) in general, have sharply tuned responses organized in a tonotopic array. Each unit in the artificial networks is connected to every unit of the next higher layer (Fig. 30), so every hidden unit has the same set of connections or (in a topological sense) the same location. Thus there is no topographic array of frequency selectivity in the artificial neural networks of this experiment. Fig. 32 shows that there were few band limited responses for the hidden units, even though the networks had no trouble processing the complex tones. This implies that lack of either tonotopy or sharp tuning curves in a nucleus (or layer of units) does not preclude processing of high resolution frequency information. Several fields in auditory cortex do not display any discernable tonotopic organization.
The behavior of this network in response to novel stimuli is analyzed formally in the Appendix. The arguments are briefly summarized here. The network was trained with and stores, the representations of 386 full harmonic complex tones. When the trained network subsequently receives an input which is a member of the training set, it produces the appropriate output (Fig. 31). When a novel input which is not a member of the training set is presented, there is no single appropriate output. If the new input is similar to a member, K, of the training set, then the new input will partially activate member K's output (see Appendix). The greater the degree of similarity between K and the novel input, the greater the degree of activation. Novel inputs are rarely similar to just one member of the training set, but to many of them. A novel input will activate the set of representations to which it is most similar, each according to the degree of similarity. One might regard this process as a Fourier-like analysis, where the input function is expanded in terms of the 386 complex tone functions instead of pure sine functions. The resultant pattern at the output of the network resembled the input pattern very closely and there were no signs of pattern completion. From the results of the testing with 3-component stimuli, it must be concluded that missing fundamental inputs are more similar to a combination of complex tones than they are to the completed harmonic series. The conclusion is that missing fundamental inputs cannot cause the corresponding full harmonic series to be recalled from the network.

*Pitch association:* In this chapter, pitch processing is treated as though it were a learned association. This approach is not novel, and has been suggested in the form of a "learning matrix" by Terhardt (1972). He hypothesized that the foundations of pitch
perception (including missing fundamental pitch) are laid down early in life via experience with harmonically structured sounds. The hypothesis that learning affects pitch perception has also been strengthened by investigations by Divenyi (1979), and by Hall and Peters (1982).

The complex tone patterns for the pitch association paradigm were chosen to cover the range of complex tone pitch which in humans gives rise to periodicity pitch, according to Ritsma (1962). These patterns and associations were successfully incorporated into the neural network by the back-propagation algorithm. This success demonstrates that the associations necessary for pitch perception, namely the harmonic complex tone with its fundamental frequency, could be formed within a distributed representation.

When this network was subsequently tested with two tone complexes or missing fundamental complexes, the resultant output was a single peak which fell at the low frequency edge of the input tone. This indicated that the most important feature of the input for making pitch matches was the low frequency edge, or lowest component of the input pattern. The use of this strategy by the network is understandable since low frequency edges of the set of complex tones with the same fundamental was their most constant common feature. This behavior is in accordance with an extrapolation of "place" theory (originated by von Helmholz, 1863) which would equate the pitch of a complex tone with its lowest component. However, experienced humans listening to comparable complex tones hear the pitch at the missing fundamental. Could networks be made to produce the same responses using paradigms different from those employed in this study?

If pitch is processed by means of previously learned spectral patterns of harmonic complex tones, as Terhardt (1972) suggests, then missing fundamental stimuli may be explicitly associated with the pitch of the appropriate fundamental. Most psychophysical experiments on humans use subjects with extensive experience in listening to missing fundamental pitches (Ritsma, 1962; Houtsma and Goldstein, 1972; Plomp, 1965). Based
on the demonstrated performance of neural networks, it is likely that one could also include missing fundamental complexes in their training set, and obtain results comparable to human pitch perception directly.

Only spectral information was used as input to the networks tested here. In principle, temporal information from models using autocorrelation or interval histograms on the sound signal (Langner, 1983; Raatgever and Bilsen, 1986; Srulovicz and Goldstein 1983) would also serve as input to a network. The similarities of input patterns with a periodic temporal structure may be more conducive to providing a basis for missing fundamental pitch extraction by PDP networks.

**Summary:** In Chap. 3 the hypothesis that pitch and complex tones may be represented in the auditory cortex by means of the firing patterns of single cells selective for pitch was investigated. No evidence for such a coding scheme was found, although it was impossible to rule out entirely. In this chapter the qualities of distributed representations as applied to complex tone processing were tested. Although significant similarities in the behavior of cortical neurons and units in distributed representations exist, such representations are not sufficient to account for cortical behavior alone.
CHAPTER 6

SUMMARY AND CONCLUSIONS

In the foregoing chapters, we have seen how harmonic complex tones are processed in a non-human primate, the rhesus monkey. Investigations were made at the behavioral level, and the single cell level in alert auditory cortex. Mathematical simulations were made of the single cell responses, and the responses of arrays of neurons to complex tones. The following list briefly summarizes the results of the previous chapters:

1. Rhesus monkeys can perform pitch matching tasks with harmonic complex tones with and without the fundamental and other low components.

2. The cortex of the alert rhesus monkey contains neurons which respond differentially to complex tones. Responses varied over a continuum of excitatory and inhibitory response types, although several basic classes were described:

   Filter neurons typically had no inhibitory sidebands and responded well when any component of a complex tone entered its pure tone receptive field. They may provide a coarse representation of the cochlear activation pattern.

   Resolver neurons had a higher frequency resolving power than filter neurons and responded well when any component of a complex tone entered its pure tone receptive field. Unlike filter neurons they could respond discretely to components of harmonic complex tones.

   Fundamental neurons exhibited similar tuning to pure tones and corresponding fundamentals of complex tones. When a complex tone series without its fundamental was presented, the neurons did not
respond, except when the physically present components entered its pure tone receptive field.

*Wide band neurons* responded more powerfully to wide band noise and complex tones than to narrow band stimuli.

*Narrow band neurons* responded more powerfully to narrow band than to wide band stimuli.

The majority of neurons had responses which combined characteristics of two or more of these classes.

3. The responses of cortical neurons could be accounted for by simple linear combinations of excitatory and inhibitory receptive fields. Inhibitory sidebands are highly effective in introducing response selectivity (feature detection), but not very efficient in extracting high resolution output from low resolution input.

4. Neural network simulations showed that excitation patterns of complex tones could be stored and retrieved from a distributed representation. Hidden units in this representation did not exhibit either band limited responses or a tonotopic arrangement, although various degrees of frequency selectivity were displayed by them. The networks did not complete the pattern of missing fundamental complexes by filling in the missing fundamental.

5. A neural network was trained to associate complex tones with pure tones at the lowest (fundamental) frequency component. The main feature of the input used by the networks to make the associations was the low frequency edge of the input excitation pattern. This was demonstrated for two-tone excitation patterns and for harmonic complex tone (missing fundamental) excitation patterns.
How do these results explain complex tone processing in the rhesus monkey? In Chap. 4 the responses of cortical cells to complex tones were modelled with linear combinations of excitatory and inhibitory frequency response areas. A comparison might be made between simple cells in visual cortex and the classes of auditory neurons discussed here (excepting wide band neurons). The excitatory receptive fields of simple cells are flanked by either one or two lateral inhibitory fields on a 2-dimensional sensory epithelium (Hubel and Wiesel, 1962). Kulikowski and Bishop (1981) analyzed the responses of simple cells in cat visual cortex to spots of light and stationary sinusoidal gratings. The responses to the gratings indicated that the majority of simple cells sum inputs from different parts of their receptive fields linearly. The success of the linear filter models in Chap. 4 suggests that the same assumption holds for the responses of auditory cortical neurons.

Mammalian cortex contains examples of neurons which seem to process a higher level of feature than just spots of light or pure tones. The inferior temporal cortex of the macaque contains a subpopulation of cells which seem to be selective to faces or features of faces (Perrett et al., 1982; Desimone et al., 1984). The responses are dependent on the configuration of specific features and the selectivity is maintained over changes in stimulus size and position. Some cells are selectively responsive to particular features such as eyes or hair (Perrett et al., 1987). In the auditory cortex of the mustached bat (Pteronotus parnellii rubiginosus) there are specific fields specialized for processing features of the bat’s echolocation call. Some cortical cells are sensitive to the delay between the call and its echo (O’Neill and Suga, 1979; Kawasaki et al., 1988), others to sounds having particular amplitudes (Suga and Manabe, 1982) and still others to the doppler-shift between the call and its echo (Suga and Jen, 1977). Even the internal frequency representation of each individual mustached bat is matched to the preferred frequency of orientation call that it emits (Suga et al., 1987).
In each of these cases the processed features are important for the animal's survival and well being. The mustached bat uses echolocation in its hunting and prey catching behavior, as well as for general orientation (Neuweiler, 1983). Facial expressions are used by rhesus monkeys to visually communicate aggression (Hinde and Rowell, 1962). This helps determine the monkey's status in the troop dominance hierarchy, which in turn affects its reproductive success and access to food. These imperatives justify the expense of having a specialized region of brain, or a subpopulation of cells, dedicated to the processing a separate set of stimulus features. More general but less precise processing schemes may be employed for less important stimuli.

The behavioral significance of pitch (and missing fundamental pitch) has not yet been ascertained in lower primates. Many macaque vocalizations are tonal (Green, 1975; Rowell and Hinde, 1962; Gouzoules et al., 1984) and consequently have the quality of pitch. Monkeys can discriminate between the coos of Japanese macaques (Macaca fuscata) based only on the starting pitch (Zoloth et al., 1979). Rhesus monkeys can discriminate between the pitches of pure tones near 2 kHz (Sinnott et al., 1987) but had frequency discrimination limens 5 to 20 times poorer than humans. The psychophysical tuning curves, which indicates the bandwidth of the peripheral filters, are as sharp in rhesus monkeys as in humans (Serafin et al., 1982). The subjects in Chap. 2 demonstrated an ability to perceive the pitch of the missing fundamental. The monkeys, however, took many training sessions to learn the task. Humans faced with the same task can readily make the required discrimination (Tomlinson, Schwarz, unpublished data). Taken together, these findings indicate that monkeys can perceive pitch, but it may be less salient to monkeys than to humans. As a result, the neuronal processing of pitch may utilize more general and basic processing schemes than is the case for bat echolocation calls or monkey face perception.
The cortical recording experiments of Chap. 3 revealed that although the basic features of complex tones were represented (components, fundamental frequencies, etc.) representations of the higher level of feature of missing fundamental pitch was not found. The fact remains however, that monkeys can perform pitch matching tasks, thus some representation of pitch must exist. Although a large number of neurons was examined within the superior temporal plane (STP), the remainder of cortex, and indeed the remainder of neurons within the STP were not investigated. The existence of feature detector neurons selective for pitch cannot be ruled out for this reason. Previous efforts at finding a topographically ordered selectivity to periodicity (extending into periods below the range of musical pitch) have been only marginally successful (de Ribaupierre et al., 1972; Schreiner and Urbas, 1986; Hose et al., 1987). An alternative to the existence of feature detectors for pitch formed the basis of the neural simulations of Chap. 5. The principle was demonstrated that the representation of complex tones and pitch may be distributed over the whole population of units with little evidence at the hidden unit level. If one equates cortical neurons with hidden units, then it is easy to see how a representation for pitch may be spread across the cortical population. However, the form of the actual representation of pitch in the cortex remains undetermined and is a question that merits further exploration.

FUTURE STRATEGIES AND QUESTIONS.

Complex tone perception in monkeys: Early psychophysical experiments in monkeys have examined perception of simple pure tone stimuli (Stebbins et al., 1966). Recently, more refined measurements of similar kinds of stimuli are being made. For example, pure tone thresholds (Owren et al., 1988), psychophysical tuning curves (Serafin et al., 1982) and frequency/intensity discrimination limens (Sinnott et al., 1985,1987) have been measured. Complex sound perception has been tested using recorded vocalizations or recorded variants of a single call type (Gouzoules et al., 1984; Symmes and Newman,
1974), or with synthetic vocal stimuli (Snowdon and Pola, 1978; Sinnott and Kreiter, 1986; Moody and Stebbins, 1987).

Despite this growing body of psychophysical data, there is little information on the perception of stimuli of intermediate complexity, harmonic complex tones. It was found in the neurophysiological recording experiments in Chap. 3, that in the population of highest frequency resolving power (the resolver neurons), resolution of components was rare above the third harmonic. This might suggest that the rhesus monkey can only hear out partials up to the third component in a harmonic complex tone. However, there are no psychophysical data to compare it to. The human behavioral data are not sufficient, since comparative studies indicate that there may be some differences between humans and macaques (Sinnott and Kreiter, 1986; Sinnott et al., 1985, 1987). If pitch is less "important" or prominent for monkeys, they might also have poorer pitch-discrimination limens. Basic quantitative work in simian complex tone perception remains to be done.

Temporal processing of complex sound stimuli: The temporal structure of the stimulus is important for several theories of pitch perception (Raatgever and Bilsen, 1986; Srulovicz and Goldstein, 1983). Temporal coding of sound in the auditory nerve is also thought to be important for speech perception (Sachs et al., 1982). In the cortex the fine time structure of the peripheral signal is present only in degraded form, however the cortex may be the beneficiary of temporal processing performed at lower levels.

Temporal processing, in the form of autocorrelation mechanisms, have been conjectured by Licklider (1951) and Raatgever and Bilsen (1986) and may occur at some site in the brainstem. Timing information from the auditory nerve is well preserved in the brainstem and is even enhanced in certain cells in the anteroventral cochlear nucleus (Rhode and Smith, 1986). The postulated mechanism for autocorrelation involves delaying input waveform by a time, T, and feeding it back to the current waveform at a coincidence detector (unit C) which fires when its two inputs peak at the same time. If a waveform
repeats itself every period T, then unit C will detect coincidences in its inputs. Thus, unit C becomes a periodicity detector for period T. Such a pitch processing mechanism requires a set of neural delay lines and coincidence detectors. Neural delay lines have been found in the barn owl (Sullivan and Konoshi, 1986) and the cat (Bojanowski et al., 1988), but the existence of arrays of such lines, with appropriate delays, has not been confirmed. Evidence of temporal processing of amplitude modulated complex tones has been reported in the brainstem of the guinea fowl (Langner, 1983). The proposed scheme compares the timing of peaks in the modulation envelope and peaks in the fine structure of the waveform with coincidence detectors and other elements to determine the periodicity. The existence of the tonotopic projection system might lead one to neglect other domains of auditory information analysis. The importance of temporal processing in speech coding (Sachs et al., 1982) and pitch analysis is becoming better understood as details of mechanism are worked out.

*Internal representation of complex sound in cortex:* The hypothesis that the representation for pitch is distributed over the whole array of cortex is difficult to verify, especially with the single unit recording technique. The hypothesis of feature detection is, by comparison, straightforward to test. Nevertheless, there are enough difficulties and problems to make this a difficult task in itself. How can one maximize the possibility of finding high-level feature detecting neurons? One approach is to select a simple, naturally occurring stimulus, which has some significance to the animal. One must consider all of the sounds in an animal's acoustic environment and identify the most important. The sound should be heard or employed frequently by the animal. Chap. 3 demonstrated another more analytical approach which employs simple signals, like sinusoids and white noise, to test the responses of the cells. Employing such signals allows the use of tools developed by engineers for systems analysis for the understanding of neural processing. Successful application of these techniques can reveal not only the existence of feature detection, but the mechanisms responsible for the selectivity.
REFERENCES


Abeles M, and Goldstein MH. (1972) "Response of single units in the primary auditory cortex of the cat to tones and to tone pairs". Brain Res. 42:337-352.


Aitkin LM. (1976) "Tonotopic organization at higher levels of the auditory pathway" in International review of physiology, Neurophysiology II, vol. 10. (R Porter, ed) University Park Press, Baltimore. pp 249-279


Bilsen FA. (1966) "Repetition pitch: Monaural interaction of sound with the repetition of the same but phase shifted sound", Acustica 17:295-300.


Delattre PC, Liberman AM, Cooper FS, and Gerstman LJ. (1952) "An experimental study of the acoustic determinants of vowel colour". Word 8:195


Evans EF. (1978) "Place and time coding of frequency in the peripheral auditory system: some physiological pros and cons". Audiol. 17:369-420.


Graybiel AM. (1973) "The thalamo-cortical projection of the so-called posterior nuclear group: a study with anterograde degeneration methods in the cat". Brain Res. 49:229-244.


Hebb, DO. "The organization of behavior". New York, Wiley.


Hinde RA, and Rowell TE. (1962) "Communication by postures and facial expressions in the rhesus monkey (Macaca mulatta)


Hubel DH, and Wiesel TN. (1963) "Shape and arrangement of columns in cat's striate cortex". J Physiol. (Lond.) 165:559-568.


Imig TJ, and Adrian HO. (1977) "Binaural columns in the primary field (A1) of cat auditory cortex". Brain Res. 138:241-257.


Itoh K, and Mizumo N. (1980) "Direct projections from the mesodiencephalic areas to the pericruciate cortex in the cat: and experimental study with the horseradish peroxidase method". Brain Res. 116:492-497.


Mendelson JR, and Cynader MS. (1985) "Sensitivity of cat primary auditory cortex (AI) neurons to the direction and rate of frequency modulation". Brain Res. 327-331-335.

Merzenich MM, and Brugge JF. (1973) "Representation of the cochlear partition on the superior temporal plane of the macaque monkey". Brain Res. 50:275-296.


Wegener, JG. (1964) "Auditory discrimination behavior of normal monkeys". J. Auditory Res. 4:81-106.


APPENDIX: The Response of a Network, Which Has Been Trained With an
Autoassociation Paradigm, To Novel Inputs.

A trained neural network can be regarded as a mapping of a set of \( N \) inputs onto \( N \) outputs,

\[ I_k \rightarrow O_k \]

where \( I_k \) and \( O_k \) are the \( k \)th input and output elements of the training set. For autoassociative networks \( I_k = O_k \). Each of the 386 inputs and outputs used for autoassociation in Chap. 5 is an array of 270 elements, and can be represented as a state vector in a 270-dimensional state space. Mapping the inputs to the outputs proceeds in two stages, corresponding to: (1) mapping the input vector \( I_k \) from the input layer to the hidden layer; (2) mapping the resultant hidden vector \( H_k \) to the output layer to produce the output vector \( O_k \). The mapping from layer to layer is performed by matrix multiplication (Jordan, 1986), followed by application of the squashing function to each unit's summed activity.

Stage 1:

\[ H_k = \text{squash}(I_k \cdot W_{IH}) \]

where

\[ \text{squash}(x) = \frac{1}{1 + e^{-x}} \]

Stage 2:

\[ O_k = \text{squash}(H_k \cdot W_{HO}) \]

\( W_{IH} \) and \( W_{HO} \) are matrices which hold the connection strengths of the input-to-hidden connections and the hidden-to-output connections respectively.

What output results when one presents a novel input to a network (i.e. an input which is not a member of the training set)? Suppose that a novel input, \( I_{\text{NOVEL}} \), is
presented to this trained network. If the N input vectors of the training set "span" the state space for all possible inputs, then by definition $I_{\text{NOVEL}}$ may be represented as a linear combination of elements of the training set $I_k$,

$$I_{\text{NOVEL}} = \sum_{k}^{N} a_k I_k$$

where $a_k$ are the coefficients of the linear combination. Since the number of state vectors in the training set ($N=386$) exceeds the number of dimensions of the vector space (270), not all input vectors will be linearly independent, and so the set of values for $a_k$ does not have a unique solution.

Let the squashing function be linear for the present example (i.e. of the form $\text{squash}(x) = mx + b$, where $m$ and $b$ are constants). This allows one to represent the entire mapping from input to output as a linear transform, $T$. The output vector, $O_{\text{NOVEL}}$, is equal to the transformed input vector,

$$O_{\text{NOVEL}} = T(I_{\text{NOVEL}})$$

expanding $I_{\text{NOVEL}}$ in terms of the set of $N$ input vectors,

$$= T(\sum_{k}^{N} a_k I_k)$$

Since $T$ is linear one may perform the operations on each input vector individually,
Since $O_k = T(I_k)$, one may substitute in for $O_k$:

\[
\sum_{k=1}^{N} a_k O_k
\]

Since this is autoassociation, $O_k = I_k$. Thus,

\[
\sum_{k=1}^{N} a_k I_k = I_{\text{NOVEL}}
\]

This demonstrates that if the novel input is a linear combination of the set of input vectors, $I_k$, then one may predict that the output from a novel input, $O_{\text{NOVEL}}$, will equal the novel input $I_{\text{NOVEL}}$ even though $I_{\text{NOVEL}}$ is not a member of the training set.

Initially the squashing function was assumed to be linear. Is this a reasonable supposition to make? When one performs a Taylor's series expansion on $\text{squash}(x)$, the first three terms are:

\[
\text{squash}(x) = \frac{1}{2} + \frac{x}{4} - \frac{x^3}{48} + \ldots
\]

For values of $x$ less than unity, the cubic and higher terms are negligible. In the interval $[-1, 1]$ this function is well approximated by a straight line with a positive slope of 0.25 and an intercept of 0.5.

The autoassociative network in Chap. 5 behaves in accordance with this demonstration; novel inputs are almost exactly equal to the resultant novel outputs (see Fig. 32, Chap. 5). Thus, the two assumptions made in the course of this demonstration are largely applicable: the set of $N=386$ input vectors in the actual training set, $I_k$,
spanned the input vector space; and the range of $x$ seen by the squashing function produced linearly related outputs.

"Pattern completion" is a phenomenon in memory where, after storing representations of a set of objects in memory, partial cues from one object (input) is sufficient to recall the entire object (output). Pattern completion implies a situation in our analysis where $I_{\text{NOVEL}}$ does not equal $O_{\text{NOVEL}}$. In the formal notation, pattern completion can be represented as the following type of mapping (for autoassociation):

$$I_{\text{NOVEL}} \implies O_k = I_k$$

where $I_k$ is a single member of the original training set. For pattern completion to occur in a linear system, $I_{\text{NOVEL}}$ must resemble $I_k$ more strongly than any other input vector. The likelihood of this grows as the number of elements, $N$, in the training set decreases. In the extreme case of one element in the training set, all novel inputs will result in scaled versions of $I_k$.

To summarize, the autoassociative network in Chap. 5 behaved in accordance with the analysis in the first part of this appendix. The training set was sufficiently large to span the input vector space and the squashing function behaved, for the most part, as though it were linear. Pattern completion was not observed in the neural networks of Chap. 5, but may be observable if the number of elements in the training set is reduced substantially.
R.W. Ward Tomlinson

Publications


9. VW Yong, M Guttman, SU Kim, DB Calne, I Turnbull, K Watabe, S Barwick, RWW Tomlinson, WRW Martin, E Walsh, BD Pate (1987) "Implantation of human fetal sympathetic neurons into hemiparkinsonian cynomologous monkeys", Schmitt Neurological Sciences Symposium on Transplantion into the Mammalian CNS, Rochester, N.Y.

